

Quarterly Progress Report

Digital Processing of Landsat MSS
and Topographic Data to Improve
Capabilities for Computerized
Mapping of Forest Cover Types

Contract No. NAS9-15508

Reporting Period: March 16, 1978-June 15, 1978

Submitted to: Exploratory Investigations Branch
NASA Lyndon B. Johnson Space Center

Prepared by: Laboratory for Applications of Remote Sensing
Purdue University
West Lafayette, Indiana 47906

Technical Monitor: Dr. D. L. Amsbury
NASA Mail Code SF5
Exploratory Investigations Branch
Houston, Texas 77058

Principal Investigator: Dr. Roger M. Hoffer
Ecosystems Program Leader
LARS/Purdue University
West Lafayette, Indiana 47906

June 1978

TABLE OF CONTENTS

- I. OVERALL STATUS AND PROGRESS TO DATE
 - Development of the Model
 - 1. Test Site
 - 2. Sampling Procedure for the Model
 - 3. Sample Identification
 - Programming Activities
 - Program EXTRACT
 - Program RANDOM
 - Program SELECT
 - Test Site Field Trip
- II. PROBLEMS ENCOUNTERED
- III. PERSONNEL STATUS
- IV. EXPECTED ACCOMPLISHMENTS

OVERALL STATUS AND PROGRESS TO DATE

During this reporting period (March 16, 1978 - June 15, 1978) three major and related research activities were accomplished. These activities involved: (1) the development of the digital forest topographic model, (2) a computer programming task, and (3) a field trip to the test site. Both the programming task and the field trip were required for the development of the model. A description of these activities is given in the following sections.

Development of the Model

Since the overall objective of this study is to develop techniques which utilize both digital topographic and spectral data to map more accurately forest cover types, that is to increase the level of mapping detail and to increase the accuracy of the obtained information, a procedure to accomplish this objective is being developed in two phases.

The first phase involves the statistical description of the distribution of the different forest cover types in terms of topographic variables (i.e. elevation, slope and aspect), which is in essence the development of the digital forest topographic model. The second phase involves the utilization of the information derived from the statistical distributions (model results) as ancillary information in the pattern recognition approach for classifying multispectral digital data. During this second phase, a series of computer-aided analysis procedures for utilizing multispectral scanner data (Landsat MSS) in conjunction with digital topographic data will be developed and evaluated. In this report, only the first phase, that is the development of the model will be described.

The procedure for developing the digital forest topographic model has been divided into five steps: (1) identification, selection and delineation of the study site, (2) description of the procedures for selecting the specific data to be used in developing the model, (3) identification of each and every data sample by its dominant cover type, (4) to conduct a regression analysis of the data samples to statistically describe the topographic distribution of the different forest cover types in the test site area, and (5) to develop the appropriate discriminant functions and evaluate the resulting model. The first three steps of this procedure will be described in detail in this report, and the last two steps, that is the regression analysis and the development of the discriminant functions will be discussed in the next quarterly report.

1. Test Site. The region designated as the primary study site includes a 34.4 mile by 43.2 mile area in the San Juan Mountains of southwestern Colorado. The area is defined by 25

adjacent U.S.G.S. 7½ minute topographic quadrangles. The location of the study site relative to the states of Arizona, Utah, Colorado and New Mexico is shown in Figure 1.

As indicated in the last quarterly report, various types of reference data are available for this site. However, only the forest type maps produced by the University of Colorado's Institute of Arctic and Alpine Research (INSTAAR) were found to be the most reliable. There are 14 INSTAAR forest type maps for the test site as illustrated in Figure 2.

According to the level of mapping detail, there are two groups of forest type maps. The first group includes the southern two rows of quadrangles (see Figure 2) which contain a considerable amount of detail (series level in Table I). The second group consists of four forest type maps in the northern two rows of quadrangles (see Figure 2) and for which the information level is more general (region level in Table I). All of these 14 type maps were produced from a set of WB57F color infrared photography and appropriate field checking.

Table I. Levels of mapping detail.

<u>Region</u>	<u>Series</u>
Coniferous Forest	Pinyon-Juniper Ponderosa Pine Douglas-fir Spruce/fir
Deciduous Forest	Aspen Oak
Grassland	Mosaic Grassland Dry Grassland Rocky Grassland
Bare Rock	Bare Rock Bare Soil
Water	

To locate the study area in the digital data base (Landsat MSS and registered digital topographic data) a dot matrix grayscale printout of Landsat band 6 at a scale of 1:24,000 was used. The coordinates (lines and columns) of the corners of the test site were then determined, and from these coordinates the location in the data base for each quadrangle was computed. (See Table II).

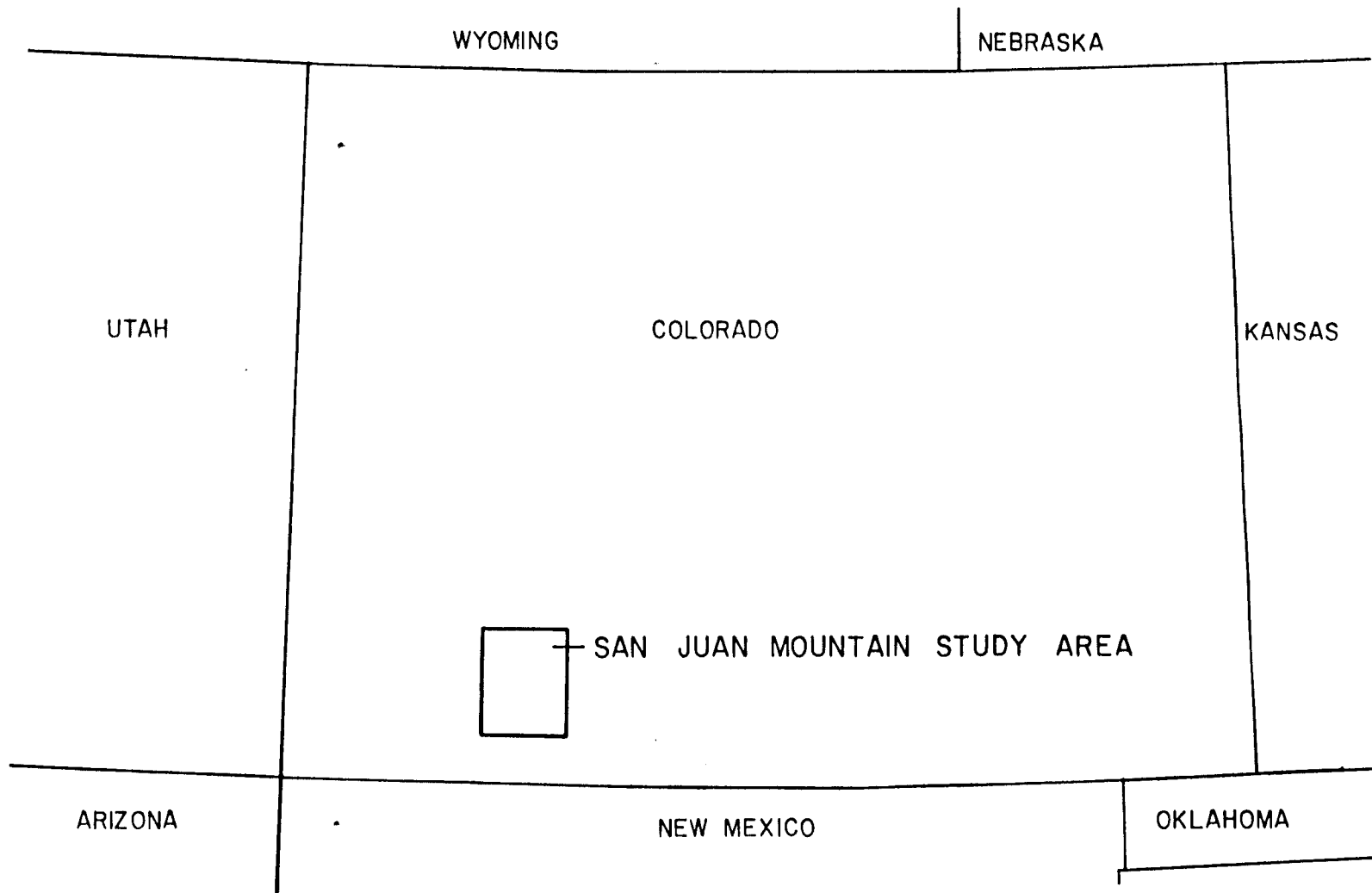


Figure 1. Location of the San Juan Mountains study area.

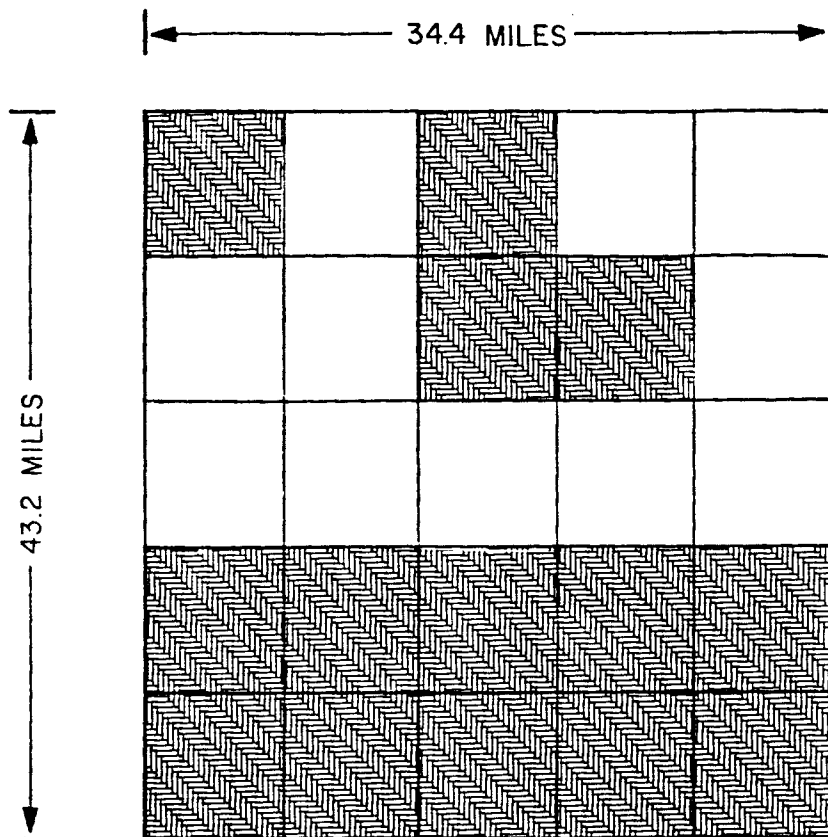


Figure 2. Location of the 14 INSTAAR forest cover type maps available for the San Juan Mountains test site.

Table II. Line and column coordinates for the 25 topographic quadrangles.

<u>Quadrangle</u>	<u>first line</u>	<u>last line</u>	<u>first column</u>	<u>last column</u>
Howardsville	205	386	610	790
Pole Creek Mtn.	205	386	791	972
Finger Mesa	205	386	973	1153
Bristol Head SW	205	386	1154	1334
Bristol Head SE	205	386	1335	1506
Storm King Peak	387	569	610	790
Rio Grande Pyramid	387	569	791	972
Weminuche Pass	387	569	973	1153
Little Squaw Creek	387	569	1154	1334
Workman Creek	387	569	1335	1506
Neetle Mtns. SE	570	751	610	790
N. Granite Peak	570	751	791	972
N. Bear Mtn.	570	751	973	1153
N. Oakbrush Ridge	570	751	1154	1334
N. Pagosa Peak	570	751	1335	1506
Vallecito Reservoir	752	934	610	790
Granite Peak	752	934	791	972
Bear Mtn.	752	934	973	1153
Oakbrush Ridge	752	934	1154	1334
Pagosa Peak	752	934	1335	1506
Ludwig Mtn.	935	1117	610	790
Baldy Mtn.	935	1117	791	972
Devil Mtn.	935	1117	973	1153
Chris Mtn.	935	1117	1154	1334
Pagosa Springs	935	1117	1335	1506

2. Sampling Procedure for the Model. An important phase of the development of the digital forest topographic model is the selection of a statistically valid sample of data points (pixels) corresponding to each cover type that would include all possible combinations of topographic positions, that is of elevation, slope and aspect.

The entire test site was stratified into a total of 91 identifiable topographic positions (strata), consisting of 7 elevation strata (300 meter interval), 4 aspect strata (N,E,S,W), 3 slope strata (1° - 7° , 8° - 17° and 18° - 70°) plus a "flat" stratum, i.e. $7 \times [(4 \times 3) + 1]$.

The first consideration in selecting the sample for the development of the model was to define the sampling unit (unit size). In this study, single pixels were defined as the basic sampling unit because they offer several advantages over larger "fields" or "clusters" of data points. The major drawback of selecting larger fields at random, is that the larger the fields (sampling units), the higher the percentage of fields that contain more than one cover type (cover type boundaries). Past experience has shown that by using a 2 x 2 pixel sampling unit, the majority of the potential fields had to be deleted because they did not represent a single homogeneous cover type. Therefore, the single-pixel cell units provide a much larger sample size (i.e. fewer cells rejected) and consequently allowing a larger inference space to be obtained for developing the model. Another advantage offered by the single-pixel sample selection is that of statistical efficiency. The higher variation in elevation between fields than within one field indicates that selecting the smallest unit possible yields the most statistically efficient results.

Out of the 14 available quadrangles containing cover type information, 7 were selected for training purposes (i.e. to actually develop the model), and the remaining 7 quadrangles were left for use at a later date for evaluation of the model.

In order to geographically spread out the sample, the 7 training quadrangles were selected using a systematic procedure with a random start. Five quadrangles were alternatively selected from the southern rows, with Vallecito Reservoir quadrangle being randomly selected over Ludwig quadrangle for the start. In the northern rows, Howardsville quadrangle was randomly selected over Wemuniche quadrangle. The location of the 7 training and 7 evaluating quadrangles are shown in Figure 3.

To equally represent each topographic position, 50 sample points were selected for each stratum or topographic position, that is, for each of the 91 strata. The 50 points for each stratum were allocated to the seven training quadrangles in proportion to the stratum size, then within each quadrangle the points were randomly selected for each stratum. The 50 sample points were selected randomly with replacement. These points comprise the training sample for the development of the model, after each of the

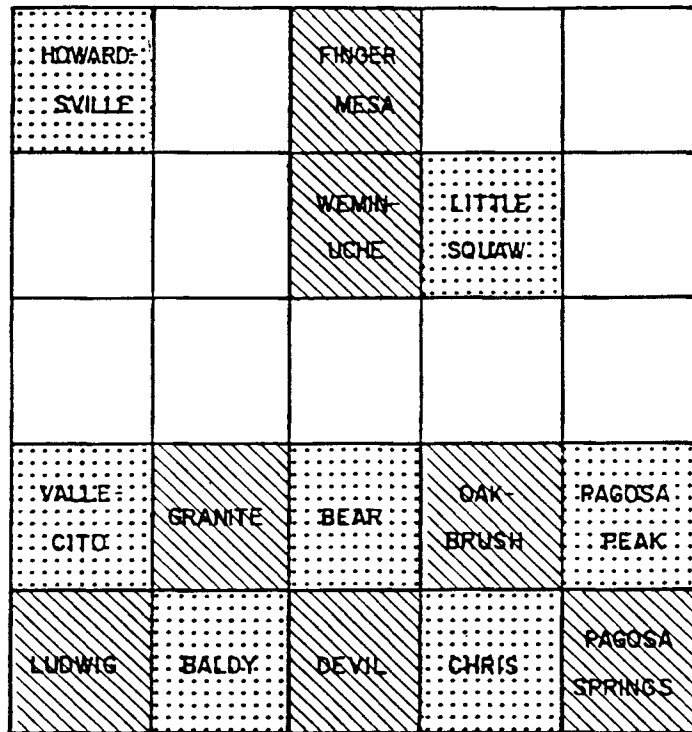


Figure 3. Location of the "training" and "evaluation" quadrangles.

points were identified as to their dominant cover type. The procedure for this identification is described in the next section.

In order to accomplish the sampling task described above, three computer programs were developed. These programs will be described in a separate section of this report.

Before the field trip to the test site, a preliminary forest topographic model was developed following the procedures outlined above. Further refinement of the model is currently being pursued using the information obtained during the field trip. Figure 4 shows one type of graphic representation of the model results, that is, the frequency distribution (histogram) of the forest cover types as a function of elevation. Another graphic representation of the model results that is in the process of being developed at the present time, is the circular or polar diagram, in which it will be possible to show the frequency distribution of forest cover types as a function of two topographic variables (slope and aspect).

The model output which will be used as input to the multi-dimensional classification (classification of the spectral and topographic data) will be in the form of a series of discriminant functions. The description of these functions and the procedures to integrate them into the multidimensional classification will be described in the next quarterly progress report.

3. Sample Identification. Each one of the $50 \times 91 = 4550$ selected points were identified using the available cover type maps as an "initial" identification. Corrections to this initial identification are currently being carried out using the information obtained during the visit to the test site. The details of the field trip are given in a separate section of this report.

Programming Activities

Three different programs were written in support of the development of the digital forest topographic model. These three programs were named EXTRACT, RANDOM and SELECT.

Program EXTRACT. This program reads the standard LARS Results Tapes where the classifications are stored. In this particular case, it reads the tapes containing the classification of the 91 topographic strata. The input parameters to this program are: 1) the class or group of classes of interest, b) the area location (block description card), 3) the "minpoints" parameter (see LARSYS Ver. 3.1 Users Manual for description of "minpoints"), and d) the tape and file number in which the result classification is stored. The output of this program consists of a listing of the coordinates of all the points belonging to each one of the requested strata or classes. This output can be in a disk file, punched card file, and/or printer file format.

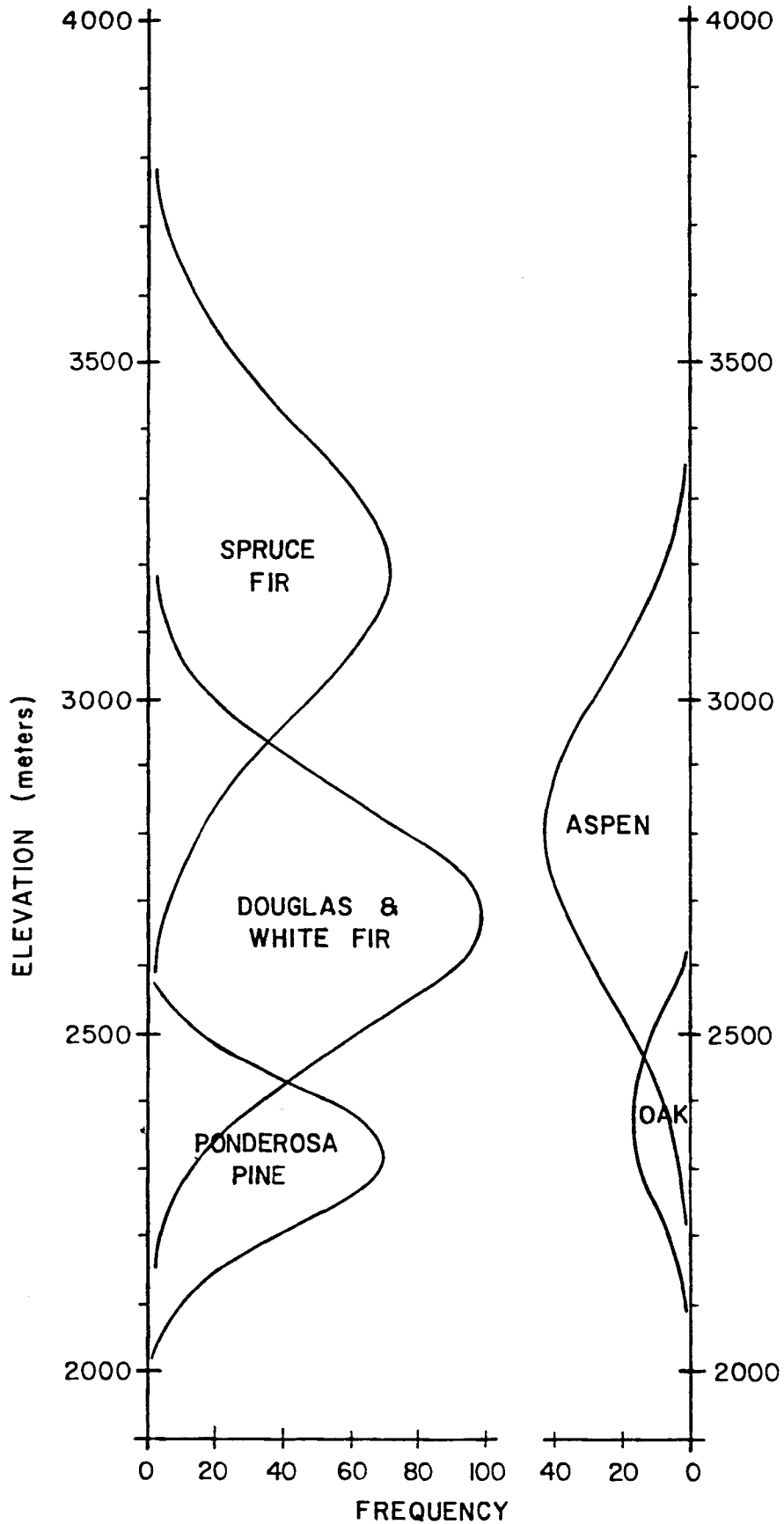


Figure 4. Graphic representation of the forest topographic model output, i.e. the frequency distribution of the forest cover types as a function of elevation.

Program RANDOM. This program is essentially a random number generator. Once the coordinates for all the points in a stratum or class have been obtained using the program EXTRACT, the program RANDOM generates a specified number of random numbers between one and the number of points in the class or stratum. The input parameters for this program are: the number of points desired, and the number of points in each stratum. The output of this program is subsequently used by the program SELECT.

Program SELECT. This program basically selects the desired number of points (which in this particular case is 50) from all the points belonging to each one of the topographic classes. The input parameters for this program are: the list of random numbers, and the list of points in the stratum. The output consists of the list of coordinates for the 50 selected random points. A standard LARS field description format (LARS-12 format) is used to represent these coordinates. This format is described in detail in the LARSYS Ver. 3.1 Users Manual.

In addition to writing the above described programs, two of the LARSYS processors were also modified. The *TRANSFERDATA processor output format was modified to accommodate the requirements of this project. Similarly, the *PRINTRESULTS processor was also modified to allow printing (displaying) each one of the randomly selected points on a quadrangle by quadrangle basis, using a "@" symbol to represent single points and a "\$" to designate points that were selected randomly more than once (caused by the selection with replacement). These printouts were then overlaid onto the forest cover type maps for identification of each one of the selected training points.

At the present time two programmers are working on the modification of the existing Layered Classifier, which will be required for the classification of the combined digital multispectral and topographic data set. Also, the programmers will write two plotting subroutines to obtain graphic representations of the digital forest topographic model.

' Test Site Field Trip

During the initial field trip (from June 5 until June 12, 1978) five objectives were accomplished. They included: (1) checking pixels from the training sample that were "most likely" to be incorrectly identified, (2) checking all pixels with questionable species identification, (3) checking the Howardsville and Little Squaw Creek quadrangles for the types of coniferous forest species, (4) familiarizing the photo-interpreter with the study site and its cover types, and (5) meeting with Forest Service personnel interested in the project.

First, the portion of the training sample to be investigated in the field had to be determined. The inaccessibility of the study site limited the extent of the field checking, thus only the pixels that were most likely to be incorrect were selected

for field checking. These pixels were identified as outliers in the distribution plots for elevation and for elevation versus aspect. The questionable outliers were rechecked using the type maps, and if needed using also the aerial photography. Then if the identification was still questionable, the location was marked in a forest map for field checking.

Each area designated for field checking was attempted to be reached on the ground. The limited field time often prevented the exact area from being visited, but in all cases, comparable topographic positions near the area in question were checked. In each case a definite determination of the dominant forest species was made. This procedure indicated that there were a few errors in the type maps, but they seemed to be easily identifiable.

The northern two training quadrangles, Howardsville and Little Squaw Creek, were checked to determine the identification of the coniferous forest species (not indicated on the type maps). In each quadrangle, the only vehicular access was by a single major east-west drainage which included the lowest elevations on a flat and on both north and south aspects with steep slopes. In both cases the relatively high minimum elevation (approximately 9,500 feet) limited the coniferous species to only Engelmann spruce. Aspen and willow were the major deciduous species. The data from these two quadrangles can now be included in the model.

Another of the activities pursued during the field trip was to familiarize the interpreter who will be photochecking the training and test sample with the area and its cover types.

The LARS field team participated in two meetings with the Forest Service personnel. The first meeting took place in Durango with personnel from the San Juan National Forest. Those present included Pete Hager (supervisor), Terry Hughes (soils scientist), Hank Bond (Range and Planning), Bob Mattson (timber inventory) and from the LARS/Purdue team Roger Hoffer, Mike Fleming and Ross Nelson. Discussions involved updating and reviewing this investigation, the current state-of-the-art in computer processing of digital data, possible Forest Service sources of information to evaluate the computer classifications, and a breakdown of cover types of interest to the Forest Service and their definitions (see Appendix I.) One possible source of evaluation data involves the Forest Service data which is a sample of photo points, of which 300 were ground checked. However, geographic identification (location) of the sample points are not readily available. Stage 2 (stand sampling) information is continually being obtained, but again the stands are indicated only on aerial photography and not tied to a base map. Thus, this source of information is not easily obtainable for use in this project. Existing Forest Service type maps are available, but are at a different scale (1:31,680), out-of-date (mapping in 1970 from 1964 photography) and mapped with very different criteria (both as far as the breakdown of classes and the level of detail). The conclusion was reached that none of the vegetative information available from the Forest Service could be effectively used to evaluate the digital classification results.

The Forest Service personnel indicated that they were currently in the process of manually digitizing the topographic data (elevation) from the 7½ minute quadrangles on a 2.5 acre cell basis.

The second meeting attended was a Forest Service discussion to evaluate procedures for obtaining the vegetation, soils, and landform data which are to be used with the R2 mapping programs. Appendix II contains a list of the people present in the second meeting. The area selected to test and evaluate the procedures for a digital data base mapping system was the Summit Peak Quadrangle.

The soils data base was manually digitized from the existing soils map (Order 3) of the area. The soil data for the entire San Juan Forest will require 4½ more years to finish with the current two man effort. The landform classes for the Summit Peak quadrangle have been mapped, hand digitized and input into the R2 mapping system. Mapping landforms over the entire San Juan Forest has just begun and will require 3 years to complete. Two methods of obtaining the vegetation information are being evaluated. The existing forest type map was manually digitized to provide one source. The second was the digital computer classification generated during the Landsat II NASA project conducted by LARS and INSTAAR. The three layers of information (soils, vegetation, and landform) are combined (overlaid) to delineate similar "Ecological Land Units" or ELU's. Two procedures are being used to combine the data: 1) manually overlaying the layers (line maps), and 2) digital processing using the R2 mapping system. Several problems were discussed, such as the time consuming procedure of manually digitizing the large quantities of map data, which was necessary for all three layers of information. Rescaling of the Landsat classification results is required to convert the data to the R2 format to avoid the manual digitization required for the vegetation and topographic data. The second difficulty noted was the large number of possible classes when the various layers are combined. The large number of classes for each of the three layers (vegetation, soils, and landform), when combined to form the ELU's results in numerous small homogeneous areas. To adequately interpret and manipulate the units, they must be combined to a much more general level of detail. On the Summit Peak quadrangle, the 21 soils classes were combined into 5, and the 16 vegetation classes into 6.

In summary, these meetings proved to be very informative concerning current thinking by Forest Service personnel about potential applications of vegetation cover type maps obtained through computer processing of Landsat data.

PROBLEMS ENCOUNTERED

The major problem encountered during this reporting period was the late start of the spring (snow melting season) in the test site. This caused a delay of the field trip from the originally planned date in early May to the second week in June when the snow had melted. Consequently, the development of the

final digital forest topographic model has been accordingly delayed.

Another problem was related to the sudden departure of the original student programmer after the University semester was over. This has delayed the Layered Classifier modification task. However, it is anticipated that this modification will be completed before the July field trip.

PERSONNEL STATUS

During the past three months, the people directly involved in this project were:

<u>Name</u>	<u>Position</u>	<u>Effort</u>
Anderson, V.	Statistician	10%
Bartolucci, L. A.	Project Manager	25%
Cain, J.	Programmer	15%
Fleming, M.	Senior Analyst	83%
Hamilton, C.	Programmer	50%
Hoffer, R. M.	Principal Investigator	15%
Nelson, R.	Grad. Student	30%
Peterson, J.	Assoc. Director of LARS	5%
Pillai, K.	Statistician	13%
Pratt, B.	Adm. Assistant	15%
Clerical	Secretaries	40%

EXPECTED ACCOMPLISHMENTS

For the next reporting period (June 16, 1978 - September 15, 1978) it is expected that the final version of the digital forest topographic model will be completed. Also the resulting discriminant functions should have been integrated in the multi-dimensional classifications of the combined digital multispectral and topographic data set. The resulting classifications will be evaluated during the next field trip to the test site, which is currently being planned for the middle portion of July.

APPENDIX I

1. Layers of information desired for Data Base
 - a. Elevation
 - b. Slope
 - c. Aspect
 - d. Landform
 - e. Vegetation
 - g. Other

2. Vegetation and Ground Cover classes of interest
 1. Spruce/fir
 2. Ponderosa pine
 3. Pinion/juniper
 4. Aspen/conifers
 5. Conifers/aspen
 6. Aspen
 7. Conifers/oakbrush
 8. Oakbrush/conifers
 9. Oakbrush
 10. Brush, mixed
 11. Wet and mosaic grassland
 12. Grassland
 13. Rocky grassland - dry tundra rocky
 14. Alpine
 15. Willow
 16. Bare rock
 17. Water

3. Crown Closure and/or Ground Cover Categories
 1. 0-30%
 2. 31-50%
 3. 51-70%
 4. 71-100%

4. Definitions of vegetation classes

Coniferous Forest is a stand with less than 20% other tree species.

Aspen Forest is a stand with less than 20% other tree species.

Coniferous/Aspen Forest is a stand with 20 to 50% Aspen

Aspen/Coniferous Forest is a stand with 20 to 50% Coniferous species.

Coniferous/Oakbrush Forest is a stand with 20 to 50% oakbrush.

Oakbrush/Coniferous Forest is a stand with 20 to 50% Coniferous species.

Brush mixed is a stand of brush with more than 30% brush species other than oakbrush.

Oakbrush is a stand of brush with less than 30% other brush species.

APPENDIX II

Skip Kowalski	Wildlife Bio.	White River N.F.
Bill Smith	Computer Spec.	Regional Office
Ron Bauer	Soil Scientist	Regional Office
Mike Fleming	Remote Sensing	LARS (Purdue)
Ross Nelson	Remote Sensing	LARS (Purdue)
Terry Hughes	Soil Scientist	San Juan N.F.
Dave Cook	Wildlife Biologist	San Juan
Hank Bond	Range & LMP	San Juan
Bob Buttery	Regional Ecologist	Regional Office
Andre Coisman	Geometronics Group Leader	Regional Office
Irv Duggan	Remote Sensing-Forestry	NFAP Houston
Jack Lowe	USFS Engineering	Regional Office