

AN EVALUATION OF THE EFFECTIVENESS
OF AN
UNSUPERVISED CLASSIFICATION OF LANDSAT IMAGERY

by

Ruth A. Norman

A Thesis

Presented to

The Faculty of Humboldt State University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science

March, 1983

AN EVALUATION OF THE EFFECTIVENESS
OF AN
UNSUPERVISED CLASSIFICATION OF LANDSAT IMAGERY

by

Ruth A. Norman

Approved by the Master's Thesis Committee

[REDACTED]

Lawrence Fox, III, Chairman

[REDACTED]

Gerald M. Allen

[REDACTED]

William R. Sise

[REDACTED]

Natural Resources Graduate Program

Approved by the Dean of Graduate Studies

[REDACTED]

Alba M. Gillespie

[Handwritten signature]

ABSTRACT

An unsupervised classification of Landsat data covering the Warner Mountain Ranger District of the Modoc National Forest, California was analyzed. Field data were in the form of Timber Stand Improvement (TSI) and Compartment Inventory Analysis (CIA) records. General stand descriptions were determined for four of the six conifer classes using mean values for stand characteristics from these timber stand examinations and spectral signatures of the classes. Discriminant analysis was applied to the average timber stand composition information paired with the six classes, from the Landsat data, identified as "conifer" by the California Department of Forestry (CDF). High standard deviations in each of the measured stand variables and poor assignments of plots to classes (41.2 to 65.2 percent) in the discriminant analysis limited the detail of the class descriptions. The results of this study indicated that unsupervised classification was not adequate for conventional timber type mapping in the study area.

ACKNOWLEDGEMENTS

I would like to thank Dr. Larry Fox for his helping me obtain the data for this study from NASA and CDF. To all the people on the Warner Mountain Ranger District, past and present, besides myself, who did CIA plots and TSI stand examinations in the south Warners I am grateful. I would also like to thank those on the district who helped when I needed information about the plots or where to get additional information about the forest in general. On campus, Dr. Gerald Allen gave me valuable assistance with the computer work and Dr. Bill Sise was always there to talk out my problems with. Finally I would like to thank my roommates who withstood my monopoly of the living room with data and rough drafts.

TABLE OF CONTENTS

	Page
ABSTRACT	iii
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	viii
INTRODUCTION	1
Relevant Case Study	2
Review of Related Literature	6
Objectives	7
STUDY AREA	9
Geology	9
Soils	10
Climate	10
Vegetation	10
MATERIALS AND METHODS	13
Spectral Data	13
Field Data	14
General Class Descriptions	17
Discriminant Analysis	19
RESULTS	22
Class Descriptions	22
Discriminant Analysis	28
DISCUSSION	35
REFERENCES CITED	38

TABLE OF CONTENTS (continued)

Page

APPENDIXES

A. Raw Data from TSI, CIA and Other Plots	40
B. Example of CIA Data Sheet	42
C. Example of TSI Data Sheet	43

LIST OF TABLES

Table		Page
1	Means and Standard Deviations of Variables Collected in TSI Stands and CIA Plots for Spectral Classes 4, 7, W and X	24
2	Results of Discriminant Analysis, Using All Data, for Spectral Classes 4, 7, W and X	29
3	Comparison of Discriminant Analyses Using Classes W and X Only. A) Using Combined TSI and CIA Data B) Using TSI Data Only C) Using CIA Data Only	31
4	Comparison of the Mean Values of the Variables Collected in the TSI Stands and in the CIA Plots for Spectral Classes W and X	33

LIST OF FIGURES

Figure		Page
1	Angle of Landsat Flight Lines in Relation to True North as the Satellite Made Successive Passes Over California	4
2	Map of the Ecological Units, Called Ecozones, Used in Classification of Pixels in Different Terrain and Vegetation Zones	5
3	Study Area Location, Warner Mountain Ranger District, Modoc National Forest, California	10
4	Comparison of Land Use Class and Spectral Class Line Printer Maps	14
5	Sphere of Influence of CIA Plots	18
6	Comparison of the Spectral Signatures of the Six Conifer Classes with One Open Shrub and One Grassland Class	20

INTRODUCTION

With the placement of satellites into earth orbit and the growth of computer technology new sources of earth resources data have become available. Landsat (formerly Earth Resources Technology Satellite) satellites were launched with sensing of the earth's resources as their primary goal. The heart of the system, the multispectral scanner (MSS), senses the reflectance of the earth's surface in four wavelengths. This information permits identification (classification) of vegetation, rock types, water and other surface conditions (Fu and Yu, 1980). The wavelength bands are associated with the colors green (0.5 to 0.6 micrometers, reflected sunlight from green vegetation) and red (0.6 to 0.7 micrometers, sensitive to chlorophyll) as well as two portions of the reflected infrared section of the spectrum (0.7 to 0.8 micrometers, sensitive to density or biomass and 0.8 to 1.1 micrometers, sensitive to leaf water content)(Taranik, 1981). Modern computer software (i.e. the EDITOR program at NASA-Ames) can be used to analyze the reflectance values (digital numbers (dn)) for any or all of the bands.

The imagery from the Landsat satellites is uniform, repetitive, available worldwide on a regular basis (every nine days) and suitable for investigations of large areas. However, a single image, representing a 115 by 115 mile (185 by 185 km) portion of the earth's surface is composed of 2,340 scan lines, each with 3,240 picture elements (pixels) or approximately 7.6×10^6 pixels (Robinove, 1979). Although it is not necessary to study a full image, the number of data points (pixels) demands the use of a computer to undertake any in depth-study.

There are three basic methods of mapping using Landsat data. The first and simplest is visual interpretation of a photo image composite. This involves simply looking at a color composite of the raw data. The second method is an improvement over the first in that the image is enhanced by increasing contrasts in a photo image or a line printer product. The third and most complex method uses the radiance values to classify terrain and produce graphic maps. This last method can be further divided into supervised and unsupervised techniques. Supervised classification requires the manual labeling of pixels with known ground characteristics (training areas). The remaining pixels are grouped, by spectral characteristics, to fit into these established classes. In unsupervised classification the analyst simply designates the number of spectrally distinct classes into which the data are to be classified. The data are then classified into the designated number of spectral classes based on the spectral values of the four wavelength bands. After this process is completed the classes are labeled by manual comparison with ground truth or some form of low altitude remote sensing. The latter technique is less time consuming and requires less detailed knowledge of the ground beforehand.

Relevant Case Study

An unsupervised classification covering a large area with the identification of hardwood, conifer and mixed forests among its goals was the "Land-Cover Classification of California using Mosaicking and High-Speed Processing" project done by NASA (National Aeronautics and Space Administration)-Ames Research Center with the California Department of Forestry (CDF) (Peterson, et al. 1979). In this study the base

data, covering the entire state, consisted of 29 Landsat-1 scenes from one date in August, 1976. Because the satellite does not fly a true north-south path over California, the data, when received, were skewed somewhat. The data were not rotated to fit the true north orientation as it was thought to have been too time consuming. Therefore the grid-lines were at an 11 degree east angle from true north at meridian 117 (Fig. 1). The 29 scenes were joined into one radiometrically smooth, geometrically acceptable mosaic of the entire state. Overlap of the images was treated through bilinear interpolation to adjust for the differences between the portions of scenes that overlapped. To the consolidated scene was added terrain data (elevation, slope gradient and aspect) from the Defense Mapping Agency (DMA) terrain tapes. The goal was to produce, in a short period of time, products useful for land use management. Bulk processing for the entire state was completed in seven months, using only unsupervised classification.

At the time this research was undertaken no one computer system was available that could completely process the data. However, by using four interconnected systems the data were processed. Many problems, especially format compatibility, were solved during this project. As time progresses it is more likely that a single system will be designed that can process the data without assistance.

The state was divided into 29 ecozones, after Küchler's Natural Vegetation Map (Küchler, 1977), to make classification of pixels more uniform within differing terrain and vegetation zones (Fig. 2). One classification would not have been adequate for an area the size of the state of California with its diversity of topography and vegetation.

The use of unsupervised classification was chosen because it was

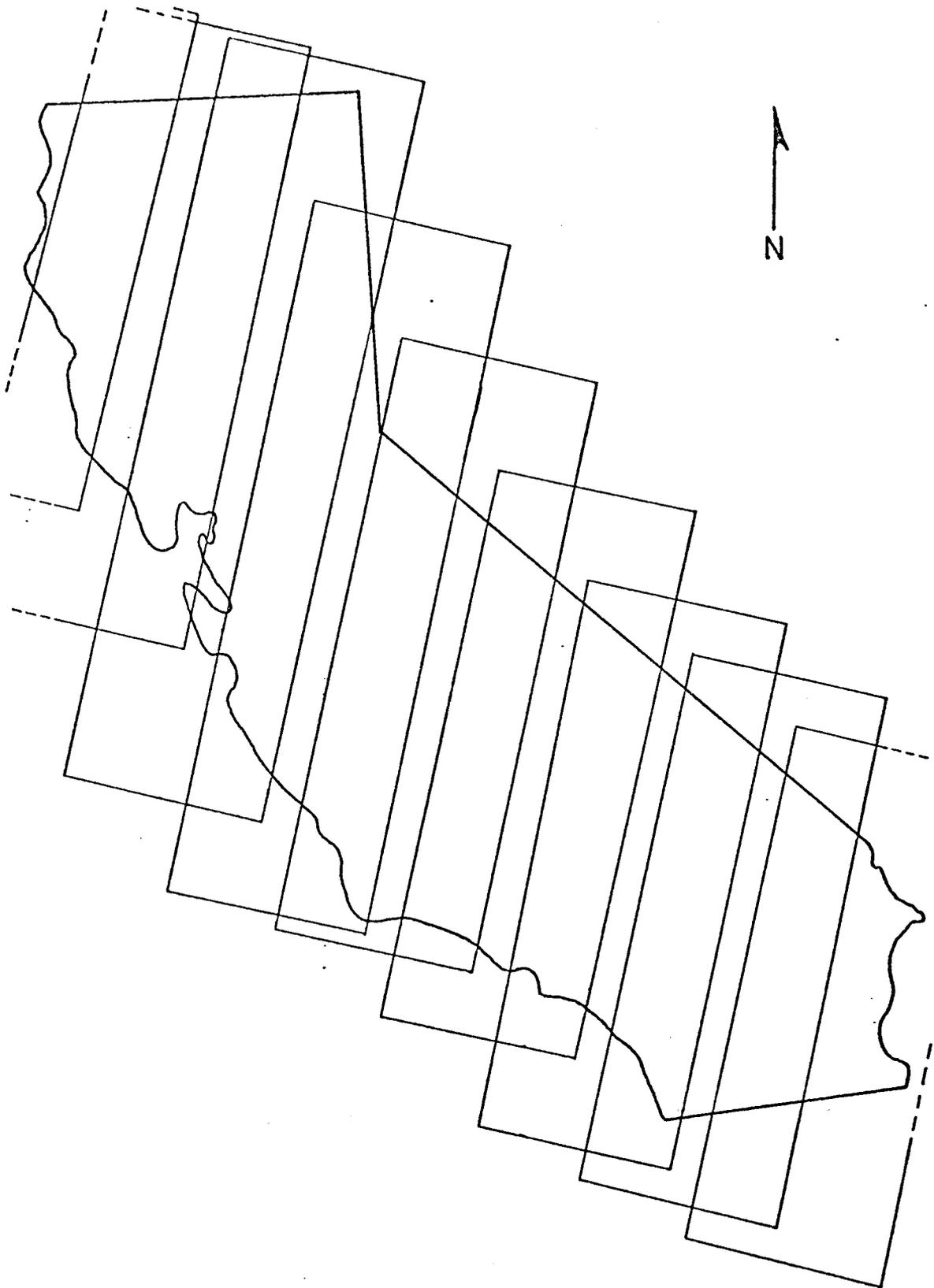


Figure 1. Angle of Landsat Flight Lines in Relation to True North as the Satellite Made Successive Passes over California. From Peterson, et al. 1979.

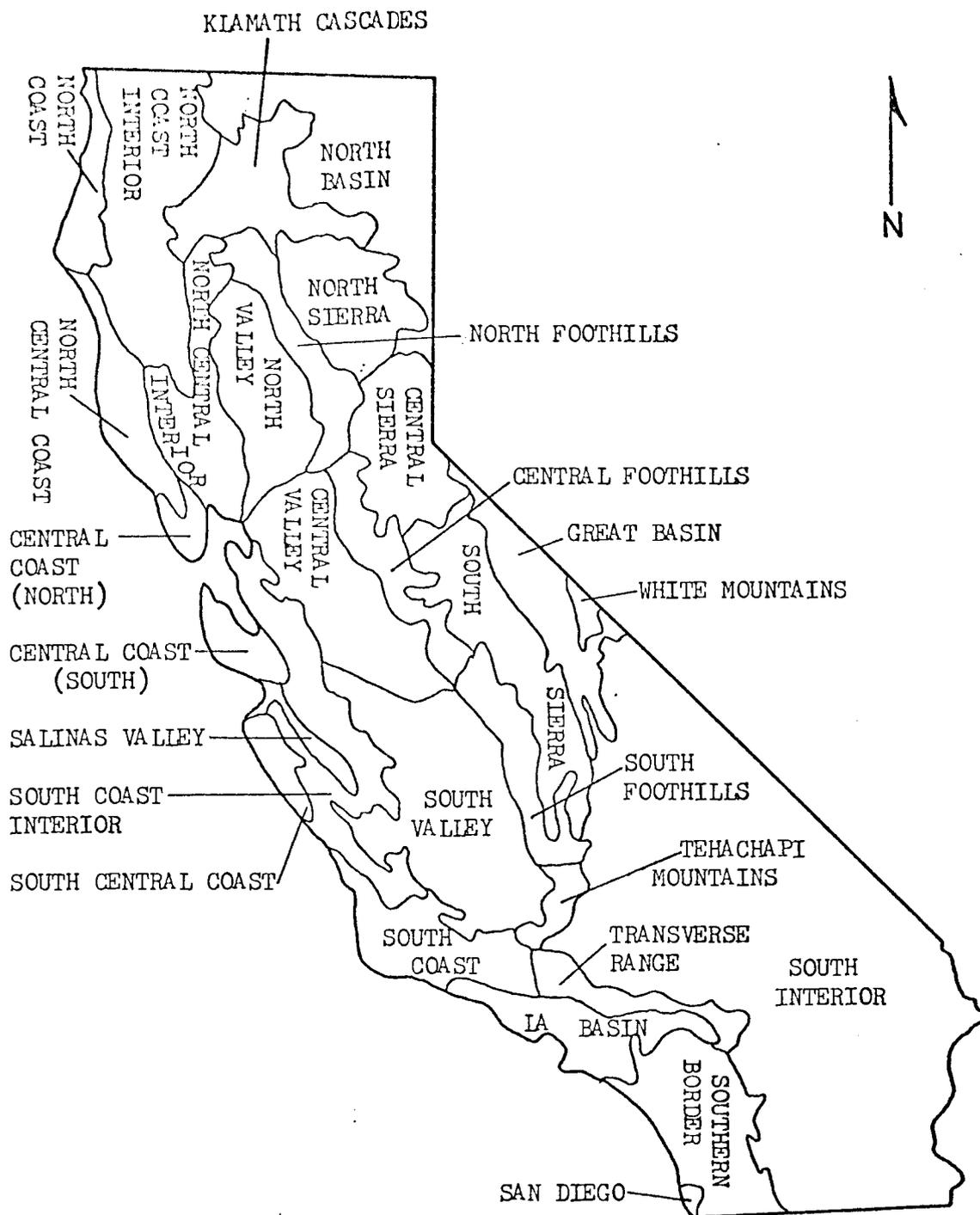


Figure 2. Map of the Ecological Units, Called Ecozones, Used in Classification of Pixels in Different Terrain and Vegetation Zones. From Peterson, et al. 1979.

less time consuming to perform than supervised. Supervised classification, guiding the results by using known information, might have been preferable. However, since this is very labor and time intensive it was not done. Unsupervised classification required only grouping and labeling the classes (i.e. classes 4, 6, 7, 8, W and X were labeled "conifer") after the classification was completed. This grouping and labeling was done by CDF personnel familiar with the ecozones. Where one class in an ecozone was found to represent more than one resource, the terrain data were also used to separate the two by evaluating elevation, slope gradient and/or aspect of the questionable pixels to determine the likelihood of each resource's occurrence.

Review of Related Literature

Although the U.S.D.A. Forest Service could make use of this kind of information derived from Landsat, little has been done with this form of remote sensing to date. In a mailing study only one of 14 National Forests, in the southeast, stated that Landsat imagery was used and then only infrequently (Mead and Rasberry, 1980). Studies done by the Forest Service Range and Experiment Stations have found that Landsat data could be used effectively to generate general information and small scale maps (Aldrich, 1979). Using unsupervised classification forested land could be separated from non-forested land and water with 95 percent accuracy. However, more detailed separations have been found to be less accurate.

In areas with conifer cover, similar to the study area, a variety of studies have been done although few utilized unsupervised classification. One of these, done on level terrain with large areas of

well defined age and species groups of forest in Canada, showed that age classes and species classes could be separated quite well (Beaubien, 1979). However, the same study methods used in a nearby area where the terrain was more rugged, with logging and fire disturbances, resulted in a much reduced ability to separate age and species classes. Stand density and exposure of slopes were recognized as having a strong influence on classification. Other stand conditions that have been found to have a large role in determining classification are; crown closure (Roller and Visser, 1980), age differences (Shimabukuro, et al. 1980), stand density (Katti, et al. 1981) and soil, rock or other background spectra (Tucker and Miller, 1977).

Generally, supervised classifications in forested areas (Oregon) have been found to yield greater accuracy than unsupervised, computer-derived classification (Walsh, 1980). However, little has been done to find out precisely what can be derived from the unsupervised classifications. A noted stumbling block has been that computer-derived classes do not correspond well with pre-defined cover classes (Merchant, 1981).

Objectives

The objectives of this specific study were:

1. to determine which, if any, of the parameters measured during routine stand examinations (Compartment Inventory Analysis and Timber Stand Improvement examinations) were characteristic of the conifer classes that had been generated in the unsupervised classification;
and
2. to describe each of the conifer classes using the information collected in the routine stand examinations and the spectral signatures.

It was hoped that objective 2 would yield descriptions roughly equal to timber types currently in use. Such descriptions, from unsupervised classifications of Landsat data, could eliminate the expensive need to obtain aerial photographs and reduce the errors of timber typing and omission frequently found in maps prepared from aerial photographs by people unfamiliar with the area. The overall process would be much less time and cost consuming than the conventional process of aerial photo-interpretation type map construction (Lowe, 1979), especially for areas the size of a National Forest or entire state.

STUDY AREA

The Warner Mountains are a short spur of the Cascade Range running north to south just west of the Nevada state line in California and extending into Oregon. The portion within California, administered by the Modoc National Forest, Warner Mountain Ranger District, is approximately 65 miles (105 km) long and 10 miles (17 km) wide, covering approximately 334,000 acres (135,000 ha) (Fig. 3). The bulk of the district is within Modoc County with the southern tip extending into Lassen County.

Geology

The Warner Mountains are part of the Great Basin province (North Basin ecozone in Figure 2) as well as being part of the Cascade Range. They are an uplifted fault block capped by an almost continuous mass of Tertiary basaltic lavas known as the Warner Mountain Basalt (U.S.G.S., 1974). Exposed along the Surprise Valley fault, which defines the eastern edge of the range, and in scattered erosion channels are tuffs and volcanic agglomerates known as the Cedarville Series. Time and erosion have only begun to modify the fault block topography. Streams are generally found in steep walled canyons with high flow gradients. Water from rain and snow falling on the southern half of the Warner Mountain Ranger District feeds the Lower Surprise Valley to the east and the Upper Pit River (a tributary of the Sacramento River) to the west. Elevations range from 5,000 feet (1,525 m) to 9,900 feet (3,020 m - Warren Peak).

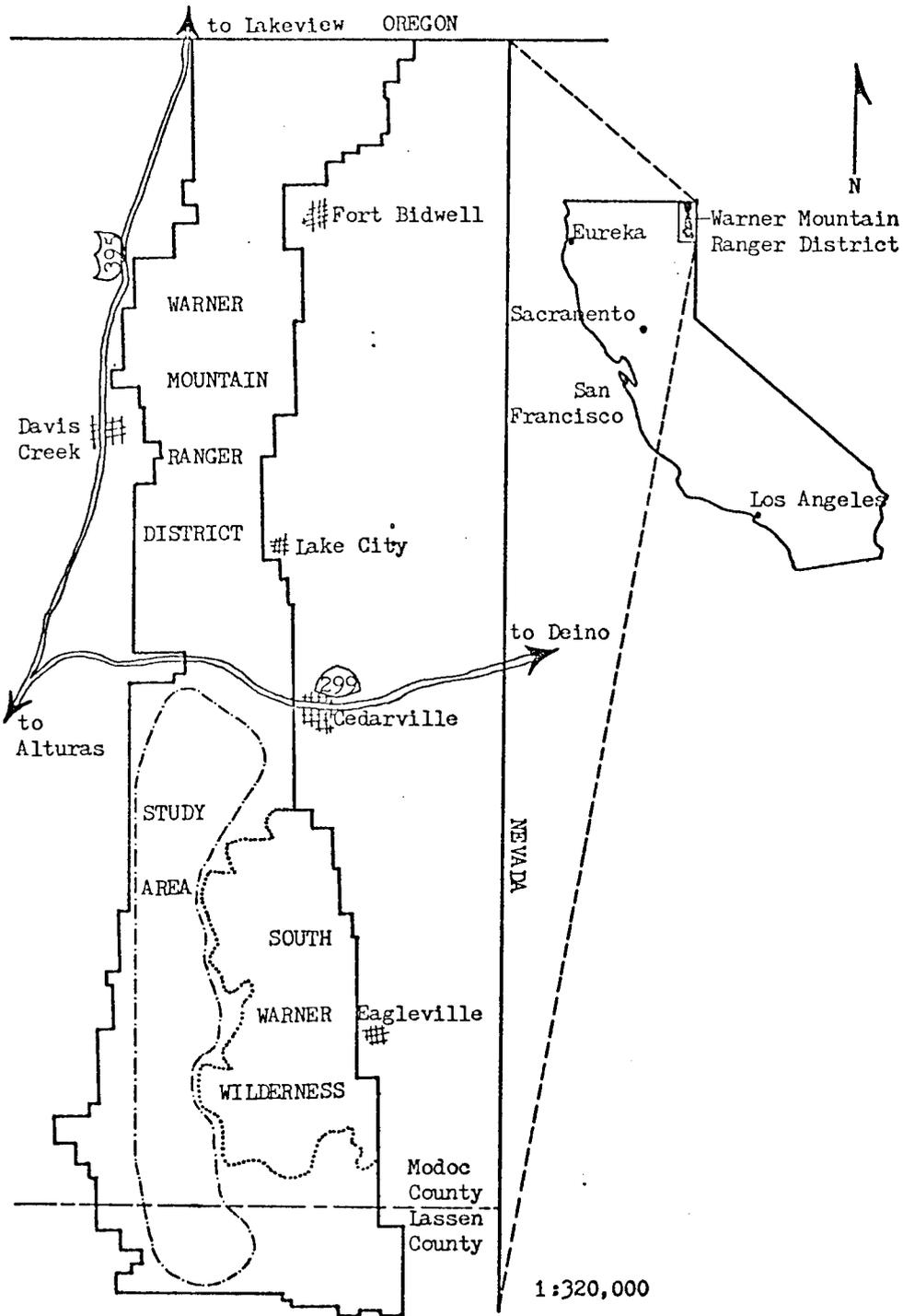


Figure 3. Study Area Location, Warner Mountain Ranger District, Modoc National Forest, California.

Soils

The forest soils are primarily derived from basalt parent material. These soils are medium to coarse in texture and are moderately deep (U.S.F.S., 1975). Because they are coarse, they have the capacity to store water for vegetation use during the summer drought.

Climate

The climate is fairly harsh with limited rainfall and a short growing season. Average annual precipitation ranges from 16 to 25 inches (41 to 65 cm). About three fourths of this is received during the winter, approximately 65 percent as snow. Winter temperatures are often 0° to minus 20°F while summers are hot and dry. All seasons are subject to unusual weather. Unexpected killing frosts have been known to occur in June and July (Brown, 1945).

Vegetation

The forests are composed almost exclusively of conifers. Western juniper (*Juniperus occidentalis*) are found on lower elevation and non-commercial forest lands. With increasing elevation, ponderosa pine (*Pinus ponderosa*) and Jeffrey pine (*Pinus jeffreyi*) are found. Above this and on north facing slopes at lower elevations white fir (*Abies concolor*) predominates. At intermediate elevations there are stands of mixed pine and fir. Limited amounts of western white pine (*Pinus monticola*) and lodgepole pine (*Pinus contorta*) are found at the highest elevations. Hardwoods cover only minor areas; species include

aspen (*Populus tremuloides*), curlleaf mountain-mahogany (*Cerocarpus ledifolius*) and willow (*Salix* sp). Vegetation in rocky open areas is dominated by big sagebrush (*Atemisia tridentata*), rabbitbrush (*Chrysothamnus* sp) and bitterbrush (*Purshia tridentata*). Grasses and broadleaf herbs are found in the wet meadows. The bulk of the forest on the Warner Mountain Ranger District has been subjected to logging and other disturbances in the past.

MATERIALS AND METHODS

Spectral Data

Unsupervised spectral classification data, in the form of line printer maps, were obtained following a multispectral analysis of August 1976 Landsat digital tapes. The classification was performed at NASA-Ames Research Center, Moffett Field, California as part of a study with the California Department of Forestry during 1978-1979 (Peterson, et al. 1979). The portion of the unsupervised classification performed in the study area consisted of 37 spectral classes (of unknown ground conditions) based on spectral signatures and topographic data. These 37 classes were grouped into 16 land use classes (conifer, hardwood, barren, water, etc.) on the basis of comparisons with field conditions, photo-interpretation and characteristics of their spectral signatures by CDF personnel. The classes were then labeled to reflect these land use categories. False color images and line printer maps were made of both the spectral classes and the land use class information by NASA-Ames.

The area of study, the southwestern portion of the Warner Mountain Ranger District (Fig. 3), was chosen because of its accessibility, familiarity, known reliability and the availability of the stand examination and Landsat data to the investigator. Both the spectral and the land use class data for the Alturas east 1⁰ by 1⁰ quadrangle were obtained in the form of alphanumeric 1:24,000 line printer maps. By comparing these maps (Fig. 4), the CDF land use class of "conifer" was seen, on the spectral class map, as being represented by spectral

classes given the names 4, 6, 7, 8, W and X during the initial classification. At this point nothing was known about these six classes other than their spectral signatures, their location on the spectral class line printer map and that CDF personnel had interpreted all of them as containing conifers.

Field Data

The Timber Stand Improvement (TSI) stands and Compartment Inventory Analysis (CIA) plots in the south Warner Mountains were used as the source of the ground data. TSI data were collected to determine the need for thinnings so the inventory data collection centered on obtaining general stand characteristics with basal area per acre (BA), average diameter breast height (DBH) and stems per acre being measured closely. Stand composition, average height and age were also measured. The CIA plots were used for timber sale planning and individual tree heights, DBHs and ages were all determined. Averages were taken of these values for each plot to make them compatible with the TSI stand data. Stand composition, BA and stems per acre were extrapolated from the data given for the five points (variable plot cruising) that made up each CIA plot. Percentages of pine and fir were determined by counting individual trees, by species, for the entire plot and dividing by the total number of trees measured in the plot. BA for the five points were averaged to get a value for the plot as a whole. Stems per acre was obtained by dividing the plot BA by the basal area of the average sized tree for the plot. The combined TSI stands and CIA plots will be referred to simply as plots from here forward.

TSI and CIA data were used for the following reasons:

1. the U.S.D.A. Forest Service has high standards for stand examinations with plots being frequently checked for accuracy;
2. the plots had been randomly located throughout the study area;
3. the plots covered many remote areas where data collection would have been limited without the use of a fourwheel drive vehicle and excessive amounts of time;
4. there was a personal involvement in and an understanding of the data collection systems by the investigator;
5. the original data sheets used when the data were collected were available;
6. the values for percentage fir, percentage yellow pine, BA, height, DBH, stems per acre and age could be derived from the plots; and
7. the plots had been accurately marked on 1:15,840 color aerial photographs.

The TSI stands and CIA plots were located on the spectral class line printer map by making a visual comparison of the map with nine by nine inch color aerial photographs (1:15,840) taken in August 1974. Each plot was located by its position relative to stand openings, bodies of water and/or unusually homogeneous stands of timber that could be identified on both the photo and the map. The CIA plot centers were located and the TSI stands outlined on the photographs. Manual color coding of the conifer, shrub, brush, grass, barren and water classes on the map aided the visual comparison. The plots, once located, were marked on the line printer map. This was done with relative ease because of the excellent separation of forest and non-forest in the classification, along with the abundance of unusually

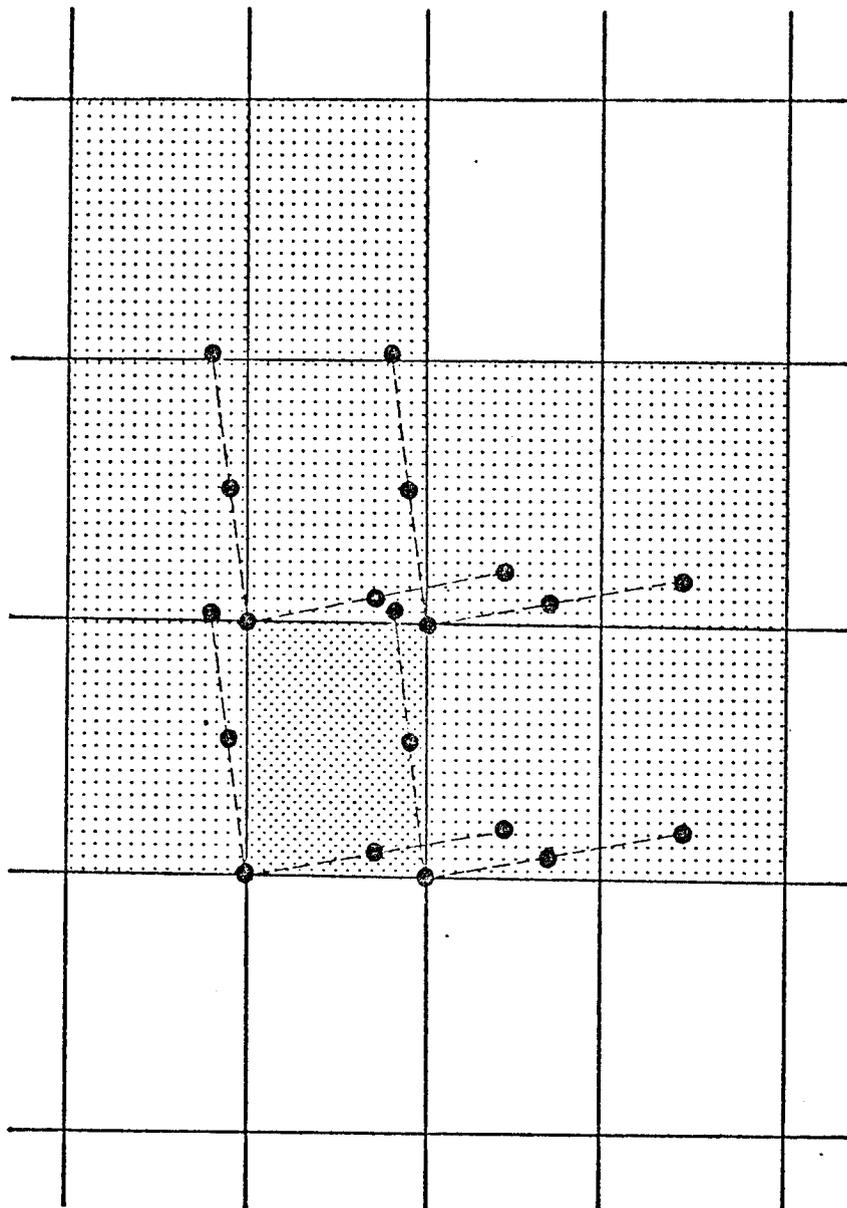
shaped openings (Fig. 4) and bodies of water found in the area. Once a plot was located and marked on the map the pixel class, line printer map coordinates and the plot information were recorded. Plot information consisted of percentage white fir, percentage yellow pine, BA, average height, DBH, stems per acre and age.

Because of the shape and dimensions of the CIA plots, they did not ever fall completely within any one pixel (Fig. 5). Therefore, only those CIA plots that fell within the illustrated "block" of pixels of the same class were recorded. Likewise, TSI stands were only recorded if it was possible to locate them on the line printer map as a single class. Where time and location permitted, large blocks of homogeneous pixels were sampled using the TSI format to provide additional data. Seventy-six plots were recorded using TSI and CIA plot information.

General Class Descriptions

The means of the variables from the plot information, as well as spectral signatures, were used in writing general descriptions of each of the classes. The mean values were useful in making general statements about the stocking, maturity, tree size and species mix of each of the classes. The spectral signatures were utilized to indicate shadow and additional vegetative characteristics.

The average reflectance, for each of the classes, in the colors green (0.5 to 0.6 micrometers), red (0.6 to 0.7 micrometers) and in near (0.7 to 0.8 micrometers) and far (0.8 to 1.1 micrometers) reflected infrared were obtained from NASA-Ames with the line printer maps. These four values for reflectance, digital numbers (dn), are collectively known as the spectral signature of the class. Each class has a unique



Full CIA plot; composed of five points, 2 and 4 chains north and west of a central point.



Pixel identified as plot center from aerial photograph



Pixels possibly affected by points of the CIA plot

Figure 5. Sphere of Influence of CIA plots. The Plot Center of a CIA Plot Could Occur at Any Place Within the Center Pixel. Plots Located at All Four Corners of This Center Pixel Show Which Pixels Could Have Had Plot Information Taken Within Them.

signature (Fig. 6). However, groups of similar classes (i.e. the "conifer" classes) generally have spectral signatures that follow similar patterns and have similar dn ranges. Differences may be caused by many factors. Dense vegetation can result in a high dn in the green band. Soil and dead vegetation can cause high red reflectances. The infrared bands show lower dns if shaded or if the vegetation has low vigor.

Discriminant Analysis

Using the variables for each of the plots paired with the spectral classes it was hoped some pattern might become apparent that would define the six spectral classes. To determine which combination of some or all of these variables defined each of the classes and how well they did so, a discriminant analysis was performed (Klecka, 1975). The variables from the 76 plots were stored in the local CYBER system at Humboldt State University to be used in conjunction with the SPSS (Statistical Package for the Social Sciences) program. Statistics generated from this package included; means, standard deviations, pooled within-groups covariance matrix, matrix of pairwise F ratios, univariate F ratios, test for equality of group covariance matrices and structure correlations.

The most important statistics generated by this program were the matrix of pairwise F ratios and the univariate F ratios. The univariate F ratios were used to determine which of the variables was most significant in separating the classes. This most significant variable was used to assign each plot to a spectral class by comparing the plot value to each of the class means. Assignment was made to the

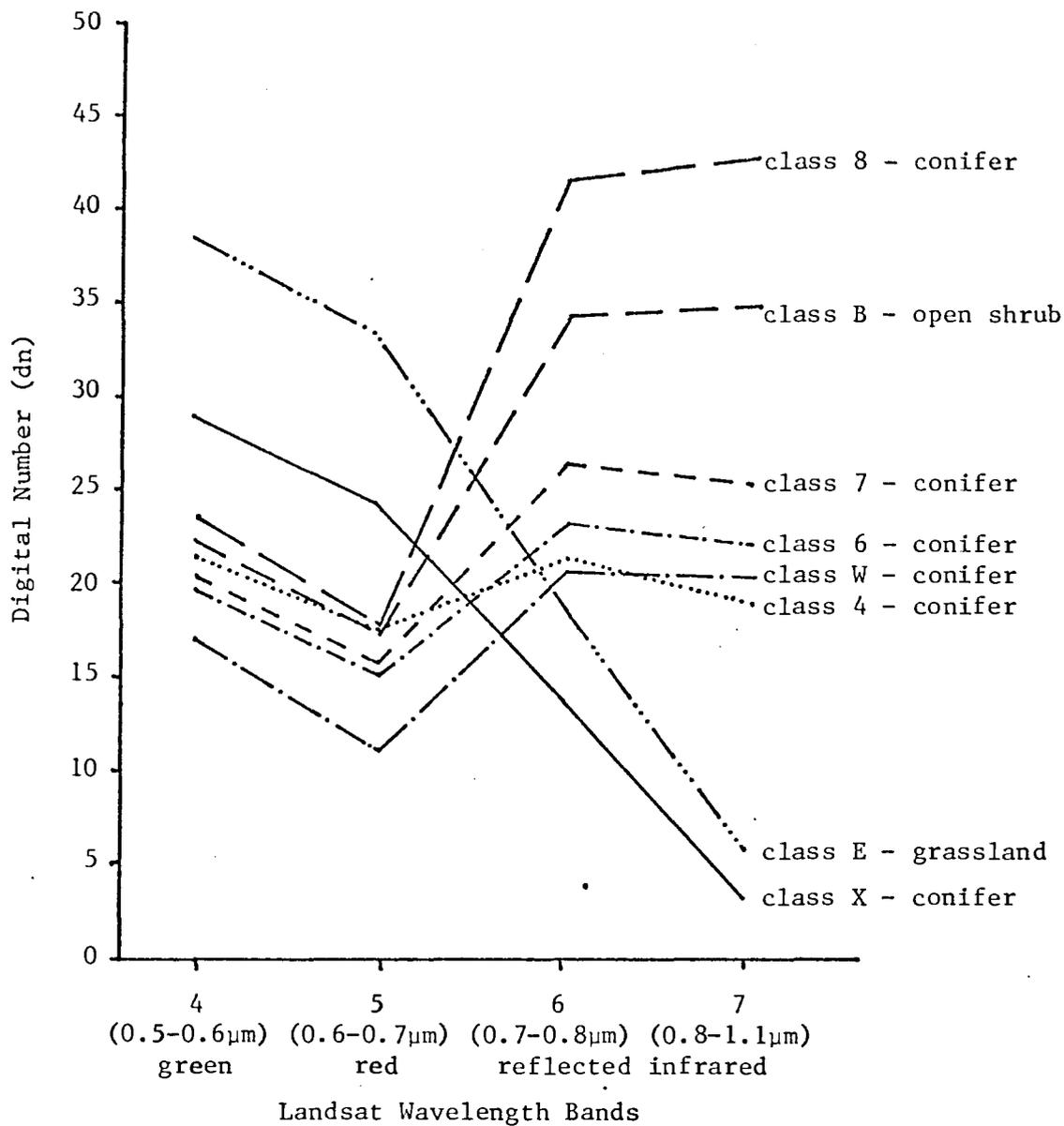


Figure 6. Comparison of the Spectral Signatures of the Six Conifer Classes with One Open Shrub and One Grassland Class.

class having the least difference. Using the six variables remaining, new F ratios were generated and a second significant variable was chosen. The two significant variables were then used to again assign the plots to classes. Pairwise F statistics for both of these assignments were used to indicate if improvement occurred by using the additional variable in the assignments. Variables continued to be added in this fashion until no improvement in the pairwise F statistics occurred, indicating that no improvement in separation could be gained by using additional variables. Using those variables that best separated the classes, each plot was placed in the class that was best defined by the plot's variables (most closely matched the means of the significant variables in that class). This was done without regard to the class the plot was associated with in the raw data. Tables were printed indicating the number and percentage of the plots that were defined by the variables as being in the class they were associated with in the raw data as well as the numbers and percentages of the plots that were placed in each of the other classes (Klecka, 1975). Separate runs were made using different combinations of classes (by eliminating those classes that were easily separable from one another) and data sources (by dividing the full data set into its TSI and CIA components). The results, in the form of tables from the various analyses, were compared and improvements or reductions in correct assignments noted. The means and standard deviations of the significant variables were examined to attempt to find explanations for the improvements or reductions.

RESULTS

The CDF "conifer" class was found to be composed of six classes on the spectral line printer map of the south Warners. These six classes were labeled 4, 6, 7, 8, W and X. Of these six classes, two were represented by three or fewer plots each and therefore could not be fully described or used in the statistical analysis. The remaining four classes were described based on the means of variables collected in the TSI and CIA plots as well as spectral signatures. They were then analyzed using discriminant analysis. From this, two classes were found to be easily separable from one another. The other two classes required analysis of their component parts to improve separation. The TSI and CIA data were identified as the source of statistical confusion in assigning the plots to classes.

Class Descriptions

Of the six classes, it was found that classes 6 and 8 were rarely associated with TSI stands or CIA plots. This was partly due to the relatively low abundance of the classes 6 and 8 on the line printer map. They were both commonly found on the boundaries between forest and grass, shrub, brush or barren areas (Fig. 4). One of these classes, class 8, represented pixels that were actually half of a pixel of forest and half a pixel of some other land use type. Since the Landsat data cannot assign half pixels, some classes could be composed of pixels that showed the combined spectral characteristics of two or more classes (Tucker and Miller, 1977). In this case the CDF personnel

labeling the land use classes identified class 8 with conifer cover more often than any of the other classes although it spectrally resembles the open shrub class. In Figure 6, class B of the open shrub class is shown. This open shrub class has a spectral signature very much like that of class 8 in general shape. No two signatures, for different classes, can be exactly the same. If they were the same they would have to represent the same class. The difference between these two signatures in the infrared (maximum separation of about 8 dn) is not of major concern when considering that the reflectance values can range from 0 to 127. Therefore class 8 can be defined as a class labeled by CDF as containing conifers but which has a spectral signature that resembles an open shrub class. This class is generally found between "conifer" pixels and pixels of non-conifer land use classes. Class 6, on the other hand, has a signature resembling the conifer classes 7 and W. Its signature lies between the signatures for these classes (Fig. 6). Class 6 is slightly "brighter" than class W but not as "bright" as class 7. What this might mean in ground conditions, beyond concluding that the class is in fact composed of conifers, cannot be said without sufficient ground data, which were not, in this case, available.

As explained above, classes 6 and 8 could not be fully described as to general stand conditions. This leaves classes 4, 7, W and X. Generalities about each of these classes can be made from either the means of the stand variables or the spectral signatures. Taking the classes in the order 4, 7, W then X some outstanding features may be noted using the means (Table 1). Class 4 has by far the greatest number of stems per acre (811.1) and the trees are the shortest (48.7 feet) of any of the classes. The outstanding features in class 7 are

Table 1. Means and Standard Deviations (in parenthesis) of Variables Collected in TSI Stands and CIA Plots for Spectral Classes 4, 7, W and X.

Spectral Class	Variables from TSI and CIA data						
	percent fir	percent pine	BA (sq.ft.)	height (feet)	DBH (inches)	stems/acre	age (years)
4	87.8 (13.9)	11.1 (14.5)	267.3 (124.8)	48.7 (18.9)	11.0 (3.7)	811.1 (348.9)	95.6 (39.2)
7	38.1 (50.0)	61.9 (50.0)	197.8 (73.8)	60.4 (15.4)	11.5 (3.4)	384.0 (256.3)	83.1 (19.8)
W	83.8 (26.7)	16.2 (26.7)	258.1 (58.5)	72.6 (18.7)	17.2 (6.0)	354.0 (311.2)	139.2 (57.4)
X	65.7 (39.3)	34.3 (39.3)	195.1 (56.7)	64.5 (13.9)	13.9 (4.3)	355.4 (332.1)	104.6 (38.0)

that the trees are the youngest (83.1 years) and that the percentage pine (61.9 percent) is the highest of the classes. Class W has trees that have average heights (72.6 feet), diameters (17.2 inches) and ages (139.2 years) that are well above the rest. Class X is quite a bit like class 7 except that the percentage pine is much less (34.3 percent) and the trees are older (104.6 years).

These quick descriptions can be expanded by using all of the variables. Classes 4 and W are the easiest to describe in this fashion because they are the most distinctive. Classes 7 and X are similar but can also be described separately.

Class 4, with the greatest number of stems per acre and the shortest trees, is easily described as an overstocked stand. The other variables tell that the stands are mostly fir (87.8 percent), have high BAs (267.3) and are composed of small stems (11.0 inches) which are fairly young in age (95.6 years). The complete description from this information would be: young, overstocked white fir stands that are stunted in growth.

Class W, besides having the largest and the oldest trees, is mostly fir (83.8 percent) with a high BA (258.1). This BA is similar to that of class 4 but there are fewer trees per acre (354.0). This is to be expected, because a given BA can be made up of many small trees or fewer large ones, as in this case. This class is composed of mature, well stocked fir stands.

Class 7, the young mixed stands with much pine, has a lower BA (197.8) than the two previous classes. The trees are small (11.5 inches) but fewer to the acre (384.0) than class 4. These stands are adequately stocked, not overstocked as in class 4. The height (60.4

feet) of the trees is intermediate between classes 4 and W. This could be due to the trees having better spacing than class 4 (overstocking can cause poor growth) especially since they are younger (83.1 years) than the trees in class 4. They would be expected to be shorter than class W because they are both younger and smaller (DBH) than the trees in that class. Class 7 is therefore described as a young, moderately stocked, mixed forest tending toward pine.

Class X, at a glance, closely resembles class 7. The BA (195.1) is nearly the same, height (64.5 feet) and DBH (13.9 inches) are only slightly larger and stems per acre (355.4) is only about eight percent less. This class is only significantly different in composition and age. But the other variables do give some useful information. The class is composed of, in general, moderately aged, almost understocked, fir stands.

From the spectral signatures some additional observations can be made. Class 4 has a similar reflectance value in each of the four bands (Fig. 6). Classes 7 and W have parallel signatures with the dn (digital numbers) making up class 7's signature being above that of class W by about five (Fig. 6). The signature for class X indicates that there is something definitely different about this class. The signature is higher in the red and green while being lower in the infrared bands than any of the other conifer classes (Fig. 6).

Class 4, the young overstocked fir class, has a signature that is similar to class 7 in the green and red reflectance values. The dn for bands 6 and 7 are both within one or two of class W. The reasons for the similarities may be different but all four values fall within the realm of the other conifer signatures. In this case the low

infrared part of the signature is most likely due to these overstocked stands (viewed in August) having poor vigor and low leaf water content (detected by band 7). To class 4's description might be added that the trees were probably under stress at the time of the study.

Class 7—the young, moderately well stocked mixed forest tending toward pine—and class W—the mature, well stocked fir class—have parallel spectral signatures. Class 7 has the "brighter" reflectance. This could be due to class 7 being younger or to the fact that class 7 is composed of more pine than class W. Age, because it affects the height and shape of trees, is most likely the cause of the difference. There would be more shadow within the tall, older trees of class W and shadow tends to reduce the reflectance in all four bands. Nothing new (for the descriptions) was seen in the spectral signatures of these two conifer classes.

Classes X is the real puzzle in the spectral signatures. It bears no resemblance to the other conifer signatures. The high value for the red reflectance indicates a source other than living vegetation. When the signature for class X is compared with that for one of the grassland classes (class E, Fig. 6) a nearly parallel trend between the two signatures can be seen. The low values for the infrared bands in the grasslands class are most likely due to the time of year that the imagery was taken. In August most meadow vegetation is dried and brown (high red reflectance). Dried vegetation has a low living biomass (band 6 is sensitive to biomass) and has low leaf water content (band 7 is sensitive to this). It seems entirely possible that meadow vegetation is present beneath the trees in class X. Because the BA is so low the trees do not fully cover the surface and much of this dried meadow

vegetation is represented in the reflectance values. So the class X description should read: moderately aged, understocked fir stands with a prominent understory of meadow vegetation.

Discriminant Analysis

Classes 4, 7, W and X were abundant on the line printer map and adequate numbers for analysis of each class were represented by CIA and TSI plots. Classes W and X dominated in both overall representation on the line printer map and in sampled plots. Class 4 was represented by nine plots, class 7 by eight, class W by 25 and class X by 34. The results of the first discriminant analysis, using these four classes, are shown in Table 2. Five of the seven variables; BA, DBH, stems per acre, percentage pine and height respectively, were used in this analysis. Percentage fir and age did not enter the computation before it terminated.

Table 2 shows the percentages of correct and misassigned classes. No one class had outstanding accuracy. There was one consistent feature; class 4 was not assigned to class 7 or vice versa. These two classes, it appears, were separated well enough that they were not confused with each other within the data set available. But significant misassignments were made to classes W and X (44.4 percent from class 4 and 37.5 percent from class 7). Class W was misassigned 40 percent of the time, half of this was to class X, the other half split unevenly between classes 4 and 7 (class 4-twelve percent, class 7-eight percent). Class X had the poorest number of correct assignments, 41.2 percent, and therefore the greatest number of misassignments. An equal number of plots of class X were assigned to classes W and 7 (23.5 percent)

Table 2. Results of Discriminant Analysis, Using All Data (TSI, CIA and Other Plots), for Spectral Classes 4, 7, W and X. Values Represent Number of Plots Assigned to Each Class. Values in Parenthesis Are Percentages of the Total Number of Plots Actually Occurring in that Class.

CLASS		Classes are assigned using values from TSI, CIA and other plots			
		W	X	4	7
Classes as identified on line printer map	W	15 (60.0)	5 (20.0)	3 (12.0)	2 (8.0)
	X	8 (23.5)	14 (41.2)	4 (11.8)	8 (23.5)
	4	1 (11.1)	3 (33.3)	5 (55.6)	0 (0.0)
	7	1 (12.5)	2 (25.0)	0 (0.0)	5 (62.5)

while 11.8 percent of the class X plots were assigned to class 4.

Because classes 4 and 7 were not confused with each other but were often confused with classes W and X, a second analysis was done using only the latter two classes. The most significant improvement was in the correct assignments to class X, which rose from 41.2 to 70.6 percent (Tables 2 & 3). An improvement in both classes was expected because those plots from classes W and X formerly misassigned to classes 4 and 7 would have been assigned to either class W or X. The 20 percent of class W misassigned to classes 4 and 7 in the first analysis were limited to either class W or class X for assignment in the second. Of this 20 percent, eight were correctly placed in class W, 12 percent in class X. Class X was misassigned as classes 4 and 7 35.3 percent of the time in the first analysis. Of this, only 5.9 percent were misassigned to class W, the remaining 29.4 percent found the correct assignment in class X (Tables 3 & 4).

A notable difference between this analysis, with only two classes, and the previous one using four classes were the variables used in the discriminant analysis. As stated above, five variables - percent pine, BA, height, DBH and stems per acre - were needed in the analysis using the four classes. To separate just the two classes, W and X, required only the variables BA and DBH respectively. The second analysis program run terminated without entering any further variables.

As the number of classes increased so did the incidence of classes having one or more similar variables (Table 1). Looking just at the mean values for each of the variables, it was easy to see that classes W and X could be separated in all variables except for stems per acre. Examination of the standard deviations showed this separation

Table 3. Comparison of Discriminant Analyses Using Classes W and X Only.

A.) Using Combined TSI and CIA Data

B.) Using TSI Data Only

C.) Using CIA Data Only

Values Represent Numbers of Plots Assigned to Each Class.

Values in Parenthesis Are Percentages of the Total Number of Plots Belonging to That Class.

A.		Assigned by Analysis of TSI and CIA Data	
Classes		W	X
From Line Printer Map	W	17 (68.0)	8 (32.0)
	X	10 (29.4)	24 (70.6)
B.		Assigned by Analysis of TSI Data Only	
Classes		W	X
From Line Printer Map	W	7 (70.0)	3 (30.0)
	X	3 (15.0)	17 (85.0)
C.		Assigned by Analysis of CIA Data Only	
Classes		W	X
From Line Printer Map	W	12 (80.0)	3 (20.0)
	X	2 (14.2)	12 (85.7)

was not nearly as clear as it had first appeared. However, BA and height were fairly well separated and were therefore used in the second analysis. The first analysis, using five variables, was much more complicated. Mean values for many of the variables were very close to one another in two or more classes. For example:

1. percentage fir and pine and BA were similar in classes 4 and W;
2. height, BA and stems per acre were similar in classes 7 and X;
3. DBH and age were similar in classes 4 and 7;
4. stems per acre was similar in classes 7, W and X; and
5. age was similar in classes 4 and X.

In each of the above examples the variable(s) listed would be a poor choice to use in trying to separate the classes. The standard deviations (Table 1) show that even in the variables where the means have fair separation there may be considerable overlap.

Two more analyses of classes W and X were made to determine if the two different methods of collecting data (TSI and CIA) caused differences in the consistency of the correct assignments in the discriminant analysis. Class W showed only minor, from 68.0 percent to 70.0 percent, improvement using TSI data only (Table 3 A & B). The correct assignment of 80.0 percent in class W using CIA data alone was a more than 10 percent improvement (Table 3 A & C). Class X showed essentially the same incidence of correct assignments using either TSI or CIA data, 85.0 and 85.7 percent respectively, as opposed to 70.6 percent for the combined data.

By comparing the two data sources side by side (Table 4), analysis of the mean values showed that the TSI stands had higher basal area per acre but the trees were shorter and smaller (DBH) resulting in more

Table 4. Comparison of the Mean Values of the Variables Collected in the TSI Stands and in the CIA Plots for Spectral Classes W and X. Standard Deviations in Parenthesis.

data source	spectral class	Variables from data						
		percent fir	percent pine	BA (sq.ft.)	height (feet)	DBH (inches)	stems/acre	age (years)
TSI	Class W	74.5 (40.3)	25.5 (40.3)	296.0 (58.9)	69.0 (15.2)	14.0 (5.4)	625.0 (307.5)	119.5 (48.2)
	Class X	63.5 (42.4)	36.5 (42.4)	213.0 (50.1)	63.7 (14.6)	12.2 (4.0)	511.7 (354.9)	93.4 (25.6)
CIA	Class W	90.0 (8.5)	10.0 (8.5)	238.8 (43.8)	75.0 (20.9)	19.3 (5.7)	173.3 (133.5)	152.3 (60.8)
	Class X	68.9 (35.7)	31.1 (35.7)	169.4 (57.5)	65.7 (13.1)	16.4 (3.4)	132.1 (67.7)	120.7 (47.3)

stems per acre than in the CIA plots. Also, the trees were generally younger in the TSI stands. Because of this, the significant variables used in the discriminant analysis were different for TSI data, CIA data and the combined data. The TSI analysis used for variables; percentage pine, BA, height and age. Likewise the CIA analysis used four variables, but they were height, DBH, stems per acre and age.

DISCUSSION

Accurate placement of plots into classes along with detailed class descriptions would have made computer generation of type maps using limited ground truth data possible. This study of an unsupervised classification of Landsat data, completed by NASA and CDF, showed that only limited information useful for timber typing could be obtained. It was hoped that the descriptions of the classes (made using the classes generated by the unsupervised classification and information from timber stand examinations) would be roughly equal to those on current timber type maps. However, only limited information about stand characteristics was gained and plot-to-class assignments were not outstanding in accuracy. The six classes were described as follows:

1. class 4 stands were young, overstocked fir stands possibly under moisture stress at the time of the study;
2. class 6 was a conifer class generally occurring next to non-forest land use classes;
3. class 7 had young, moderately stocked mixed stands tending toward pine;
4. class 8 was a combination class, possibly containing conifers but having a spectral signature of an open shrub class;
5. class W had mature, well stocked fir stands; and
6. class X was a moderately aged, understocked fir class with a prominent understory of meadow vegetation.

The use of the mean values of stand characteristics of each class and spectral signatures figured heavily in generating these stand

descriptions. The variances of the stand variables in each of the classes ruled out any descriptions more specific than those given above. Also, the conditions causing a specific reflectance signature may possibly have had more to do with the vegetative vigor, stand openness, background reflectance, sun angle and/or some other factors not yet studied than the average size, species, age and number of trees per acre gathered as TSI and CIA data.

The discriminant analyses were performed to see if some combination of the variables, not apparent by just looking at the mean values, might be used to assign plots to classes accurately. Using all the data, the first analysis could not separate the plots into classes with greater than 62.5 percent accuracy. Two classes, 4 and 7, were separable from each other but not from W and X. This seems to indicate that if two components of a forest are different enough (class 4 was overstocked fir, class 7 moderately stocked with quite a bit of pine) Landsat data could be used to accurately put them into groups using unsupervised classification. For the other classes and other analyses the results indicated that unless the forest was measured uniformly (i.e. all TSI examinations or all CIA examinations) plots could not be assigned to classes with outstanding accuracy. The analyses using just the CIA or TSI data showed the highest accuracies, up to 85 percent, of plot assignments because the data were taken using the same techniques and measuring the same variables, without extrapolations.

In forest conditions as diverse as those on the Warner Mountains stands ranged from those just planted to oldgrowth. This entire range of stands was represented by only six spectral classes. Had the area

not been so diverse, or had there been a greater number of spectral classes representing conifers, it may have been possible to garner more specific information about stand character from the classes and plots.

This study found little use for unsupervised classifications for detailed timber management. Much more detail and more accurate assignments would be necessary to be compatible with timber type maps currently in use. It is possible that the general classes that were obtained may be of some use to some other discipline such as wildlife management or other non-timber oriented forest management. The excellent separation of forest and non-forest could also be used by those interested in acreages of forested and non-forested lands.

REFERENCES CITED

- Aldrich, R.C. 1979. Remote sensing of wildland resources: A state-of-the-art review. *General Technical Report RM-71*. Rocky Mountain Forest and Range Experiment Station. 56 pp.
- Beaubien, J. 1979. Forest type mapping from Landsat digital data. *Photogrammetric Engineering and Remote Sensing*. 45:1135-1144.
- Brown, W.S. 1945. *History of the Modoc National Forest*. Pp. 40-49.
- Fu, F.S. and T.S. Yu. 1980. *Statistical Pattern Classification Using Contextual Information*. Research Studies Press, New York, N.Y. 191 pp.
- Katti, R.K., M.G. Sardar, T.V. Pavate and P. Venkatachalam. 1981. Utilization of remote sensing in resources identification and land use in India - An integrated approach. *Proceedings Fifteenth International Symposium on Remote Sensing of Environment*. Ann Arbor, Mich. Pp. 1487-1497.
- Klecka, W.R. 1975. Discriminant Analysis. In N.H. Nie, C.H. Hull and J.G. Jenkins (Eds.), *Statistical Package for the Social Sciences*. McGraw-Hill Book Company, San Francisco, CA. Pp. 434-467.
- Küchler, A.W. 1977. *Natural Vegetation of California*. William and Heintz Map Corp., Washington, D.C. 20027.
- Lowe, D.S. 1979. Use of Landsat in computer data bases. In *Computer Mapping in Natural Resources and the Environment*. Harvard University. Pp. 69-76.
- Mead, R.A. and D.A. Rasberry. 1980. Current use of remote sensing by foresters in the south. *Southern Journal of Applied Forestry*. 4(3):143-147.
- Merchant, J.W. 1981. Utilization of spatial complexity in computer classified Landsat MSS data for multi-factoral thematic mapping. *Proceedings Fifteenth International Symposium on Remote Sensing of Environment*. Ann Arbor, Mich. Pp. 905-914.
- Peterson, D.L., N. Tosta-Miller, S. Norman, D. Wierman and W. Newland. 1979. Land-cover classification of California using mosaicking and high-speed processing. *Proceedings Thirteenth International Symposium on Remote Sensing of Environment*. Ann Arbor, Mich. Pp. 279-305.

- Robinove, C.J. 1979. Integrated terrain mapping with digital Landsat images in Queensland, Australia. *U.S.G.S. Professional Paper 1102*. Washington, D.C. 39 pp.
- Roller, N.E.G. and L. Visser. 1980. Accuracy of Landsat forest cover type mapping in the Lake States region of the U.S. *Proceedings Fourteenth International Symposium on Remote Sensing of Environment*. Ann Arbor, Mich. Pp. 1511-1520.
- Shimabukuro, Y.E., P.H. Filho, N.F. Koffler and S.C. Chen. 1980. Automatic classification of reforested pine and eucalyptus using Landsat data. *Photogrammetric Engineering and Remote Sensing*. 46:209-216.
- Taranik, J.V. 1981. Advanced aerospace remote sensing systems for global resource applications. *Proceedings Fifteenth International Symposium on Remote Sensing of Environment*. Ann Arbor, Mich. Pp. 2-30.
- Tucker, C.J. and L.D. Miller. 1977. Soil spectra contributions to grass canopy spectral reflectance. *Photogrammetric Engineering and Remote Sensing*. 43(6):721-726.
- U.S.F.S. 1975. *Modoc Timber Management Plan of 8/14/75*. Alturas, Calif. P. 12.
- U.S.G.S. 1974. *Soil Survey, Surprise Valley-Home Camp Area, Calif.-Nevada*. Pp. 124-128.
- U.S.G.S. 1980. *Soil Survey of Modoc County, Calif., Alturas Area*. Pp. 148-149.
- Walsh, S.J. 1980. Coniferous tree species mapping using Landsat data. *Remote Sensing Environment*. 9:11-26.

APPENDIX A. Raw Data from TSI, CIA and Other Plots.

Class	% fir	% pine	BA	height (feet)	DBH (inches)	stems/acre	age	(photo) TSI / CIA compartment	plot
4	100	0	500	65	12	950	125	TSI ¹	299
4	70	30	120	85	18	750	175	TSI	271
4	95	5	260	50	14	450	85	TSI	256
4	60	40	220	43	13	350	115	TSI	335
4	95	5	280	35	9	1150	75	TSI	334
4	100	0	120	25	7	900	45	TSI	338
4	95	5	420	60	10	900	105	TSI	249(2) ⁴
4	90	0	240	30	6	1400	65	TSI ₂	250(2)
4	85	15	246	45	10	450	70	Mile ²	12
7	95	5	126	40	11	190	65	Refuge	4
7	100	0	180	73	11	250	60	TSI	284
7	0	100	56	90	18	32	120	Jess	2
7	0	100	280	50	8	800	90	(47 14 201) ³	
7	0	100	240	50	8	700	80	(47 14 201)	
7	0	100	220	60	13	400	80	(47 14 201)	
7	10	90	240	60	9	400	70	(48 14 100)	
7	100	0	240	60	14	300	100	(50 17 210)	
W	100	0	340	45	7	1350	85	TSI	298
W	100	0	320	60	10	850	100	TSI	292
W	70	30	240	60	13	350	90	TSI	345
W	100	0	216	55	8	600	125	Refuge	1
W	100	0	204	45	13	250	80	Refuge	12
W	90	10	152	45	12	200	75	Refuge	17
W	80	20	198	70	18	150	115	Refuge	18
W	100	0	248	110	26	100	225	Refuge	21
W	80	20	216	70	12	300	105	Refuge	23
W	100	0	248	100	26	100	250	Mahogany	44
W	100	0	230	85	21	100	105	Mile	5
W	85	15	256	80	26	100	205	Mile	9
W	95	5	204	100	21	100	175	Mile	10
W	80	20	198	85	21	100	215	Mile	13
W	85	15	234	75	23	100	170	Mile	17
W	90	10	288	90	23	100	200	Mile	18
W	80	20	336	70	21	150	180	Mile	19
W	85	15	264	45	18	150	60	Mile	20
W	90	10	320	85	24	500	200	(48 14 97)	
W	100	0	360	65	10	800	100	(47 14 201)	
W	95	5	360	90	18	400	180	(48 14 100)	
W	90	10	330	90	20	600	100	(48 14 100)	
W	100	0	270	70	12	400	100	(50 17 210)	
W	0	100	200	70	16	400	180	(50 17 210)	
W	0	100	220	55	10	600	60	(50 17 210)	
X	0	100	240	85	15	300	93	TSI	206
X	0	100	240	85	15	300	93	TSI	207
X	100	0	240	70	15	550	100	TSI	257
X	100	0	140	75	26	100	150	TSI	258(2)
X	5	95	60	48	11	200	58	TSI	282

APPENDIX A. Raw Data from TSI, CIA and Other Plots (continued). 41

Class	% fir	% pine	BA	height (feet)	DBH (inches)	stems/acre	age	(photo)	plot
								TSI / CIA compartment	
X	100	0	220	45	7	600	60	TSI	291
X	65	35	240	73	11	500	58	TSI	294(1)
X	15	85	220	45	7	1250	55	TSI	294(1)
X	90	10	180	48	10	1450	70	TSI	295
X	95	5	260	55	6	900	100	TSI	303
X	50	50	220	55	16	750	100	TSI	337
X	100	0	260	55	14	450	85	TSI	340
X	70	30	230	50	16	165	90	Mile	14
X	85	15	210	60	24	100	230	Mile	16
X	65	35	174	75	16	125	130	Refuge	5
X	50	50	132	65	18	75	85	Refuge	6
X	95	5	296	50	16	148	115	Refuge	11
X	95	5	168	65	21	71	190	Refuge	14
X	0	100	114	65	12	145	70	Refuge	15
X	95	5	254	55	15	207	80	Refuge	19
X	100	0	116	55	17	74	90	Refuge	20
X	30	70	108	95	16	78	70	Refuge	22
X	0	100	60	80	18	34	95	Mahogany	31
X	95	5	160	70	10	293	160	Mahogany	41
X	85	15	200	55	15	162	130	Mahogany	42
X	100	0	240	80	16	173	145	Mahogany	45
X	80	20	200	80	21	84	140	Mahogany	56
X	90	10	280	80	12	600	110	(48 14 94)	
X	95	5	240	75	12	300	106	(48 14 94)	
X	90	10	200	80	10	400	80	(47 14 197)	
X	0	100	150	50	18	200	110	(47 14 197)	
X	0	100	220	45	8	300	90	(47 14 197)	
X	100	0	210	60	9	500	110	(49 14 46)	
X	95	5	240	65	11	500	100	(49 14 46)	

- 1 TSI in the compartment column indicates that the data is from a TSI stand
- 2 Jess, Mahogany, Mile or Refuge are compartment names for CIA plots
- 3 Numbers in parentheses are flight line and photo numbers for plots taken independent of TSI or CIA plots
- 4 Numbers in parentheses in the plot column are aggregates of TSI stands

APPENDIX C. Example of TSI Data Sheet.

(1:25,000 map
with TSI stand
outlined)

AGGREGATION DESCRIPTIONS	AG. NO.	STAND	LAYER	%	SPECIES COMPOSITION								B.A./A. (11.2)	HT. (ft.)	D.B.H. (in.)	STEMS /ACRE	AGE	TREATMENT/REMARKS			
					Cover	Symbol %		Symbol %		Symbol %		Basis						SITE TREE HT.	AGE	RINGS/IN.	SPECIES
						Symbol	%	Symbol	%	Symbol	%										
EXISTING			1																		
			2																		
			3																		
			1																		
			2																		
			3																		
PLANNED			1																		
			2																		
			3																		
			1																		
			2																		
			3																		