Reprinted from

Symposium on Machine Processing of Remotely Sensed Data

June 27 - 29, 1979

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana 47907 USA

IEEE Catalog No.
79CH1430-8 MPRSD

MACHINE PROCESSING OF LANDSAT MSS DATA AND DMA TOPOGRAPHIC DATA FOR FOREST COVER TYPE MAPPING

MICHAEL D. FLEMING, AND ROGER M. HOFFER
Purdue University

I. INTRODUCTION

In forestry, as in many other disciplines involving land management, there exists a definite need for timely, reliable information on which to base resource management decisions. This was emphasized in 1974 through the Forest and Rangeland Renewable Resources Planning Act in which the U.S. Congress mandated the U.S. Forest Service to provide information on the condition and productivity of approximately 1.6 billion acres of land, every 10 years. The synoptic view that can be obtained through data from spacecraft altitudes is proving to be of considerable value in developing resource bases, particularly where information over extensive geographic areas is needed. The launch of Landsat-1 in 1972 initiated a new era for land managers by proving that high-quality data can be obtained from satellites at reasonably frequent intervals for nearly any portion of the earth's surface. However, the ability to collect data from satellites far surpasses existing capabilities to analyze and interpret the data in a timely, reliable manner. If computer-aided analysis techniques are to be effectively utilized in conjunction with MSS satellite data on a routine, operational basis, it is important to define and develop the most effective analysis techniques and to determine the level of detail and the reliability of information that can be obtained with such techniques.

A. BACKGROUND

In recent years, a series of investigations have been conducted at the Laboratory for Applications of Remote Sensing (LARS), Purdue University, which have indicated many of the capabilities and limitations of Landsat data and various analysis techniques for classifying and mapping forest cover in regions of significant topographic relief. The results of a Landsat-1 investigation (1) indicated that deciduous and coniferous cover, as well as other major cover types, could be classified and mapped with a reasonable degree of accuracy (75-85%). However, the classification and mapping accuracies for individual forest cover types were much lower. A Skylab experiment (2) evaluated the potential for increasing the spectral separability of the individual forest cover types, through the use of narrower wavelength bands located over a broader portion of the spectrum, but the relatively poor quality of the digital MSS data negated any increase in accuracy obtained by the additional wavelength bands. However, as part of the Skylab investigation, software was developed to geographically register the MSS data, both Skylab and Landsat, with digital topographic data obtained from DMA digital terrain tapes. A Landsat-2 study (3), was a cooperative study between the U.S. Forest Service, INSTAAR of the University of Colorado, and LARS, Purdue University, to evaluate the analysis techniques developed in the previous investigations for purposes of mapping a U.S. Forest Service Planning Unit. Landsat data for the entire Southern San Juan Mountain Planning Unit, an area of nearly 1/2 million hectares, was classified into major cover types, the results of which were then evaluated by the U.S. Forest Service. Part of this same study involved the development of a digital data base in which the cover type map (derived through the computer classification of Landsat data) was overlayed with DMA topographic data, and each data cell corresponded to a 0.45 hectare resolution element in the Landsat data. From this digital data base, selected combinations of cover type, elevation, slope and aspect could be displayed as maps and acreage tables. Such a post-classification utilization of the combined spectral/topographic data proved very effective, providing the U.S. Forest Service with a flexible method of obtaining "customized" resource management maps for areas of particular interest (3). However, we also found that combining the topographic data with the spectral data in the actual classification process (i.e., cover type identification) produced mixed results. This led to the current project in which it was hypothesized that since many forest cover types in this area of mountainous terrain seem to be topographically distributed, a more effective method could be defined to take advantage of such topographic distribution in order to more accurately identify and map individual forest cover types.

CH1430-8/79/0000-0377$00.75 © 1979 IEEE

B. OBJECTIVE

The objective of this research was to develop, test, and document a digital processing technique for using Landsat MSS data in combination with topographic data (elevation, slope, and aspect) to accurately and reliably map individual forest cover types in regions of mountainous terrain.

C. APPROACH

The initial phase of this study involved the development and evaluation of different, alternative techniques for using a digital data base of Landsat MSS and topographic data to increase the accuracy of mapping forest cover types. Two very different general approaches for using topographic data in conjunction with Landsat data have been developed and evaluated. These are referred to as the Topographic Distribution Model approach and the Reflectance Geometry Correction Model approach.

The Topographic Distribution Model approach involved the development of a quantitative description of the distribution of each of the forest cover types in the study site as a function of elevation, aspect, and slope. The model thus provided a quantitative indication of the probability of occurrence of any species for any topographic location. Statistical characterization of the topographic distribution of the various cover types could then be combined with the statistical data describing the spectral characteristics of the various cover types. This provided the basis for the training data necessary for the classification of the combined spectral/topographic data set. Previous studies had shown that many different approaches can be followed in developing the training statistics, and many different classification algorithms can be used for the actual classification process. In this study, two very different techniques were used to develop the training statistics and two distinctly different procedures were used in the classification step. In addition, several variations on these basic classification procedures were also tested.

The Reflectance Geometry Correction Model approach involved the alteration of the reflectance values contained in the Landsat scanner data in order to remove spectral variations resulting solely from topographic effects. Knowledge of the geometric relationships between the sun, the ground, and the satellite were used to "correct" the spectral data by calculating correction coefficients to remove the effects of the topographic position on the spectral values. The "corrected" spectral data then represents the responses of a hypothetical, horizontal surface, and all remaining spectral differences are, in theory, a function of the earth surface materials present. The details involved in developing this model are beyond the scope of the current paper, but some of the major results are discussed and compared to the results obtained from the Topographic Distribution Model.

II. TEST SITE AND DATA DESCRIPTION

A. TEST SITE DESCRIPTION

The study described in this paper involved the San Juan Mountain test site, a twenty-five 7½-minute U.S.G.S. quadrangles area (approximately 34 x 43 miles) in the center of the rugged San Juan Mountains of southwestern Colorado. The area straddles the continental divide and includes portions of two National Forests, the San Juan National Forest and the Rio Grande National Forest.

The study site is characterized by a diverse and complex mixture of land forms and vegetation types. Elevation within the area ranges from approximately 2200 meters (7,200 feet) to 4000 meters (13,000 feet). The climate in this area is typical of the Colorado Rockies, with very low relative humidity, abundant sunshine, cool summers with frequent afternoon showers and heavy winter snows. Wide daily temperature fluctuations are normal. The annual precipitation varies with elevation and ranges from 30.5 to 127.0 centimeters (12-50 inches) per year. More than half of this falls as snow during the winter months, remaining on the ground well into June in fairly extensive areas at the upper elevations and year around in some small areas.

The test site area consists primarily of Tertiary volcanics with the topographic expression of a maturely dissected plateau, further modified by extensive valley glaciation. This area is characterized by numerous glacial lakes, meadows, and commercial stands of spruce and fir. Narrow strips of aspen or Gambel oak extending down the side of a mountain often mark the paths of former landslides, avalanches, or forest fires. Extensive areas of mine tailings are evidence of the former importance of the area as a mineral-producing region, particularly for silver. At the higher elevations, steep slopes, rugged peaks, and rock outcrops are frequent. This rugged topography and the related local climatic regimes within the San Juan Mountains result in a diversity of vegetation and wildlife communities within a relatively small geographic area.

In areas where man, animals, fire, avalanches, landslides, or other influences have not caused major changes in the vegetative cover, the naturally occurring vegetation is determined by a complex interaction of edaphic, topographic and climatic influences. Climatic conditions in particular are influenced by differences in elevation. As elevation increases, the mean annual air temperature decreases and, in general, precipitation levels increase. Similarly, aspect and the steepness of a slope influence the micro-climatic conditions of a particular area and therefore also have a distinct impact on the vegetation occurring there. The result of this interaction among the edaphic, topographic and climatic influences is a distinct distribution of vegetative cover types within various elevation ranges. Figure 1 graphically displays the general distributions. Further-

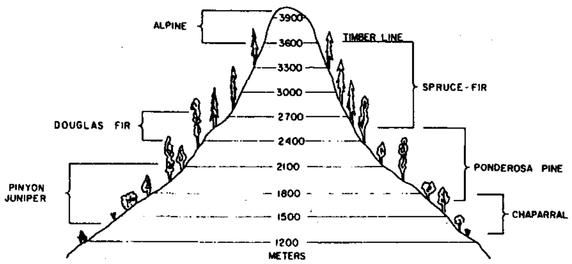

Figure 1. Relationships between elevation and distribution of vegetative cover types.

more, in a single elevation range, the frequency with which a species may appear is affected by the aspect and slope characteristics of the area.

The following paragraphs provide a general description of the altitudinally defined vegetation zones within the San Juan site. The elevation ranges indicated for the individual species reflect generalized estimates of elevation as given in existing literature and are not specific to the San Juan Mountains.

Alpine Tundra. The Alpine area occurs above the timberline, at about 3400 meters (11,000 feet) and above. Because of the short frost-free growing season and the possibility of frost at any time of the year, the vegetation is limited to short grasses and sedges, hardy forbes, alpine willows, and other low shrubby plants.

Spruce/Fir. The spruce/fir zone extends from approximately 2700 meters (9,000 feet) to the timberline, with the dominant tree species being Engelmann spruce (*Picea engelmannii*), subalpine fir (*Abies lasiocarpa*), and aspen (*Populus tremuloides*). Aspen is often an indicator of a disturbed site; areas burned within the previous 50 years frequently have dense aspen stands, often with a coniferous understory which will eventually overtop and shade out the aspen.

Douglas-fir/White fir. Below the spruce/fir zone is an elevation belt dominated by Douglas-fir (*Pseudotsuga menziesii*) and white fir (*Abies concolor*). Dense stands containing both species are found on north-facing slopes at the lower ranges and in all aspects at higher elevations. Aspen continues as the dominant hardwood, forming pure stands and mixtures with the Douglas-fir/white fir throughout the zone on all aspects. White fir, sometimes a disturbance indicator, and Douglas-fir are commercially harvested.

Ponderosa pine. The ponderosa pine (*Pinus ponderosa*) zone extends from an elevation of about 1800 meters (6,000 feet) to 2900 meters (9,500 feet), mixing with Douglas-fir at the upper extent of this range and with the pinion/juniper cover at the lower extent. Stands of ponderosa pine seldom have more than 70% crown closure and are characteristically rather open with grass or mixtures of brush forming the understory of vegetation. Aspen generally occurs on northern slopes in small patches, interspersed among the ponderosa pine or in pure stands. Gambel oak (*Quercus gambelii*) appears in mixture with the ponderosa pine and in large, sparse, shrubby stands at the lower elevations.

Pinion/Juniper. The elevation belt immediately below the ponderosa pine contains pinion pine (*Pinus edulis*) and juniper, especially the Utah juniper (*Juniperus osteusperma*), Rocky Mountain juniper (*Juniperus scopulorum*) and one-seed juniper (*Juniperus monosperma*). These semi-arid areas are much lower, dryer and warmer and have more sparse understory vegetation. Pure Gambel oak stands and mixed shrub stands are found on all aspects within this elevation zone.

B. LANDSAT, TOPOGRAPHIC, AND SUPPORT DATA USED

The spectral data used in this investigation was Landsat MSS data (Frame No. 1407-17193 obtained on Sept. 3, 1973) which was geometrically corrected and re-scaled to a 1:24,000 line-printer scale through LARS' preprocessing routines (4).

Digital elevation data were obtained from the Topographic Center of the U.S. Defense Mapping Agency (DMA), Washington, D.C. To produce these data, DMA used a table digitizer to manually digitize the contour lines of a 1:250,000 scale U.S.G.S. map having contour intervals of 61 meters (200 feet). Since it was necessary to produce a uniform grid of elevation data, the values for cells through which no contour line passes were interpolated. The resulting digital elevation data tape has a cell size of 64 meters square. These DMA elevation data were registered with the Landsat data at LARS, using a nearest neighbor fit and then added to the Landsat data tape as channel 5.

Since the data analysis process required slope and aspect information on a pixel-by-pixel basis, the elevation data were numerically differentiated to produce an estimate of the gradient vector at each pixel location. The magnitude of the vector defines the slope angle, and the direction defines the aspect angle. A total of four channels of slope and aspect data were added to the data tape. Slope was recorded as degrees in channel 6 on the tape. Channels 7 and 8 represented one way of encoding the aspect information. Channel 7 expressed aspect such that north = 0, east and west = 90, and south = 180 degrees. Channel 8 then represented the east/west flag, east = 0 and west = 1. Channel 9 was the azimuth direction, north = 0 and 360, east = 90, south = 180, and west = 270.

The support data used in this project consisted of $7\frac{1}{2}$-minute U.S.G.S. topographic maps, color infrared aerial photography, and forest cover type maps. Topographic maps were used to assess the characteristics and quality of the DMA elevation data. The aerial photography used was color infrared photography at a scale of 1:120,000 obtained by NASA's WB-57 on August 4, 1973. This photography was of excellent quality and provided the data needed to verify the accuracy of the forest type maps and to identify cover types at individual selected pixels. These identifications were needed both for training the computer and for evaluating the classification results.

The forest cover type maps, available for 14 quadrangles, had been produced by INSTAAR, University of Colorado, using WB-57F color infrared photography and field checking. Four of the 14 quadrangle maps are located in the Rio Grande National Forest; ten are located in the San Juan National Forest; they contain information at the "series" and "sub-series" level as defined in Table 1. While the series level of detail defines the informational classes desired, many of the series-level cover types actually occur in mixtures. Therefore, the cover type maps and the test pixel identifications use cover type classes that are more detailed than the series level. These detailed cover type classes are referred to as Level IV or sub-series classes.

Table 1. Breakdown of cover types at various levels of detail.

Region or Level II	Series or Level III	Sub-Series or Level IV
Coniferous Forest	Spruce-Fir (SF)	SF
		SF/Aspen
	SF/DWF	SF/DWF
		SF/DWF/Aspen
	Douglas and White Fir (DWF)	DWF
		DWF/Aspen
	DWF/PP	DWF/PP
		DWF/PP/Aspen
	Ponderosa Pine (PP)	PP
		PP/Oak
	PP/PJ	PP/PJ
		PP/PJ/Oak
	Pinyon-Juniper (PJ)	PJ
		PJ/Oak
Deciduous Forest	Aspen	Aspen
	Oak	Oak
Herbaceous	Alpine	Alpine Willow
		Xeric Tundra
		Mesic Tundra
		Hydric Tundra
	Grassland	Xeric Grassland
		Mesic Grassland
		Hydric Grassland
Non-Vegetated	Barren	Exposed Rock and Soil
	Urban	Urban
	Water	Water

Of the fourteen quadrangles for which cover type maps existed, seven were designated as "training" quadrangles and were used to develop the topographic and spectral statistics used in the classifications. The remaining seven quadrangles were designated "evaluation" quadrangles and were used to evaluate the accuracy of the classifications. The procedure for sub-dividing the quadrangles between training and test areas was an alternating selection with a random start. Figure 2 indicates the general location of the test site in SW Colorado, and shows the specific quadrangles used for developing the training and test data sets used throughout the study.

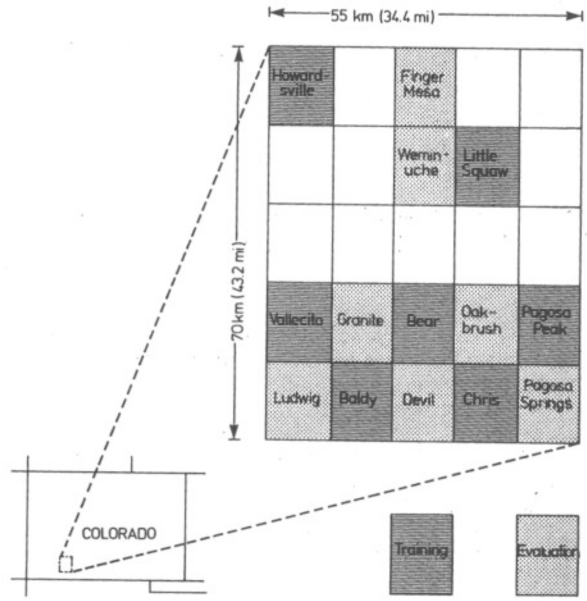

Figure 2. Quadrangles selected for "training" and evaluation.

III. TOPOGRAPHIC DISTRIBUTION MODEL

As previously stated, the overall objective of this study was the development and testing of techniques which utilize both digital topographic data and spectral data in order to map forest cover types at a greater level of mapping detail and with increased accuracy than is possible with spectral data alone. To meet this objective, a key requirement was the development of a procedure to quantify the topographic characteristics of the vegetation types and then utilize the information in the classification process, a two-phase procedure.

The first phase was the development of the statistical description of the distribution of each forest cover type in terms of topographic variables (i.e., elevation, slope and aspect). This was in essence, the development of the digital forest topographic model. The second phase was the utilization of the information derived from the topographic distributions (model results) and analysis of the spectral data in a pattern recognition procedure for classifying digital data.

A. BACKGROUND

Ecological studies in the western coniferous forests have demonstrated the existence of vegetational gradients along changes in altitude, soil moisture, parent material, climate, and other ecological factors. These factors create a gradual sequence of changes in forest composition and structure, as well as some relatively abrupt transitions from one community to another. Topography alone influences plant distributions indirectly through its control of many environmental parameters including insolation, temperature, atmospheric pressure, precipitation, relative humidity, wind velocity, evaporation, and soil characteristics. Daubenmire (5) recognized these relationships and pointed out that topographic position accounts for most of the climatological and vegetative deviations from the ideal altitudinal gradient. He concluded that rigidly defined altitudinal belts do not exist throughout the Rocky Mountains, but rather, one finds a regularly repeated series of distinct vegetation types, each of which bears a constant altitudinal or topographic relationship to contiguous types. Previous descriptions, however, have been qualitative evaluations of the general topographic distribution of the various cover types. For purposes of this study a quantitative characterization was needed, such that for any given topographic position (i.e., elevation, slope and aspect), the probability of occurrence for any particular species could be given. No such quantitative information was known to exist. Therefore, the first phase of this study was directed at the development of the Topographic Distribution Model in order to obtain the required information in the format needed.

The Topographic Distribution Model is a mechanism for combining point-by-point information about forest species, elevation, slope, and aspect to describe quantitatively the topographic positions of the major forest cover types. The input used to develop the model for this study was information obtained from forest cover type maps, aerial photography, and topographic/spectral data tapes. The output from the model is a quantitative characterization of the topographic distribution of each forest type by graphical displays (i.e., histograms, polar plots and regression lines) and in terms of estimated multivariate normal distributions (means and covariance matrix). It should be noted that these characterizations are specific descriptions of the vegetation of the San Juan Mountain area and cannot be said to describe the topographic distribution of the forest cover in other mountainous areas of the North American continent or even of the entire Rocky Mountain area. However, the basic techniques used to develop these plots can be applied to other mountainous areas for which cover type maps and elevation information is available.

B. DEVELOPMENT OF TOPOGRAPHIC DISTRIBUTION MODEL

The objective of the topographic distribution model is to determine the probability of each vegetation type occurring at any given elevation, slope and aspect. Thus, the data sample must statistically describe the various cover types as a function of their topographic distribution rather than their spectral response. Development of the topographic training statistics is optimized when developed through a random sampling procedure stratified by topographic positions. This insures that the sample represents the distribution of the cover types, and is not a function of the frequency of the topographic positions. This type of sampling procedure applies for utilization of any type of ancillary information (i.e., soils, parent material, geology), not just topographic information.

Selection of a statistically valid sample of data points was a prime concern in constructing the topographic distribution model. The first consideration was to define the size of the sampling unit. In this study, units corresponding to single Landsat pixels were selected because of the statistical efficiency and simplicity in handling. An additional advantage of using single pixel cells over groups of pixels or "fields" was the minimization of the many edge effect problems inherent in using larger cells.

The most important consideration in developing the topographic distribution model is to equally represent the entire range of topography (elevation, slope and aspect). To accomplish this the study area was divided into 91 topographic positions or sampling strata. Elevation was stratified into seven 300 meter zones; each elevation zone was then divided into 13 slope/aspect strata, which consisted of a flat class and three slope classes (1-7, 8-17, 18-70), with each slope class being divided into four aspect classes (N, E, S, and W) $(7\text{x}[1+\{3\text{x}4\}])$.

It was estimated that 50 points, randomly selected within each strata, would provide an adequate representation of each topographic position. The points were allocated among the seven training quadrangles as a function of the proportion of the topographic position present in the quadrangle. The result was a total of 4,550 stratified random sample points (50x91).

Each of the 4,550 selected points was initially identified using the available cover type maps. This identification was then verified through photointerpretation of color infrared photography. In addition, 20-30% of the selected training points were checked on the ground. Sample points that fell on cover type boundaries and therefore could not be defined as belonging to any single cover type class were excluded from the sample, as were points that could not be reliably identified. Non-forest points that fell on water or bare rock were also excluded from the model data, since they were not needed for characterizing the topographic distribution of vegetative cover types. The result was a total of 3,379 sample points that were actually utilized in development of the topographic distribution model of cover types. The comparison of type maps, aerial photos, and field checking insured accurate identification of the data used to develop the model.

After the cover type for each of the points had been identified, it was necessary to go back to the topographic data to obtain the actual elevation, aspect, and slope for each sample. These data describing the actual topographic position were then utilized in the development of the final topographic distribution model for each cover type.

C. EVALUATION OF THE TOPOGRAPHIC DISTRIBUTION MODEL

One of the major results of this research was the development of quantitative descriptions of the distribution of each of the primary coniferous, deciduous, and herbaceous cover types present in the San Juan Mountain study area. The distribution of the various cover types relative to topography can be presented graphically in a number of ways. For each of ten cover types or cover type mixtures involved in this study, five types of graphical displays were generated, and the figures evaluated. These five formats included: a histogram that shows the distribution of the cover types as a function of elevation; polar diagrams that show the distribution as a function of either elevation and aspect, or slope and aspect; a regression line that presents the distribution of each type as a function of elevation and the 0-180 aspect channel; and a plot of the Gaussian distribution of cover types as a function of elevation. Figures 3-7 are examples of these plots.

In Figure 3, the elevation is divided into 50 meter bins. Examination of the histograms indicates the degree of normality of each distribution. The polar plot such as Figure 4, displays the distributions as a function of elevation and aspect, and indicates that the "typical" elevation range varies for each species as a function of aspect as well as elevation. The slope and aspect polar diagrams, of which Figure 5 is an example, were used to verify that the model contains a good representation of the sample points for the full range of aspects and slopes, and that each cover type is found on all slopes and aspects at some elevation. The regression line graphs (e.g., Figure 6), display the key information for the two most significant variables obtained in regression analysis--elevation and aspect (0-180 channel). Figure 7 shows the normalized distribution of the various cover types as a function of elevation, and in this figure the distribution is shown for all three major coniferous cover types.

In evaluating the histogram graphs for each vegetation type, it was noted that the elevation distribution for each species is generally normal, except for ponderosa pine and Gambel oak. The apparent skew in the ponderosa pine and oak data

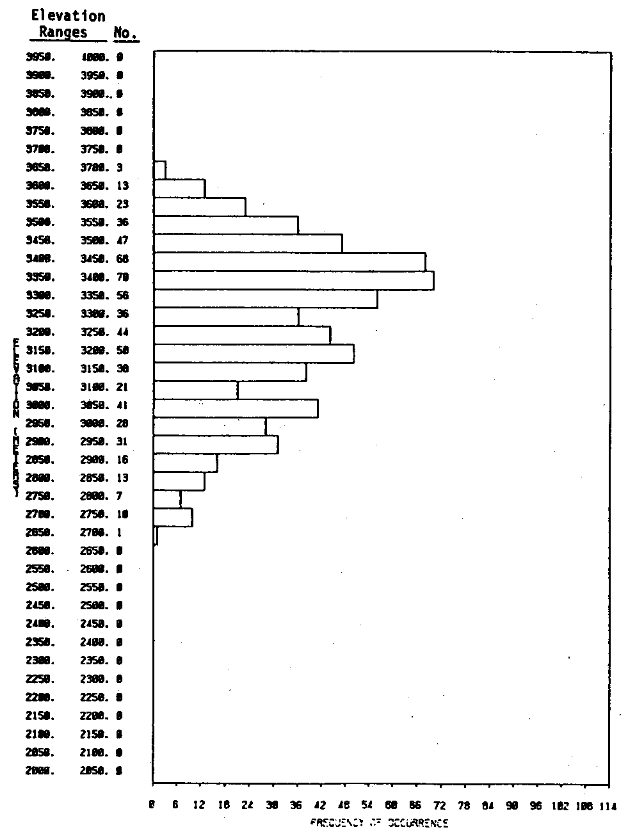

Figure 3. Histogram of Engelmann spruce/subalpine fir as a function of elevation.

was caused by the absence in sample points below 2,225 meters due to the range of elevation existing in the study site. However, this did not seem to cause any problems in the classification of the data.

The polar diagrams of elevation and aspect showed that for all cover types, the average elevation is higher on the southern aspects than on northern aspects, and that there seems to be very little difference in average elevation betwee east and west aspects. Average elevation for each species varies as a function of aspect by an average of 70 meters (225 feet), with Douglas-fir and white fir having the greatest aspect-dependent variations among the coniferous species. The polar diagrams of slope and aspect indicated that very few data points occurred on slopes greater than 40^{o}. When the slope data was divided into the 1-7^{o}, 8-17^{o}, and 18-70^{o} groups, each vegetation type was found to be rather uniformly distributed over all combinations of slope and aspect. In other words, none of the vegetation types were found only at certain combinations of slope and aspect.

To simplify the distribution of the species as a function of aspect as well as elevation, the aspect data was collapsed to a linear scale

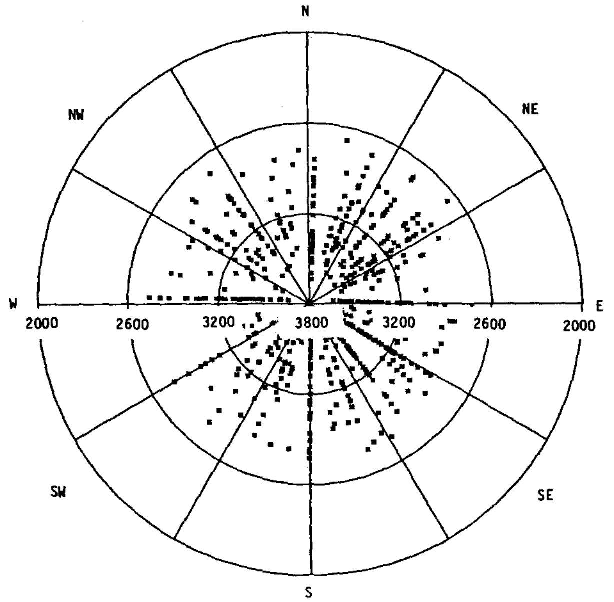

Figure 4. Polar plot of Engelmann spruce/sub-alpine fir as a function of elevation and aspect.

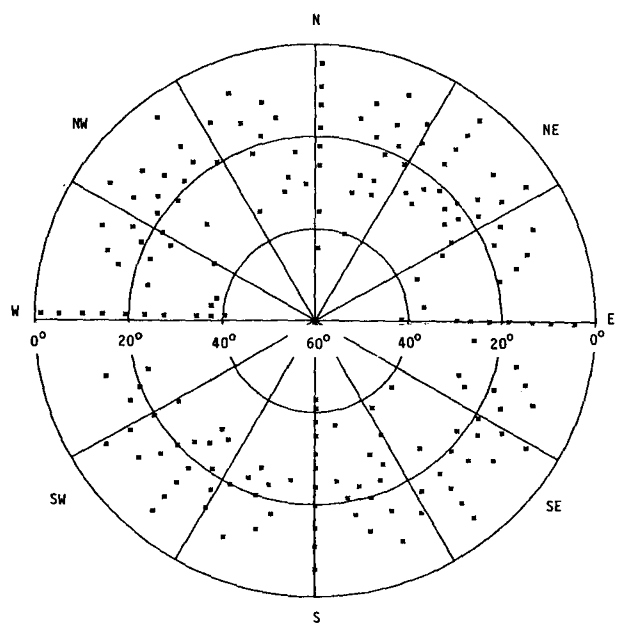

Figure 5. Polar plot of Engelmann spruce/sub-alpine fir as a function of slope and aspect.

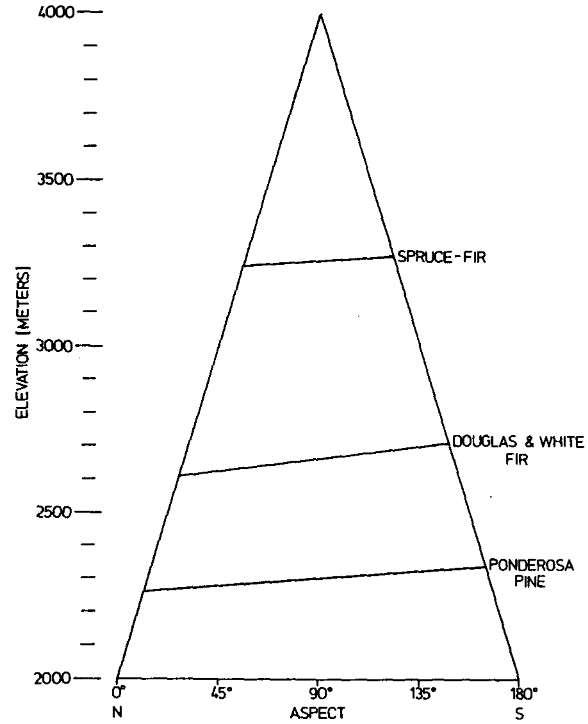

Figure 6. Regression lines for distributions as a function of elevation and aspect.

(north = 0, south = 180, with east and west both 90). A regression analysis of these data confirmed in a more quantitative manner that the average elevation of individual species is higher on the southern aspects than on the northern. A regression analysis of elevation versus aspect was also conducted for each cover type on the three slope classes (1-7°, 8-17°, and 18-70°). The results indicated that slope appears to have no significant effect on the distribution of the various cover types.

Further analysis of the Topographic Distribution Model data was conducted to determine which variables are statistically significant in distinguishing among the various species, and to determine the potential for using topographic data alone to distinguish among the species within each major cover type group. A discriminant analysis approach was used, in which the vegetative cover types were grouped into the three Level II categories (i.e., coniferous forest, deciduous forest, and herbaceous), and the SPSS discriminant function (6) was run on each category. (In general terms, this involved a principal components transformation of the data, followed by a maximum likelihood classification.) To double-check the results of the previous regression analysis of the topographic data, topo-

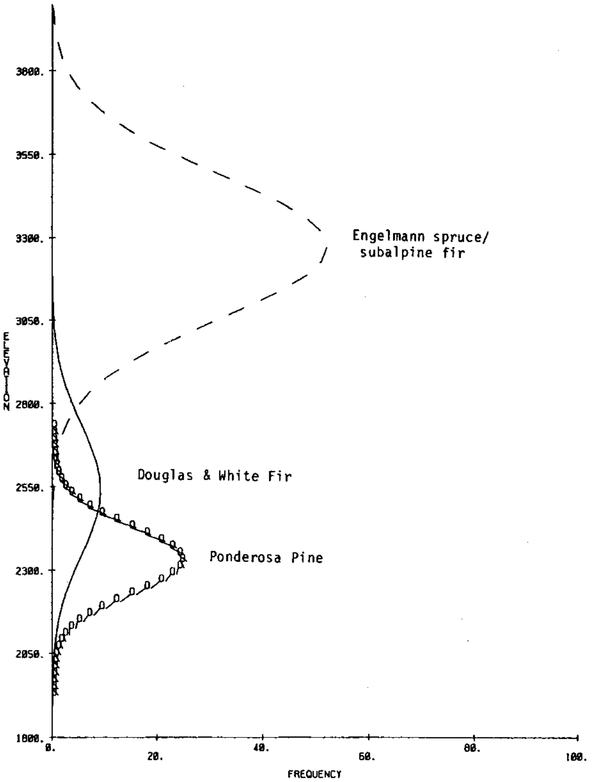

Figure 7. Gausian curves of frequency along the elevational gradient.

graphic variables for all sample points within each Level II category were input to the discriminant analysis function. The processor was allowed to select the significant variables and perform the classification of the sample points.

The results of this discriminant analysis indicated that the most significant variable was elevation, explaining almost 98% of the variation between cover types (within each of the major [Level II] cover type groups). The second, and only other significant variable was aspect, specifically the 0-180^{o} channel. The range of training accuracies for the various species was from 70.8% for aspen to 100% for grass, with the average being 89.5%. In all cases, species in the middle elevation ranges (i.e., Douglas-fir/white fir, aspen, and meadow), were classified less accurately than the other two classes in each group, mainly because each middle class is flanked by transitional zones. These results indicated that the topographic data should be helpful in distinguishing among the individual Level III cover types within each of the major or Level II cover type categories—a differentiation which could not be accurately accomplished using spectral data alone (1).

These accuracy figures however, would not be expected to be representative of those obtained in the final classifications, when Level III categories would be differentiated using both spectral and topographic data.

IV. APPLICATION OF TOPOGRAPHIC DISTRIBUTION MODEL

A. TRAINING PROCEDURES

In developing the training statistics for classifying a combination of spectral data and topographic data, a primary consideration was that two very different types of data were involved in the analysis; spectral data and topographic data. Consideration had to be given to the development of one set of training statistics that was appropriate for the topographic data, and that an independent set of training statistics might need to be developed which would be appropriate for the spectral data. In previous work with the Skylab MSS data (2), the modified cluster technique had been used to develop the spectral training statistics. It was concluded that although these training statistics did characterize the spectral characteristics of the various cover types present, they did not adequately define their topographic characteristics. The result was that the addition of topographic data increased classification accuracies for some of the cover types but lowered them for others. It was hypothesized, however, that if the various forest cover types were significantly distributed as a function of their topographic position (particularly elevation), one could use the spectral data to identify major cover types (i.e., coniferous forest, deciduous forest, herbaceous vegetation, rock and soil, and water) and then depend upon the topographic characteristics to classify and delineate individual forest cover types. Such a procedure would require two sets of training statistics—one to define the spectral reflectance characteristics of the cover types and another set to define the topographic characteristics of the cover types. Two different methods for creating these different types of training statistics were developed and tested in the current study. These are referred to as the Multi-Cluster Blocks (MCB) approach (5), and the Topographic Stratified Random Sample (TSRS) approach.

The Multi-Cluster Blocks approach, previously referred to as the modified clustering technique (8) was used to develop only the spectral training statistics. In this technique, Landsat color composite imagery and small-scale color infrared aerial photography of the area were used to select a number of relatively small blocks in the data. Each block was approximately 40 x 40 pixels in size and contained a diversity of spectral features. Careful selection of these training blocks enables spectral characterization of all cover types in the entire area. Each training block was clustered independently into 16 spectral classes. Since each cover type may occur in more than one of the training blocks, the spectral characteristics for the different training blocks were then merged to produce a single set of training statistics to describe the spectral characteristics of the cover types.

The TSRS (Topographic Stratified Random Sample) approach was the procedure for developing the topographic distribution model discussed in the previous section. The site was stratified into numerous topographic positions and an equal size sample of single pixels was selected from each strata. The result was a statistically valid sample which described quantitatively, the topographic distribution of each vegetation type. However, since the sample was stratified by topography, particularly slope and aspect, it also provided an excellent sample from which the spectral characteristics of each cover type could be estimated. Every slope and aspect combination was represented for each vegetation type, so all variation in spectral response for each type due to topography was included. Also, the sample size for each cover type was thought to be large enough that any variation in density would also be represented in the sample. This was not true for the water class which occupied a very small percentage of the study site. Although this approach required considerable effort to identify the relatively large sample of pixels, it could be used to develop training statistics for both the topographic data and the spectral data.

One point to note is that the sample defined by the TSRS procedure enabled the spectral characteristics of individual forest cover types to be determined. The multi-cluster blocks approach, on the other hand, could be used to effectively describe only major cover types because the clustering procedure determines the natural groups of spectral features in the study site, and in this area, only major cover types could be reliably separated on the basis of spectral characteristics alone. Therefore, even though the TSRS approach could be used to obtain spectral data for individual cover types, the actual spectral separability of these individual forest cover types was rather poor. Thus, the TSRS approach had no real advantage over the MCB approach for developing the spectral training statistics in terms of the level of detail that could actually be defined on the basis of spectral separability.

B. CLASSIFICATION OF SPECTRAL AND TOPOGRAPHIC DATA

Once the statistical distributions (training statistics) have been developed, the classification of the data set can be accomplished by any of several different approaches. The purpose of the classification step is to logically combine the spectral and topographic classes to obtain the desired informational classes. The major difficulty encountered is that the spectral classes and topographic classes do not necessarily match the informational classes. In other words, there is not always one topographic class and one spectral class for each informational class.

The two types of training procedures described in the last section provide two different techniques for classifying the data. Therefore, during this study, both of these techniques for classifying the data were developed and evaluated. One technique utilizes the MCB statistics to train the spectral data, and the TSRS statistics to train the topographic data; a multi-stage or layered classification algorithm is then used to combine the spectral and topographic data. This procedure will be referred to as the "Layered" approach. The second procedure utilizes just the TSRS statistics to train both the spectral and topographic data in a straight "supervised" type approach. This technique will be referred to as the Single-stage approach. Both classification algorithms make maximum likelihood decisions on a pixel-by-pixel (perpoint) basis, but differ in the logic sequence for making the classification decisions. It should be noted that numerous other approaches for utilizing the MCB and TSRS training samples and the classification algorithms are possible, but will not be discussed in this paper.

The main interest of this study was to evaluate the addition of topographic data in the classification process. Once the different approaches have been developed to classify the combined data, the type(s) of topographic data utilized can be varied to determine the incremental improvement in classification accuracy made possible by the addition of the topographic data. The three major levels compared in this study were: (1) the spectral data only; (2) spectral data plus elevation; and (3) spectral data plus all topographic data (elevation, slope and aspect). The classification using only the spectral data was the "base-line" classification, against which the other classifications were compared. It was anticipated that results from the "base-line" classification would be comparable to previous studies in this area in which only the Landsat spectral data had been utilized to classify cover types. The spectral plus elevation data classification would indicate the improvement in results that could be achieved from using the elevation data in combination with the spectral data. The spectral plus all topographic data would indicate the maximum contribution that the DMA topographic data would make when combined with the Landsat spectral data.

The Layered approach, developed for classifying the spectral and topographic data, uses the MCB statistics to train the spectral data and the TSRS statistics to train the topographic data. A multi-stage classification algorithm then combines the spectral and topographic data to extract the desired informational classes. This optimizes the capabilities to extract information from both the spectral and topographic data. The procedure for combining the spectral and topographic data is defined for the algorithm by a decision tree. The decision tree developed for the Layered classification in this study is shown in Figure 8. At each decision node, the

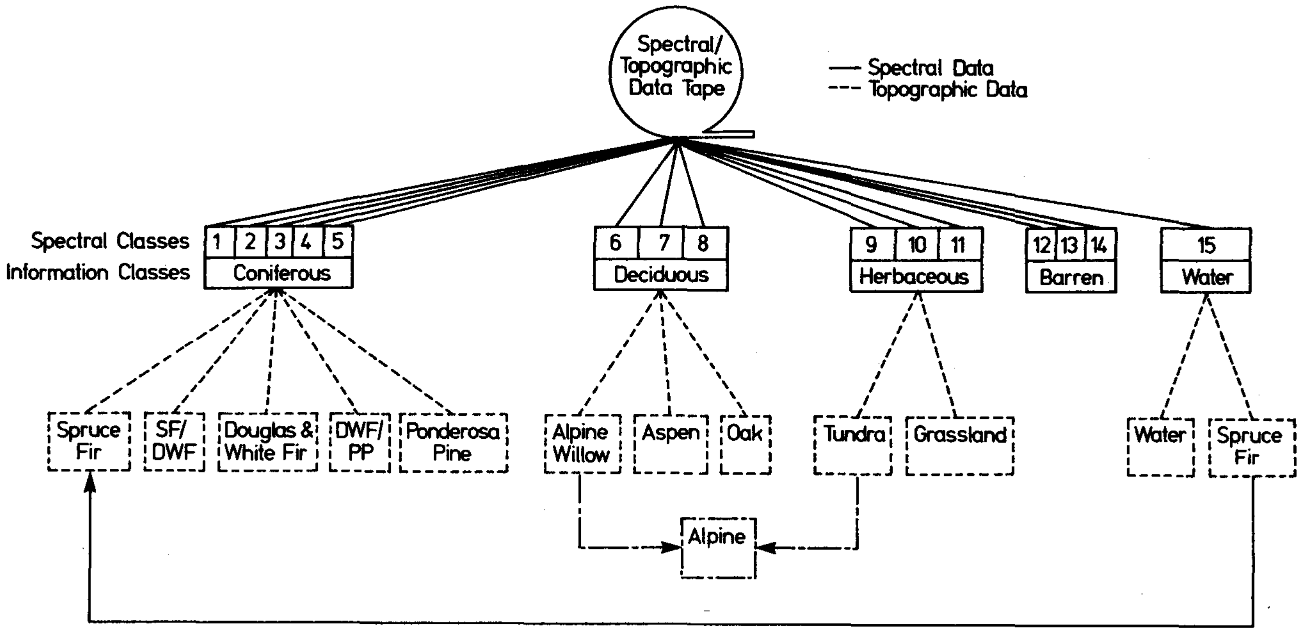

Figure 8. Tree structure for Layered classification approach to combine spectral and topographic data.

classifier uses the kind of data that provides the best information for making that decision. First, the spectral data alone was used to distinguish between the five primary ground cover types present in the study area: coniferous forest, deciduous forest, herbaceous areas, barren areas, and water. Then the major vegetation types were subdivided into the individual forest cover type classes, which could best be accomplished with the topographic information. In this way, information about the topographic distribution of each pixel could be used to differentiate among the individual forest cover type groups that were not spectrally separable. One could attempt to differentiate among crown closure percentages of each forest type as a third stage in the classification, using spectral information, but that level of detail was not attempted in this phase of the project.

The topographic information could also be used in the second layer of the classification procedure to check the pixels that had been spectrally identified as water to see if they were located on a flat surface. By determining if the slope was zero, any topographic shadow points (usually coniferous forest or north and northwest slopes) that were identified as water can be changed to spruce-fir.

The Single-stage classification approach is basically a straight supervised type analysis procedure. In this study, the training statistics were calculated from the topographic stratified random sample of pixels using all channels of data—both MSS and topographic. The topographic and spectral characteristics of each informational class were developed without attempting to group the classes into spectrally or topographically distinct classes. Thus, there was one and only one set of statistics, containing both spectral and topographic data, for each informational category of interest. This enabled the "standard" single-stage, maximum likelihood perpoint classification algorithm to be used. Although this procedure is simple and straight forward, it is not a completely logical approach. Some classes (specifically non-vegetated) cannot be characterized topographically, particularly in terms of elevation. For example, the water class does not have a characteristic elevation zone, and utilization of the elevational channel in classifying water should (and did) cause errors. The same problem occurs with the spectral data in that some cover types have more than one spectral class, and it seems logical that limiting each cover type to just one spectral class would result in errors. Both of these type problems could be avoided by using a different classification approach (e.g., Layered), but still using the TSRS statistics to train the topographic and spectral data.

C. PROCEDURES FOR COMPARING THE CLASSIFICATION RESULTS

In order to compare quantitatively the results of the various classification approaches, a test data set was developed. Because of the complexity of the test site and the ensuing complexity of statistical sampling procedures, the best approach was to use individual Landsat pixels for the test samples. In order to estimate the overall classification accuracy for each quadrangle within ± 5% at a 90% confidence interval, it was determined that 200 test pixels would be required for each quadrangle. To be sure that a minimum of 200 test pixels would be available in each quadrangle after the photointerpretation and field work, an initial set of 300 pixels per quadrangle was randomly selected, giving a total of 2100 pixels over the seven test quadrangles. The location of these pixels in the test quadrangles was then plotted by computer on 1:24,000 scale line-printer maps. Next, a three-step process was used to identify the cover type associated with each test pixel. First, a tentative identification was made for each of the pixels using the INSTAAR cover type maps. Then, as many of the test pixels as possible were field checked; approximately 20% were checked during the time available in the field. Following the field work, detailed photointerpretation was undertaken to establish positive identification of all test pixels. The areas which had been field checked were used to establish confidence in the photointerpretation activity. The photointerpretation was conducted by using the Zoom Transfer Scope to align the color infrared photos with 1:24,000 scale printouts of each quadrangle on which the test pixels had been plotted. Each test pixel was then located on the aerial photography and interpreted, with stand density as well as cover type recorded. During the photointerpretation process, pixels which were located too close to a border between two cover types to positively identify were excluded from the test data set. Pixels that fell on clouds or cloud shadows on the aerial photography were also excluded. This resulted in a decrease from the potential 2100 test pixels to 1539 pixels actually being defined as the test data set. This test sample size was still sufficient to achieve a ± 5% error in the overall classification accuracy by quadrangle at the 90% confidence level.

D. RESULTS OF TOPOGRAPHIC AND SPECTRAL ANALYSIS

The techniques for incorporating the topographic data into the classification process were evaluated by comparing the classification accuracy and the computer cost (CPU time) of each procedure. The two factors involved in the classification procedure which were evaluated are; (a) the type of topographic data utilized by the classification algorithm and its effect on the classification, and (b) the techniques for developing the training statistics and classifying the data. To effectively evaluate each factor, all other parameters were held constant. The test sample discussed in the previous section was used to estimate the accuracy of all classifications discussed. The sample provides unbiased estimates of the overall classification accuracy for each of the seven test quadrangles. By post-stratifying the sample into cover types classes, the overall accuracy for each cover type can also be estimated.

The first major factor evaluated was the effect topographic data would have on classification performance. To provide a baseline classification with which to evaluate the inclusion of topographic data, the TSRS statistics were used to define the spectral classes for each Level III cover type. The TSRS statistics were the only training data that could provide spectral classes at the Level III degree of detail. The computer classification time was 55.5 seconds and the classification accuracy was estimated to be 49.4 percent at Level III and 71.1 percent at Level II. These overall accuracies are comparable to those reported in other investigations for relatively large sites in the San Juan Mountains (2) and (7). It should be noted that the more statistically valid method of defining the test data set used in this study should have eliminated any analyst bias, which may have been present in the manually selected test data sets used during the previous investigations. To confirm this, a manually selected sample of test fields were defined, and the overall classification performance for these test fields was 79.6% at Level II. This is considerably higher than the 71.1% obtained by the random sample of individual test pixels. This indicates that there may be a significant positive bias in the estimates of classification performance when measured using manually selected test fields, even in situations where the analyst is trying to define a representative, unbiased set of test fields.

To determine the improvement in classification performance that can be achieved by adding the topographic data to the spectral data, the data were classified using the spectral data plus elevation and using the spectral data plus three topographic channels (elevation, slope, and aspect). To make a valid comparison with the baseline classification, the TSRS statistics were used to train both the spectral and topographic data. Table 2 summarizes the overall classification accuracy by quadrangle for the three combinations of data channels utilized: (a) spectral only (baseline), (b) spectral plus elevation, and (c) spectral plus topographic. The classification accuracies are given at a Level III (individual forest types) degree of detail. An Analysis of Variance (ANOVA) was performed on a arcsin square root transformation of the data in Table 2. The results indicated a significant difference between the various combinations of data channels and also a significant difference between quadrangles. A Newman-Keuls range test indicated that the inclusion of topographic data--either just elevation, or elevation, slope and aspect--significantly increased the classification accuracies over using just the spectral data. There was, however, no significant difference between using just elevation and using all three topographic parameters. The lack of improvement when slope and aspect data were used in the classification

Table 2. Classification performance showing the impact of topographic data by quadrangle for Level III cover types.

		Data Utilized		
	Sample Size	Spectral Only	Spectral + Elevation	Spectral + Topographic
Oakbrush	199	43.7	50.8	56.3
Finger Mesa	214	38.6	67.0	63.7
Granite Peaks	202	56.9	79.7	80.2
Pagosa Springs	237	49.6	66.4	63.9
Devil Mountain	233	51.9	60.5	65.7
Weminuche	212	59.0	73.6	74.1
Ludwig Mountain	242	45.9	59.5	55.4
Overall	1539	49.4	65.6	65.9

may be due to inaccuracies in the topographic data and/or improper characterization of the slope/aspect effects on the distributions (i.e., a different gradient using a combination of both slope and aspect may be more effective). The significant difference between quadrangles was somewhat unexpected, but indicates that when analyzing relatively large, complex sites, some quadrangles are much easier to classify accurately than others. Evaluation of the quadrangles with the worst accuracies indicated them to be the most complex areas, the result of numerous vegetative disturbances. If the computer analysis was of a smaller area (e.g. one quadrangle), the accuracy has been shown to be much higher, even in complex areas (1). To further evaluate the addition of topographic data, the test sample was summarized by cover type for each combination of data channels. An ANOVA of the data confirmed the significant improvement in classification performance when topographic data is added to the spectral data. The ANOVA also showed a significant difference between cover types, indicating that not only are some areas easier to classify than others, but that some cover types are easier to classify than others.

The computer time required to perform the classification is known to increase with the number of channels used (3). The CPU times measured in this study were 55.5 seconds for the spectral data only (4 channels), 169.6 seconds for the spectral data plus elevation (5 channels), and 211.0 seconds for the spectral data plus topographic (7 channels). There was a considerable increase in computer time when going from 4 channels to 5 and from 5 channels to 7.

The second major effect studied was the approach used to accomplish the computer-aided analysis using both the spectral and topographic data. The two different approaches evaluated in this study for developing the training statistics and classifying the data were: (a) supervised TSRS statistics to train both the spectral and topographic data, and using a single-stage classification algorithm, and (b) TSRS statistics to train the topographic data and the MCB statistics to train the spectral data, and using a multi-stage classification algorithm. Since using the spectral data and all three topographic channels should give the maximum classification accuracy, all seven channels of data were utilized in both analysis procedures. Table 3 summarizes the overall classification performance by quadrangle for both analysis procedures. An ANOVA of the Level III results in Table 3 indicated no significant difference between the two analysis procedures, but again a significant difference between quadrangles. Both analysis procedures performed equally well on all quadrangles, except Devil Mountain. The computer times for the two analysis procedures were 50.6 seconds for the Layered approach and 211.0 seconds for the Single-stage approach. Thus, even though the classification performances were not significantly different, the Layered approach was over four times faster than the Single-stage procedure.

Table 3. Classification performance of two analysis techniques by quadrangle for Level III cover types.

Quadrangle	Sample Size	Layered Analysis Procedure (MCB for spectral, TSRS for topographic, and multi-stage algorithm)	Single-Stage Analysis Procedure (TSRS for spectral, TSRS for topographic and single-stage algorithm)
Oakbrush	199	57.3	56.3
Finger Mesa	214	64.5	63.7
Granite Peaks	702	78.2	80.2
Pagosa Springs	237	60.3	63.9
Devil Mountain	233	53.6	65.7
Weminuche	212	73.6	74.1
Ludwig Mountain	242	54.5	55.4
Total	1539		
Overall Performance		63.6%	65.5%

The test sample was also summarized by cover type for each analysis procedure, as shown in Table 4. An ANOVA of the results again indicated no significant difference between analysis procedures, but a significant difference between cover types. The effects of the different logic sequences in the two analysis procedures show up in this table. By making several decisions in a series of steps, the accuracy was improved over the single-stage approach for the three cover types with the lowest classification performance. The deciduous cover types (oak and aspen) both increased in accuracy as the result of making the spectral decision separate from the topographic decision using the Layered classification. When the decisions are made at one time in a single-stage classification procedure, the accuracy is

Table 4. Classification performance showing the impact of topographic data by Level III cover type.

Forest Cover Type (Level III)	Sample Size	Layered Analysis Procedure (MCB for spectral, TSRS for topographic, and multi-stage algorithm)	Single-Stage Analysis Procedure (TSRS for spectral, TSRS for topographic and single-stage algorithm)
SF	313	89.1	88.5
SF/DWF	156	58.3	75.0
DWF	39	46.2	48.7
DWF/PP	144	80.6	72.9
PP	265	60.0	71.3
Aspen	110	43.6	35.5
Oak	97	46.4	39.2
Alpine	79	70.9	78.5
Grassland	245	50.6	51.4
Barren	86	46.5	37.2
Water	5	60.0	80.0
Total	1539	63.6	65.5

reduced by confusion with other topographically similar cover types. The non-vegetated barren class does not have a characteristic topographic distribution, and therefore the use of the topographic data in the single-stage classification causes confusion and results in a decreased accuracy. The water test class had a rather low classification performance because a topographic line crossed one of the reservoirs and two of the five water test pixels happend to fall on this topographic line. The larger coniferous cover types decreased in accuracy in the Layered classification, offsetting the increases in other classes and resulting in no significant difference between analysis procedures. But in general, the Layered approach may be considered better since it raised the accuracy of a number of the worst classes, provides a logical sequence of decision points, and significantly reduces computer time. The key to the problem of using the Layered classifier is in developing appropriate statistics to train the multi-stage classification algorithm. Another logical classification procedure would be to use the TSRS statistics to train both the spectral and topographic data, but use the multi-stage classification algorithm. Other analysis procedures are also possible, some of which are in the process of being evaluated.

V. CONCLUSIONS

An effective computer-aided analysis technique has been developed and applied to mapping individual forest cover types in the rugged terrain of southern Rocky Mountains. This study represented a detailed analysis of a very complex area in terms of both vegetation types and topography. The results were evaluated using a random sample of single pixels to provide a statistically valid estimation of the overall classification accuracy. The following are the major conclusions of this study.

1. The stratified random sample approach for developing the topographic distribution model proved to be very effective, and provided a statistically valid quantitative description of the topographic distribution of the various vegetation types. The model is a quantitative approach to describe the ecological characteristics of the various cover types in a particular geographic area. Also, the topographic stratified random sample (TSRS) approach appeared to be effective, for developing spectral training statistics provided that the various cover types are represented by an adequate sample of data points.

2. The use of topographic data significantly improved the overall classification accuracy of forest cover types as compared to using only spectral data. Elevation is particularly significant in helping to improve the classification accuracy. Use of elevation data in addition to the spectral data provided an improvement in overall classification accuracy of approximately 15%. The addition of slope and aspect data did not significantly improve the classification performance over using elevation and spectral data. This was believed to be due to the characteristics of the DMA elevation data and the way in which the elevation data was digitized. Better quality topographic data would probably show an improvement in the classification performance due to aspect and perhaps slope.

3. No significant difference in the overall classification accuracy was found between the two methods of developing training statistics and classifying the data. The Layered approach showed the most promise since the multi-stage classification algorithm minimized the complexity of the classification process by allowing a sequence of logical decisions to be made. This approach both increased the accuracy of a number of important classes, and greatly reduced the amount of computer time required.

4. Although not discussed in detail in this paper, a reflectance geometry correction model was developed and tested, and, based upon analyst evaluation of grayscale maps, seemed to "correct" the Landsat data to remove topographic variation in spectral response. However, a classification accuracy evaluation did not indicate any improvement due to the "correction" model. This was thought to be due to errors in the slope and aspect for many of the individual pixels, caused by the characteristics of the DMA elevation data.

5. The Defense Mapping Agency digitized elevation data is reasonably accurate and is effective for situations in which only generalized elevation information is required or is utilized. However, the slope and aspect data generated from the elevation data had limitations due to the procedures involved in the digitization process. These procedures caused a large percentage of the elevation cells (pixels) to be assigned to the

61 meter (200 foot) contour interval bins. As a result, the interpolation process inaccurately defined the slope and aspect, causing a relatively large number of errors in the slope and aspect data sets. Therefore, for purposes of generating slope and aspect data, the DMA elevation data has serious limitations.

VI. REFERENCES

(1) Fleming, M.D., et al., 1975. Ecological Inventory, Section A of Natural Resource Mapping in Mountainous Terrain by Computer-Analysis of ERTS-1 Satellite Data, by R.M. Hoffer and Staff. Agricultural Experiment Station, Research Bulletin 919, Purdue University, West Lafayette, Indiana 47907.

(2) Hoffer, R.M. and Staff, 1975b. Computer-Aided Analysis of Skylab Multispectral Scanner Data - Mountainous Terrain for Land Use, Forestry, Water Resource, and Geologic Applications. LARS Information Note 121275. Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana 47907. Final Report on NASA Contract #NAS 9-13380, Skylab EREP Project 398.

(3) Krebs, Paula V. and Staff. 1976. Multiple Resource Evaluation of Region 2 U.S. Forest Service Lands Utilizing Landsat MSS Data. Final Report for NASA Contract NAS 5-20948. Institute of Arctic and Alpine Research, University of Colorado, Boulder, Colorado.

(4) Anuta, P.E., 1973. Geometric Correction of ERTS-1 Digital Multispectral Scanner Data. LARS Technical Report 103073. Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana 47907. 23 pp.

(5) Daubenmire, R.F. 1943. Vegetational Zonation in the Rocky Mountains. Bot. Rev., 9:326-393.

(6) Nie, N.H., Hull, C.H., et al. 1975. SPSS: Statistical Package for the Social Sciences, Second Edition, McGraw-Hill Book Company, New York.

(7) Fleming, M.D. , 1977. Computer-Aided Analysis Techniques for an Operational System to Map Forest Lands Utilizing Landsat MSS Data. Masters Thesis and LARS Information Note 112277, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana 47907. 236 pp.

(8) Fleming, M.D., J.S. Berkebile, and R.M. Hoffer, 1975. Computer-Aided Analysis of Landsat-1 MSS Data: A Comparison of Three Approaches, Including a "Modified Clustering" Approach. Proceedings of the Symposium on Machine Processing of Remotely Sensed Data. pp. 13-54 - 13-61 (LARS Information Note 072475).

(9) Landgrebe, D.A., 1976. SRT Final Report, NASA Contract NAS 9-14016, June 1, 1975 - May 31, 1976, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana 47907. pp. 2.7-4 to 2.7-17.

ACKNOWLEDGEMENTS

This research was supported by NASA Contract NAS 9-15508, and we thank Mr. David Amsbury, our NASA Technical Monitor for his assistance and advice. The authors also wish to thank several colleagues at LARS, particularly Dr. Luis Bartolucci, Mrs. Shirley Davis, Mr. Ross Nelson, Mr. Paul Anuta, Mr. Luis Lang, and Dr. Virgil Anderson, for their contributions to this research project.

Michael Fleming is a Research Associate at the Laboratory for Applications of Remote Sensing and a Ph.D. Candidate in the Department of Forestry and Natural Resources, Purdue University. He received a B.S. in Forest Management from Northern Arizona University (1973) and a MSF in Forest Remote Sensing from Purdue University (1977). Since 1973, Mr. Fleming has participated in numerous remote sensing projects, including Landsat-1, Skylab, U.S.F.S. and Landsat Follow On. His emphasis has been on the design, development and statistical evaluation of computer-aided analysis techniques for applications to natural resource management. Professional memberships include the Society of American Foresters, the American Society of Photogrammetry and several Honorary societies.

Roger Hoffer is Professor of Forestry, and Leader, Ecosystems Research Programs, LARS, Purdue University. He was a co-founder of LARS in 1966; has lectured and participated in remote sensing projects in various countires throughout South America, Asia, and Europe; has served as a principal investigator on Landsat, Skylab, and other remote sensing projects; has authored over 100 scientific papers on remote sensing. Dr. Hoffer teaches three courses in Remote Sensing; is a member of the American Society of Photogrammetry (where he has served as Director of the Remote Sensing and Interp. Div., and Assoc. Editor of Photogrammetric Engineering and Remote Sensing), Society of American Foresters, Sigma Xi, Phi Kappa Phi, and other professional and honorary societies. He is a Certified Photogrammetrist and is listed in American Men and Women in Science.