INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.
2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.
3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.
4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.
5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.

University
Microfilms
International
300 N. Zeeb Road
Ann Arbor, MI 48106

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

8303735

Spencer, JoEllen Page

A COMPARISON OF THE EFFECTS OF ANALYSIS TECHNIQUES AND COMPUTER SYSTEMS IN REMOTE SENSING TECHNOLOGY AND A REFERENCE DATA COLLECTION TECHNIQUE

University of Alaska Ph.D. 1981

University
Microfilms
International 300 N. Zeeb Road, Ann Arbor, MI 48106

A COMPARISON OF THE EFFECTS OF ANALYSIS TECHNIQUES AND COMPUTER SYSTEMS IN REMOTE SENSING TECHNOLOGY AND A REFERENCE DATA COLLECTION TECHNIQUE

A

THESIS

Presented to the Faculty of the
University of Alaska in partial fulfillment
of the requirements
for the Degree of

DOCTOR OF PHILOSOPHY

By
JoEllen Page Spencer, B.S. Biology; M.A. Plant Ecology
Fairbanks, Alaska
December, 1981

A COMPARISON OF THE EFFECTS OF ANALYSIS TECHNIQUES
AND COMPUTER SYSTEMS IN REMOTE SENSING TECHNOLOGY
AND A REFERENCE DATA COLLECTION TECHNIQUE

RECOMMENDED:

Bonita J. Nieland
Major Professor, Advisory Committee

Charles W. Slaughter

Paula V Krebs

Lidia L Selkregg

P. Jan Cannon

John D. Fox

Bonita J. Nieland
Dean - School of Agriculture and
Land Resource Management

APPROVED: K. B. Mather
Vice Chancellor for Research and
Advanced Study

December 9, 1981
Date

ABSTRACT

A technique for collecting and recording reference data which considers the spectral and spatial characteristics of Landsat data, the computer system being used, and the gradient nature of wildland vegetation was developed and described. Different analysis techniques for four critical factors affecting the accuracy of computer-aided analysis products were evaluated.

Comparisons were made on the basis of accuracy evaluations of two methods of data/analyst interface, three methods of deriving training statistics, three methods of spectral class descriptions, and two levels of map category detail. The primary data set used was digital Landsat multispectral data for a study area around Fairbanks, Alaska. Reference data were developed from field work and photo-interpretation. The training methods compared were supervised, unsupervised, and modified clustering. The three spectral class description methods were: 1) labels derived from the training data; 2) the color display screen; and 3) from ground plot data. Community level cover types were compared with generalized map categories. The effect of post-classification stratification was evaluated.

The reference data technique provides geographically located stands and cover types identifications with a flexible coding system that can be aggregated to correspond to the spectral data categories. No difference in classification accuracy was found for an experienced analyst using a printout oriented system such as EDITOR or a screen oriented system such as IDIMS. The modified cluster method of developing training statistics was more effective and efficient than supervised or unsupervised training methods. The use of ground plot data and subsequent stratification improved the descriptions of spectral classes. Generalized mapping categories were more accurate than detailed mapping categories. Knowledge of the ecologic, floristic, and spectral characteristics of the cover types in the study area is necessary to develop spectral class descriptions and stratification criteria.

TABLE OF CONTENTS

TABLE OF CONTENTS
(cont'd)

TABLE OF CONTENTS
(cont'd)

Appendix

LIST OF FIGURES

LIST OF FIGURES
(cont'd)

LIST OF TABLES

LIST OF TABLES
(cont'd)

ACKNOWLEDGMENTS

My deepest appreciation is extended to Dr. Bonita J. Neiland and Dr. Paula V. Krebs. Drs. Neiland and Krebs have been my advisors, teachers and friends for many years, always encouraging, questioning, and pushing me. Each has contributed in her unique way to my finding and developing my professional career and their influence will guide me as I continue to learn. Thanks are due Tom George whose constant discussions, suggestions, and arguments played a large part in this study.

This research was unfunded and depended on the contributions of expertise, computer time, facilities and equipment by many people. Dr. Dale Lumb, NASA WRAP, and Len Gaydos, USGS Geography Program, made the IDIMS and EDITOR computer systems and the facilities of the Ames Research Center available to me for a two month internship at Ames. Richard Johnson, the late Randy Ennis, and Leslie Morrissey taught me how to use IDIMS and EDITOR and helped me finish the classifications used in this study. Nancy Larson and Holly Kerr taught me to use the University of Alaska computer network for the analysis of spectral class descriptions and the accuracy evaluation. Ray Thomas allowed me to use the aerial photographs and interpretation equipment of the BLM/Branch of Photogrammetry. Mike Fleming and Pat Fox assisted with the statistical aspects of this study. Nancy Crozer did much of the drafting of the figures.

I appreciate the time and expertise given me by my graduate committee at the University of Alaska: Drs. Bonita Neiland (major professor), Paula Krebs, Jan Cannon, Lidia Selkregg, John Fox, and Charles Slaughter. And finally, my love and appreciation to my parents, Eloise and Dave Spencer, who began teaching me a love of the land and ecology before my formal education started, and who have always encouraged me toward my personal and professional goals.

CHAPTER I

INTRODUCTION AND BACKGROUND

The science of acquiring and interpreting data from remote sensing devices has come a long way since the first aerial photograph of Paris was taken from a balloon in 1857 and cameras were strapped to carrier pigeons. Sensors mounted on spacecraft have passed Saturn and are headed toward Pluto before leaving this solar system. Computers are used to analyze the data from these sensors to show images of places currently beyond the scope of human eyes. In a similar fashion, sensors on unmanned spacecraft orbit the earth and telemeter data to receiving stations. Computer-aided analysis of these data give humans a perspective from altitudes not available with balloons and pigeons. Interpretations of information from computer-aided analysis and other analysis methods of remote sensor data provide tools for inventorying and managing natural resources, monitoring land use changes and weather patterns, identifying natural hazards, locating mineralized zones, and mapping natural and human-created features on the surface of the earth.

Prior to the launch of Landsat I in July 1972, only small quantities of multispectral digital data were available. Early development of computer-aided analysis methods was done using digital data from airborne multispectral scanners. With the launch of Landsat I, large quantities of digital multispectral scanner (MSS) data became available. Rapid progress has been made in the development of computer-aided analysis systems to interpret MSS digital data. Concurrent with Landsat there have been several legislative mandates calling for resource inventories over large areas in a short time frame (Krebs, 1976). In the rush to process and utilize Landsat data for inventories, few careful evaluations or comparisons have been made of the different computer-aided analysis systems. Much of the remote sensing literature is based on classification results presented as a thematic map for a specific area. These maps are usually interpreted for a specific application. Occasionally a quantitative evaluation is made of the classification performance. In the rush to make computer-aided classification of Landsat data "operational", few researchers have made comparisons of the tools used for classification.

Many land managers use landcover data combined with other types of resource data. Information interpreted from these data sources is used to define and solve resource management problems. Landcover maps are one of the primary output products of computer-aided analysis of Landsat digital data. These landcover maps are combined with other information sources for solving many resource planning and management problems. Some of the common uses of landcover maps are for detection of wildlife habitat, range for domestic livestock, wildfire fuel types, agricultural crop diseases and crop yields, flood hazard mapping, soils, monitoring and planning urban development, road and transmission line corridor locations, watershed planning, snowpack monitoring, forestry, and basic studies involving vegetation, hydrology, or geomorphology.

The validity of interpretations from remote sensor data to provide landcover maps are dependent on several factors. Remote sensor data are a record of spectral information reflected or emitted from an object or landscape. The quality of the remote sensor data is one factor which affects the accuracy of the final interpretations. The second major factor affecting the final results of remote sensing interpretations are the characteristics of the landcover types in the study area and the analysts' knowledge of these types. The analyst should have an ecological perspective of the study area. This provides an understanding of the cover types present in the area, their location, distribution, and pattern of occurrence with other cover types, the environmental factors which influence their distribution, the successional patterns, the disturbance history and effects, and the composition of the various communities. The analyst should have a background in the discipline of the features being mapped as well as knowledge of remote sensing. The procedure for collection of field data for a remote sensing project should consider the characteristics of the remote sensor data and the computer-aided system to be used.

The third important factor is how the analyst uses the tools and techniques available to him to integrate his knowledge of the spectral characteristics of the landcover types with his ecological perspective of the study area. There are many steps during a computer-aided analysis process where this integration critically impacts the accuracy of the final landcover products.

The study reported here compares the use of several different tools and techniques of computer-aided analysis in the environment of the boreal forest of interior Alaska. The results from the interpretations of digital Landsat data are compared on the basis of the final accuracies of the landcover maps.

1.1 Objectives and Justification

There are four main objectives of this study:

1. Development of a technique for collecting field data and coding the vegetation stands. This system must map stands of vegetation and landcover types and describe the vegetation with a flexible label. This system should provide a flexible set of reference data which is not tied to a static vegetation framework. The location and size of the field plots should consider the characteristics of the Landsat data and the computer-aided analysis system.
2. Comparison of the IDIMS and EDITOR digital classification systems. These systems have different approaches for the analyst/data interactions. The IDIMS system uses a color display screen for the primary analyst/data interface. EDITOR uses printouts of the data and digitizing from maps for the analyst to interact with the data.
3. Comparison of three methods of training set selection and manipulation. These three methods were supervised, unsupervised, and modified cluster.

4. Comparison of three different methods of describing the spectral classes of a classification. The three methods investigated were training labels, identification from the color display screen, and ground plot descriptions.

Quantitative evaluations were made of the classification performance for each method. Evaluations of the various methods were made by comparing the classification accuracies.

The use of remote sensor data does not preclude the need for field work and the use of other reference data such as aerial photographs or topographic maps. Results from a computer-aided analysis can be used to make the field work for applications more efficient and effective. A landcover map can separate forested from non-forested lands so that a forestor need only conduct timber surveys on forested lands. Detailed productivity data for range can be collected on potential rangelands. However, field work is necessary during the computer-aided analysis before landcover maps are completed.

Field work for a computer-aided anlysis project must address the characteristics of the Landsat data and the computer-aided analysis system. The Landsat sensor averages reflectance data over a pixel slightly larger than one acre. Field data for digital analysis of Landsat data should cover areas larger than one acre, preferably larger than five acres. It is very difficult to precisely locate one pixel in the field using the 1:63,360 or smaller scale maps which cover most of Alaska. Detailed data from one meter square plots or even one quarter acre plots are not suitable for computer-aided analysis of Landsat data. These detailed data are valuable for post-classification interpretations and applications.

In addition to considerations of the areal extent of a Landsat pixel, a flexible coding system is necessary for recording stand composition. Map category composition and crown closure will vary for different classifications or even different sets of spectral statistics for the same set of Landsat data. The field data can then be merged to fit the spectral classes instead of committing the analyst to a static set of map categories which may not exist in the study area vegetation or the Landsat data.

Many of the techniques used for field work can also be applied to photo-interpretation of aerial photographs. During a computer-aided analysis, a sample of training data is selected from the entire Landsat data set to "train" the computer to classify the whole study area. These training data should represent the range of spectral variability present in the data set. Similarly, the field effort should expose the analyst to the entire range of landcover types in the study area.

Computer-aided analysis techniques are used for the classification of digital spectral data. There are several steps in the analysis procedure in which the analyst/data interactions critically impact the quality of the final classification. These are: 1) geometric correction, 2) development of training statistics, 3) statistics editing, 4) spectral class descriptions, and 5) accuracy evaluation. At each of these steps the analyst makes decisions about the data or the statistics using his experience and discipline expertise.

The format of the data and representation of the statistics affect the analyst's decisions. There are two major approaches for the analyst/data interactions (Phillips and Swain, 1978). One is through a color display screen used by systems such as IDIMS and Image 100. The other method is through line printer products used by EDITOR and LARSYS analysis systems.

It is very difficult to overlay reference data such as aerial photographs, topographic maps, and other map-based data such as landcover data from field work directly onto images displayed on the color display screen (Fleming and Hoffer, 1977). Location of control points, training fields of known cover type stands, and other features tend to be based on the spectral data instead of the exact locations on maps. Printout products can be directly registered with reference maps or overlaid using a zoom transfer scope. This makes feature location more accurate with respect to the base maps used. The EDITOR and IDIMS systems were compared to determine if the approximate location of features affects the final classification accuracy.

A sample of the MSS digital data is used to train the classifier in the computer system. The training statistics calculated from this sample are used by the classifier to process the entire data set. Each pixel is assigned to the training cluster which it most closely resembles (Section 1.4.).

There are two major methods of training set selection for digital classification: supervised and unsupervised. In addition, there are several methods of training set selection which combine both approaches. When using supervised training the analyst has a set of reference data from which to draw the training fields. The training fields are delineated in the digital data set and identified as to landcover type before clustering. Training data for the unsupervised method are selected by sampling the digital data and clustering without knowing the information descriptions of the spectral classes. Various combinations of these techniques have been developed.

The modified cluster method was developed at the Laboratory for Applications of Remote Sensing (LARS) and compared with supervised and unsupervised training methods (Fleming, et. al., 1975). The modified cluster method takes several small blocks of data and clusters each block separately. The analyst identifies each spectral class in each block by cover type, and then merges the statistics from all the blocks to form the final statistics file. The various combinations of techniques were tested on a mountainous area in Colorado (Fleming and Hoffer, 1977). One objective of the study presented here was to evaluate the three methods discussed above using data from interior Alaska. All three methods were implemented on both IDIMS and EDITOR.

There are two major factors which influence classification accuracy. One is the separability of the spectral classes. Separable spectral classes result in a minimum of confusion between the classes when the data set is classified. The best situation is characterized by landcover classes which correspond to spectral classes derived from the clustering and editing processes of the analysis. The second important factor affecting classification accuracy is the description of the spectral classes during the analysis and in the final classification. The descriptions label each spectral class as a landcover type.

This study evaluated three methods of spectral class descriptions. 1. Spectral classes were described during the training process. 2. On the IDIMS system the completed classifications of the entire study area were displayed on the color display screen. Each spectral class was described by visually comparing the reference data with the classification displayed on the screen. 3. Ground plots were located and identified using data from field work and photo-interpretation. Printout maps at the scale of 1:24,000 were made for each classification. The ground data point was located in the classification on the printout maps and the spectral classes described using the ground plot information.

1.2 Description of Study Area

1.2.1 Physical Features

The western half of the Fairbanks D-2 quadrangle was selected as the study area. Figure 1.1 shows the study area outlined on band 7 of the Landsat data. The study area dimensions are 28.2 by 12 km. (17.5 X 7.5 miles) covering approximately 34,150 hectares (84,000 acres). The study area is located near the northern edge of the Tanana Valley at

Figure 1.1. Fairbanks study area on band 7 (.8-1.1um) of Landsat scene 5470-19553.

the junction of the Chena and Tanana rivers.

The major drainages in the study area are the Tanana River, Goldstream Creek to the west and Our Creek to the north. The elevation rises from 60m near Salchaket Slough to 650m on the ridge above O'Conner Creek (USGS, 1949). The ridges are composed of Birch Creek shist overlain with a loess mantle. The valleys have alluvial deposits of sand and gravel up to 300m thick (Selkregg, 1975). The Chena and Tanana rivers have meandered near their current channels.

The study area is located in a zone of discontinuous permafrost (Selkregg, 1975). Slopes with southern exposures and the floodplains of large rivers are generally free of permafrost. North facing slopes and areas of slope break between the ridges and valleys have discontinuous permafrost. Poorly drained valleys and lowlands have nearly continuous permafrost near the surface.

The area is in the continental climatic zone (Selkregg, 1975) with temperatures ranging from -50°F in winter to +90°F in summer. The mean January temperature is -11°F and the mean July temperature is +60°F (Lutz, 1956). The photoperiods are controlled by the northerly latitudes with short days in the winter months and long days in the summer (Selkregg, 1975).

1.2.2. Vegetation

The study area encompasses a wide range of vegetation communities in a small area. The general vegetative patterns are part of the subarctic - subalpine or taiga region (Daubenmire, 1978). The major factors controlling the distribution of cover types appear to be permafrost, microclimate, topography, soil moisture, fire, and human activity. (Neiland and Viereck, 1977; Selkregg, 1975).

Prior to the discovery of gold in 1902 on Pedro Creek, there was little human activity in the area (Selkregg, 1975). For 15 years after the discovery of gold, mining activities greatly affected vegetation patterns in the area. Much of the timber in the area was cut for building materials, railroad ties and large quantities of firewood. Many stream valleys were mined with shafts. Later large dredges and hydraulic mining removed the vegetation and organic layer, and exposed bare rocks and gravel. The small communities of Chena, Happy and Ester were established in the study area. The University of Alaska was established in 1921 and fields were cleared near its campus for the Agricultural Experiment Station. The towns of Fairbanks and College coalesced, especially during and after World War II. From the late 1940's through the 1960's homesteads were established throughout the study area. Homesteads were chiefly located in the Goldstream Valley and between the Chena and Tanana rivers. A few clearings are scattered throughout the rest of the study area. Many of the fields were cleared to prove up on the homesteads and then allowed to regrow with the natural vegetation of the area. A few fields are still planted every year, especially on the University of Alaska Agricultural Experiment Station lands. In the past ten years, with the construction of the Trans-Alaska Pipeline and the growth of Fairbanks, many new residences have been built throughout the study area. The only locations in the study area without strong human influences are those south of the Tanana River, on the north sides of Chena Ridge and Ester Dome, and on the ridge to the north of the Goldstream Valley.

There have been several recent fires in the study area. There was a fire on the ridge to the east of O'Conner Creek. A fire burned several hundred acres south of the Ballaine Road and north-west of Goldstream Creek. A small fire burned near the top of the ridge above Farmer's Loop in 1978, which was after the acquisition of the Landsat data used in this study and before the field season. A fire in 1975 burned in the bog area south of the Tanana River.

1.3. Descriptions of Major Cover Types

There are several major cover types in the study area. Each of these cover types is qualitatively described. Vascular plant nomenclature follows Hulten (1968). These landcover descriptions correspond to levels II and III of the fifth version of the Provisional Framework of Alaskan Vegetation (Viereck and Dyrness, 1980). Mapping categories were developed from these descriptions (Section 2.5.2).

Black spruce (*Picea mariana* (Mill.) Britt., Sterns & Pogg.) stands (Figure 1.2) grow in the broad stream valleys, near the slope break on south facing slopes, and on north facing slopes. Most black spruce grow on poorly drained sites underlain with permafrost. Most stands

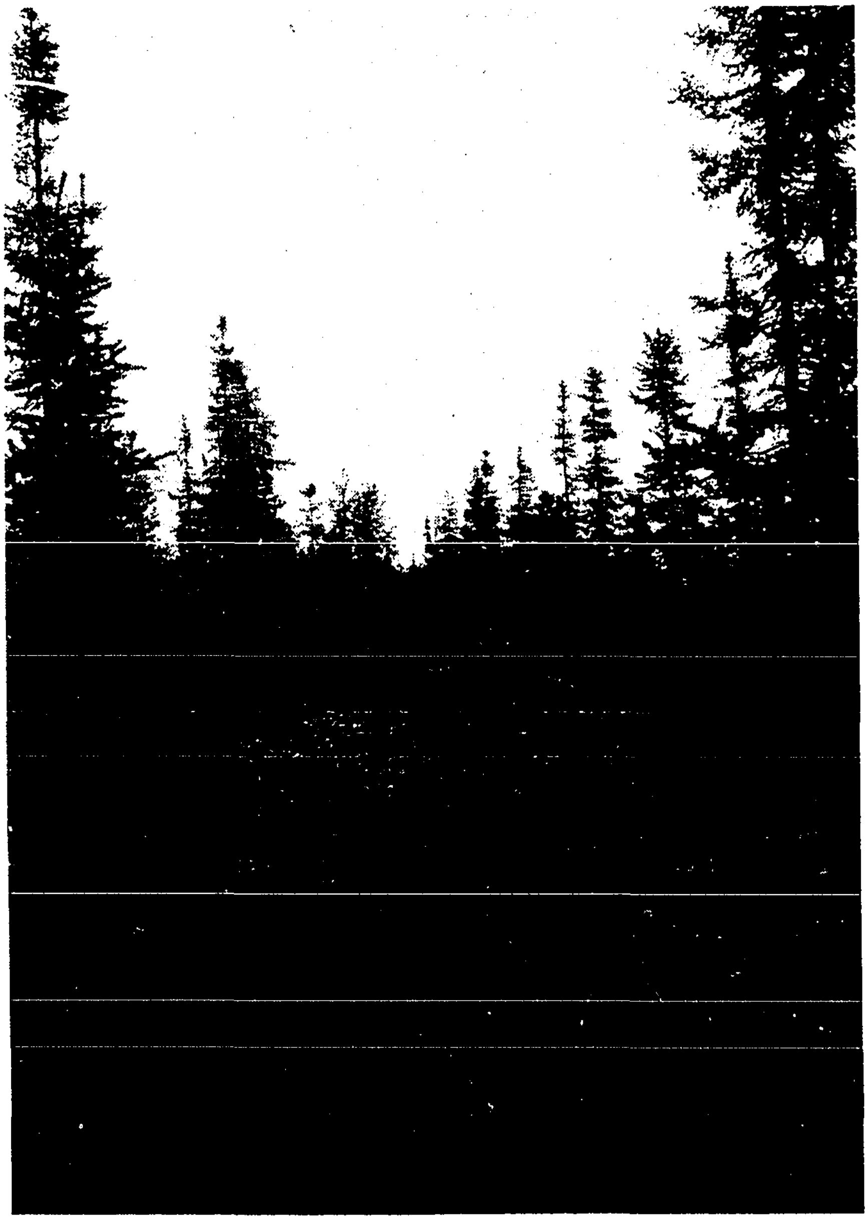


Figure 1.2. Typical black spruce stand.

are sparse (80 percent crown closure of black spruce is rare) with an understory of dwarf birch (Betula nana L.), willow (Salix spp.), ericaceous shrubs, mosses, lichens, and some herbs. Some black spruce stands have scattered tamarack (Larix laricina (Du Roi) K. Koch) but tamarack is not a dominant species anywhere in the study area. Black spruce stands have inclusions of small stands of birch and brush areas. Some black spruce had a growth form with spreading lower branches and a spindly top which made them difficult to interpret on color infrared aerial photographs (Section 2.5.1).

White spruce (Picea glauca (Moench) Voss) stands grow almost exclusively along the floodplains of the Tanana and Chena rivers. They grow in well drained gravels where the permafrost is lower beneath the surface than on the surrounding uplands outside the floodplains (Viereck, 1970). Most stands are narrow and linear except in the maze of old meanders on the eastern side of the Tanana River where it turns south off of the end of Chena Ridge. White spruce is scattered through the deciduous stands on south-facing hills and occurs in varying mixes with birch and aspen on the slopes above O'Conner Creek. White spruce is mixed with cottonwood (Populus balsamifera L.) near the Tanana River. Figure 1.3 shows a white spruce stand on the banks of the Tanana River.

Birch (Betula papyrifera Marsh), aspen (Populus tremuloides Michx.) and cottonwood are the three major deciduous tree species in the study area. Aspen generally grows on the warmest and driest south facing slopes with no permafrost (Rieger, et. al., 1963). Birch grows

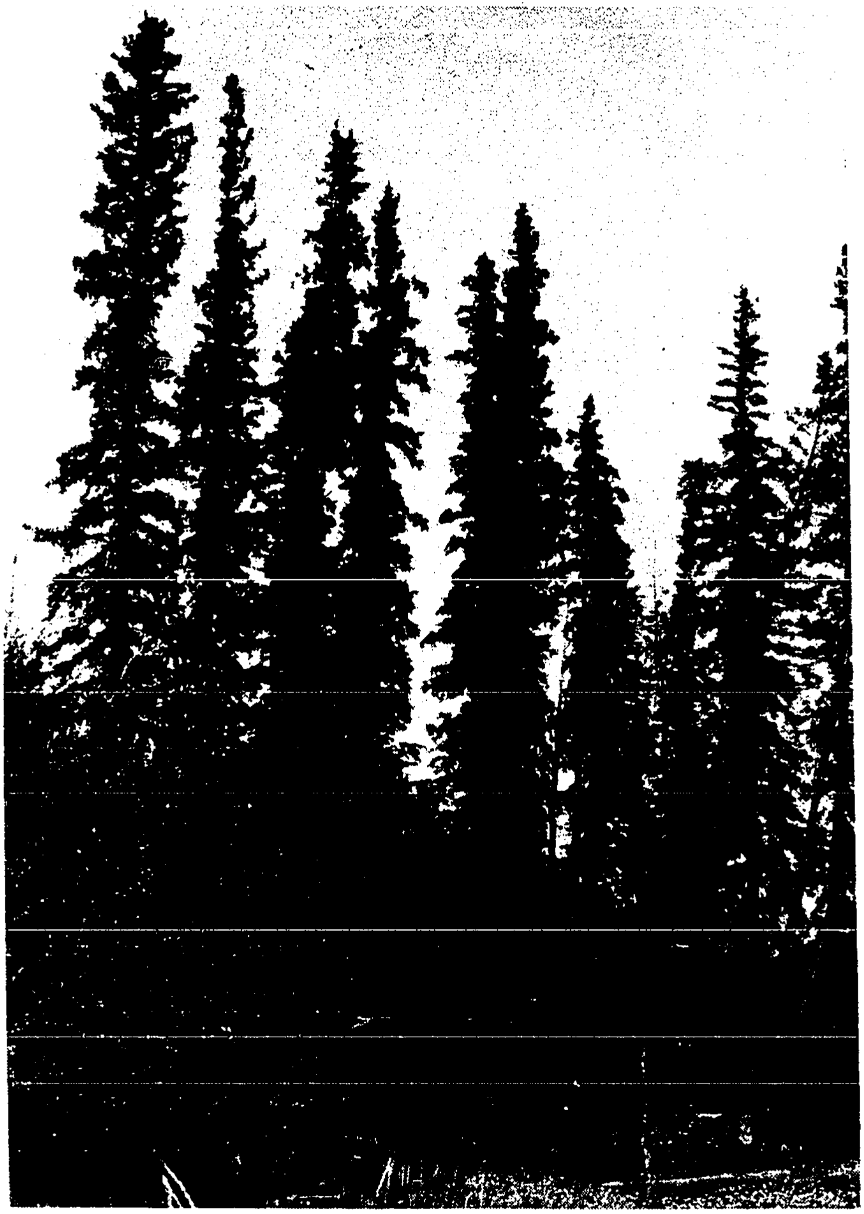


Figure 1.3. White spruce on Tanana River floodplain.

near the tops of hills on north facing slopes, on south facing slopes and on moderately drained sites in the valleys with discontinuous permafrost. Cottonwood grows primarily on the floodplains of major drainages with scattered stands on south facing slopes. The distributions of all three species overlap to some extent. The three species could not be reliably differentiated from a distance in the field during the summer or on the color infrared aerial photographs, so they were combined into the deciduous category. Large stands of predominately birch and aspen grow on Chena Ridge, Ester Dome and the ridges on both sides of Goldstream Valley. Cottonwood stands grow on the islands and banks of the Tanana River and occasionally mixed with birch and aspen on south facing slopes.

There are several locations in the study area where birch, aspen, and cottonwood saplings form a major part of the cover. These areas were considered as brush unless the saplings were greater than 5m (16.4 ft.) high. Figure 1.4 shows a mature deciduous stand. Deciduous stands have inclusions of brush (regrowing fields), conifers, and barrens (roads and residences).

Mixed forest stands (Figure 1.5) incorporate varying combinations of coniferous and deciduous species. The major stands are mixtures of birch, aspen, and white spruce on the slopes above O'Conner Creek and Big Eldorado Creek. There are a few stands of mixed white spruce and cottonwood along the Chena and Tanana Rivers. Some mixed black spruce and birch stands occur on lower slopes of the valleys in the study area.

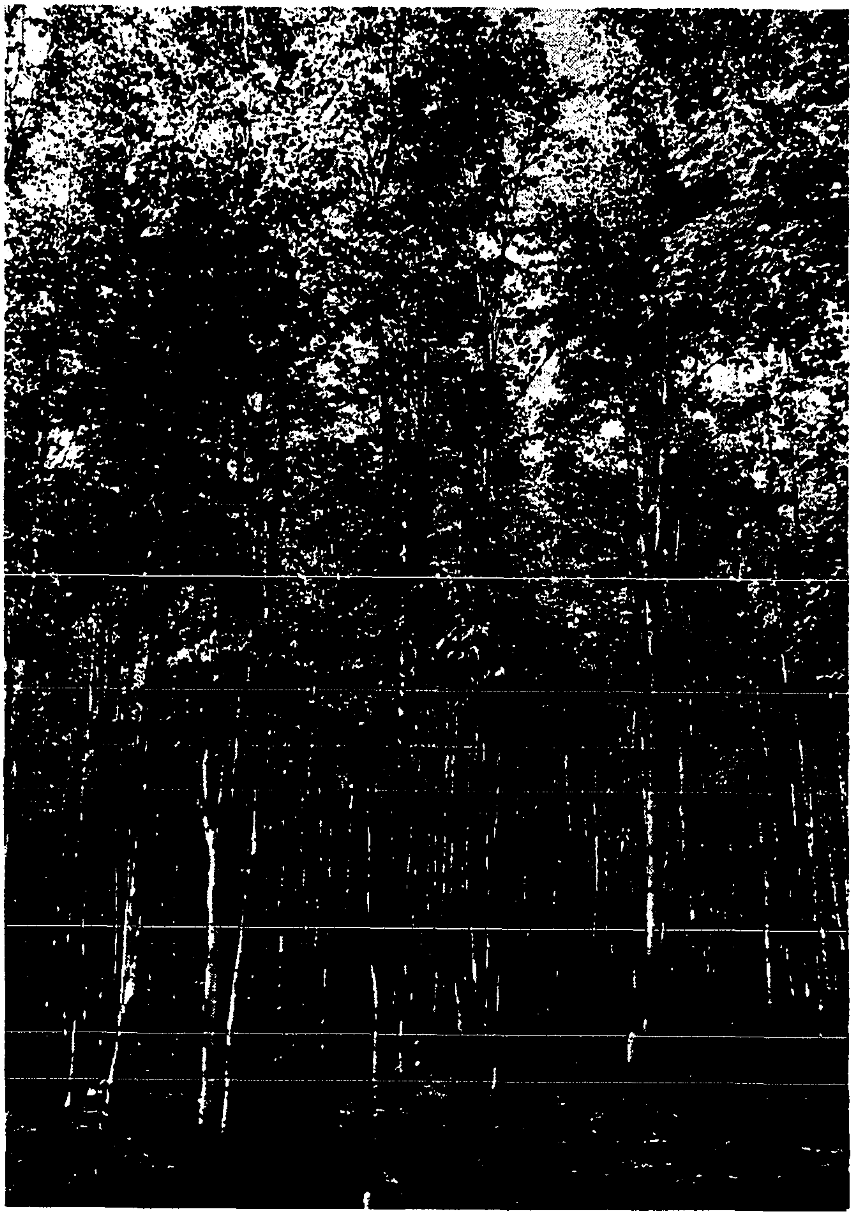


Figure 1.4. Mature birch stand typical of the deciduous cover types near Fairbanks.

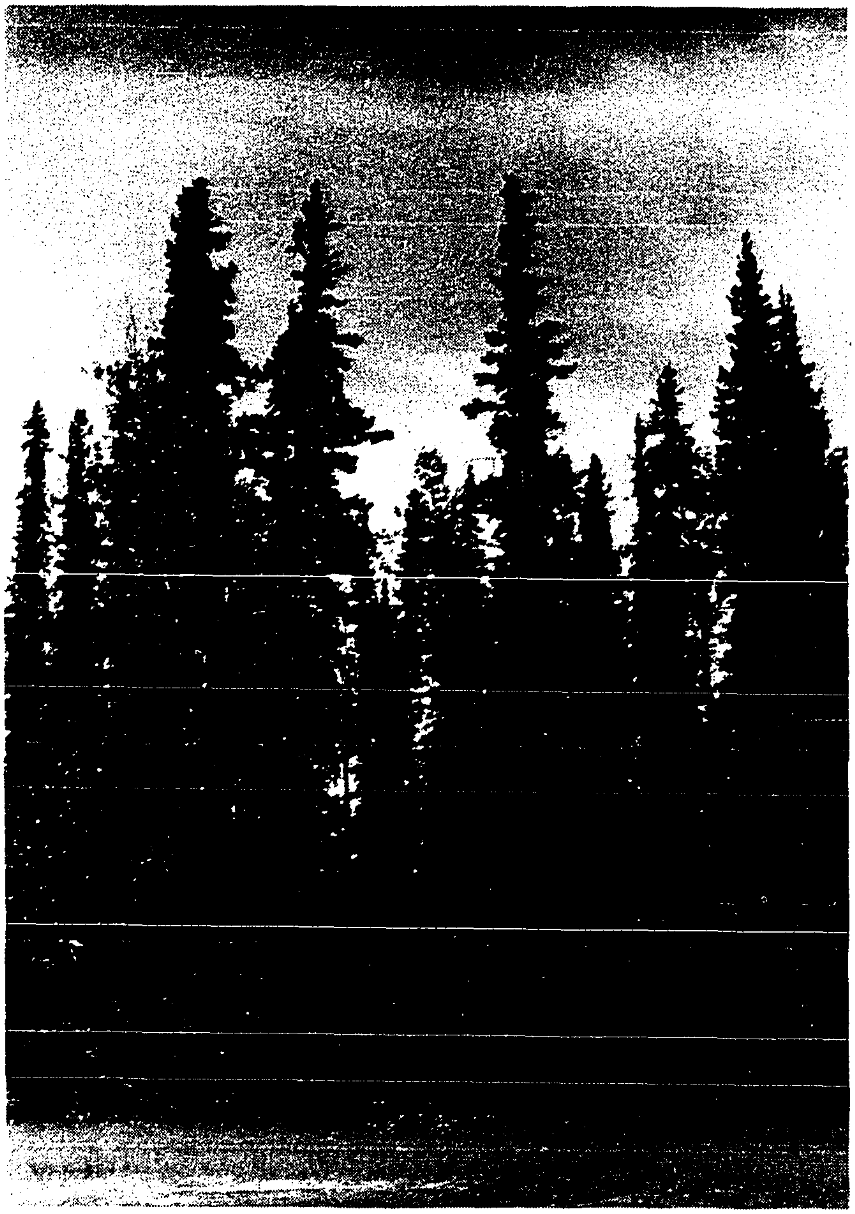


Figure 1.5. Typical stand of mixed white spruce and cottonwood.

There are two major brush communities in the study area (Figures 1.6 and 1.7). One is the sapling regrowth stands noted above. The birch, aspen, and cottonwood saplings are usually mixed with willow or alder (*Alnus crispa* (Ait.) Pursh). This type is usually found on well-drained sites cleared to mineral soil with permafrost well below the surface (Lutz, 1956). The primary locations are on sandbars along the Tanana River, old clearings, fire areas in the Goldstream Valley, some areas east of Chena Ridge, and on old gold dredge tailings near Ester and along Little Dome Creek. The other brush community is dominated by dwarf birch and willows with *Ledum palustre* L., *Vaccinum vitis-idaea* L., *V. uliginosum* L., *Equisetum* sp., *Rubus chamaemorus* L., grasses and *Carex* spp., and a ground cover of mosses such as *Sphagnum* spp. and *Hylocomium* sp. and fruticose and foliose lichens. Sparse black spruce or tamarack grow in these brush stands; and this type grades into the black spruce type. This brush community grows on sites which have not had recent major disturbance which removed the peat layer. These sites are generally poorly drained, acidic and underlain by discontinuous permafrost (Rieger, et. al., 1963). The wet brush community occurs on many flat lands in valleys, especially areas of the Goldstream valley which were not cleared by bulldozers during summer; on lower slopes of O'Conner and Big Eldorado drainages; between Ester Dome and Sheep Creek Road; along Our Creek; between College and Farmer's Loop roads; and between the Chena and Tanana rivers where recent disturbance has not been severe.

Pure grassland types (Figure 1.8) are fairly scarce in the study

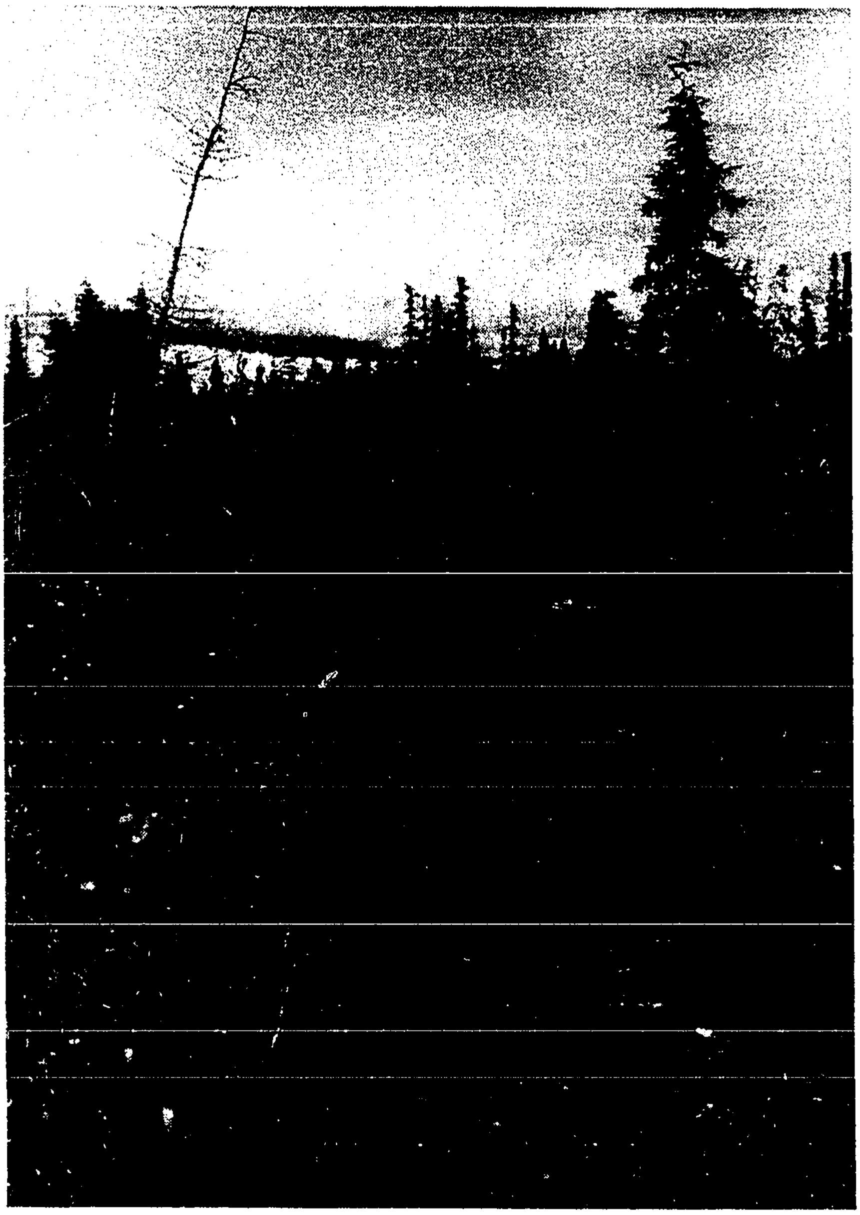


Figure 1.6. Wet brush cover type found in low areas with permafrost near the surface.

Figure 1.7. Successional brush community typical of cleared fields or burns.

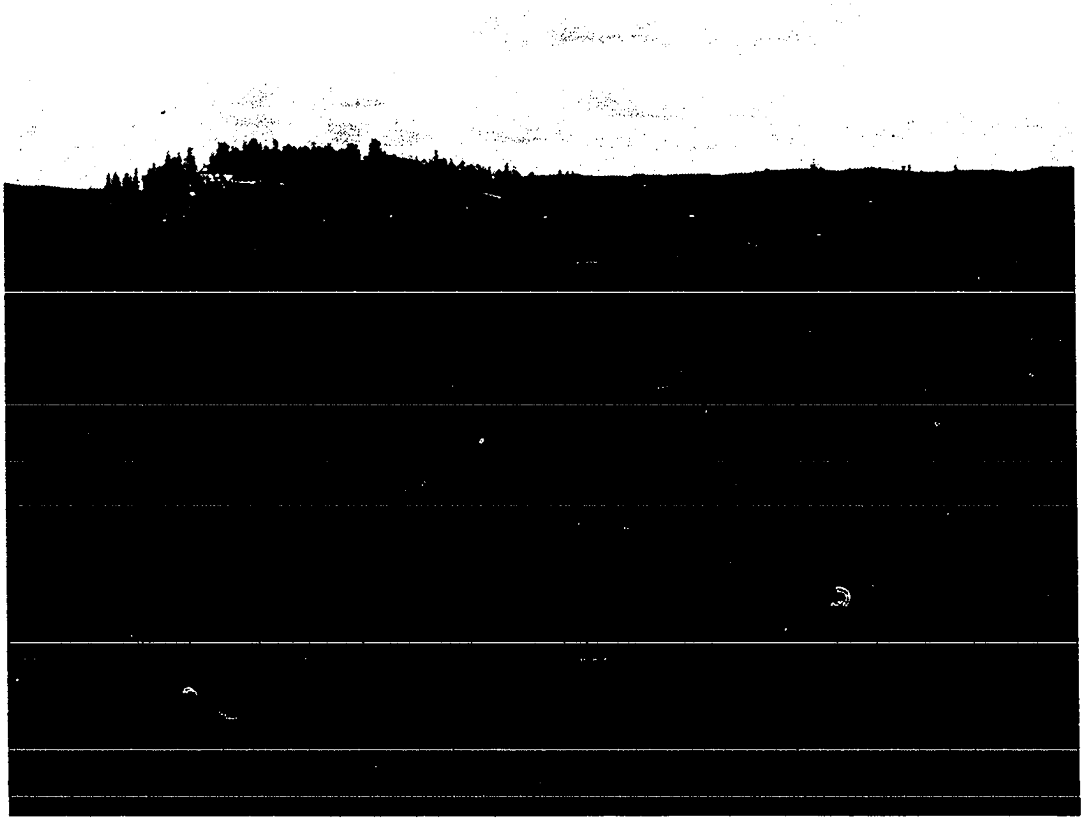


Figure 1.8. Grassland community in a recently cultivated field.

area. This type is composed of pasture and agricultural crop fields near the University of Alaska and on the ridge south of Goldstream Creek and a golf course on Farmer's Loop Road. Grasslands are difficult to identify and map because the planting pattern varies from year to year and percent cover varies throughout the growing season. These areas may have been mapped as grasslands when they were actually barren or sparsely vegetated in August 1976 when the Landsat data were acquired.

Sparsely vegetated communities (Figure 1.9) are mixtures of grass or brush with barrens. Sparse brush stands grow on sandbars in the Tanana River and on some of the tailings piles near Ester and Little Dome Creek. Sparse grass and tree communities are in dense residential areas with a mix of roads, residences, lawns, and trees. This type occurs primarily between College Road, Chena Pump Road, and the Tanana River.

Barren areas (Figure 1.10) have little or no vegetative cover. This type includes sandbars, gravel pits, tailing piles, roads, residences, airstrips, and recent clearings. Patches of barrens occur throughout the study area, but the major concentration is along the Tanana River and the urban area bounded by Chena Pump and College Roads and the Tanana River.

Silty water flows in the Tanana River due to the high silt load from glaciers at its headwaters (Selkregg, 1975). Clear water is found in the Chena River and various lakes and impoundments throughout the study area. Figure 1.11 shows clear water in an impoundment

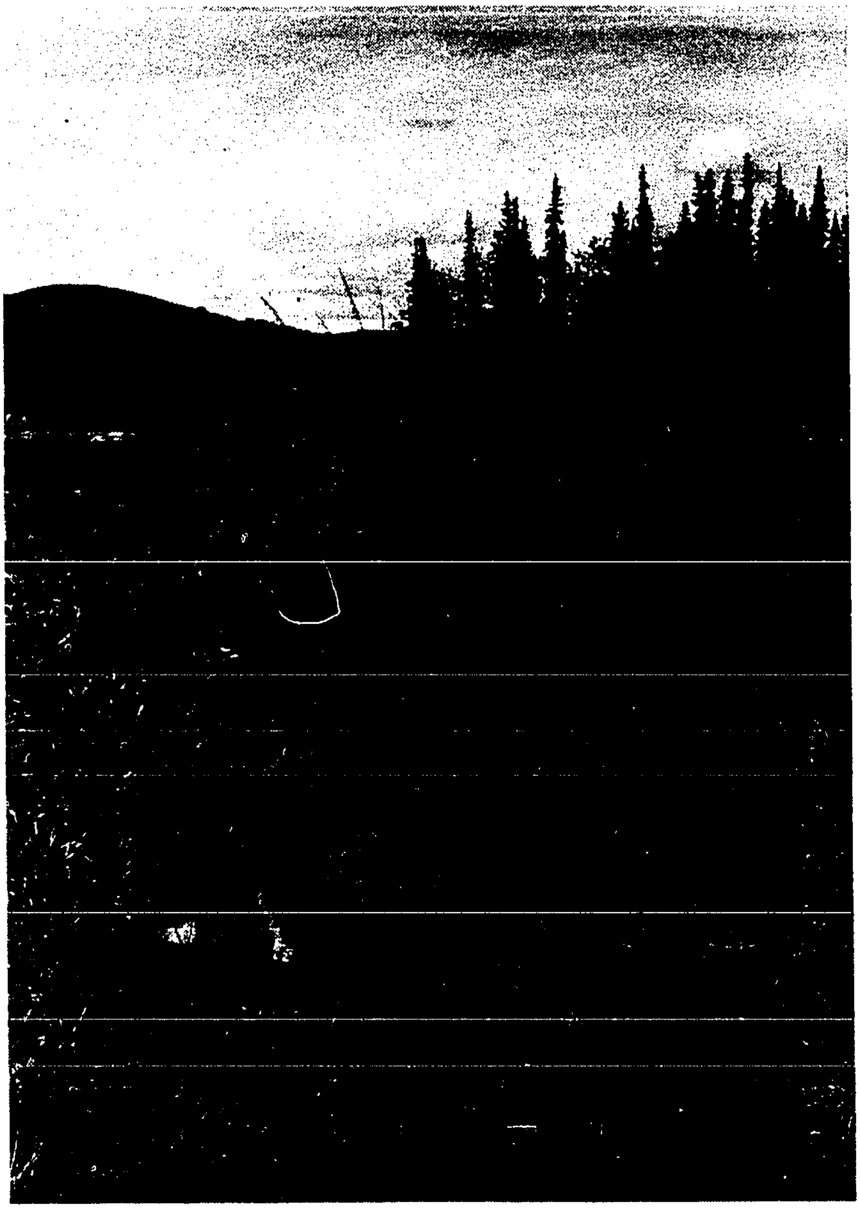


Figure 1.9. Sparsely vegetated stand on a sandbar in the floodplain of the Tanana River.

Figure 1.10. Tailings pile in Little Dome Creek illustrating the barren cover type.

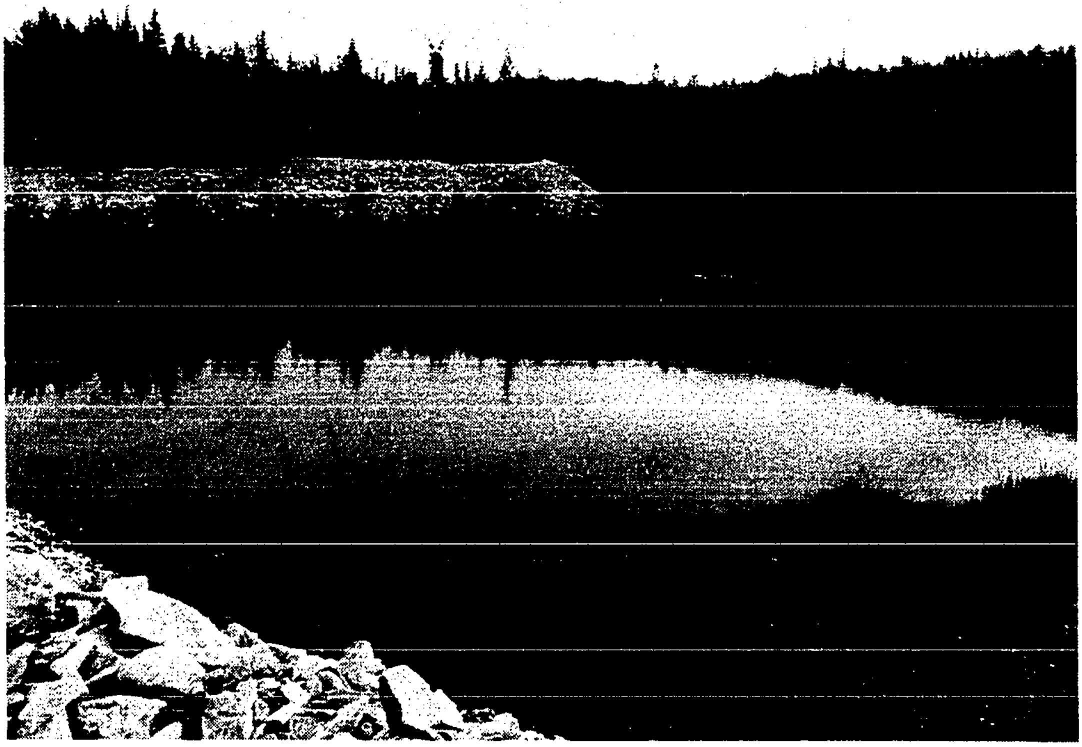


Figure 1.11. Clear water impoundment in Little Dome Creek.

resulting from dredging in Little Dome Creek.

The 1975 burn on the Fort Wainwright Military Reservation was stratified into a separate cover type during the post-classification analysis. The burned area was within the range of variability for brush with scattered patches of birch and black spruce on the aerial photographs and the Landsat image. The area was initially mapped as brush with birch and black spruce inclusions. However, during the spectral class descriptions spectral confusion was evident in the area of the burn, so it was stratified into a separate cover type category. When examined during the 1979 field season from light aircraft the area appeared to be hummocky with the wet brush type having an extensive component of grasses or sedges. The other burns in the study area were mapped as the current cover type. The burns were covered primarily by brush or sparse forest communities. The cover type of a recent burn is dependent on the severity of the fire, the length of recovery time, environmental conditions in the area of the burn, the vegetation prior to the fire and characteristics of the species which are established in the burn area after the fire.

1.4 Primary and Ancillary Data

There was good access in the study area to most of the cover types. The recent urban development in the area has resulted in a network of roads, trails, powerlines, clearings, and survey lines to allow access on foot. The only area not available for foot access was south of the Tanana River. This area was observed from low flying aircraft during the field season. The analyst had lived in the study area for approximately five years prior to the computer-aided analysis. This familiarity with the study area, knowledge of the history of human activity, and an awareness of the distributions of the cover types was very valuable during the analysis. Field data were collected during the summer of 1979.

Good Landsat data and reference data were available for the study area. Scene number 5470-19533, acquired on August 1, 1976, had good quality data in all four bands. The study area is located in the northwest corner of the scene. The August 1 date is during the peak of vegetative development. The computer-compatible tape of the digital data was available at the Landsat Library of the Geophysical Institute, University of Alaska. The scene had been digitally enhanced using the EDIES process at the EROS Data Center. The EDIES process used contrast stretch and edge enhancement techniques on the digital data before an image was generated. This enhancement applied only to the images and did not affect the digital data used for the computer-aided analysis. Manual interpretations from enhanced images are better than interpretations from standard images (Spencer, 1977). Good winter Landsat data were also available for use as reference data. Varying snow depths on scenes throughout the winter helped to separate brush and sparse black spruce, especially in wetland areas.

Color infrared photographs were available for most of the study area prior to the computer-aided analysis in November - December 1978.

Table 1.1. List of aerial photographs and Landsat data used for study.

Aerial Photographs

Date	Scale	Film Type	Roll	Frames
1972	1:30,000	Color IR	Mission 209	153-160
1974	1:130,000	Color IR	1814	1123-4
1974	1:130,000	Color IR	1809	9247-8
1977	1:65,000	Color IR	2506	1755-6
1977	1:130,000	Color IR	2505	5917-8
1978	1:60,000	Color IR	23	146-8 318-20
1979	1:65,000	Color IR	2800	3747-9

Landsat

Scene ID	Date	Band(s)	Format
5470-19533	8/1/76	4,5,7	Color Composit
		4,5,6,7	Computer-compatible tape

Approximately two thirds of the study area was covered by large scale natural color and color infrared photographs acquired in 1972. Excellent quality 1:60,000 color infrared photographs of the entire study area, acquired during July - August 1978, were available for the field season and photo-interpretation of the ground plots. The aerial photographs used for this study are listed in Table 1.1.

Topographic base maps at scales of 1:24,000, 1:63,360, and 1:250,000 were available. The 1:24,000 scale was used for the field work, photo-interpretation and the printout maps of the classifications.

1.5 Overview of Computer-aided Analysis

Computer-aided analysis of Landsat multispectral scanner data is available on many different computer systems. The exact analysis procedures vary among computer systems, analysts, and the circumstances of each project. However, most follow the general data flow discussed below.

1. Project planning defines the objectives of the entire project and the data and products needed to meet these objectives. The products desired often have to be adjusted for the time and funds available. The combination of data types and analysis methods should be selected which will best fit the situation (Philipson, 1980). Many projects using remote sensing data have suffered because the user/ interpreter over-extended or improperly used the data (Sabins, 1978).

The remainder of this discussion assumes that computer-aided analysis of digital Landsat data has been selected, and that a good data set has been acquired over the study area.

2. The digital data are reformatted to be compatible with the computer system being used and the study area is extracted from the entire data set. On some systems these steps are accomplished concurrently.

3. Landsat data are skewed and rotated towards the northwest during acquisition. The geometric irregularities are due to the near-polar orbit of the spacecraft, rotation of the earth during data collection, the attitude and altitude of the spacecraft (Anuta, 1973).

Geometric correction of the data is necessary for training and spectral class descriptions on many systems (Hoffer and staff, 1974). Geometric correction is essential if the final products are to be used as maps. Bulk corrections can usually be done automatically using the latitude of the center of the study area. Bulk corrections on EDITOR were printed out as greyscales and overlaid with a 1:24,000 topographic map. Greyscales are displays of one band of the spectral data with the reflectance values shown in symbols resembling shades of grey. The greyscales of bulk corrected data registered within 0.5km with the topographic map. Precision geometric corrections use control points located throughout the study area to register the Landsat data to a map base (Anuta, 1973). Often a data set is bulk corrected and control points selected from this data set for a precision correction.

4. Training of the classifier is the core of computer-aided

analysis techniques. The objective of the training process is to identify the range of reflectance values for each of the land cover types in the study area. Training statistics are usually derived from a sample of the study area. The training data are selected to cover the spectral variability in the study area. The pixels in the sample which have been selected for training are clustered. During the clustering process the training pixels are divided into groups of similar spectral reflectance. Each group is called a cluster or spectral class. Each cluster is defined by a set of statistics that is based on the spectral characteristics of the training pixels grouped into that cluster. These statistics are means and variances for each of the four wavelength bands and a covariance matrix between bands.

There are two major methods for selecting the sample of training pixels: supervised and unsupervised (Hoffer, 1972). Techniques using combinations of both of these methods are frequently used (Fleming and Hoffer, 1977; Nelson and Hoffer, 1979). The selection of the method used in a project depends on many factors, especially the availability of reference data and the timing of the field season. A final set of statistics is derived from the sample clusters by combining various sets, pooling, and deleting some clusters (Fleming and Hoffer, 1977).

5. The set of statistics derived from the training is applied to data from the entire study area during classification. There are several types of classification algorithims (Hixson, et. al., 1980), but each generally examines the reflectance values for each pixel and classifies that pixel into the training spectral class with which it

is most similar. In this fashion all pixels in a data set are grouped into one of the spectral classes from the training data set.

6. Each spectral class must be assigned to an information class if the final products are to be more than a map of symbols. The time and expertise invested in this process directly affects the quality of the final products. If an unsupervised method of training was used, each class must be identified and described according to the cover type(s) it represents. If some training method has been used where the spectral classes have been identified prior to classification, these descriptions should be qualitatively assessed and refined as necessary. If several classes represent one cover type they can be combined into one map category. Often one spectral class will contain two or more cover types. More work with the training statistics and reclassification may help separate two similar cover types. Post classification stratification may help define the cover types more precisely. Stratification procedures incorporate additional data such as topographic data or vegetation distribution patterns to more accurately describe the spectral classes. Spectral class descriptions and other work using reference data usually are most efficient and effective when done by an experienced photo-interpreter with substantial field expertise from the study area and experience in spectral reflectance patterns.

7. Effective use of data from the final classification is difficult if the data reliability is unknown. A qualitative evaluation is generally made after classification, especially as an

aid in refining spectral class descriptions. A quantitative evaluation is made by comparing the final classification with an unbiased sample of reference data and calculating the proportion correct. Several different methods have been developed for sampling and accuracy calculation (vanGenderen and Lack, 1977; Rosenfield and Melley, 1980; Fitzpatrick-Lins, 1978; Krebs and staff, 1976). The method selected usually depends on access in the study area, and time and fiscal constraints of the project.

The data from computer-aided analysis can be output in a variety of final products. Maps can be made at the scale desired as printout or photographic products. Tabular estimates of acreage or percentage can be made for the entire study or digitized subsections. Digital products have been entered into a data bank and overlaid with other data bases such as topographic data, soil types, or land ownership. New maps can be made using combinations of these parameters as they are needed.

1.6 Background

Remote sensing is generally defined as the process of collecting, recording, and interpreting information about an object or landscape without coming in direct contact with that object (Landgrebe, 1978). Remote sensing is usually limited to methods which record electromagnetic radiation which has been reflected or emitted from an object. The major sources of remote sensing data are photographic and scanner sensors on board aircraft and spacecraft platforms.

1.6.1 History

The development of remote sensing technology has proceeded in a series of jumps and plateaus since the first aerial photograph was taken in 1858 (Fischer, 1975). Early photographs were taken from balloons, kites, and pigeons using bulky cameras and unstable films. The photographs were primarily used to develop topographic maps and to sell for souvenirs. In 1871 film was developed with an emulsion of silver halid grains that could be processed after returning to ground. In 1909 the first photograph was taken from an airplane. These early photographs were taken with panchromatic black and white film. Natural color film was commercially available in 1935. It was not widely used until the mid-1960's because of the slow film speeds and problems with haze penetration. Color infrared film was in use by 1942 as a military camouflage detection film, but has evolved as a tool for natural resources inventories due to the ease of photo-interpretation of vegetation types. Multiband photography has been used sporadically since 1855. In the mid-1960's NASA began a program to test the usefulness of multiband photography for resource applications.

Aerial photoreconnaissance was used extensively during World War II for detection of enemy movements and terrain mapping. There were two major benefits to resource inventories which resulted from this: 1) improvement of films and interpretation techniques and 2) training of a large number of excellent photo-interpreters. Aerial photographs were used for geologic interpretations beginning in 1920. Use for agricultural and forestry surveys began in the 1930's and aerial photographs are still a major component in these inventories.

In 1957 another advancement was made with the launch of Sputnik I which was followed by other spacecraft in orbit around the earth. Early manned space missions did not emphasize the acquisition of remote sensor data. Later Gemini flights and Apollo missions acquired large quantities of photographs. SKYLAB missions acquired both photographs and 13 channel multispectral scanner data. Perhaps the most important advance in earth-orbiting platforms was the launch of Landsat I in 1972. Two additional Landsats were launched in 1975 and 1978. These satellites carry multispectral scanners and the return beam vidicon systems. Vast quantities of multispectral data in a digital format has been recorded from these sensors.

Computer-aided analysis techniques are used to analyze numerical or digital scanner data. Early development of computer-aided analysis was done using data from airborne scanners. One approach involving the use of pattern recognition theory applied to scanner data began in 1966 at the Laboratory for Agricultural (later Applications of) Remote Sensing (Fleming and Hoffer, 1977). This process involves a machine/analyst interaction "whereby the man will train the computer (utilizing data collected over a limited geographic area), and then the computer will continue to map and analyze data collected over a large geographic area at a much faster rate than would be possible for the man if he were using normal image interpretation techniques" (Hoffer, 1972). The basic supervised and unsupervised training methods were developed before the launch of Landsat I (Hoffer, 1972). Much of the preliminary research was devoted to developing basic methods of data analysis. This research involved selection of optimum wavelength bands (Coggeshall and Hoffer, 1973), adequacy of the maximum likelihood classifier (Wacker and Landgrebe, 1971), development and comparison of training methods, and determining if cover types of interest could be reliably separated on the basis of spectral characteristics (Hoffer, 1972).

The corn blight infestation of the corn fields of the midwestern states in 1971 provided an opportunity for a practical application of computer-aided analysis of multispectral data from aircraft. Data acquired throughout the growing season and processed at LARS provided information about the spread and the severity of the corn blight. This monitoring program resulted in cost savings to the farmers and demonstrated that computer-aided analysis techniques could provide reliable data in a timely manner (Bauer, et. al., 1971). Rohde and Olson (1972), working in Michigan, mapped deciduous and coniferous forest types with 85 percent accuracy using six bands of data. This study utilized a supervised training approach. Several studies were undertaken to detect diseased or stressed trees with varying degrees of success (Aldrich, 1979). Problems were noted involving geometric corrections and spectral separability of the diseased cover types.

Preliminary analysis of data from one of the first Landsat scenes indicated "a great deal of potential in the analysis and interpretation of this ERTS (later Landsat) imagery" (Landgrebe, et. al., 1972). In 1972 and 1974 NASA funded a series of studies to investigate the potential of various interpretations of Landsat data to practical problems of user agencies. There was a push to make Landsat operational and to have the data used in many fields including geologic mineral exploration, forest resources inventories, land use monitoring, and archeological investigations (NASA, 1975). These programs and other projects supported by universities, government agencies, and private companies have provided a great impetus to the development of remote sensing analysis techniques and the acceptance of the data by resource managers. Many of the techniques have been identified in conjunction with mapping or inventorying projects to meet the objectives of the original study. A large number of techniques and processing systems are now available. The analyst can select or modify the techniques which fit his particular project.

Currently, Landsats 2 and 3 are providing repetitive coverage over the earth. Landsat D is scheduled for launch in 1982 and is scheduled to carry the seven band thematic mapper with higher spectral and spatial resolution than the current multispectral scanner of Landsats 1, 2, and 3 (Doyle, 1978). Other exciting possibilities are offered by the Space Shuttle and Stereosat. Analysis techniques will have to be developed further to meet the challenges of new data types and greater data volume.

1.6.2. Literature Review

Published studies concerned with field techniques specifically developed for digital analysis remote sensing projects are few. A wide variety of field techniques have been developed for the disciplines of plant ecology, forestry and range management which address specific problems or information needs. Many of the data collection techniques incorporate field work and aerial photographs to acquire information for a study or management concern (Aldrich, 1979). Specialized field work and reference data must be acquired for the analysis of digital Landsat data before the landcover maps can be used for further interpretations.

Hoffer (1971) discussed the need for "ground truth" data in a project involving multispectral scanner or radar data. He emphasized that an analyst needs an understanding of the energy-matter relationships recorded in the data. "Knowledge of these energy-matter interactions allows the spectral characteristics of the materials to be predicted and the remote sensor imagery to be accurately interpreted." He lists the primary causes for spectral variation within and between vegetation types: differences in ground cover, maturity, cultural practices, disease, moisture stress, insect infestations, geometric configuration, and environmental variables. Experience at LARS showed that a truck mounted aerial platform was effective for repetitious observations of small, well known field sites. Most of the research involved agricultural lands. Hoffer also discussed the utility of combining several sources of field and remote sensor data for the most effective set of reference data at the least cost.

A.T. Joyce (1978) has prepared a handbook which addresses field procedures for training fields to be used in a supervised classification. This handbook is for use by a remote sensing staff or field teams of land managment agency personnel who will be using the final products of the classification. Criteria are given for the selection of training fields based on landcover and topographic variability, number of plots, and size and shape of stand. Joyce recommends locating potential sites and tentatively identifying them using aerial photographs. The photographs and maps are used by the field crews to locate the stand. A "forest uniform verification pattern" is walked through forest stands with a compass to ensure that the stand is uniform and a homogenous vegetation type. Agricultural fields may be spot checked at several points by walking into the field. Ground truth forms were prepared for major landcover types. These forms are filled out by the crew. A variety of data is recorded for each training field including major species, sparse or dense crown cover, training field size, general condition, disease or flooding conditions, age class, and land use. This system requires that every training field be verified on the ground prior to clustering in the computer-aided analysis. Joyce recommends that training fields be 40 to 160 acres in size, with 10 acre fields the minimum. The larger plots are more easily and accurately located in the Landsat data using a color display screen. The supervised approach to training the computer classifier can use a set of desired mapping categories for the classification. Therefore, a flexible coding system for the vegetation plots is not necessary.

Many factors control the design of field and reference data collection techniques for computer-aided analysis projects. Craighead (1980) collected detailed and extensive field data in support of a digital analysis of grizzly bear habitat. The field data were collected using land/vegetation categories based on ecological characteristics. The resulting Landsat classification of moderately detailed map categories had accuracies from 85 to 91 percent correct. When ecotone areas were accounted for in the accuracy calculations, an accuracy of 89 percent was increased to 93 percent.

Many projects use a combination of field work and photo-interpretation similar to that used by Fleming and Hoffer (1977) and Mayer, et. al. (1979) to make most efficient use of a short and expensive field season. Some projects in remote areas with few or no aerial photographs must use spot landings with helicopter and observations from light aircraft (George, 1981).

A study along the Denali Hiway in Alaska (ESL, 1978) used a complex multistage sampling technique to collect field and reference data to support a computer-aided classification. The field data were collected on one tenth square meter plots for ground cover, four square meter plots for shrubs, and a circle with a fifteen meter radius for trees. Thirty-five mm photo plots were interpreted using the detailed ground data. The photo plots varied in size from one fourth to one acre. Problems were encountered in accurately locating the field and photo plots in the Landsat data and in making correlations between the different levels of data: field work, 35 mm aerial photographs, and the Landsat data.

A field technique similar to the one discussed in Section 2.5 was used by the author in Colorado (Krebs and staff, 1976). The field work in the Colorado study was designed to serve as the sole source of reference data if aerial photographs were not acquired that summer. Data were collected for homogenous stands of 5 to 200 acres. Preselected points were located in the field and the extent of the stand represented at that point was outlined on a map.

Another important consideration in planning the field work for a digital analysis project is a flexible recording system for the landcover types in the study area. Poulton (1972) has developed a comprehensive legend system which incorporates characteristics of the remote sensor data and the ecological characteristics of the landscape. The numerator of his legend concerns the vegetation of an area. The detail of the vegetation increases from a resource class to physiognomic type to a specific ecosystem. The denominator depicts the physical characteristics of the environment. The hierarchal physical legend increases in detail from macrorelief to landform to surficial geology to soil taxa. This legend system is flexible and responsive to manual interpretations of remote sensor data. However, the vegetation categories are predefined by species mixture and crown closures. A review of the current literature shows that the field work for most studies is based on a predefined set of vegetation categories.

Comparisons of factors affecting classification accuracies are rare in the literature. This may partially be due to the relatively young age of computer-aided analysis techniques and the emphasis on applications by funding agencies. Early comparisons involved the selection of which spectral bands and the number of spectral bands of data to be used. These studies were conducted on aircraft data with 12 channels (Rohde and Olson, 1972), Skylab data with 13 channels (Hoffer and staff, 1975) and with multi-temporal sets of Landsat data (Hoffer and staff, 1974). These studies sought to find the trade-off between accuracy and cost.

There are several different classification algorithims which use the training statistics developed by the analyst and classify the entire data set into spectral classes. Bauer, et. al., (1977) compared the accuracies, CPU costs, and analyst times for five different classification algorithims. The classifiers tested were 1) per point maximum likelihood, 2) ECHO (Kettig and Landgrebe, 1973), 3) minimum distance, 4) layered, and 5) levels (parallelepiped). All five were used with the same training statistics, study area, and test data. The levels classifier had a lower accuracy with higher cost and time requirements than the other four classifiers. No significant difference of accuracies was found between the maximum likelihood, ECHO, minimum distance, and layered classifier but there were some variations of cost and time involvement. Hixson, et. al. (1980) compared five classifiers for crop identification. The classifiers were similar to the study of Bauer, et. al. (1977); 1) per point maximum likelihood, 2) sum-of-normal densities (developed for LACIE), 3) minimum distance, 4) layered, and 5) ECHO. Two different training sets were used for the sum-of-normal-densities classifier: 1) the automatic clustering procedure which is part of the usual analysis and 2) ISOCLS clustering. Accuracies for the five classifiers using the normal training methods (not the ISOCLS method) were not significantly different from each other at the .95 confidence level. The sum-of-normal-densities using ISOCLS resulted in significantly lower accuracies. The classifiers were ranked according to cost and ease of use with the minimum distance being the least expensive and easiest to use. The authors concluded that development of training statistics is more important than selection of classifier. Both studies indicate that the analyst should select the classifier which will best serve the individual project.

The supervised and unsupervised training methods were developed using aircraft multispectral scanner data. There were difficulties in using each of these methods so additional training methods were developed using Landsat data. One of the initial methods was called modified cluster and combined parts of both supervised and unsupervised (Fleming et. al., 1975). Comparison of the modified cluster with unsupervised and modified supervised showed accuracies of 78.5 percent, 70.0 percent, and 84.7 percent respectively. Further investigations by Fleming and Hoffer (1977) used six different training methods: 1) supervised, 2) modified supervised or multicluster fields, 3) mono-cluster fields, 4) non-supervised, 5) mono-cluster blocks and 6) multi-cluster blocks (modified cluster of Fleming, et. al., 1975). The six methods were compared in terms of CPU time, analyst time, accuracy, and support data requirements. Both of these studies found the modified cluster or multi-cluster blocks approach was the most accurate overall and used reduced time and cost when compared to the other methods. This accuracy was achieved using less reference data than methods 1, 2, 3, or 4. This is often an important consideration for a project.

The multi-cluster approach was compared with the Procedure 1 method developed during LACIE (Nelson and Hoffer, 1979). The multi-cluster and Procedure 1 training methods were used to develop training statistics for the sum-of-normal-densities and the maximum likelihood classifiers. A variety of clustering methods were also used: 1) ISOCLS unseeded with 10 iterations, 2) ISOCLS seeded with 3 iterations, and 3) 1 iteration and 4) CLUSTER algorithim - a different method of clustering from ISOCLS. A total of 12 classifications were evaluated. The Procedure 1 with unseeded ISOCLS clustering and multi-cluster blocks with CLUSTER clustering were not significantly different and had the highest average classification accuracies. The researchers recommend that the analyst select the method which fits the study area considerations of size and type of reference data. No significant difference was found for the comparison of the two classifiers;

maximum likelihood and sum-of-normal-densities. The Procedure 1 generally had higher accuracies using ISOCLS clustering and the multi-cluster block method had higher accuracy with the CLUSTER clustering algorithim.

Few researchers have addressed the problems of methods of spectral class descriptions. In 1972 Landgrebe et. al., outlined the problem of relating spectral classes with resource categories. "One analysis step most in need of study at the present state of the remote sensing art appears to be the refinement of a straight-forward technique to relate the spectral classes present to the significant categories of interest defined by the users." Seven years later Isaacson, et. al., (1979) reiterated the need for research involving the techniques of spectral class descriptions. "A current limitation of Landsat multispectral scanner data in wildland resource inventories is the complexity of interrelationships among spectral and resource information classes. Although there is a clear need to develop a systematic methodology to be used in examining these relationships, this area is little studied, and the number of classes generated by the various classifiers exceeds the capability of the data analyst to provide attendant resource-oriented descriptions." These authors reported the developing of one technique for spectral class descriptions. This method involved constructing a co-occurrence table and interpreting the positive associations between the spectral and resource classes. The resultant mapping aggregations were not evaluated using separate test data.

Most facilities for remote sensing interpretations do not have access to multiple computer-aided analysis systems. This has contributed to the paucity of research directly comparing different computer-aided analysis systems. At facilities which do have more than one system, analysts generally make a qualitative evaluation of the systems and use the parts of each system which best serve each project. Morrisey and Ennis (1981) used both the IDIMS and EDITOR systems to complete a Landsat classification of NPR-A on the North Slope, Alaska.

Carter, et. al., (1977) developed summary tables for comparing various processing systems and their hardware features and capabilities. This is a useful starting point, but it only lists the hardware and software capabilities comprising each system and does not attempt to evaluate the results from the various systems. Phillips and Swain (1978) give a general procedure for evaluating an analysis system. This method considers the capabilities of a system which are considered important by the analyst.

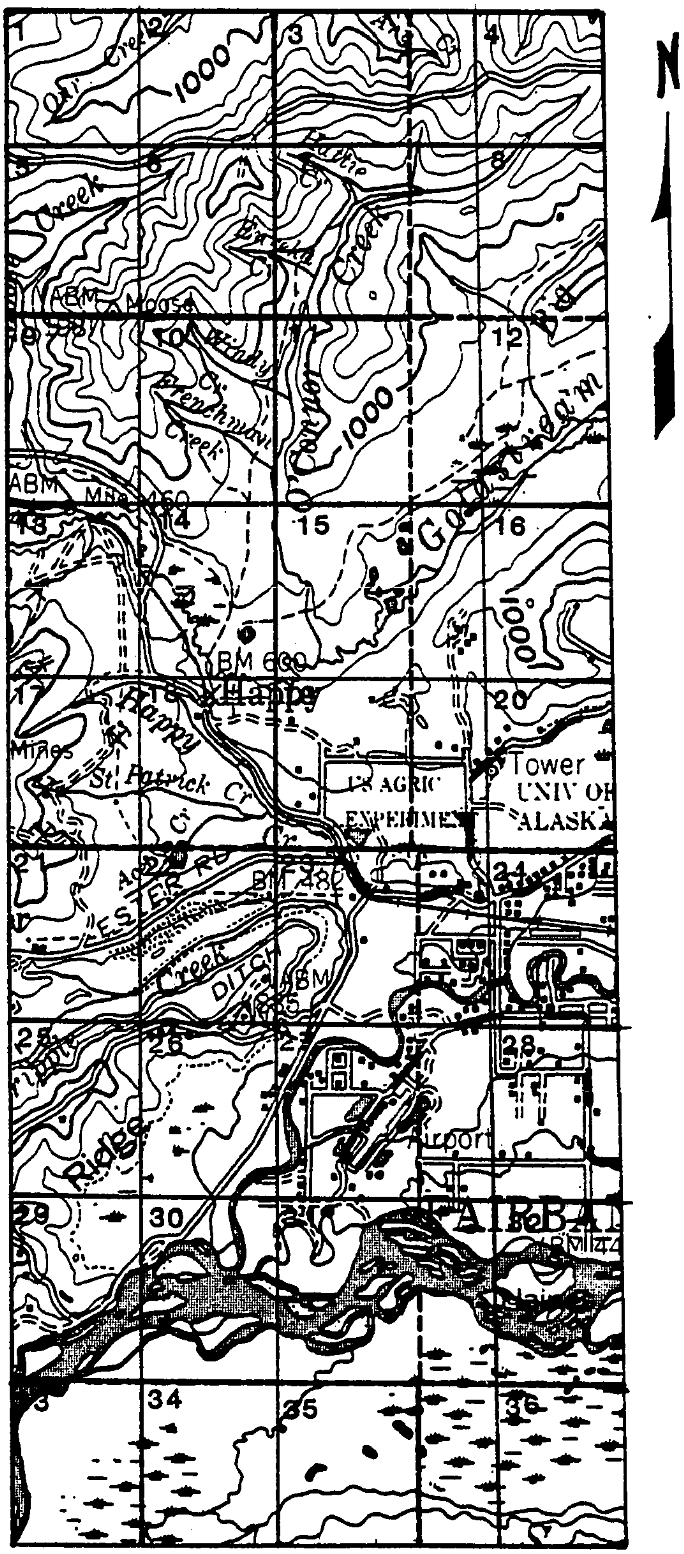
CHAPTER II

METHODS

The primary foci of this study were the comparison of computer systems and analysis techniques and the development of a field method and coding system suitable for digital remote sensing projects in Alaskan wildlands. The field and photo-interpretation techniques and a flexible coding system are discussed. The methods outlined in this chapter permitted direct comparisons of the various analysis factors being studied. Twenty-six different classifications were derived and evaluated to provide accuracy data for the comparisons. A color display oriented system (IDIMS) was compared with a printout oriented system (EDITOR) to test the effect of two methods of data/analyst interactions. Supervised, unsupervised, and modified clustering training techniques were compared with each other. Three techniques were evaluated for describing spectral classes: training, screen, and ground plot. The generalized and community levels of detail of vegetation categories were compared. Additional testing was done to determine the effect of post-classification stratification on classification accuracies.

The study area was divided into 36 data blocks for the analysis and evaluation phases. Figure 2.1 shows the data blocks outlined for the study area. Most of the blocks were four square miles, with border blocks being smaller. Every other block in a checkerboard pattern was used for training and spectral class description. The

Figure 2.1. Topographic map of the study area showing training and test blocks.


N
1
2
3
4
5
6
8
9
10
12
13
14
15
16
17
18
20
21
22
23
24
25
26
27
28
29
30
34
35
36
Creek
1000
Creek
Moose
ABM
Windy Cr.
Frenchman Creek
Canyon
1000
Goldstream
ABM
460
BM 600
1000
Happy
Happy
Mines
St. Patrick Cr.
U.S. AGRIC
EXPERIMENT
Tower
UNIV OF
ALASKA
ESTER RD
BM 482
Creek
DITCH
BM
Ridge
Airport
FAIRBANKS

remaining 18 blocks were used as test data for the accuracy evaluation. Each set of blocks (training and test) covered the full range of cover types represented in the study area. The training and test blocks were combined during the field work and photo-interpretation to establish reference data having consistent information content.

2.1 Systems Overview

IDIMS (Interactive Digital Image Manipulation System) is a stand-alone system of software packages run on a HP 3000 - Series II mini-computer. IDIMS is developed and marketed by ESL, Inc. and is available throughout the United States through government agencies and private firms. The IDIMS software is a powerful and flexible package allowing full analysis of digital data from original Landsat computer-compatible tapes through the production and manipulation of a multi-data information system (ESL, 1976). Peripherals at most facilities include terminals, a color display screen, tape drives, discs, a digitizer, various hardcopy output devices and a high speed array processor for large jobs. The IDIMS system is oriented toward the color display screen for analyst/data interactions. The IDIMS software provides capabilities for data entry, radiometric and geometric corrections, a variety of methods for developing training statistics and classifying data, registering ancillary data sets, data manipulation, and output products.

The EDITOR system of software was developed by the Center for Advanced Computation at the University of Illinois. Development and maintenance of EDITOR are now handled by the Institute of Advanced Computation. EDITOR is available through ARPANET, the national computer network of the Advanced Research Project Agency of the Department of Defense. EDITOR provides interactive image processing, file manipulation and batch processing capability through the ILLIAC IV and TENEX computer systems. The EDITOR software is an hierarchical system of commands incorporating a prompt system which is logical and easy to use. Peripherals for using the EDITOR system include terminals, a plotter and a digitizer. The EDITOR system does not incorporate a color display screen, but is oriented toward printer and plotter outputs for the analyst/data interface. EDITOR provides capability for data entry, geometric corrections, a variety of methods for developing training statistics and classifying data, data manipulation and output products. EDITOR provides a versatile and powerful capability for developing classification statistics.

2.2 Training Methods

Three different methods of developing training statistics were used for each of the analysis systems. These methods were: 1) supervised, 2) unsupervised, and 3) modified cluster (Fleming et. al., 1975) or multi-cluster block (Fleming and Hoffer, 1977).

Supervised training uses training fields of known extent and cover type to develop statistics for the classifier. The analyst selects training fields of classes of interest (cover types) using reference data. The training fields are located in the raw data set. The raw data for all training fields of the same cover type are clustered to produce spectral statistics. After editing and merging, the final statistics file is used to classify the entire study area. This method has been generally used when the analyst has prior knowledge and reference data for the study area.

The supervised training fields used in this study were selected from the training and descriptive set of blocks (Figure 2.1). Each training field was identified and outlined on color infrared aerial photographs. The photographs used are listed in Table 1.1. Each training field was transferred to the 1:24,000 base map using a Zoom Transfer Scope.

Unsupervised training uses a random or systematic sample of the raw data to develop spectral statistics. No information is assigned to the spectral classes until after the raw data set is classified. This method is frequently used when the analyst has no knowledge or reference data for the study area prior to classification.

A 10 percent sample was taken within the training/descriptive blocks. The sampled data from all 18 blocks were clustered together to produce spectral statistics. The statistics files were edited and used to classify the data for the study area.

The modified cluster or multi-cluster block approach uses a combination of supervised and unsupervised techniques to develop spectral statistics. Several blocks of data are selected, each block containing several cover types. Each block is clustered separately and each spectral class is identified in each block. The statistics from each block are merged and edited to produce the final statistics file to classify the data set for the study area.

Four of the training/descriptive blocks were used to develop the statistics for the modified cluster training. These were blocks 3, 15, 23, and 31 (Figure 2.1). These blocks covered the range of landcover types represented in the study area. The data in each block were clustered. Each spectral class within a training/descriptive block was identified using color infrared aerial photographs. The statistics files from each of the four blocks were then merged and edited to develop the final training statistics for classification.

2.3. IDIMS Analysis

2.3.1. Data Entry

The coordinates of the study area are approximated in the digital data set by using an acetate overlay of coordinates on a 1:1,000,000 image of the data set. The raw data were reformatted for IDIMS and the study area extracted with the ERTSENTR program. The data were bulk geometrically corrected with CONTROL1 and REGISTER. This process resamples the data to square pixels, rotates the data to north and removes most of the skew caused by data collection. Each band is registered separately and the four bands merged using UNITE to create a four banded image. The data were bulk geometrically corrected to facilitate locating features on the screen. Locating features on uncorrected data is difficult on a display screen. The corner points of the study area and the corners of each of the training/descriptive blocks were located in the data set. The coordinates of each block were recorded for use in the training process.

2.3.2. Supervised Training

The supervised training fields were outlined in TSSELECT using the display screen and trackball cursor. Each training/descriptive block was an ISA (intensive study area). Each ISA was displayed on the screen as a false color infrared image. The training fields were located in the Landsat data as closely as possible using the map and aerial photography. Table 2.1 is a list of the cover types which had training fields. The data from all the training fields of each cover type were clustered together using ISOCLS. For example, all data from deciduous training fields from all the training/descriptive blocks were clustered together to derive spectral statistics for deciduous vegetation. The training cover types are more detailed than the cover types used in the classification and evaluation procedure (Table 2.1), especially in the sparsely vegetated classes. These detailed cover types were not spectrally separable and were

Table 2.1. Cover types of training fields used in supervised training and groupings to community cover types.

Training Field Label	Community Cover Type
Black spruce	Sparse conifer
Dense black spruce	Dense conifer
Black spruce bog	Sparse conifer
White spruce	Dense conifer
White spruce mix	Mixed forest
Mixed forest/bog	Mixed forest
Mixed forest	Mixed forest
Mixed/barren	Mixed forest
Birch	Deciduous
Aspen	Deciduous
Deciduous	Deciduous
Shrub/bog	Shrubland
Shrub/barren	Sparsely vegetated
Grassland	Grassland
Grass/barren	Sparsely vegetated
Herbaceous bog	Grassland
Gravel	Barren
Asphalt	Barren
Clear water	Silty water
Silty water	Silty water

combined for the classification.

The raw data for each cover type were clustered using the ISOCLS routine in TSSELECT. Table 2.2 shows the clustering parameters used in clustering each of the training methods. DIVERGE was run on each statistics file to create a weighted divergence table of each class with every other class in each statistics file. Weighted divergence is a measure of the spectral separability between the classes. Bispectral or COMPARE plots were not made of any of the statistical files because the plotter was not functioning.

The statistics files were merged and edited to create statistics files for each of the community cover types (i.e., asphalt and gravel merged to form barren). These statistics files of each cover type were then merged and edited to form the final statistics file incorporating all the cover types represented in the study area. The final statistics file for the IDIMS supervised training method has 24 spectral classes. The statistics for this file are in Appendix A. The bulk geometrically corrected raw data for the study area were classified using the final statistics file by the IDIMS function CLASFY. CLASFY uses a maximum likelihood algorithm to assign each pixel to a spectral class based on the reflectance values in the four wavelength bands.

2.3.3 Unsupervised Training

The line/column coordinates of the training/descriptive blocks

Table 2.2 ISOCLS parameters for IDIMS clustering

Prompts	Supervised	Unsupervised	Modified Cluster
DLMIN	1	1	1
SEP	0	0	0
STDMAX	2.5	1.5	1.5
ISTOP	12-20	20	20
LNCAT	1	1	1
NMIN	10-30	30	30
MAXCLS	8-15	50	30
KRN	11-19	11	11
CHNTHS	2	1	1

were located in the data set using the display screen. The training/description blocks were copied and mosaicked together (IDIMS function MOSAIC) to form a single image. This image was a four banded image of the raw data from the training/descriptive blocks.

The 10 percent random sample was taken of the data in the mosaicked image. The data in the 10 percent sample were clustered using the IDIMS function ISOCLS. The ISOCLS parameters used are listed in Table 2.2. Weighted divergence values were calculated for all combinations of the spectral classes. The original statistics file of 46 spectral classes was reduced to 21 classes by deleting or combining classes which overlapped each other spectrally. The statistics for this file are in Appendix A. The final statistics file was used to classify the entire data set for the study area.

2.3.4 Modified Cluster Training

Four of the training/descriptive blocks were used to develop a statistics file for the modified cluster training method. These were blocks 7, 15, 23, and 31 (Figure 2.1). Each block was clustered separately with ISOCLS. The clustering parameters are listed in Table 2.2. The spectral classes in each training/descriptive block were identified by displaying the classified image on the screen. The displayed image was compared with color infrared aerial photographs. The spectral classes in each block were edited to increase spectral separability and then the classes from all of the blocks merged.

Similar spectral/informational classes were combined or deleted to create the final statistics file. This file has 17 spectral classes which are listed in Appendix A. This final statistics file was used to classify the entire data set.

These training and classification procedures resulted in three separate IDIMS classifications using supervised, unsupervised, and modified clustering training methods. These classifications were done on the bulk geometrically corrected raw data set. The classifications had to be precision geometrically corrected to produce map products that can be registered to base maps.

2.3.5. Precision Geometric Correction

All three of the classifications were precision geometrically corrected using the GES software and the IDIMS function REGISTER. Forty control points were selected throughout the study area on 1:24,000 scale base maps. The raw data were displayed and enlarged on the display screen as a false color infrared image. The displayed image was compared with the base map and the control points located in the data set. The analyst recorded the line/column coordinates and created a text file of the image coordinates for each control point. The control points were digitized from the map and stored in a GES file. This procedure results in two files of control points: 1) image coordinates and 2) digitized map coordinates. The digitized control points from the map were converted to a 57m grid using

ALLCOORD. TRNSFORM was run using the digitized control points from the map as the source, and the text file of control point image coordinates as the destination. Eleven control points were deleted because location errors on the map or image contributed to large residual values. The final transformation used was a third order transformation between 29 control points from the map and image. The transformation is a link between the line/column location of the control points in the Landsat data and the latitude/longitude locations of the control points on the base map. This transformation was applied to each of the three classifications using the function REGISTER. This process results in precision geometrically corrected classifications.

2.3.6. Final Products

Each classification was scaled to 1:24,000 using CONTROL1 and REGISTER. Line printer maps at 1:24,000 were made using LPMAP. The Fairbanks D-2 NW and SW quadrangles were each printed separately. Every spectral class was shown with a different symbol. Line printer maps were produced for all three of the IDIMS classifications. The line printer maps were used for the spectral class description using ground plot data and for the accuracy evaluations.

2.4. EDITOR Analysis

2.4.1. Data Entry

Three classifications were completed using the EDITOR analysis system available at the Ames Research Center. The EDITOR software is implemented on the BBN TENEX system. Pre- and post-classification processing is done with programs outside EDITOR using the Ames IBM 360 or the ILLIAC IV. The EDITOR commands have a hierarchical structure. The analyst selects one of six major groups of commands and works down to the command needed for a particular analysis step. The command structure for the programs used in this study are fully listed in Table 2.3.

The program RECTF3 was used to reformat the original Landsat data, subsection the study area, and perform a bulk deskewing and north rotation based on latitude of the study area. The HSKEW and VSKEW parameters used for RECTF3 were calculated by EDITOR.

A precision calibration file was constructed to provide a link between the Landsat data and the base maps. Greyscales were printed of the bulk geometrically corrected data output from RECTF3. The same control points which were used for the IDIMS precision geometric correction were used for the EDITOR precision correction. The control points were located on the greyscale printouts of bands 5 and 7 and the line/column coordinates recorded. The control points were digitized from the map and the greyscale coordinates were typed in

Table 2.3. Command hierarchy for programs used in EDITOR processing (Institute for Advanced Computation, 1978).

Geometric Correction
- Parameters for RECTF3:
  - ! Registration and Digitization Functions
    - 2! Skew Correction Parameter Approximation
- Digitize control points and enter grey scale coordinate:
  - ! Registration and Digitization Functions
    - 2! Control Point Location and Analysis

Supervised Training
- Digitize and label training fields:
  - ! Registration and Digitization Functions
    - 2! Segment Digitization
- Plot training fields:
  - ! Plot Functions
    - 2! Segment Field Plots
- Create pack files:
  - ! Ground Truth Analysis Functions
    - 2! Field Selection for Analysis
- Specify number of clusters desired:
  - ! Raw Data Analysis Functions
    - 2! Modify a Window Header
- Clustering:
  - ! Raw Data Analysis Functions
    - 2! Cluster
      - 3! Ordinary Cluster
- Plot bispectral plots for statistics files:
  - ! Plot Functions
    - 2! Ellipse Plots of Statistics Files
- Statistics file editing:
  - ! Raw Data Analysis Functions
    - 2! Statistics File Editing
- Classifier:
  - ! Raw Data Analysis Functions
    - 2! Classify

Table 2.3 (Con't). Command hierarchy for programs used in EDITOR processing (Institute for Advanced Computation, 1978).

```
Unsupervised Training
     Specify training blocks
          ! Registration and Digitization Functions
               2! Window Computation Functions
                    3! Maps Digitization to Specify Windows

     Sample and extract data for clustering:
          ! Subwindow of Window Files
               2! Coordinates
               2! Sample
               2! Write

     Clustering:
          ! Raw Data Analysis Functions
               2! Cluster
                    3! Ordinary Cluster

     Separability matrix and statistics editing:
          ! Raw Data Analysis Functions
               2! Statistics File Editing

     Classifier:
          ! Raw Data Analysis Functions
               2! Classify

Modified Cluster Training

     Training block cornors used from unsupervised training.

     Extract training block data into window files:
          ! Subwindow of Window Files
               2! Coordinates
               2! Write
               2! Close Output File and Continue

     Clustering:
          ! Raw Data Analysis Functions
               2! Cluster
                    3! Ordinary Clustering

     Print out classified maps of each block:
          ! Print Window Files
```

Table 2.3 (Con't). Command hierarchy for programs used in EDITOR processing (Institute for Advanced Computation, 1978).

Merge and edit statistics files:
- ! Raw Data Analysis Functions
    - 2! Statistics File Editing

Classifier:
- ! Raw Data Analysis Functions
    - 2! Classify

for each point using the EDITOR function for control point location and analysis. The control point pairs were analyzed via a least-squares polynominal. The control points which were contributing the greatest error to the fit of the polynominal were deleted. When the errors of fit were less than ± one pixel, the polynominal was output to an image calibration file. This calibration file was tested for accuracy using points of known latitude/longitude and line/column coordinates. This process does not resample the raw Landsat data set to produce a precision geometrically corrected data set. Whenever the analyst needs to select points or polygons in the Landsat data, the points are digitized from maps or photographs. The calibration file is used to calculate the map coordinates in the Landsat data. The Landsat data are bulk corrected similar to the data set used for training on the IDIMS. The calibration file allows the analyst to interact with the data using geometrically correct base maps. The resampling for the precision correction was done on the final classifications.

2.4.2 Supervised training

The training fields were mapped on 1:24,000 base maps in the training/description blocks. The same training fields and cover types were used for EDITOR as were used in the IDIMS analysis. The cover types are listed in Table 2.1. These training fields were digitized and labeled by cover type using EDITOR functions for segment digitization. The training field digitizing was checked by plotting the digitized fields.

The Landsat data from all the training fields of a cover type are extracted from the full data set and copied together into one pack file for each cover type using a function for field selection. Pack files were created for all the cover types. The number of clusters desired are specified for each pack file as part of its header information. The clustering algorithm in EDITOR differs from the ISOCLS algorithm in IDIMS. The EDITOR clustering algorithm parameters are: number of clusters desired (provided with data file), minimum number of categories after merging, percent convergence (percent of pixels not shifting between clusters with each iteration), and maximum number of iterations to achieve desired convergence. In the EDITOR clustering process the number of clusters and percent convergence are the controlling parameters. The number of clusters requested varied with variability of the cover type and the number of pixels in the pack file. Ninety-nine percent convergence was requested for each run with as many iterations as necessary to achieve the convergence. Each pack file, representing one cover type, was clustered. The Swain-Fu separability between clusters is similar to the weighted divergence calculated for IDIMS statistics files. Bispectral plots were plotted for each statistics file for bands 4 and 6. Bispectral plots are graphic representations of the clusters. Each cluster, defined by its mean and variances in the appropriate bands, is plotted for two bands. This is a valuable aid for seeing the size, position, and overlap of

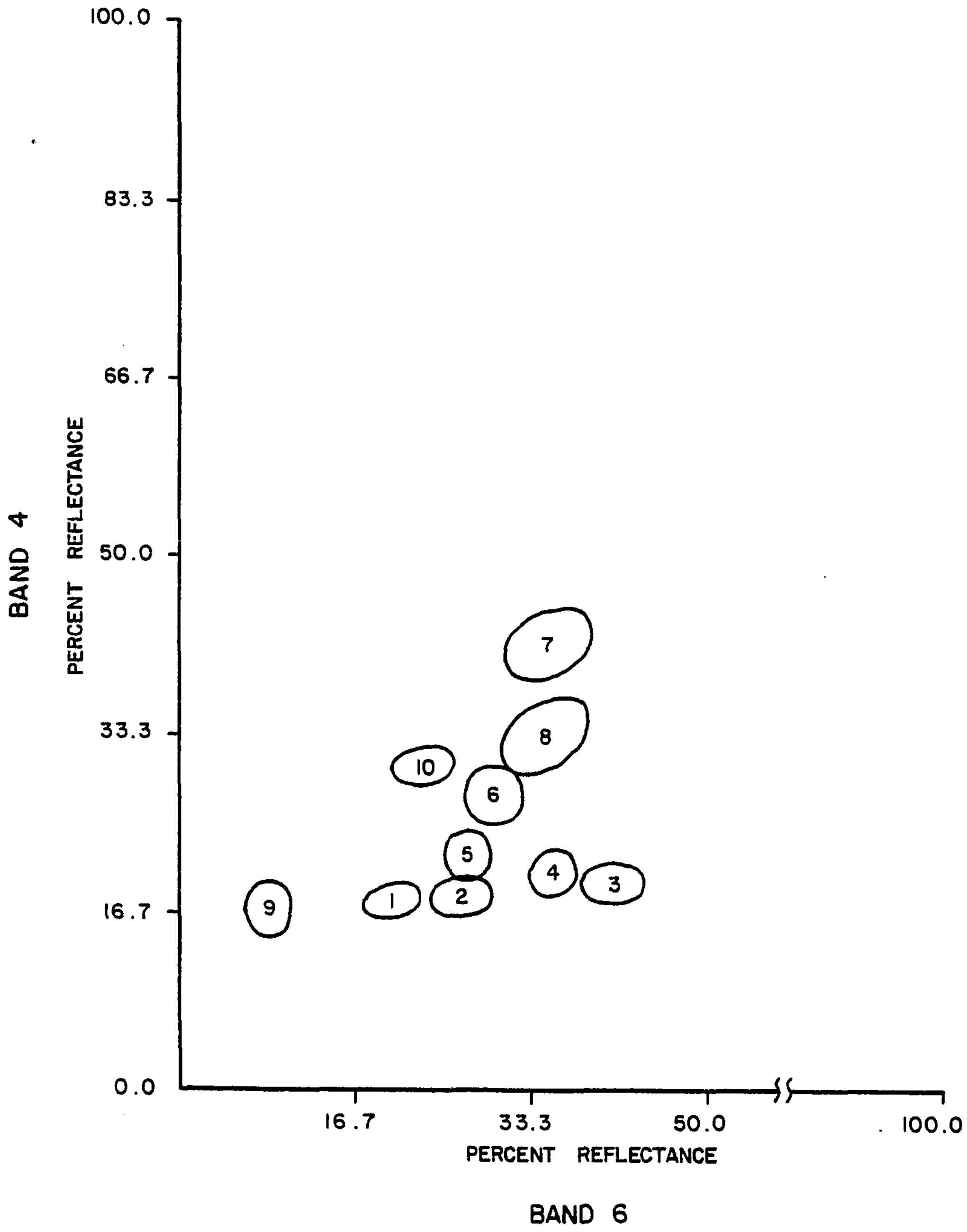


Figure 2.2. Bispectral plot of final statistics used in EDITOR supervised classification.

clusters in a statistics file. The individual cover type statistics files were edited and combined into the cover type groups in Table 2.1. The community cover type files were merged and edited to create the final statistics file for classification. The statistics for this file are listed in Appendix A. The bispectral plot of the final statistics file are shown in Figure 2.2. This statistics file was used to classify the Landsat data for the study area using the Gaussian maximum likelihood classifier.

2.4.3. Unsupervised Training

The training/description blocks were digitized using the 1:24,000 base maps. The corner coordinates of each block were calculated using the precision calibration file. The Landsat data for the training/description blocks were extracted, sampled every third line and column and written to a window file. The resulting data file contained an 11 percent sample of the Landsat data in the training/description blocks. This data set corresponded to the 10 percent random sample used for the IDIMS unsupervised training. The sampled data set was clustered with the EDITOR ordinary clustering function. Twenty-five classes were requested with 100 percent convergence and as many iterations as necessary to attain the 100 percent convergence. A separability matrix was calculated using Swain-Fu distance and bispectral plots made of band 4 with band 6. The statistics file was edited to 14 classes by deleting and combining classes to increase spectral separability.

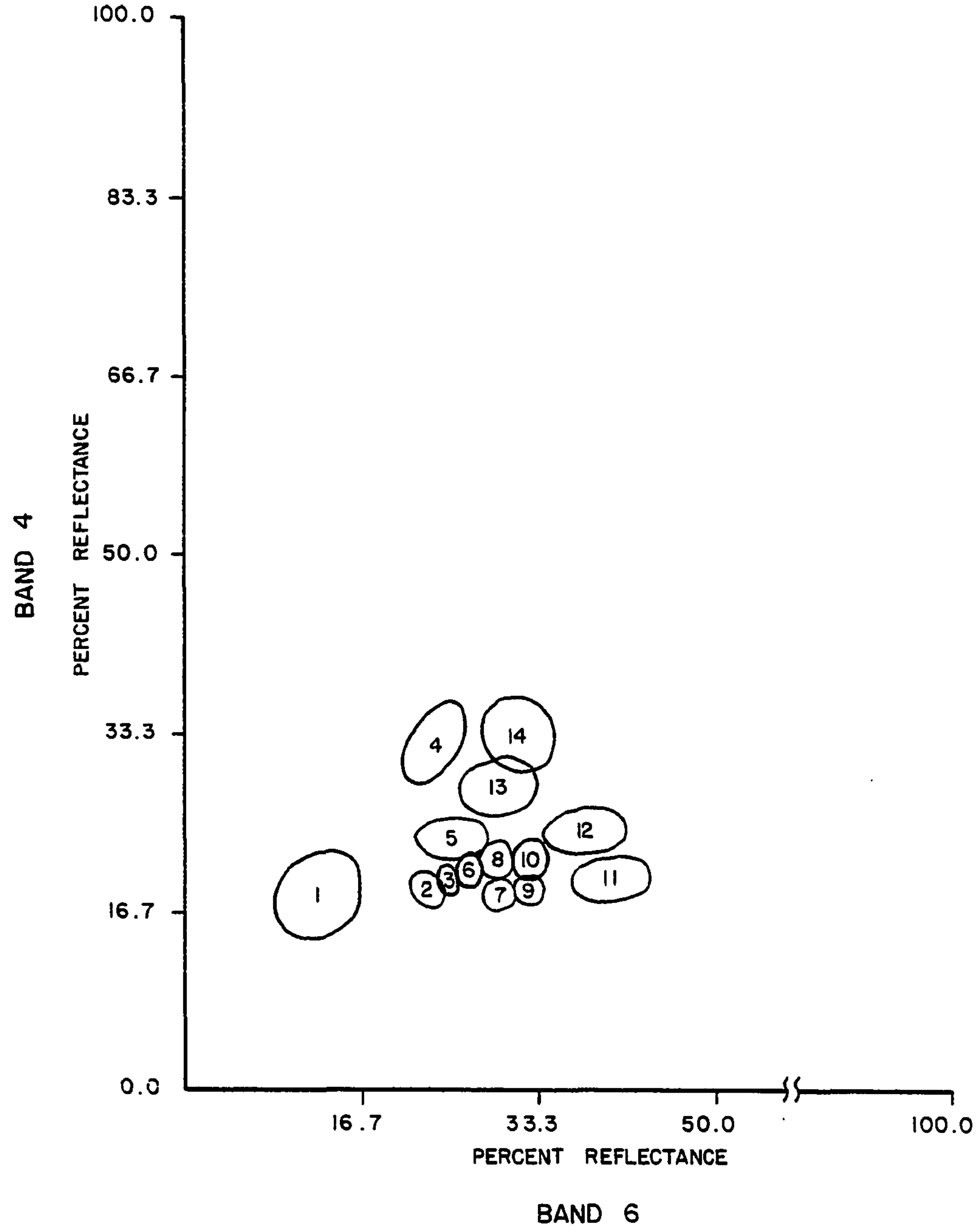


Figure 2.3. Bispectral plot of final statistics used in EDITOR unsupervised classification.

The final statistics file used for the classification is listed in Appendix A. The bispectral plot for the final statistics is shown in Figure 2.3. The Landsat data set for the study area was classified using the final statistics file and the Gaussian maximum likelihood classifier.

2.4.4. Modified Cluster Training

Four of the training/description blocks were used to develop the modified cluster training statistics. These were blocks 7, 15, 23, and 31 (Figure 2.1) and were the same blocks used in the IDIMS modified cluster training. Each block contained approximately 2300 pixels. A window file was created from the Landsat data for each of the training/description blocks used for modified cluster training. The coordinates for the blocks were calculated for the unsupervised training. Landsat data for each of the four blocks were created into separate window files. The data for each window file was clustered separately. The number of clusters requested varied with the vegetation complexity of the training/descriptive block. The number of clusters per block varied from 10 to 16. Ninety-nine percent convergence was requested for each block, with as many iterations as necessary to achieve the 99 percent convergence. Each block was clustered several times with different numbers of clusters requested each time.

A categorized window file was created for each block in the clustering function. This makes a mini-classification of the Landsat data in the block using the statistics from the clustering process. Each categorized training/description block was printed out with every spectral class. Each spectral class was identified with color infrared aerial photographs using a Zoom Transfer Scope. Bispectral plots and Swain-Fu separability tables were made for each statistics file. The set of clusters for each training block which best followed the vegetation patterns was selected for editing and inclusion in the final statistics file. The statistics file for each training/description block was edited separately. The four statistics files were then merged and similar spectral/information classes pooled or deleted to create the final statistics file. The final file has 11 spectral classes and is listed in Appendix A. The bispectral plot for the final statistics file is shown in Figure 2.4. The final statistics file was tested on small blocks of the Landsat data before the entire data set was classified with the Gaussian maximum likelihood classifier.

2.4.5. Precision Geometric Correction and Final Products

The procedures discussed above resulted in three classifications of the bulk corrected Landsat data over the study area. The actual classifications were not precision geometrically corrected. The classified images were transferred from the BBN Tenex which had done the EDITOR processing to the ILLIAC IV at Ames. The precision reformatting (geometric correction) of all three classifications was

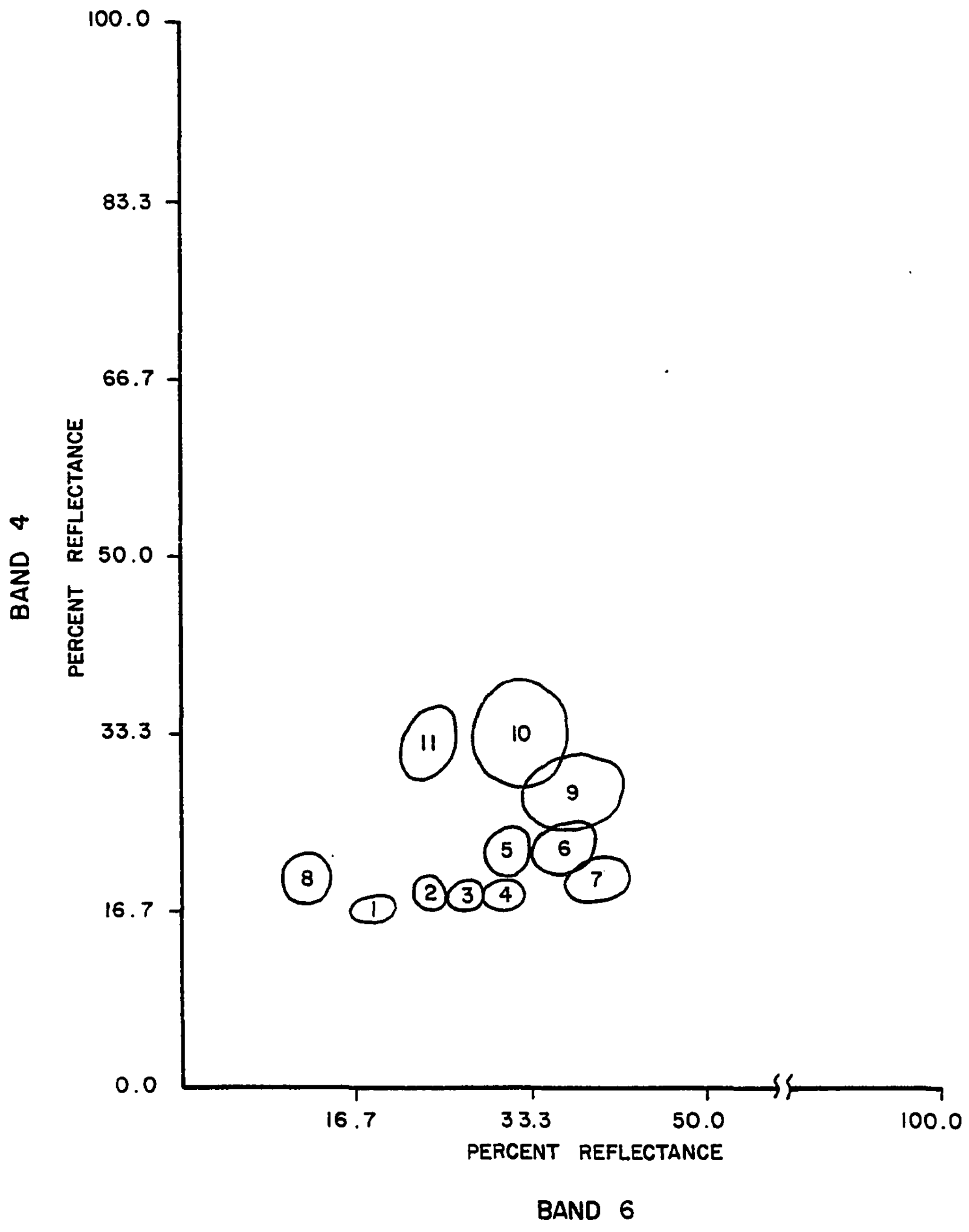


Figure 2.4. Bispectral plot of final statistics used in EDITOR modified cluster classification.

done on the ILLIAC IV. Printouts were made of the study area for every spectral class at the scale of 1:24,000 for all classifications. These line printer maps were used for the spectral class descriptions using ground plot data and for the accuracy evaluations.

2.5. Reference Data

Reference data are used throughout a computer-aided analysis project. A reconnaissance is conducted during the early part of a project with Landsat images and aerial over-flights. Topographic maps are used to select and digitize control points for geometric correction. All sources of reference data are used for training field selection, cluster identification, spectral class descriptions, stratification, and for the test data to evaluate the accuracy. Additional information is combined with the finished classification for specialized interpretations.

A Landsat data set is only a matrix of reflectance data. It is entirely possible to produce a color coded classified map without reference data. The result is no better than a pretty picture. The value of the map comes from accurate descriptions of the spectral classes in terms of resource classes of interest. The analyst needs to be familiar with the spectral characteristics of the cover types in the study area, the computer system being used, the distribution patterns of the cover types and factors influencing this distribution. Reference data are necessary for the analyst to become familiar with the cover types in the study area, their location, and distribution throughout the study area. There are many different sources of reference data for a computer-aided analysis. These data types include field work, aerial photographs, Landsat images, over-flights, topographic and thematic maps, and reports by previous investigators.

Different types of reference data are necessary for the various phases of the analysis procedure. Field data gathered for some other project such as inventory or timber survey generally are not satisfactory for a digital analysis. The characteristics of the Landsat data and scanner system must be considered when planning acquisition of reference data:

1. Landsat sensors only record light reflected from the uppermost layer of vegetation or landcover type. Reference data should be gathered for the uppermost 100 percent crown cover. Understory and ground cover information is only useful for preparing detailed spectral class descriptions if these species are consistently associated with the overstory.

2. The Landsat scanner averages reflectance information for one pixel - approximately 1.12 acres. Reference data should be gathered for areas larger than one pixel and preferably five acres or more. Plot or releve data such as gathered for community analysis or inventories are not valid for Landsat analysis projects.

3. Landsat data are only spectral reflectance data. The cover types of interest have to be defined by spectral characteristics as well as floristic characteristics to be mapped from Landsat data.

Very detailed landcover categories cannot be separated using Landsat. Reference data of Level V of the Alaskan Vegetation Framework (Viereck and Dyrness, 1980) will have to be generalized to broader categories to be valid for a Landsat analysis project.

All sources of reference data listed above were used for this study. The reference data for landcover information were derived from a combination of two data sources: direct observation from the ground and light aircraft, and photo-interpretation of 1:60,000 color infrared aerial photographs which are available for the entire study area. Identification of cover types is best accomplished from ground observation, and determinations of crown closure and extent of stand are best accomplished from photo-interpretation. The purpose of the field work was to give the photo-interpreter a sound ecological basis from which to identify and map the cover types occurring in the study area from aerial photographs. Specific sites were ground visited, the cover type identified and the extent of the stand mapped.

2.5.1. Landcover Coding System

When a project begins, the analyst usually has a list of desired landcover categories. This list or framework is usually based on non-spectral characteristics such as vegetation communities, land use, or habitat types. In the ideal situation, the breaks in the spectral data or classes would correspond to the original list of landcover categories. This seldom happens. The analyst may then force the spectral classes into the landcover category that the class most closely resembles. This process generally results in low accuracies. The alternative is for the analyst to completely describe the cover types that occur in a spectral class. This process generally results in higher accuracies. A flexible coding system is needed to record landcover information during field work and photo-interpretation. The analyst can combine these reference data into categories that correspond to the spectral classes for more accurate interpretations during the analysis procedure. Such a flexible coding system is particularly valuable for reference data used for training, spectral class descriptions, stratification, and the accuracy evaluation. A good coding system will not commit the analyst to a static set of landcover categories.

The coding system developed in this study identifies the cover type in a stand by the major overstory species present, the total crown cover, and the crown cover of each species in a mixed species stand. A numeric or alpha-numeric key similar to that in Table 2.4 is developed for the major overstory species or landcover types in the study area. The code in Table 2.4 has two levels of detail: a generalized level for non-vegetated lands, grassland types, shrubs and forest lands, and a detailed species level.

Each major species or cover type has a unique identifier. A preliminary list is prepared after reconnaissance and before the field work or photo-interpretation. If additional species or landcover types are encountered, they can be added to the list without invalidation of

Table 2.4 Cover type code for field work and photo-interpretation.

00		Bare rock, soil and asphalt
01		Water
1		Grasslands
	-1	Sedges/mosses/ericaceous shrubs
	-2	Grasslands - monospecies
	-3	Herbaceous - herbs/grasses
2		Shrublands
	-1	Cottonwood/willow - riparian regrowth
	-2	Alder/willow - drainages
	-3	Birch/aspen - regrowth
	-4	Dwarf birch/low willow
3		Forestlands
	-1	Deciduous - birch/aspen
	-2	Cottonwood - floodplain
	-3	White spruce
	-4	Black spruce
	-5	Tamarack

Example:

3 - 13 (10) 7, 3

- mixed forest of deciduous (70% crown closure) and white spruce (30% crown closure) for a total of 100 percent crown closure.

- scattered residences in stand.

the data collected before the new item was added. If a large number of species are anticipated, they can be arbitrarily divided into alpha and numeric designators or units of some other system. When this coding system was used in the San Juan Mountains, Colorado (Krebs and staff, 1976), there were nine coniferous species with numeric codes and three deciduous species with alpha codes.

The ground cover and brush types can be coded by individual species or community types based on the analyst's knowledge of the vegetation of the study area. When determining whether brush, herbaceous species, or mosses and lichens should be recorded by species or by community, the analyst needs to consider: 1) the physical size of the plant in the study area, and how the plant size and shape relate to the working resolution of the remote sensing data source, 2) whether the species grow in well defined communities or mix with each other in an infinite variety of plant assemblages, and 3) whether the species in question form part of the uppermost 100 percent crown cover or grow as understory to other species.

The forest trees are usually large enough to be identified separately in the field and on aerial photographs. Conifer species can usually be reliably identified. Deciduous species in interior Alaska could not be reliably separated during a summer field season, so birch, aspen, and cottonwood were combined into one deciduous category. Some species such as black and white spruce have large ranges in growth form and ecological conditions. Their distinguishing traits cannot be accurately determined at a distance in the field or on medium scale aerial photographs. The analyst may define "ecological taxa" based on growth form and environmental conditions rather than taxonomic features. For example, in this study, black and white spruce were defined by growth form and location rather than cone length and pubescence on young twigs.

Non-vegetated landcover types must be considered during preparation of the code. Depending on the study area, categories such as snow, ice, water, and barrens should be included. These can be further subdivided into categories of clear and silty water, sand, gravel, bedrock, and asphalt. On some remote sensor data, categories should also be included for non-landcover units such as clouds, haze, smoke, and cloud or topographic shadow.

Categories such as land use or habitat generally do not work well in computer-aided classifications, and should not be the primary categories in the reference data coding system. Land use and habitat are interpreted from the base landcover maps. "Residential" does not tell the landcover type found; and the mixture of driveways, lawns and sparse trees may be spectrally similar to a sparsely vegetated sandbar. Airports cannot be separated from parking lots or some bedrock types. Pastures are similar to golf courses (recreational) and herbaceous tundra types. Wetlands are also complex mosaics of several vegetation communities and are easily confused spectrally with non-wetland types. If further interpretations are to be made from the landcover map, notes of different land uses or habitats should be made, but these categories should not be the only identification for a stand.

After a stand has been identified by landcover type or vegetation, the total crown closure and the crown closure of each species should be recorded. Experience has shown that an experienced ecologist or photo-interpreter can differentiate 10 percent differences in crown closure between species and for total crown closure. Cover determinations by experienced ecologists are usually within 10 percent of each other, with occasional records differing by 20 percent. For less detailed work, 25 percent crown cover breaks may be adequate. Crown cover should not be recorded in predetermined categories such as 1-25 percent, 26-40 percent, 41-75 percent and 76-100 percent. Such a set of categories forces the data into static units which may or may not correlate to the spectral classes.

The cover type coding developed in this study has two parts: 1) a total crown closure for the major species, in parentheses, followed by 2) crown closure for each species, in decreasing order of percent crown closure. For stands combining forest, brush, and/or ground layers of vegetation, a separate line was recorded for each layer with the total crown closure for all layers equal to 100 percent.

Miscellaneous notes were also recorded for each stand. These notes include the kind of inclusions, other species (usually trees) with cover less than 10 percent, roads and residences, visible disturbances, and other observations. An example of the information recorded for a stand is in Table 2.4.

2.5.2. Field Work

Field work is the foundation of all the other reference data types. A project can be completed without aerial photographs or thematic data if adequate field work has been done. It is one of the most expensive parts of a remote sensing project and needs to be carefully planned to be efficient and effective. Field work is used alone or combined with photo-interpretation for training, spectral class descriptions and refinement, stratifications, and the accuracy evaluation. There may be several levels of intensity required for field work.

1. Reconnaissance, to gain knowledge of the entire area, including an overview of cover types, their general location and major features. The reconnaissance is best accomplished in a light aircraft, or a helicopter, if the project is well funded.
2. The medium level of intensity involves actual ground work. Specific stands of vegetation are identified and located on maps and/or aerial photographs. These data are used during the actual analysis procedure.
3. Additional field work may be necessary to calculate estimates for attributes such as biomass. This is usually very detailed work on small areas and is extrapolated to pixel sized units. Information such as biomass, productivity, board feet, or carrying capacity can be combined with area estimates from a completed classification.

The field data points for this study were selected by overlaying a map of the Fairbanks study area with a grid of points 1.2 km (.75 mile) apart. The points were marked on 1:24,000 base maps for location. Each point was assigned a unique number for the study area. These points were considered random with reference to the Landsat data and the vegetation since neither of those data sources influenced the location of the points. There is a total of 360 data points for the entire study area. Due to time and budget constraints, only a sample of the points could be ground visited. A preliminary sample of 35 points was field checked and then photo-interpreted. The probability of correct photo-interpretation as compared with ground reference data was calculated using formula 2.1.

$$\hat{p} = \frac{\text{\# of species correctly photo-interpreted}}{\text{total \# of species occurrences in sample}} \qquad \text{Formula 2.1}$$

$$\hat{p} = \frac{69}{73^*} = .945$$

*Most of the data points had more than one species present.

The number of samples to be field checked was calculated using formula 2.2.

$$n = \frac{Z_{\alpha/2}^{2}\,\hat{p}\hat{q}}{e^2} \qquad \text{Formula 2.2}$$

Where $Z_{\alpha} = .05 = 1.96$

$q = 1-\hat{p}$

e^2 = allowable error

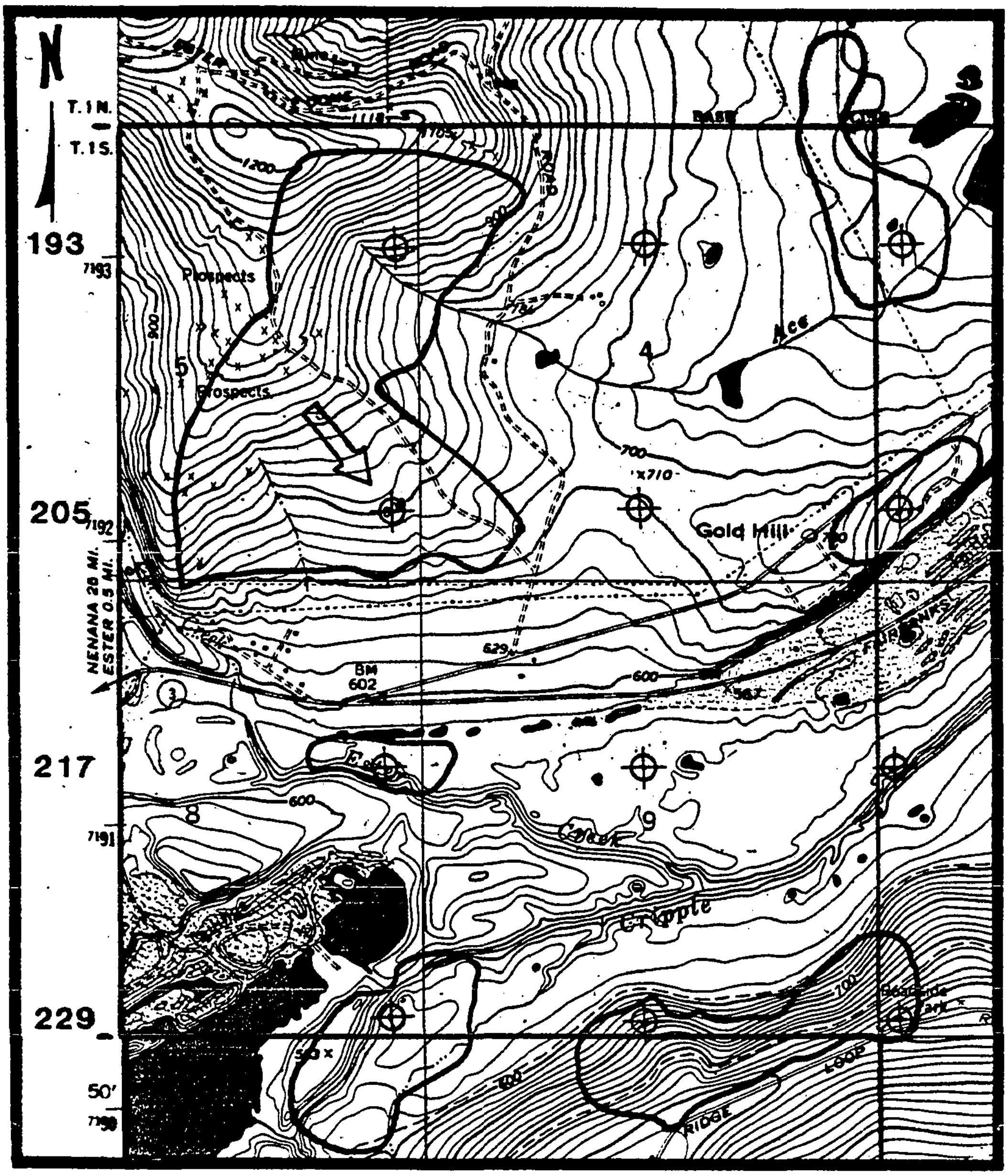
$$n = \frac{(1.96)^2\,(.95)\,(.05)}{(.05)^2} = 72.99 = 73 \text{ points}$$

Seventy-five ground plots were selected randomly from the 310 points north of the Tanana River. If any of the original randomly selected points could not be observed in the field they were replaced by additional randomly selected points. Seventy-five additional points were selected subjectively for field visitation on the basis of accessibility and the need to completely cover vegetation types that were not adequately represented in the random sample. The subjectively selected points included areas south of the Tanana River which were observed from low flying light aircraft. The subjective points were not all chosen prior to the field work but were selected or added during the course of the field season to fill gaps or to strengthen interpretation of some cover types.

The color infrared aerial photographs over the study area were enlarged to 1:24,000 scale. The data points were located by corresponding topographic and vegetation features on the ground with the 1:24,000 scale maps and enlarged photographs. Data were collected for each point by going to the actual site of the point, by observing the area from a high vantage point, or by both methods.

After a data point was located on the ground, the cover type represented by that point was determined. The vegetation cover was identified by major tree and brush species or ground cover. Total crown closure and species crown closures were estimated in 10 percent increments. The cover type was given a numeric designation according to the cover type code developed for the study (Table 2.4) and recorded (Section 2.5.1). For each data point, the cover type, total crown closure, crown closure by species breakdown, understory if the overstory is less than 100 percent, data point number, and miscellaneous notes were recorded. A boundary was drawn on the map showing the observed extent of the cover type represented by the data point. This boundary was drawn using ground observations and the enlarged aerial photographs while the observer was in the field. The cover type may extend beyond the boundary shown, but the area enclosed indicated that it had been directly observed. The area within the boundary was homogeneous with respect to species composition, total crown closure and community composition. Inclusions of other cover types of one acre were often included within the boundary but were noted in the miscellaneous notes for each point. An example of the mapped cover type and recorded data for each point is shown in Figure 2.5. Field data were collected for 193 points during the summer of 1979.

Several potential problems were recognized during the field work which would influence photo-interpretation accuracy. One problem concerned identification of sparse grass, forbs and shrubs on sandbars in the Tanana River. There was more vegetation cover on some of these sandbars than could be seen on the photographs. Destruction and formation of sandbars resulted in many locational differences between the base map made with 1949 photographs, the Landsat data collected in 1976, the infrared aerial photographs acquired in 1978, and the field work conducted in 1979. The sandbars were mapped as accurately as possible using ground observations and the 1978 aerial



point #: 205 . date: 6/16/79

cover type code: 3-1(10)

large deciduous stand with scattered white spruce

Figure 2.5. Example of the data collected in the field for each sample point. The cover type code follows Table 2.4. This block is near Ester.

photographs.

Cover types in cleared areas in the Goldstream valley, areas east of Chena Ridge, and between the Chena and Tanana rivers were difficult to identify from the aerial photographs. The areas were cleared by bulldozers, apparently in the late 1950's and early 1960's, for homesteads or burned fairly recently. The clearings and burns have regrown to a variety of cover types from dense birch and aspen regrowth stands of varying heights on better drained sites to moss, lichen and ericaceous shrubs with varying densities of black spruce in the poorly drained sites.

Another problem encountered was in determining crown closure of black spruce from the aerial photographs. The sparse crown closure and growth form of some black spruce with spindly tops and spreading branches near the ground made interpretation of black spruce stands difficult. Even with stereo viewing under magnification, sparse stands of black spruce can not be identified and dense stands look like sparse stands. The analyst needed additional field work in areas of black spruce to calibrate the identification of black spruce densities on the aerial photographs. Additional ground plots were visited in these problem areas to improve the interpretation of these cover types on the color infrared aerial photographs.

2.5.3 Photo-interpretation

The major photo-interpretation effort was after the 1979 field season. The primary data source was 1:60,000 color infrared aerial photographs acquired under the Alaska High Altitude Aerial Photography Program during July and August 1978. Frame numbers 146 - 148, 318 - 320 from roll number 23 and frames 5606-7 from accession number 2655 were used. Ancillary photographs are listed in Table 1.1.

Second generation positive color transparencies from the Bureau of Land Management Branch of Photogrammetry were used for the photo-interpretation. A Bausch and Lomb variable magnification stereoscope on a MIMS-3 light table was used for stereoscopic viewing. The grid of data points was transferred from the base maps used for field work to the corresponding location on 1:24,000 acetate duplicates of the base map. The 1:24,000 enlargements of the aerial photographs were valuable aids in the photo-interpretation phase.

Each data point was located on the 1:60,000 color infrared photographs using stereo viewing, by corresponding topographic features from the base maps, and color and tone characteristics from the enlargements. After the point was located, the cover type represented by the point was identified. Identification was made on the basis of color, tone, texture, spatial feature, location of topographic features and the interpreter's knowledge of the vegetation patterns and the controlling ecological parameters. The data and maps from the field work were consulted throughout the photo-interpretation. The extent of the cover type at each point was mapped on the acetate base map. The acetate base map was registered to the 1:24,000 enlargement and the polygon drawn on the map corresponding to the stand on the photo enlargement. The delineated cover type unit was reasonably homogeneous with respect to species, total crown closure and percent cover of each species as in the field work mapping. All 360 of the data points located in the study area were identified and mapped from aerial photographs.

After the photo-interpretation was completed the cover type identifications were generalized to the community cover types listed in Table 2.5. The photo-interpreted identifications were compared with the field work cover type identifications. The photo-interpretation was considered wrong if the different species changed the community cover type categories. Ninety-two percent of the points were correctly identified. As anticipated, the problems were with the black spruce cover estimates. The photo-interpreted identifications were corrected where necessary.

2.6 Spectral Class Descriptions

Description of the spectral classes is one of the most important steps in a project involving computer-aided analysis techniques. This step requires that the analyst merge his knowledge of the machine processing, spectral response patterns of landcover types, and ecological knowledge of the vegetation in the study area. The analyst assigns information labels to the spectral classes which the computer has developed through the clustering process. The spectral/informational classes may then be grouped into map categories.

Table 2.5. Community cover type categories.

Covertype	Label	Description
Dense conifer	A	70-100% black or white spruce, includes up to 20% deciduous, 0-30% understory.
Sparse conifer	B	40-60% spruce, includes up to 20% deciduous, remainder understory. Predominately black spruce with scattered tamarack.
Deciduous	C	Greater than 40% deciduous trees, up to 20% conifer. Most stands were greater than 60% cover with herbaceous and shrub understory.
Mixed Forest	E	Includes all mixed conifer/deciduous forests greater than 40% cover. 70% conifer/30% deciduous through 70% deciduous/30% spruce.
Burn	F	1975 burn south of Tanana River.
Shrublands	G	Shrub areas with less than 40% tree cover. Primarily dwarf birch/willow in lowland areas and birch/aspen/cottonwood regrowth in clearings and floodplain.
Grasslands	H	Herbaceous ground cover with less than 40% tree cover. Primarily cultivated or recently cleared fields.
Sparsely vegetated	I	Grass and shrublands with 20-50% ground cover. Primarily sandbars and residential areas.
Barren	K	Barren areas with less than 20% vegetative cover. Gravel, sand and asphalt areas.
Silty water	L	Tanana River with heavy silt load.
Clear water	M	Clear water streams, lakes and impoundments.

There are several ways to develop descriptions for spectral classes. The clusters may be identified during the training procedure as in supervised training. A classified image can be viewed on an interactive color display screen and the spectral classes identified from the screen. A sample of pixels may be taken from the classification and those pixels described. The classification may also be compared with a sample of ground based reference data and descriptions developed for each spectral class. Spectral class descriptions use qualitative intuitive methods, elaborate quantative methods, or combinations of both. This study compares three methods of deriving spectral class descriptions: 1) training, 2) identification on the color display screen, and 3) identification with known ground plots using scaled printout maps. For each method the cover type identifications were structured into the community level cover type categories in Table 2.5. The community level categories were further grouped into the generalized level categories in Table 2.6. This resulted in community and generalized categories for the three spectral class description methods for each of the three training methods for the IDIMS and EDITOR analysis systems, or a total of 26 sets of spectral class descriptions. The class by class descriptions for each of the techniques are in Appendix B.

2.6.1. Spectral Class Descriptions from Training

With supervised training techniques, spectral classes are

Table 2.6. Cover type Groupings.

Generalized	Community
1. Conifer Forest	A. Dense conifer forest B. Sparse conifer forest
2. Deciduous Forest	C. Deciduous forest
3. Mixed Forest	E. Mixed forest
4. Brush and Grasslands	G. Shrublands H. Grasslands
5. Burn	F. Burn
6. Sparsely Vegetated & Barrens	I. Sparse grass and brush J. Sparse grass K. Barren
7. Water	L. Silty water M. Clear water

identified before clustering. The techniques used for identifying training fields in supervised training for IDIMS and EDITOR are discussed in Section 2.2. The cover type assigned to the original training field was structured into the community cover type category (Table 2.5). The clusters in the final statistics file were described by the cover type of the original training field. This technique does not involve any further refinement of the spectral class descriptions after the initial labeling of the training fields.

In modified cluster training the spectral classes are described after the clustering but before the editing and classification of the entire data set. The process of modified cluster training for IDIMS and EDITOR is discussed in Sections 2.3.4 and 2.4.4. The community cover type labels assigned to each cluster were used to describe the spectral classes after classification. There was no further work on the spectral class descriptions after the initial labeling of the training blocks. The spectral classes for each community cover type were grouped to develop the generalized cover type categories. The spectral class descriptions are listed in Appendix B.

Unsupervised techniques do not describe the spectral classes until after the classification of the study area. There were no spectral class descriptions from training for the unsupervised classifications.

2.6.2. Spectral Class Descriptions from the Color Display Screen

The primary data/analyst interface on the IDIMS system is through the color display screen. The supervised, unsupervised, and modified cluster classifications from IDIMS were each displayed on the color display screen. The color display screen is a versatile method for interpreting data sets. All or part of the data set can be viewed at various enlargements. Spectral classes can be uniquely color coded, either singly or combined as the analyst desires.

The entire study area was displayed to interpret general patterns for each spectral class in each classification. Detailed work was done by displaying and interpreting the classifications in the training/description blocks. Preliminary screen descriptions were attempted using aerial photographs before the intensive field season. After the field season, the classifications were re-displayed. The spectral class descriptions were verified or changed as necessary based on the experience of the field season and the 1978 aerial photographs.

The community cover types were grouped for the generalized descriptions. The burn south of the Tanana River was stratified as an additional cover type. Spectral classes which included the burn area and another cover type were stratified by location in the study area. This meant that one class could be labeled a cover type such as dense conifer for most of the study area and labeled as burn in the burned area. All other spectral classes were assigned one cover type. The EDITOR classifications were not displayed on the color screen as this is not available in the standard EDITOR processing.

2.6.3. Spectral Class Descriptions from Ground Plots

The spectral classes from the supervised, unsupervised, and modified cluster classifications for both the IDIMS and EDITOR systems were described using reference data from the ground points. The information from the photo-interpretation and field work was used for the training/description blocks. The cover type identifications were structured into the community categories. The 1:24,000 scale acetate base maps were marked with the data points, the surrounding cover type polygons and the cover type identifications. The 1:24,000 printout maps of each classification were registered with the acetate base map (Figure 2.6). Every spectral class was represented by a different symbol on the maps.

The symbols (spectral classes) occurring within each cover type polygon were recorded with each point number and the reference cover type class. If a spectral class distribution pattern corresponded closely with the cover type pattern or was a dominant class in the polygon it was recorded as a plus. Other symbols (spectral classes) were recorded as a check.

The programs of the Statistical Package for Social Sciences (SPSS) (Nie et. al., 1975) were used to reduce and analyze the ground plot descriptive data. SPSS was developed at Stanford University, California and the National Opinion Research Center at the University of Chicago. The set of programs is available on the University of Alaska Computer Network. The primary functions used in the analysis

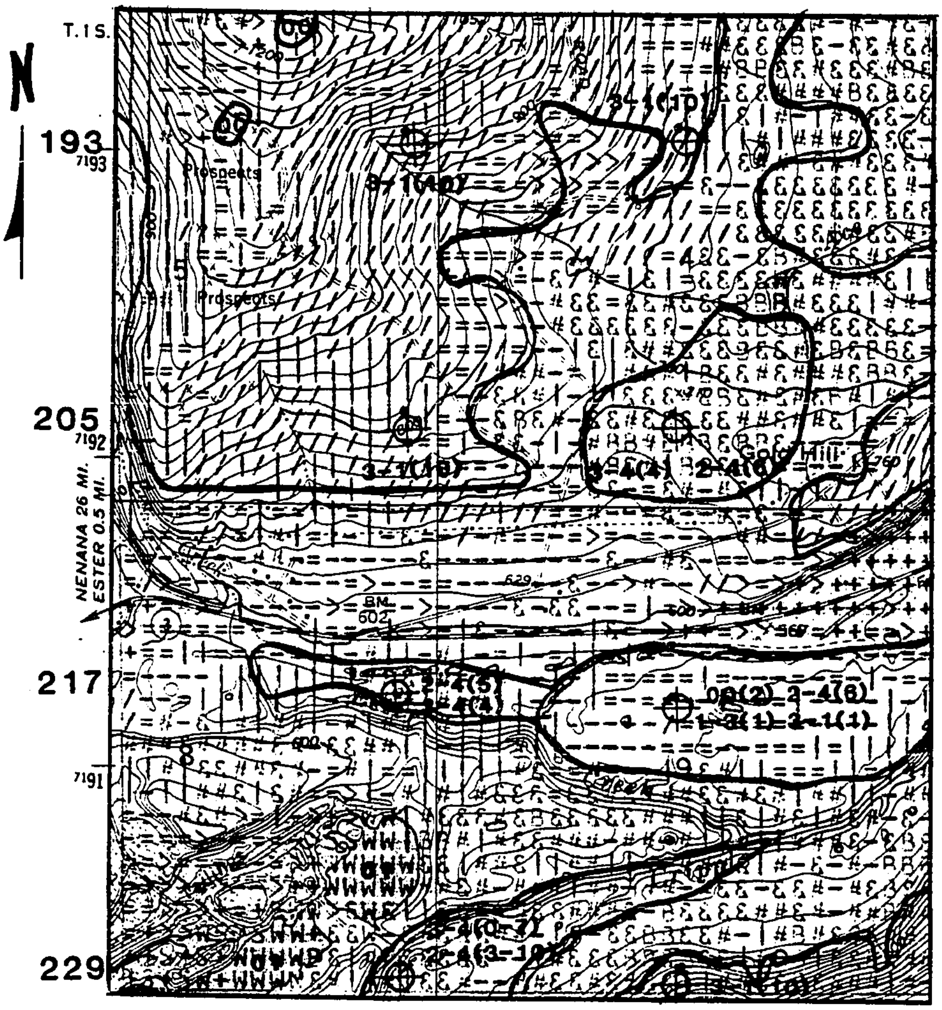


Figure 2.6. Example of the printout maps of the classification registered with the reference data. This block is located near Ester. The cover type code follows Table 2.4. This example shows the EDITOR modified cluster classification.

were IF, RECODE, FREQUENCIES, and CROSSTABS.

A confusion matrix or cross-tabulation table was created for each of the six classifications being considered. Plus and check values were tabulated separately. An example of the cross-tabulation tables is shown in Table 2.7. The data from the plus and check cross-tabulation tables were combined into a single table. The number of plus occurrences of a spectral class in each cover type was weighted by two. This weighted value was added to the number of check occurrences of the spectral class in each cover type. These calculations resulted in an occurrence value for every spectral class in each cover type. An example is shown in Table 2.8. Each spectral class was described as the cover type with the highest value. The burn south of the Tanana River was stratified as an additional cover type in that area and added to the spectral classes where it was strongly represented. All other spectral classes were assigned to one cover type. If the occurrence values were equal between two cover types for a spectral class, the cover type with the highest frequency of ground points was assigned to the class. In cases where the spectral classes did not adequately distinguish between similar cover types, the occurrence values are similar. These cases are good indications of low classification performance results when the classifications are evaluated.

The process was repeated for the generalized level of cover type detail by combining the reference cover type calls according to Table 2.6. The ground data identification for each plot were grouped into

Table 2.7. Example of cross-tabulation tables used for ground plot spectral class descriptions. These data are plus occurrences for the EDITOR Modified Cluster classification, generalized mapping categories.

	Ground Data Cover Type						
Spectral Classes	Conifer Forest	Deciduous Forest	Mixed Forest	Brush & Grasslands	Burn	Barren & Light Veg.	Water
1	15		1				
2	20		2	1	6		
3	9	1	4	1	6		
4		15	8	1	1		
5	6	1	2	10	12		
6		7		5			
7		18	2				
8							3
9		2		3		2	
10		1		3		10	1
11						2	4

Table 2.8. Example of occurrence data used to describe spectral classes with ground plot data. These data are from the EDITOR Modified Cluster classification with generalized map categories. All plus values have been weighted by 2.

	Ground Data Cover Type						
Spectral Classes	Conifer Forest	Deciduous Forest	Mixed Forest	Brush & Grasslands	Burn	Barren & Light Veg.	Water
1. plus	30		2				
check	16		3	1	1		2
Total	46		5	1	1		2
2. plus	40		4	2	12		
check	25	6	12	7	8	2	
Total	65	6	16	9	20	2	
3. plus	18	2	8	2	12		
check	29	18	13	9	6		
Total	47	20	21	11	18		
4. plus		30	16	2	2		
check	12	26	9	11	8		
Total	12	56	25	13	10		
5. plus	12	2	4	20	24		
check	13	17	5	11	5	2	
Total	25	19	9	31	29	2	
6. plus		14		10			
check	3	29	5	21	7		
Total	3	43	5	31	7		
7. plus		36		4			
check	3	19	4	9	4		
Total	3	45	4	13	4		

Table 2.8. (Cont'd)

Spectral Classes	Conifer Forest	Deciduous Forest
8. plus		
check		
Total		
9. plus		4
check	1	6
Total	1	10
10. plus		2
check	3	3
Total	3	5
11. plus		
check		1
Total		1

Ground Data Cover Type				
Mixed Forest	Brush & Grasslands	Burn	Barren & Light Veg.	Water
		1 1		6 1 7
2 2	6 7 13	1 1	4 3 7	
2 2	6 1 7		20 1 21	2 1 3
1 1		1 1	4 3 7	8 6 14

generalized categories. The cross-tabulation table was constructed for the grouped cover types and each spectral class. The spectral class assignments were made as for the community level of detail.

2.7 Accuracy Evaluations

The objective of the evaluation was to assess the accuracy of each classification and each spectral class description method. The various methods could then be compared to demonstrate if some techniques are more effective than others. Test field data from the test blocks (Figure 2.1) were used for the evaluation. This was a completely separate set of data than the data which were used for the training and description phases of the analysis.

The test data points were identified and mapped at the same time as the descriptive data points during the field season and photo-interpretation (see Section 2.5). After the mapping of all points was completed, the points were separated into descriptive and test data sets based on location in the description and test blocks. The cover type categories of the test points were grouped into the community and generalized levels of detail (Table 2.6).

A cluster of pixels in the cover type polygon for each test data point was used to compare the classifications with the reference data. A one pixel wide border was left around each cluster. This buffer was used for two reasons. First, the buffer nullifies minor locational errors of the cover type boundary around the test point. The test field had to be entirely within the polygon boundary. Second, the buffer allows for difference between the geometric corrections of classifications from IDIMS and EDITOR. The data in the IDIMS classification were offset slightly from north/south axis. The buffer was necessary to ensure that the test field cluster had four complete pixels for the evaluation.

A comparison was made of various test field sizes using the acetate base maps with the cover type polygons and a template of various test field sizes. The sizes of test fields which were compared were 6X6 pixels, 5X5, 4X4, 3X3, and less than 3X3 pixels. The test field size template was overlaid on the cover type polygon. If the test field fit within the polygon the corresponding data point could be used. If the test field template did not fit, the data point was discarded. If a test point was located on the boundary of two cover types, it was discarded because a test point could not be assigned two cover types for the accuracy evaluation.

The frequency of data points in each cover type category was calculated for the entire study area. The distribution of test fields by cover type was tallied for each test field size. The distribution of the test field data for each size was compared with the total data point distribution. The test field size with the frequencies most similar to the total data point distribution was selected for the evaluation. This was the 4X4 pixel test field; 1 pixel buffer and a 2X2 pixel cluster for the actual evaluation. Larger test fields emphasized cover types occurring in large polygons such as deciduous forest and missed cover types in small or linear polygons such as water and barrens.

The test fields were drawn on the acetate base map. For each test point with a large enough polygon to contain the test field, a 4X4 pixel test field was outlined. If the cover type polygon was large, the test field was placed on or near the test data point. If the polygon was irregularly shaped, the test field was placed wherever it would fit while maintaining the north/south orientation. One hundred and six valid test fields were located in the test blocks. Table 2.9 shows the distribution of the test fields by cover type.

The acetate base map with the cover type polygons and test fields outlined on it was registered to the printout maps of each classification. For each test field the following data were recorded: test point number, reference data cover type, and symbols of the spectral classes for the four pixels of the evaluation cluster. This process was repeated for the six classifications of the study area. Figure 2.7 shows an example of the test fields registered with the printout classification map.

The data from the evaluations were analyzed using SPSS. The major functions used were IF, RECODE, FREQUENCIES, and CROSSTABS. Accuracy evaluations were calculated for 26 different combinations of analysis methods and levels of detail (Table 2.10).

The spectral classes were grouped into cover types according to the spectral class descriptions used for each method. The acceptable spectral classes for each cover type in each evaluation are listed in

Table 2.9. Frequency of test fields by cover type.

Cover type	Frequency
Dense Conifer	52
Sparse Conifer	52
Deciduous	136
Mixed Forest	64
Burn	48
Brush	32
Grassland	4
Sparse Vegetation	8
Barren	12
Silty Water	12
Clear Water	4

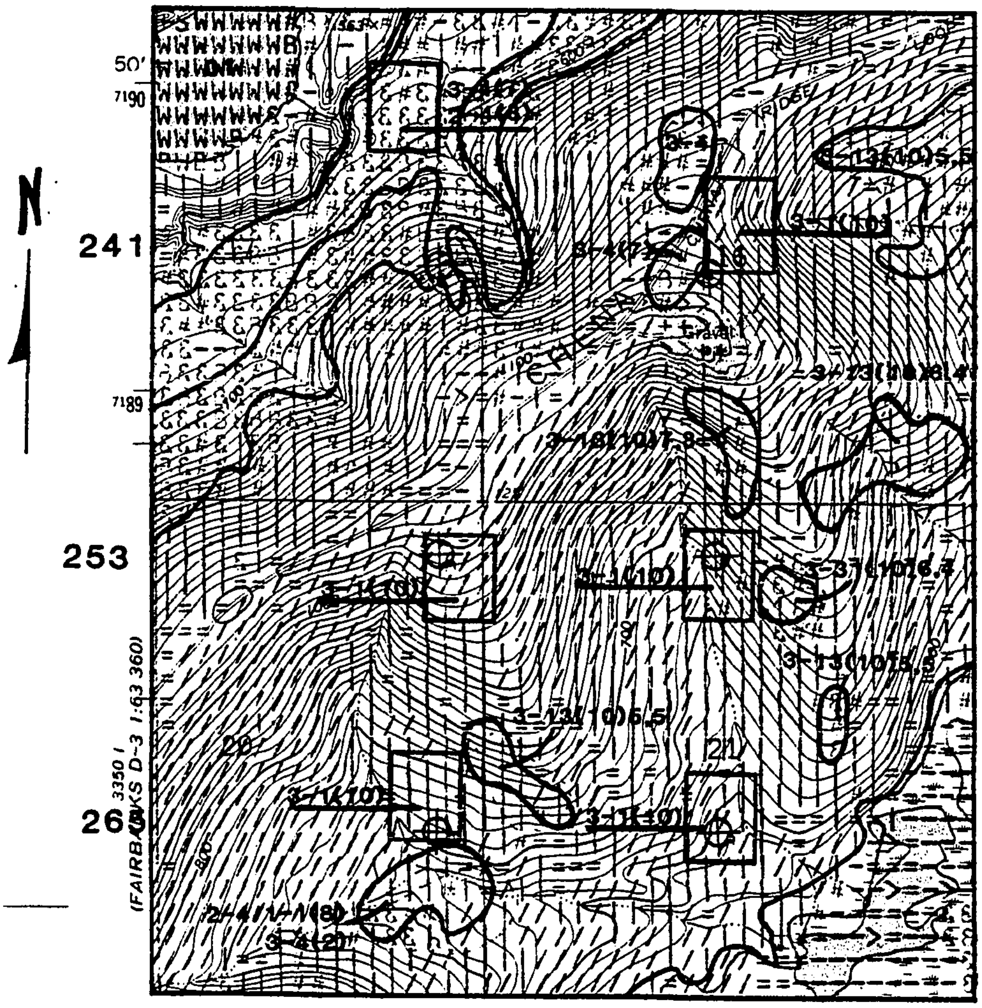


Figure 2.7. Example of the test fields registered with the printout map. The cover type code follows Table 2.4. This block is on Chena Ridge south of Ester. The printout is part of the EDITOR modified cluster classification.

Table 2.10. List of the different systems, analysis methods and levels of detail evaluated for this study.

IDIMS Supervised training descriptions, community level
IDIMS Supervised training descriptions, generalized level
IDIMS Supervised screen descriptions, community level
IDIMS Supervised screen descriptions, generalized level
IDIMS Supervised ground plot descriptions, community level
IDIMS Supervised ground plot descriptions, generalized level

IDIMS Unsupervised screen descriptions, community level
IDIMS Unsupervised screen descriptions, generalized level
IDIMS Unsupervised ground plot descriptions, community level
IDIMS Unsupervised ground plot descriptions, generalized level

IDIMS Modified Cluster training descriptions, community level
IDIMS Modified Cluster training descriptions, generalized level
IDIMS Modified Cluster screen descriptions, community level
IDIMS Modified Cluster screen descriptions, generalized level
IDIMS Modified Cluster ground plot descriptions, community level
IDIMS Modified Cluster ground plot descriptions, generalized level

EDITOR Supervised training descriptions, community level
EDITOR Supervised training descriptions, generalized level
EDITOR Supervised ground plot descriptions, community level
EDITOR Supervised ground plot descriptions, generalized level

EDITOR Unsupervised ground plot descriptions, community level
EDITOR Unsupervised ground plot descriptions, generalized level

EDITOR Modified Cluster training descriptions, community level
EDITOR Modified Cluster training descriptions, generalized level
EDITOR Modified Cluster ground plot descriptions, community level
EDITOR Modified Cluster ground plot descriptions, generalized level

Appendix B. The reference data for community types were grouped into generalized categories as shown in Table 2.6. The classified spectral classes were grouped into generalized categories using the same methods for each technique as used for the spectral class descriptions (Section 2.6.3).

When running the SPSS programs the spectral classes were combined into their appropriate cover types. Tables of cross tabulation were constructed for the reference data by cover type with the classified data (spectral classes) grouped by cover types. Table 2.11 shows an example of the cross-tabulation tables. The tables for all evaluations are in Appendix C. Tables of cross tabulation were also constructed to show the number of pixels correctly classified in each test field.

A pixel was considered correctly classified if the spectral class cover type and the reference cover type matched. The estimated accuracy or proportion correctly classified ($\hat{p}$) was calculated using Formula 2.3 (Scheaffer, et. al., 1979).

$$\hat{p} = \frac{\sum_{n=1}^{i} a_i}{\sum_{n=1}^{i} m_i} \qquad \text{Formula 2.3}$$

where a_i = number of pixels in the i^{th} cluster correctly classified.

m_i = number of pixels in the i^{th} cluster

percent correct = $\hat{p} \times 100$

Table 2.11. Example of the cross-tabulation tables used to calculate classification performance. These data are from the EDITOR Modified Cluster classification using ground plot spectral class descriptions with generalized map categories.

	Ground Data						
Classification	Conifer Forest	Deciduous Forest	Mixed Forest	Brush & Grasslands	Burn	Barren & Light Veg.	Water
Conifer Forest	98	4	25	4	7		
Deciduous Forest	2	130	39	18	5		
Mixed Forest							
Brush and Grasslands	4	6		10		10	
Burn					36		
Barren and Lt. Vegetation						6	1
Water						4	15

$\frac{295}{424} =$

69.5% correct

The estimates of percent correct were calculated for all the analysis methods and techniques being evaluated. The results are shown at the bottom of the cross-tabulation tables in Appendix C and in Tables 3.2 - 3.5 in Chapter 3.

The various techniques were compared using the Newman - Keuls Range Test (Landgrebe, 1976). The evaluation of a computer-aided analysis is a binomial distribution as tested pixels are either correct or not correct. The proportion correct ($\hat{p}$) can be transformed using ($\arcsin\sqrt{\hat{p}}$) to approximate the normal distribution. The standard error of the mean ($S\bar{y}$) was calculated with formula 2.4.

$$S\bar{y} = \frac{821}{n} = 1.39 \qquad \text{Formula 2.4}$$

where n = number observations per mean = 424 test pixels

The tabular ranges (Newman - Keuls Range Test) was calculated with Formula 2.5.

$$\text{Range} = (\text{Studentized Range}_{(df = \infty)})\ (S\bar{y}) \qquad \text{Formula 2.5}$$

$$\text{Range} = (2.77)\ (1.39) = 3.8$$

The Studentized Range value used (Snedecor and Cochran, 1967) was for comparing two treatments at the .95 confidence level.

The various classifications and techniques were compared two at a time. If the difference between the transformed $\hat{p}$ values was greater than 3.8 units apart, the difference was significant at the .95 level of confidence. If the difference was less than or equal tc 3.8 units, the two classifications were not significantly different. Comparisons were made for the IDIMS and EDITOR analysis systems, training methods, spectral class description techniques, and level of detail of mapping categories. The results are shown in Tables 3.2 - 3.5 and discussed in Chapter 3.

CHAPTER III

RESULTS AND DISCUSSION

A technique for collecting and recording reference data for computer-aided analysis projects has been developed and described. The data collection technique can be used with field work or air photo-interpretation. The coding system provides a flexible set of data which is not dependent on a static set of map categories, but which can be aggregated to describe the spectral classes of a computer-aided classification. Comparisons of different techniques were made for four different factors which affect classification accuracy. These factors are: 1) analysis system, 2) training method, 3) spectral class descriptions, and 4) map category detail. When making comparisons, only the techniques being evaluated for each factor were varied. All other factors were held constant. The same Landsat and reference data sets were used throughout the study to maintain consistency. The products for this study were developed strictly as a means to make comparisons and are not intended to be used for actual applications.

3.1. Reference Data Collection and Coding

One of the major considerations in a computer-aided analysis project is the collection and recording of reference data from field work and photo-interpretation. The reference data for digital analysis should consist of a map or photograph with the areal extent of the stands outlined on it, and the cover type of that stand recorded with a flexible coding system so that the data may be aggregated as necessary during the analysis. The system for collecting and recording reference data described in Section 2.5 meets these criteria and is summarized here. The analyst has a set of preselected points to locate in the field or on aerial photographs. The extent of the stand at a data point is outlined on the map or photograph. The stand enclosed by the boundary or polygon is relatively uniform for the major species present, the percentage of each species in a mix, and overall crown cover. The cover type identification is recorded for each stand using a code for each species or landcover type present and 10 percent increments in crown closure. Data for each stand or data point include major species, total crown closure, crown closure for each species, and miscellaneous items such as inclusions, disturbance history or understory.

A set of points should be selected prior to the field work or photo-interpretation for reference data collection. The author's experience has shown that a definite plan for locating and mapping stands in the field is more effective than random wandering through the study area. The points may be distributed throughout the study area using whatever sampling scheme has been selected. Additional points may be added during the reference data collection if necessary to adequately include the cover types of the study area for training or spectral class descriptions. The data collection points may be in a systematic grid as was used for this project, a random sample, or some stratification based on availability of aerial photographs, accessibility, or vegetation complexity. If statistical estimators are used for calculating accuracy, the reference data sampling scheme must address the assumptions made by the estimators.

This data collection and recording system does not require elaborate equipment or large field crews to collect sufficient data for a computer-aided analysis. The equipment needed for the field work includes a good pair of binoculars, a field notebook, maps, aerial photographs, and mosquito repellant. Plot frame equipment and complicated data recording sheets are not necessary.

The analyst who is doing the computer-aided analysis should participate in the major reference data collection effort. He should be familiar with the major species in the study area and be able to reliably identify them at a distance. The ability to read topographic maps, to interpret aerial photographs, and to locate ground points on the maps and photographs are essential. The analyst needs the ability to consistently estimate total crown closures for a stand and for the individual species crown closures within the stand. Vertical aerial photographs can be used to calibrate ocular estimates of crown closure. Work by the author with other ecologists shows that estimates by two experienced interpreters are usually within 10 percent of each other for total crown closure and mix percentages.

The use of predefined, static landcover categories for field work often results in differences in identification of stands, even between experienced ecologists. These differences are partially caused by vague or inadequate descriptions of the landcover categories and differences of interpretations between analysts with varying backgrounds. Another factor contributing to different identification of a stand is the highly variable nature of much vegetation which makes it difficult to find "typical" stands matching a specific landcover category. The wildland vegetation of an area often consists of a complex mosaic of "pure" types grading into each other in response to complicated patterns of environmental and vegetational characteristics and historical events. The coding system described in Section 2.5.1 requires only that the analyst be able to identify the species and estimate crown closures for a relatively homogeneous stand. The data can later be aggregated into the mapping categories which correspond to the spectral classes of the Landsat data.

Development of the data collection and coding systems used during this project began in 1975 in Colorado by this investigator and others (Krebs and staff, 1976). The system was very successful in mountainous wildland terrain of the San Juan Mountains, Colorado in a complex of cover types influenced by topography, fire, mining, and logging. Since that time, the author has used adaptations of this coding system for remote sensing projects in Mesa Verde National Park, Colorado, for Landsat image interpretation along the Denali Hiway, Alaska, for a computer-aided classification of the Susitna River and Cook Inlet basins of south-central Alaska, for reconnaissance field work for a computer-aided analysis of the Nulato Hills, Alaska, for fire fuels mapping in interior Alaska, and finally, for the study reported here. The landcover types in these studies have included the semi-arid sagebrush and pinyon/juniper forests of Mesa Verde, the complex mosaic of conifer forest types, brush and pastures in the San Juan Mountains, intricate mosaics of forest and brush types resulting from fire and other environmental factors in interior and south-central Alaska and the complex ecotone communities between arctic or alpine tundra and the boreal forest along the Denali Hiway and the Nulato Hills. In all cases, the coding system provided an efficient and effective method for recording landcover information which could be aggregated as necessary for interpretations of remote sensing data. This system also avoids the confusion resulting from subjective decisions by several interpreters as to the vegetation category of a particular stand.

3.2 Classification Accuracies

The results of the accuracy evaluations for all twenty-six analysis methods are shown in Table 3.1. The accuracies range from 31.1 percent for the community level of the EDITOR supervised method with training descriptions to 70.5 percent for the generalized level of the IDIMS modified cluster with ground plot descriptions. The average accuracy at the community level was approximately 50 percent and the average accuracy at the generalized level was approximately 60 percent. The most commonly used vegetation map for Alaska is the

Table 3.1. Results of Accuracy Evaluations.

Classification

System	Training	Description	Detail	Correct	arcsin $\sqrt{\hat{p}}$
IDIMS	Supervised	Training	Community	45.8	42.6
			Generalized	55.2	48.0
		Screen	Community	44.1	41.6
			Generalized	57.3	49.2
		Ground Plots	Community	59.0	50.2
			Generalized	69.8	56.7
	Unsupervised	Screen	Community	42.5	40.7
			Generalized	51.2	45.7
		Ground Plots	Community	54.7	47.7
			Generalized	64.4	53.4
	Modified Cluster	Training	Community	45.0	42.1
			Generalized	55.7	48.3
		Screen	Community	46.5	43.0
			Generalized	58.0	49.6
		Ground Plots	Community	58.7	50.0
			Generalized	70.5	57.1
EDITOR	Supervised	Training	Community	31.1	33.9
			Generalized	40.3	39.4
		Ground Plots	Community	56.1	48.5
			Generalized	65.6	54.1
	Unsupervised	Ground Plots	Community	58.5	49.9
			Generalized	70.0	56.8
	Modified Cluster	Training	Community	50.9	45.5
			Generalized	62.5	52.2
		Ground Plots	Community	63.7	53.0
			Generalized	69.6	56.5

Major Ecosystem map (LUPC, 1973) which was interpreted from Spetzman's original map produced for the military moves in Alaska during World War II (Spetzman, 1957-1963). The Major Ecosystems map has been qualitatively evaluated as 20-40 percent correct (Krebs, 1980). Although the map products from the study reported here are more accurate and detailed than the Major Ecosystem map, these products are not intended for use as a resource management tool. Krebs (1976) reported "... accuracy figures of 40-60 percent doomed any attempts for the U.S. Forest Service to actually use the data in planning efforts." Accuracy figures of 70-90 percent are frequently reported for quasi-operational classifications (Krebs, 1976; Hoffer and staff, 1975; George, 1981).

The low overall accuracies are due to a combination of several factors. One important factor is the composition of the mapping categories.

The choice of mapping categories for computer-aided analysis is a two fold problem. The first problem is the whole dilemma of separating continuously varying systems of vegetational units into discrete mapping units. There may be natural breaks in the vegetation continuum due to changes in species composition and crown closure. These breaks must first be recognized by the ecologist. They may or may not correspond to the mapping categories needed by a resource manager. The second problem involves finding the breaks in the continua of spectral reflectance values. With the proper parameters, the clustering process should result in clusters which break the spectral data into separable units. Ideally, the breaks in the spectral data correspond to the user specified mapping categories. This is seldom a viable assumption. (Occasionally the spectral classes correspond with the natural breaks in the vegetation continua.) If there is a high correlation between the spectral classes and the desired mapping categories and if the spectral classes are accurately described, the classification accuracy will be high. If the correlation between spectral classes and mapping categories is low, the user will have to use the vegetation classes found on the spectral data or low classification accuracies.

One method for defining mapping categories is to develop a framework of the desired cover types and then describe each spectral class as the cover type it most closely resembles. This method tends to result in lower accuracies because it makes no allowances for the spectral characteristics of the study area. There is often confusion between spectrally similar cover types. Another method is to have a "wish list" of desired mapping categories and try to achieve these cover types with the training procedures. After the classification is finished, the spectral class descriptions are refined to more precisely define the cover type(s) they actually represent. This method usually results in higher accuracies because the final mapping categories incorporate both spectral and vegetational characteristics of the study area. The first method was used for this study because comparable cover types were necessary for making direct comparisons.

The low accuracies were also due to problems with the mixed forest cover type class. In all spectral class description methods, only a few spectral classes were described as mixed forest. The mixed forest type was often confused with both coniferous and deciduous cover types, so the spectral class was described as conifer or deciduous instead of mixed forest. However, 15 percent of the test fields were mixed forest, and most of the pixels in these test fields were mis-classified as either conifer forest or deciduous forest.

Other factors also may have contributed to the low accuracies. The identification of the reference data was approximately 92 percent correct (see Section 2.5) and cannot be considered 100 percent correct ground truth (Smedes, date unknown). Radiometric corrections of the sensor miscalibration in band 7 may have improved the final classification. Further work with the training statistics and reclassifications may also have improved the accuracies.

3.3. Systems Comparison: IDIMS and EDITOR

The results of the systems comparison are shown in Table 3.2. These results show that there are no real differences between the IDIMS and EDITOR analysis systems. These results are interpreted as meaning there is no significant difference between the printout and color display screen interface methods for the analyst and the data. IDIMS is significantly better than EDITOR for the supervised classification with training spectral class descriptions for both the

Table 3.2. Comparisons of IDIMS and EDITOR Analysis Systems.

Classification - Level	$\arcsin\sqrt{\hat{p}}$ I = IDIMS; E = EDITOR				Δ $\arcsin\sqrt{\hat{p}}$	interpretation	Δ % correct
Supervised training - Community	I	42.6	E	33.9	8.7*	I>E	14.7
Supervised training - Generalized	I	48.0	E	39.4	8.6*	I>E	14.9
Supervised ground plot - Community	I	50.2	E	48.5	1.7	I=E	2.9
Supervised ground plot - Generalized	I	56.7	E	54.1	2.6	I=E	4.2
Unsupervised ground - Community	E	49.9	I	47.7	2.2	E=I	3.8
Unsupervised ground - Generalized	E	56.8	I	53.4	3.4	E=I	5.6
Mod. cluster training - Community	E	45.5	I	42.1	3.4	E=I	5.9
Mod. cluster training - Generalized	E	52.2	I	48.3	3.9*	E>I	6.8
Mod. cluster ground - Community	E	53.0	I	50.0	3.0	E=I	5.0
Mod. cluster ground - Generalized	I	57.1	E	56.5	.6	I=E	.9

* - significant differences at 95% confidence level.
The significant difference is the Δ $\arcsin\sqrt{\hat{p}}$ greater than 3.8.

community and generalized levels of detail. This is due to the extremely low accuracies of the EDITOR supervised training classifications (31.1 percent and 40.3 percent). EDITOR is significantly better than IDIMS for the generalized level of the modified cluster, training classification. These results do not rate one system consistently better than the other.

The lack of difference between results from IDIMS and EDITOR is partially due to the clustering techniques used in both systems, and editing of the statistics files. The use of the color display screen often results in the inclusion of contaminant pixels around the border of a training field of a pure cover type for supervised training. These contaminant pixels are usually separated into different clusters during the iterative clustering process. These extraneous classes are deleted during the statistics editing on the basis of having reflectance value means which do not match the cover type being considered. This process eliminates the effects of locational errors of the training fields during training. Each cluster block for modified cluster training includes several different cover types. Extra pixels are clustered into one of the classes for the major cover type or an additional class.

The EDITOR system is primarily available to government and University users through the ARPANET system. Pre- and post-classification processing (radiometric and geometric corrections) must be done on an ancillary computer system. EDITOR has flexible and powerful options for developing and fine-tuning training statistics.

EDITOR is also simple to learn and use due to the hierarchical command system and the question mark prompt (?) which lists or explains options.

The IDIMS software is implemented on a stand alone HP-3000 computer. These systems are marketed by a private firm, ESL, Inc., and are available at several government and private facilities throughout the country. All processing for a project can be completed on the IDIMS system through an extensive and flexible software package. The development of training statistics on the IDIMS is fairly automated and the analyst has to work around the programs to get specialized results. The clustering package has many parameters which the analyst must set. This provides the potential of fine-tuning the clusters. The study area characteristics and analyst experience are very important factors. IDIMS is more difficult to learn than EDITOR, but is versatile when used by an experienced analyst.

In the modified cluster training statistics, the spectral classes in small cluster blocks are identified. The classes derived from the EDITOR clustering were more consistent with the vegetation patterns and were easier to identify than the spectral classes from the IDIMS clustering algorithm ISOCLS. The researcher also has had difficulty identifying classes from ISOCLS in other study areas when using the modified cluster training technique. Other analysts have observed that spectral clusters from ISOCLS are more difficult to identify than are clusters from the LARSYS CLUSTER algorithm (Fleming, pers. comm.; Nelson and Hoffer, 1979). The LARSYS and EDITOR clustering algorithms are similar. More experience with ISOCLS parameters may improve the clustering for the modified cluster training method.

Another factor to consider when comparing these two systems is the number of spectral classes derived from the clustering procedure. The number of spectral classes derived should reflect the variability in the data set. The spectral classes are then combined to form information classes. Both systems have the capability to cluster a data set into a large or small number of clusters according to analyst-specified parameters.

The IDIMS classifications had 24 classes for supervised, 21 classes for unsupervised, and 17 classes for modified clustering training methods. The EDITOR classifications had 10 classes for supervised, 14 classes for unsupervised and 11 classes for the modified cluster training methods. A larger number of classes did not give higher accuracies for this study. The one exception to this trend was the EDITOR supervised classification. The low accuracies are a result of excessive editing of statistics and resulting confusion of cover types. Every spectral class has to be described using one or a combination of several methods of spectral class descriptions. The greater the number of classes, the more time and reference data are needed to adequately describe all the classes. The multiple spectral classes for one cover type are aggregated when the final products are produced. A balance needs to be found for each project so that the spectrally separable cover types are classified without having a large number of spectral classes for each cover type.

Projects using IDIMS tend to have large numbers of spectral classes, while projects using the EDITOR or similar LARSYS clustering have fewer clusters. Krebs (1978) used 43 and 38 classes for two scenes in Southcentral Alaska. ESL, Inc. (1978) used 56 classes for the Denali ASVT Project, and Rohde and Miller (1980) used 76 classes in Arizona. These projects used IDIMS clustering to develop the training statistics. Projects using EDITOR or LARSYS clustering include Krebs (1976) with 25 classes in mountainous Colorado, Morrissey and Ennis (1981) with 10-25 classes per scene on the North Slope, Alaska, and Gaydos and Newland (1978) using 37 classes for an urban and wildland study area in Puget Sound. Martin (1981) compared several parameters affecting number of clusters in the IDIMS clustering algorithm ISOCLS and found in every case that the more clusters, the higher the accuracy. Theoretically, a larger number of compact spectral classes will correspond to more detailed cover types. This is limited by the spectral separability of the cover types of interest (Hoffer, 1972).

3.4 Comparison of Training Methods

The final results of the comparison of training methods are summarized in Table 3.3. Supervised training is not significantly different from unsupervised training in six comparisons (.95 level of confidence). Modified cluster training is significantly better than unsupervised training in one out of five comparisons, and better than supervised training for three out of eleven comparisons. Supervised

Table 3.3. Comparisons of supervised, unsupervised and modified cluster training methods.

Classification - Level	$\arcsin\sqrt{\hat{\rho}}$ S = supervised U = unsupervised M = modified cluster				Δ $\arcsin\sqrt{\hat{\rho}}$	interpretation	Δ % correct
IDIMS training - Community	S	42.6	M	42.1	.5	S=M	.8
IDIMS training - Generalized	M	48.3	S	48.0	.3	M=S	.5
IDIMS screen - Community	M	43.0	S	41.6	1.4	M=S	2.4
IDIMS screen - Community	M	43.0	U	40.7	2.3	M=U	4.0
IDIMS screen - Community	S	41.6	U	40.7	.9	S=U	1.6
IDIMS screen - Generalized	M	49.6	S	49.2	.4	M=S	.7
IDIMS screen - Generalized	M	49.6	U	45.7	3.9*	M>U	6.8
IDIMS screen - Generalized	S	49.2	U	45.7	3.5	S=U	6.1
IDIMS ground - Community	S	50.2	M	50.2	.2	S=M	.3
IDIMS ground - Community	M	50.0	U	47.7	2.3	M=U	4.0
IDIMS ground - Community	S	50.2	U	47.7	2.5	S=U	4.3
IDIMS ground - Generalized	M	57.1	S	56.7	.4	M=S	.7
IDIMS ground - Generalized	M	57.1	U	53.4	3.7	M=U	6.1
IDIMS ground - Generalized	S	56.7	U	53.4	3.3	S=U	5.4

Table 3.3. (cont'd) Comparisons of supervised, unsupervised and modified cluster training methods.

Classification - Level	arcsin $\sqrt{\hat{\rho}}$ S = supervised U = unsupervised M = modified cluster				Δ arcsin $\sqrt{\hat{\rho}}$	interpretation	Δ % correct
EDITOR training - Community	M	45.0	S	33.9	11.1*	M>S	19.8
EDITOR training - Generalized	M	52.2	S	39.4	12.8*	M>S	22.2
EDITOR ground - Community	M	53.0	S	48.5	4.8*	M>S	7.6
EDITOR ground - Community	M	53.0	U	49.9	3.1	M=U	5.2
EDITOR ground - Community	S	48.5	U	49.9	1.4	U=S	2.4
EDITOR ground - Generalized	M	56.5	S	54.1	2.4	M=S	4.0
EDITOR ground - Generalized	M	56.5	U	56.8	.3	U=M	.4
EDITOR ground - Generalized	S	54.1	U	56.8	2.7	U=S	4.4

* - significant differences at .95% confidence level. The significant difference is the Δ arcsin $\sqrt{\hat{\rho}}$ greater than 3.8.

and unsupervised training is not significantly better than modified cluster training in any of the comparisons. Modified cluster accuracies were an average 6.4 percent higher than supervised or unsupervised accuracies for 13 out of 16 comparisons. Although this difference is not large enough to be significant at the .95 confidence level, these results show that modified cluster training gives consistently higher classification accuracies than either supervised or unsupervised training.

Perhaps as important as accuracy of the final classification is the amount of reference data and analyst time necessary to complete a project. Reference data, especially for field work, are expensive. Often reference data simply are not available or must be carefully planned within a strict budget. The modified cluster training method used approximately one-fourth as much reference data as the supervised method for developing the training statistics. The supervised method took approximately four times as much analyst time to develop the training statistics as the modified cluster method.

3.5. Comparison of Spectral Class Description Methods

Three methods of deriving spectral class descriptions were compared (Table 3.4). In all comparisons in this study, the ground plot descriptions were significantly better than either training or screen descriptions, and the training descriptions were not significantly different from the screen descriptions.

Table 3.4. Comparison of three methods of spectral class descriptions. Spectral classes were described using training, screen and ground plot data.

Classification - Level	arcsin $\sqrt{\hat{\rho}}$ T - training S - screen G - ground plot				Δ arcsin $\sqrt{\hat{\rho}}$	interpretation	Δ % correct
IDIMS Supervised - Community	G	50.2	T	42.6	7.6*	G>T	13.2
IDIMS Supervised - Community	G	50.2	S	41.6	8.6*	G>S	14.9
IDIMS Supervised - Community	S	41.6	T	42.6	1.0	T=S	1.7
IDIMS Supervised - Generalized	G	56.7	T	48.0	8.7*	G>T	14.6
IDIMS Supervised - Generalized	G	56.7	S	49.2	7.5*	G>S	12.5
IDIMS Supervised - Generalized	S	49.2	T	48.0	1.2	S=T	2.1
IDIMS Unsupervised - Community	G	47.7	S	40.7	7.0*	G>S	12.2
IDIMS Unsupervised - Generalized	G	53.4	S	45.7	7.7*	G>S	13.2
IDIMS Mod. Cluster - Community	G	50.0	T	42.1	7.9*	G>T	13.7
IDIMS Mod. Cluster - Community	G	50.0	S	43.0	7.0*	G>S	12.2
IDIMS Mod. Cluster - Community	S	43.0	T	42.1	.9	S=T	1.5
IDIMS Mod. Cluster - Generalized	G	57.1	T	48.3	8.8*	G>T	14.8
IDIMS Mod. Cluster - Generalized	G	57.1	S	49.6	7.5*	G>S	12.5

Table 3.4. (cont'd) Comparison of three methods of spectral class descriptions. Spectral classes were described using training, screen and ground plot data.

Classification - Level	arcsin $\sqrt{\hat{\rho}}$ T - training S - screen G - ground plot				Δ arcsin$\sqrt{\hat{\rho}}$	interpretation	Δ % correct
IDIMS Mod. Cluster - Generalized	S	49.6	T	48.3	1.3	S=T	2.3
EDITOR Supervised - Community	G	48.5	T	33.9	14.6*	G>T	25.0
EDITOR Supervised - Generalized	G	54.1	T	39.4	14.7*	G>T	25.3
EDITOR Mod. Cluster - Community	G	53.0	T	45.5	7.5*	G>T	12.8
EDITOR Mod. Cluster - Generalized	G	56.5	T	52.2	4.3*	G>T	7.1

* - significatn difference at 95% confidence level. The significant difference is the Δ arcsin $\sqrt{\hat{\rho}}$ greater than 3.8.

Much of the higher accuracies of the ground plot descriptions are due to the stratification of the 1975 burn south of the Tanana River. Examination of the Landsat data acquired one year after the burn did not indicate that special training should be used for a burn class. The variations within the burn area were less than the overall variations for other brush and sparse black spruce stands in the study area. Fuller and Rouse (1979) found little difference in reflectance between recent burns and mature lichen/black spruce forest in the visible wavelengths. The mature forest had approximately 10 percent higher reflectance in the near infrared wavelengths due to the lichen component. This difference was not evident on the Landsat data set used for this study, probably because the lichen component is not as dominant in the study area.

When the IDIMS classifications were examined on the color display screen, confusion of cover types was evident, but the burn pattern was not clear. The IDIMS unsupervised classification showed confusion between water and brush in the burn area, and the corresponding spectral classes were stratified. This stratification did not improve the accuracy as compared with ground plot descriptions (Table 3.4). The IDIMS supervised and modified cluster did not show clear patterns of confusion and were not stratified.

The extent of the confusion in the burn area was not evident until the CROSSTABS matrix was constructed for the spectral classes and the ground plot reference data. Some spectral classes had 40 percent or more of their description data in the burn area. These spectral classes were stratified by assigning them two cover type descriptions. For the burn area they were identified as burn. For the remainder of the study area these spectral classes were identified by their primary cover type.

Stratification is a process for redefining a spectral class into two or more informational classes through the use of ancillary data. It is particularly valuable when one spectral class corresponds to different cover types in different locations of the study area or on different elevations and/or aspects. Stratification procedures may be used prior to the classification, during the classification, or post-classification. In all stratification procedures it is imperative that the analyst knows how the vegetational distribution varies with the stratification parameters.

Stratification has been used as a method to improve classifications on several projects. Rohde (1978) reports that stratification of flooded lands prior to classification improved the accuracy 5 to 10 percent. Stratification of a study area near Denali, Alaska improved classification of barren and water classes. Morrisey and Ennis (1981) stratified riparian brush on the north slope, Alaska. George (pers. comm.) stratified a classification of range types on the Seward Peninsula.

Two different interpretations may be made using the results from the comparison of spectral class description methods:

1. Spectral class descriptions using the ground plot method

results in higher classification accuracies.

2. Stratification of the spectral classes results in higher classification accuracies.

Further investigations were made as part of the comparison of spectral class descriptions. The test fields located in the burn area were removed from the total test data set to form an abbreviated test data set. This was called the unstratified data set.

The ground plot descriptions were compared using the unstratified data set and the test data with the stratified burn area. The results are shown in Table 3.5. There were no significant differences between the two data sets when the errors caused by unstratified spectral descriptions were removed from the comparison. The stratified accuracies were consistently higher than the unstratified accuracies. The stratified accuracies were an average of 3.3 percent higher than the unstratified accuracies.

The three methods of spectral class descriptions were compared using the unstratified test data set (Table 3.6). There were no significant differences between the training, screen, and ground plot description methods (.95 level of confidence). The screen descriptions were generally better than the training descriptions (average 2.2 percent higher) and the ground plot descriptions were always higher than the screen or training description methods (average of 4.9 percent).

Table 3.5. Comparison of ground plot descriptions using test data with burn plots (stratified) and test data with burn plots removed (unstratified).

Classification - Level	arcsin $\sqrt{\hat{p}}$ S = stratified U = unstratified				Δ arcsin $\sqrt{\hat{p}}$	interpretation	Δ % correct
IDIMS Supervised - Community	S	50.2	U	47.9	2.3	S=U	3.9
IDIMS Supervised - Generalized	S	56.7	U	54.8	1.9	S=U	3.1
IDIMS Unsupervised - Community	S	47.7	U	45.6	2.1	S=U	3.6
IDIMS Unsupervised - Generalized	S	53.4	U	52.7	.7	S=U	1.1
IDIMS Mod. Cluster - Community	S	50.0	U	46.9	3.1	S=U	5.3
IDIMS Mod. Cluster - Generalized	S	57.1	U	54.8	2.3	S=U	3.7
EDITOR Supervised - Community	S	48.5	U	45.6	2.9	S=U	5.0
EDITOR Supervised - Generalized	S	54.1	U	51.8	2.3	S=U	3.9
EDITOR Unsupervised - Community	S	49.9	U	48.7	1.2	S=U	2.1
EDITOR Unsupervised - Generalized	S	56.8	U	56.9	.1	U=S	.2
EDITOR Mod. Cluster - Community	S	53.0	U	50.9	2.1	S=U	3.4
EDITOR Mod. Cluster - Generalized	S	56.5	U	56.1	.4	S=U	.7

The significant difference at the .95% confidence level is Δ arcsin $\sqrt{\hat{p}}$ greater than 3.9.

$$S\bar{y} = \frac{\frac{821}{2}}{(1/424 + 1/376)} = 1.44$$

Table 3.6. Comparison of spectral class description methods using test data without the burn fields (unstratified).

Classification - Level	arcsin $\sqrt{\hat{\rho}}$ T = training S = screen G = ground plots				Δ arcsin $\sqrt{\hat{\rho}}$	interpretation	Δ % correct
IDIMS Supervised - Community	G	47.9	T	45.9	2.0	G=T	3.5
IDIMS Supervised - Community	G	47.9	S	44.8	3.1	G=S	5.4
IDIMS Supervised - Community	S	44.8	T	45.9	1.1	T=S	1.9
IDIMS Supervised - Generalized	G	54.8	T	52.1	2.7	G=T	2.4
IDIMS Supervised - Generalized	G	54.8	T	53.5	1.3	G=T	2.1
IDIMS Supervised - Generalized	S	53.5	T	52.1	1.4	S=T	2.4
IDIMS Unsupervised - Community	G	45.6	S	43.6	2.0	G=S	3.5
IDIMS Unsupervised - Generalized	G	52.7	S	49.3	3.4	G=S	5.9
IDIMS Mod. Cluster - Community	G	46.9	T	45.4	1.5	G=T	2.6
IDIMS Mod. Cluster - Community	G	46.9	S	46.4	.5	G=S	1.0
IDIMS Mod. Cluster - Community	S	46.4	T	45.4	1.0	S=T	1.6
IDIMS Mod. Cluster - Generalized	G	54.8	T	52.4	2.4	G=T	4.0
IDIMS Mod. Cluster - Generalized	G	54.8	S	54.0	.8	G=S	1.4
IDIMS Mod. Cluster - Generalized	S	54.0	T	52.4	1.6	S=T	2.6

Table 3.6. (cont'd) Comparison of spectral class description methods using test data without the burn fields (unstratified).

Classification - Level	arcsin $\sqrt{\hat{p}}$ T = training S = screen G = ground plots	Δ arcsin $\sqrt{\hat{p}}$	interpretation	Δ % correct
EDITOR Supervised - Community	G 45.6 T 36.3	9.3*	G>T	16.0
EDITOR Supervised - Generalized	G 51.8 T 42.4	9.4*	G>T	16.2
EDITOR Mod. Cluster - Community	G 50.9 T 49.3	1.6	G=T	2.9
EDITOR Mod. Cluster - Generalized	G 57.1 T 56.1	1.0	G=T	1.6

* - significant differences at .95% confidence level. The significant difference is Δ arcsin $\sqrt{\hat{p}}$ greater than 4.1.

$$Sy = 821/379 = 1.48$$
$$R2 = (1.48)\,(2.77) = 4.1$$

The comparison of the three spectral class description methods using test data without the burn indicate that the high accuracies from the ground plot description method are primarily due to the effects of stratifying the 1975 burn (Table 3.4). Ground plot descriptions give an average of 4.9 percent higher accuracies than training or screen description methods (Table 3.6), but the differences were not significant at the .95 confidence level. The final interpretation for the data on comparing methods of spectral class descriptions is a combination of the two interpretations suggested above. The ground plot description method identified the problem with the burn and a stratification technique was used to improve the descriptions.

The ground plot method is valuable because it locates stands of each cover type and matches them with the corresponding spectral classes. When using the screen the analyst tends to interpret according to spectral class patterns and approximate locations. Errors such as confusion of water and shadow are easy to identify using a screen. Errors involving broad cover types such as forest or brush classes in a wildland situation are much more difficult to pinpoint using the screen. For these cases the ground plot data can be used to identify the cover types which are being confused with each other.

A combination of screen and training descriptions with ground plot data can be used to develop good spectral class descriptions and identify any stratification necessary. In this proposed method, the classification is viewed on the screen and the spectral classes are identified (or refined) by general cover types. The ground plot data are then used to refine the descriptions and identify classes that may benefit by stratification. The actual stratification involves using the screen to identify general trouble areas if they can be seen, then use of field work, aerial photo-interpretation and spectral factors of the study area to outline the area for stratification and re-identify the spectral classes. Digital topographic data may offer potential in some stratification efforts.

3.6 Comparison of Mapping Category Detail

The results of comparing a community level and a generalized level of mapping categories are in Table 3.7. As expected, the generalized mapping categories are significantly more accurate than the community level of detail. The primary importance of these results is to illustrate how detail of mapping categories affects accuracy. When planning a project, resource managers often request very detailed cover types which are not necessary for the problems being addressed. Broad cover types may be adequate and give a higher accuracy. If detailed cover types are necessary, the user often sacrifices some accuracy.

Table 3.7. Comparisons of community and generalized levels of detail of mapping categories.

Classification - Level	arcsin $\sqrt{\hat{\rho}}$ C = community G = generalized				Δ arcsin $\sqrt{\hat{\rho}}$	interpretation	Δ % correct
IDIMS Supervised - Training	G	48.0	C	42.6	5.4*	G>C	9.4
IDIMS Supervised - Screen	G	49.2	C	41.6	7.6*	G>C	13.2
IDIMS Supervised - Ground plot	G	56.7	C	50.2	6.5*	G>C	10.8
IDIMS Unsupervised - Screen	G	45.7	C	40.7	5.0*	G>C	8.7
IDIMS Unsupervised - Ground plot	G	53.4	C	47.7	5.7*	G>C	9.7
IDIMS Mod. Cluster - Training	G	48.3	C	42.1	6.2*	G>C	10.7
IDIMS Mod. Cluster - Screen	G	49.6	C	43.0	6.6*	G>C	11.5
IDIMS Mod. Cluster - Ground plot	G	57.1	C	50.0	7.1*	G>C	11.8
EDITOR Supervised - Training	G	39.4	C	33.9	5.5*	G>C	9.2
EDITOR Supervised - Ground plot	G	54.1	C	48.5	5.6*	G>C	9.4
EDITOR Unsupervised - Ground plot	G	56.8	C	49.9	6.9*	G>C	11.5
EDITOR Mod. Cluster - Training	G	52.2	C	45.5	6.7*	G>C	11.6
EDITOR Mod. Cluster - Ground plot	G	56.5	C	53.0	3.5	G=C	5.9

* - significant differences at .95 confidence level. The significant difference is the Δ arcsin $\sqrt{\hat{\rho}}$ greater than 3.8.

CHAPTER IV

SUMMARY AND CONCLUSIONS

The key to computer-aided analysis is the word "aided." The portions of the analysis which involve repetitive computations for a large amount of data are automated. There are many phases which require intensive interaction of the analyst and the data. The quality of these interactions depends on the analyst's experience with the computer analysis, the vegetation of the study area, the reference data available, and the spectral response patterns of the cover types involved. This study compared the IDIMS and EDITOR analysis systems, three methods of deriving training statistics, three methods of spectral class descriptions, and two levels of detail for mapping categories on the basis of classification accuracy. A system for collecting and recording reference data from field work and air photo-interpretation was developed and described.

Reference data for a computer-aided analysis project should consider the spectral and spatial characteristics of the Landsat data, the analysis procedures of the computer system being used, and the complex and variable characteristics of wildland cover types. A technique has been developed which meets these criteria. Preselected data points are located in the field or on the aerial photographs. The cover type at each point is identified and the extent of the stand is delineated on a map or aerial photograph. The stand is considered to be homogeneous with respect to total crown closure, species present, and species mix. The cover type identification is recorded by species present and crown cover in 10 percent increments. These reference data have geographically located stands of the various cover types in the study, each recorded with a flexible coding system. The data can be aggregated easily to correspond to the spectral categories of the digital data set. The analyst is not committed to a static set of mapping categories.

There were no significant differences between the IDIMS and EDITOR systems. The IDIMS system provides a powerful analysis package with a stand-alone computer. EDITOR is easier to learn and use. IDIMS clustering tends to produce more spectral classes than EDITOR for a data set. This increases the time necessary for analysis and spectral class descriptions and did not result in higher accuracies in this study.

The supervised, unsupervised, and modified cluster (multi-cluster block) training methods were compared. The supervised and unsupervised were not significantly different. The modified cluster method was an average of 6.4 percent higher than both the supervised and unsupervised methods, but these results were not consistently significant at the .95 confidence level. The modified cluster training method used approximately one-fourth as much reference data and analyst time as the supervised method. This may be an important consideration to a project with limited time, budget, or reference data.

Spectral classes were described using the labels from training, identifications on the color display screen, and ground plots identified with field work and photo-interpretation. The training and screen description methods were not significantly different from each other. The ground plot descriptions were significantly better than either training or screen descriptions. Most of the difference was due to stratification of a burn south of the Tanana River. The need for stratification was not evident except with the confusion matrix generated for the ground plot descriptions. Properly applied stratification procedures can significantly increase classification accuracy.

Generalized cover types were significantly more accurate than the more detailed community land cover types. This has implications for project planning and trade-off of detail and classification accuracy.

There is probably no one best set of methods or a cookbook for computer-aided analysis. The methods used for each phase of an analysis should be selected for each specific project. Factors affecting choice of methods include vegetation and spectral variability in the study area, analyst experience in the study area aerial photographs and other reference data available, the computer system available for the project, accuracies and mapping categories desired, and budget available for the project.

The following conclusions have been made from the results of this study:

1. Reference data which consider spectral and spatial

characteristics of the Landsat data and the floristic and spatial patterns of the landcover types of a study area are effective for a computer-aided analysis project. The reference data collection and recording technique has geographically located stands and cover type identifications in a flexible coding system that can be aggregated to correspond to the spectral data categories.

2. The analyst must have a knowledge of the ecologic, floristic and spectral characteristics of the cover type in the study area to develop the spectral class descriptions and stratification criteria.
3. For an experienced analyst, there is no difference in classification accuracy from using a printout oriented system such as EDITOR or a screen oriented system such as IDIMS.
4. The modified cluster method of developing training statistics is more effective and efficient than supervised or unsupervised training methods.
5. The use of ground plot data and subsequent stratification improves the descriptions of spectral classes.
6. Generalized mapping categories are more accurate than detailed mapping categories.

REFERENCES CITED

Aldrich, R.C. 1979. Remote Sensing of Wildland Resources: A State-of-the-Art Review. General Technical Report RM-71, Rocky Mt. Forest and Range Experiment Station, U.S.D.A. Forest Service, Ft. Collins, Colo. 56 p.

Anuta, P.E. 1973. Geometric Correction of ERTS-1 Digital Multispectral Scanner Data. LARS Information Note 103073. Laboratory for Applications of Remote Sensing, Purdue Univ., W. Lafayette, Ind. 23 p.

Bauer, M.E., P.H. Swain, R.P. Mroczynski, P.E. Anuta, and R.B. MacDonald. 1971. Detection of Southern Corn Leaf Blight by Remote Sensing Techniques. Pages 693-704 in Proceedings of the 7th International Symposium on Remote Sensing of the Environment, Univ. of Michigan, Ann Arbor, Mich.

Bauer, M.E., L. Silva, R.M. Hoffer, and M. Baumgardner. 1977. Agricultural Scene Understanding. LARS Contract Report 112677. Laboratory for Applications of Remote Sensing, Purdue Univ., W. Lafayette, Ind. 173 p.

Carter, V.P., F. Billingsley, and J. Lamar. 1977. Summary Tables for Selected Digital Image Processing Systems. U.S.G.S. Open File Report 77-414. 45 p.

Coggeshall, M.E. and R.M. Hoffer. 1973. Basic Forest Cover Mapping Using Digital Remote Sensor Data and ADP Techniques. M.S. Thesis, Purdue Univ., W. Lafayette, Ind. 157 p.

Craighead, J.J. 1980. Grizzly Bear Habitat Analysis; Section III, Landsat I Multispectral Imagery and Computer Analysis of Grizzly Bear Habitat. Wildlands-Wildlife Institute, Univ. of Montana, Missoula, Mont. 274 p.

Daubenmire, R. 1978. Plant Geography. Academic Press, New York, N.Y. 338 p.

Doyle, F.A. 1978. The Next Decade of Satellite Remote Sensing. Photog. Eng. 44(2):155-164.

ESL. 1976. IDIMS User's Guide. Technical Memorandum ESL-TM705. ESL Inc., Sunnyvale, Ca. 2 vols. n.p.

ESL. 1978. Final Report for Phase I, Wildland Vegetation Resources Inventory - System Integration and Demonstration. Engineering Report ESL-ER208. ESL Inc., Sunnyvale, Ca. 557 p.

Fischer, W.A. 1975. History of Remote Sensing, Chapt. 2. Pages 27-50 in Manual of Remote Sensing, Reeves, R.G. (ed.). Keuffel and Esser Co. USA. 2144 p.

Fitzpatric-Lins, K. 1978. Accuracy of Selected Land Use and Land Cover Maps in the Greater Atlantic Region, Georgia. Jour. Research U.S. Geological Survey 6(2):169-173.

Fleming, M.D., J.S. Berkebile, and R.M. Hoffer. 1975. Computer-aided Analysis of Landsat 1 MSS Data: A Comparison of Three Approaches Including a Modified Cluster Approach. LARS Information Note 072475, Laboratory for Applications of Remote Sensing, Purdue Univ., W. Lafayette, Ind. 9 p.

Fleming, M.D. and R.M. Hoffer. 1977. Computer-aided Analysis Techniques for an Operational System to Map Forest Lands Utilizing Landsat MSS Data. LARS Technical Report 112277, LARS, Purdue Univ., W. Lafayette, Ind. 236 p.

Fuller, S.P. and W.R. Rouse. 1979. Spectral Relectance Changes Accompanying a Post-Fire Recovery Sequence in a Subarctic Spruce Lichen Woodland. Remote Sensing of the Envir. 8(1):11-23.

Gaydos, L. and W.L. Newland. 1978. Inventory of Land Use of the Puget Sound Region Using Landsat Digital Data. Jour. Research U.S. Geological Survey 6(6):807-819.

George, T. 1981. Verification of Range Inventory Maps. Northern Remote Sensing Laboratory, Geophysical Institute, Univ. of Alaska, Fairbanks, Alaska. 9 p.

Hixon, M.D., D. Scholz, N. Fuhs, and T. Akiyama. 1980. Evaluation of Several Schemes for Classification of Remotely Sensed Data. Photog. Eng. 46(12):1547-1553.

Hoffer, R.M. 1971. The Importance of "Ground Truth" Data in Remote Sensing. LARS Print 120371. Laboratory for Applications of Remote Sensing, Purdue Univ., W. Lafayette, Ind. 12 p.

Hoffer, R.M. 1972. ADP of Multispectral Scanner Data for Land Use Mapping. LARS Information Note 080372. Laboratory for Applications of Remote Sensing, Purdue Univ., W. Lafayette, Ind. 24 p.

Hoffer, R.M. and Staff. 1974. An Interdisciplinary Analysis of Colorado Rocky Mountain Environments Using ADP Techniques. Research Bulletin 919 (LARS Information Note 061575). Laboratory for Applications of Remote Sensing, Purdue Univ., W. Lafayette, Ind. 124 p.

Hoffer, R.M. and Staff. 1975. Computer-aided Analysis of SKYLAB Multispectral Scanner Data in Mountainous Terrain for Land Use, Forestry, Water Resource and Geologic Applications. LARS Information Note 121275. Laboratory for Applications of Remote Sensing, Purdue Univ., W. Lafayette, Ind. 381 p.

Hulten, E. 1968. Flora of Alaska and Neighboring Territories. A Manual of the Vascular Plants. Stanford Univ. Press, Stanford, Ca. 1008 p.

Institute for Advanced Computation. 1978. EDITOR User's Handbook. IAC Tech. Memo No. 5662, IAC Project No. 78-24. Earth Resources Program. Ames Research Center, Moffit Field, Ca. n.p.

Isaacson, D.L., C.J. Alexander, B.J. Schrumf, and R. Murray. 1979. Analysis of Association of Landsat Spectral Classes with Ground Cover Classes in Wildland Inventories. Pages 180-191 in Proceedings of Remote Sensing for Natural Resources. Univ. of Idaho, Moscow, Id.

Joyce, A.T. 1978. Procedures for Gathering Ground Truth Information for a Supervised Approach to a Computer-Implemented Land Cover Classification of Landsat-Acquired Multi-spectral Scanner Data. NASA References Publication 1015. NASA Earth Resources Laboratory, Slidell, La. 48 p.

Kettig, R.L. and D.A. Landgrebe. 1973. Automatic Boundary and Sample Classification of Remotely Sensed Multi-spectral Data. LARS Information Note 041773. Laboratory for Applications of Remote Sensing, Purdue Univ., W. Lafayette, Ind. 32 p.

Krebs, P.V. (ed.) 1980. ASVT Project, Phase I, Denali, Alaska Study Area - Technical Report of the Evaluation. U.S.D.I. Bureau of Land Management Alaska State Office, Anchorage, Alaska. 152 p.

Krebs, P.V., J.P. Spencer, K.G. Dean, and S.E. Rawlinson 1978. Natural Resources Maps of Southcentral Alaska: Landforms and Surficial Deposits, Geologic Hazards, and Landcover. Final Report: User's Guide. BLM Contract YA-512-CT8-97. Geophysical Institute, Univ. of Alaska, Fairbanks, Alaska. 46 p.

Krebs, P.V. and Staff 1976. Multiple Resource Evaluation of Region 2, U.S. Forest Service Lands Utilizing Landsat MSS Data. Final Report for NASA Contract NAS5020498. INSTAAR, Univ. of Colo., Boulder, Colo. 298 p.

Landgrebe, D.A., R.M. Hoffer, E. Goodrich and Staff. 1972. An Early Analysis of ERTS-1 Data. LARS Information Note 092972. Laboratory for Applications of Remote Sensing, Purdue Univ., W. Lafayette, Ind. 21 p.

Landgrebe, D.A. 1976. Final Report, NASA Contract NAS9-14016. Laboratory for Applications of Remote Sensing, Purdue Univ., W. Lafayette, Ind. 173 p.

Landgrebe, D.A. 1978. The Quantitative Approach: Concept and Rationale, Chp. 1. Pages 1-20 in Remote Sensing, the Quantitative Approach, Swain and Davis (eds.). McGraw-Hill Int'l Book Co., New York, N.Y.

LUPC (Joint Federal-State Land Use Planning Commission for Alaska). 1973. Major Ecosystems of Alaska. U.S. Geological Survey, Washington, D.C. 1 map.

Lutz, H.J. 1956. Ecological Effects of Forest Fire in the Interior of Alaska. Tech. Bull. No. 1133. U.S.D.A., Washington, D.C. 112 p.

Martin, G. 1981. Effect of Varying ISOCLS Parameters on Classification of Landsat MSS Data. U.S.G.S. Contract No. 14-08-001-16439, 20 p.

Mayer, K.E., L. Fox, and J.L. Webster. 1979. Forest Condition Mapping of the Hoopa Valley Indian Reservation Using Landsat Data. Pages 217-242 in Proceedings of Remote Sensing for Natural Resources, A Symposium. Sept. 10-14, 1979. Univ. of Idaho, Moscow, Ind.

Morrissey, L.A. and R.A. Ennis. 1981. Vegetation Mapping of the National Petroleum Reserve - Alaska Using Landsat Digital Data. U.S.D.I. - U.S. Geological Survey Open File Report 81-315. 53 p.

NASA. 1975. Abstracts of the NASA Earth Resources Survey Symposium. National Aeronautics and Space Administration. Johnson Space Center, Houston, Texas. 315 p.

Neiland, B.J. and L.A. Viereck. 1977. Forest Types and Ecosystems. Pages 109-136 in Proceedings of North American Forest Lands at Latitudes North of 60 Degrees. Univ. of Alaska, Fairbanks, Alaska, Sept. 19-22, 1977.

Nelson, R.F. and R.M. Hoffer. 1979. Computer-Aided Processing of Landsat MSS Data for Classification of Forest Lands. LARS Tech. Report 102679. Laboratory for Applications of Remote Sensing, Purdue Univ., W. Lafayette, Ind. 94 p.

Nie, N.H., C.H. Hull, J.G. Jankins, K. Steinbrenner, and D.H. Bent. 1975. Statistical Package of the Social Sciences, 2nd Edition. McGraw-Hill Book Co., New York, N.Y. 675 p.

Philipson, W.R. 1980. Problem Solving with Remote Sensing. Photog. Eng. 46(10):1335-1338.

Phillips, T.L. and P.H. Swain 1978. Data-Processing Methods and Systems, Chp. 4. Pages 188-226 in Remote Sensing, the Quantitative Approach. McGraw-Hill Int'l Book Co., USA.

Poulton, C.E. 1972. A Comprehensive Remote Sensing Legend System for the Ecological Characterization and Annotation of Natural and Altered Landscapes. Pages 393-408 in Proceedings of the 8th InternationalSymposium on Remote Sensing of the Environment. Oct. 2-6, 1972, Ann Arbor, Mich.

Rieger, S.A., J.A. Benet, and D. Sanders 1963. Soil Survey of Fairbanks Area, Alaska. U.S.D.A. Soil Conservation Services, Series 1959, No. 25. U.S. Gov't Printing Office, Washington, D.C. 41 p.

Rohde, W.G. 1978. Improving Land Cover Classification by Image Stratification of Landsat Data. Pages 729-741 in Proceedings of the 12th International Symposium on Remote Sensing of the Environment. Environmental Research Institute of Michigan, Ann Arbor, Mich.

Rohde, W.G. and C.E. Olson, Jr. 1972. Multispectral Sensing of Forest Tree Species. Photog. Eng. 38(12):1209-1215.

Rohde, W.G. and W.A. Miller 1980. AVRI Project Final Report. U.S.G.S. EROS Office Internal Report to BLM, Denver. EROS Data Center, Souix Falls, S.D. 200 p.

Rosenfield, G.H. and M.L. Melley 1980. Applications of Statistics to Thematic Mapping. Photog. Eng. 46(10):1287-1294.

Sabins, F.F. 1978. Remote Sensing Principles and Interpretations. H.S. Freeman and Co., San Francisco, Ca. 426 p.

Scheaffer, R.L., W. Mendenhall, and L. Ott 1979. Elementary Survey Sampling, 2nd ed. Duxbury Press, North Scituate, Mass. 278 p.

Selkregg, L. (ed.) 1975. Alaska Regional Profiles, Yukon Region. Arctic Environmental Information and Data Center, Univ. of Alaska, Anchorage, Alaska. 346 p.

Smedes, H.W. (date unknown). The Truth About Ground-Truth Maps. U.S. Geological Survey, Denver, Colo.

Snedecor, G.W. and W.G. Cochran 1967. Statistical Methods. The Iowa State Univ. Press, Ames, Iowa. 593 p.

Spencer, J.P. 1977. Landsat Photo-interpretation for Vegetation, Final Report to ESL for Denali ASVT Project, Geophysical Institute, Univ. of Alaska, Fairbanks, Alaska. 20 p.

Spetzman, L.A. 1957-1963. Vegetation-Type Maps. Report for Military Geology Branch, U.S. Geological Survey, Washington, D.C.

vanGenderen, J.L. and B.F. Lack 1977. Testing Land Use Map Accuracy. Photog. Eng. 43(10):1135-1137.

Viereck, L.A. 1970. 'Forest Succession and Soil Development Adjacent to the Chena River in Interior Alaska. Arc. and Alp. Res. 2(1):1-26.

Viereck, L.A. and C.T. Dyrness 1980. A Preliminary Classification System for Vegetation of Alaska. General Technical Report PNW-106, Pacific Northwest Forest and Range Experiment Station, U.S. Dept. of Agriculture, Forest Service. 38 p.

Wacker, A. and D. Landgrebe 1971. The Minimum Distance Approach to Classification. LARS Tech. Report 100771. Laboratory for Applications of Remote Sensing, Purdue Univ., W. Lafayette, Ind. 361 p.

U.S.G.S. 1949. Fairbanks D-2 Topographic Map. U.S. Geological Survey, Denver, Colo. 1 map.

APPENDIX A

STATISTICS FOR CLASSIFICATION

Table A.1. Final Statistics - IDIMS Supervised Classification

Cluster Means

Class	Band	4	5	6	7*
1		41.4	45.1	41.5	29.6
2		30.9	27.6	28.0	21.0
3		38.5	36.8	37.6	27.1
4		31.3	29.8	35.7	28.7
5		36.9	33.4	29.4	19.3
6		18.8	11.1	37.4	36.6
7		19.2	10.9	40.2	42.4
8		18.0	9.8	25.8	25.9
9		18.3	9.9	32.1	32.7
10		19.0	11.4	31.5	34.9
11		18.7	11.2	22.0	19.2
12		17.7	9.7	17.4	14.9
13		17.4	10.0	11.0	7.0
14		15.0	6.9	5.4	1.6
15		18.0	11.1	9.6	3.0
16		19.1	11.9	14.3	10.4
17		33.8	30.1	23.9	12.2
18		21.5	13.5	32.8	28.9
19		20.4	14.0	28.2	24.7
20		22.4	15.1	39.8	38.7
21		20.7	12.9	33.5	34.4
22		24.2	17.0	32.6	32.7
23		28.2	23.7	42.4	37.0
24		24.7	21.3	28.9	26.0

*Reflectance values in band 7 are doubled.

Table A.1. (Cont't). Final Statistics - IDIMS Supervised Classification

Cluster Variances

Class	Band	4	5	6	7
1		2.41	2.81	11.52	15.87
2		5.70	10.54	3.56	6.63
3		3.23	6.78	19.30	20.53
4		4.44	6.69	8.16	7.63
5		4.11	2.62	3.75	2.44
6		1.07	1.61	2.00	1.82
7		.94	1.61	3.14	8.16
8		1.35	1.82	1.96	6.03
9		.68	.93	3.34	8.04
10		.77	1.93	.71	.99
11		1.06	1.98	4.33	3.93
12		.38	.56	2.41	3.17
13		1.55	2.74	2.96	1.35
14		.56	.94	1.95	2.58
15		2.00	1.68	2.11	1.00
16		1.83	4.50	3.98	3.36
17		1.81	2.01	3.02	2.42
18		.75	1.06	1.11	1.75
19		1.74	2.62	1.05	2.17
20		3.09	4.25	6.52	4.59
21		1.64	.64	2.25	3.88
22		1.31	3.00	3.09	2.63
23		1.80	5.25	4.33	7.86
24		1.74	4.91	6.06	9.42

Table A.2. Final Statistics - IDIMS Unsupervised Classification

Cluster Means

Class	Band 4	5	6	7*
1	18.7	11.5	28.4	26.2
2	24.9	18.6	30.1	25.8
3	18.9	11.6	27.6	22.4
4	32.9	29.1	23.6	12.8
5	20.1	12.6	33.7	35.8
6	20.1	12.2	38.9	38.0
7	18.9	11.9	23.5	21.9
8	19.0	12.0	24.0	24.1
9	22.3	15.8	31.3	29.1
10	22.2	15.4	27.6	23.9
11	18.3	10.5	30.3	28.6
12	17.3	9.7	17.2	14.1
13	18.1	9.9	26.6	28.0
14	20.2	13.0	26.8	28.0
15	20.2	13.4	36.1	32.1
16	17.5	9.9	21.6	19.1
17	27.9	23.9	32.8	28.3
18	20.2	12.7	31.9	31.6
19	34.0	32.2	33.0	25.1
20	23.3	16.3	35.8	36.0
21	18.9	10.5	35.0	37.8

*Reflectance values for band 7 have been doubled.

Table A.2 (Con't). Final Statistics - IDIMS Unsupervised Classification

Cluster Variances

Class	Band	4	5	6	7
1		.84	.94	1.36	.29
2		.71	2.44	3.61	4.49
3		.57	1.59	1.34	2.52
4		2.69	1.73	2.71	2.22
5		.71	1.05	1.13	.92
6		1.46	1.98	2.36	1.58
7		.84	.73	.99	.22
8		.76	.56	1.29	.11
9		1.02	1.00	3.46	2.62
10		.66	2.37	1.28	1.77
11		.55	.88	.93	1.12
12		1.35	2.37	12.10	19.43
13		.68	.68	.96	1.73
14		.84	.72	.52	1.03
15		1.01	1.07	2.91	1.06
16		.67	.81	1.98	1.42
17		2.08	3.82	2.13	3.11
18		.81	.64	.41	1.27
19		11.57	22.52	10.63	10.42
20		.91	1.50	3.59	2.82
21		.54	1.00	2.00	.40

Table A.3. Final Statistics - IDIMS Modified Cluster Classification

Cluster Means

Class	Band	4	5	6	7
1		19.5	11.0	39.5	42.8
2		19.2	12.0	26.0	25.3
3		18.3	10.7	20.8	18.6
4		20.5	13.4	31.3	32.6
5		18.3	10.1	30.3	29.7
6		24.3	17.7	32.0	30.1
7		32.2	29.7	36.1	31.0
8		28.8	25.2	40.8	35.8
9		25.4	19.5	27.8	22.7
10		32.0	28.3	29.6	21.8
11		19.7	12.7	11.5	6.6
12		24.4	18.5	36.5	34.9
13		40.8	39.0	39.3	28.6
14		19.7	12.3	35.0	36.2
15		33.6	29.7	23.7	12.2
16		20.5	14.3	23.7	21.7
17		29.3	25.5	21.93	13.1

*Reflectance values in band 7 have been doubled.

Table A.3 (Con't). Final Statistics - IDIMS Modified Cluster Classification

Cluster Variances

Class	Band	4	5	6	7
1		1.30	1.60	4.42	7.86
2		1.70	2.93	2.45	2.49
3		1.33	2.02	8.44	13.11
4		1.29	1.57	1.04	.96
5		.79	1.03	5.07	2.58
6		2.45	3.96	2.09	4.57
7		3.96	4.08	4.85	5.00
8		4.22	6.97	5.34	6.91
9		4.21	7.59	16.89	18.15
10		4.95	8.89	6.69	5.45
11		7.70	9.90	4.00	8.36
12		1.43	2.59	4.21	5.72
13		15.91	22.61	28.02	22.41
14		1.38	2.15	1.65	.34
15		1.33	1.54	1.75	1.52
16		1.72	1.07	.80	2.93
17		1.48	4.29	1.66	2.62

Table A.4. Final Statistics - EDITOR Supervised Classification

Cluster Means

Class	Band	4	5	6	7
1		17.8	10.3	20.3	9.3
2		18.1	10.2	26.9	13.1
3		19.4	11.0	41.4	21.4
4		20.4	12.7	35.7	18.3
5		22.0	16.7	27.5	12.7
6		27.6	23.4	30.0	12.0
7		41.8	39.7	35.2	11.7
8		33.3	32.3	34.9	13.3
9		16.9	9.2	8.6	2.2
10		30.3	26.6	23.3	6.8

Cluster Variances

Class	Band	4	5	6	7
1		1.01	1.23	2.98	1.41
2		1.26	1.74	3.36	1.73
3		1.23	1.84	3.56	3.62
4		1.78	2.37	1.97	1.87
5		2.00	2.00	2.00	1.74
6		2.65	3.65	3.00	2.20
7		4.01	4.76	6.49	1.13
8		4.80	8.41	6.81	2.93
9		2.50	2.00	1.75	1.90
10		1.24	2.03	3.51	1.39

Table A.5. Final Statistics - EDITOR Unsupervised Classification

Cluster Means

Class	Band	4	5	6	7
1		18.1	10.7	12.5	4.4
2		18.6	11.3	22.9	10.7
3		19.4	12.5	24.9	1.7
4		32.3	28.5	23.6	6.6
5		23.4	17.2	25.3	10.8
6		20.4	13.4	26.9	13.2
7		18.1	10.0	29.7	14.5
8		21.4	15.0	29.5	13.5
9		21.4	14.1	32.6	15.5
10		20.3	12.8	35.6	18.0
11		19.6	11.4	40.3	20.9
12		24.1	17.8	37.8	18.2
13		28.2	23.5	29.6	11.8
14		33.1	30.4	31.5	11.8

Cluster Variances

Class	Band	4	5	6	7
1		6.05	10.40	6.26	3.74
2		1.05	2.19	1.24	1.04
3		0.75	0.71	0.56	0.99
4		5.42	5.04	3.43	1.44
5		1.36	2.78	4.99	2.24
6		0.91	1.08	0.71	0.93
7		0.79	0.74	1.05	1.18
8		1.12	1.32	1.13	1.06
9		1.30	1.36	1.08	1.47
10		0.93	1.02	2.28	1.81
11		1.57	2.33	5.28	3.32
12		1.63	4.29	5.97	2.58
13		2.65	2.16	5.05	3.20
14		4.28	5.79	4.49	2.71

Table A.6. Final Statistics - EDITOR Modified Cluster Classification

Cluster Means

Class	Band	4	5	6	7
1		16.8	9.3	18.5	8.4
2		18.3	10.7	23.7	11.2
3		18.0	10.0	27.2	13.0
4		18.1	9.8	30.9	15.6
5		22.3	15.4	31.2	14.4
6		22.6	15.6	36.5	17.3
7		19.6	11.3	39.7	21.1
8		19.6	12.3	12.1	3.9
9		27.8	23.8	37.4	16.7
10		33.3	30.5	32.4	12.3
11		32.5	28.6	23.8	6.7

Cluster Variances

Class	Band	4	5	6	7
1		0.62	0.79	2.00	1.01
2		1.00	1.78	0.97	1.13
3		0.74	0.99	1.27	1.38
4		0.74	0.76	1.63	1.75
5		1.95	1.45	1.81	1.61
6		2.17	3.45	3.95	1.91
7		1.65	2.22	3.55	2.28
8		2.00	1.70	2.10	1.00
9		4.37	5.83	8.49	2.88
10		9.23	9.77	7.39	3.78
11		4.32	5.29	2.75	1.29

APPENDIX B

SPECTRAL CLASS DESCRIPTIONS

Table B.1. IDIMS Supervised spectral class descriptions. See Table 2.5 for key to cover type characters.

Spectral Class	Spectral Class Description Method: Training Community	Training Generalized	Screen Community	Screen Generalized	Ground Plots Community	Ground Plots Generalized
1	K	6	K	6	K	6
2	K	6	K	6	K	6
3	K	6	K	6	K	6
4	K	6	K	6	H	4
5	K	6	K	6	K	6
6	C	2	C	2	C	2
7	C	2	C	2	C	2
8	C	2	B	1	B/F	1/5
9	C	2	C	2	C	2
10	C	2	C	2	C	2
11	A	1	B	1	A/F	1/5
12	A	1	A	1	A	1
13	M	7	M	7	M/F	7/5
14	M	7	M	7	No Plots	No Plots
15	M	7	M	7	M	7
16	M	7	M/F	7/5	M	7
17	L	7	L	7	L	7
18	B	1	B	1	G/F	4/5
19	G	4	G	4	B/F	1/5
20	I	6	E	3	C	2
21	G	4	G	4	C	2
22	I	6	I	6	G	4
23	H	4	H	4	H	4
24	H	4	H	4	H	4

Table B.2. IDIMS Unsupervised spectral class descriptions. See Table 2.5 for key to cover type characters.

Spectral Class Description Method

Spectral	Training		Screen		Ground Plots	
Class	Community	Generalized	Community	Generalized	Community	Generalized
1	Not Applicable		A	1	B/F	1
2			I	6	C	2
3			G	4	B/F	1/5
4			L	7	L	7
5			C	2	C	2
6			C	2	C	2
7			B	1	B/F	1/5
8			B	1	B/F	1/5
9			G	4	C/F	2/5
10			G	4	G/F	4/5
11			E	3	C/F	2/5
12			M/F	7/5	A/F	1/5
13			C	2	C	3
14			G	4	B/F	1/5
15			C	2	C	2
16			B	1	A	1
17			H	4	No Plots	No Plots
18			C	2	C/F	2/5
19			K	6	C	6
20			C	2	C	2
21			C	2	C	2

Table B.3. IDIMS Modified Cluster spectral class descriptions. See Table 2.5 for key to cover type characters.

Spectral Class Description Method

Spectral Class	Training		Screen		Ground Plots	
	Community	Generalized	Community	Generalized	Community	Generalized
1	C	2	C	2	C	2
2	B	1	B	1	B/F	1/5
3	A	1	A	1	A/F	1/5
4	G	4	G	4	C/F	4/5
5	E	3	E	3	C	2
6	I	6	G	4	C/F	2/5
7	H	4	H	4	H	4
8	H	4	G	4	H	4
9	G	4	K	6	G	4
10	K	6	K	6	K	6
11	M	7	M	7	M	7
12	H	4	G	4	C	4
13	K	6	K	6	K	6
14	C	2	C	2	C	2
15	L	7	L	7	L	7
16	G	4	H	4	B/F	1/5
17	K	6	K	6	L	7

Table B.4. EDITOR Supervised spectral class descriptions. See Table 2.5 for key to cover type characters.

Spectral Class Description Method

Spectral Class	Training Community	Training Generalized	Screen Community	Screen Generalized	Ground Plots Community	Ground Plots Generalized
1	A	1	Not Applicable		A/F	1/5
2	C	2			C/F	1/5
3	C	2			C	2
4	G	4			C	2
5	H	4			G/F	4/5
6	I	6			K	6
7	K	6			K	6
8	K	6			K	6
9	M	7			M	7
10	L	7			L	7

Table B.5. EDITOR Unsupervised spectral class descriptions. See Table 2.5 for key to cover type characters.

Spectral	Spectral Class Description Method					
	Training		Screen		Ground Plots	
Class	Community	Generalized	Community	Generalized	Community	Generalized
1	Not Applicable		Not Applicable		A	1
2					A	1
3					B/F	1/5
4					L	7
5					G	4
6					B/F	1/5
7					C	2
8					G/F	4/5
9					C/F	2/5
10					G/F	4
11					C	2
12					C	2
13					C	4
14					K	6

Table B.6. EDITOR Modified Cluster spectral class descriptions. See Table 2.5 for key to cover type characters.

Spectral Class Description Method

Spectral	Training		Screen		Ground Plots	
Class	Community	Generalized	Community	Generalized	Community	Generalized
1	A	1	Not Applicable		A	1
2	B	1			B/F	1/5
3	E	3			B/F	1/5
4	C	2			C	2
5	G	4			G/F	4/5
6	G	4			C	2
7	C	2			C	2
8	M	7			M	7
9	H	4			C	4
10	K	6			K	6
11	L	7			L	7

APPENDIX C

ACCURACY EVALUATIONS FOR ALL CLASSIFICATIONS

Table C.1. Supervised Training Evaluation - Community Level

Covertype	Reference Data										
Classified	Dense Conifer	Sparse Conifer	Deciduous Forest	Mixed Forest	Burn	Brush	Grasslands	Sparse Vegetation	Barren	Silty Water	Clear Water
Dense Conifer	40	32	-	15	25	-	-	-	-	-	-
Sparse Conifer	-	1	2	-	1	6	-	-	-	-	-
Deciduous Forest	9	5	121	41	6	10	-	-	-	-	-
Mixed Forest	-	-	-	-	-	-	-	-	-	-	-
Burn	-	-	-	-	-	-	-	-	-	-	-
Brush	3	12	12	7	15	10	-	-	-	-	-
Grassland	-	-	-	-	-	-	1	-	1	-	-
Sparse Forest	-	-	-	-	-	3	-	4	-	-	-
Barren	-	-	-	-	-	-	-	4	7	2	-
Silty Water	-	-	-	-	-	-	-	-	3	10	-
Clear Water	-	2	-	-	-	-	-	-	1	-	4

$\frac{194}{424}$ = 45.8% Correct

Table C.2. IDIMS Supervised Training Evaluation - Generalized Level

Covertype	Reference Data						
Classified	Conifer	Deciduous	Mix	Brush and Grass	Burn	Barren and Lt. Veg.	Water
Conifer	73	2	15	6	26	-	-
Deciduous	14	121	11	10	6	-	-
Mixed Forest	-	-	-	-	-	-	-
Brush & Grass	15	12	7	11	15	1	-
Burn	-	-	-	-	-	-	-
Barren & Lt. Veg.	-	1	1	9	1	15	2
Water	2	-	-	-	-	4	14

$\frac{234}{424}$ = 55.2% Correct

Table C.3. IDIMS Supervised Screen Evaluation - Community Level

Covertype	Reference Data										
Classified	Dense Conifer	Sparse Conifer	Deciduous Forest	Mixed Forest	Burn	Brush	Grasslands	Sparse Vegetation	Barren	Silty Water	Clear Water
Dense Conifer	2	5	-	-	-	-	-	-	-	-	-
Sparse Conifer	47	32	7	22	32	9	-	-	-	-	-
Deciduous Forest	-	1	116	34	-	7	-	-	-	-	-
Mixed Forest	-	-	1	1	1	3	3	-	-	-	-
Burn	-	-	-	-	-	-	-	-	-	-	-
Brush	3	12	12	7	15	10	-	-	-	-	-
Grassland	-	-	-	-	-	-	1	-	1	-	-
Sparse Vegetation	-	-	-	-	-	3	-	4	-	-	-
Barren	-	-	-	-	-	-	-	4	7	2	-
Silty Water	-	-	-	-	-	-	-	-	3	10	-
Clear Water	-	2	-	-	-	-	-	-	1	-	4

$\frac{187}{424}$ = 44.7% Correct

Table C.4. IDIMS Supervised Screen Evaluation - Generalized Level

Covertype	Reference Data						
Classified	Conifer	Deciduous	Mix	Brush and Grass	Burn	Barren and Lt. Veg.	Water
Conifer	86	7	22	9	32	-	-
Deciduous	1	116	34	7	-	-	-
Mixed Forest	-	1	1	6	1	-	-
Brush & Grass	15	12	7	11	15	1	-
Burn	-	-	-	-	-	-	
Barren & Lt. Veg.	-	-	-	3	-	15	2
Water	2	-	-	-	-	4	14

$\frac{243}{424}$ = 57.3% Correct

Table C.5. IDIMS Supervised Ground Plot Evaluation - Community Level

Covertype	Reference Data										
Classified	Dense Conifer	Sparse Conifer	Deciduous Forest	Mixed Forest	Burn	Brush	Grasslands	Sparse Vegetation	Barren	Silty Water	Clear Water
Dense Conifer	40	32	-	15	-	-	-	-	-	-	-
Sparse Conifer	12	16	8	14	-	6	-	-	-	-	-
Deciduous Forest	-	1	126	35	5	17	3	-	-	-	-
Mixed Forest	-	-	-	-	-	-	-	-	-	-	-
Burn	-	-	-	-	43	-	-	-	-	-	-
Brush	-	1	2	-	-	9	-	4	-	-	-
Grassland	-	-	-	-	-	-	-	3	2	-	-
Sparse Vegetation	-	-	-	-	-	-	1	-	1	-	-
Barren	-	-	-	-	-	-	-	1	5	2	-
Silty Water	-	-	-	-	-	-	-	-	3	10	-
Clear Water	-	2	-	-	-	-	-	-	1	-	1

$\frac{250}{424}$ = 59.0% Correct

Table C.6. IDIMS Supervised Ground Plot Evaluation - Generalized Level

Covertype	Reference Data						
Classified	Conifer	Deciduous	Mix	Brush and Grass	Burn	Barren and Lt. Veg.	Water
Conifer	100	8	29	6	-	-	-
Deciduous	1	126	35	20	5	-	-
Mixed Forest	-	-	-	-	-	-	-
Brush & Grass	1	2	-	10	-	10	-
Burn	-	-	-	-	43	-	-
Barren & Lt. Veg.	-	-	-	-	-	6	2
Water	2	-	-	-	-	4	11

$\frac{296}{424}$ = 69.8% Correct

Table C.7. IDIMS Unsupervised Screen Evaluation – Community Level

Covertype	Reference Data										
Classified	Dense Conifer	Sparse Conifer	Deciduous Forest	Mixed Forest	Burn	Brush	Grasslands	Sparse Vegetation	Barren	Silty Water	Clear Water
Dense Conifer	-	7	3	9	5	2	-	-	-	-	-
Sparse Conifer	28	21	-	13	19	2	-	-	-	-	-
Deciduous Forest	-	5	118	23	3	20	4	-	-	-	-
Mixed Forest	-	3	15	14	1	1	-	-	-	-	-
Burn	-	-	-	-	1	-	-	-	-	-	-
Brush	2	1	1	-	12	3	-	2	-	-	-
Grassland	10	2	3	2	7	-	-	2	2	-	-
Sparse Vegetation	-	-	-	-	-	-	-	2	-	-	-
Barren	-	-	-	-	-	-	-	2	5	-	-
Silty Water	-	-	-	-	-	-	-	5	12	-	-
Clear Water	12	13	-	3	-	-	-	-	-	4	-

$\frac{180}{424}$ = 42.5% Correct

Table C.8. IDIMS Unsupervised Screen Evaluation - Generalized Level

Covertype	Reference Data						
Classified	Conifer	Deciduous	Mix	Brush and Grass	Burn	Barren and Lt. Veg.	Water
Conifer	56	3	22	4	24	-	-
Deciduous	5	118	23	24	3	-	-
Mixed Forest	3	15	14	1	1	-	-
Brush & Grass	15	4	2	3	19	6	-
Burn	-	-	-	-	1	-	-
Barren & Lt. Veg.	-	-	-	-	-	9	-
Water	25	-	3	-	-	5	16

$\frac{217}{424}$ = 51.2% Correct

Table C.9. IDIMS Unsupervised Ground Plot Evaluation - Community Level

Covertype	Reference Data										
Classified	Dense Conifer	Sparse Conifer	Deciduous Forest	Mixed Forest	Burn	Brush	Grasslands	Sparse Vegetation	Barren	Silty Water	Clear Water
Dense Conifer	31	26	-	11	-	-	-	-	-	-	4
Sparse Conifer	9	15	3	14	-	5	-	-	-	-	-
Deciduous Forest	-	8	134	37	1	23	4	6	5	-	-
Mixed Forest	-	-	-	-	-	-	-	-	-	-	-
Burn	-	-	-	-	40	-	-	-	-	-	-
Brush	2	1	-	-	-	-	-	-	-	-	-
Grassland	-	-	-	-	-	-	-	-	-	-	-
Sparse Vegetation	-	-	-	-	-	-	-	-	-	-	-
Barren	-	-	-	-	-	-	-	-	-	-	-
Silty Water	-	-	-	-	-	-	-	-	5	12	-
Clear Water	-	-	-	-	-	-	-	-	-	-	-
NO PLOTS	10	2	3	2	7	-	-	7	2	-	-

$\frac{232}{424}$ = 54.7% Correct

Table C.10. IDIMS Unsupervised Ground Plot Evaluation - Generalized Level

Covertype	Reference Data						
Classified	Conifer	Deciduous	Mix	Brush and Grass	Burn	Barren and Lt. Veg.	Water
Conifer	81	3	25	5	5	-	4
Deciduous	7	133	34	25	1	2	-
Mixed Forest	1	-	3	-	-	-	-
Brush & Grass	3	1	-	2	-	2	-
Burn	-	-	-	-	35	-	-
Barren & Lt. Veg.	-	-	-	-	-	7	-
Water	-	-	-	-	-	5	12

$\frac{273}{424}$ = 64.4% Correct

Table C.11. IDIMS Modified Cluster Training Evaluation - Community Level

Covertype	Reference Data										
Classified	Dense Conifer	Sparse Conifer	Deciduous Forest	Mixed Forest	Burn	Brush	Grasslands	Sparse Vegetation	Barren	Silty Water	Clear Water
Dense Conifer	38	33	2	8	13	2	-	-	-	-	-
Sparse Conifer	9	11	2	13	11	3	-	-	-	-	-
Deciduous Forest	-	-	11	14	-	8	3	-	-	-	-
Mixed Forest	1	3	37	23	-	1	-	-	-	-	-
Burn	-	-	-	-	-	-	-	-	-	-	-
Brush	4	2	4	6	19	8	-	6	2	-	-
Grassland	-	-	3	-	-	3	1	1	-	-	-
Sparse Vegetation	-	3	1	-	5	3	-	1	-	-	-
Barren	-	-	-	-	-	-	-	-	6	4	-
Silty Water	-	-	-	-	-	-	-	-	2	8	-
Clear Water	-	-	-	-	-	-	-	-	2	-	4

$\frac{191}{424}$ = 45.0% Correct

Table C.12. IDIMS Modified Cluster Training Evaluation - Generalized Level

Covertype Classified	Reference Data Conifer	Deciduous	Mix	Brush and Grass	Burn	Barren and Lt. Veg.	Water
Conifer	91	4	21	5	24	-	-
Deciduous	-	91	14	11	-	-	-
Mixed Forest	4	37	23	1	-	-	-
Brush & Grass	6	7	6	12	19	9	-
Burn	-	-	-	-	-	-	-
Barren & Lt. Veg.	3	1	-	3	5	7	4
Water	-	-	-	-	-	4	12

$\frac{236}{424}$ = 55.7% Correct

Table C.13. IDIMS Modified Cluster Screen Evaluation - Community Level

Covertype Classified	Reference Data: Dense Conifer	Sparse Conifer	Deciduous Forest	Mixed Forest	Burn	Brush	Grasslands	Sparse Vegetation	Barren	Silty Water	Clear Water
Dense Conifer	38	33	2	8	13	2	-	-	-	-	-
Sparse Conifer	9	11	2	13	11	3	-	-	-	-	-
Deciduous Forest	-	-	91	14	-	8	3	-	-	-	-
Mixed Forest	1	3	37	23	-	1	-	-	-	-	-
Burn	-	-	-	-	-	-	-	-	-	-	-
Brush	-	5	8	5	9	14	1	2	-	-	-
Grassland	4	-	-	1	15	-	-	-	-	-	-
Sparse Vegetation	-	-	-	-	-	-	-	-	-	-	-
Barren	-	-	-	-	-	-	-	6	8	4	-
Silty Water	-	-	-	-	-	-	-	-	2	8	-
Clear Water	-	-	-	-	-	-	-	-	2	-	4

$\frac{197}{424}$ = 46.5% Correct

Table C.14. IDIMS Modified Cluster Screen Evaluation - Generalized Level

Covertype	Reference Data						
Classified	Conifer	Deciduous	Mix	Brush and Grass	Burn	Barren and Lt. Veg.	Water
Conifer	91	4	21	5	24	-	-
Deciduous	-	91	14	11	-	-	-
Mixed Forest	4	37	23	1	-	-	-
Brush & Grass	9	8	6	15	24	2	-
Burn	-	-	-	-	-	-	-
Barren & Lt. Veg.	-	-	-	-	-	14	4
Water	-	-	-	-	-	4	12

$\frac{246}{424}$ = 58.0% Correct

Table C.15. IDIMS Modified Cluster Ground Plot Evaluation - Community Level

Covertype	Reference Data										
Classified	Dense Conifer	Sparse Conifer	Deciduous Forest	Mixed Forest	Burn	Brush	Grasslands	Sparse Vegetation	Barren	Silty Water	Clear Water
Dense Conifer	38	33	2	8	-	2	-	-	-	-	-
Sparse Conifer	13	11	2	14	-	3	-	-	-	-	-
Deciduous Forest	1	8	136	42	-	23	3	2	-	-	-
Mixed Forest	-	-	-	-	-	-	-	-	-	-	-
Burn	-	-	-	-	48	-	-	-	-	-	-
Brush	-	-	-	-	-	-	-	6	2	-	-
Grassland	-	-	-	-	-	-	1	-	-	-	-
Sparse Vegetation	-	-	-	-	-	-	-	-	-	-	-
Barren	-	-	-	-	-	-	-	-	3	4	-
Silty Water	-	-	-	-	-	-	-	-	5	8	-
Clear Water	-	-	-	-	-	-	-	-	2	-	4

$\frac{249}{424}$ = 58.7% Correct

Table C.16. IDIMS Modified Cluster Ground Plot Evaluation - Generalized Level

Covertype	Reference Data						
Classified	Conifer	Deciduous	Mix	Brush and Grass	Burn	Barren and Lt. Veg.	Water
Conifer	95	4	22	5	-	-	-
Deciduous	7	129	37	15	-	1	-
Mixed Forest	-	-	-	-	-	-	-
Brush & Grass	2	7	5	12	-	9	-
Burn	-	-	-	-	48	-	-
Barren & Lt. Veg.	-	-	-	-	-	3	4
Water	-	-	-	-	-	7	12

$\frac{299}{424}$ = 70.5% Correct

Table C.17. EDITOR Supervised Training Evaluation - Community Level

Covertype	Reference Data										
Classified	Dense Conifer	Sparse Conifer	Deciduous Forest	Mixed Forest	Burn	Brush	Grasslands	Sparse Vegetation	Barren	Silty Water	Clear Water
Dense Conifer	33	35	-	10	10	2	-	-	-	-	-
Sparse Conifer	-	-	-	-	-	-	-	-	-	-	-
Deciduous Forest	16	17	59	43	27	4	4	-	-	-	-
Mixed Forest	-	-	-	-	-	-	-	-	-	-	-
Burn	-	-	-	-	-	-	-	-	-	-	-
Brush	-	-	81	10	2	21	-	2	-	-	-
Grassland	3	-	-	1	9	1	-	2	-	1	-
Sparse Vegetation	-	-	-	-	-	-	-	4	3	-	-
Barren	-	-	-	-	-	-	-	-	3	3	-
Silty Water	-	-	-	-	-	-	-	-	6	8	-
Clear Water	-	-	-	-	-	-	-	-	-	-	4

$\frac{132}{424}$ = 31.1% Correct

Table C.18. EDITOR Supervised Training Evaluation - Generalized Level

Covertype	Reference Data						
Classified	Conifer	Deciduous	Mix	Brush and Grass	Burn	Barren and Lt. Veg.	Water
Conifer	68	-	10	2	10	-	-
Deciduous	33	59	43	8	27	-	-
Mixed Forest	-	-	-	-	-	-	-
Brush & Grass	3	81	11	22	11	4	1
Burn	-	-	-	-	-	-	-
Barren & Lt. Veg.	-	-	-	-	-	10	3
Water	-	-	-	-	-	6	12

$\frac{171}{424}$ = 40.3% Correct

Table C.19. EDITOR Supervised Ground Plot Evaluation - Community Level

Covertype	Reference Data										
Classified	Dense Conifer	Sparse Conifer	Deciduous Forest	Mixed Forest	Burn	Brush	Grasslands	Sparse Vegetation	Barren	Silty Water	Clear Water
Dense Conifer	33	35	-	10	-	2	-	-	-	-	-
Sparse Conifer	-	-	-	-	-	-	-	-	-	-	-
Deciduous Forest	16	17	140	53	2	25	4	2	-	-	-
Mixed Forest	-	-	-	-	-	-	-	-	-	-	-
Burn	-	-	-	-	46	-	-	-	-	-	-
Brush	3	-	-	1	-	1	-	2	-	1	-
Grassland	-	-	-	-	-	-	-	-	-	-	-
Sparse Vegetation	-	-	-	-	-	-	-	-	-	-	-
Barren	-	-	-	-	-	-	-	4	6	3	-
Silty Water	-	-	-	-	-	-	-	-	6	8	-
Clear Water	-	-	-	-	-	-	-	-	-	-	4

$\frac{238}{424}$ = 56.1% Correct

Table C.20. EDITOR Supervised Ground Plots Evaluation - Generalized Level

Covertype	Reference Data						
Classified	Conifer	Deciduous	Mix	Brush and Grass	Burn	Barren and Lt. Veg.	Water
Conifer	101	32	46	6	-	-	-
Deciduous	-	108	17	25	2	2	-
Mixed Forest	-	-	-	-	-	-	-
Brush & Grass	3	-	1	1	-	2	1
Burn	-	-	-	-	46	-	-
Barren & Lt. Veg.	-	-	-	-	-	10	3
Water	-	-	-	-	-	6	12

$\frac{278}{424}$ = 65.6% Correct

Table C.21. EDITOR Unsupervised Ground Plot Evaluation - Community Level

Covertype	Reference Data										
Classified	Dense Conifer	Sparse Conifer	Deciduous Forest	Mixed Forest	Burn	Brush	Grasslands	Sparse Vegetation	Barren	Silty Water	Clear Water
Dense Conifer	38	32	-	15	9	1	-	-	1	-	4
Sparse Conifer	12	12	-	5	-	3	-	-	-	-	-
Deciduous Forest	1	3	134	43	3	14	4	1	-	-	-
Mixed Forest	-	-	-	-	-	-	-	-	-	-	-
Burn	-	-	-	-	36	-	-	-	-	-	-
Brush	1	5	6	1	-	10	-	3	1	-	-
Grassland	-	-	-	-	-	-	-	-	-	-	-
Sparse Vegetation	-	-	-	-	-	-	-	-	-	-	-
Barren	-	-	-	-	-	-	-	4	7	1	-
Silty Water	-	-	-	-	-	-	-	-	3	11	-
Clear Water	-	-	-	-	-	-	-	-	-	-	-

$\frac{248}{424}$ = 58.5% Correct

Table C.22. Unsupervised Ground Plot Evaluation - Generalized Level

Covertype	Reference Data						
Classified	Conifer	Deciduous	Mix	Brush and Grass	Burn	Barren and Lt. Veg.	Water
Conifer	94	-	20	4	1	9	4
Deciduous	3	132	43	12	3	-	-
Mixed Forest	-	-	-	-	-	-	-
Brush & Grass	7	8	1	16	3	5	-
Burn	-	-	-	-	33	-	-
Barren & Lt. Veg.	-	-	-	-	-	11	1
Water	-	-	-	-	-	3	11

$\frac{297}{424}$ = 70.0% Correct

Table C.23. EDITOR Modified Cluster Training Evaluation - Community Level

Covertype	Reference Data										
Classified	Dense Conifer	Sparse Conifer	Deciduous Forest	Mixed Forest	Burn	Brush	Grasslands	Sparse Vegetation	Barren	Silty Water	Clear Water
Dense Conifer	15	18	-	3	-	-	-	-	-	-	-
Sparse Conifer	29	23	-	17	25	4	-	-	-	-	-
Deciduous Forest	-	2	128	39	5	-	2	-	-	-	-
Mixed Forest	7	6	4	5	7	-	-	-	-	-	-
Burn	-	-	-	-	-	-	-	-	-	-	-
Brush	1	3	8	-	11	23	1	2	-	-	-
Grassland	-	-	-	-	-	1	1	6	2	-	-
Sparse Vegetation	-	-	-	-	-	-	-	-	-	-	-
Barren	-	-	-	-	-	-	-	-	6	1	-
Silty Water	-	-	-	-	-	-	-	-	4	11	-
Clear Water	-	-	-	-	-	-	-	-	-	4	-

$\frac{216}{424}$ = 50.9% Correct

Table C.24. EDITOR Modified Cluster Training Evaluation - Generalized Level

Covertype	Reference Data						
Classified	Conifer	Deciduous	Mix	Brush and Grass	Burn	Barren and Lt. Veg.	Water
Conifer	85	-	20	4	25	-	-
Deciduous	2	128	39	2	5	-	-
Mixed Forest	13	4	5	-	7	-	-
Brush & Grass	4	8	-	26	11	10	-
Burn	-	-	-	-	-	-	-
Barren & Lt. Veg.	-	-	-	-	-	6	1
Water	-	-	-	-	-	4	15

$\frac{265}{424}$ = 62.5% Correct

Table C.25. EDITOR Modified Cluster Ground Plot Evaluation - Community Level

Covertype Classified	Reference Data Dense Conifer	Sparse Conifer	Deciduous Forest	Mixed Forest	Burn	Brush	Grasslands	Sparse Vegetation	Barren	Silty Water	Clear Water
Dense Conifer	15	18	-	3	-	-	-	-	-	-	-
Sparse Conifer	36	29	4	22	-	4	-	-	-	-	-
Deciduous Forest	-	2	130	16	5	16	3	6	2	-	-
Mixed Forest	-	-	-	23	-	-	-	-	-	-	-
Burn	-	-	-	-	43	-	-	-	-	-	-
Brush	1	3	6	-	-	8	-	2	-	-	-
Grassland	-	-	-	-	-	-	1	-	-	-	-
Sparse Vegetation	-	-	-	-	-	-	-	-	-	-	-
Barren	-	-	-	-	-	-	-	-	6	1	-
Silty Water	-	-	-	-	-	-	-	-	4	11	-
Clear Water	-	-	-	-	-	-	-	-	-	-	4

$\frac{270}{424}$ = 63.7% Correct

Table C.26. EDITOR Modified Cluster Ground Plot Evaluation - Generalized Level

Covertype	Reference Data						
Classified	Conifer	Deciduous	Mix	Brush and Grass	Burn	Barren and Lt. Veg.	Water
Conifer	98	4	25	4	7	-	-
Deciduous	2	130	16	18	5	-	-
Mixed Forest	-	-	23	-	-	-	-
Brush & Grass	4	6	-	10	-	10	-
Burn	-	-	-	-	36	-	-
Barren & Lt. Veg.	-	-	-	-	-	6	1
Water	-	-	-	-	-	4	15

$\frac{318}{424}$ = 75.0% Correct