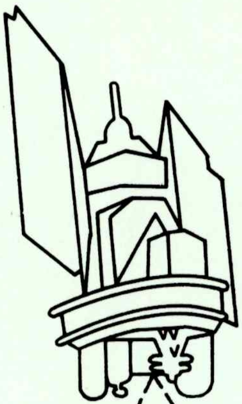
LARS Technical Report 101880



Evaluation of Landsat Data Analysis for Forest Survey

R. P. Mroczynski
R. M. Hoffer
R. F. Nelson

Department of Forestry & Natural Resources
in cooperation with the
Laboratory for Applications of Remote Sensing
Purdue University West Lafayette, Indiana 47906 USA

1980

EVALUATION OF LANDSAT DATA ANALYSIS FOR FOREST SURVEY

by

R. P. Mroczynski
R. M. Hoffer
R. F. Nelson

Department of Forestry and Natural Resources
in cooperation with the
Laboratory for Applications of Remote Sensing
Purdue University
West Lafayette, Indiana 47906

Final Report

on Cooperative Research Agreement No. 13-571

Submitted to

North Central Forest Experiment Station
U.S. Forest Service
St. Paul, Minnesota 55108

TABLE OF CONTENTS

List of Figures

List of Tables

1.0 INTRODUCTION

The Forest Service of the U.S. Department of Agriculture was directed by the McSweeney-McNary Forest Resources Act of 1928 to periodically review the timber supply demand situation and outlook in the United States. These surveys are to be conducted periodically by the Regional Forest Experiment Stations on a state-by-state basis. Results of these surveys are published as Forest Service Research Bulletins. New legislation, the Forest and Rangeland Resources Planning Act of 1974 and the National Forest Management Act of 1976, expand the amount of information needed about the status of our renewable natural resources beyond that required in the past. Furthermore, the frequency to update this information has been mandated by Congress to a 10-year cycle.

This legislation places increased demand on the Forest Service to provide timely information on the status of the nation's renewable resources. The series of earth-orbiting resource satellites--Landsats 1, 2, and 3--which were first launched by NASA in 1972, provide a means for collecting data on our renewable resources in a timely manner. The satellites collect data from a sun-synchronious orbital altitude of 900 kilometers using a 4-band multispectral scanner. These data are available in either image or computer compatible digital tape format. Each Landsat frame contains data for an area of approximately 185 x 185 kilometers. Computer processing of this data offers a potential for rapidly and effectively providing pertinent information concerning the forest resources for an area of interest. The literature describes a number of investigations which have utilized computer processing techniques and Landsat satellite data to provide various types of information related to the renewable natural resources. Generally, these articles and reports indicate a reasonable high correlation between computer classifications of the Landsat data and the ground reference for a forest/non-forest degree of detail.

1.1 Background and Objectives

During 1976, personnel in Purdue University's Forestry and Natural Resources Department, working at the Laboratory for Applications of Remote Sensing (LARS), studies the utility of satellite remote sensor data for application to the 1974 Resources Planning Act. This study involved an evaluation of Landsat classification results for assessing forest acreage estimates using a test site that comprised 158 counties in 16 Survey Units within four states (Figure 1). The results indicated the potential of computer-aided analysis of Landsat data for Forest Survey (Hoffer et al., 1978).

With these results as background, Purdue's Department of Forestry and Natural Resources entered into a cooperative agreement with the U.S.D.A. Forest Service, North Central Forest Experiment Station (Cooperative Agreement No. 13-571). The objectives of this research were to:

1. Compare Landsat classification results to current survey statistics for forest/non-forest classes in the northeastern aspen-birch survey unit in Minnesota. These results will be used to assist the feasibility of these techniques to provide:
    a. The first step in a land change monitoring system.
    b. Estimates of cost of local and regional land use statistics obtained from Landsat.

c. Information regarding the degree of information that can be extracted from Landsat.

2. Determine the possibility of detecting disturbance on survey plots so that more precise updates can be made using the North Central FREP (Forest Resource Evaluation Program) system.

3. Determine if forest survey ground plots can be used as training for computer-aided Landsat classifications.

2.0 METHODS AND MATERIALS

Our approach involved testing various machine-assisted analysis techniques of Landsat data, and the evaluation of their suitability to address the stated objectives. A review of the materials and methods used to conduct this study appear in Sections 2.1 and 2.2 respectively.

2.1 Materials

The test site, Figure 2, was defined as the Northeastern Aspen-Birch Survey Unit in Minnesota. Carlton County was selected for all classification work, because it was the only one of the five counties in the unit that was totally covered by a single Landsat scene.

The Landsat data (Scene ID: 1345-16313) collected 7/03/73 was available in-house at LARS. This data has been formatted to meet the specific requirements of a previous project. Because of the format required, 27% of the data cells had been deleted from the original data tape, and then, unfortunately, the original CCT had not been retained in the LARS data bank. Eventually, this format would cause us problems in precisely identifying survey plots in the data.

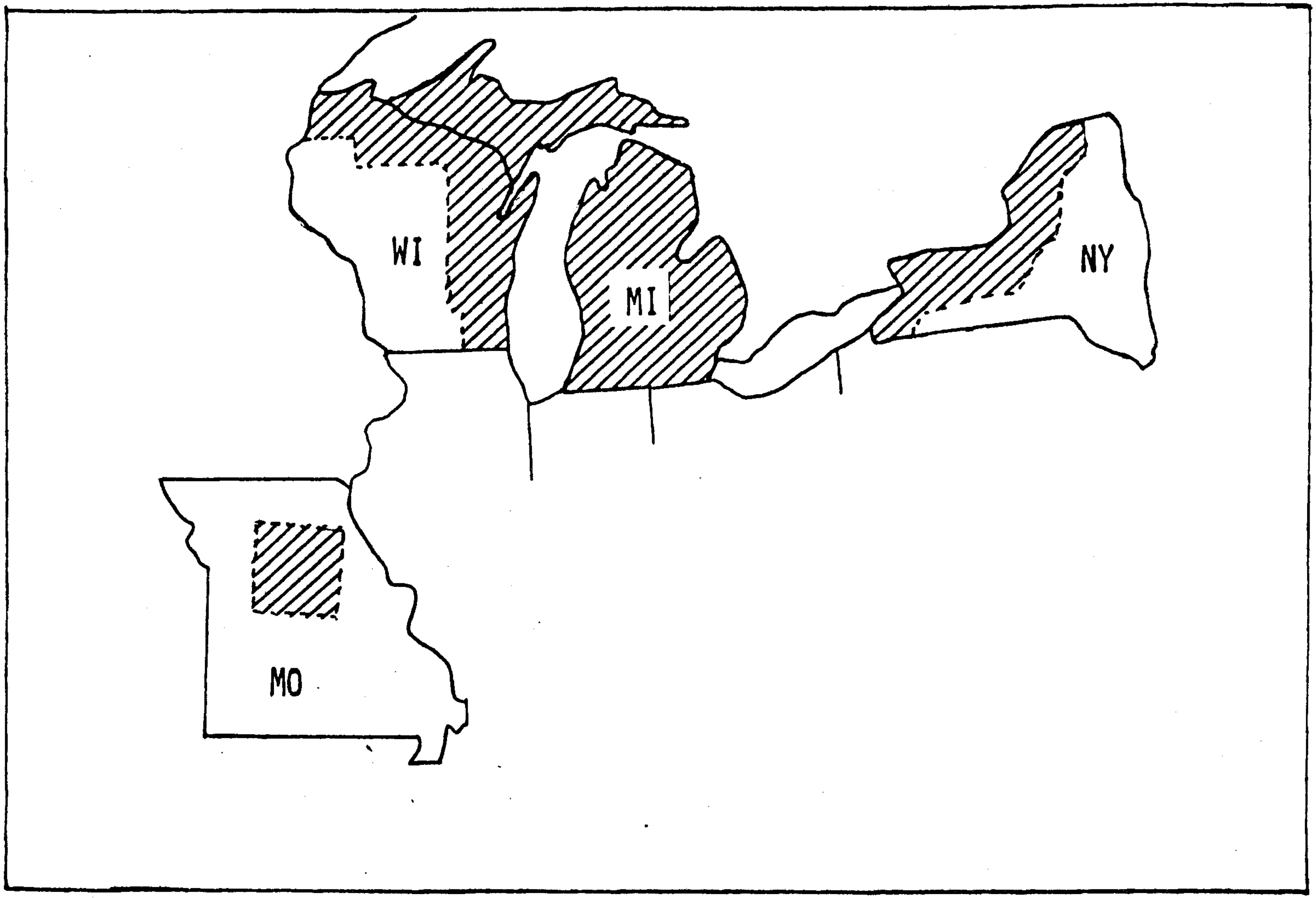


Figure 1. Areas in states where both Landsat and Forest Survey results were available for county comparisons.

To aid in classifying this data, the Forest Survey provided the following information:

- Microfilm records of photo location and photo-interpretation information.
- Computer listing of survey plot information, including:
  - a) plot classification listings,
  - b) stereo classification listing, and
  - c) plot variance listings.
- Photo dot records and township uncontrolled mosaics for Carlton County.
- Quadrangle center, blue line ortho-photographs with type annotations.

These materials represent a cross section of information available from a typical forest survey.

Analysis of Landsat multispectral scanner data may be accomplished with various computer-aided techniques. Four different approaches were evaluated as part of this study. The concepts underlying these various approaches are discussed in Section 2.2 .

2.2 Numerical Classification Overview

The concept behind most computer-aided analysis techniques involves a man/machine interaction, whereby the man "trains" the computer to recognize specific combinations of numbers that represent reflectance measurements in each of several wavelength bands, for the cover types of interest. This training process usually involves fairly limited areas for which accurate information exists concerning the types and condition of the ground cover. After a representative set of training statistics have been developed, the computer is programmed to classify the reflectance values for each resolution

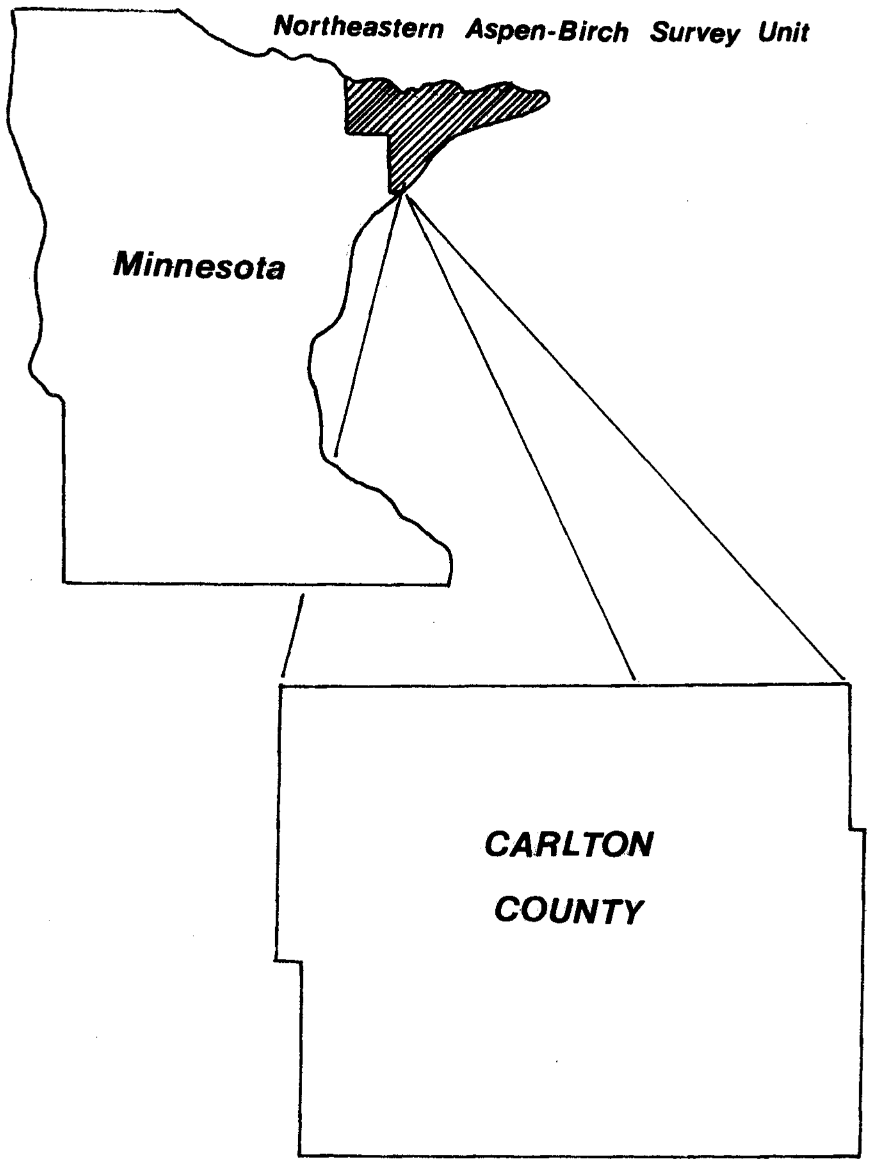


Figure 2. The area selected for study was the Northeastern Aspen-Birch Survey Unit in upper Minnesota. Carlton County was selected for intensive investigation.

element (or a statistical sub-sample thereof) in the entire data set. In this way, the speed of the computer is used to advantage and a large geographic area can be mapped and acreage tabulations obtained at a much faster rate than would be possible using standard image interpretation techniques.

2.2.1 The Supervised Technique

A common approach used for developing training statistics is the "supervised training fields" technique, whereby the analyst designates to the computer the X-Y coordinates of "training fields" of the various cover types which have informational value. A data deck of training statistics is developed, based upon the spectral characteristics of the designated training fields. The entire area of interest is then classified using any one of several different classification algorithms available. One of the most frequently used and statistically rigorous algorithms is the "maximum likelihood algorithm" which is based upon an assumption of a Gaussian distribution of the data. The supervised training technique has been used quite effectively for agricultural mapping (Bauer and Cipra, 1973), and has also been utilized for several forestry applications, but with varying degrees of success (Bryant and Dodge, 1976; Williams and Haver, 1976; Mead and Meyer, 1977). Previous experience at LARS with Landsat data indicated that for classification of forest cover, the cover types of interest are often not spectrally homogeneous, and therefore the supervised technique does not yield acceptable accuracy or reliability. The primary reason for this is the difficulty for the analyst to define locations in the data that represent all significant variations in spectral response for every cover type of interest.

2.2.2 The Non-Supervised Technique

A second approach to analyzing MSS data is referred to as "clustering" (sometimes called the "nonsupervised" technique). There are different clustering algorithms available which vary in terms of the input parameters which the analyst must specify, but one of the simplest and most effective is the LARSYS CLUSTER algorithm. With this algorithm, the analyst simply designates the area to be classified and the number of spectrally distinct classes into which the data should be divided. The computer is programmed to classify the data into the designated number of spectral classes and prints out a map indicating which resolution elements in the data belong to which spectral classes. The analyst then relates this classification output map to aerial photos or surface observation data, and determines which resources are represented by each of the spectral classes. The appropriate spectral classes can then be combined to print out maps and tables showing the informational categories of interest. Experience at LARS (and elsewhere) has shown that this technique effectively overcomes the primary limitation of the "supervised" approach, but when working with large areas, the amount of computer time involved in the interactive clustering sequence makes this technique very expensive (Fleming, et al., 1975). In addition, the number of spectral classes defined is often very large, since a single cover type of interest is usually represented by several spectral classes. In areas where the vegetative cover is complex (e.g., small stands, variations in stand density, species mixtures, etc.) it is often difficult to reliably relate each of the spectral classes defined by the computer to the vegetative cover type.

2.2.3 The Multi-Cluster Blocks (MCB) Technique

In addition to the two common and rather basic analysis techniques just described, a hybrid approach called the "Multi-Cluster Blocks Technique" has been developed at LARS in which various aspects of the "supervised" and the "clustering" techniques are combined (Fleming and Hoffer, 1977). In this method, several small blocks of data are located, each of which contains several cover types and spectral classes. Each data block is individually clustered, and the spectral classes defined in the clustering process are related to informational classed by standard photo interpretation methods. Software has been developed which allows the analyst to merge the spectral classes for all cluster areas into a single data deck, which should statistically describe the spectral characteristics of all cover types in the entire study site. In most situations, each cover type of interest is represented by several spectral classes in the MSS data. Research has shown that the MCB technique requires significantly less analyst time and computer time, and that the classification results are more accurate (as much as 14% higher) than use of either the Supervised or Unsupervised techniques (Fleming and Hoffer, 1977). The MCB approach also requires minimal reference data as input (usually a small sample of aerial photos) and allows an effective man/machine interaction with the data throughout the analysis process.

2.2.4 The Procedure-1 (P-1) Technique

The fourth technique for developing training statistics to be discussed is known as Procedure-1. This technique involves a grid sample of point data of known cover type which is used to automatically label the spectral classes defined by a clustering algorithm. Because a gridded sample of point data is the basis for the entire Forest Survey procedure, the P-1 technique would seem to offer considerable potential as a method to effectively utilize Forest Survey data in the analysis of Landsat satellite data. The P-1 technique was developed by personnel at the NASA Johnson Space Center (Wills, Gardner and Aucoin, 1977) during the LACIE (Large Area Crop Inventory Experiment) program.

In the P-1 approach, an array of X-Y coordinates is defined and the cover types or informational classes of interest are determined, using image interpretation or some other source of information. The entire area (or some sub-sample thereof) is then clustered using the ISOCLAS cluster algorithm, and the cover type information from the array of X-Y coordinates is used to automatically label the spectral cluster classes (i.e., attach an informational class label to each of the spectral classes defined by the clustering algorithm). The spectral data associated with the X-Y coordinates can also be used to "seed" the clustering algorithm in order to improve the efficiency of the clustering process. Once the cover type associated with the array of X-Y coordinates has been defined, P-1 is a very automated approach, requiring very little man/machine interaction. P-1 was thoroughly tested and applied to agricultural situations during LACIE, but has not been tested for possible applicability to forest cover mapping.

2.3 Classification Techniques Used

Modified versions of the Supervised, Multi-Cluster Blocks, and Procedure-1 techniques were used in this study. The specific procedures followed are described below in Sections 2.3.1, 2.3.2, and 2.3.3 . In addition, as an outgrowth of some of the results obtained with these three techniques, a fourth technique for relating Landsat data and Forest Survey data was developed and tested during the current study. This fourth technique is based upon ratios of Landsat spectral data, and is described in detail in Section 2.3.4 .

To provide an overview of the four analysis techniques actually used in this study, Table 1 summarizes the characteristics, advantages and disadvantages, and an indication of relative cost differences.

2.3.1 Supervised LARSYS

To take advantage of the available Forest Survey data, it was decided that the Supervised approach would be utilized, with the training fields being statistically defined by the Survey point locations. Each training field would be at least 3 x 3 pixels in size, centered upon the Survey data point. The training data locations were defined for Carlton County using the Black and White microfilm records of plot locations and 1:62,500 greyscale Landsat maps of the County, bands 4 and 7 (Figure 3).

Ninety-seven of 208 forested points shown on the microfilm were located. These points were used both for training the classifier and testing the classifier performance. These points were identified by:

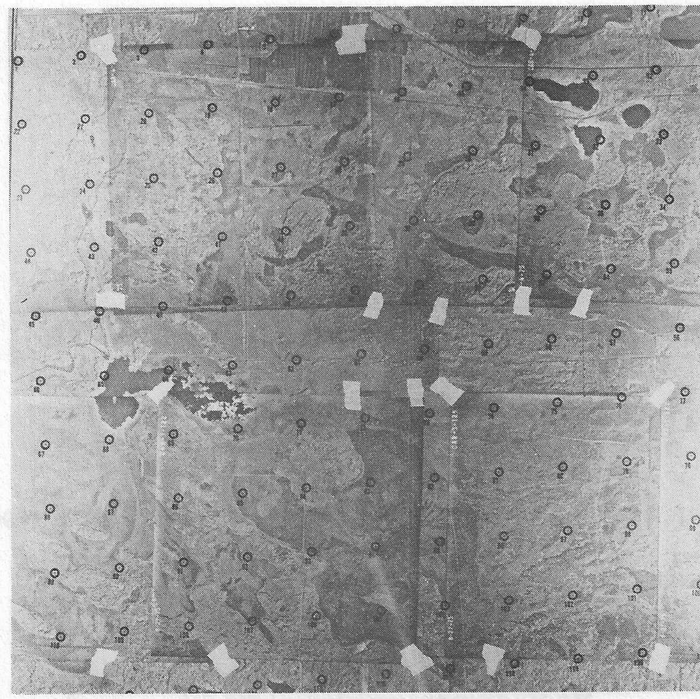
a. locating a feature identifiable both on the photo and on the Landsat maps,

b. measuring vertical and horizontal photo distances with the aid of a microfilm reader and engineer's scale,

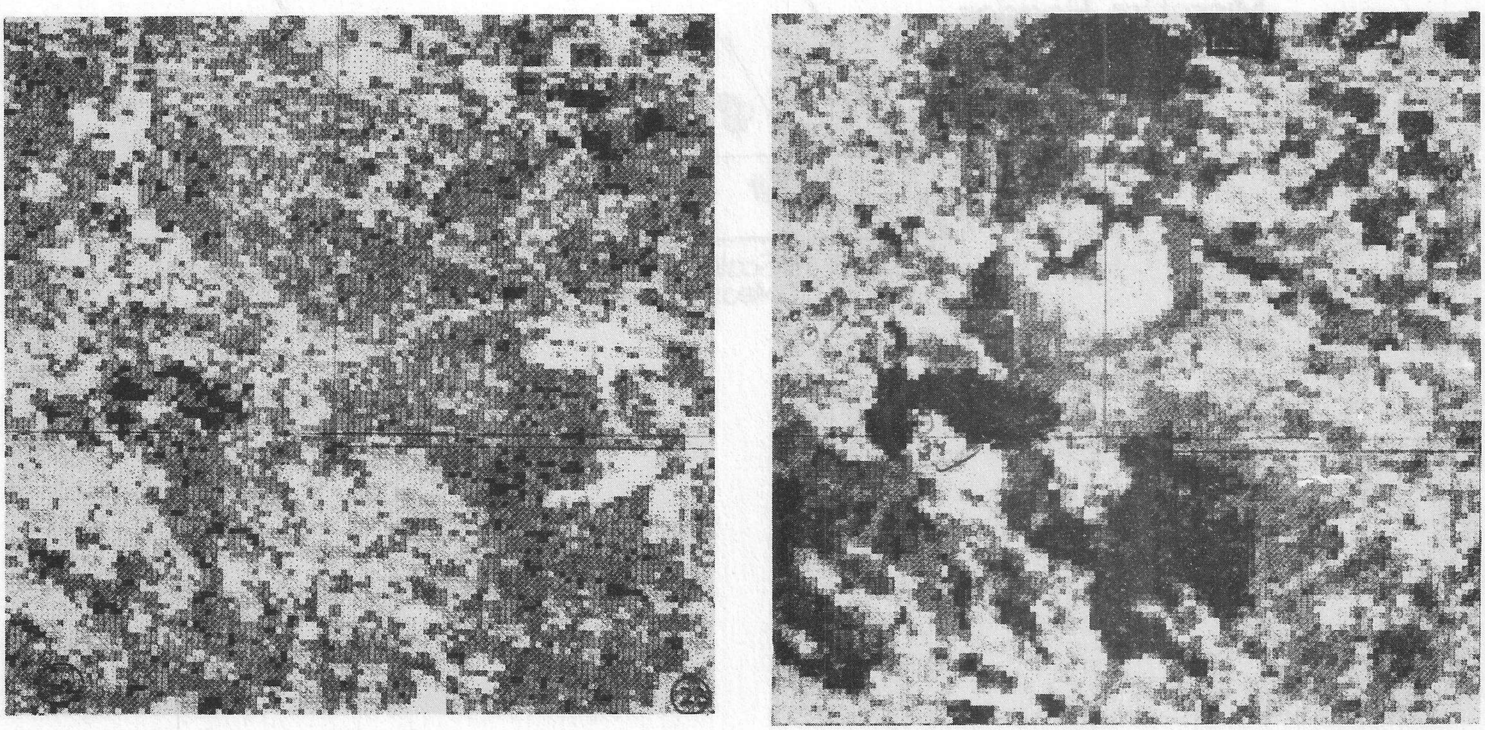
Table 1. General Summary of Analysis Techniques Used.

Analysis Techniques	Characteristics	Advantage	Disadvantage	Relative Cost*
Supervised Technique	Analyst selects homogenous training fields	Suitable where large training and test fields can be identified in data, and analyst has knowledge of site.	Training fields must be homogenous. Difficult to find large fields in Landsat data.	4
Multicluster Blocks Technique	Analyst selects training blocks that exhibit a diversity of cover	Applicable to most land cover mapping projects where blocks containing known diversity can be identified.	Requires good local knowledge of scene diversity for cluster blocks.	2
Procedure-1	Known points of information are used to seed cluster algorithm	Minimizes analyst input into training classifier.	Programs limited to size of test area. Analyst must have good knowledge of composition of seed points.	3
Ratio	Survey points located in data and IR/visible ratio calculated	Minimizes analyst input and eliminates some training bias.	Must be able to transform line/column information to UTM to precisely locate points.	1

*Relative Cost; 1 to 4, where 1 = lowest cost relative to 4



a



b

Figure 3. Example of data types available to the analyst:
a) Photo mosaic of T48N, R19W showing sample grid, and
b) Electrostatic printer/plotter grayscales of Landsat band 4 (0.5-0.6 μm) (left) and band 7 (0.8-1.1 μm) (right) of the same area as shown in a.

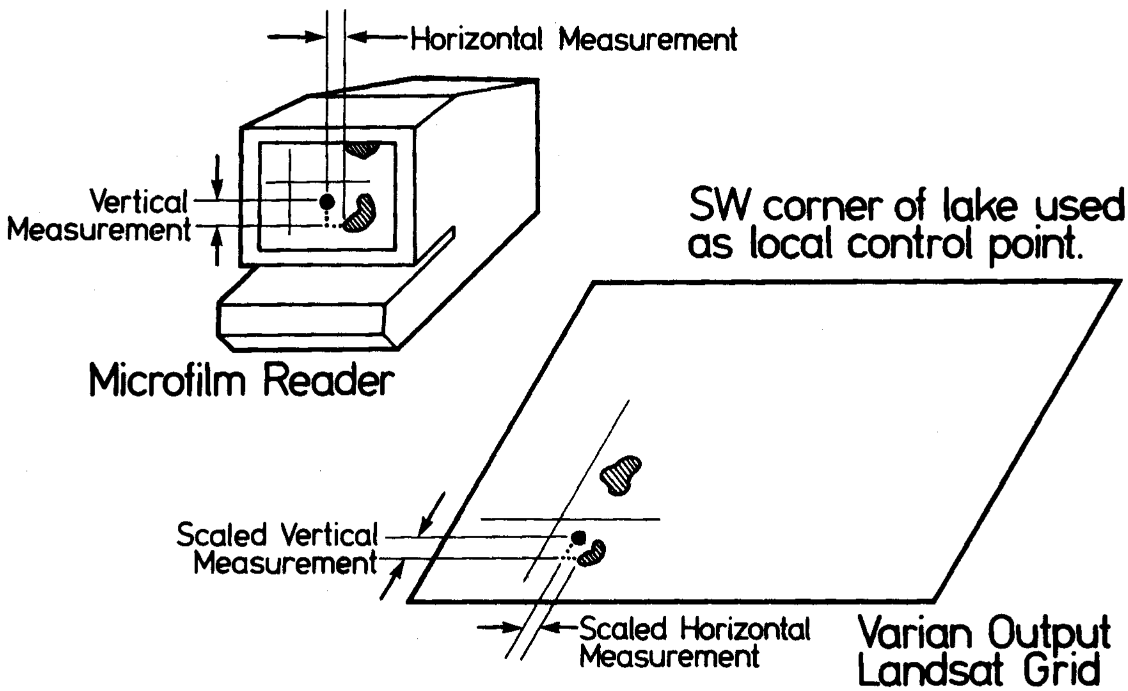


Figure 4. Method of locating plots on Landsat data using USFS microfilm records of photo point locations.

c. converting photo scale distances to map scale distances,

d. measuring vertical and horizontal distances from the control point on the Landsat map to the point in question (Figure 4),

e. determining that the point was in the middle of a homogeneous 3 x 3 pixel (10 acres) block.

One hundred eleven points could not be used because:

a. the photos did not have locatable features identifiable on the Landsat map,

b. the point was not located in the middle of a 3 x 3 pixel (10 acres) block.

Of the 97 points located, 82 were identified by photo-interpretation as hardwood points, predominantly aspen-birch, the remaining points were softwood. Blocks established around the points varied in size from a minimum block of 3 x 3 pixels to a maximum block of 5 x 5 pixels. Hardwood and softwood training statistics were generated using the LARSYS statistics processor. Because the procedure described did not enable an adequate set of non-forest statistics to be defined, the regular supervised approach was modified somewhat, and non-forest statistics were developed by clustering large blocks (40 x 40 pixels or larger) in which the non-forest classes could be readily identified using the available data.

The forest and non-forest statistics were then merged into a single set of training statistics and every other pixel of every other line of the data which comprised the Carlton County data were classified (Figure 5, Classification I).

Due to considerable uncertainty concerning the exact location of the training pixels in the first classification and how well they were correlated with the photo-interpreted plot and ground plot locations, it was decided

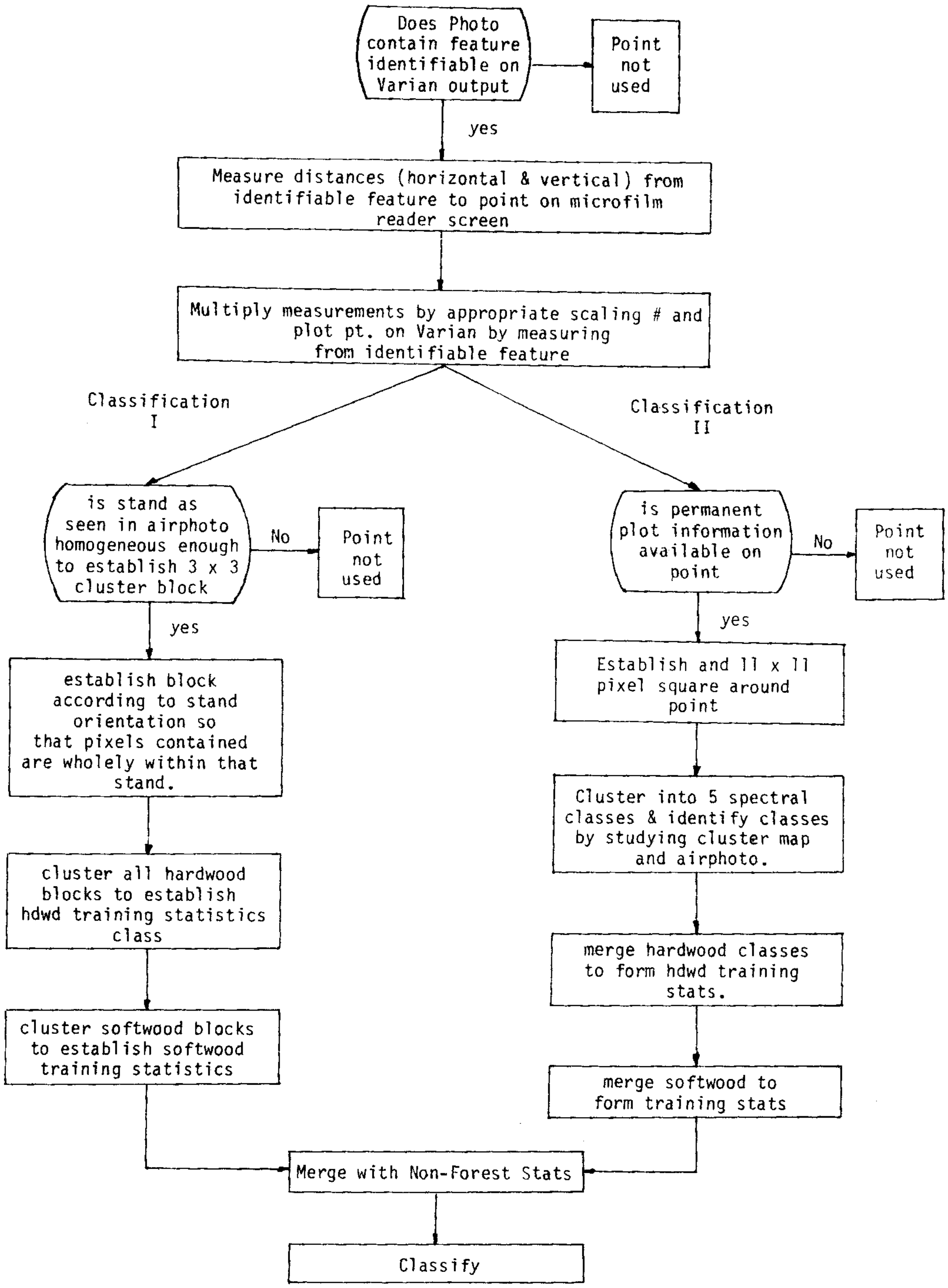


Figure 5. Flowchart of the procedure used to relate Survey plot locations to Landsat data and the two approaches used for classifying the data.

that a technique should be used for which the geometric characteristics of the data would not be as critical. Therefore, the second classification involved the Multi-Cluster Blocks technique.

2.3.2 Modified Multi-Cluster Blocks Approach

The second classification test involved a modification of the "multi-cluster blocks" approach. Usually, in the Multi-Cluster Blocks approach, the analyst manually selects a number of heterogeneous cluster blocks. Neither the size nor the number of blocks is fixed, but the analyst attempts to define blocks that, taken together, will be representative of the entire study site. In this classification, it was decided to use the MCB approach for clustering and grouping the data, but the cluster blocks to be used would be 11 x 11 pixels in size, and centered on the Forest Survey plot locations. Thirty-nine plots had been field checked, so were used for the cluster blocks. After each of these 39 blocks were clustered, the spectral classes were identified by comparing the 11 x 11 pixel cluster map with the B&W airphoto mosaic of the same area. The hardwood and softwood spectral classes from each block were pooled to form the forest training statistics.

The non-forest training statistics used in the first classification were again used in this second classification, since they were believed to be representative, and had been developed using a clustering algorithm. The non-forest and forest spectral classes were then merged to form the final training statistics deck used for the classification (see Figure 5, Classification II).

2.3.3 Modified P-1 Approach

The third classification involved only a relatively small area in Carlton County. As originally developed, the Procedure-1 technique is based on the identification of pixels by an analyst that are located in a regular array of X-Y coordinates. This coordinate array is a rectangular block approximately the size of a normal township. Since the Forest Survey information is collected on a regular grid of points, by township, the Procedure-1 approach seemed especially suitable to existing Forest Service data.

Township 46N, Range 19W contained many identifiable features including lakes, major highways, towns, road intersections, and distinctively shaped agricultural fields. These features helped to correctly orient the 121-point grid used by Forest Survey to identify photo points and which we used to identify seed pixels for the Procedure-1 classification.

Photo interpreted information was available for all 121 grid points, but only 108 were used as input to Procedure-1. Thirteen were excluded because (a) the point fell outside the Forest Service's inventory area (penciled "O" on the grid tally sheed), or (b) the identification of the point was questionable (green "Q" on the tally sheet), or (c) the point was identified as a windrow of trees - a mixed pixel point (purple "I" on tally sheet). Figure 6 is an example of the photo-interpretation tally sheet.

The grid point locations were transferred to the Landsat data by measuring distances between points from the mosaic grid and converting into a row-column grid that could be laid over the township area delineated on the 1:24,000 line-printer output (see Figure 7). Landsat data coordinates of the 121-points were obtained with the aid of a table digitizer after the analyst had defined the Landsat coordinates of the township's corner points.

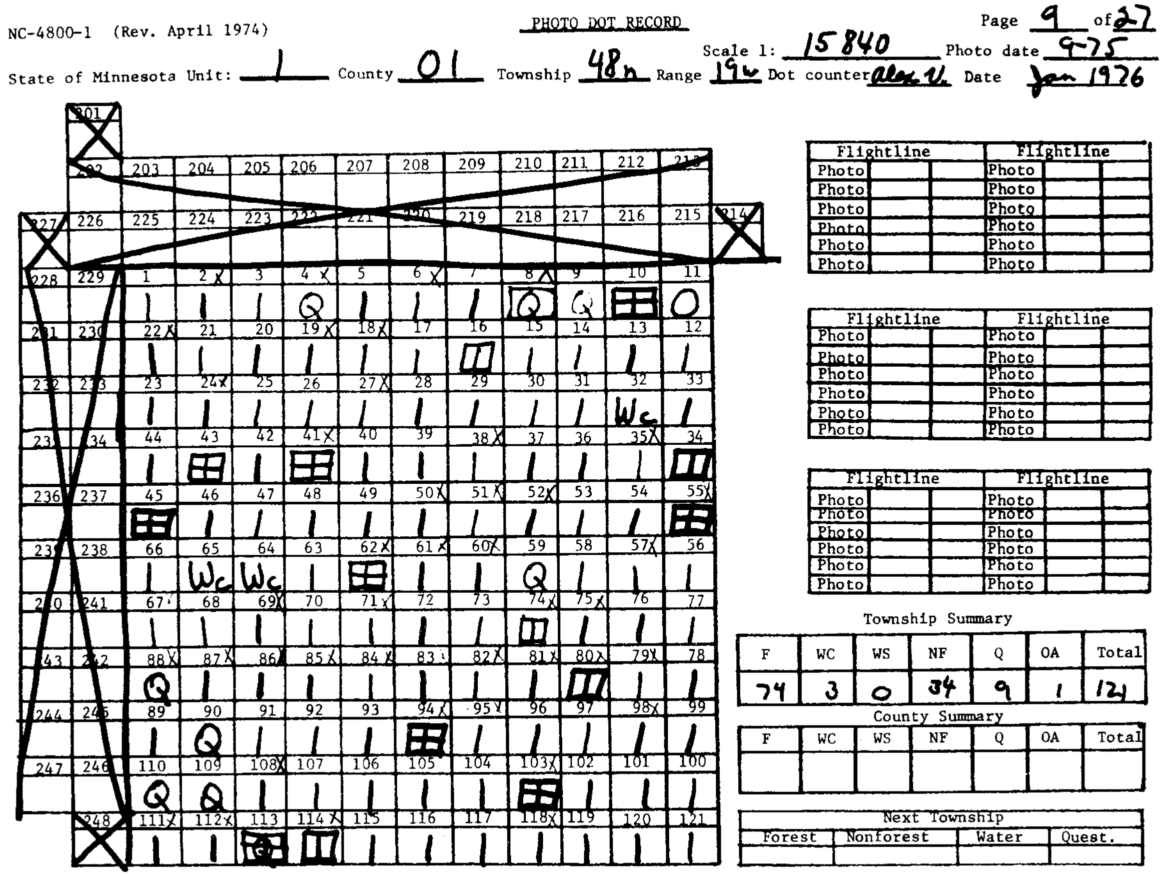


Figure 6. An example of the Photo Dot Record form that the analyst used to identify classes for Procedure-1.

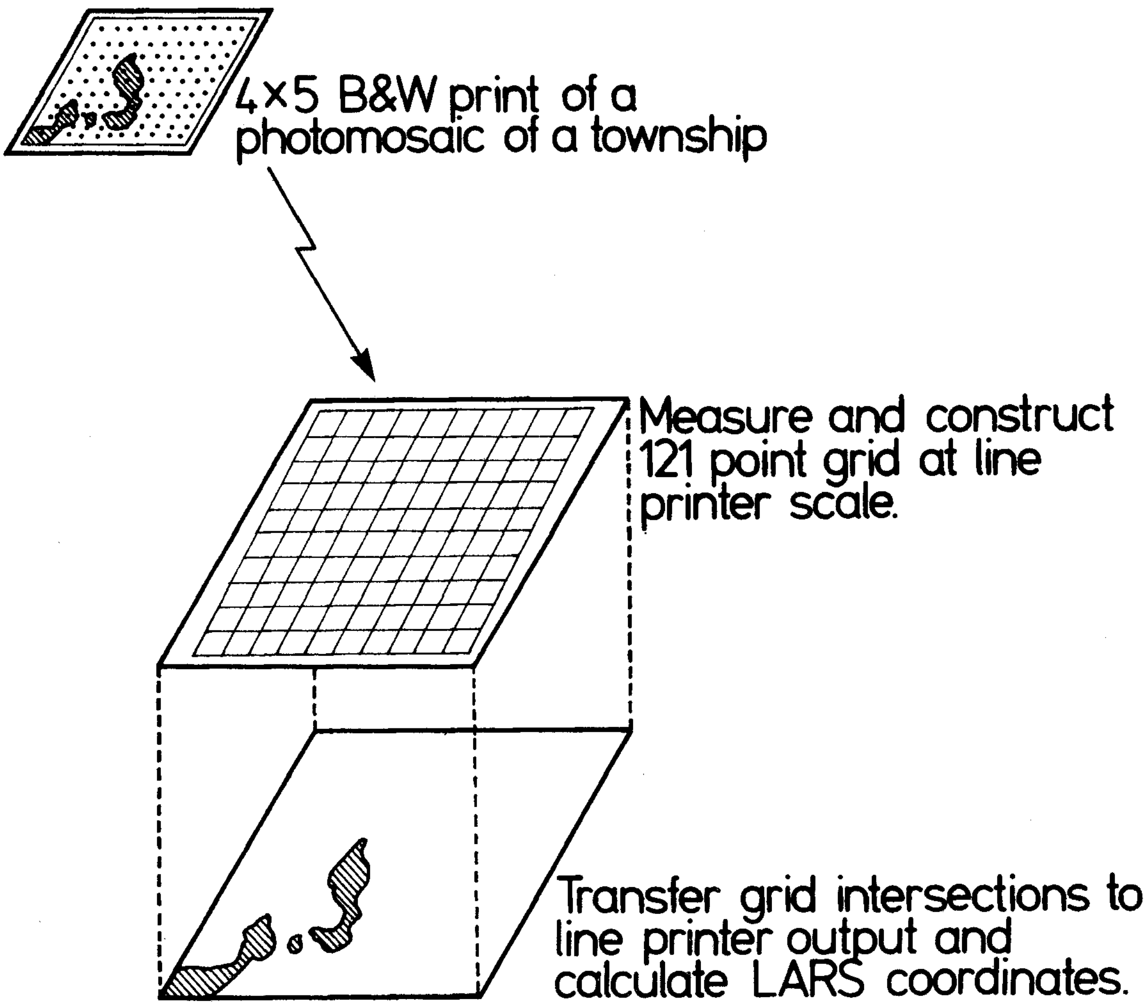


Figure 7. Sequence followed to locate Forest Survey plot locations on the line printer output of Landsat data.

Landsat coordinates obtained were within ± 2 pixels (approximately 160 meters) of their true ground location. This is an estimated error based on the type of geometric correction applied to the data set after it was received from the EROS Data Center. A flow chart for the steps used in classifying data with Procedure-1 is given in Figure 8.

2.3.4 Ratio Evaluation Scheme

A fourth approach that was evaluated during this study involved use of a ratio technique. The concept tested involved evaluation of the ratio values of select pixels (i.e., Forest Survey plots) on anniversary Landsat data. A change in the ratio from one date to another would indicate a change in status of the forest point. Special software was developed and tested during this study to determine if the concept is valid and potentially useful to the Forest Survey.

Information was obtained from 167 survey points located in Carlton County. The points were located in forested lands, non-forested lands, and in water. Plot sequence numbers and land-use and type information for those points was supplied by the Forest Survey. UTM coordinates were also supplied for each plot number. Hence, for each survey point, the following information was available: 1) Plot number; 2) Land Use; 3) Species type - if applicable; 4) UTM coordinates.

Our first task involved locating the 167 points on the Landsat grid. Control points, points identifiable both on the Landsat grid and on USGS topographic maps (with UTM markings), were established (at least one control point per township) and adjusted to reduce possible location error. A second program then took the UTM coordinates of any given field point, and calculated the Landsat line and column coordinates.

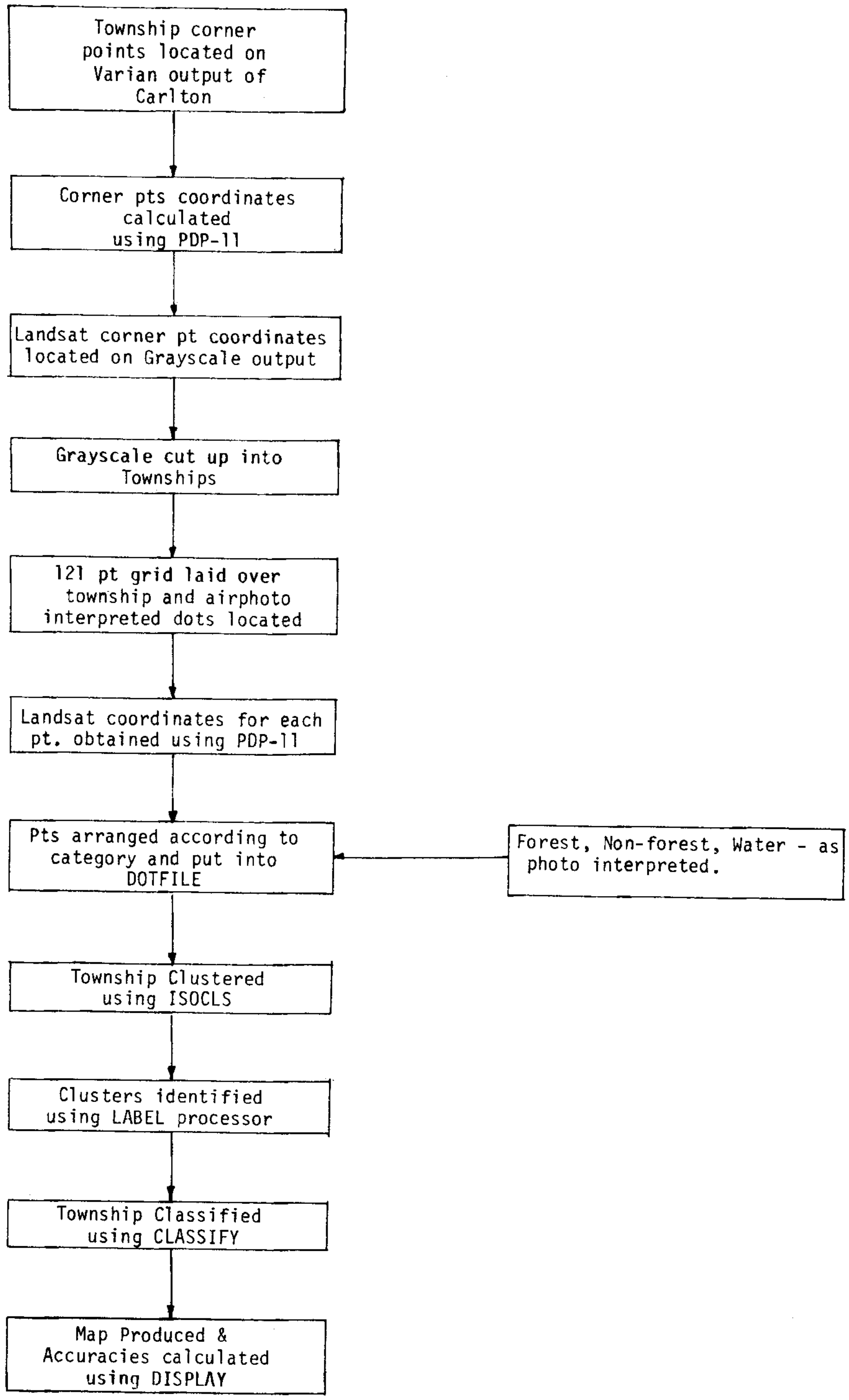


Figure 8. Flowchart of the steps used for Procedure-1 classification.

Knowing the Landsat coordinates for the 167 points, a set of programs were written to:

1. compile the 4 channel multispectral reflectance values associated with each point, and
2. compile the IR/Visible reflectance ratio.

Table 2 lists the programs developed, their functions and the actual time required to run these programs on this data set. The IR/Visible ratio is computed as:

$$\text{Ratio} = \frac{R_6 + R_7}{R_4 + R_5}$$

where

R_n = reflectance value for band n

n = reflectance values for band 4, 5, 6 or 7

Such a ratio provides a dimensionless number in the range of 0.2 to 3.5 for most naturally-occurring earth resource materials. For example, since green vegetation has a relatively high reflectance in the reflective infrared portion of the spectrum, but relatively low reflectance in the visible wavelengths, a ratio of approximately 2.0 or greater is often obtained from vegetative cover. On the other hand, water absorbs a great deal of the incoming reflective infrared radiation relative to the amount visible radiation absorbed. Hence the IR/Visible ratio for water is very low, commonly less than 0.50. Exposed soil areas commonly have a ratio of about 1.0 (Kristof and Zachary, 1971; Tarnoci and Kristof, 1976).

A significant shift in ratio values from one date to the next indicates a change <u>may</u> have taken place in the cover class. However, in addition to a change in cover class or condition, ratio values between two dates may

Table 2. Names and functions of programs developed to compare Survey plot ratio values, along with the computer times required to run the data for 167 plots.

Program Name	Function	time (sec)*
CHECK	Establish control point coordinates	15.92
CLOSE	Calculates line column coordinates for survey plot location from closest control point	4.57
TRANSFERDATA	Retrieves 4-channel reflectance values from coordinates provided by CLOSE	93.95
RATIO	Calculates IR/VIS ratio for points	5.48
COMPARE	Compares ratios from one date to another	8.11

*time is based on seconds required to execute programs on an IBM 370/148 computer.

also be affected by time of growing season and, to a lesser extent, atmospheric condition. By using reasonable good quality anniversary data, ratios provide a means whereby points which may have undergone change can be flagged.

If the procedure of comparing pixel ratio values from subsequent years of Landsat data were followed, Forest Survey would have the capability to:

1) Track gross forestland losses between inventory dates, and
2) Have a capability to update forest area/volume statistics between periods.

A test of this ratio concept required the comparison of two data sets. Unfortunately, a second data set of suitable quality (e.g., cloud-free during the growing season) was not available. A simulated data set was created by incrementing the 167 original coordinate pairs by two. This essentially moves the point over and down (east and south) approximately 650 feet (1 pixel $\approx$ 1.1 acres). Software was developed to compare the ratio of the reflectance values of the new point to the original ratio. If the new ratio changed by more than 20% of the first ratio, the point was flagged. Figure 9 represents a flow chart of the procedure followed.

The 20% criterion for determining significant ratio changes was selected as a beginning point. In the future, options other than a strict percentage of a ratio should be considered. One possibility is to flag any point whose ratio changes by more than one standard deviation for a given land use and type. Another possibility is to analyze both the ratios and magnitudes (sume of the data values for all channels) in order to detect changes.

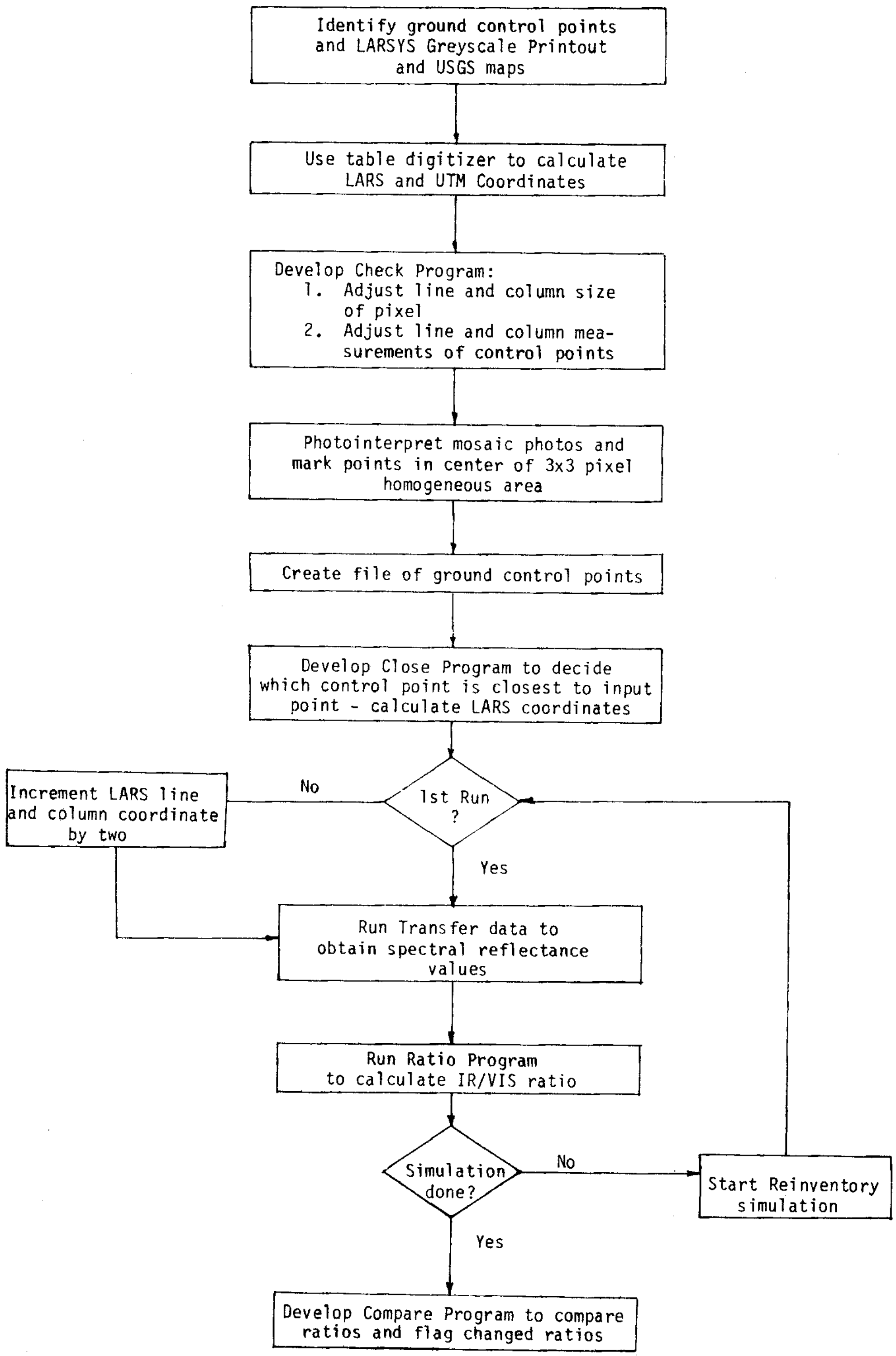


Figure 9. Flowchart of the approach defined for determining a potential change in forest cover on Survey Plot locations.

3.0 RESULTS

The Landsat data for Carlton County was classified using the methods described in Section 2.3 . Results of the various classification approaches were evaluated to determine if the methodology provided information suitable for use by the Forest Survey. Two different evaluation criteria were employed. The first involved the calculation of classification performance based on an evaluation of test fields. The second was a comparison between forest acreage classified on the Landsat data and the most recent Forest Survey estimate. Table 3 gives the results of both comparisons. Because the ratio approach was not used as a classifier, results using this technique are not presented in this Table, but are discussed in detail in Section 3.1.4 .

3.1 Discussion

3.1.1 Supervised LARSYS

Eleven spectrally separable classes were defined, four hardwood, two conifer, one water, and four Other classes (1 Urban, 1 Cloud, and 2 Agricultural). Only 97 plots were successfully located on the Landsat grid, and therefore the same areas had to be used to develop the training statistics and to test the classification results. Hardwoods were classified correctly 85% of the time, softwoods 75.6% of the time. The overall training field performance was 84.1%, the average training field performance by class was 80.3%.

Although these results are fairly high, previous work would indicate that they are not as high as one might expect for a classification that is

Table 3. Comparison of classification accuracy and area estimates for three classification schemes tested on Carlton County, MN. USFS Survey data were used to assess accuracy and for areal comparisons.

Approach	Classification[A] Performance (Percent)	Difference in AREAL ESTIMATE from USFS Survey Data (Percent)
Supervised Technique		
Forest	84.1	+14.1
Hardwood	85.0	+ 2.9
Softwood	75.6	+11.2
Multi-Cluster Blocks Technique		
Forest	81.5	+ 9.6
Hardwood		+ .3
Softwood		+ 9.3
Procedure-1[B]		
T46N $19W	84.9	-
Forest		+ 3.6
Non-Forest		+ 3.9

A. Classification Performance = Percent pixels correctly identified as forest.

B. Results for this classification scheme are reported for one township only.

no more complex than simply defining hardwood and coniferous forest cover types. The primary limiting factor to the classification performance was the fact that only 97 points could be used for developing the training statistics. Previous work involved in developing training statistics indicate that several hundred pixels are necessary to thoroughly characterize the mean and variance of the spectral response for the various cover type classes of concern. Therefore, improved classification and mensurational results over those presented in Table 3 could be achieved through the use of more training data.

It should be noted that the performance figures shown in Table 3 may not, in fact, represent true classification accuracy. The reason is that the figures to which the classification is compared are themselves only estimates. We have, for purposes of this study, assumed that the Forest Survey acreage estimates are "correct." Minimally, one would hope to see a high degree of correlation and reasonably similar acreage estimates between the data previously obtained by the Forest Survey and the classification estimates obtained from the Landsat analysis. Although the estimates are fairly close for the hardwood category, there is considerable discrepancy in the conifer category. We believe this is probably due to the small number of softwood points (15) used for training.

3.1.2 Multi-Cluster Blocks

The spectral classes obtained from each of the five cluster blocks were merged to form one Hardwood class, one Conifer class, one Water class and one class to include Other land uses. Classification results were evaluated using the same test fields as used in the Supervised classification.

The major problem encountered in using the Multi-Cluster Blocks technique involved the lack of detailed information concerning the cover type actually present within the blocks being clustered to develop the training statistics. However, we believe that the clustering approach using blocks of data around the ground sample points offers a great deal of potential as an effective means of developing training statistics. However, in this study, the ability to do accurate and detailed photo interpretation from the microfilm data prevented an effective definition of training statistics. Furthermore, detailed sub-groups of the major cover types were not defined in the initial training phase and later evaluation of the results indicated that this probably decreased the classification accuracies. Based on this experience, we feel that an effective method could be defined using the unsupervised classification or clustering technique in conjunction with the aerial photos of the areas around permanent sample plot locations. We believe that an accurate and effective classification could be achieved based upon this approach.

3.1.3 Procedure-1

Table 4 indicates that in the first attempt involving use of the P-1 technique, 94% of the points originally identified as forest were classified as forest, 29.2% of the non-forest points were correctly classified, and 100% of the water points were correctly identified. The low performance in the non-forest catefory (i.e., agricultural classes) raised questions about the points defined as non-forest. Previous experience with classification of agricultural scenes suggests that these results should have been higher. Therefore, we refined the statistics of the non-forest data points, and reclassified the data set. The classification accuracies of the second classification showed that 84.9% of the forested points, 79.4% non-forest, and 100% of the water points were correctly identified (see Table 4).

Although the Procedure-1 approach has a great deal of potential for automatically developing effective training statistics, it is very much dependent upon the accurate location of known cover type classes. In this case the Forest Survey data provides the cover type class identifications; the limiting factor was the ability to define the location of the training pixels in the Landsat data to the Forest Survey photo-interpreted or ground plot locations. We had considerable doubt concerning the reliability of the location of many of the data points because of some of the geometric problems in the data set. Therefore, the classification results achieved with the Procedure-1 approach are believed to be somewhat questionable. More importantly, this study showed that the Procedure-1 technique takes a relatively large amount of computer time even for classifying relatively small areas. This may be a function of the clustering algorithm used and the fact that it is not optimized for our computer.

3.1.4 Ratio Technique

The ratio technique, as applied to the Carlton County data, did not result in a classification *per se*. That is, the technique was not used to identify every pixel in the scene of interest, produce a map and sum the pixels by class to produce tabular acreage summaries. The approach followed was to locate and evaluate only those pixels that corresponded to Forest Survey plots. Conceptually, pixel ratios for survey plots would be compared from the Landsat data for year 1 with the Landsat data for year 1 with the

Table 4. Classification results (two iterations) for T46N, R19W, Carlton County, Minnesota using the Procedure-1 approach.

	Accuracy (%)	
Cover Type	First P-1	Second P-1
Forest	94.8	84.9
Non-Forest	29.2	79.4
Water	100	100
#training points	108	88
#clusters formed	30	20

Landsat data for year N. If a significant change in the ratio value occurred between the data sets, that area would be a candidate for a change in cover type.

Table 5 is an example of the output from the ratio process. This example is for the Forest Survey Type 70 or the Elm-Ash-Cottonwood type. The results presented have been simulated for two years of Landsat data. This particular example indicates that for the six survey points of Type 70, one point may have changed. This point is identified in the table by the flag: "**Following Ratio Differences Exceed Limit, Possible Type Change**". For purposes of demonstration, we define that a change may have occurred if the percent change in the ratio values between time one and time n + 1 (where n + 1 is an anniversary Landsat data set), is 20% or greater. Using this criteria, the third point on Table 5 exhibits a change of negative 20.35% between the two ratio values and therefore has been identified as a possible change.

4.0 COST EVALUATION

An important component of this study revolved around the projected cost of implementing this technology in the Forest Survey scheme. To provide a meaningful comparison of costs, we have assumed that the image processing software and appropriate hardware are available and that the important items to compare are computer time and human resource time. Throughout the course of this study, we maintained a detailed record of the time required for each classification approach. The computer CPU times and analyst times required for both the computer and analyst were then converted to costs using a rate of $250/hr for CPU time and $10/hr for analyst time.

Table 5. Change Detection Output for Elm-Ash-Cottonwood.

	Line	Column	Channels 1	2	3	4	Ratio IR/VIS	Data Sum	Sequence	Land Use	Type	Quotient of Ratios	Percent Change
Original	44	428	22.	13.	61.	37.	2.80	133.	5166	10	70	94.29	-5.71
Incremented			23.	15.	40.	33.	2.64	131.					
Original	152	97	24.	14.	51.	29.	2.11	118.	5198	10	70	116.59	16.59
Incremented			23.	14.	58.	38.	2.48	135.					
* * * * * * * * * * Following Ratio Difference Exceeds Limit. Possible Type Change * * * * * * * * * *													
Original	141	557	23.	15.	55.	33.	2.26	127.	5242	10	70	79.65	-20.35
Incremented			26.	20.	54.	29.	1.80	129.					
Original	312	396	22.	12.	58.	34.	2.54	124.	5297	10	70	96.85	-3.15
Incremented			24.	15.	58.	38.	2.46	135.					
Original	443	194	23.	13.	59.	37.	2.67	132.	1430	10	70	95.13	-4.87
Incremented			24.	13.	57.	37.	2.54	131.					
Original	445	645	23.	16.	53.	32.	2.23	126.	5385	10	70	91.03	-8.97
Incremented			20.	13.	42.	25.	2.03	100.					

Mean IR/VIS Ratio and Standard Deviation for
Land Use 10 and Type 70

Mean	2.40
Standard Deviation	.27
Number of Cases	6

Table 6 clearly indicates that the supervised technique requires the most analyst time whereas the Multi-Cluster Blocks technique is much more efficient in terms of the analyst time required. However, the more automated Procedure-1 technique required considerably less analyst time than the Multi-Cluster Blocks. On the other hand, the Procedure-1 approach required considerably more computer time than either the Multi-Cluster Blocks or Supervised technique, primarily because of the cluster algorithm utilized by Procedure-1. A total estimated cost is therefore much higher for Procedure-1 than either of the other approaches, and the Multi-Cluster Blocks technique has the lowest overall cost.

Our past experience in classifying large areas of Landsat data also indicates that the Multi-Cluster Blocks approach is the most effective technique in terms of the accuracy of the classification. It should be noted, however, that a classification of the entire North-Central region would cover an area of approximately 420 thousant square miles and therefore even at a cost of approximately $40 per hundred square miles, the overall classification costs for the entire area would be considerable. In addition, because of the time required to select, obtain, and classify the number of Landsat scenes required to cover the entire North-Central region on an annual basis, it would appear that a complete classification of the entire data set for the region could be done only periodically and that none of the three classification techniques evaluated would be appropriate for an annual assessment of significant changes in the forest cover types.

Since the ratio approach does not require the extensive mathematical calculations that are utilized in any of the classification algorithms, a massive number of point ratios can be calculated with a relatively small expenditure of computer time. Likewise, analyst time is minimal since there

Table 6. Comparison of Estimated Resource Requirements necessary to classify Landsat MSS data for USFS Forest Survey purposes.

Approach	CPU time/100[1/] Sq. mi. (minutes)	Man-hrs/100[2/] Sq. mi.	Cost Estimates ($/100 sq. mi.)[3/] Data Acquisition & Reformatting	CPU Cost	Analyst Cost	Estimated Total Cost
Supervised Technique	1.7	6.32	$2.60	$ 7.08	$63.20	$ 72.88
Multi-Cluster Blocks Technique	2.7	2.57	$2.60	$ 11.25	$25.70	$ 39.55
Procedure-1	39.5	0.73	$2.60	$164.58	$ 7.30	$174.48

1/ Computer time used is based on analysis using an IBM 370/148 computer.

2/ Includes all the time associated with classifier training and classification.

3/ Cost estimates are based on rates of $250/hr CPU time for the computer, $10/hr for analyst time, and $335/Landsat frame for data acquisition and reformatting costs ($200/frame=acquisition; $135/frame= reformatting). Each frame of Landsat data contains 13,242 sq. mi.

is no requirement to identify training fields or training pixels. Thus, the value of the ratio technique offers a great deal of potential for minimizing computer and analyst costs to a point where even an annual evaluation of significant change in cover type for all Forest Survey points in the North-Central region might be realistic.

Because of recent changes in Landsat processing techniques, data obtained from the EROS Data Center in the future will be geometrically and radiometrically corrected. This should allow the UTM coordinates of Forest Survey points to be located quickly and accurately and should allow a ratio evaluation of the Survey points to be carried out in a relatively automated manner. A key point for such an operational scheme, however, would hinge on the availability of data containing only the UTM coordinate points of interest to the Forest Service. At present, each tape of Landsat data contains over 7.5 million individual one-acre pixels and costs $200 when obtained from the EROS Data Center. Over 60 such frames of data would be required to cover the entire North-Central region. The number of Forest Survey plots would only involve 121 pixels per township or roughly 33,000 per tape or 1.4 million for the entire region. Therefore, if the Forest Service could obtain a preprocessed tape from the EROS Data Center containing only the points of significance and of interest, this would provide a much more realistic capability for carrying out such a change detection analysis on a routine, operational basis. We would hope that with the designation of Landsat as an operational system under the direction of NOAA, procedures will be placed into effect for providing user agencies such as the Forest Service with data sets appropriate to their particular needs.

5.0 SUMMARY AND CONCLUSIONS

Key results and conclusions in relation to the classification of Landsat data can be summarized as follows:

1. Landsat data and computer-aided analysis techniques can provide for a 100 percent sample of cover type acreages within Forest Survey Units.
2. Forest Survey Plots must be accurately located in the Landsat data in order to provide effective training and evaluation data.
3. Ground plot data alone cannot adequately train the classifier. The plot data in conjunction with additional photo-interpretation would provide sufficient training if a Multi-Cluster Blocks sampling approach were used to develop the training statistics.
4. Ground and photo plots located adjacent to cover type boundaries are not suitable for training a supervised classifier or evaluating classification results, due to variability in precisely locating the plots in the Landsat data.
5. Ground and photo plots that are not adjacent to cover type boundaries could provide a method of using Forest Survey data to train the classifier using the Procedure-1 classification approach. However, this technique is expensive when compared to the other techniques evaluated in this investigation.
6. Only geometrically corrected Landsat data sets should be used for land and forest cover classifications.
7. A Multi-Cluster Blocks sampling approach is recommended as the best method to develop training statistics.

The following statements relate to the results of this study that were directed toward the use of a ratio technique for evaluation of major changes in land use, rather than actually classifying the data.

1. UTM coordinates of Survey plots are extremely useful in locating the line/column address of the plot in the Landsat data.
2. Programs have been written to relate Landsat data to UTM coordinates by (a) use of a set of control points locatable on both the Landsat data and maps, or (b) use of NASA-registered data sets (to be available in the future).
3. Geometric correction of Landsat data is not required, so the ratio procedure would be relatively economical and effective even with uncorrected data sets.
4. Using the ratio approach, anniversary data sets do not have to be digitially overlayed, again enabling this approach to be cost effective.
5. The ratio technique is computationally simple, thereby enabling it to be very fast and cost effective even when large quantities of data are involved.
6. Although the ratio approach appears to be an effective tool to monitor Survey Plot status, there needs to be more work in assessing the geometric relationship of the pixel to the plot before an operational procedure can be developed.
7. The significance of various ratio values in terms of cover types, and the difference in ratio values required to indicate a significant type of change in land cover must be studied further.

Our overall conclusions would be that accurate geometric correction of Landsat data is a prerequisite to effective classification if one is to use existing Forest Survey plot locations as input into the training or test phase of the analysis sequence. Secondly, the Multi-Cluster Blocks approach appears to be the most promising for developing an effective set of training statistics, although the Procedure-1 does have some potential if it can be made more cost effective. Use of ground control points from the Forest Survey data for development of training statistics as the only source of ground truth does not appear to be a feasible approach if a maximum likelihood classifier is to be used for analyzing Landsat data. Classification of Landsat data does not need to involve the classification of every pixel in the Landsat data set. However, development of an effective set of training statistics would allow classification of a sub-set of the Landsat data which could improve the sample efficiency currently being used by the U.S.F.S. to estimate acreages of major cover types.

A ratio technique, in which UTM coordinates of Forest Survey plot locations are used to evaluate cover type characteristics and to monitor cahnges from one year to another, appears to offer a great deal of potential as a cost effective method for evaluating major changes in land use patterns over extensive areas at frequent intervals. It is believed that the ratio approach has distinct advantages in terms of calibration requirements of the Landsat data and efficiency in computer analysis costs. However, further work needs to be done to evaluate the significance of the various ratio values obtained.

6.0 RECOMMENDATIONS FOR FURTHER WORK

The significance of the ratio values as a function of cover type and the importance of changes in solar altitude and atmospheric conditions on the ratio values themselves needs to be investigated further. Such a study should evaluate the temporal variability of spectral response values in the Landsat data and how this changes as a function of season.

A methodology for circumventing cloud cover problems in portions of the data during a period of several weeks in the summer would need to be developed in order to obtain a nearly cloud-free data set showing all or most of the Survey Plot locations during a single summer-long time period.

A prototype system to locate Forest Service Survey plot locations in Landsat data that has been geometrically corrected, and to determine ratio values for these plots for anniversary data sets of Landsat data should be developed and evaluated. It would appear that one could utilize Landsat data that has been geometrically corrected but has not been overlayed, and define and locate only those points where significant changes in the ratio value have occurred in anniversary data sets. Such a system could be very effective in providing the Forest Service with a method for monitoring major changes in land use at frequent intervals in a cost effective manner. The economic value and potential of such a system would seem to indicate that further work toward the development of such a procedure should be pursued.

7.0 REFERENCES

Tarnocai, C., and S. J. Kristof. 1976.

"Computer-aided classification ofIland and water bodies using Landsat data, Mackenzie Delta Arca, N.W.T., Canada." Artic and Alpine Research, Vol. 8, No. 2, pp. 151-159.

S. J. Kristof and A. L. Zachary. 1974.

"Mapping soil features from multispectral scanner data." Photogram. Engr., Vol. XL, No. 12, pp. 1427-1434.

Hoffer, R. M., S. C. Noyer, and R. P. Mroczynski. 1978.

A Comparison of Landsat and Forest Survey Estimates of Forest Cover." in Proc.: Fall Technical Meeting of the American Society of Photogrammetry, Albuquerque, New Mexico. pp. 221-231.