1-1-1980

Procedure 1 and Forestland Classification Using Landsat Data

Ross F. Nelson
National Aeronautics and Space Administration (NASA)

Roger M. Hoffer
Purdue University

Nelson, Ross F. and Hoffer, Roger M., "Procedure 1 and Forestland Classification Using Landsat Data" (1980). *LARS Symposia.* Paper 384.
http://docs.lib.purdue.edu/lars_symp/384

Reprinted from

Symposium on

Machine Processing of

Remotely Sensed Data

and

Soil Information Systems

and

Remote Sensing and Soil Survey

June 3-6, 1980

Proceedings

The Laboratory for Applications of Remote Sensing

Purdue University
West Lafayette
Indiana 47907 USA

IEEE Catalog No.
80CH1533-9 MPRSD

Copyright © 1980 IEEE
The Institute of Electrical and Electronics Engineers, Inc.

PROCEDURE 1 AND FORESTLAND CLASSIFICATION USING LANDSAT DATA

ROSS F. NELSON
NASA/Goddard Space Flight Center

ROGER M. HOFFER
LARS/Purdue University

ABSTRACT

Procedure 1 approaches to developing land cover classifications were compared with the Multicluster Blocks process on a 15,000 hectare forested area in southwestern Colorado. Results showed that P-1 (using the clustering processor in an unseeded, iterative mode) performed as well as the Multicluster Blocks approach on the rugged study area. The average accuracies of classification for the best P-1 method and the McB approach were 77.8 and 75.3 percent respectively; overall accuracies were 88.3 and 87.4 percent respectively.

These results may interest the forestry community or any resource discipline which has available to it ground-checked or photointerpreted point (or plot) information. Procedure 1 can use this information directly to output a land cover classification with little analyst interaction.

I. INTRODUCTION

A number of computer-aided analysis techniques have been developed which utilize Landsat MSS data in order to produce land-use classifications of agricultural or forested areas. Two such techniques were compared in this study. Both require that training statistics be developed so that the classifier may statistically recognize spectral-informational classes of interest. However, the two methods used to develop the statistics differ markedly. The Multicluster Blocks approach to developing training statistics requires that the analyst select relatively small, heterogeneous blocks in the study area, cluster each block individually, identify the spectral classes in each block, and merge the statistics from the blocks to form the final statistics deck used by the classifier. The Procedure 1 approach requires that pixels (also called dots) of known identity must be located in the study scene. The entire area is clustered and the spectral classes formed are identified using the dots. The analyst need only locate and identify the dots (assuming processor parameters are properly set); the rest of the work is done by the computer.

The objective of this study was to evaluate the effectiveness of P-1's automated approach in a complex forestland situation. The forestry community may be interested in the Procedure 1 concept for two reasons. First, P-1 provides a method whereby forest inventory information may be utilized directly in the computer-aided classification process. The inventory information, currently available on private, state, and national forests, can be located in the Landsat scene and used by the computer to identify the spectral classes formed by the P-1 clustering processor. Second, P-1 is a semi-automated approach which merits scrutiny. Perhaps the ultimate goal in Landsat data processing is the development of a completely automated procedure which would output an accurate, reproducible classification of a given scene[1]. Though this goal may never be realized due to the complexities and variability of nature, steps may be taken to reduce analyst interaction in the classification process. Analysis techniques such as P-1 may allow much of the analyst interaction found in many current techniques (for instance, Multicluster Blocks or the supervised techniques) to be significantly reduced.

Procedure 1 was developed for the Large Area Crop Inventory Experiment (LACIE) and has been used extensively to develop land-use classifications of agricultural areas[2,3] with varying degrees of success. Reeves (1978) used

P-1 to accurately discriminate rangeland from nonrangeland[4]. A study using Procedure 1 to classify forested areas has been completed by Lockheed personnel at the Johnson Space Center, but the results of that study are not available at this writing.

In its narrowest context, Procedure 1 refers specifically to the method of operation outlined for LACIE where

1. 209 random dots (pixels) are systematically located within a 5 x 6 mile segment of a Landsat scene and identified as to cover type by image interpreters,

2. a portion of the dots are used as initial clustering centers by the P-1 cluster processor,

3. the spectral reflectance values of dots and clusters are compared by the computer (geometric distance) and a cluster is given the identity of the closest dot,

4. the segment is classified using a Sum-of-Normal-Densities classifier,

5. a portion of the 209 dots are used to evaluate classification accuracy, and

6. all P-1 processors--DOTDATA, ISOCLS, LABEL, CLASSIFY, and DISPLAY--have fixed parameter levels.

This experiment attempted to follow and build on the P-1 concept developed during LACIE. However, it should be noted that Procedure 1 (as used in this report) does not necessarily refer to the specific P-1 sequence used in LACIE (listed above), but refers to a more general approach. This general approach is one where

1. dots (pixels) are located in the study area, and are identified using ancillary information,

2. the dots are used in both the clustering and labeling steps in the analysis (in this more generalized approach, processor parameters and inputs may be adjusted--see Figure 1),

3. a Sum-of-Normal-Densities classifier or some other algorithm may be used for the classification, and

4. test fields are used to obtain classification accuracies.

In order to investigate P-1's capabilities, various P-1 approaches were compared to the Multicluster Blocks

(McB) approach to developing training statistics. McB has been shown to be an effective method of developing training statistics on complex forested areas in a number of studies conducted at Purdue[5,6] and at Humbolt State University[7]. The McB approach was used as a yardstick against which the merits of P-1 were measured. The analysis sequence used in the two approaches, McB and P-1, are outlined in Figure 1.

II. STUDY AREA

The study site was located in the eastern half of the 844,000 hectare (2,086,484 acre) San Juan National Forest in southwestern Colorado. The Devil Mountain quadrangle, encompassing 15,156 hectares (37,450 acres), lies approximately 48 kilometers (30 miles) east of Durango in the San Juan Mountains. The terrain on the quadrangle is rugged (elevations vary from 2,000 to 3,100 meters) and the forest cover types are strongly influenced by topographic position.

Ponderosa pine (Pinus ponderosa) flanks the lower elevations of the Piedra River Valley which roughly bisects the quadrangle. The ponderosa pine may be dense enough to generate a parklike grasslands understory, but in less dense stands it is often found in conjunction with Gambel oak (Quercus gambeli). As elevation increases or aspect changes from south to north, pine is found with, and is gradually replaced by, quaking aspen (Populus tremuloides), Douglas fir (Pseudotsugo menziesii), Englemann spruce (Picea engelmannii), and subalpine fir (Abies lasiocarpa). Aspen offtimes forms large pure stands, perhaps as a result of small lightning fires or harvesting activities. Clearcuts in various stages of regeneration are noticeable, and relatively small rock outcrops are present.

III. SUPPORTING INFORMATION
AND FACILITIES

A cover type map developed by the Institute for Arctic and Alpine Research (INSTAAR), University of Colorado, provided baseline data on the vegetative characteristics of the Devil Mountain quadrangle. A 1:24,000 USGS 7½ minute quadrangle map had been used as a base for cover type delineation. The type lines were the result of photointerpretation efforts at INSTAAR using CIR photos in conjunction with field checks.

The Landsat 1 data used in the analysis were obtained on June 5, 1973, at 9:20 a.m. local time. The data were

geometrically corrected and rescaled so that 1:24,000 scale lineprinter products could be produced. No clouds were present at the time of the overpass. 1:120,000 scale color infrared photos taken by NASA on June 6, 1973, were also available.

The entire analysis was done on the LARS (Laboratory for Applications of Remote Sensing) IBM 370/148 medium speed computer. Procedure 1 processors are part of the EODLARSYS software system developed at Johnson Space Center, and the five processors used in the Multicluster Blocks approach are part of the LARSYS software system. Both software packages are easily accessed on the 370/148.

IV. METHODS

In order to provide a reasonable comparison of the P-1 method for developing training statistics with the Multicluster Blocks approach, a preliminary processor parameter study was done. This parameter study determined the effects of various clustering and spectral class labeling parameters (ISOCLS and LABEL processors respectively) on the number of clusters formed, CPU time used, and classification accuracies. This initial study yielded a set of parameters thought to be appropriate for developing P-1 training statistics on the Devil Mountain quadrangle.

Three items were necessary to proceed with the comparison of P-1 and McB on the Devil Mountain quadrangle: dots, training blocks, and test fields. The dots were used by various P-1 processors; the training blocks were used in the McB approach. Test fields were common to both.

Dots are individual pixels from within the scene to be classified located according to some unbiased statistical sampling procedure and identified by cover type. In this study, dots were located using a grid[1] (121 points/township) laid atop an alphanumeric lineprinter map of the quadrangle. Because some cover types were inadequately represented by the grid, additional dots in these cover types were selected by the analyst. One hundred ninety-six dots were identified to the level of detail expected in the final classification

_____

1. The grid was fashioned after one used by the USFS to locate photoinventory points in Minnesota[8].

(hardwood, conifer, grass, and barren) using aerial photography and a Zoom Transfer Scope. These dots (line-column coordinates and identity) were input to the first P-1 processor, DOTDATA, which compiled a listing of the multispectral values and associated identities (a dotfile) for use in proceeding P-1 processors (see Figure 1).

Three training blocks, ranging in size from 841 to 1804 pixels, were established in heterogeneous areas on the Devil Mountain quadrangle. These were individually clustered and the spectral classes identified using the CIR aerial photography and a Zoom Transfer Scope. The identified spectral classes were merged into one final statistics deck to form the Multicluster Blocks statistics used by the classifiers.

Manually selected test fields were located in homogeneous areas using the same materials that were used to select the training blocks. The same test fields were used to evaluate each of the classification results listed in Table 1. The test fields contained no dots and were not located in any of the training blocks.

Three P-1 approaches to developing training statistics were compared to the McB approach. Procedure 1 compiles a listing of pixels (dots) identified by the analyst (described above), then clusters the entire area or a systematic subsample of the area using the ISOCLS processor. This processor is both versatile and complicated. ISOCLS as implemented on the EODLARSYS software system puts the data through a certain number of split and combine iterations, the number and order subject to analyst control. On a split iteration, clusters which exhibit standard deviations (in one or more channels) greater than a given parameter (STDMAX) are split, i.e. separated into two new clusters. On a combine iteration, pairs of clusters are evaluated to see if their separability is less than a specified parameter (DLMIN). If the calculated separability is less than DLMIN, the clusters are merged. The analyst also has the option of introducing initial cluster centers from the dotfile. If specific data vectors (dots) are used as initial cluster means, ISOCLS is said to be seeded. The alternative is to let ISOCLS derive cluster centers by splitting the data set[9].

Three different ISOCLS parameter sets were used, a LACIE parameter set[1] (essentially a grouping algorithm using seed dots), a limited iteration parameter set (three split and one combine iterations) using seed dots, and a parameter set which did not use seed dots and performed ten splits and one combine iteration. The four sets of training statistics developed (three P-1, one McB) were input to two different classifiers, a standard maximum-likelihood (ML) classifier and a Sum-of-Normal-Densities (SoND) classifier available as an option on the P-1 classification processor. The experimental design is shown in Table 1. The parameters used for each P-1 run are given in the Appendix.

V. RESULTS

Three P-1 approaches (each using different ISOCLS clustering parameters) were compared to the McB approach. The effectiveness of a given method for developing training statistics was based solely on the quantitative classification performance indicated by the test field results. The results are shown in Table 2.

A Newman-Keuls Range Test was run on the classification performance results of each of the four methods[2]. The results of the statistical ranking are given in Table 3.

The results indicate that the P-1 approach using ISOCLS in an unseeded, iterative mode, and the Multicluster Blocks approach were the two best approaches to developing training statistics. The P-1 approach using LACIE parameters, which essentially makes ISOCLS a one pass grouping algorithm, fared poorly. Evidently the increased heterogeneity of a forestland situation could not be adequately characterized by a clustering processor using parameters developed for more simplistic

_____

1. The LACIE parameter set assigns each pixel in the area being clustered to the "closest" (i.e. minimum euclidian distance) seed dot, then attempts one split and one combine iteration. The LACIE parameters are set so that no splitting or combining is done (see Appendix).

2. The Newman-Keuls procedure for this type of data involved a series of steps described by Landgrebe (1976), p. 2.7-11[10].

agricultural scenes. The limited iteration, seeded clustering approach which gives the analyst a great deal of control over the number and identity of the clusters output, showed promise.

The differences in classification accuracies obtained and CPU time used for the two classifiers (maximum-likelihood and Sum-of-Normal-Densities) were minimal and were not statistically significant, even at the 0.10 alpha level (paried-t test).

VI. FINAL STATEMENT

Procedure-1 has been shown (on one study site) to be a potentially useful method for classifying forested areas, performing as well as the Multicluster Blocks approach. P-1's clustering processor, ISOCLS, is complicated, and it takes no small amount of effort to develop appropriate clustering parameters for a particular site. The parameter set required to obtain reasonably satisfactory results seems to be specific to a particular site and for a given date. Based upon these results, it would appear that a similar number of spectral classes cannot be expected when different study areas (even within the same data set) are clustered using ISOCLS. This experiment did not look into the effects of using the same parameters on the same study area in two anniversary data sets. Multitemporal applications of P-1 to the same study area should be investigated, for P-1's potential usefulness increases greatly if ISOCLS performance is little affected using the anniversary data.

As computer technology improves, the amount of analyst time required to output a given product becomes more and more important (since the cost of the analyst becomes a larger percentage of the overall cost of the final product). Hence the cost-effective classification schemes of the future will be those that minimize the amount of analyst interaction. When used in appropriate situations, the P-1 approach markedly reduces the analyst input requirements. Such a situation might be one where:

1. Procedure 1 is used to monitor a forest tract on a yearly basis to detect major changes. Once parameter levels are set for a given area, gross changes in classification results would most likely be due to changes in the forest cover conditions.

2. Point (plot) information is available on the monitored forest track and can be machine located in the Landsat data. If the forest points or plots have been located on a geographical coordinate system (UTM, state plane, latitude-longitude) then transformations can be calculated using tie points (a simple registration procedure).

Classification maps generated from Landsat data could be used to monitor forest conditions. Such classifications (done yearly) might be used to supplement (not replace) inventory information taken every 5 to 10 years. Hence the forest manager might be better able to detect major changes (burns, major disease or insect infestations, cuttings, or blowdowns) that could affect his management decisions. Procedure 1 provides a potentially useful method whereby such information could be obtained quickly with minimal analyst involvement.

VII. REFERENCES

1. Schubert, J.S. 1978. Computer Processing of Landsat Data for Canada Land Inventory Land Use Mapping. Report #13, Gregory Geosciences Ltd., Ottawa, Ontario. 70 pp.

2. MacDonald, R.B. 1976. Large Area Crop Inventory Experiment. 2nd Annual William T. Pecora Memorial Symposium, Sioux Falls, South Dakota.

3. Heydorn, R.P., R.M. Bizzel, J.A. Quirein, K.M. Abotteen, and C.A. Sumner. 1978. Classification and Mensuration of LACIE Segments. Proceedings of Technical Sessions, Volume 1, the LACIE Symposium, October 1978, JSC, Houston, Texas, pp. 78-86.

4. Reeves, C.A. 1978. Procedure-1: Applicability to Rangeland Classification, Final Report. Report #AD-631737-5335-02, Lockheed Electronics Co., Houston, Texas.

5. Hoffer, R.M., and Staff. 1975. Natural Resource Mapping in Mountainous Terrain by Computer Analysis of ERTS-1 Satellite Data. Research Bulletin 919, Agricultural Experiment Station, Purdue University, W. Lafayette, Indiana. 124 pp.

6. 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 Information Note #112277,
Laboratory for Applications of Remote Sensing, Purdue University, W. Lafayette, Indiana. 235 pp.

7. Fox, L., and K.E. Mayer, 1979. Using Guided Clustering Techniques to Analyze Landsat Data for Mapping Forest Land Cover in Northern California. Proceedings, 1979 Machine Processing of Remotely Sensed Data Symposium, Purdue University, W. Lafayette, Indiana. pp. 364-367.

8. U.S. Forest Service. 1974. Photo Sampling Instructions for the Fourth Minnesota Forest Survey. USDA, Forest Service, North Central Forest Exp. Station, St. Paul, Minnesota. 21 pp.

9. Stewart, J., and P.J. Aucoin. 1978. Earth Observations Division Version of the Laboratory for Applications of Remote Sensing (EOD-LARSYS) Users Guide for IBM 370/148, Vol. I--System Overview. Report #LEC-12563, Lockheed Electronics Co., Inc., Houston, Texas.

10. Landgrebe, D.A. 1976. Final Report, NASA Contract NAS9-14016, June 1, 1975-May 31, 1976. Laboratory for Applications of Remote Sensing, Purdue University, W. Lafayette, Indiana.

APPENDIX

Three Procedure-1 processors are involved in the production of training statistics: DOTDATA, ISOCLS, and LABEL. The same dotfile was used throughout the study. The dotfile contained 196 dots, 117 were conifer, 52 hardwood, 17 grass, and 10 barren pixels. The differences between the P-1 approaches lie with the parameter settings for ISOCLS and LABEL. The parameters used for these two processors for each of the three approaches are detailed below. The clustering results of the fourth approach, Multicluster Blocks, are given for comparison purposes.

A. Approach 1: P-1

ISOCLS: 10 split and 1 combine iterations, not seeded with dots.

Parameters:		
	STDMAX	2.0
	DLMIN	3.2
	PERCENT	90
	CLUS	25
	ISTOP	10
	SEQUEN	SC
	All other parameters default.	

ISOCLS produced 20 spectral classes on the Devil Mountain quadrangle (8372 pixels clustered).

LABEL: Using 10 Nearest Neighbor and L2 distance, LABEL identified 8 spectral classes as conifer, 6 hardwood, 4 grass, and 2 barren.

B. Approach 2: P-1

ISOCLS: 3 split and 1 combine iterations, seeded with 23 dots (10 conifer, 7 hardwood, 4 grass, 2 barren).

Parameters: STDMAX 2.8
DLMIN 3.2
PERCENT 90
CLUS 25
ISTOP 3
SEQUEN SC
All other parameters default.

ISOCLS produced 20 spectal classes.

LABEL: Using 10 Nearest Neighbor and L2 distance, LABEL identified 8 conifer classes, 6 hardwood, 5 grass and 1 barren.

C. Approach 3: P-1; LACIE Parameters

ISOCLS: After assigning the study area pixels to one of the seed dots (based on minimum euclidean distance), one split and one combine iterations are done. However STDMAX is set so high that no splitting is done, and DLMIN is 0; no combining is done.

Parameters: STDMAX 15.0
DLMIN 0
PERCENT 80
CLUS 60
ISTOP 1
SEQUEN SC
SEP 1.0
All other parameters default.

LABEL: Using 1 Nearest Neighbor and L1 distance, LABEL identified 13 conifer classes, 4 hardwood, 3 grass, and 2 barren.

D. Approach 4: McB

Each of 3 training blocks was clustered into 15 spectral classes. These 45 spectral classes were identified by the analyst (using airphotos and a Zoom Transfer Scope) and merged. The resultant 18 spectral classes consisted of 8 conifer, 4 hardwood, 3 grass, and 3 barren classes.