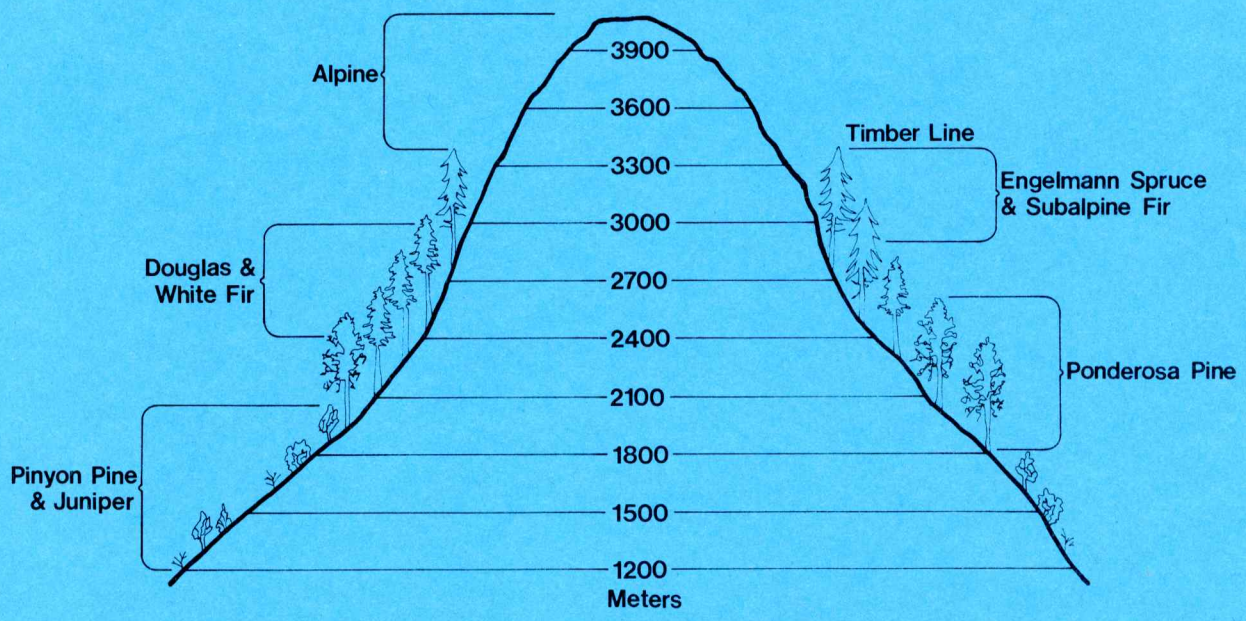
LARS Technical Report 011579

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

R.M. Hoffer
M. D. Fleming
L. A. Bartolucci
S. M. Davis
R. F. Nelson



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

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

R.M. Hoffer
M.D. Fleming
L.A. Bartolucci
S.M. Davis
R.F. Nelson

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

LARS Technical Report 011579

DIGITAL PROCESSING OF LANDSAT MSS AND TOPOGRAPHIC DATA TO IMPROVE CAPABILITIES FOR COMPUTERIZED MAPPING OF FOREST COVER TYPES

by

R.M. Hoffer, M.D. Fleming, L.A. Bartolucci, S.M. Davis and R.F. Nelson

Laboratory for Applications of Remote Sensing
in cooperation with
Department of Forestry and Natural Resources
Purdue University
West Lafayette, Indiana 47907

Annual Report on NASA Contract NAS 9-15508
Submitted to D.L. Ausbury, Technical Monitor
NASA, JSC Houston, Texas

Star Information Form

1. Report No. LARS Technical Report 011579	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle: Digital Processing of Landsat MSS and Topographic Data To Improve Capabilities for Computerized Mapping of Forest Cover Types		5. Report Date: January 15, 1979 6. Performing Organization Code
7. Author(s): R.M. Hoffer, M. D. Fleming, L.A. Bartolucci S.M. Davis, and R.F. Nelson		8. Performing Organization Report No. 011579
9. Performing Organization Name and Address: Laboratory for Applications of Remote Sensing/Purdue University 1220 Potter Drive West Lafayette, IN 47906		10. Work Unit No. 11. Contract or Grant No. NAS 9-15508
12. Sponsoring Agency Name and Address: NASA/Johnson Space Center Houston, TX 77058		13. Type of Report and Period Covered: Annual Report 12/16/78-1/15/79 14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This study involved the development and evaluation of two techniques for using a digital data base of Landsat MSS and digital topographic data to increase the accuracy of mapping forest cover types. The first approach, involved the development of a topographic distribution model, a quantitative description of the distribution of each of the forest cover types as a function of elevation, aspect, and slope, thereby providing the topographic training statistics. Spectral training statistics were developed using two different procedures and the combined spectral/topographic data set was then classified using several methods. The use of topographic data in combination with the spectral data significantly improved the overall classification accuracy of forest cover types. The elevation data alone improved the classification accuracy by over 15%.

The second approach, a reflectance geometry correction model, involved the alteration of the reflectance values of the Landsat scanner data to remove spectral variation due to topography. The reflectance geometry correction model removed a great deal of the spectral variations due to topography. However, classification accuracy estimates did not indicate any improvement due to the "correction" model.

17. Key Words (Suggested by Author(s)): Remote Sensing, Digital Terrain Model, Digital processing, Forest type mapping, Spectral Analysis, Ancillary topographic data, Reflectance geometry, Landsat	18. Distribution Statement		
19. Security Classif. (of this report): Unclassified	20. Security Classif. (of this page): Unclassified	21. No. of Pages	22. Price*

*For sale by the National Technical Information Service, Springfield, Virginia 22161

NASA - JSC

Acknowledgements

In addition to the people directly involved in the research and the preparation of this report, a number of other individuals made significant contributions, which we gratefully acknowledge: Paul Anuta, for his assistance in geometric correction and overlay of Landsat and topographic data sets and the description of the processes involved in the preparation of these data; Dr. Virgil Anderson and Dr. K. S. Pillai for their assistance in the statistical design of the experiment; Mr. Louis Lang, for long hours spent in revision of computer programs to meet the requirements of this investigation; Mrs. Jeanne Etheridge, for her invaluable assistance in programming activities; and to many other colleagues at LARS and elsewhere who have offered helpful suggestions, advice and criticisms. We also wish to thank Mr. Hank Bond, U.S. Forest Service, Durango, Colorado, for his help during the field work activities and for help in defining the U.S. Forest Service information needs. Finally, we wish to thank Mr. David Amsbury, the NASA Technical Monitor for this investigation, for his help and suggestions throughout the project.

DIGITAL PROCESSING OF LANDSAT MSS AND TOPOGRAPHIC DATA TO IMPROVE CAPABILITIES FOR COMPUTERIZED MAPPING OF FOREST COVER TYPES

Table of Contents

List of Figures

Section III (continued) Page

List of Tables

Section V (continued) Page

I. DESCRIPTION OF STUDY

A. Introductory Statement

In forestry, as in many other discipline areas involving land management, there exists a distinct need for timely, reliable information concerning the resource base with which one is working. 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, as is the case in management of the world's forest resources. The launch of Landsat-1 in 1972 initiated a new era for land managers by proving that high-quality data can be obtained from satellite altitudes at reasonably frequent intervals for nearly any portion of the earth's surface. However, the ability to collect data from satellite altitudes far surpasses existing capabilities to analyze and interpret the data in a timely, reliable manner. As the demand and potential for more effective utilization of Landsat data have developed, many questions have been raised concerning the accuracy, reliability and limitations of various analysis techniques to extract pertinent information from the masses of satellite data.

Many studies have been conducted at Purdue University/LARS and elsewhere using Landsat and Skylab multispectral scanner data and various computer-aided analysis techniques; these studies have clearly shown the value of this combination of numerical data and quantitative analysis techniques. Several of these studies were directed at mapping forest cover types, but they involved study sites where topographic relief is minimal. Even a cursory examination of small-scale aerial photos or Landsat imagery indicates that slope and aspect have considerable influence on the spectral reflectance characteristics of forest cover. Furthermore since much of the forest land in the U.S. and elsewhere in the world is in areas of significant topographic relief, it is important that research not be confined to areas where topographic relief is minimal.

It has been, and continues to be, our belief that 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 the most effective analysis techniques and to determine the level of detail

and the reliability of information that can be obtained with such techniques. It is toward these goals that the current project is directed.

B. Background and Rationale

In 1974, the U.S. Congress mandated the U.S. Forest Service to inventory, every ten years, the extent and condition of all forest and rangeland resources throughout the United States (Renewable Resources Act of 1974). NASA and the U.S. Forest Service are both keenly interested in the potential application of remote sensing technology for meeting the requirements of this Act. The development of such techniques will in turn enable resource management personnel and agencies (such as the U.S. Forest Service) to obtain accurate and reliable forest cover type maps that are vital for effective resource management.

Prior to this project, a series of investigations had been conducted at the Laboratory for Applications of Remote Sensing (LARS), Purdue University, which indicated many of the capabilities and limitations of various analysis techniques for classifying and mapping forest cover in regions of significant topographic relief:

First, Landsat-1 investigation had been conducted in the San Juan Mountains of southwestern Colorado--an area of rugged mountain terrain and complex vegetative cover types. The results of this investigation (Hoffer, 1975a) indicated that deciduous and coniferous cover, as well as other major cover types, could be classified and mapped with a reasonably high degree of accuracy (80-85%); the classification and mapping accuracies for individual forest cover types, however, were much lower. Detailed statistical analyses of spectral responses led to the conviction that if satisfactory accuracies were to be obtained for individual forest cover types, there would need to be developed analysis techniques which account for topographic variability of spectral response.

Second, an investigation using Skylab data had been carried out in the same general portion of the San Juan Mountains used for the Landsat-1 investigation (Hoffer, 1975b). One phase of the Skylab investigation had involved the development of a digital overlay procedure to geometrically correct Landsat and Skylab data and

overlay them with digital elevation data on a single data tape. Analyses of the combined topographic data and spectral satellite data with conventional analysis techniques indicated that utilizing data vectors that included both spectral and topographic data in a standard, maximum-likelihood classifier would not consistently increase classification accuracy.

Third, during a Landsat-2 investigation carried out in conjunction with the Institute of Artic and Alpine Research (INSTAAR), University of Colorado, and the U.S. Forest Service, Region 2, a series of cover type classifications had been generated for large portions of the San Juan, Rio Grande, and Carson National Forests (Krebs et al., 1976). In addition, a set of software had been especially designed and developed to combine spectral Landsat classification maps with digital topographic data to create products useful in various management decisions. This software allows the generation of products on a 7½-minute quadrangle-by-quadrangle basis in formats suitable for meeting the specific requests and needs of Forest Service personnel. The major limitation of these combined cover-type/topographic-parameter maps was the level of cover type detail which could be accurately and reliably classified. It was concluded that variation in spectral response due to topography and forest stand density significantly reduced the capability to reliably classify individual forest cover types when using Landsat spectral data alone.

Based upon the results of these Landsat and Skylab investigations and the associated field work, it was clear that the occurrence of different forest cover types was significantly influenced by elevation and aspect and furthermore, that the aspect, slope, and stand density all have a significant influence on the spectral response of the various forest cover types. We concluded, however, that the influences of topography on species composition could be quantified, and that computer-aided analysis techniques could be used to combine the topographic data with the Landsat spectral data in the classification procedure, to provide more accurate, reliable forest cover type maps for forest management purposes. This led to the development of a proposal to NASA which resulted in the funding of the current project. This report summarizes the results of the first year's activities and findings.

During the second year, the research will involve refinement and definition of a recommended analysis technique, testing this technique on a data set from a totally different geographic location, and preparing a final report.

C. Objective

The objective of this research is 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.

D. Approach

The first year of this project has been devoted to the development and evaluation of different techniques for using a digital data base of Landsat MSS and topographic data to increase the accuracy of mapping forest cover types. Two different 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 involves 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 provides a quantitative probability of occurrence for each species for all topographic locations. Statistical characterization of the topographic distribution of the various cover types can then be combined with the statistical data describing the spectral characteristics of the various cover types. This provides the basis for the training data necessary for the classification of the combined spectral/topographic data set. 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 different techniques were used to develop the training statistics and two different procedures were used in the classification step. Additional variations on these basic classification procedures were also tested. The various combinations of different training and classification procedures resulted in a set of twelve classifications being obtained, compared, and evaluated. The results are discussed in detail in Section III.

The reflectance geometry correction model involves the "correction" 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 positions of the sun, the ground, and the satellite was used to "correct" the spectral data by calculating correction coefficients to remove the effects of the topographic position of 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. This approach is discussed in detail in Section IV.

The project was designed to generate several products that will have significance in the future development and use of computer-aided analysis techniques for forest inventory. The most significant of these are:

a. a topographic distribution model that quantitatively defines the relationship between the occurrence of forest cover types in the study area and their topographic position (elevation, aspect, and slope);

b. documentation of a tested technique for computer-aided analysis of Landsat data that uses topographic data to improve classification accuracy and reliability.

II. STUDY AREA DESCRIPTION AND DATA CHARACTERISTICS

A. Study Area Description

This first phase of the two-year study involved fourteen 7½-minute U.S.G.S. quadrangles within the San Juan Mountain study area, an area of approximately 34 x 43 miles in the center of the rugged San Juan Mountains of southwestern Colorado (Figure 1). 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 area 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) at the town of Pagosa Springs 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 study 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 or avalanches. 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.

The San Juan Mountain area has long been grazed by both cattle and sheep. Because cattle have a tendency to feed on certain palatable grass species, over-grazing of the area removes these grasses and encourages the

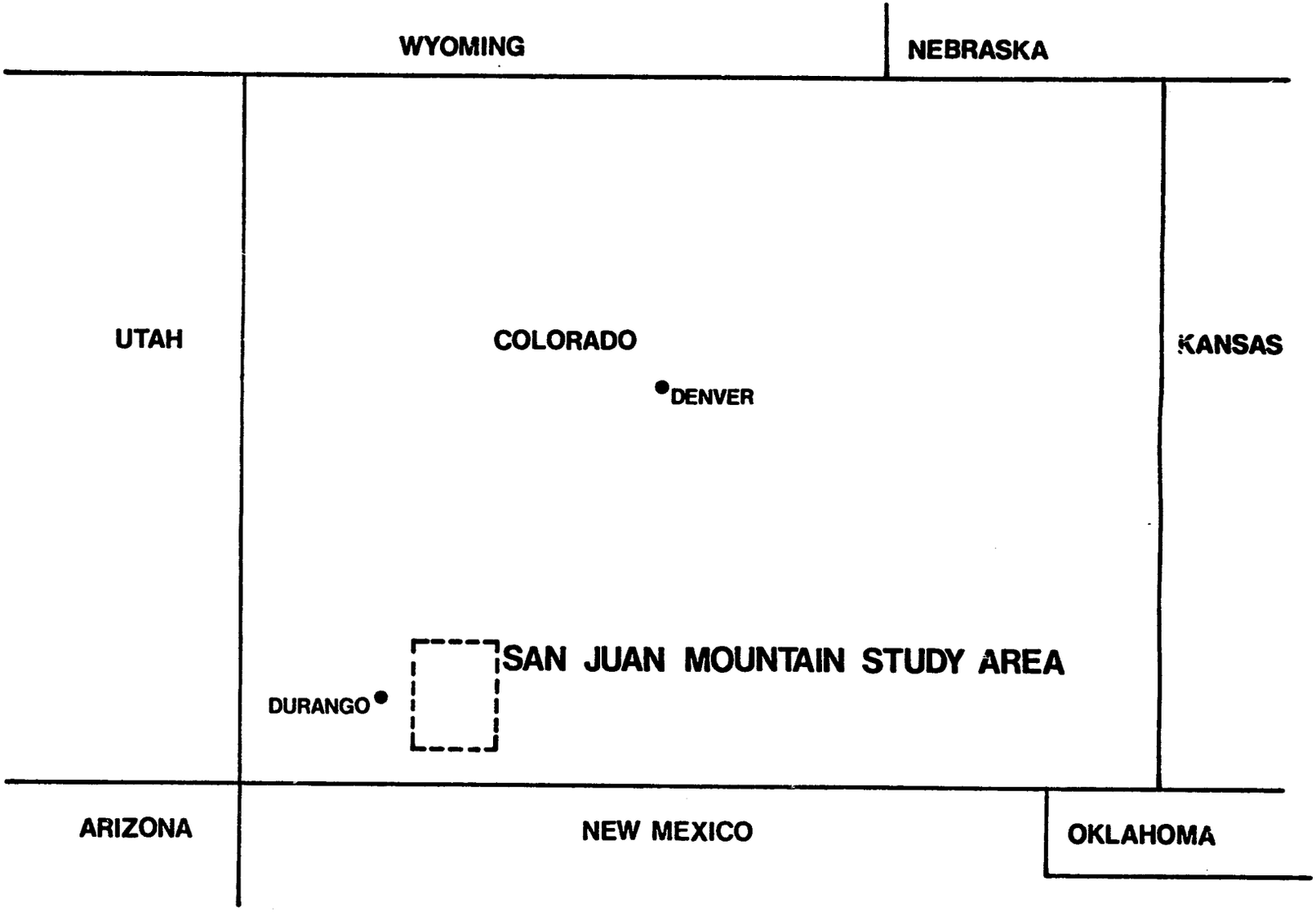


Figure 1. Location of the San Juan Mountain study area.

growth of unpalatable forbes or the invasion of sagebrush and Gambel oak. Over-grazing by sheep above the timberline seriously increases the erosion potential.

In areas where man, animals, fire, landslides, or other influences have not caused major changes in the vegetative cover, the naturally occurring vegetation is not only influenced by but is, in fact, determined by a complex interaction of edaphic, topographic and climatic factors. For example, at the higher elevations the soil mantle is thin and very poorly developed, air temperatures are generally low, and the growing season is very short. This combination of factors creates an environment suitable only for the short-season grasses and forbes found in tundra areas.

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, a complex relationship also exists between elevation and the quantity and quality of the solar radiation. While there is some tendency at high altitudes toward increased cloud cover, the solar radiation that is received at the earth's surface is of greater intensity and has a larger component of the high-energy shorter wavelengths than is found at lower elevations, since there is less atmospheric attenuation at the higher elevations. Both the 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 2 graphically displays the generalized distribution of cover types in the S.W. United States as a function of elevation. Within 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 describe in a general way the altitudinally defined vegetation zones within the San Juan Mountains of Colorado.

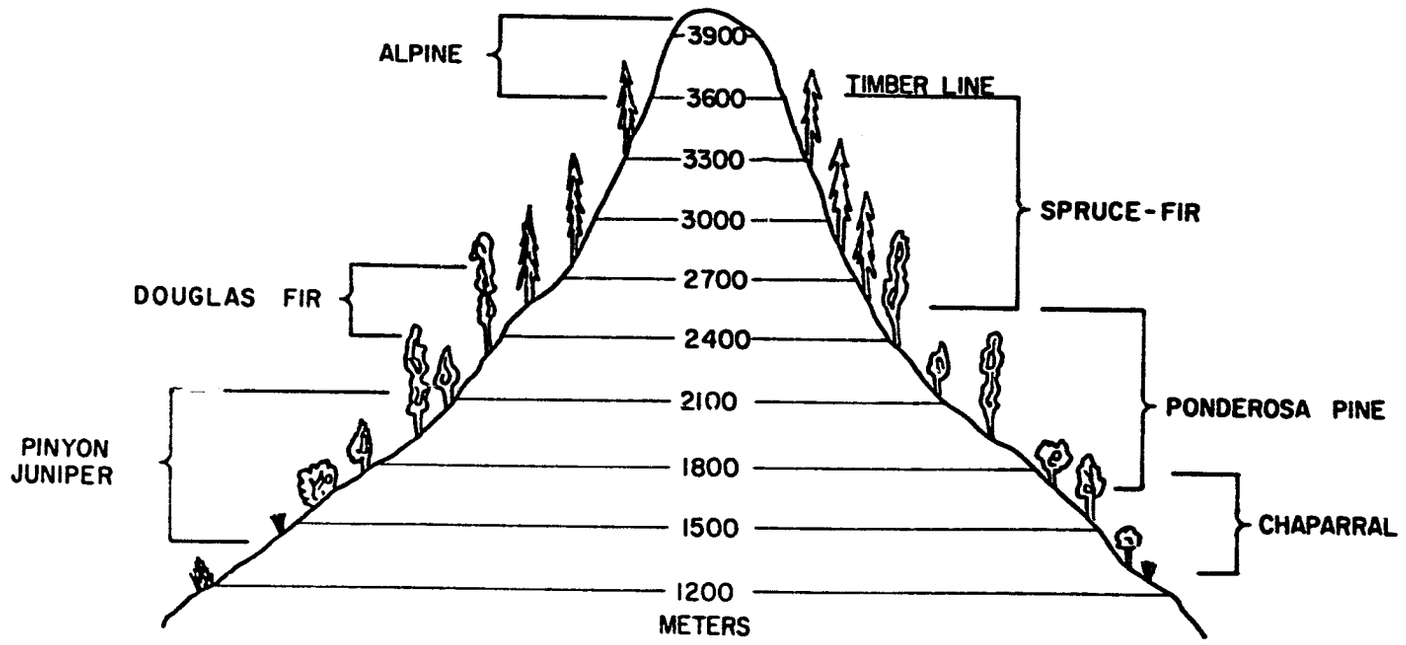


Figure 2. Relationships between elevation and distribution of vegetative cover types in the Rocky Mountains (after Little).

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.

Engelmann spruce and subalpine fir form the most extensive coniferous forest in the study area, extending from the timberline down to the Douglas-fir/white fir zone. At timberline, Engelmann spruce forms a dense climax cover as krummholz between the forest and alpine tundra. Here the growth is very stunted and twisted by the harsh weather conditions. At the lower elevations, however, Engelmann spruce and subalpine fir are very valuable timber resources and are logged extensively. Interspersed among the spruce and firs are numerous subalpine wet mountain meadows and grassland areas which characteristically are rather park-like with lush growths of grasses and grass-like plants and forbes.

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. Characteristics of Landsat and Topographic Data

The spectral data used in this investigation were Landsat MSS data which had been geometrically corrected and re-scaled to a 1:24,000 line-printer scale through LARS' preprocessing routines (Anuta, 1973). A detailed description of the data set is shown in Table 1.

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 has a cell size of 64 meters square. These DMA elevation data were registered with the Landsat data at LARS, using a nearest-neightbor fit, and then added to the Landsat data

Table 1. Location and description of spectral and topographic data.

	Spectral Data	Topographic Data
Source	Landsat Scene Number 1407-17193	Defense Mapping Agency (Rescaled at LARS at which time the aspect and slope channels were generated)
Date Collected	3 September 1973	--
Tape/File Number	2634/1	2629/1
LARS Run Number	73034309	73034311
Lines/Interval	1 - 1398/1	1 - 1398/1
Columns/Interval	1 - 1512/1	1 - 1512/1
Channel Descriptions	(1) 0.5-0.6 μm (2) 0.6-0.7 μm (3) 0.7-0.8 μm (4) 0.8-1.1 μm	(5) elevation (10-meter contour intervals) (6) slope (0-90° in 1° increments) (9) aspect (0-360° in 1° increments)

tape as channel 5 (Table 2). (See Appendix A for additional details). Figure 3 is a gray-level printer display of a portion of this elevation data.

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. These data added two channels to the data tape, channel 6 for slope and channel 9 for aspect. Figures 4 and 5a are gray-level representations of the slope data in channel 6 and the aspect data in channel 9, respectively. Channels 7 and 8 were added in order to express the aspect information in a different format: 0 - 180^{o}, and a 0-1 flag indicating direction (0 = East aspect, 1 = West aspect). Figure 5b represents this data.

The representation of the actual topographic character of the scene by this topographic data set is limited in several ways. The mountain tops that extend above contour lines but do not reach the next higher contour are truncated to the elevation of the contour line they reach; second, in areas where the elevation changes rapidly within short distances, a large number of pixel locations are defined as contour-line elevations and relatively few fall at gradations between the lines. These attributes of the data are a direct result of the procedures used to digitize the elevation information and the size of the grid cells used in creating this digital elevation data set. Details describing the procedures for elevation interpolation, registration, and derivation of slope and aspect data are included in Appendix A.

C. Reference Data

The reference data used in this project consist of 7½-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 data and of the interpolated topographic data. The aerial photography used is color infrared photography at a scale of 1:120,000 obtained by NASA's WB-57 on August 4, 1973. This photography is of excellent quality

Table 2. Characteristics of combined spectral and topographic data set used in analysis.

Source:	Landsat MSS and DMA digital topographic data (described in Table 1)
Tape/File Number	4827/3
LARS Run Number	73057711
Lines/Interval	1 - 1398/1
Columns/Interval	1 - 1512/1
Channel Descriptions	(1) Landsat 0.5-0.6 μm
	(2) Landsat 0.6-0.7 μm
	(3) Landsat 0.7-0.8 μm
	(4) Landsat 0.8-1.1 μm
	(5) Elevation (in 10-meter contour intervals)
	(6) Slope (0 - 90° in 1° increments)
	(7) Aspect (1 - 180° in 1° increments)
	(8) Aspect (0 or 1; e.g., E or W)
	(9) Aspect (1 - 360° in 1.41° increments)
	(10) Inverse of reflectance geometry correction factor
	(11) Corrected spectral values for 0.5-0.6 μm channel
	(12) Corrected spectral values for 0.6-0.7 μm channel
	(13) Corrected spectral values for 0.7-0.8 μm channel
	(14) Corrected spectral values for 0.8-1.1 μm channel

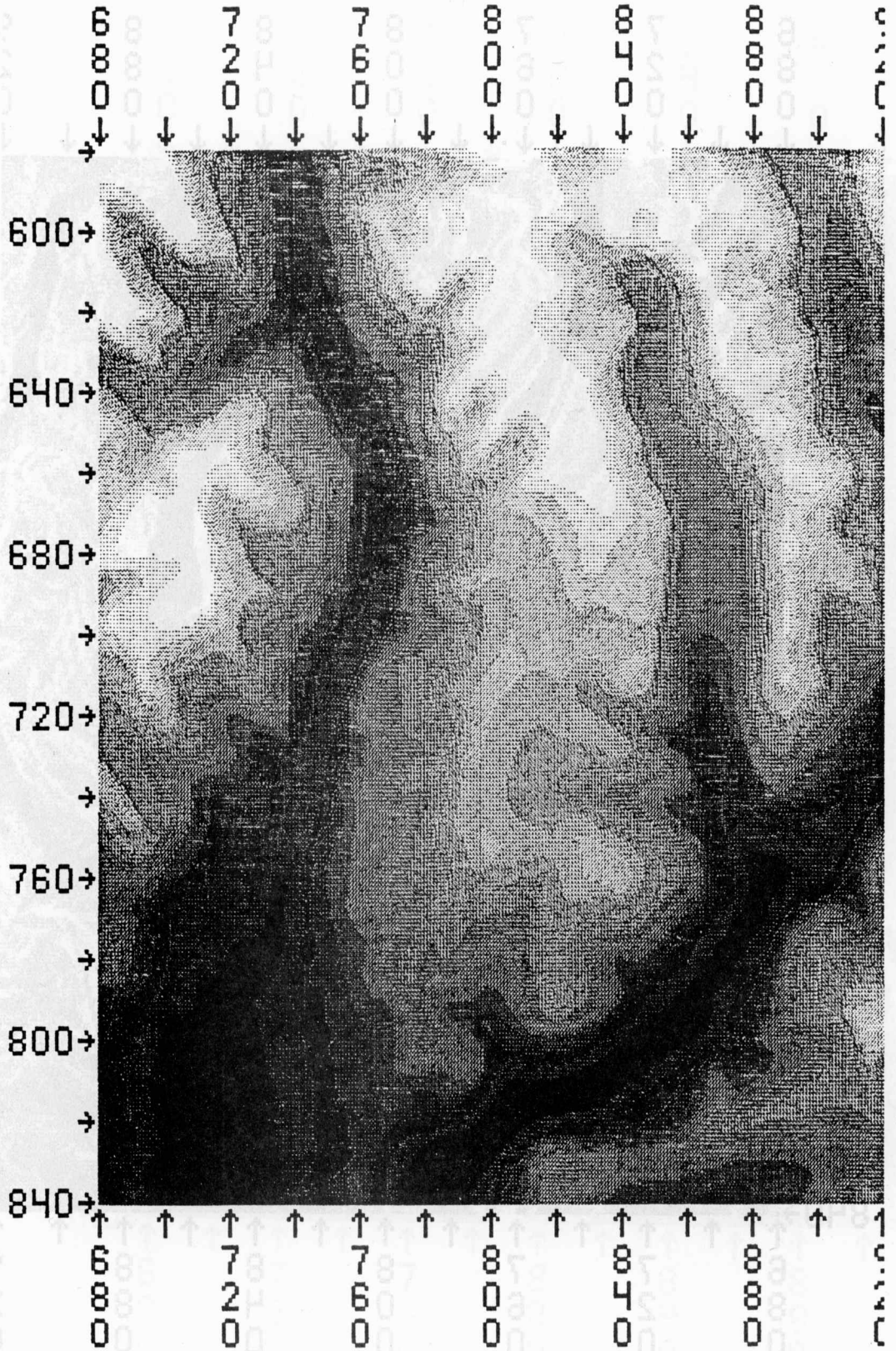


Figure 3. Example of digital elevation data. The area shown includes approximately 57,000 acres north of the Vallecito Reservoir. Lighter tones represent higher elevations.

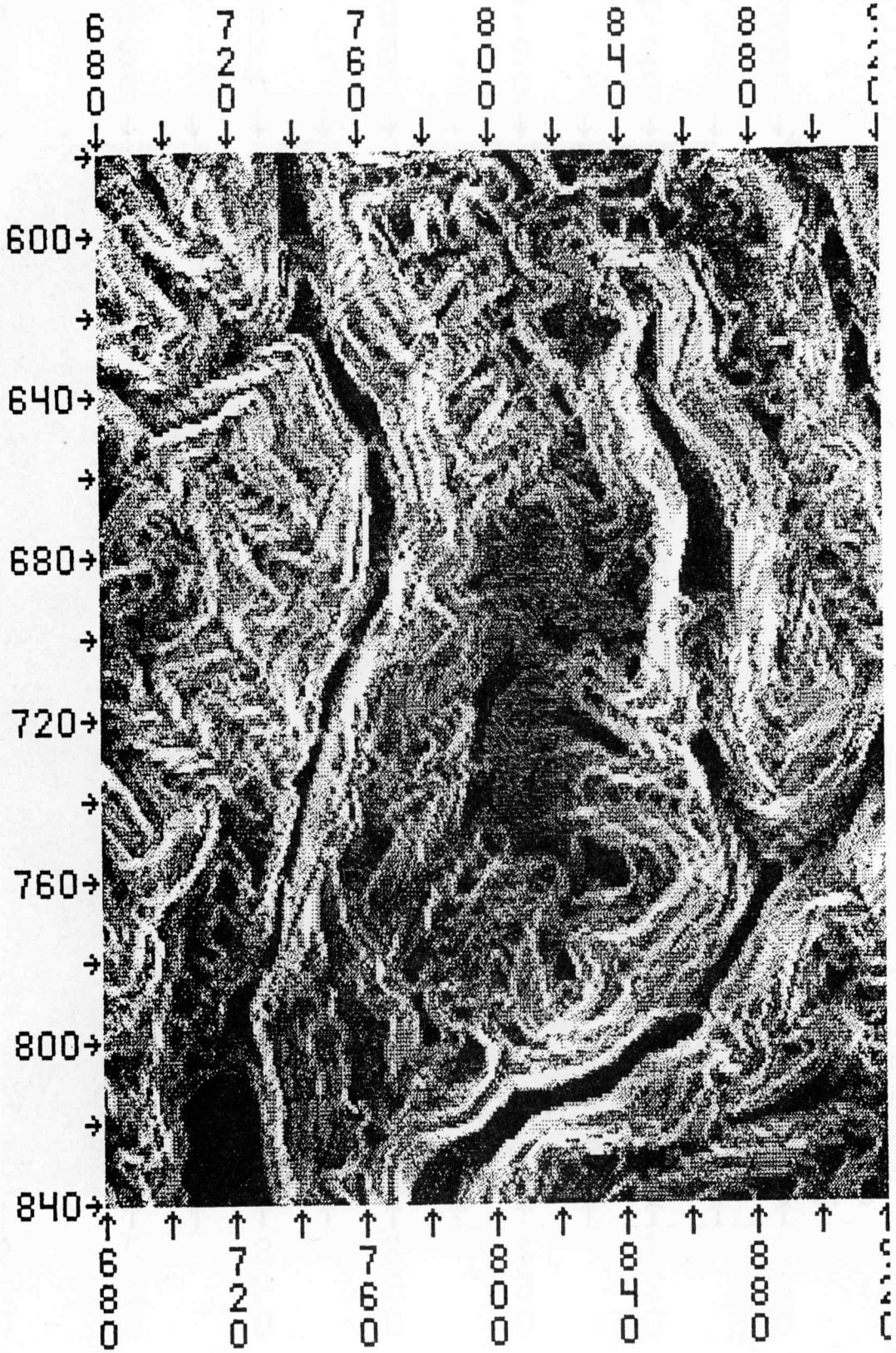


Figure 4. Example of the digitally interpolated slope data for the same area shown in Figure 3. Sixteen slope groups are represented: white is used for the steepest slope, black represents the 0^{o} (or flat) slope group.

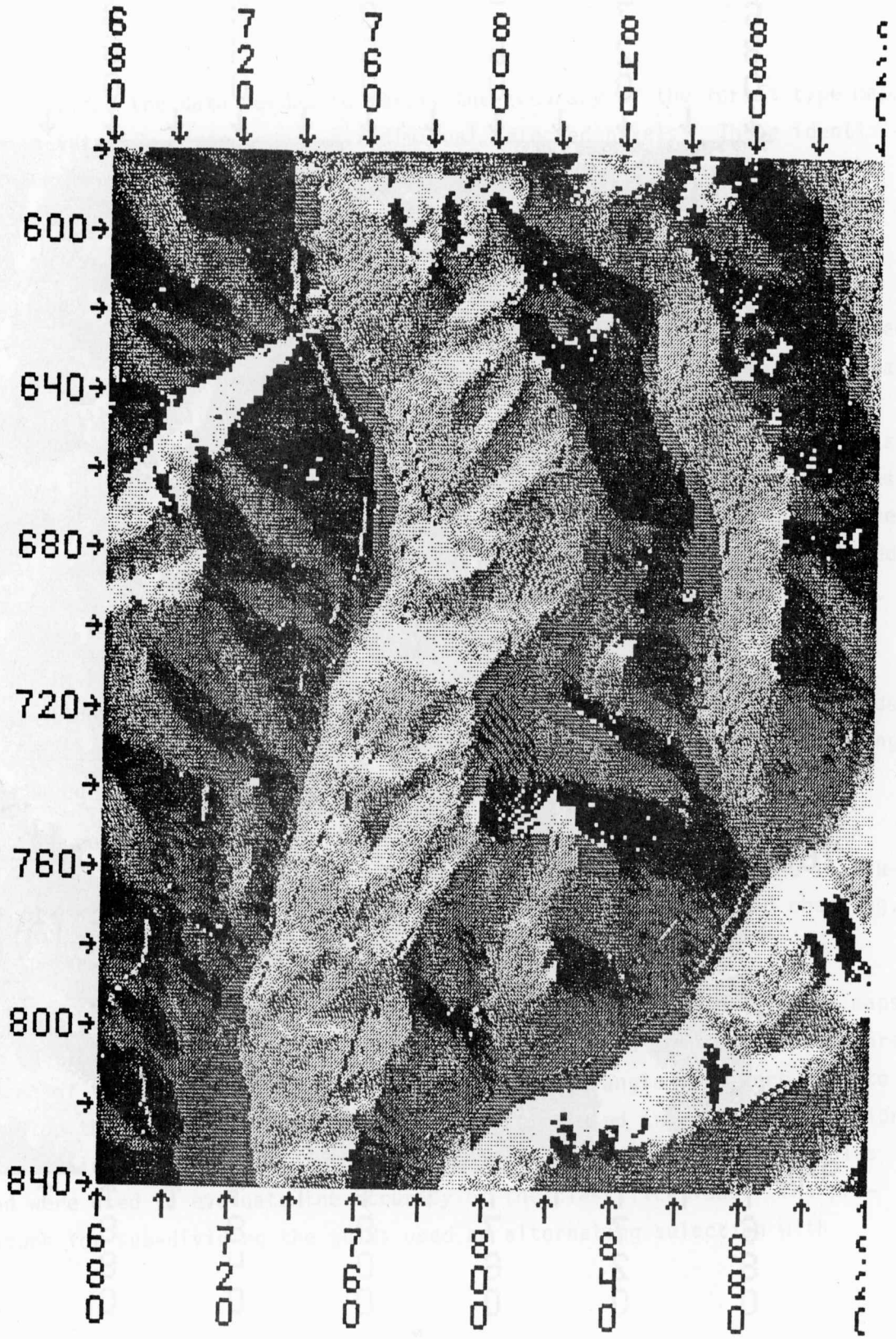


Figure 5a. Example of digitally interpolated aspect data for the same area shown in Figure 3. Aspects ranging from 0^{o} (North) to 360^{o} are displayed in shades of gray. Lighter tones are west and northwest-facing slopes.

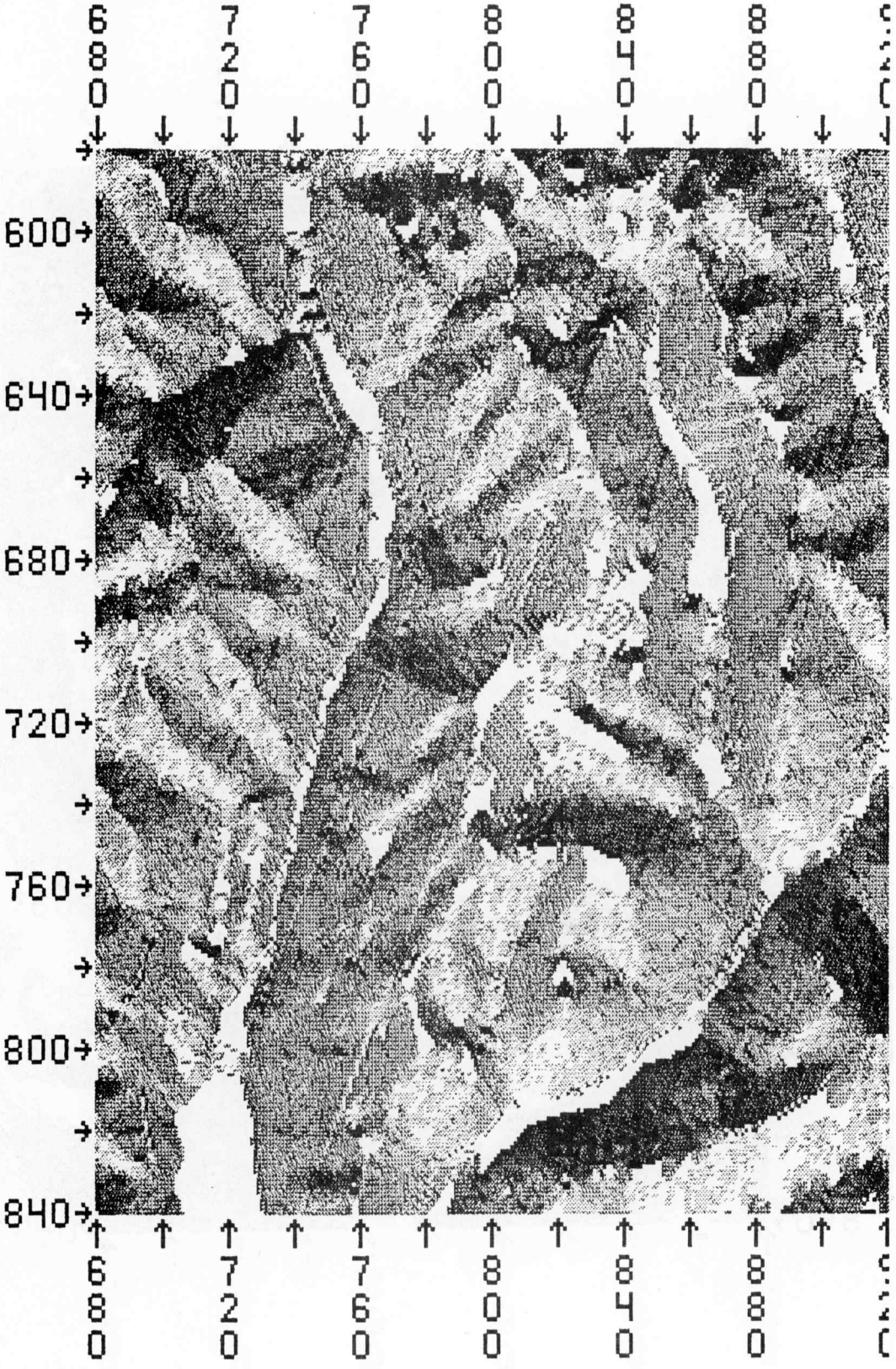


Figure 5b. Example of digitally interpreted aspect data for the same area shown in Figure 3. Aspects ranging from 0^{o} (North) to 180^{o} (South) are displayed in shades of gray. Dark tones are north-facing aspects and light tones are south-facing aspects.

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. All contain information at the "sub-series" level as defined in Table 3. While the series level of detail defines the informational classes desired, many of these 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 will be referred to as Level IV or sub-series classes. An example of one of the type maps from the San Juan National Forest is shown in Figure 6. Appendix B lists the code numbers and corresponding cover types shown on the type maps developed by INSTAAR; the original code designations were modified in the summer of 1978 by the field team working on the current project.

In general the INSTAAR cover type maps were reasonably accurate in stand identification, but some boundaries between cover types were inaccurate. These maps were refined during this project through field checking.

D. Training and Evaluation Quadrangles

As part of the reference data described in the previous section, maps of the forest cover were available for 14 quadrangles within the study area. Seven of these 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 quads used an alternating selection with

Table 3. Levels of mapping 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 & 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
	Alpine Willow	Alpine Willow
Herbaceous	Tundra	Xeric Tundra
		Mesic Tundra
		Hydric Tundra
	Grassland	Xeric Grassland
		Mesic Grassland
		Hydric Grassland
Non-Vegetated	Barren	Exposed Rock and Soil
	Urban	Urban
	Water	Water

Figure 6. Cover type map at the series and sub-series level for part of the Vallecito Reservoir quadrangle. The area shown is a portion of the area in Figure 3, corresponding approximately to lines 752-840 and columns 680-790. Identification of cover type code numbers appears in Appendix B.

a random start. Figure 7 shows the location of the quadrangles within the study area, and Table 4 lists the line and column coordinates for each quadrangle.

Table 4. Line and column coordinates for the 14 quadrangles used in the investigation.

Training quadrangles	Lines (first & last)	Columns (first & last)
Howardsville	205-386	610-790
Little Squaw Creek	387-569	1154-1334
Vallecito Reservoir	752-934	610-790
Bear Mountain	752-934	973-1153
Pagosa Peak	752-934	1335-1506
Baldy Mountain	935-1117	791-972
Chris Mountain	935-1117	1154-1334
Evaluation quadrangles		
Finger Mesa	205-386	973-1153
Weminuche Pass	387-569	973-1153
Granite Peak	752-934	791-972
Oakbrush Ridge	752-934	1154-1334
Ludwig Mountain	935-1117	610-790
Devil Mountain	935-1117	973-1153
Pagosa Springs	935-1117	1335-1506

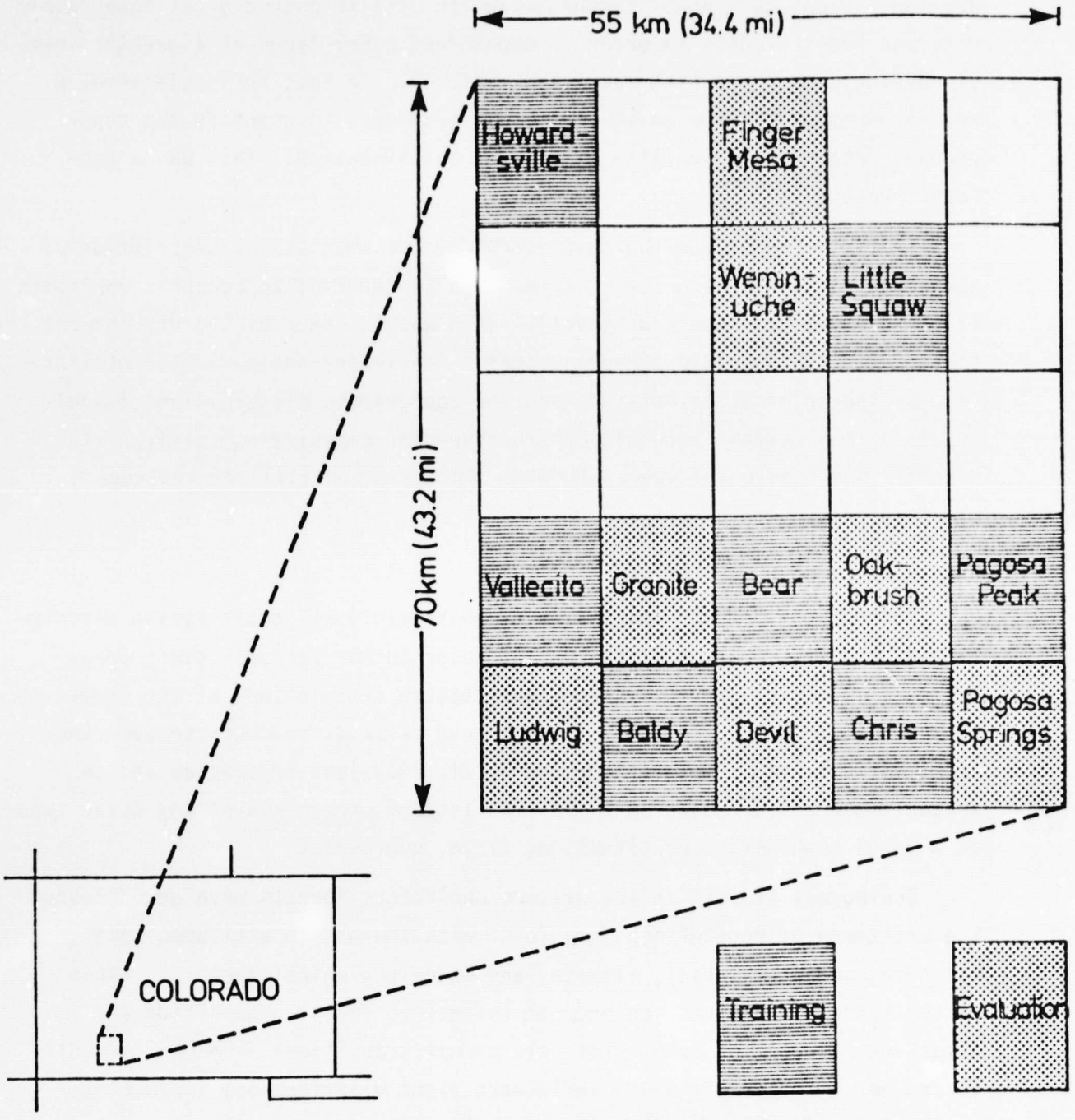


Figure 7. Quadrangles in the study area designated for "training" and "evaluation."

III. TOPOGRAPHIC DISTRIBUTION MODEL APPROACH

As previously stated, the overall objective of this study is 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. To meet this objective, a key requirement was the development of a procedure to quantify the topographic data and then utilize it in the classification. This was a two-phase process.

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) in a pattern recognition procedure for classifying multivariate digital topographic and spectral data. This section (III) of the report discusses these two phases.

A. Background and Literature Survey

Prior to this work, there existed no statistical, quantitative description of the distribution of the forest cover in the San Juan study area. The available literature contained qualitative descriptions of the elevation strata, as summarized in Table 5. The need existed, however, to describe quantitatively the complete topographic distributions of species and to assemble more information on the probability of occurrence of any cover type at a given combination of elevation, slope, and aspect.

Ecological studies in the western coniferous forests have demonstrated the existence of vegetational gradients with 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 (1943) recognized these relationships and pointed out that topographic position

Table 5. Summary of literature on topographic distribution of forest cover in the western mountains.

Author	Rasmussen	Pearson	Merriam	Whitfield	Woodbury		Marr	Fleming	
Date	1941	1920	1890	1933	1947		1961	1975	
Location	Kiabab Plateau, Northern Arizona	Southwestern Arizona/New Mexico	San Francisco Peaks, Northern Arizona	Pike's Peak, Central Colorado	Southeastern Utah		Boulder, Colorado	San Juan's SW Colorado	
Alpine tundra zone	-	above 11,500	above 11,500	above 11,500	North above 10,000	South above 11,000	above 11,300	above 11,500	
Engelmann spruce and subalpine fir zone	above 8,200 *	9,500 to 11,500	9,200 to 11,500	9,000 to 11,500	North 7,000 to 10,000	South 8,000 to 11,000	9,300 to 11,000	10,000 to 12,000	
Douglas and white fir zone	North down to 6,800	8,300 to 9,500	8,200 to 9,200	8,000 to 9,000			8,000 to 9,000	North 6,500 to 9,000	South 7,500 to 10,000
Ponderosa pine zone	6,800 N down to 6500 S up to 8,800	6,700 to 8,300	7,000 to 8,200	6,500 to 8,000	North 5,000 to 7,000	South 7,000 to 8,000	6,000 to 7,000	North 6,250 to 8,000	South 7,000 to 9,500
Pinyon/juniper zone	5,500-6,800 S up to 7,250 N 5,000-6,500	5,000 to 6,700	6,000 to 7,000	below 6,500	North 4,000 to 5,000	South 4,000 to 7,000		North 5,000 to 6,500	South 5,000 to 7,000
Oak/mountain mahogany	4,500 to 5,500	3,000 to 5,000	4,000 to 6,000	-	-	-	-	-	-

* not recognized as distinguishable communities

- not included in study

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.

Many researchers have described the location and characteristics of various vegetation zones in the Rocky Mountains. Daubenmire (1943) evaluated the existing descriptions and distinguished six major vegetation zones: the Alpine tundra zone; the Engelmann spruce/Subalpine fir zone; the Douglas-fir/White fir zone; the Ponderosa pine zone; the Pinyon/Juniper zone; and the Oak/Mountain mahogany zone.

The literature previously summarized in Table 5 indicates that the major forest communities in the Rocky Mountains do not seem to vary considerably in their elevation ranges from central Colorado south to northern Arizona and New Mexico. Each vegetation type has a characteristic elevation range which is adjusted locally by a combination of slope and aspect. Several authors did differentiate between two distinct classes, northern exposures and southern exposures. The warmer, drier southern exposure raises the elevation range of a species whereas the cooler and moister northern exposure lowers the elevation range. However, it must be noted that with the exception of Fleming et al. (1975a), none of these studies is specific to the San Juan Mountains and thus none can serve as a definitive statement of species distributions in that location. Furthermore, the 1975 work by Fleming et al. (1975a) which is specific to the San Juan Mountains, has been modified through the current study.

B. Development of the Topographic Distribution Model

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 the forest cover type maps, the aerial photography, and topographic data tapes. The output from the model is a quantitative characterization of the topographic distribution of each major forest species in terms of means and variances; these data and statistics can be presented graphically

as histograms, polar plots, regression line plots, and normal distribution curves.

This section of the report describes the procedures used to develop the topographic model and the techniques developed for displaying the results. Section IIIC and Appendix C present graphically the topographic distribution of each major species in the San Juan study area. These characterizations, however, are specific descriptions of the vegetation of the San Juan 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 region. However, the basic techniques used to develop these plots can be applied to other mountainous areas for which cover type maps and elevation information are available.

1. Stratification to define topographic positions. The procedure for developing the topographic distribution model involves: a) stratifying the test site into 300-meter elevation zones (resulting in a total of 7 strata for the San Juan site); b) stratifying each elevation stratum into three slope zones (1-7^{o}, 8-17^{o}, and 18-70^{o}) and a "zero slope" zone; and c) stratifying each point with a non-zero slope into one of four aspect zones (N,S,E,W). This process provides for the definition of 91 (7 x [(3 x 4) + 1]) distinct topographic positions for the study site.

2. Sampling procedure for the topographic model. Selection of a statistically valid sample of data points was the next step in constructing the 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 sampling efficiency and simplicity in handling. An additional advantage of using single-pixel cells over groups of pixels is the minimization of many of the edge effect problems inherent in using the larger cells.

In order to represent equally each of the 91 topographic positions in the study area, 50 randomly defined X-Y coordinates were selected in each of the 91 topographic positions. It had been estimated that 50 points would provide an adequate representation of each topographic position. The points were allocated among the training quadrangles as a function of the proportion of the 91 positions present in the quadrangle. This yielded a total of 4,550 stratified random sample points (50 x 91). Selection of these stratified

random points was carried out through three computer programs developed as part of this project: EXTRACT, RANDOM, and SELECT. EXTRACT is a program that, in essence, classifies the elevation, slope, and aspect channels of the data base, assigns each pixel to one of 91 topographic position classes and then lists the X-Y coordinates of all the points in each of the 91 topographic position classes. RANDOM is basically a random-number generator that provides a set of random numbers to SELECT which uses them to select the desired number of points (in this case 50) from among all the points in each topographic position class. Detailed descriptions of these programs appear in Appendix D.

3. Identification of sample points. 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. This resulted in 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 a high degree of accuracy in the identification of the data used to develop the model.

It should be noted at this point that the 3,379 training sample points actually utilized in the model served simply as representatives of the 91 topographic classes. After the cover type for each of the points had been identified, the topographic information from the data tape was used to define 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. In summary, while a stratification procedure had been used to obtain a sample of all topographic positions present, the model itself was developed using the actual topographic location of each of the sample points.

4. Statistical characterization of the distributions. Regression analyses of the samples were run to describe statistically the topographic distribution

of each cover type. The frequency of the various species along an elevational gradient can be easily plotted to show the basic shape and characteristics of the distribution. Examples of these plots appear in the next section. 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 (north = 0, south = 180, with east and west both 90). A regression analysis was also conducted on the three slope classes (1-7^{0}, 8-17^{0}, and 18-70^{0}) for each species.

5. Procedures for displaying the distributions. The distributions of the various cover types relative to topographic position can be presented graphically in a number of ways, and during this project, computer software was developed to accomplish this. Figures 8, 9, 10 and 11 are examples of four of these formats: a histogram that shows the distribution of cover types as a function of elevation; polar diagrams that shows the distribution as a function of elevation and aspect, or of slope and aspect; and a regression line that combines elevation and aspect to present in another way the distribution of each cover type. In Figure 8, the elevation is divided into 50-meter zones, and in each zone the number of pixels assigned to each major species is counted. Examination of the histograms reveals the degree of normality of each distribution. The polar plots, which display distribution as a function of elevation and aspect or slope and aspect, indicate the extent that the "typical" elevation range varies for each species as a function of aspect or slope. The polar diagrams also serve to verify that the model contains a good representation of the sample points for the full range of aspects and slopes. The regression line (Figure 11) displays the key information about the two most significant variables, elevation and aspect, in a format that clearly shows their relationships.

C. The Model for the Topographic Distribution of Forest and Herbaceous Cover Types in the San Juan Study Area.

1. Graphical characterization of the topographic distribution of forest-land cover types. One of the products of this study is a quantitative description of the distribution of each of the three major coniferous species, the three major deciduous species, and the two major herbaceous cover types in the San Juan study area. Figures in Appendix C display those distributions

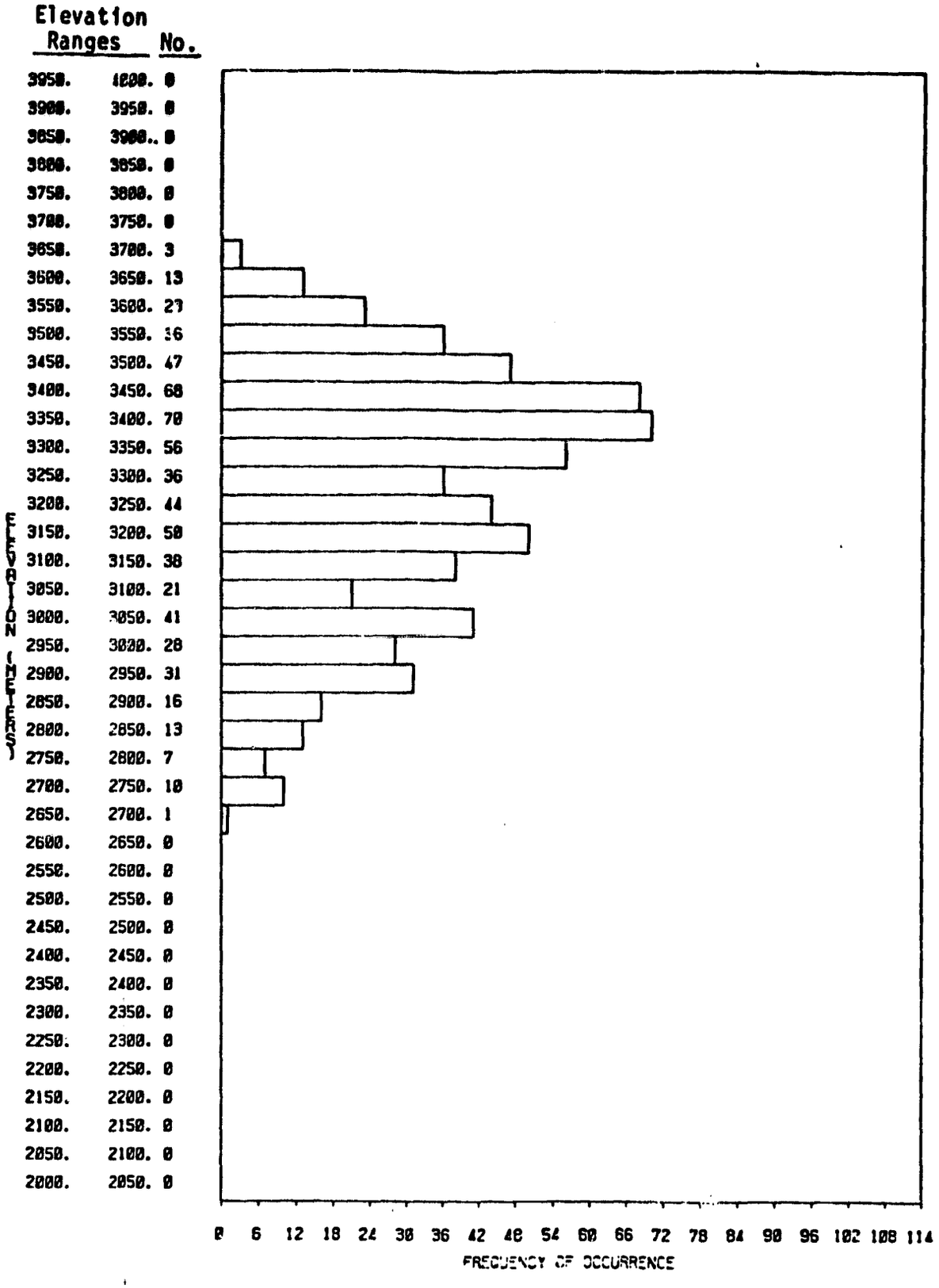


Figure 8. Sample histogram graph showing the distribution of Engelmann spruce/subalpine fir as a function of elevation.

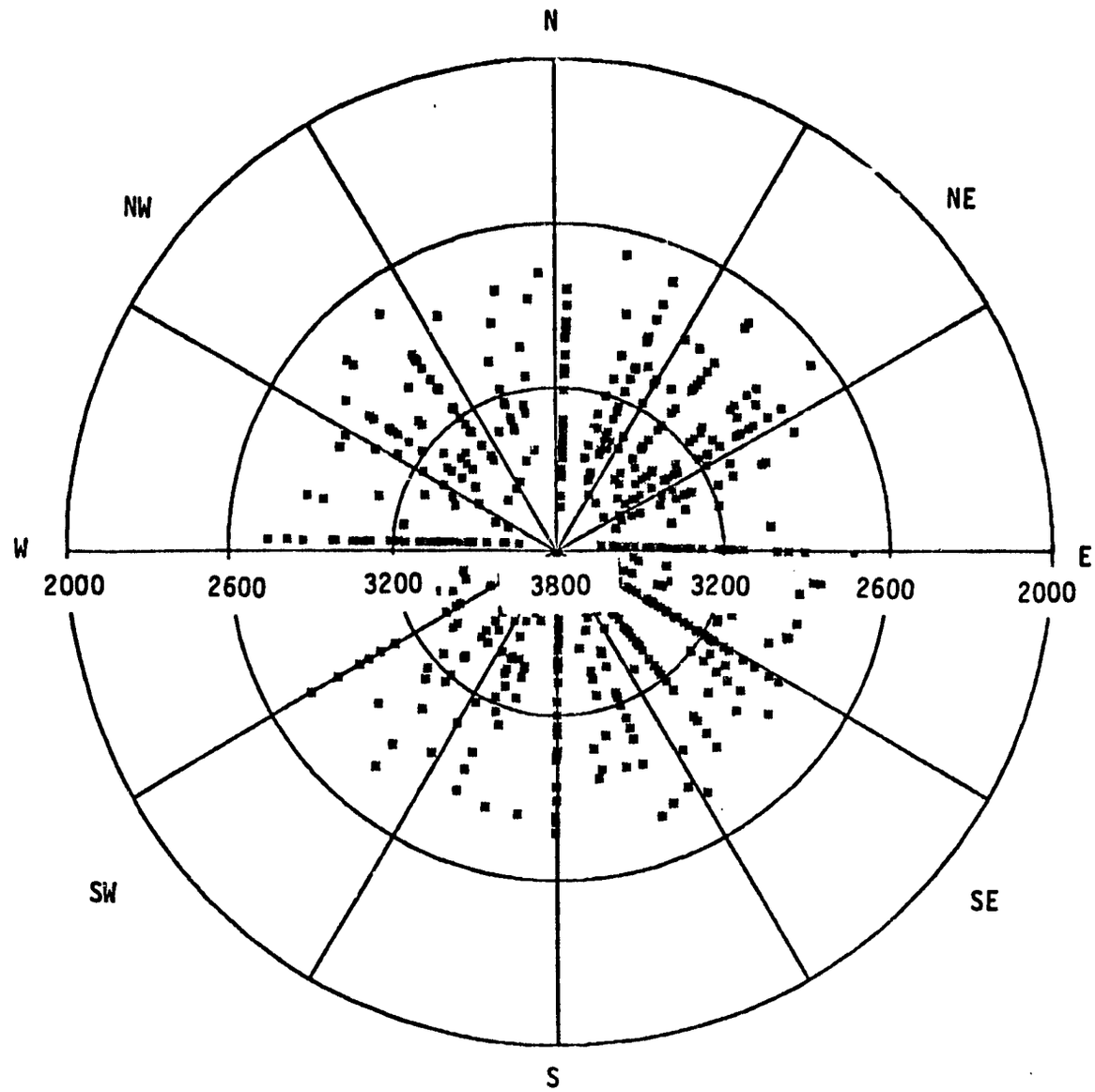


Figure 9. Sample polar plot displaying the distribution of Engelmann spruce/subalpine fir as a function of elevation (in meters) and aspect.

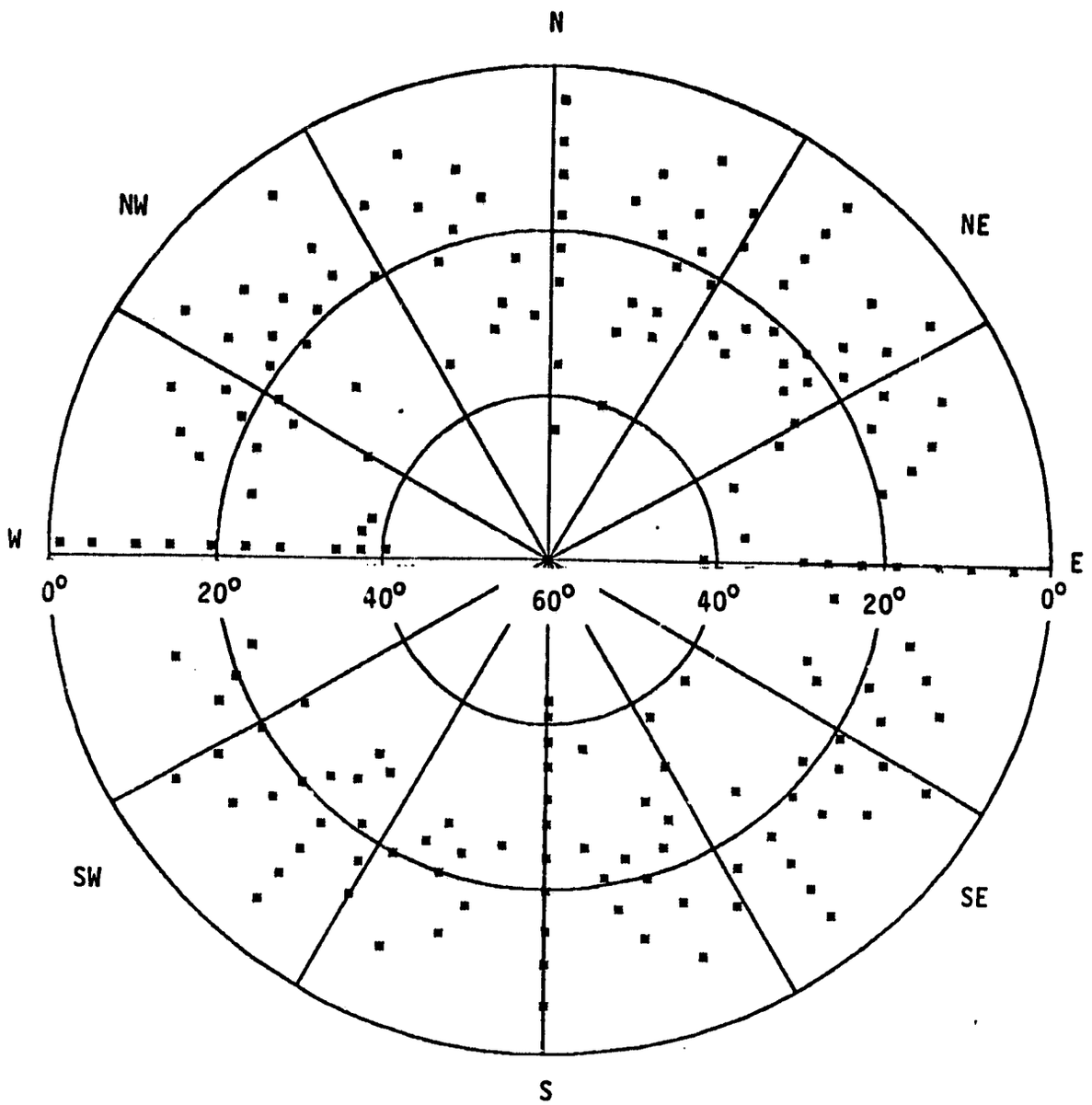


Figure 10. Sample polar plot displaying the distribution of Engelmann spruce/subalpine fir as a function of slope and aspect.

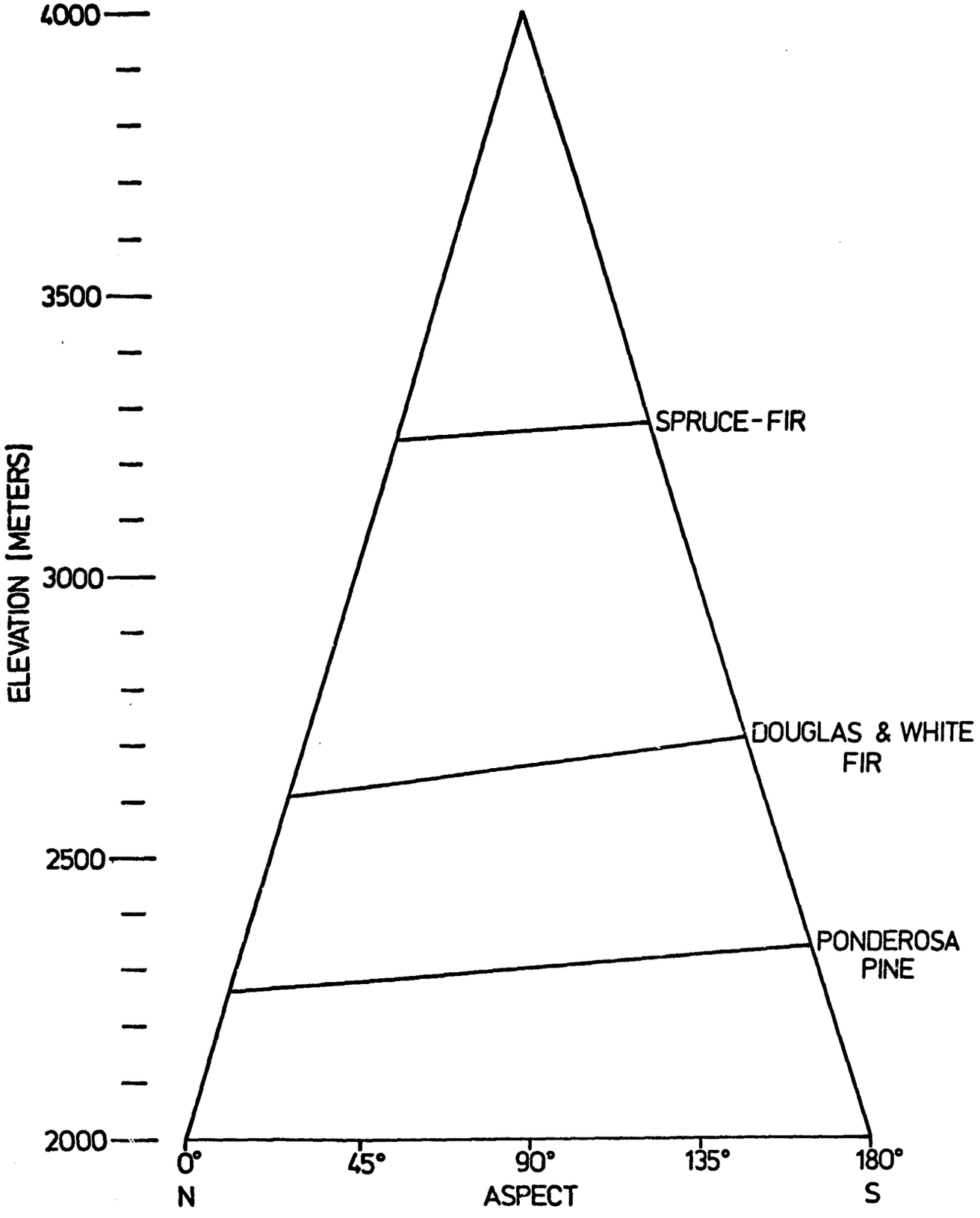


Figure 11. Sample regression line plots showing the relationship between elevation and aspect for spruce/fir, Douglas/white fir, and ponderosa pine for all slopes.

for each species relative to elevation, slope, and aspect. The normalized curves of the distribution of the various cover types as a function of elevation are shown, by group, in Figures 12, 13, and 14. From these figures it is evident that the various species have statistically different mean elevations but also that some overlap of elevation ranges does exist between the species, creating transition zones.

As compared to previous information concerning the topographic position and distribution of the various cover types (as shown in Figure 2), the figures appearing in Appendix C and statistically summarized in Figures 12, 13, and 14 represent the actual distribution of the various cover types, based upon the information generated through the topographic distribution model. Previous information has been largely qualitative, whereas these figures quantitatively characterize the topographic positions and ranges of the various cover types. This approach has shown some differences in the elevation range of some species as compared to the information available in the literature. For example, the topographic distribution model data shows that the spruce/fir cover type extends to a higher elevation than previously thought.

In comparing the histogram data shown in Appendix C with the normalized data summary displayed in Figures 12, 13, and 14, 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 original data for these species was caused by the deficiency in sample points below 2,225 meters. The topographic distribution model had been truncated at the lower elevations due to the range of elevation existing in the study site. (Ponderosa pine as well as pinyon/juniper cover types are present at lower elevations in southwest Colorado outside of the test site area.) It could also be noted that the histogram data for grassland shown in Appendix C displays a broad range of elevations without a distinct single mean. It is possible that this is because the data shown summarizes data for many different species of grass, each of which may in fact have distinct elevational distributions. However, the fact that the normalized curve for grassland shown in Figure 14 displays a more pronounced mean than the original data displayed does not indicate a need to separate grassland into sub-groups having different elevational ranges. Because the major use of the elevation

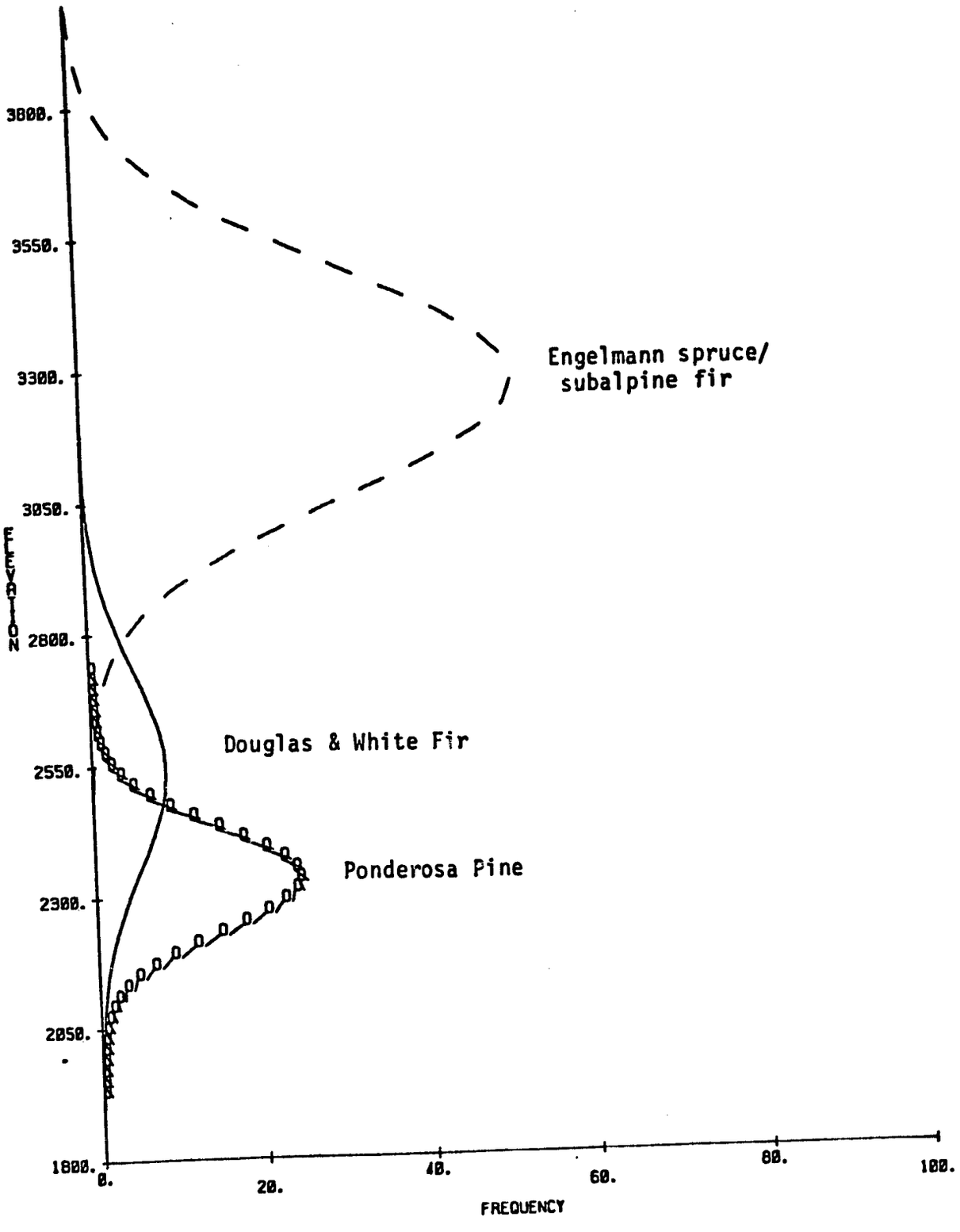


Figure 12. Gaussian curves of frequency along the elevational gradient for the major coniferous cover types. (Elevation in meters)

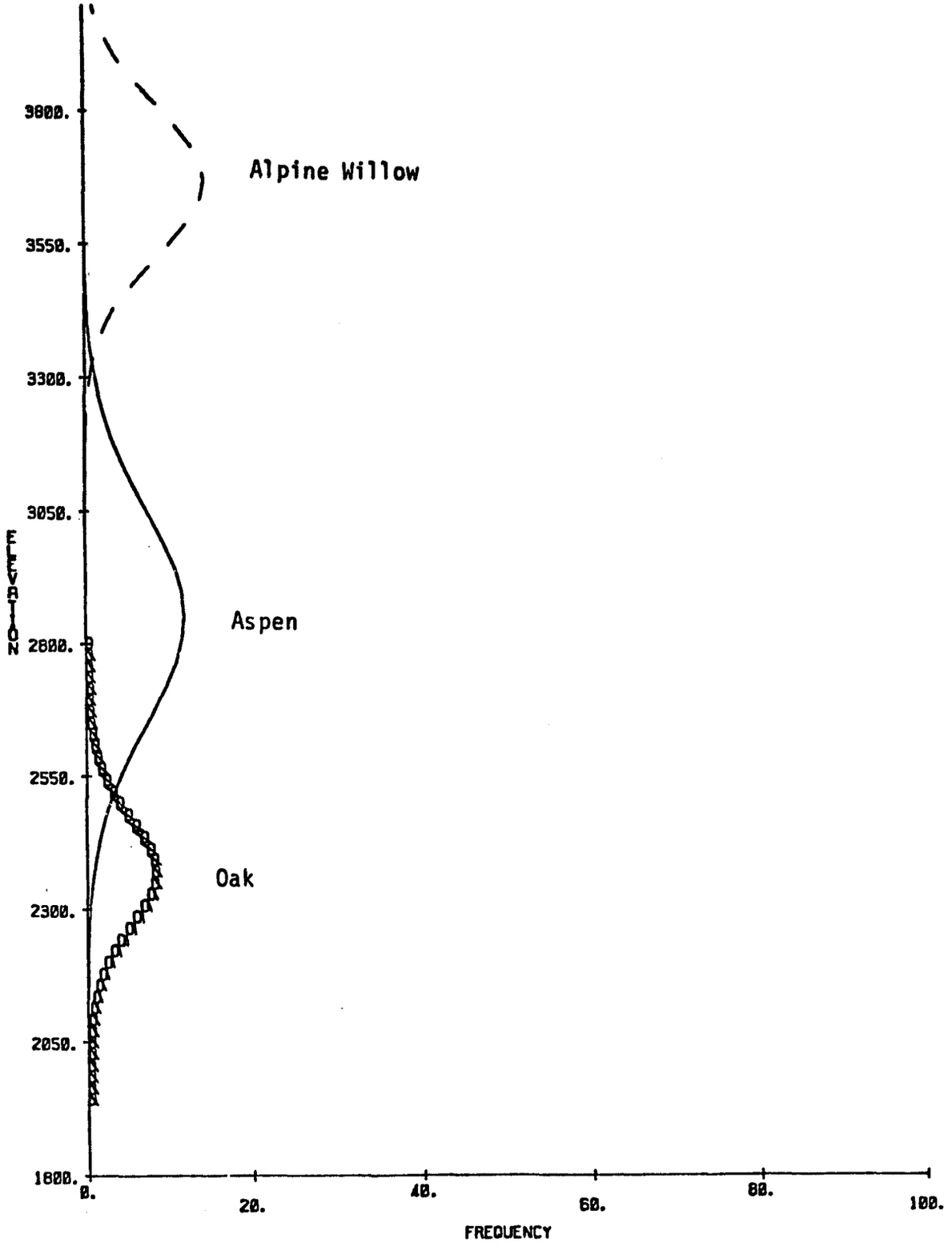


Figure 13. Gaussian curves of frequency along the elevational gradient for the major deciduous cover types. (Elevation in meters)

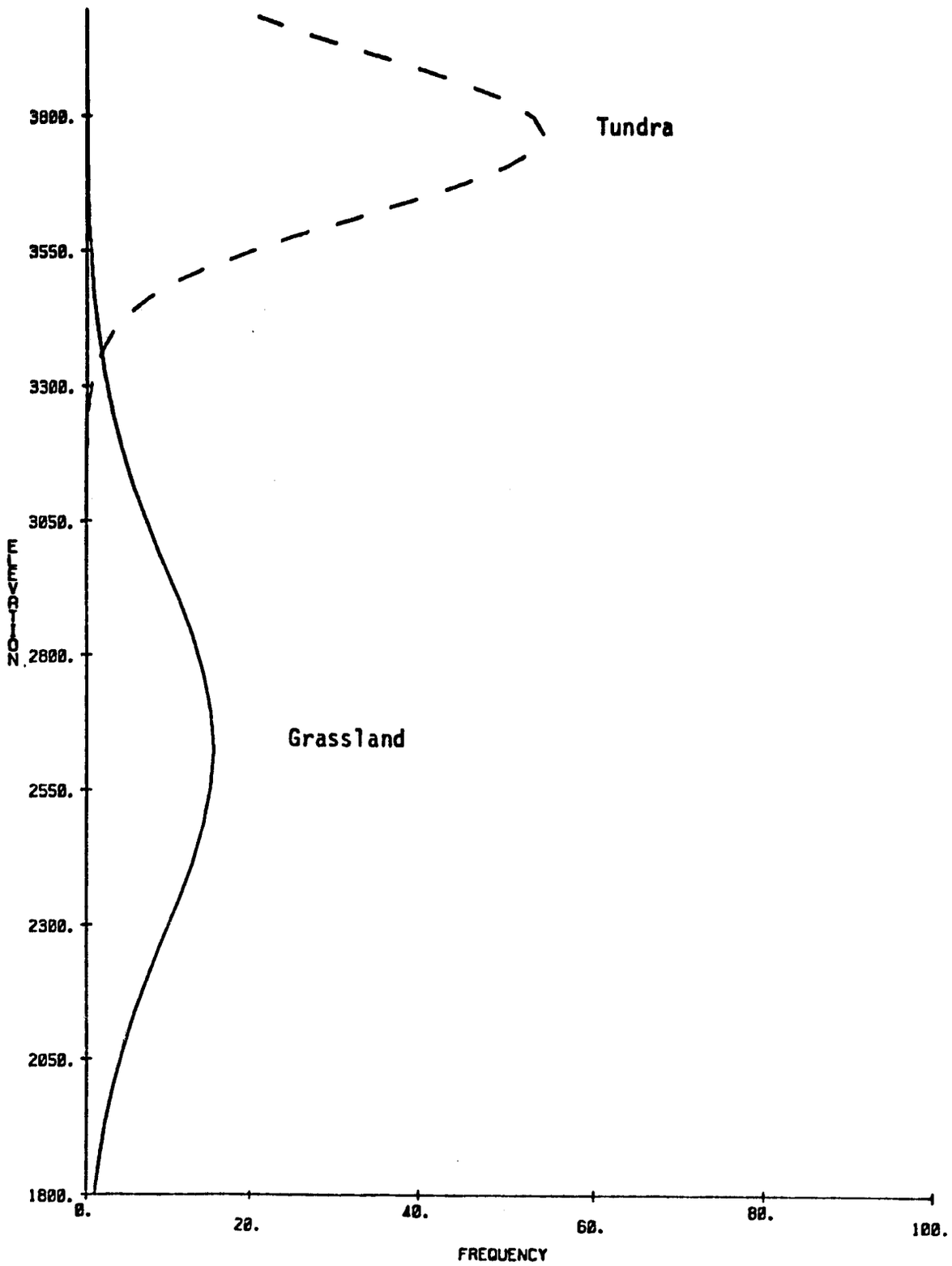


Figure 14. Gaussian curves of frequency along the elevational gradient for the major herbaceous cover types. (Elevation in meters)

data for grassland was to separate grassland from tundra, the topographic distribution model data does clearly show that grassland can be separated from tundra as a function of elevation.

Figure 11 (shown previously), Figures 15 and 16 show the regression line plots for the three coniferous cover types, the three deciduous, and the two herbaceous cover types, respectively. As one would expect, the average elevation is significantly higher on the southern aspects than on the northern. The data also show that there is very little difference in average elevation between east and west aspects. Average elevation for each species varies as a function of aspect by approximately 70 meters (225 feet), with Douglas-fir and white fir having the greatest aspect-dependent variations among the coniferous species. In addition, analysis of the data showed that slope is not a significant factor affecting the distributions of any of the forest species. This is indicated for the coniferous forest cover types in Figure 17, which is an example of the results of the regression analyses of the three slope classes.

2. Analysis of the topographic distribution model. A further analysis of the results was conducted to determine statistically which variables are significant in distinguishing among the various species and the accuracy of using topographic data alone to distinguish among the species within each major cover type group. To carry out the discriminant analysis, 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 (Nie et al., 1975) 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, topographic variables for all sample points in 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.

Table 6 shows the results of classifications of the sample points in the three categories when (1) equal a priori probabilities were assigned to each class and (2) when the probabilities were weighted. In the former case, the range of accuracies for the various species is from 70.8% for aspen to 100% for grass, with the average near 89.5%. In all cases, species in the

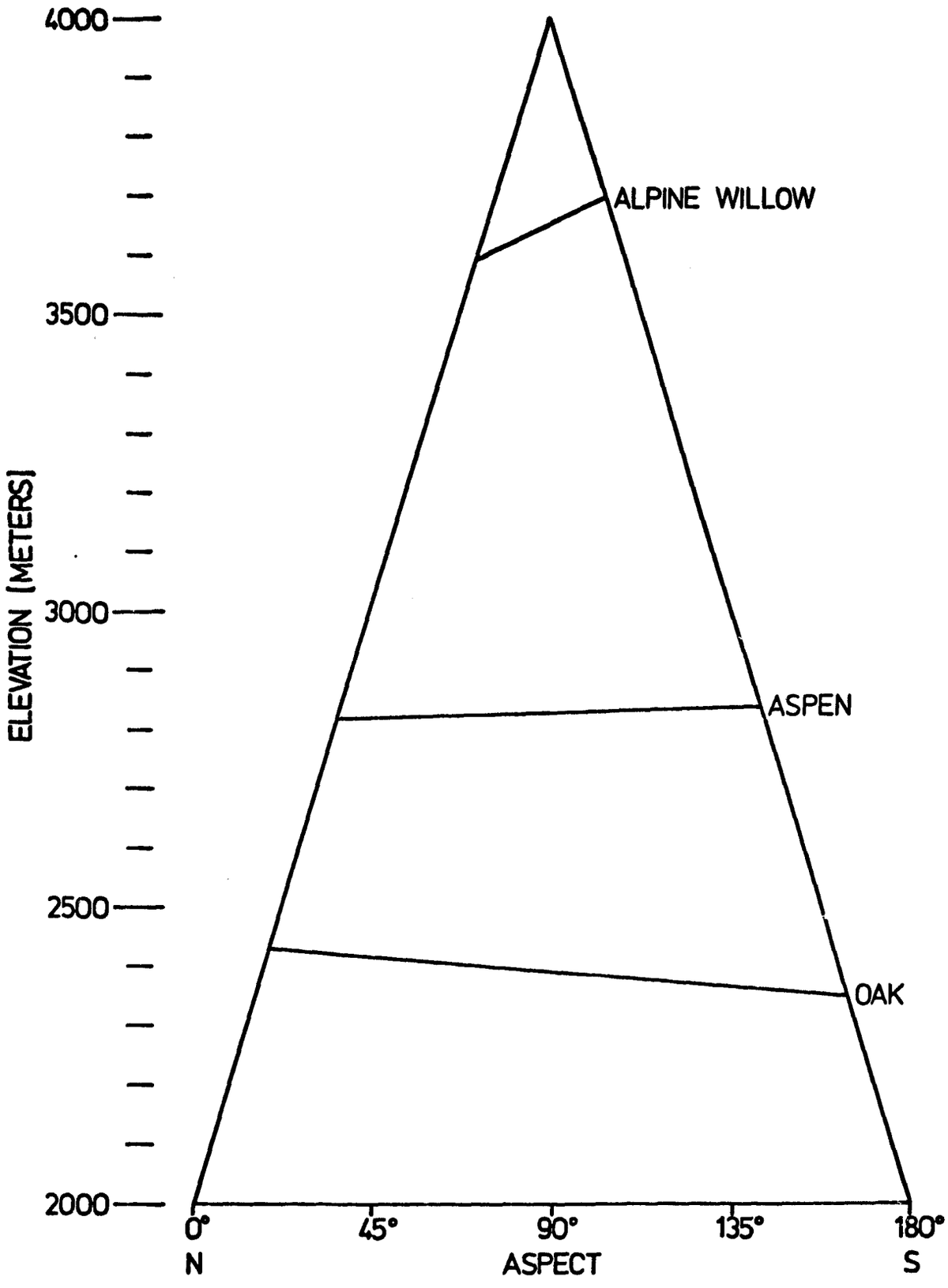


Figure 15. Sample regression line plots showing the relationship between elevation and aspect for alpine willow, aspen and oak.

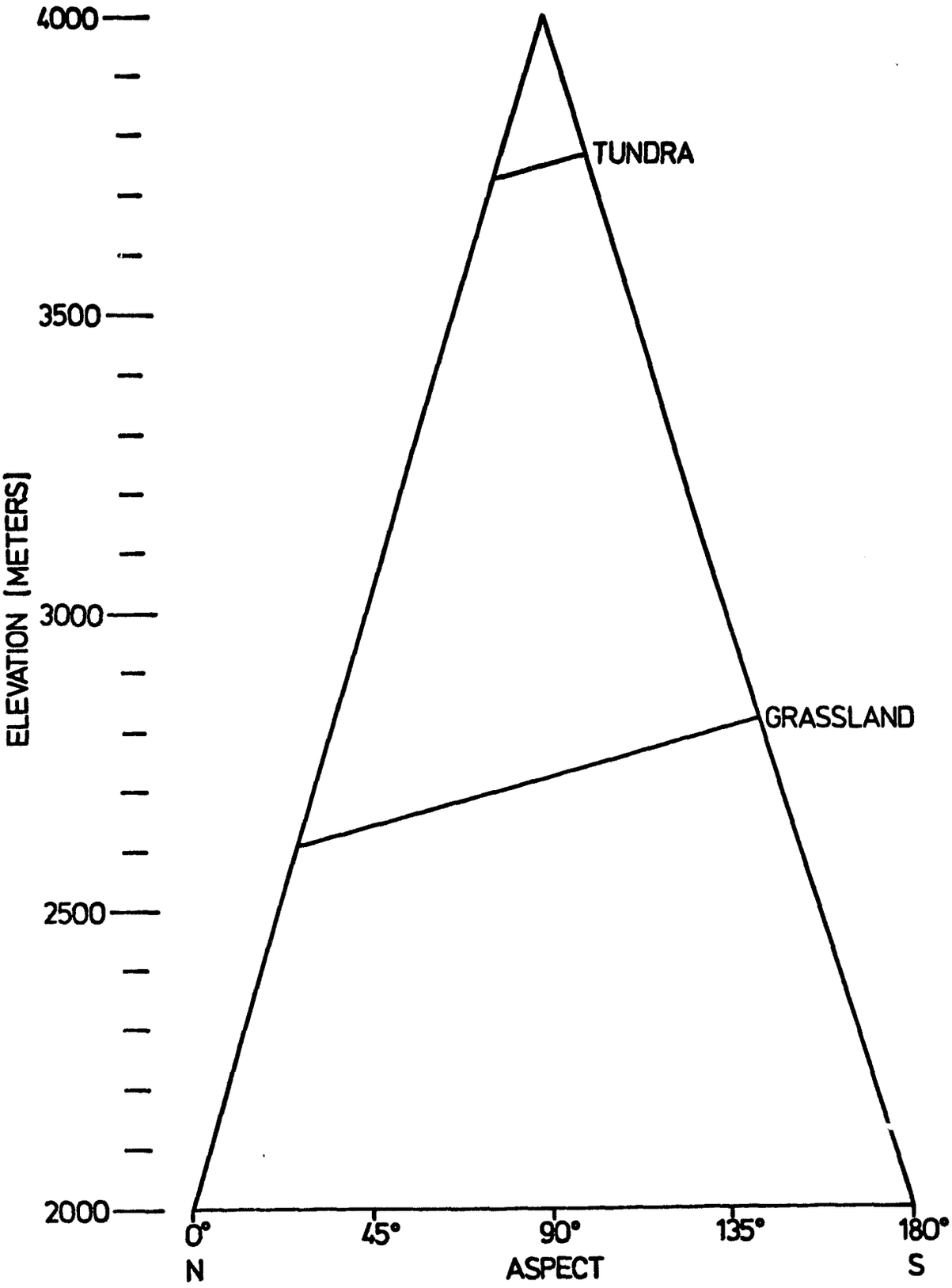


Figure 16. Sample regression line plots showing the relationship between elevation and aspect for tundra and grassland.

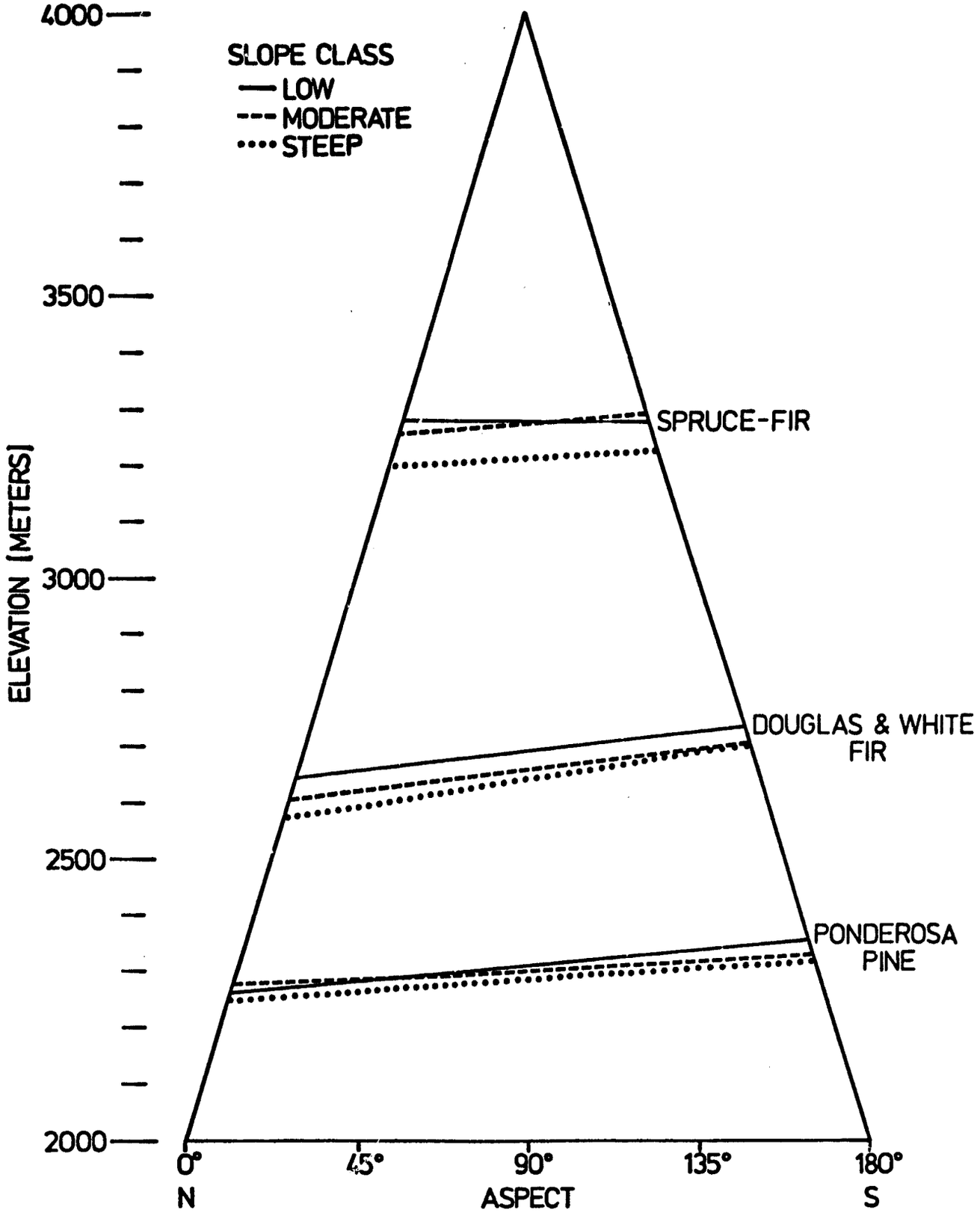


Figure 17. The relationship between elevation and aspect by slope class for the major coniferous cover types.

Table 6. Training sample discriminant analysis results when using only topographic data.

	No. of Cases	Class Assignment: Equal Probability			Class Assignment: Weighted Probability[1]		
Actual Coniferous Groups		Spruce/ fir	Doug-white fir	Ponderosa pine	Spruce/ fir	Doug-white fir	Ponderosa pine
Spruce/Fir	806	729 90.4%	77 9.6%	0 0.0%	740 91.8%	66 8.2%	0 0.0%
Douglas & White Fir	617	21 3.4%	513 83.1%	83 13.5%	24 3.9%	526 85.3%	67 10.9%
Ponderosa Pine	440	0 0.0%	39 8.9%	401 91.1%	0 0.0%	43 9.6%	397 90.2%
Percent of "grouped" cases correctly classified			88.19%			89.26%	
Actual Deciduous Groups		Alpine Willow	Aspen	Oak	Alpine Willow	Aspen	Oak
Alpine Willow	232	219 94.4%	13 5.6%	0 0.0%	213 91.8%	19 8.2%	0 0.0%
Aspen	432	51 11.8%	306 70.8%	75 17.4%	36 8.3%	373 86.3%	23 5.3%
Oak	111	0 0.0%	1 0.9%	110 99.1%	0 0.0%	8 7.2%	103 92.8%
Percent of "grouped" cases correctly classified			81.94%			88.90%	
Actual Herbaceous Groups		Grass	Meadow	Tundra	Grass	Meadow	Tundra
Grass	99	99 100%	0 0.0%	0 0.0%	99 100%	0 0.0%	0 0.0%
Meadow	108	17 15.7%	87 80.6%	4 3.7%	17 15.7%	81 75.0%	10 9.3%
Tundra	787	0 0.0%	30 3.8%	757 96.2%	0 0.0%	19 2.4%	768 97.6%
Percent of "grouped" cases correctly classified			94.87%			95.37%	

[1]Number and percentage of sample points in each group assigned to each class.

middle elevation ranges (Douglas-fir/white fir, aspen, and meadow) were classified less accurately than the other two classes in the group, mainly because each middle class is flanked by transitional zones. To help refine the training procedure, the sample size for each class was used as a weight in the classification, a Baysian-type classifier. These results for the weighted classification are shown on the righthand side of Table 6. In each category the overall classifications are slightly improved, although individual species classification accuracies were decreased in some cases. The lowest percentage for any species was raised from 70.8% (aspen) in the unweighted classification to 75% (meadow) in the weighted classification with most percentages tending to be much closer to the average of 90%.

Care must be taken when interpreting these results since they indicate only the ability to distinguish among the various species within each of the major categories, using topographic data alone. Also it must be emphasized that the points classified in this test were solely training data. Therefore, it is not expected that these accuracy figures would be representative of those obtained in the final classifications when Level III categories and species would be differentiated using both spectral and topographic data. The results do indicate, however, that the topographic data should be helpful in distinguishing among the various species (Level III) within one category (Level II) -- a differentiation which could not be accurately accomplished using spectral data alone (Hoffer et al., 1975a).

D. Techniques for Using the Topographic Distribution Model in Conjunction with Spectral Data.

Prior work in analysis of satellite multispectral scanner data revealed that the simple addition of topographic channels to the spectral data vector for each pixel and multivariate analysis of the augmented data set did not result in consistent, significant improvements in classification accuracy (Hoffer et al., 1975b). A major part of the work completed under the current project has been the development and evaluation of a wide range of procedures for conducting analyses using topographic data as ancillary information in the analysis of spectral data. Various methods for developing training statistics are discussed first; these are followed by a discussion of several classification strategies.

1. Development of training statistics. In developing the training statistics for use in classifying a combination of spectral and topographic data, a primary consideration is that two very different types of data are involved in the analysis: spectral data and topographic data. Consideration must be given to the development of one set of training statistics that is appropriate for the spectral data, and perhaps an independent set of training statistics must be developed which is appropriate for the topographic data. In previous work with the Skylab MSS data (Hoffer et al., 1975b), the modified cluster technique had been used to develop the training statistics. Although these training statistics did characterize the spectral characteristics of the various cover types present, they did not define their topographic characteristics. The result was that the addition of topograhic data increased classification accuracies for some of the cover types but lowered them for some others. Therefore, it was determined that for the current study two sets of training statistics would need to be developed, one to define the spectral reflectance characteristics of the cover types and one to define the topographic characteristics of the cover types. Two different techniques for creating the training statistics have been developed and tested in the current study. These will be referred to as the "MCB" (Multi-Cluster Blocks) and the "TSRS" (Topographic Stratified Random Sample) techniques.

The MCB technique[1/] (Fleming and Hoffer, 1978) was used to develop the spectral training statistics only. 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. In this study, two such blocks were defined in each of the seven training quadrangles. Each block was approximately 40 x 40 pixels in size and contained a diversity of spectral characteristics. (Careful selection of these training blocks enables spectral data characterizing all cover types in the entire area to be included in the training sample.) Each training block was clustered independently into 16 spectral classes, and the cover type associated with each cluster class was identified. Spectrally similar cluster classes for the different training blocks were then pooled to produce a single set of training statistics to describe the spectral characteristics of the cover types present.

[1/] Previously referred to as the modified cluster technique, Fleming et al., 1975b.

In this study, there were fifteen spectral classes in the final set of training statistics.

Previous work with Landsat and Skylab data had indicated that it was very difficult to get an effective set of training statistics that would spectrally discriminate among some of the individual forest cover types present in the San Juan Mountains. This was particularly true for spruce/fir and Douglas/white fir. Preliminary analysis of the data used in this study again verified that training statistics could not be developed on the basis of spectral data alone which would reliably separate individual forest cover types in this area of complex topographic and vegetative characteristics. Therefore, the MCB technique was used to define spectral training statistics only for major cover types (i.e., Level II groupings). It was hypothesized, however, that if the various species and forest cover types were 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 separate and classify individual forest cover types.

A TSRS (Topographic Stratified Random Sample) technique had been used earlier in the study to develop the Topographic Distribution Model, which characterized the topographic distribution of the various cover types. Since the cover type as well as topographic characteristics of each of the pixels used to develop the model had been determined, it was apparent that the same data set could be used to define the topographic training statistics.[1/]

As previously described, the TSRS (Topographic Stratified Random Sample) technique involved stratifying the site into numerous (i.e., 91) topographic positions, followed by selecting and identifying an equal-size sample of single pixels (i.e., 50) from each strata. The result was a statistically valid sample which described, quantitatively, the topographic

---

[1/] Whereas spectral training statistics quantitatively describe the spectral characteristics of each cover type of interest, topographic training statistics were required to quantitatively describe the topographic characteristics of each cover type. Therefore, development of the topographic training statistics required input data which would statistically describe the various cover types as a function of their topographic position rather than their spectral response.

distribution of each vegetation type.

In addition to using the TSRS technique to define topographic training statistics, it was also apparent that this technique could be used to define spectral training statistics, since the cover type of each of the pixels sampled had been determined. Because the data had been stratified by topographic position, every slope and aspect combination was represented for each vegetation type. Therefore, all variations in spectral response for each cover type due to topography should have been represented. 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 area.) Although the TSRS approach requires considerable effort to identify the location and the cover type associated with the relatively large sample of pixels, once the data set is developed, it can be used to generate training statistics for both the topographic data and the spectral data.

It should be noted that any statistically defined random sample of data points could have been used to provide a set of X-Y coordinates for developing the spectral characteristics of the individual forest cover types. Since the topographic distribution model data was available and had been developed using an appropriate statistical sampling design, use of this data set eliminated the requirement to photo-interpret and field check an additional set of X-Y coordinates which would be used only for developing the training statistics.

A key point to note is that 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 it determines the natural groupings of spectral characteristics in the study site.

2. Approaches to classifying combined spectral and topographic data. The objective of the classification step is to integrate the spectral and topographic distributions into a logical classification sequence. Once the statistical distributions (training statistics) have been developed, the classification of the data set can be accomplished by any one of several

different approaches. The major difficulty encountered is that the spectral classes and topographic classes do not necessarily match the information classes. In other words, there is not always one topographic class and one spectral class for each information class. (Informational classes could be considered to be individual forest cover types, for example.) The purpose of the classification step is to logically combine the spectral and topographic data to define the desired informational classes.

The classification procedure can vary in several ways depending on the mathematics, the logic, and the type of data used by the algorithm. In this study, two basic types of algorithms were used: single-stage and multi-stage (or layered) classifiers. Both are maximum likelihood per-point classifiers which differ only in the logic for making the classification decisions. The single-stage classifier is the commonly used "standard" LARSYS algorithm known as *CLASSIFYPOINTS (Phillips, 1973). The layered classifier has been developed over the last several years at LARS by Wu, Swain, Landgrebe, and Hauska (Wu et al., 1974; Swain et al., 1975).

In addition to the different procedures that can be used in the classification process, different combinations of data can be used by the classification algorithm. Although many combinations are possible, the three major variations compared in this study are: (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 "baseline" classification, against which the other classifications were compared. It was anticipated that results from the "baseline" 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 classification would indicate the maximum accuracy achievable when using Landsat spectral data and DMA topographic data.

To combine (a) the two different methods for defining training statistics (MCB and TSRS), (b) the two types of training statistics (spectral and topographic), and (c) the two different classification algorithms (single-stage and layered), the current study defined and evaluated two different analysis

techniques. These are referred to as "Analysis Technique A" and "Analysis Technique B."

Analysis Technique A involved spectral training statistics developed using the MCB approach and topographic training statistics obtained using the TSRS approach. A layered (or multi-stage) classifier was then used in the classification. The first stage of the classification utilized only the spectral data and classified only Level II cover types; it was followed by the second stage in which only the topographic data was used in attempting to identify Level III Forest Cover Types.

Analysis Technique B used topographic training statistics obtained by the TSRS approach and spectral statistics obtained from the same set of X-Y coordinates. A single-stage classification involving both the spectral and topographic training statistics was used in attempting to identify Level III Forest Cover Types.

Both the layered and single-stage classifiers are capable of using weighting factors in the classification process, thereby resulting in a Baysian type of classifier. For purposes of this study, we therefore evaluated (a) the technique used to combine the spectral and topographic data (i.e., Analysis Techniques A and B), (b) the use of weights in the classification step and (c) the impact of using the topographic data in addition to the spectral data. Thus, results obtained from twelve different classification sequences were compared: (1) layered classification, with (a) equal and (b) weighted probabilities of occurrence, and (2) single-stage classification, with (a) equal and (b) weighted probabilities of occurrence. Each of these four combinations was classified using (a) only spectral data, (b) spectral plus elevation data, and (c) spectral plus elevation, slope, and aspect data. By way of a summary, Table 7 shows a matrix of the possible data combinations and classification procedures that were involved in this study. For ease in communication, a classification number was assigned to each combination, as shown in the table.

As indicated in the above discussion, Analysis Technique A uses the layered classifier. A critical aspect in using the layered classifier is the development of the decision tree. For this study, the decision tree which is based upon use of both spectral and topographic data by the layered

Table 7. Numerical designations of various classification procedures used.

Type of Data Used	Analysis Technique: A[1/] Equal Probabilities[3/]	A[1/] Weighted	B[2/] Equal Probabilities	B[2/] Weighted
Spectral Only (i.e., "Baseline classifications")	1	2	3	4
Spectral and Elevation only	5	6	7	8
Spectral and Topographic (i.e., Elevation & Aspect & Slope	9	10	11	12

[1/] Analysis Technique A utilizes spectral statistics derived using the Multi-Cluster Blocks (MCB) procedure and topographic statistics obtained using the Topographic Stratified Random Sample (TSRS) procedure. The classifications involve the Layered Classifier in which the first stage utilizes only spectral data and classifies only Level II cover types and the second stage utilizes only topographic data in attempting to identify Level III Forest Cover Types.

[2/] Analysis Technique B utilizes topographic statistics obtained using the TSRS procedure and spectral statistics are obtained from the same set of pixels (matching training pixel locations). The classifications involve a singel-step classification based on both the spectral and topographic training statistics (if called for) to identify Level III Forest Cover Types.

[3/] Equal (i.e., unweighted) or weighted probabilities of occurrence of the individual cover types. Weights were generated using data derived from the Topographic Distribution Model.

classifier, is shown in Figure 18. The input data tape contains both spectral and topographic data in an overlayed "database" format. At each decision node the type of data that provides the best information for making a particular decision can be used. For example, as indicated by Figure 18, spectral data alone was used to distinguish between the five primary ground cover types present in the study area: coniferous forest, deciduous forest, herbaceous cover, barren areas, and water. Based upon the discriminant analysis results using only topographic data to separate individual forest cover types (shown in Table 6), it was thought that the sub-division of the major forest cover types into the individual forest cover types present in the study area could be effectively accomplished through the use of topographic data alone.

As previously mentioned, the Multi-Cluster Blocks approach was utilized to develop the spectral training statistics for the layered classifier. The cluster classes from the fourteen training blocks defined by the analyst were pooled, resulting in the final set of fifteen spectral/informational classes. As indicated in Figure 18, five of these spectral classes were identified as belonging to the coniferous forest cover type, three as deciduous, three as herbaceous, three as barren, and one as water.

The next stage in the multi-stage classification involved the utilization of topographic data to divide the major cover types into individual forest cover types. In this sequence, an attempt was made to define specific cover type groupings which had been identified by the U.S. Forest Service as being of interest. Therefore, five individual forest cover types were distinguished within the coniferous group, three within the deciduous forest cover type, and two herbaceous classes were separated. The alpine willow class was combined with the tundra class to identify a single group which should be identified as alpine. In developing the decision tree, it was also found that some areas in the spruce-fir cover type were being misclassified as water on the basis of spectral data alone. Since water would logically occur only in topographic positions having a 0% slope, at the second level in the decision tree, pixels which had initially been classified as water were divided into two groups, based upon slope. If the slope was 0% in the topographic data, the pixel remained identified as water, but if it was not 0%, the pixel classification was changed to spruce-fir. These examples indicate how the spectral and topographic data can be used <u>in combination</u> to more

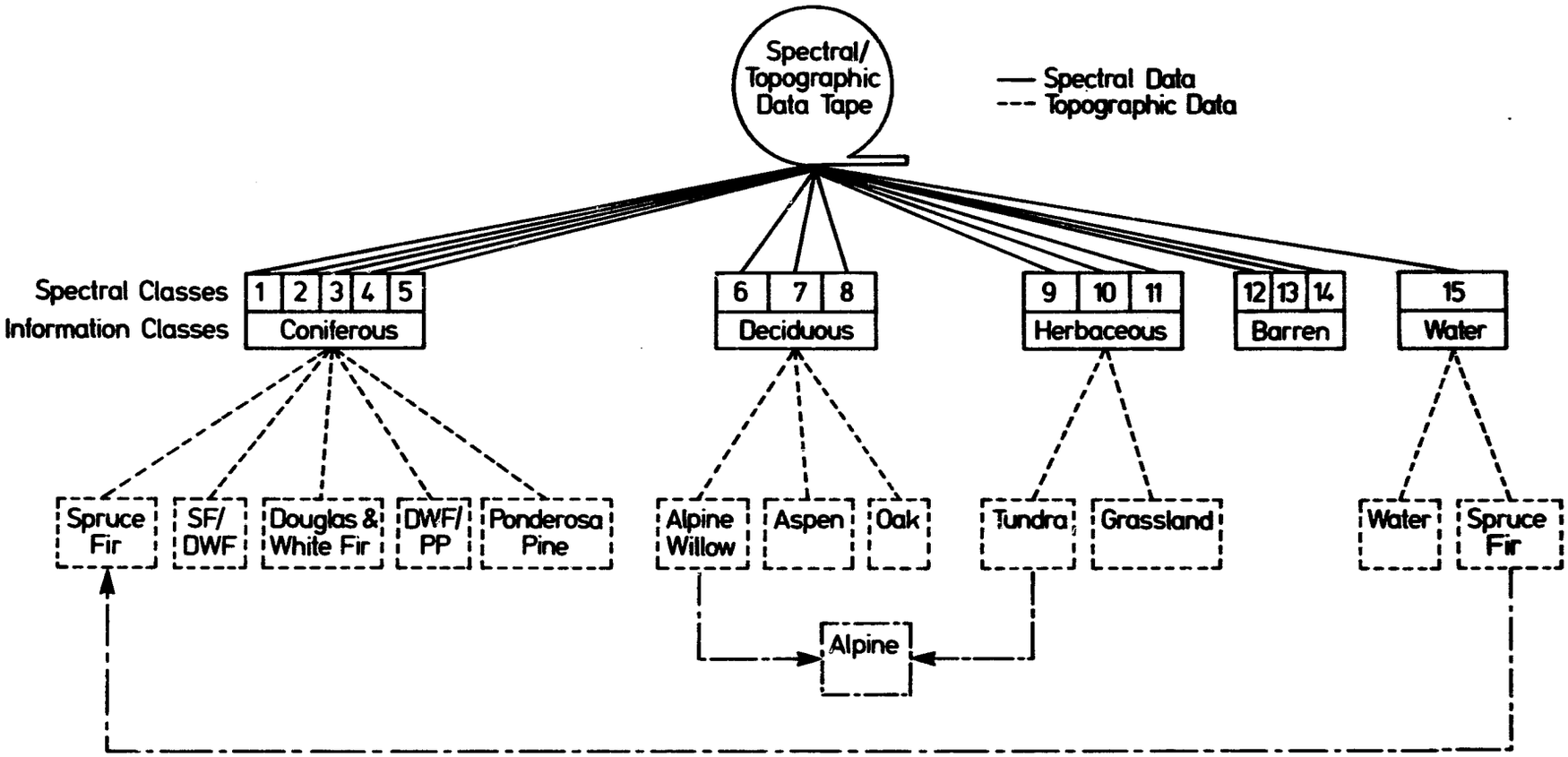


Figure 18. Decision tree used with the layered classifier for combining spectral and topographic data.

effectively identify individual forest cover type groups of interest to the Forest Service. One could attempt to differentiate among crown closure percentages of a single species as a third stage in a layered classification process, using spectral information alone, but that level of detail was not attempted in this phase of the project.

It should also be noted that a major programming effort was necessary in this part of the project in order to refine the layered classifier so that the algorithm would accept both spectral and topographic training statistics decks and also would allow weights to be used in the classification.

3. Evaluation procedures for comparing the classification results. In order to compare quantitatively the results of the various classifications and strategies, a test data set was developed. Because of the complexity of the test site and the ensuing complexity of statistical sampling procedures, it was decided that the best approach would be to use individual Landsat pixels for the test data set. In order to estimate the overall classification performance for each quadrangle within $\pm$ 5% at the 95% confidence level, 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 as a line-printer output. 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 of the test quadrangles, and the tentative identifications were compared to the aerial photos. Next, the July 1978 field trip was used to locate and field check as many of the test pixels as possible; 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 1:24,000 printouts of each quadrangle on which the test pixels had been plotted with the color infrared photos of the same area. The X-Y coordinates of each test pixel were

then located on the aerial photo and interpreted. Stand density as well as cover type were recorded for each of the test pixels. During the photointerpretation process, pixels which were located too close to borders between two cover types to allow positive identification were excluded from the test pixel data set (i.e., the population being considered), thus reducing the inference space of the accuracy estimates. Also pixels that fell on clouds or cloud shadows on the aerial photography were excluded. This resulted in a decrease from the 2100 potential test pixels to 1539 pixels actually defined as the test data set. This test sample size was still sufficient to achieve a $\pm$ 5% error of estimate at the 95% confidence level in evaluating the overall classification accuracy by quadrangle.

IV. REFLECTANCE GEOMETRY CORRECTION MODEL APPROACH

The spectral response of forest cover types in areas of rugged terrain is influenced not only by the intrinsic spectral properties (reflectivity) of the cover types themselves, but to a large extent by their relative topographic position. The relative topographic position of the terrain with respect to the position of the sun and the location of the sensor system affects the spectral response of forest cover types as the result of differential insolation rates. A simple observation of a Landsat scene collected over rugged terrain clearly reveals the effects of differential insolation rates. The surfaces directed towards the sun appear lighter in tone and those directed away from the sun appear darker in tone up to a critical position in which a complete topographic shadow is produced.

In order to use the spectral information contained in the Landsat data to discriminate more accurately among different ground cover types, it would be desirable to eliminate, as much as possible, the topographic effects which cause differential insolation rates. In other words, one should normalize the relative spectral response of each Landsat spatial resolution element (pixel) to that of a flat surface. To accomplish this normalization, a simplified model of the reflection geometry was used to calculate the appropriate correction factor for each Landsat pixel.

A. Development of the Correction Model

The intensity of radiation reflected by a surface at a given wavelength (R_λ) is essentially controlled by two factors, i.e., the intrinsic spectral reflectance characteristics of the surface (ρ_λ) and the amount of incoming spectral radiation flux (Q), which is defined as the total amount of radiant energy at a given wavelength (or wavelength band) that crosses a unit area of an intercepting surface per unit time. This functional relationship is expressed by Equation (1).

$$R_\lambda = f(\rho_\lambda Q) \qquad \text{Equation (1)}$$

In remote sensing applications, the ideal situation would exist when the reflected radiation is a function of the spectral characteristics of

the cover type only, that is, when the term Q in equation (1) is a constant. This situation could be reached only by the normalization of the incoming radiation, and thus equation (1) would become:

$$R_\lambda = Q[f(\rho_\lambda)] \qquad \text{Equation (2)}$$

where Q would be a constant for all pixels in the scene, regardless of the topographic position of the different ground cover types. Therefore, let us consider now in detail the term Q, which is for all practical purposes entirely composed of the solar radiation intercepted by the earth. At a given instant in time and at a particular geographic location, the amount of solar radiation incident on the top of the atmosphere (Q_s) can be described as a function of the solar azimuth angle and the solar zenith angle. The solar radiation intercepted by the earth-atmosphere system will be either absorbed or returned to space by scattering and/or reflection. According to Sellers (1972), the disposition of the solar radiation Q_s can be described by the following mathematical expression:

$$Q_s = C_r + A_r + (Q + q)\rho + C_a + A_a + (Q + q)(1 - \rho) \qquad \text{Equation (3)}$$

where C_r = radiation reflected and/or scattered by clouds

A_r = radiation reflected and/or scattered by air molecules, dust, or water vapor

C_a = solar radiation absorbed by clouds

A_a = solar radiation absorbed by air molecules, dust, or water vapor

Q = direct incoming solar radiation

q = diffuse incoming solar radiation

ρ = reflectance characteristics of the surface

This expression indicates that the solar radiation incident on a horizontal surface at the top of the atmosphere can be reflected and scattered back to space by clouds (C_r), by dry air molecules, dust and water vapor (A_r) or by

the earth's surface ($(Q + q)\rho$), where Q and q are respectively the direct and diffuse solar radiation incident on a horizontal area at the ground and ρ is the reflectance characteristic of the surface. Alternatively, this solar radiation can be absorbed by clouds (C_a), by dry air molecules, dust and water vapor (A_a), or by the earth's surface ($(Q + q)(1 - \rho)$).

Since the satellite multispectral scanner system will record the intensity of radiation being reflected by the "earth-atmosphere" system, all the absorption terms in Equation (3) can be deleted. Similarly, since the Landsat scene utilized in this study is cloud-free, the term C_r in Equation (3) is therefore zero. Thus, Equation (3) is greatly simplified and can be rewritten as follows:

$$R = A_r + (Q + q)\rho$$
$$= (A_r + q\rho) + Q\rho \qquad \text{Equation (4)}$$

where R is the amount of energy reflected by a surface on the earth.

In this study, the term $(A_r + q\rho)$ in Equation (4) will be considered constant and relatively small in comparison to the term $Q\rho$. In practice it will be calculated from areas within complete topographic shadow, that is, in areas with absence of direct incoming solar radiation. This term is, in essence, the amount of radiation reflected or scattered by the atmosphere plus the diffuse radiation being reflected by the target surface. This amount of non-direct radiation is subtracted from the actual radiation measured by the Landsat scanner before the correction for topographic effects is applied to the direct reflectance of each spatial resolution element (pixel) in the scene.

To simplify the model and consequently the calculations of the reflectance geometry correction coefficients, the reflecting features on the earth's surface will be assumed to be perfectly diffusing surfaces (Lambertian surfaces). Struve et al. (1977), who have also studied the effects of topography on the spectral response of earth-surface features, have indicated that this assumption is not completely valid; they stated, however, that in the absence of definitive data, non-Lambertian effects

could not be included in their calculations. Therefore, the reflectance characteristics of earth-surface features can be simplified from this complex, non-Lambertian expression:

$$\rho = f(\lambda, \alpha, \beta) \qquad \text{Equation (5)}$$

where α and β are respectively the angles between the position of the sun and the position of the sensor system with respect to the normal to the reflecting surface, to a more simple, Lambertian expression:

$$\rho = f(\lambda) \qquad \text{Equation (6)}$$

which indicates that the reflectance characteristic ρ is a function of wavelength only. This simplification is very important in that it allows one to disregard the position of the sensor system (satellite position and look-angle). Therefore, from this point on, only the positions of the sun and that of the reflecting surfaces will be considered for developing the reflectance geometry corrections.

Since the amount of direct solar radiation reaching a surface and consequently reflected by the surface is a function of the position (slope and aspect) of the surface with respect to the direction of the incoming radiation, the intensity of the direct solar radiation illuminating surfaces sloping at different angles and oriented in different directions can be described by the cosine law of spherical trigonometry as illustrated by Figure 22 and described by Equation (7):

$$\begin{aligned} Q_i' &= Q_n' \cos Z' \\ &= Q_n'(\cos Z \cos i + \sin Z \sin i \cos (a - a')) \end{aligned} \qquad \text{Equation (7)}$$

where Q_i' is the amount of direct solar radiation reaching a given surface with slope i. Z is the solar zenith angle, a is the azimuth angle of the sun, and a' is the azimuth angle of the normal to the surface (aspect); Q_n is the intensity of direct solar radiation on a surface normal to the sun's rays, and Z' is the angle between the incident solar rays and the

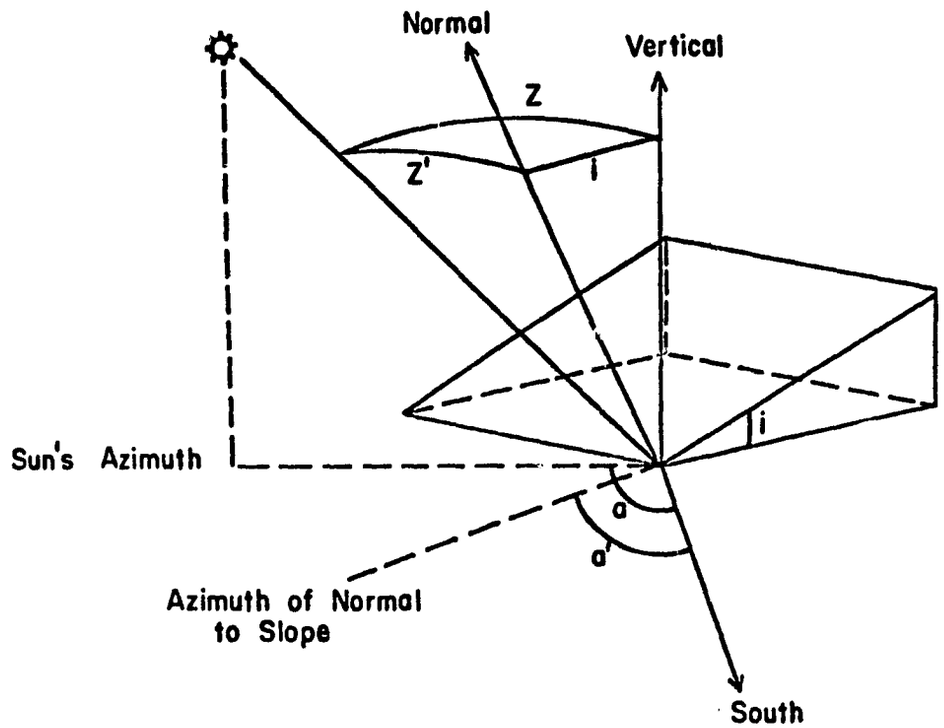


Figure 19. Relation of the solar zenith angle Z to the energy incident on a sloping surface. (After Sellers, 1972.)

perpendicular to the sloping surface. For the particular cases of representing the intensity of direct solar radiation on a horizontal surface (Q'_h) and on a vertical surface (Q'_v), Equation (7) can be simplified into the following expressions:

$$Q'_h = Q'_n \cos Z \qquad \text{Equation (8)}$$

$$Q'_v = Q'_n \sin Z \cos (a - a') \qquad \text{Equation (9)}$$

It is not difficult to see that Equations (8) and (9) are obtained directly from Equation (7) if one remembers that a horizontal surface has a slope angle (i) equal to zero and that a vertical surface has a slope angle (i) equal to 90^0, and finally that $\cos 0^0 = 1$, $\sin 0^0 = 0$, $\cos 90^0 = 0$, and $\sin 90^0 = 1$.

In order to normalize the direct incoming solar radiation reaching any surface on the ground to that of a horizontal surface (flat pixel), one has to substitute Q'_n of Equation (7) into Equation (8). This substitution yields Equation (10):

$$Q'_h = Q'_i \left(\frac{\cos Z}{\cos Z'}\right) \qquad \text{Equation (10)}$$

which simply gives the amount of radiation reaching a horizontal surface (Q'_h) as a function of the amount of direct radiation reaching a surface with a certain slope (i) and a certain aspect (a') times a correction (normalization) factor $\left(\frac{\cos Z}{\cos Z'}\right)$.

From Equations (2) and (4) it becomes evident that the reflectance of a surface is equal to the product of the incoming direct solar radiation Q times the intrinsic spectral characteristic (reflectivity, ρ) of the reflecting surface, i.e.:

$$R = Q \rho \qquad \text{Equation (11)}$$

And furthermore, the reflectance of a horizontal surface would be:

$$R_h = Q_h' \rho$$
$$= Q_i' \rho \left(\frac{\cos Z}{\cos Z'}\right) \qquad \text{Equation (12)}$$

where Q_i' in this case is the amount of direct incoming radiation reaching the surface, which in practice is essentially the total radiation measured by the Landsat scanner minus the background non-direct radiation ($A_r + q\rho$); ρ is the intrinsic spectral characteristic of the surface, and $\left(\frac{\cos Z}{\cos Z'}\right)$ is the correction factor needed for normalizing the reflectance of each Landsat spatial resolution element to that of a horizontal surface.

In the next section of this report, the actual calculations of the correction coefficients for the particular Landsat data set used in this study are described in detail.

B. Calculation of the Correction Coefficients

The primary aim of this part of the investigation is to normalize the spectral response values measured and recorded by the Landsat MSS system in order to eliminate the variations resulting from topographic effects. These altered spectral response values were then analyzed and classified using the LARSYS system, and the resulting classification compared with a baseline classification of the same but unaltered Landsat data set.

To calculate the radiation geometry correction coefficients $\left(\frac{\cos Z}{\cos Z'}\right)$ for the particular Landsat data set used in this study, Equation (7) had to be solved. This Landsat data set had been collected over an area located at a latitude of 37^{o} 35' and a longitude of 107^{o} 15', at 17 hours and 19.3 minutes Greenwich Mean Time (GMT). From this information one can calculate the solar zenith angle (Z) and the solar azimuth angle (a).

To calculate the solar zenith angle (Z), we first calculated the solar elevation angle (E), which is in essence the complementary angle of Z and is given by the following formula (from Doan and Sanford, 1970):

$$E = \sin^{-1}(\sin L \sin D + \cos L \cos D \cos H) \qquad \text{Equation (13)}$$

where D is the solar declination, L is the latitude, and H is the local hour angle (in degrees).

The solar declination D was obtained from the Nautical Almanac for the year 1973, which for this particular case is equal to 7^0 25.5' or 7.43^0; the latitude is known to be L = 37^0 35' or 37.58^0, and the local hour angle H = -27^0 18.3' or -27.308^0. Substituting these values in Equation (13) yielded the solar elevation E = 50^0 59.6' = 51^0. Since by definition E is the complementary angle of Z, then Z = 90 - E = 39^0.

Calculation of the solar azimuth angle (a) was also accomplished using the formula given by Doan and Sanford (1970):

$$a = \cos^{-1}(\sin D - \sin E \sin L/\cos E \cos L). \qquad \text{Equation (14)}$$

Substituting the appropriate values in Equation (14) yielded the solar azimuth angle a = 133^0 42.3' or 133.7^0.

Substituting Z = 39^0 and a = 133.7^0 in Equation (7), one could easily calcualte the differential insolation rate (which is described by cos Z') for any given slope (i) and any given aspect (a'). Table 8 shows the values of cos Z' for several combinations of slopes (from 0^0 - 85^0) and aspect angles (0^0 - 360^0).

To normalize the reflectance values of every spatial resolution element of the Landsat scene used in this study, a series of procedures were followed:

1. Calculate the correction coefficients ($\frac{\cos Z}{\cos Z'}$) for every pixel in the scene.
2. Calculate the non-direct reflectance (background radiation) from areas within complete shadows.
3. Subtract this non-direct reflectance (assumed constant for the entire test site) from the actual Landsat data values to obtain the direct reflectance values.
4. Multiply the direct reflectance value of each pixel by its appropriate correction factor.
5. This normalized (corrected) data set is then classified using the LARSYS processing and analysis techniques.

Table 8. Percent of direct radiation illuminating a horizontal surface.

	SLOPE																	
ASPECT	0.0	5.0	10.0	15.0	20.0	25.0	30.0	35.0	40.0	45.0	50.0	55.0	60.0	65.0	70.0	75.0	80.0	85.0
0	1.000	0.947	0.888	0.821	0.748	0.670	0.586	0.498	0.406	0.311	0.214	0.115	0.015	0.0	0.0	0.0	0.0	0.0
30	1.000	0.979	0.951	0.916	0.874	0.825	0.770	0.709	0.643	0.571	0.496	0.416	0.334	0.249	0.162	0.073	0.0	0.0
60	1.000	1.016	1.024	1.025	1.017	1.002	0.980	0.949	0.912	0.868	0.817	0.760	0.697	0.628	0.555	0.478	0.397	0.313
90	1.000	1.047	1.086	1.117	1.140	1.154	1.159	1.155	1.142	1.121	1.091	1.053	1.007	0.953	0.892	0.824	0.750	0.670
120	1.000	1.065	1.121	1.170	1.209	1.239	1.259	1.270	1.272	1.263	1.245	1.218	1.181	1.136	1.081	1.019	0.948	0.871
150	1.000	1.064	1.120	1.167	1.206	1.235	1.255	1.265	1.266	1.257	1.238	1.210	1.173	1.127	1.072	1.010	0.939	0.861
180	1.000	1.045	1.082	1.111	1.131	1.143	1.146	1.140	1.126	1.103	1.071	1.032	0.985	0.930	0.868	0.799	0.725	0.645
210	1.000	1.013	1.018	1.016	1.005	0.987	0.962	0.929	0.889	0.843	0.790	0.731	0.666	0.597	0.522	0.444	0.363	0.278
240	1.000	0.976	0.945	0.907	0.862	0.810	0.752	0.689	0.620	0.546	0.469	0.388	0.303	0.217	0.129	0.039	0.0	0.0
270	1.000	0.945	0.883	0.814	0.739	0.659	0.573	0.483	0.390	0.293	0.194	0.094	0.0	0.0	0.0	0.0	0.0	0.0
300	1.000	0.928	0.848	0.762	0.671	0.574	0.473	0.368	0.260	0.151	0.040	0.0	0.0	0.0	0.0	0.0	0.0	0.0
330	1.000	0.928	0.850	0.765	0.674	0.578	0.477	0.373	0.266	0.157	0.047	0.0	0.0	0.0	0.0	0.0	0.0	0.0
360	1.000	0.947	0.888	0.821	0.748	0.670	0.586	0.498	0.406	0.311	0.214	0.115	0.015	0.0	0.0	0.0	0.0	0.0

SOLAR ELEVATION = 51.00 SOLAR AZIMUTH = 133.71

V. RESULTS AND DISCUSSION

A. Results of Using the Topographic Distribution Model

Testing and evaluating the usefulness of the topographic distribution model involved a number of different classifications, as previously shown in Table 7. In this section, the classification results are examined and the major variables affecting these results are evaluated individually. These variables include:

1. level of detail of the classification;
2. method of developing training statistics and classifying the MSS data;
3. use of elevation data and use of topographic data (elevation + slope + aspect) in conjunction with spectral data;
4. use of a priori probability;
5. computer time required for each technique; and
6. use of different data sets for evaluation of classification results.

A brief description of the key elements of each classification is given in this section, followed by a more detailed discussion of the significance and implications of these results in Section V-B.

Initially, it is desirable to estimate the accuracy that can be obtained using only the spectral information to classify the area. Table 9 contains the "baseline" classification results for the test data set. In this classification, the Multi-Cluster Blocks approach was used for developing the training statistics that resulted in the definition of 15 spectral classes describing the spectral characteristics of the five major cover types (Level II) present in the test site. Previous work and further analysis of the spectral data during the current study indicate that there is considerable spectral similarity among individual forest cover types within the coniferous or deciduous categories. Conditions of stand density or topographic position can cause a similar spectral response to be obtained from different species, as well as distinct differences in spectral response from a single species or cover type. For these reasons, the MCB technique was used to classify only Level II major cover types when only spectral data was involved.

Table 9. "Baseline" classification test results and error matrix for Level II cover types, using the MCB technique.

(Classification No. 1[1/]: Spectral data only; Spectral training by Multi-Cluster Blocks; Equal weights; Only the first stage of the Layered classification sequence was involved, resulting in a Level II degree of detail.)

			No. Samples Classified As:				
Cover Type	Sample Size	Percent Correctly Classified	Conif-erous	Decid-uous	Herba-ceous	Barren	Water
Coniferous	917	80.2	735	115	42	20	5
Deciduous	252	55.2	74	139	35	4	0
Herbaceous	279	51.6	35	77	144	23	0
Barren	86	46.5	10	5	31	40	0
Water	5	60.0	2	0	0	0	3
Total	1539						
Overall Performance[2/]		68.9%					

[1/] As indicated on Table 7.

[2/] Overall Performance = $\frac{\text{No. Correctly Classified Samples in Each Class}}{\text{Total No. Test Samples}} \times 100$, i.e.,

$$\frac{735 + 139 + 144 + 40 + 3}{1539} \times 100 = 68.9\%$$

As Table 9 indicates, the coniferous cover type was classified with approximately 80% accuracy. However, among the other cover types there was considerable spectral confusion. The fact that over 25% (74 samples) of the deciduous forest cover was classified as coniferous is attributable to the vegetative complexity of the area since many of the deciduous stands contain significant numbers of coniferous trees. It is suspected that topographic influences on the spectral response of the deciduous forest cover also contributed to the misclassifications into coniferous forest. The misclassification of coniferous forest cover into deciduous is believed due to the presence of aspen in many of the coniferous stands. The random selection of 200 test samples per quadrangle resulted in only five test samples in the water class, which is not an adequate sample to effectively evaluate the classification performance. Two of the five water pixels were incorrectly classified as coniferous forest, suggesting a spectral similarity between the water and the very low response of coniferous stands on northern aspects and in topographic shadows.

The overall performance of the "baseline" classification is 68.9%, considerably lower than the classification results for major cover types previously reported under the Landsat-1 investigation [Hoffer, 1975a]. It is believed that the difference in accuracy is largely the result of using a more statistically reliable method of defining the test data set in the current investigation, thereby minimizing the human bias which we now suspect may have been present in the development of the test data sets used during the earlier study. The overall accuracy reported here is comparable to that reported during the Landsat Follow-on investigation of the San Juan Mountain area for data obtained in early August (Krebs et al., 1976). That study also utilized a statistically defined set of test data for evaluating the Landsat classifications. Therefore, although the classification results shown on Table 9 appear low, they are generally similar to the previous classification performance figures when only spectral data obtained in August was used and when the evaluation was based on a statistically defined test data set. Use of data from earlier in the summer might have improved the classification performance, but most of the higher elevation areas were still snow covered in the data sets that were available.

Table 10 summarizes the classification results of the "baseline classi-

Table 10. "Baseline" classification test results and error matrix for Level III forest cover types.

(Classification No. 3[1]; Spectral data only; Based on Stratified Random Sample training; Equal weights; Single-stage classification.)

			Number of Samples Classified As:										
	Sample Size	Accuracy	SF	SF/DWF	DWF	DWF/PP	PP	Aspen	Oak	Alpine	Grass-land	Barren	Water
SF	313	70.9	169	53	30	10	20	2	5	5	13	6	0
SF/DWF	156	66.7	12	74	18	8	12	17	12	1	2	0	0
DWF	39	71.8	9	17	6	5	1	0	1	0	0	0	0
DWF/PP	144	47.2	16	45	17	4	47	2	11	1	1	0	0
PP	265	54.0	12	15	12	9	134	6	33	9	33	2	0
Aspen	110	21.8	4	22	4	0	18	24	28	7	3	0	0
Oak	97	33.0	2	6	0	1	13	13	32	8	21	1	0
Alpine	79	25.3	0	1	1	8	0	10	9	20	29	1	0
Grassland	245	38.0	0	4	1	2	31	26	41	39	93	8	0
Barren	86	26.7	1	2	0	0	11	0	1	17	31	23	0
Water	5	60.0	1	0	0	0	0	0	0	0	0	1	3
Total	1539												
Overall Performance		49.4%											

[1] As indicated on Table 7.

fication" for individual forest cover types using only spectral data. In this case, the Level-III degree of detail was achieved by developing training statistics using the TSRS approach. The cover types of the pixels initially used to develop the topographic distribution model had been identified, so spectral training statistics for each of the forest cover types could be calculated. This resulted in what was basically a "supervised" classification procedure, in which the analyst utilized a set of spectral statistics for individual forest cover types, even though such cover types could not be reliably separated on the basis of spectral response.

Table 11 summarizes the classification results shown on Table 10 at a Level-II degree of detail. This allows us to compare (in Tables 9 and 11) the impact of the two methods used for developing training statistics. The classifications summarized in Tables 9 and 11 both used only spectral data, but the former was based on the Multi-Cluster Blocks method of developing training statistics whereas the latter represents the Stratified Random Sample method of developing training statistics. It is worth noting that while the overall classification performance is slightly lower for the TSRS approach, the accuracy of some cover types is significantly lower with this approach. For example, there was a decrease of about 8% in classification accuracy for deciduous forest cover and of approximately 20% for the barren class when the training statistics were developed using the Stratified Random Sample approach.

Tables 12 and 13 contain some of the key results of the investigation and show the impact of adding elevation and topographic data to the spectral data to improve classification performance. Table 12 summarizes the results on a quadrangle-by-quadrangle basis, whereas Table 13 summarizes the results over all the quadrangles on the basis of the individual forest cover types. The results in both tables are based upon use of the Stratified Random Sample training approach. Both Table 12 and 13 show that the addition of elevation data to the spectral data improved the classification performance considerably (i.e., about 15%). However, the use of all topographic data (elevation + aspect + slope) did not improve the classification performance beyond that achieved using just the elevation and spectral data, although the classifications vary from quadrangle to quadrangle and from cover type to cover type.

Table 11. "Baseline" classification test results and error matrix for Level II cover types, using the TSRS technique.

(Classification No. 3; Spectral data only, based on Stratified Random Sample training; Equal weights; Single-stage classification.)

			No. Samples Classified As:				
	Sample Size	Accuracy	Conif-erous	Decid-uous	Herba-ceous	Barren	Water
Coniferous	917	82.3	755	93	61	8	0
Deciduous	252	49.6	78	125	47	2	0
Herbaceous	279	47.0	40	100	131	8	0
Barren	86	26.7	14	3	46	23	0
Water	5	60.0	1	0	0	1	3
Total	1539						
Overall Performance		67.4%					

Table 12. Classification test results showing impact of topographic data for Level III forest cover types, by quadrangle.

(Classifications 3, 7, and 11: Training by Stratified Random Sample; Equal weights; Single-stage classifier.)

		Percent Correct Classification of Test Pixels		
Quadrangle	Sample Size	Spectral Data Only	Spectral + Elevation Data	Spectral + Topographic Data
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
Total	1539			
Overall Performance		49.4%	65.6%	65.9%

Table 13. Classification test results showing impact of topographic data for Level III forest cover types, summarized over all quadrangles.

(Classifications 3, 7, and 11: Training by Stratified Random Sample; Equal weights; Single-stage classifier.)

Forest Cover Types	Sample Size	Percent Correct Classification of Test Pixels		
		Spectral Data Only	Spectral + Elevation Data	Spectral + Topographic Data
SF	313	70.9	88.2	88.5
SF/DWF	156	66.7	70.5	75.0
DWF	39	71.8	61.5	48.7
DWF/PP	144	47.2	68.1	72.9
PP	265	54.0	71.3	71.3
Aspen	110	21.8	39.1	35.5
Oak	97	33.0	46.4	39.2
Alpine	79	25.3	82.3	78.5
Grassland	245	38.0	47.3	51.4
Barren	86	26.7	41.9	37.2
Water	5	60.0	60.0	80.0
Total	1539			
Overall Performance		49.4%	65.3%	65.5%

An Analysis of Variance (ANOVA) was performed on the arcsin square root transformation of the data in Table 12 and in Table 13 (Landgrebe, 1976) and are summarized in Appendix F. The results indicated a significant difference between the different combinations of topographic data used, a significant difference between quadrangles and a significant difference between cover types. A Newman-Keuls multiple 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 only elevation or using all three topographic parameters. We can conclude that the use of elevation data in conjunction with spectral data significantly improves classification performance over that obtained using spectral data alone.

The difference in classification performances among the various quadrangles (Table 12) merits additional attention. When only spectral data is used, classification performance differs by over 20% (Finger Mesa vs. Weminuche quadrangles). The addition of elevation data increased the classification performance for tne Finger Mesa quadrangles by almost 30% but only by 7% for the Oakbrush quadrangle. However, there are still differences of as much as 19% between quadrangles when elevation data is used in conjunction with spectral data (Granite Peaks vs. Oakbrush). Thus, classifications carried out and evaluated over limited test areas may be significantly influenced by the vegetative and topographic characteristics of the particular area, and may not, in fact, be representative of the classification performance that can be expected over a larger geographic region. Therefore, a person should be cautious concerning the conclusions drawn from a classification obtained over a relatively limited geographic area, particularly in regions of complex topography and vegetative cover.

The classification results in Table 13 indicate that most of the individual cover types can be classified with a much higher degree of accuracy through the use of elevation data in conjunction with the spectral data. The only decrease is in the Douglas/white fir class, a result which may be due, in part to the rather small number of training and test pixels in this cover type. While the addition of slope and aspect to the spectral and elevation data brought improvement in some of the individual forest

cover types, it also caused a decrease in performance for others. One interesting point shown in Table 13 is that one of the test pixels for water that had been incorrectly classified when only the spectral data was used was correctly classified through the use of the topographic data in conjunction with the spectral data. When using spectral data alone, the sample had been classified as coniferous forest. Use of the slope data helped in the classification, since water does not occur on slopes other than 0%.

Tables 14 and 15 show the impact of using a priori probabilities of occurrence (i.e., weights) in conjunction with the classification. If no weights (a priori probabilities) are specified to the classification algorithm, each spectral class is considered to have an equal probability of occurrence. The use of weights allows the classifier to favor those classes known to cover a larger areal extent. Table 14 shows very little change in performance of individual quadrangles, and less than a 1% difference in overall classification performance between the weighted and unweighted test results. Table 15 shows that the use of the weights did have an effect on the classification performance for certain cover types. In particular, the use of a priori probabilities seems to cause some of the mixed deciduous and coniferous stands to be classified as coniferous rather than deciduous. Thus, the classification accuracy for both the spruce fir and the spruce fir/Douglas white fir categories were increased through the use of the a priori probabilities, largely at the expense of the aspen and oak cover types. The use of weights increases the classification performance for grassland rather significantly (i.e., about 9%), whereas the classification accuracy for the barren class was decreased. In general, the use of a priori probabilities tended to increase the classification for cover types that are found over extensive areas (larger weights). An ANOVA of the data in Table 14 and in Table 15 is summarized in Appendix F. The results indicated no significant difference between using a priori probabilities and using equal weights. However, significant difference between quadrangles and between cover types was again indicated. In summary, it does not appear that the use of weights is particularly beneficial in improving overall classification performance.

Tables 16 and 17 also contain key results from the investigation. They compare results from the two major procedures used for developing the training statistics and classifying the data. An ANOVA of the data in

Table 14. Classification test results showing impact of using *a priori* probabilities (i.e., weights) for Level III forest cover types, by quadrangle.

(Classifications 11 and 12: Spectral + Topographic Data[1/]; Training by Stratified Random Sample; Single-stage classification.)

Quadrangle	Sample Size	Percent Correct Classification of Test Pixels: Equal Weights (i.e., Unweighted)	Percent Correct Classification of Test Pixels: With Weights
Oakbrush	199	56.3	59.3
Finger Mesa	214	63.7	63.3
Granite Peaks	202	80.2	77.7
Pagosa Springs	237	63.9	63.1
Devil Mountain	233	65.7	61.8
Weminuche	212	74.1	74.1
Ludwig Mountain	242	55.4	53.7
Total	1539		
Overall Performance		65.5%	64.4%

[1/] Topographic Data = Elevation + Aspect + Slope

Table 15. Classification test results showing impact of *a priori* probabilities (i.e., weights) for Level III forest cover types, summarized over all quadrangles.

(Classifications 11 and 12: Spectral + Topographic data; Training by Stratified Random Sample; Single-stage classification.)

Forest Cover Type (Level III)	Sample Size	Percent Correct Classification of Test Pixels	
		Equal Weights (i.e., Unweighted)	With Weights
SF	313	88.5	93.0
SF/DWF	156	75.0	82.1
DWF	39	48.7	51.3
DWF/PP	144	72.9	51.4
PP	265	71.3	69.1
Aspen	110	35.5	25.5
Oak	97	39.2	28.9
Alpine	79	78.5	75.9
Grassland	245	51.4	60.0
Barren	86	37.2	33.7
Water	5	80.0	60.0
Total	1539		
Overall Performance		65.5%	64.4%

Table 16. Classification test results showing impact of training and classification procedures, for Level III forest cover types, by quadrangle.

(Classifications 9 and 11: Spectral + Topographic data; Equal weights.)

Quadrangle	Sample Size	Percent Correct Classification of Test Pixels	
		Analysis Technique A[1/]	Analysis Technique B[2/]
Oakbrush	199	57.3	56.3
Finger Mesa	214	64.5	63.7
Granite Peaks	202	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%

[1/] Spectral training statistics developed by the Multi-Cluster Blocks technique; Topographic training statistics developed using the Stratified Random Sampling approach; Layered classification.

[2/] Stratified Random Sample approach used to develop training statistics for both spectral and topographic data; Single-stage classification.

Table 17. Classification test results showing impact of training and classification procedures, for Level III forest cover types, summarized over all quadrangles.

(Classifications 9 and 11: Spectral + Topographic data; Equal weights.)

Forest Cover Type (Level III)	Sample Size	Percent Correct Classification of Test Pixels	
		Analysis Technique A[1/]	Analysis Technique B[2/]
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		
Overall Performance		63.6%	65.5%

[1/] Spectral training statistics developed by the Multi-Cluster Blocks technique; Topographic training statistics developed using the Stratified Random Sampling approach; Layered classification.

[2/] Stratified Random Sample approach used to develop training statistics for both spectral and topographic data; Single-stage classification.

Table 16 and in Table 17 are summarized in Appendix F. The results indicated no significant difference between the two analysis procedures, but again a significant difference between quadrangles and between cover types. Table 16 shows that for most quadrangles, similar classification performances were achieved with either method. The exception was Devil Mountain quad, which shows approximately a 12% decrease in classification performance with Analysis Technique A. This indicates that the training statistics developed using the Multi-Cluster Blocks approach may not be completely representative of the spectral characteristics of this particular quadrangle. The Stratified Random Sample approach to developing training statistics produced slightly less variation among quadrangles, thereby indicating the merit of using a topographically stratified random sample set of training data.

Table 17 indicates the differences in classification performance among the individual forest cover types for each analysis technique. In some cases, classification performance is considerably better using the Analysis Technique A (which is based on the Multi-Cluster Blocks approach for developing training statistics) and in other cases the reverse is true. Analysis Technique A did somewhat better for the deciduous forest and barren classes but did not do as well for the alpine cover types. Performance for coniferous forest cover types tended to vary considerably. Therefore, neither technique can be defined as "best" on the basis of the classification performance of this data set.

Table 18 indicates that the layered classification approach requires considerably less computer time than single-stage classifications. At any stage during the multi-stage classification relatively few spectral classes are involved, making the classification procedure relatively efficient. The layered classification, using either the elevation-plus-spectral or the topographic-plus-spectral data sets can be completed in about the same amount of CPU time as the single-stage classification using only the spectral data and random sample training statistics. The single-stage classification using spectral-plus-elevation data required considerably more CPU time, and the use of the spectral-plus-topographic data, even a larger amount of CPU time. Thus the efficiency of a multi-stage classifier could be important when classifying large geographic areas utilizing both topographic and spectral data, particularly since (as shown in Tables 16 and 17) the classi-

Table 18. Comparison of computer CPU time (in seconds) required for each classification.

Type of Data Used	Computer Time Required (seconds)[2/]			
	Analysis Technique A (Layered Classifier)		Analysis Technique B (Single-Stage Classifier)	
	Without Weights	With Weights	Without Weights	With Weights
Spectral Only	[1/]	[1/]	55.5	50.1
Spectral + Elevation	50.9	51.0	169.6	156.4
Spectral + Topographic	50.6	53.7	211.1	200.6

[1/] CPU times for these two classifications are not included in the comparison since they involved only Level II major cover types and only the first stage of the classification sequence.

[2/] CPU time in seconds for the classification only (i.e., time to develop the training statistics is not included in this comparison).

fication performance is similar for the two classifiers tested. Use of weights did not significantly change the amount of CPU time used by either of the classification approaches.

Table 19 represents an additional evaluation of the classification results beyond those originally planned. Since the initial classification performances based upon spectral data alone were somewhat lower than we had expected, we decided to evaluate the impact of using randomly selected, individual test pixels instead of analyst-selected test fields, particularly since the latter is such a commonly used procedure and was the method used in the Landsat-1 investigation (Fleming et al., 1975a). There are at least three possible reasons for differences in classification performance due to the method of selecting test data. First, in using the test field method, the analyst often tends to select relatively pure, homogeneous test areas to represent the various cover type classes, causing a bias in the test data set which generally results in a higher classification performance than may truly be representative of the entire study area. Second, the ability to define the precise location of individual Landsat test pixels on the aerial photography is critical and difficult. If there is a slight misregistration between the location of a pixel on Landsat data and on the aerial photos, it is possible for the photo-interpreter to incorrectly identify the one-acre cell as belonging to an adjacent category. For example, a slight shift in apparent location could cause a mixture of deciduous and coniferous forest to be identified as primarily aspen or, if slightly more conifer and less aspen occurred, as coniferous. These problems tend to be minimized when one identifies a larger block of cover types. By using a larger area, slight variations in stand composition on any one-acre cell are not as significant nor is the slight misregistration in the location of a particular individual pixel. In this study, we attempted to minimize such misregistration of individual pixels by requiring a one-pixel buffer strip of the same cover type to be present around the designated test pixel. The third reason involves possible misregistrations between the Landsat and topographic data and also the possible errors in designating the elevation, slope, and aspect of individual pixels. Because both spectral and topographic data were used in the forest cover type classifications, it would seem that an incorrect designation of the topographic characteristics of a particular

Table 19. Comparison of classification performance based on different test data sets, for Level III forest cover types.

(Classification No. 11: Training by Stratified Random Sample; Spectral + Topographic data; Equal weights; Single-stage classification.)

Forest Cover Type (Level III)	Random Pixels[1/] Test Sample			Manually Selected Fields[2/] Test Sample		
	No. Pixels	Percent Correctly Classified	Percent[3/] of Total Pixels	No. Pixels	Percent Correctly Classified	Percent of Total Pixels
SF	313	88.5	20.3	3152	97.3	23.8
SF/DWF	156	75.0	10.1	1737	85.8	13.1
DWF	39	48.7	2.5	240	76.3	1.8
DWF/PP	144	72.9	9.4	695	77.0	5.2
PP	265	71.3	17.2	2096	79.2	15.8
Aspen	110	35.5	7.1	979	56.6	7.4
Oak	97	39.2	6.3	633	39.2	4.8
Alpine	79	78.5	5.1	405	81.2	3.1
Grassland	245	51.4	15.9	2504	53.6	18.9
Barren	86	37.2	5.6	573	59.3	4.3
Water	5	80.0	0.3	248	58.1	1.9
Total	1,539		99.8%	13,262		99.9%
Overall Performance		65.5%			74.6%	

1/ Statistical sample, based on 300 samples per quadrangle, located using a table of random numbers. Points falling on a cover type boundary and therefore not belonging to any single cover type class were excluded from the sample.

2/ Test fields defined by the commonly used procedure of locating areas (i.e., fields) of individual cover types on aerial photos and/or type maps, and then locating the same area in the Landsat data prior to the classification. In using this procedure, the total number of pixels in each cover type should be limited to approximately the proportion of that cover type in the study area.

3/ Since a statistically defined random sampling procedure was used, these percentages should be fairly representative of the amount of each of these cover types throughout the test site.

pixel or a slight misregistration of the data sets could cause classification errors of the individual pixels.

Table 19 shows that there is a marked difference in the assessment of classification performance, depending on whether randomly selected test pixels or manually selected test areas are used. Overall performance is approximately 9% higher when test fields were used to evaluate the classification. The classification of oak was the same using both sets of test data, but for all other cover types, with the exception of water, the performance was higher when test fields were used. The water test areas indicated a classification performance of only 59%; the reason for this apparently poor classification performance is not clear and is being investigated. It is known, however, that the use of the topographic data is causing most of the classification errors in the water class, since these same water test fields had a 91.1% correct classification performance when only the spectral data was used.

It is also important to note from Table 19 that the percentage of pixels in each of the individual forest cover types was approximately the same for the two test data sets. The test field data provided a significantly larger number of pixels, but from a statistical standpoint, the sample size (i.e., the number of test locations--either pixels or fields) is much larger when using individual pixels. It was for purposes of achieving maximum statistical validity that the individual pixel test data set was initially defined and used for evaluating the classification results throughout this study.

Table 20 is a comparison of the classification performance of major cover types for 1) the training data, 2) the test data set based on randomly selected pixels, and 3) the test data set based on manually selected test fields. There are several key elements to be observed. First, the percentage of pixels used for developing the training statistics varies from one class to the next but these differences are not directly related to differences in the areal extent of the various cover types, since the sample was stratified by topographic position rather than cover type class. However, the number of test pixels belonging to each of the cover type classes should be a good representation of the actual areal distribution of cover types within the test quadrangles, since this test data set was obtained by randomly

Table 20. Comparison of Level II classification performances for training data and two different test data sets.

(Classification No. 11: Spectral + Topographic data; Training by Stratified Random Sample; Equal weights; Single-stage classification.)

Major Cover Type (Level II)	Training Data			Random Pixels Test Sample			Manually Selected Fields Test Sample		
	No. Pixels	Percent Correctly Classified	Percent of Total Pixels	No. Pixels	Percent Correctly Classified	Percent of Total Pixels	No. Pixels	Percent Correctly Classified	Percent of Total Pixels
Coniferous	1953	89.8	51.6	917	85.2	59.6	7920	93.9	59.7
Deciduous	297	49.2	7.8	207	43.0	13.4	1612	59.6	12.2
Alpine	950	86.1	25.1	79	78.5	5.1	405	81.2	3.0
Grassland	214	57.5	5.7	245	51.4	15.9	2504	53.6	18.9
Barren	364	53.0	9.6	86	37.2	5.6	573	59.3	4.3
Water	6	83.3	0.2	5	80.0	0.3	248	58.1	1.9
Total	3784		100%	1539		99.9%	13,262		100%
Overall Performance		80.3%			71.1%			79.6%	

selecting 300 pixels in each of the seven test quadrangles. (Border and indeterminate pixels were eliminated, thereby resulting in the final sample of approximately 200 samples per quadrangle.) Thus, one sees on Table 20 that approximately 25% of the training pixels were used to define the alpine cover type class, but only 5% of the single test pixels represent alpine. On the other hand, 7.8% of the training pixels were used to define the deciduous forest cover, whereas 13.4% of the single test pixels belong to the deciduous class. These differences between the training and test data sets may be due in part to the fact that a different set of quadrangles were used for developing the training and the test data sets.

The second key point to be observed on Table 20 is that, for both the random pixel and the test field approach for defining the test sample, the Level II overall classification performance is approximately 5% higher than was achieved for Level III (as shown on Table 19). Since alpine, grassland, barren and water classes were treated the same for both Level II and Level III, the difference between Level II and Level III overall classification is due only to the classification performances of the deciduous and coniferous forest cover types.

A third observation from Table 20 is that the test field results are higher than the training pixel results for both the coniferous and deciduous forest cover types, as well as for the barren class. Such a result may indicate some peculiarities in the training data set or may be indicative of a bias in the test field data set. This question is also being investigated further. We believe, however, that the use of test fields tends to give an upward bias to the classification results. Use of randomly defined test pixels is statistically much better than analyst-selected test fields, but use of individual pixels may have resulted in some errors in the test data set, as discussed above. In this case, the classification results based on individual pixels could be somewhat lower than is actually the case. Thus, the true classification performance is probably somewhere between the values obtained on the basis of test fields and individual test pixels.

B. Results of Using the Reflectance Geometry Correction Model

Another potential application of digital topographic data is to spectrally "correct" the MSS data to remove the variation in spectral response due to topography. Two classifications were compared to evaluate the "correction" model, one using the original MSS data and the second using the "corrected" MSS data. In both classifications the TSRS data were used to calculate the statistics for each cover type and the single-stage classifier was used to classify the spectral and topographic data. The classification results showing the effect of the reflectance geometry correction model on the Landsat data are summarized by quadrangle in Table 21 and by Level III cover type in Table 22.

An ANOVA of transformed (arcsin square root) data in both tables indicated that there is no significant difference between using the "corrected" data and using the "original" Landsat data. The ANOVA also indicated that there is a significant difference among quadrangles and a significant difference among cover types, which is consistent with the previous results. Although the results indicated no significant difference between the "corrected" and "original" data, there was a slight decrease in accuracy after the spectral data had been "corrected." The lack of improvement in classification accuracy was surprising, especially since a qualitative evaluation based upon analysis of grayscale images of the "corrected" and "original" data (Figure 20a and 20b) had indicated that the correction model seemed to remove the topographic variations in spectral response in the Landsat data. In some cases, differences in spectral response were apparent and seemed to indicate that the model was not working. But the variations which were apparently due to topography were actually differences in cover types, caused by topography. As a result, in some areas the correction model did not seem to be effective, but the variation in spectral response was caused by differences in cover type instead of topography. Overall, the variation in spectral response caused by topography does appear to be reduced by application of the reflectance geometry correction model. However, the quantitative evaluation of the classification accuracies did not indicate any improvement due to the "correction" model.

There are several possible reasons for this result. One reason seems to be caused by errors in the slope and aspect data for the individual pixels,

Table 21. Classification test results for Level III forest cover types, by quadrangle, showing impact of applying the Reflectance Geometry Correction Model to the Landsat data.

(Spectral + Topographic data; Training by Stratified Random sample; Equal weights; Single-stage classification.)

Quadrangle	Sample Size	Uncorrected (i.e., "Original") Landsat Data	"Corrected" Landsat Data[1/]
Oakbrush	199	56.3	56.3
Finger Mesa	214	63.7	65.6
Granite Peaks	202	80.2	82.2
Pagosa Springs	237	63.9	65.3
Devil Mountain	233	65.7	58.8
Weminuche	212	79.1	70.8
Ludwig Mountain	242	55.4	54.1
Total	1539		
Overall Performance		65.5%	64.1%

[1/] Landsat data "corrected" prior to the classification using the Reflectance Geometry Correction Model.

Table 22. Classification test results for Level III forest cover types, summarized over all quadrangles, showing impact of applying the Reflectance Geometry Correction Model to the Landsat data.

(Spectral + Topographic data; Training by Stratified Random sample; Equal weights; Single-stage classification.)

Forest Cover Type (Level III)	Sample Size	Uncorrected (i.e., "Original") Landsat Data	"Corrected"[1/] Landsat Data
SF	313	88.5	90.1
SF/DWF	156	75.0	66.0
DWF	39	48.7	51.3
DWF/PP	144	72.9	75.7
PP	265	71.3	70.9
Aspen	110	35.5	32.7
Oak	97	39.2	42.3
Alpine	79	78.5	75.9
Grassland	245	51.4	45.3
Barren	86	37.2	39.5
Water	5	80.0	60.0
Total	1539		
Overall Performance		65.5%	64.1%

[1/] Landsat data "corrected" prior to the classification using the Reflectance Geometry Correction Model.

Figure 20a. Uncorrected Landsat Band 7 (0.8-1.1 μm) imagery of the Vallecito Reservoir area in the San Juan Mountain test site.

Figure 20b. Corrected Landsat Band 7 (0.8-1.1 μm) imagery of the same area shown in Figure 20a. Differences between these two illustrations are due to the application of the Reflectance Geometry Correction Model.

caused by the characteristics of the DMA elevation data. The coarsely defined steps in elevation cause inaccuracy in the slope data, particularly in areas of lower relief. Also, the characteristics of the elevation data caused a high percentage of the data to have North, East, South, and West aspects. These problems with the topographic data were also evident in the quantitative results; quadrangles with the roughest terrain tended to be classified better with the "corrected" data and the quadrangles with minimal relief (i.e., higher probability of errors in the slope and aspect data) tended to be classified less accurately with the "corrected" data.

A second problem with the "correction" model is the violation of some of the model's assumptions, specifically the assumption of Lambertian reflecting surface and the assumption of no variation in spectral response due to indirect radiation. Both assumptions, particularly the former, are not completely valid and must be further evaluated.

C. Discussion of Classification Results

In order to summarize the results shown in Section V. A., Table 23 shows the overall classification performance for each of the 12 classifications previously described in Table 7. The footnotes in Table 23 indicate some of the key aspects of these classifications in terms of the training and classification techniques used.

The most significant result shown in Table 23 is the fact that the use of elevation data in addition to the spectral data causes a significant increase in classification performance at the Level-III degree of detail. The use of the aspect and slope data did not cause an increase in classification performance over that achieved by using just elevation plus spectral data. This latter result was somewhat surprising since aerial photos reveal that aspect often has a distinct influence on cover type characteristics. It is thought that the reason for the apparent lack of improvement in the classification when aspect and slope data are included is largely due to the quality of the topographic data. In evaluating these classification results, we did some further evaluations of the characteristics of the topographic data and found some interesting results. A key element affecting the DMA elevation data involves the digitization and interpolation process. In doing the digitization, the DMA used a 0.01-inch grid on the original 1:250,000 scale USGS topographic maps. This is a relatively fine grid, but when applied to maps having such a small scale it resulted in a cell size on the ground of 64 meters. In an area such as this test site in Colorado where there is a significant amount of topographic relief, the 200-ft. contour intervals are rather close in many places on the map. In the digitization process, if any part of a map cell falls on the contour line, the entire cell is then coded as having the elevation of that contour line. To define the elevation of cells falling between contour lines, an interpolation process was used by DMA in which the three closest cells to the point in question were used to define the elevation of the unknown point. This procedure caused additional cells to be "grouped" into 200-ft. contour levels. In evaluating the topographic data, a histogram was generated showing the number of digitized grid cells occurring at each elevation level throughout the test site (Figure 21). This histogram shows a significant increase in the number of cells at each 200-ft. or 64-meter contour line interval. In fact, approximately one-half of the total data set is defined as being on a 200 ft. contour!

Table 23. Summary for the matrix of twelve classifications using different analysis procedures.[1/]

Type of Data Used	Analysis Technique A[2/]		Analysis Technique B[3/]	
	Equal	Weighted[4/]	Equal	Weighted
Spectral Only	No. 1[5/]	No. 2	49.4%	42.9%
Spectral + Elevation	63.8%	62.2%	65.3%	64.5%
Spectral + Topographic (elevation, aspect and slope)	63.6%	59.1%	65.5%	64.4%

[1/] Compare to Figure 18. Figures shown in the table are the Overall Classification Performance values, based on Level III Forest Cover Type classifications.

[2/] Analysis Technique A utilized spectral statistics derived using Multi-Cluster Blocks Technique and topographic statistics obtained using the Topographic Stratified Random Sample technique (i.e., independent training data locations). The classifications involved the Layered Classifier in which the first stage utilized only spectral data and classified only Level II cover types and the second stage utilized only the topographic data to identify Level III Forest Cover Types.

[3/] Analysis Technique B utilized topographic statistics obtained using the Topographic Stratified Random Sample method, and spectral statistics obtained from the same set of X-Y coordinates (matching training pixel locations). The classifications involved a single-step classification using both the spectral and training statistics to identify Level III Forest Cover Types.

[4/] Equal or Weighted probability of occurrence of individual cover type classes.

[5/] Classification Nos. 1 and 2 involved only Level II (Major) Cover Types, since only the spectral data was used in the first stage of the Layered classification procedure, and individual forest cover type classes could not be effectively defined using only the spectral data. All other classifications (Nos. 3-12) did involve the Level III Forest Cover Type degree of detail. Overall classification performances were 68.9% and 69.9% for Classification Nos. 1 and 2, respectively.

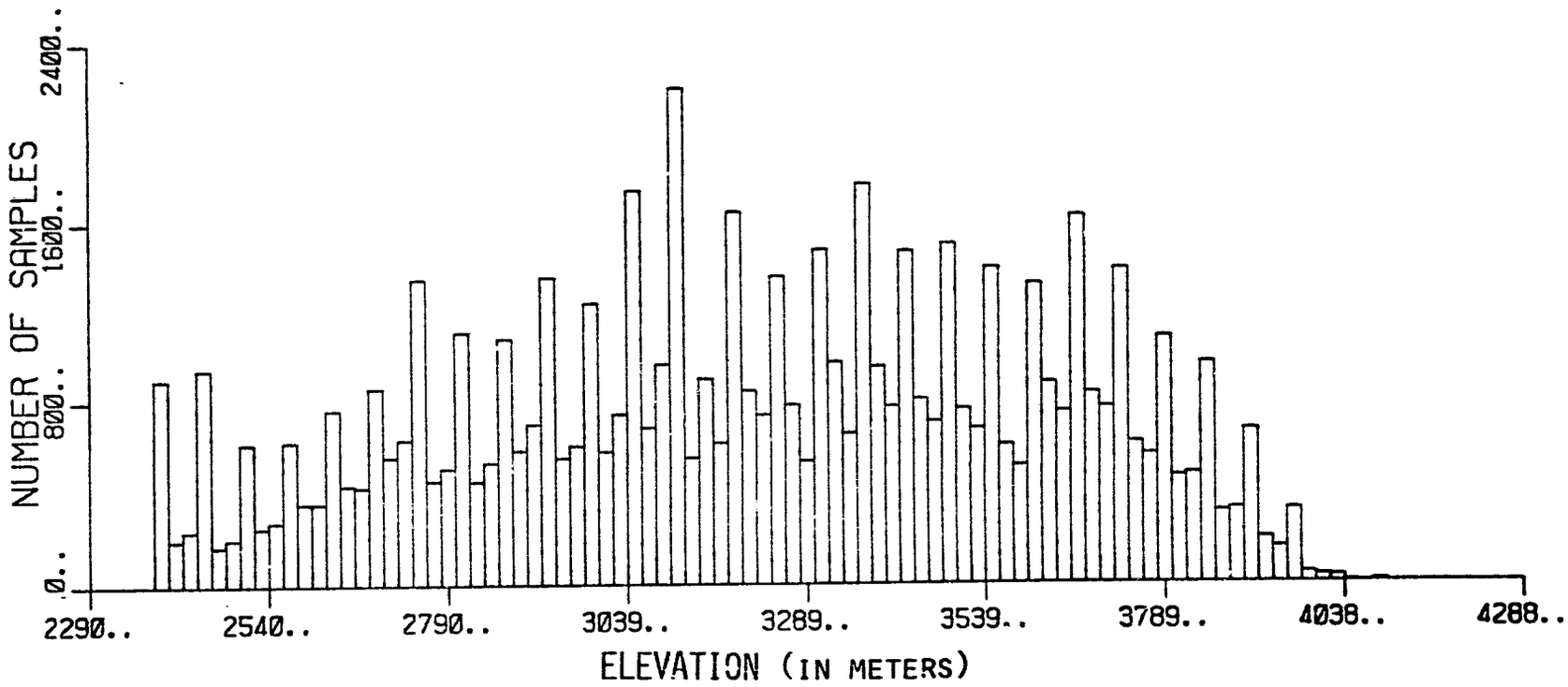


Figure 21. Histogram of the DMA elevation data for the San Juan Mountain test site area.

To further check the quality of the elevation data, 341 X-Y coordinates were randomly selected and located in the digital data. The locations of the pixels were then determined on 7½-minute USGS topographic quadrangle maps, and the elevation obtained from the 7½-minute topographic map was compared to the elevation given on the DMA data tape. Differences in elevation between the two data sets were tabulated and are shown in histogram format in Figure 22. As this figure indicates, the DMA data was within ± 25 meters of the elevation shown on the 1:24,000 scale maps for 128 out of the 341 pixels (37.5%), 217 pixels (63.6%) were within ± 50 meters, and 285 pixels (83.6%) were within ± 100 meters. It would therefore appear that the elevation data on the DMA data tapes compares reasonably well with the elevation defined for the same location on the 1:24,000 scale maps, and is generally adequate for the purpose of differentiating individual forest cover types.

The problem with the data occurs in the process of developing the aspect and slope data files from the digitized elevation data. In this process, a linear interpolation procedure is used. Because the digitized cell size is relatively large and there is considerable amount of topographic relief in the area, there are many areas where there may be only one or two cells falling between the cells located on the contour lines. In such situations the slope that is defined for those cells is much greater than is actually the case, and the aspect is courser than it should be. This causes the quality of the topograhic data to be relatively poor for many pixels. In summary, it would appear that the 1:250,000 scale DMA data should be primarily used for problems involving only elevation effects or where the topography is not as rugged. When slope and aspect data must be generated, larger scale maps should be used for digitizing the elevation data, if at all possible, using the smallest cell size possible. Where available, digital data obtained during the production of orthophotos could be utilized.

Another key aspect of the results shown on Table 23 is that the use of the weights (e.g., a priori probability of occurrence) did not significantly influence the overall classification performance. The reasons for this are not clear. For informational classes that are spectrally similar the use of an accurate set of a priori probabilities should have improved classification performance. However, since the training data were developed on one set of

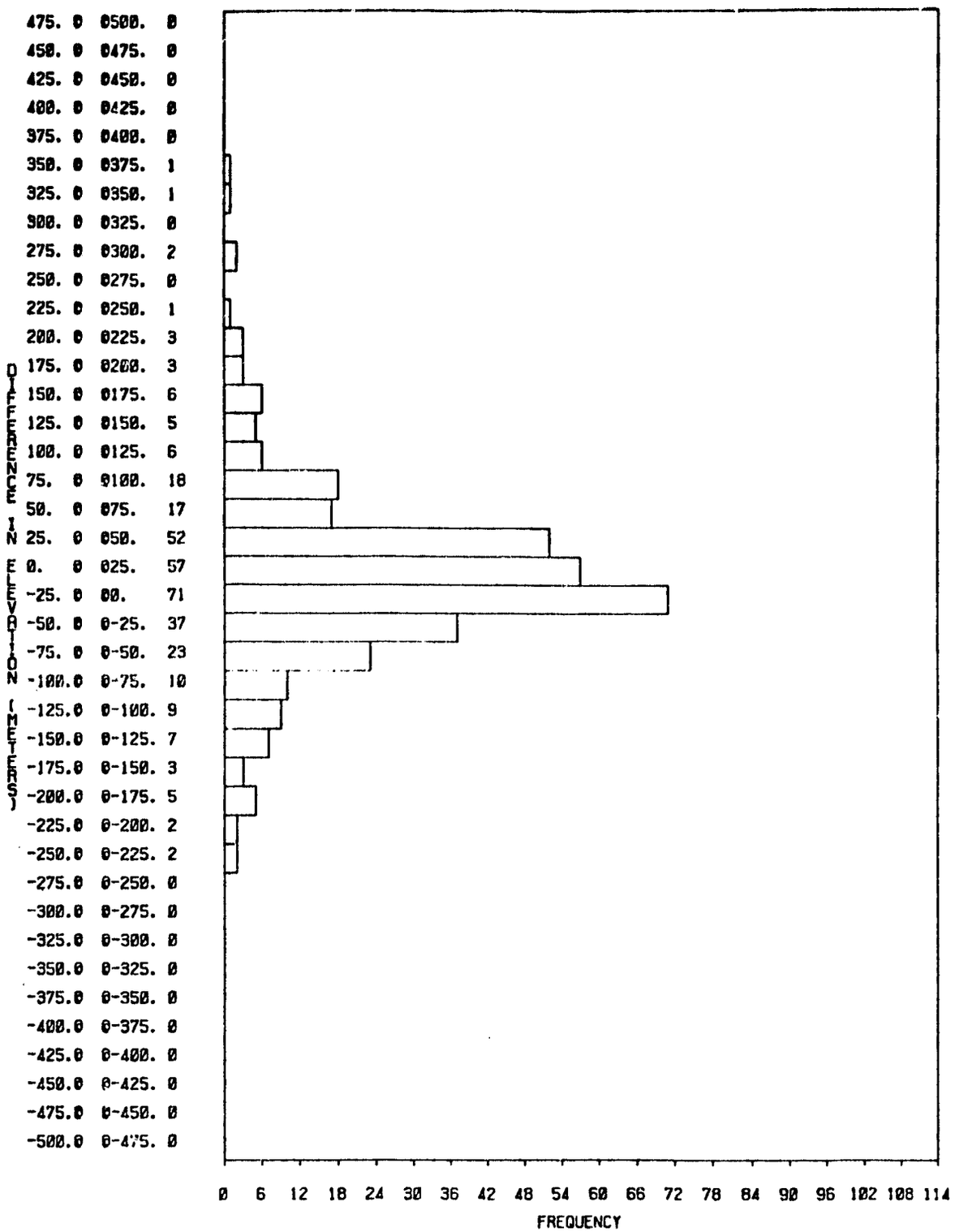


Figure 22. Histogram based upon a random sample of 341 pixels showing the difference between elevation in the DMA digital data and on USGS 7½' topographic maps.

quadrangles and the test data came from a different set of quadrangles, if there were significant differences in the areal extent of the cover types in the training quadrangles as compared to the test quadrangles, such differences could have affected the results. The overall results of the current study would indicated that, at least in some cases, the use of the a priori probabilities is not particularly effective. Based upon these results, it would appear that the time and effort involved in developing the a priori probability values and applying them to the computer processing procedure would not be warranted during the next phase of this study.

The other major result that is indicated in Table 23 involved the comparison between the training and classification techniques used to classify the data. Approximately the same overall classification performance was obtained using each technique, but each approach has some distinct advantages and limitations. Of particular importance is the fact that the Multi-Cluster Blocks procedure used in Analysis Technique A allows spectral training statistics to be developed that are based on the natural spectral groupings of the data. Previous work at LARS had indicated that the Multi-Cluster Blocks technique is the most effective approach for developing training statistics in terms of (a) analysis time, (b) computer classification time involved, and (c) classification performance achieved (Fleming, 1977). The Multi-Cluster Blocks technique is particularly useful in situations where the amount of reference data available at the beginning of the analysis is very limited. Thus, in developing the spectral training statistics, aerial photos obtained from a relatively few locations over the test site can be used to relate spectral cluster classes to the informational classes of interest. However, if knowledge concerning the cover type already exists for a statistically defined array of data cells (which is the situation for the GRIDS data set in the state of Washington and for U.S. Forest Service lands where Forest Survey plot locations have been defined and typed), one of the biggest limitations in the use of a statistical array of individual training data cells is covercome, and the amount of analyst time involved in developing the training statistics becomes more reasonable. The key to the potential effectiveness in utilizing such an existing data set for developing training statistics involves the type and quality of the information on each of the grid cell locations, and the ability to relate the

location of the points from which the existing information was obtained to the same location in the Landsat data. Thus, it would appear that the choice of methods for developing training statistics would largely be a function of the situation in which one is working, the amount of information that is available to the analyst when he is starting to develop his training statistics, and the geometric quality and characteristics of the data sets available. Depending on the type and characteristics of the data sets and information available, either the Multi-Cluster Blocks or a statistical sample of individual data cells would be appropriate.

The layered classification procedure offers the distinct advantage of computational efficiency as compared to a single-stage classification. A major factor of the layered classifier is that it is much simpler because one is dealing with fewer spectral classes at any individual step in the classification sequence. However, once a particular resolution element is classified into a major cover type category, it remains within that major cover type category in all subsequent levels of the classification tree. Therefore, the accuracy of the initial classificatior into major cover type groupings is very important.

The method of defining the test data set appears to have a significant impact on the quantitative evaluation of classification performance. The results in Table 20 appear to indicate that the use of manually selected test fields does cause a bias in the classification performance. However, as previously discussed, it is also possible that some of the individual test pixels were not correctly identified or may have topographic characteristics that were in error. Therefore, in considering the complexity of the forest cover (stand size, density variations, and composition), and also the geometric complexity of the data set (geometric correction of Landsat data, plus a 90^{0} rotation and overlay of the DMA data), and the coarse resolution and other characteristics of the DMA data, it would seem that in the future a statistically defined set of test areas rather than individual pixels would provide the best test data set to use for quantitative evaluation of the classification results. However, for statistical purposes, each test area would need to be treated as though it were a single pixel regardless of the actual number of pixels present within the designated area. It seems

clear that additional work is needed in defining effective and statistically valid methods of defining test data sets.

The application of the Reflectance Geometry Correction Model to the Landsat data does appear to be an effective approach for reducing the spectral variability caused by topography. Evaluation of map display outputs of the corrected and uncorrected data indicates that a significant amount of topographic variation was removed through the application of the correction model. In many areas, topographic characteristics still appear to be evident, but much of this appearance is due to the difference in cover type classes which occur in different topographic positions (e.g., North vs. South aspects). However, the quantitative classification results did not show an improvement in classification performance when the data had been corrected using the Reflectance Geometry Correction Model. It is believed that the difference between the qualitative evaluation of the uncorrected and corrected data sets and the quantitative evaluation of the test data set reflects some of the problems that may exist in the slope and aspect values for individual pixels in the digital topographic data. Therefore, we feel further work should be done with this Reflectance Geometry Correction Model approach, but if possible, such work should be based upon topographic data that has been generated from a more detailed and accurate data source than that involved in generating the Defense Mapping Agency data tapes. Data sets such as the digital topographic data generated by USGS in conjunction with producing orthophoto quad sheets would be suitable, or the digitization of 1:24,000 scale 7½-minute quad sheets could provide a more detailed and useful set of digital topographic data.

In discussing the overall results of this analysis effort, one additional point that should be emphasized is that the TSRS procedure allowed the ecological characteristics of the study area to be quantitatively characterized and the Topographic Distribution Model to be developed. Figure 2 (p. 9) showed a generalized relationship between elevation and cover type for the Rocky Mountain region. As pointed out, however, this data was not sufficiently accurate to characterize the relationships between elevation and cover types in the San Juan Mountains. Figure 23 is a modified version of Figure 2 in which the Topographic Distribution Model data were used to define the relationship between elevation and cover type for the San Juan Mountains.

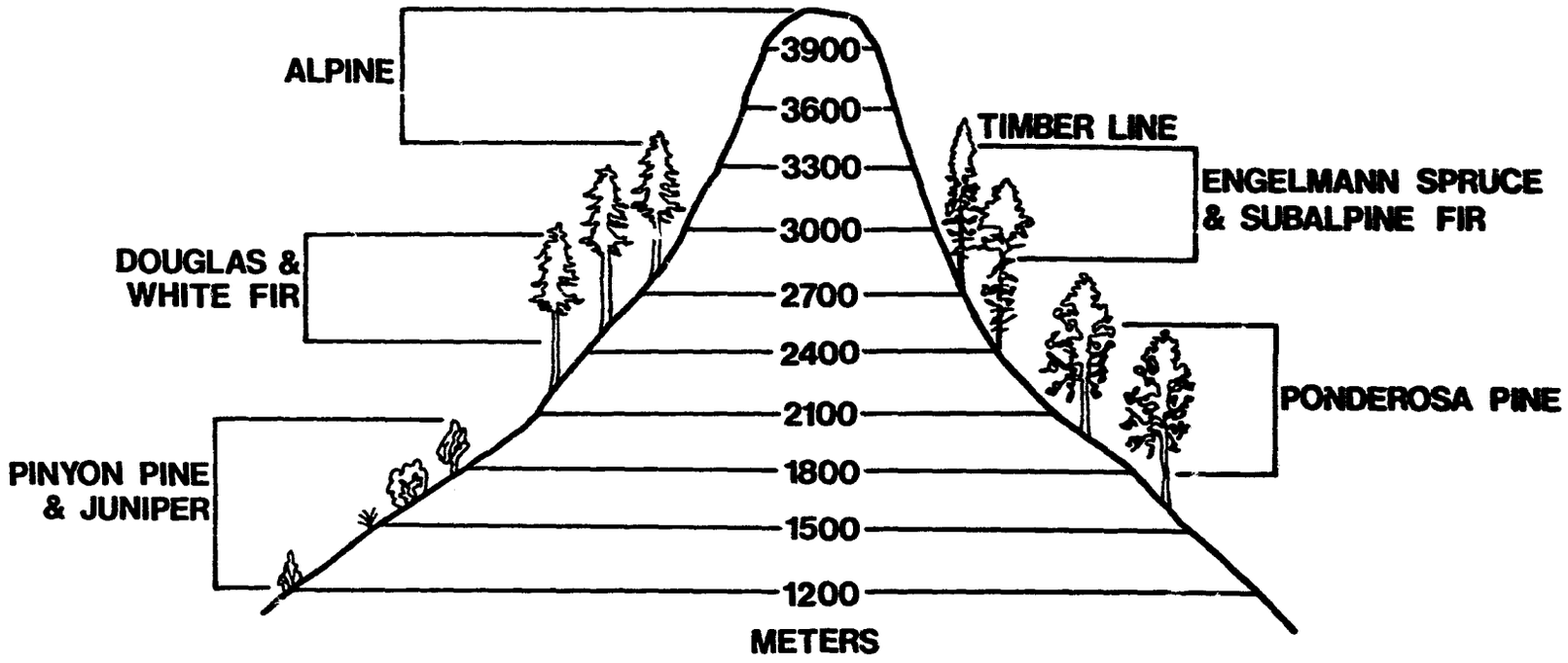


Figure 23. Relationships between elevation and distribution of cover types in the San Juan Mountains. The Topographic Distribution Model was used to develop these relationships, which are quantitative and specific to the San Juan Mountains (as opposed to the data shown in Figure 2 which is qualitative and generalized for the entire Rocky Mountain region).

VI. CONCLUSIONS AND RECOMMENDATIONS

A. Conclusions

The most significant conclusions to date from this spectral/topographic study can be summarized as follows:

1. Using topographic data in addition to spectral data does significantly improve classification performance over using spectral data alone. Elevation data is particularly important in improving classification performance of individual forest cover types. In this study, use of elevation data in conjunction with the spectral data improved overall classification performance by approximately 15%. Aspect and slope did not appear to improve classification performance, but this is believed to be related to the specific characteristics of this topographic data set. Topographic data of better quality and a modified approach to analyzing the data would probably result in significant improvement in classification performance, particularly with regard to aspect.

2. The Stratified Random Sample approach to developing the Topographic Distribution Model data proved to be effective and provided a statistically valid, quantitative description of the distribution of cover types as a function of topography. This is believed to be the first detailed quantitative attempt ever at describing the topographic distribution of the various cover types. The Topographic Distribution Model data was essential for developing the topographic training statistics. Once developed for a given geographic area, such data is relatively static, and so only spectral training statistics are needed for classification of new Landsat data from the same geographic area.

3. The Multi-Cluster Blocks procedure for developing the spectral training statistics is recommended and the Layered classification technique is recommended for the classification of combined spectral and topographic data. The Multi-Cluster Blocks approach for developing spectral training statistics is particularly helpful in (a) minimizing the reference data required and (b) allowing the

training statistics to be developed as a function of the spectral rather than informational, classes present. The Layered Classification technique reduces the complexity of the classification process by dividing it into separate stages and therefore requires considerably less computer time.

4. A Stratified Random Sample approach for developing spectral training statistics appears to be effective, provided that an adequately large sample of data points is used. A regularly defined set of grid cells based upon random start could also be utilized. Existing data sets (such as the U.S. Forest Service Forest Survey sample plot data and the State of Washington GRIDS data) may therefore be of value in developing spectral as well as topographic training statistics. Data sets involving individual pixels can be effectively utilized for developing spectral training statistics only if the cover type associated with the pixel location is already known, if the data set is of adequate size to spectrally characterize the cover types of interest, and if the location of the X-Y coordinates in the Landsat data can be accurately defined. If such a data set is not available, time considerations would suggest the development of spectral training statistics using the Multi-Cluster Blocks approach.

5. The Maximum Likelihood algorithm is not an appropriate algorithm for classifying cover types that do not have a Gaussian distribution as a function of elevation, e.g., water, exposed rock, or grasslands. For these cover types, an algorithm such as contained in the Levels Classifier would probably be more effective.

6. Use of a priori probabilities of occurrence (i.e.,"weights) did not improve the classification performance. The reasons for this are not clear, but may be related to differences in the areal extent of cover types in the training and test quadrangles.

7. The Reflectance Geometry Correction Model did "correct" the Landsat data for topographic effects, if judged on comparison of map printouts of the "corected" and "uncorrected" data; however, the

tabular classification results did not indicate any improvement from use of the model. This lack of improvement may have been due, in part, to errors in the slope and aspect data which were due to the characteristics of the DMA elevation data and the interpolation procedures used.

8. The method used in selecting test data sets can significantly influence the quantitative results of the classification. Manual selection of "test fields" can cause significant upward bias in classification performance figures. On the other hand, a random statistical sample of individual test pixels can cause a downward bias due to registration errors, between the Landsat and topographic data, in accuracies in the topographic data, and spatial variability of the cover type. In addition, the use of randomly defined individual test pixels required a great amount of analyst time to accurately locate the pixel on the aerial photography and identify the cover type. The best test data for quantitative evaluation appears to be a randomly defined set of test fields. This approach minimizes registration and identification problems associated with a single test pixel and also minimizes analyst bias by statistically defining the location of the test areas.

9. The Defense Mapping Agency topographic data has some limitations due to the size of the digitized cell and the interpolation process used for defining the elevation of cells located between contour lines. An unproportionally large percentage of the cells were placed at the 61-meter (200-foot) contour elevation levels, and this distribution, in turn, caused errors in defining slope and aspect. These errors in slope and aspect (e.g., as seen in reservoirs) are believed to be the cause for the lack of improvement in classification accuracy when slope and aspect data were used.

10. This study represents a very detailed analysis of an area that is very complex, from the standpoint of both vegetation and topography. In this first phase of the work, many insights into

analysis techniques and the complexities of relating spectral and topographic data were gained, providing a base of knowledge for more effective approaches in the next phase of this study.

B. Recommendations

It is recommended that: 1) "Analysis Technique A" (i.e., Multi-Cluster Blocks for developing the spectral training statistics, the Topographic Stratified Random Sample for developing topographic training statistics, and the Layered Classification procedure) be further tested and refined during the next phase of this investigation; 2) "Analysis Technique B" be modified to use a Layered classification approach rather than a single stage classification; 3) both "Analysis Technique B" and the "Reflectance Geometry Correction Model" be further evaluated on the new data set if time and resources permit; 4) the Topographic Stratified Random Sample approach be used to develop the Topographic Distribution Model for the new test site; 5) the new test site for Phase II be a 24-Township area in the west half of the Okanogon quadrangle in north-central Washington.

A major reason for selecting this area is the existence of a data set representing a 10% sample of the state-owned land. Information for each of these "GRIDS" sample locations includes primary and secondary species of forest cover, size-class, basal area, height, and many other characteristics of the forest cover, as well as elevation, slope and aspect data.

References Cited

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.

Anuta, P.E., and M.E. Bauer. 1973. "An Analysis of Temporal Data for Crop Species Classification and Urban Change Detection." LARS Technical Report 110873. Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana 47907. 86 pp.

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

Doan, L.R. and B.P. Sanford. 1970. "Solar Elevation, Depression and Azimuth Graphs." Environmental Research Paper No. 313, AF CRL-70-0086, Air Force Cambridge Research Laboratory, Bedford, Massachusetts.

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. Research Bulletin 919, Agricultural Experiment Station, and Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana 47907. pp. 13-54.

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

Hoffer, R.M. and Staff. 1975a. Natural Resource Mapping in Mountainous Terrain by Computer-Analysis of ERTS-1 Satellite Data. Research Bulletin 919, Agricultural Experiment Station, and Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana 47907. 124 pp.

Hoffer, R.M. and Staff. 1975b. Computer-Aided Analysis of Skylab Multispectral Scanner Data in Mountainous Terrain for Land Use, Forestry, Water Resource, and Geologic Applications. LARS Technical Report 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.

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.

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. 217-4 to 2.7-17.

Marr, J.W. 1961. Ecosystems of the East Slope of the Front Range in Colorado. University of Colorado Studies, Series in Biology No. 8, University of Colorado Press, Boulder, Colorado. 134 pp.

Merriam, C.H. 1890. Results of a Biological Survey of the San Francisco Mountain Region and the Desert of the Little Colorado, Az. U.S. Dept. Agr. No. Am. Fauna 3:1-136.

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

Pearson, G.A. 1920. Factors Controlling Distribution of Forest Types. Ecology 1:289-308.

Phillips, T.L. 1973. LARSYS Version 3. User's Manual , Laboratory for Applications of Remote Sensing, Purdue University, June 1973, 3 Volumes.

Rasmussen, P. Irwin. 1941. The Biotic Communities of the Kaibab Plateau, Arizona. Ecological Monographs, Vol. 11, pp. 229-275.

Sellers, 1972. Physical Climatology, The University of Chicago Press, Chicago and London

Struve, H., W.E. Grabau, and H.W. West. 1977. "Acquisition of Terrain Information using Landsat Multispectral Data; Correction of Landsat Multispectral Data for Extrinsic Effects", Report I, M-77-2, Mobility and Environmental Systems Laboratory, U.S. Army Engineer Waterways Experiment Station, Vicksburg, Mississippi.

Swain, P.H., C.L. Wu, D.A. Landgrebe and H. Hauska. 1975. "Layered Classification Techniques for Remote Sensing Applications," Proceedings of the Earth Resources Survey Symposium, Houston, Texas, June 9, 1975. Vol. I-B, pp. 1087-1097.

U.S. Naval Observatory. 1973. The Nautical Almanac. Superintendent of Documents, U.S. Government Printing Office, Washington, D.C.

Whitfield, C.J. 1933. The Ecology of the Vegetation of Pike's Peak Region. Ecological Monographs 3:75-105.

Woodbury, A.M. 1947. Distribution of Pigmy Conifers in Utah and Northeastern Arizona. Ecology 78:113-126.

Wu, C.L., D.A. Landgrebe and P.H. Swain. 1974. "The Decision Tree Approach to Classification," Ph.D. Thesis, Electrical Engineering Department, Purdue University, West Lafayette, Indiana 47907, 194 pp.

Appendix A

INTERPOLATION AND REGISTRATION OF TOPOGRAPHIC DATA

Digital elevation data for the San Juan site were obtained from the U.S. Defense Mapping Agency (DMA), Topographic Center, Washington, D.C. These data were derived from the 1:250,000 scale U.S.G.S. topographic map of this area (NJ 13-7 W, Durango, Colorado) which has a contour interval of 61 meters (200 feet). The range of elevation in this particular study area is from 1805 meters to 4344 meters. The contours were digitized by hand by the DMA on a table digitizer, and the resulting data points were interpolated using a "planar" algorithm which fits a plane to a triangle of three data points to define new points within the triangle (Noma, 1974). In this manner, a uniform grid of elevation values was obtained from the unequally spaced contour samples. The digitizing increment is .254mm in the x and y directions. On a 1:250,000 map this corresponds to 63.5 meters. The output grid cell was therefore defined as 64 meters square in order to coincide with this sampling resolution.

The elevation data was written on tape in 16-bit words (15 bits plus sign, or 2^{15} = 32,768 levels). At LARS, the data was reformatted to LARSYS format which uses eight-bit words. Therefore, the quantization level of the original data is at best:

$$\frac{(4344-1805) \text{ meters}}{32{,}768 \text{ discrete levels}} = .08 \text{ meters.}$$

The actual range spread over these levels is unknown, but it will be nominally in the range of 0.1 m per bin. The point here is that the original quantization error is minor. In order to fit this range into 8 bits (0-255), the data had to be rescaled, resulting in a quantization of:

$$\frac{(4344-1805) \text{ meters}}{256} = 9.9 \text{ meters.}$$

A significant quantization is thus introduced by the LARSYS representation with respect to the contour interval of 61 meters; on a percentage basis, however, this is only 0.4% for the 2539-meter range of elevation in the test site, which is not an unreasonable error. The accuracy of the original elevation data is not known, but since those elevations were interpolated from contours having an interval of 61 meters, it seems reasonable that the 0.4% error is no worse than that obtained in the process of digitizing the elevations from the original map.

Another reformatting consideration concerned the designation of rows and columns on the DMA topographic data tape. The rows of the topographic data were oriented north-south on the DMA tapes, and the row direction (i.e. scan lines) in the reformatted Landsat data is east-west. Thus a transposition of the topographic data array on the tape was required.

The final LARSYS elevation data tape contains one channel of eight-bit values on a grid of 64 meters for the west half of the Durango quadrangle, which covers a rectangle of one degree of latitude and longitude. In order to retrieve the true elevation values from the eight-bit words, the lower and upper limits of elevation (1805m and 4344m in this case) are stored in full precision format on the tape identification record and used to rescale the eight-bit data to the original range when the data is read from tape into the computer. Thus, the elevations printed out by LARSYS are within the 0.4% quantization error of the original elevations recorded by DMA. The DMA-LARSYS elevation data then had to be registered or digitally overlayed onto the Landsat data.

The normal procedure at LARS for the registration of two digital data sets is to manually determine the approximate location of potential control points that are clearly recognizable on the images of both data sets. A numerical correlation procedure is then used to define the precise X-Y coordinates of the control points in the data sets to be registered. Manual techniques are required for the initial phase of these registrations due to the dissimilar nature of the data involved. The topographic data, in general, will not correlate with the Landsat data, even when the Landsat data have been rotated approximately 12^{o} counterclockwise to a north-south orientation. Therefore, matching points in each data set to be registered are defined

visually using images from the LARS Digital Display and the computer line-printer. The coordinates of these points are punched on cards and processed by a least squares, bi-quadratic polynominal approximation program (Anuta and Bauer, 1973) to define coefficients for use by the registration program.

The registration algorithm uses a nearest neighbor rule to define output points which are required between existing input data points. Since the topographic grid spacing is 64 meters square and the reference grid is 79 meters square, the registered topographic data will have position errors which range from zero to 32 meters. This is an error characteristic of the method used, but in all cases this positional error between the Landsat and topographic data sets is less than one pixel.

In summary the topographic elevation channel registered to the Landsat has two types of errors: value error and position error. The error in the value of the elevation is due to:

1. Inaccuracy in the value of the contour line on the original map.
2. Error due to the action of the planar interpolation algorithm used by the DMA.
3. Quantization error on the DMA data set.
4. Quantization error due to representation in LARSYS format.

Errors 3 and 4 have been shown to be small relative to the contour interval. The errors due to 1 and 2 are unknown; however, it seems clear that between contours the error could be no more than one-half the interval, i.e. $\pm$ 100 ft., and at extreme points, peaks and sinks, the error could be as high as one contour interval. Past experience with quantized and interpolated data sets

indicated that these errors tend to have zero mean and be uniformly distributed over the quantization interval. The standard deviation for this case is $Q/\sqrt{12}$ where Q is the quantization interval. For this case the standard deviation would be σ = 200/3.46 = 57.7 ft. or 17.6 meters. Based on this assumption the standard deviation of errors due to 2 and 4 would be $\sqrt{17.6^2 + 9.9^2}$ = 20.2 meters.

Previous work (Hoffer et al., 1975a) had indicated the desirability of utilizing the slope and direction of slope (i.e. aspect) information as part of the analysis process, if these data could be made available on a pixel by pixel basis as additional registered channels on the data tape. This requirement was met by numerically differentiating the topographic data to produce an estimate of the gradient vector at each pixel location. The magnitude of the vector is then used to derive slope angle, and the direction is used as the aspect angle. The approximate gradient at line i and column j is computed as:

$$\vec{\nabla Z} \approx \vec{I}(z_{i-1,j} - z_{i+1,j}) + \vec{J}(z_{i,j-1} - z_{i,j+1}) \qquad \text{(A.1)}$$

where $\vec{\nabla Z}$ is the gradient vector,

z_{ij} is the topographic elevation value at i,j,

i,j are line and column coordinates, and

$\vec{I}$ and $\vec{J}$ are line and column unit vectors.

The slope angle is computed from the magnitude of gradient. The $|\vec{\nabla Z}|$ value is the vertical change in elevation over one unit of pixel distance.

Thus the slope is:

$$s_{ij} = \tan^{-1} \frac{\sqrt{(z_{i-1,j} - z_{i+1,j})^2 + (z_{i,j-1} - z_{i,j+1})^2}}{\Delta d} \qquad (A.2)$$

where: s_{ij} is the slope angle at point i,j with $0 \leq s < 90$ degree, and

Δd is the pixel spacing.

The aspect angle is derived from the vector direction of the gradient:

$$\alpha = \tan^{-1} \frac{(z_{i-1,j} - z_{i+1,j})}{(z_{i,j-1} - z_{i,j+1})} \qquad (A.3)$$

where α is the direction of slope. The actual implementation is more complex than this formula indicates.

Since only positive values from 0-255 can be represented on LARSYS format tapes, the aspect angle is recorded on a range of zero to 180 in one channel to keep a resolution of one degree and an additional channel is used which has only the values zero or one. If the slope,faces to the east the zero-one channel will have a value of zero, and if the slope faces the west the zero-one channel will have a value of one. Thus a pixel having a slope facing toward the east will have an aspect value of 90^{o} and a flag value of zero. The resolution of the slope and aspect angles is one degree. A fourth topographic channel is included which contains aspect on a $0\text{-}360^{o}$ scaled so that 360^{o} equals a value of 255. Therefore, the aspect resolution of this channel is 1.4^{o}.

The slope and aspect angle derivation was then implemented in a program (SLOPE) which adds these channels to a data tape as four additional channels, registered to the topographic elevation channel (and in this study, also registered to the Landsat data on a pixel-by-pixel basis). The channels containing the various Landsat wavelength bands and the topographic data are summarized in Table 2 of this report.

Appendix B

IDENTIFICATION OF CODES USED IN INSTAAR COVER TYPE MAPS

Number Code	Category
00.	Non-vegetated
00.	Exposed soil
01.	Water
02.	Urban
110	Grasslands
121	Colorado blue spruce
122	Cottonwood-willow
130	Montane/subalpine meadow
141	0-30% vegetative cover tundra
142	30-70% vegetative cover tundra
143	70-100% vegetative cover tundra
144	Graminoid wet meadow, usually tundra
145	Alpine shrub
151	Wet shrub
152	Dry shrub
153	Oak
211	Aspen
221	Pinyon pine/Rocky Mountain juniper
222	Ponderosa pine
222.1	Ponderosa pine with shrub
223	Ponderosa pine/Rocky Mountain juniper
224	Ponderosa pine/Douglas fir
225	Engelmann spruce-subalpine fir
Zipatone	Krummholz
225.1	Engelmann spruce/Douglas fir
226	Lodgepole pine
227	Limber pine/bristlecone pine
228	Douglas fir/white fir
229	Mixed coniderous (DF/WF/ESP/PP)
231	Douglas fir/Ponderosa pine/Aspen
232	Douglas fir/White fir/Aspen
233	Lodgepole/Aspen

Number Code	Category
234	Mixed coniferous-deciduous
235	Engelmann spruce/Subalpine fir/Aspen
161	Pasture
162	Cultivated crop
163	Cultivated pasture

Appendix C

QUANTITATIVE DESCRIPTION OF FOREST AND HERBACEOUS COVER TYPES IN THE SAN JUAN STUDY AREA

The thirty figures that appear in Appendix C show graphically the cover type distribution data assembled as part of this study. The cover types described are:

Coniferous cover types and cover type mixtures
- Spruce/fir
- Spruce/fir and Douglas/white fir
- Douglas/white fir
- Douglas/white fir and ponderosa pine
- Ponderosa pine

Deciduous cover types
- Aspen/willow
- Aspen
- Oak

Herbaceous cover types
- Tundra
- Grassland

For each cover type, a histogram is used to depict the frequency of occurrence of the class as a function of elevation. Polar plots for each cover type are used to display the distribution as a function of aspect and elevation, and also of aspect and slope.

The data represented by these figures is the only known quantitative description of the distribution of forest and herbaceous cover types in the San Juan Mountains.

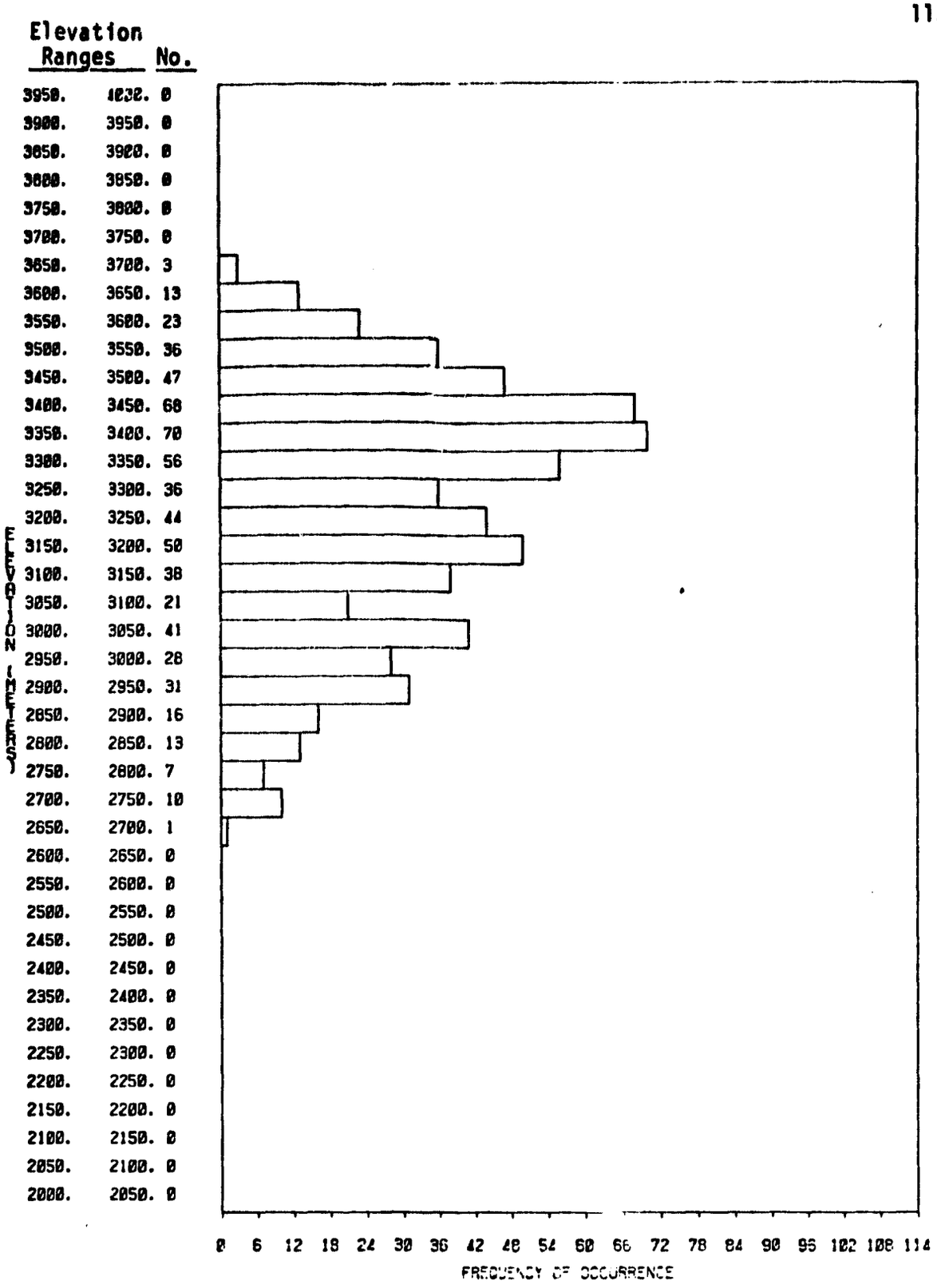


Figure C.1 Distribution of spruce/fir as a function of elevation.

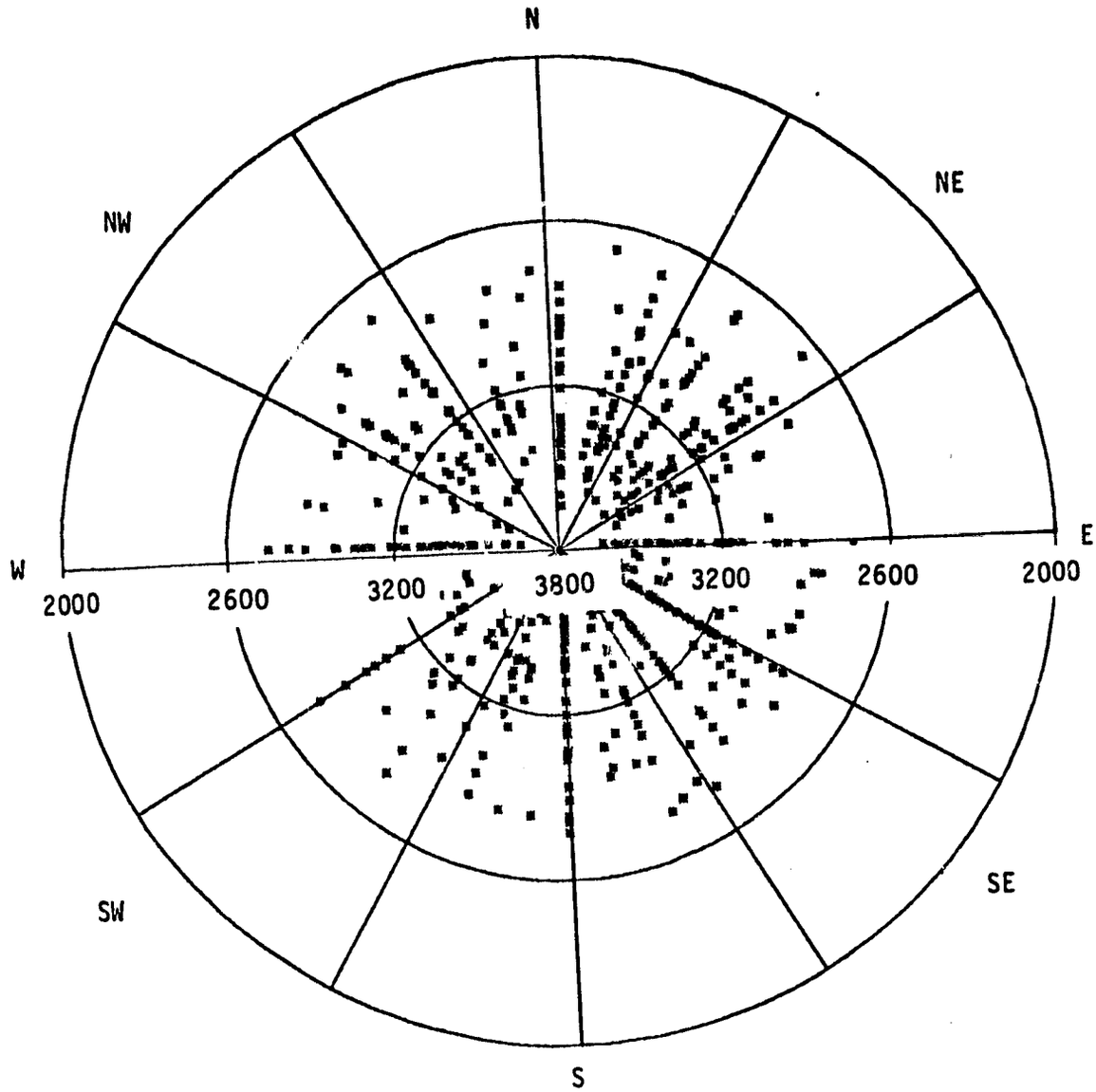


Figure C.2 Distribution of spruce/fir as a function of elevation (in meters) and aspect.

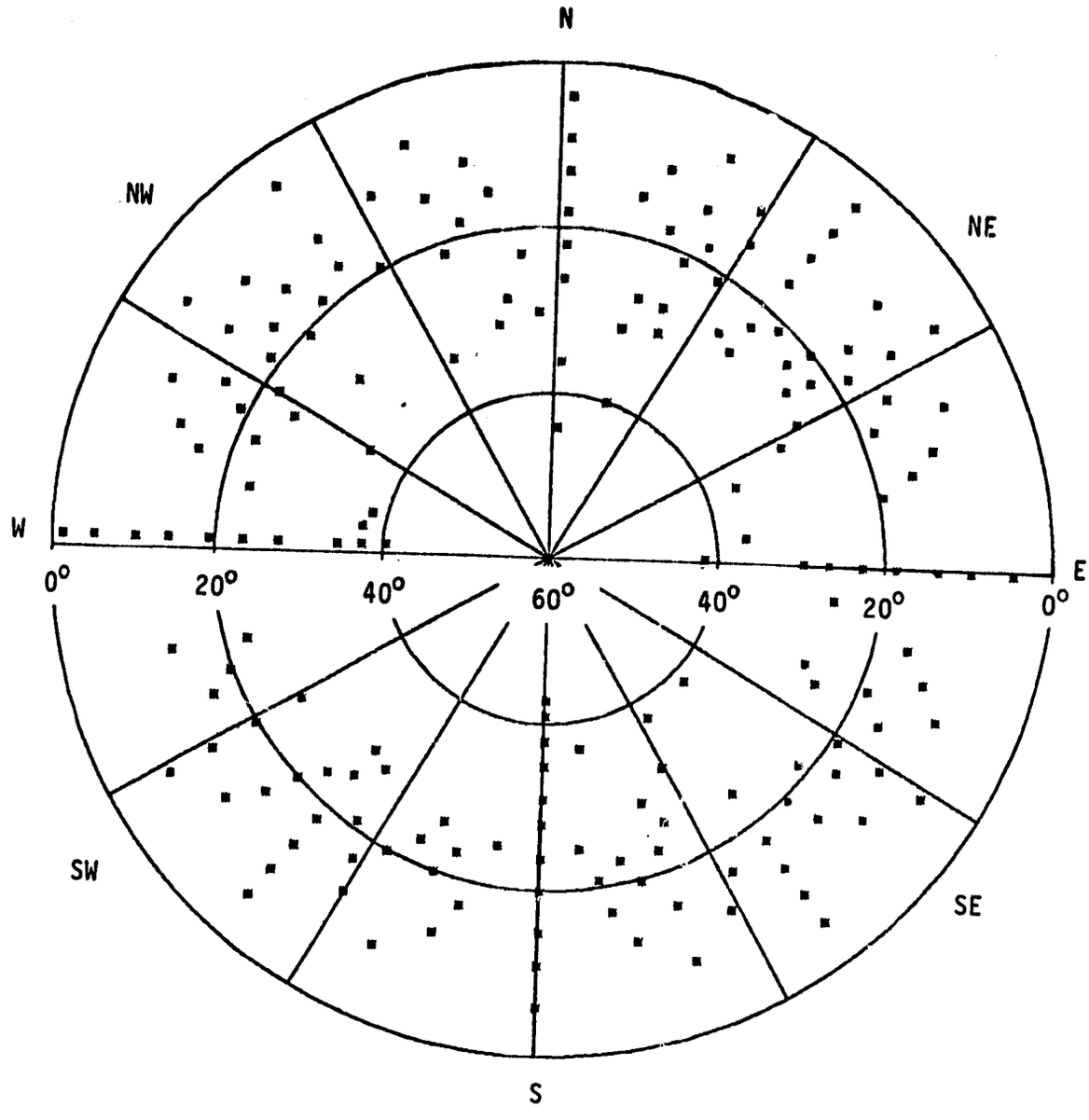


Figure C.3 Distribution of spruce/fir as a function of slope and aspect.

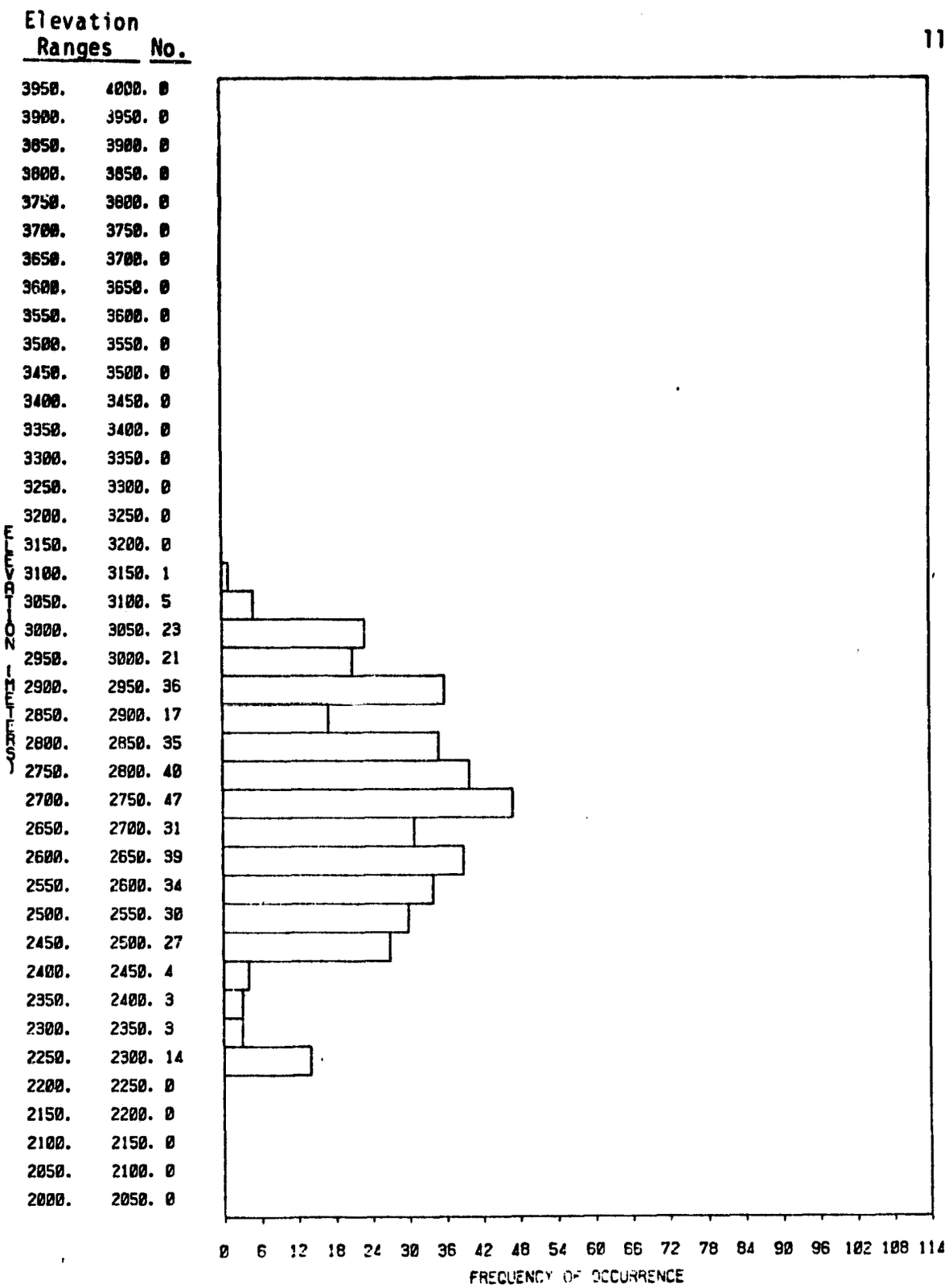


Figure C.4 Distribution of mixture of spruce/fir and Douglas/white fir as a function of elevation.

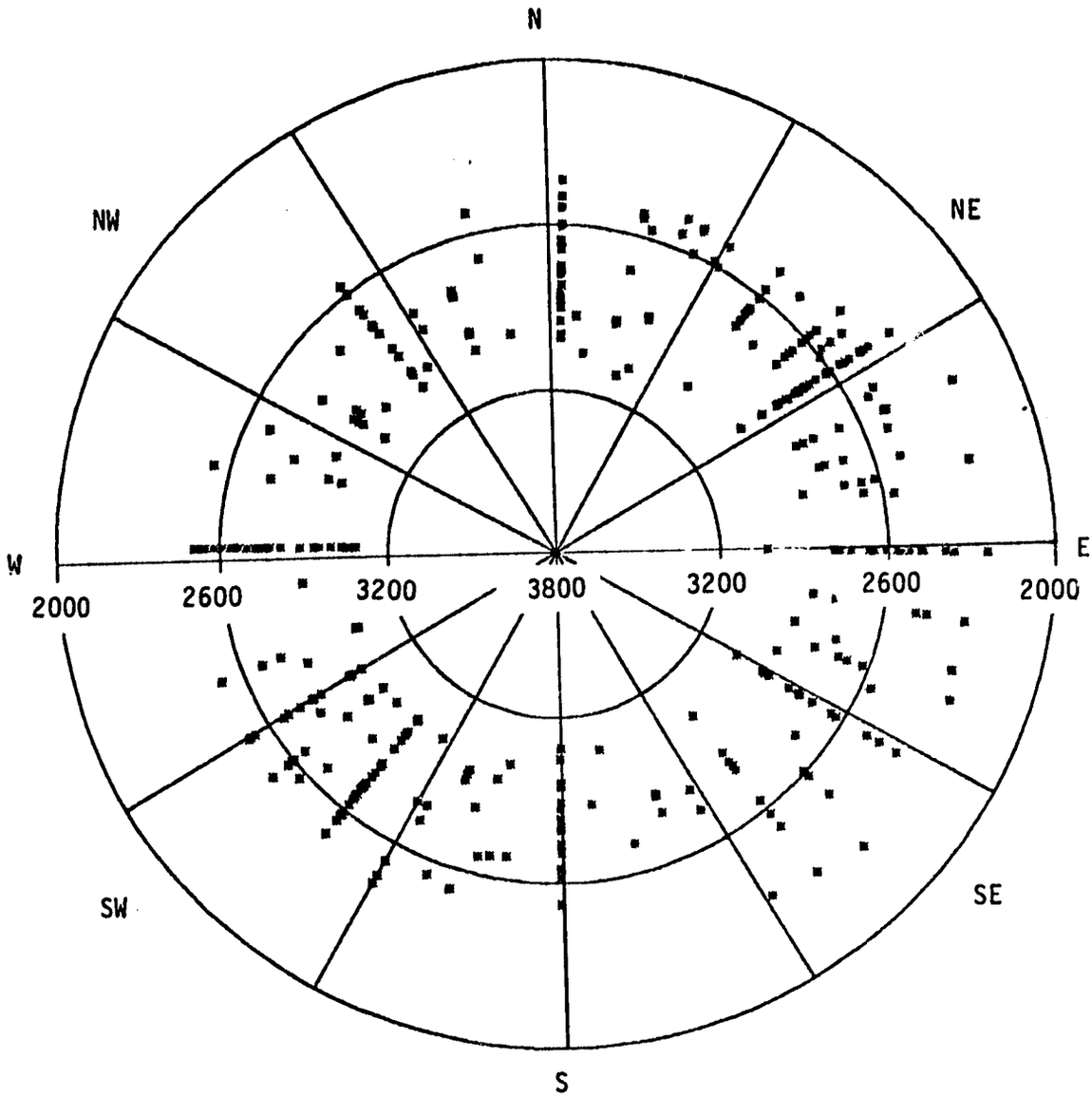


Figure C.5 Distribution of mixture of spruce/fir and Douglas/white fir as a function of elevation (in meters) and aspect.

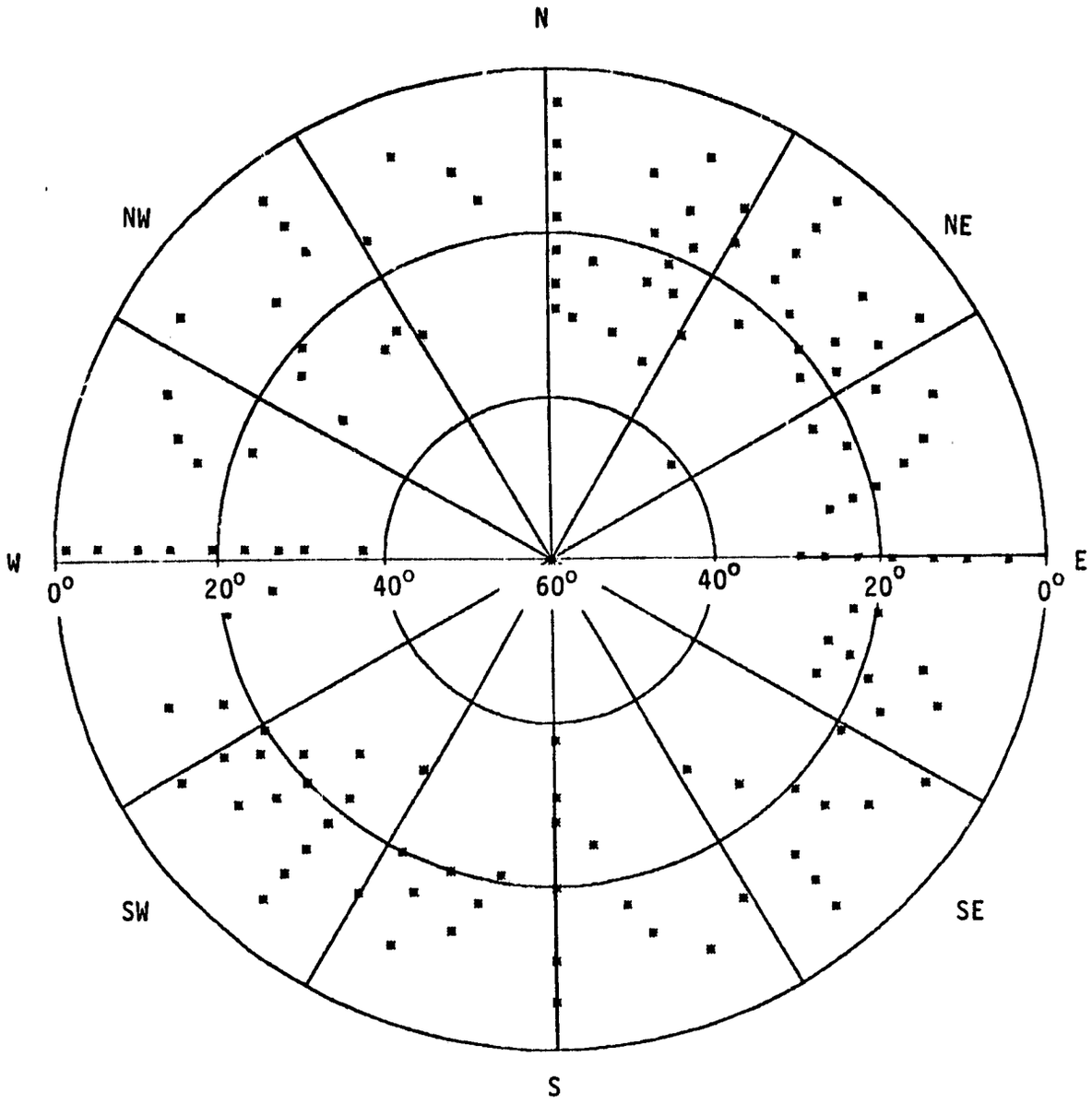


Figure C.6 Distribution of mixture of spruce/fir and Douglas/white fir as a function of slope and aspect.

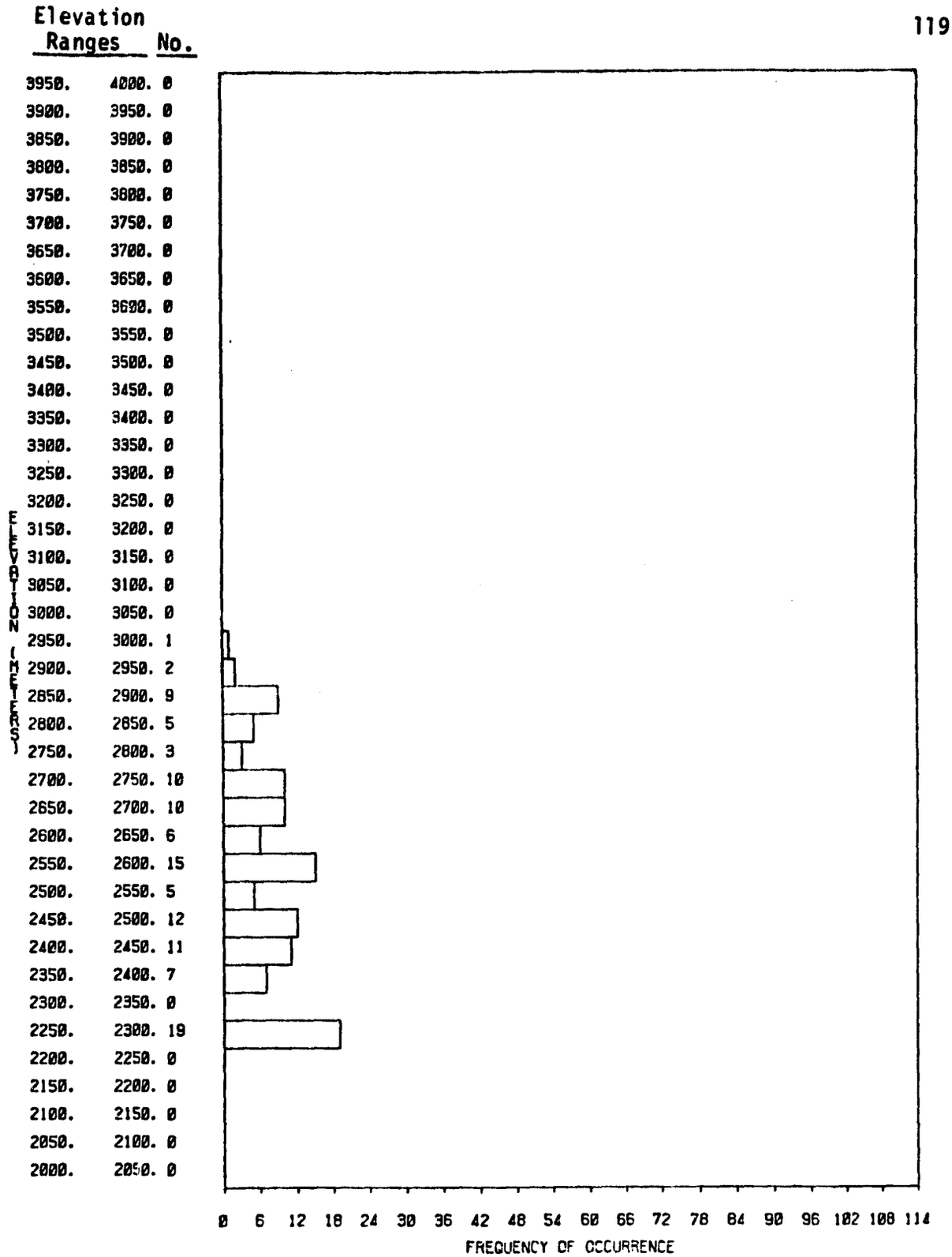


Figure C.7 Distribution of mixture of Douglas/white fir as a function of elevation.

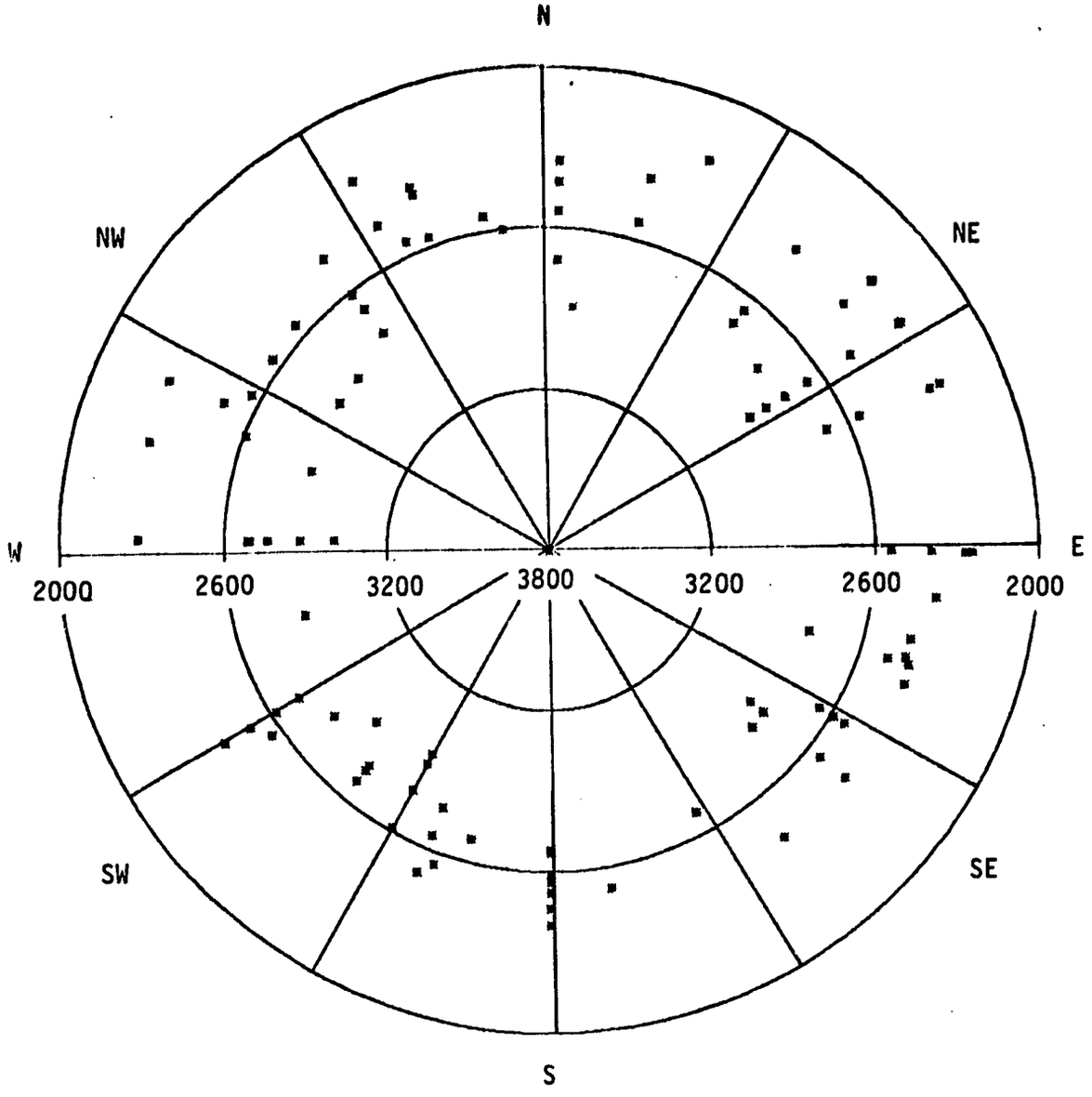


Figure C.8 Distribution of mixture of Douglas/white fir as a function of elevation (in meters) and aspect.

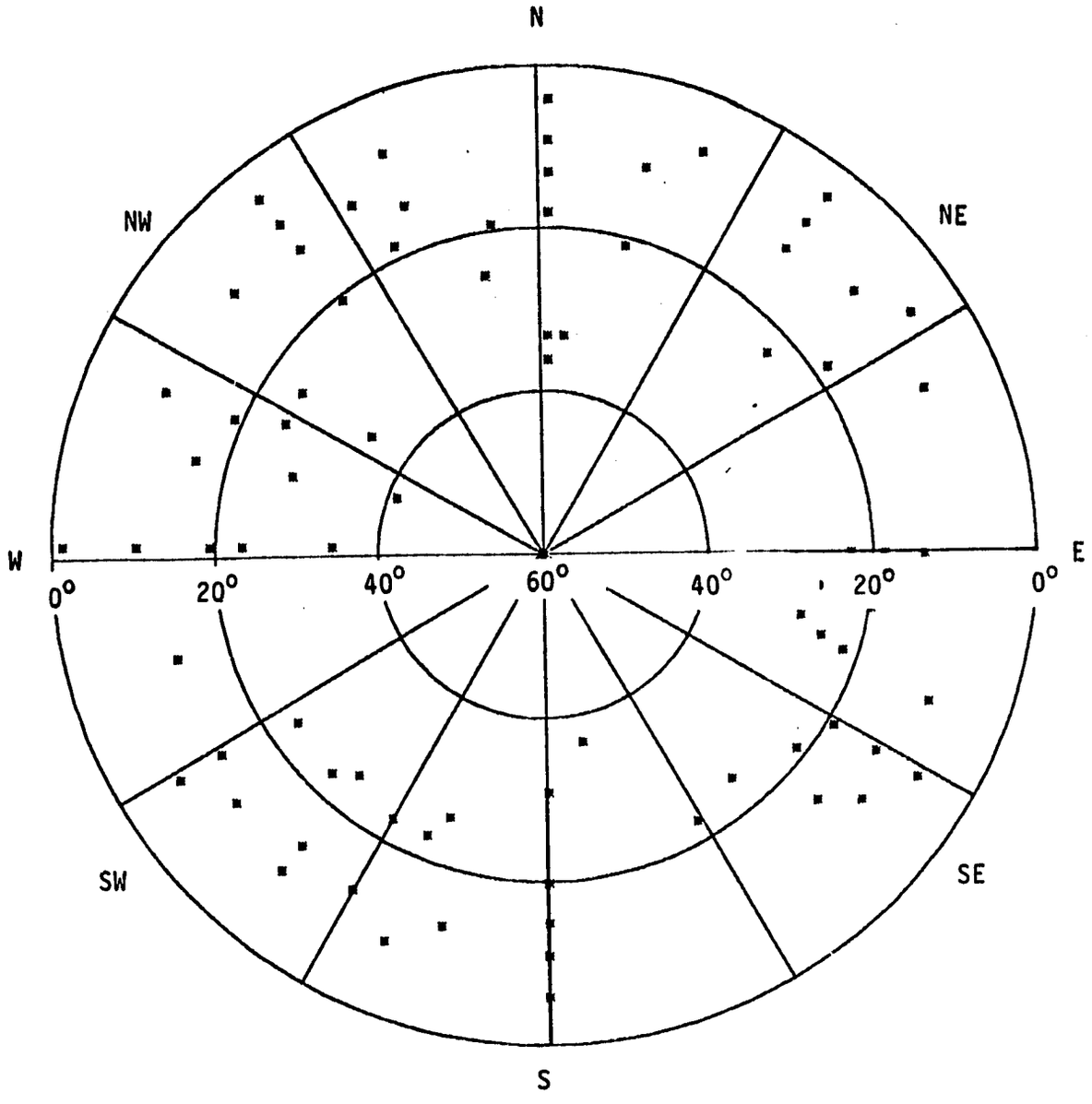


Figure C.9 Distribution of mixture of Douglas/white fir as a function of slope and aspect.

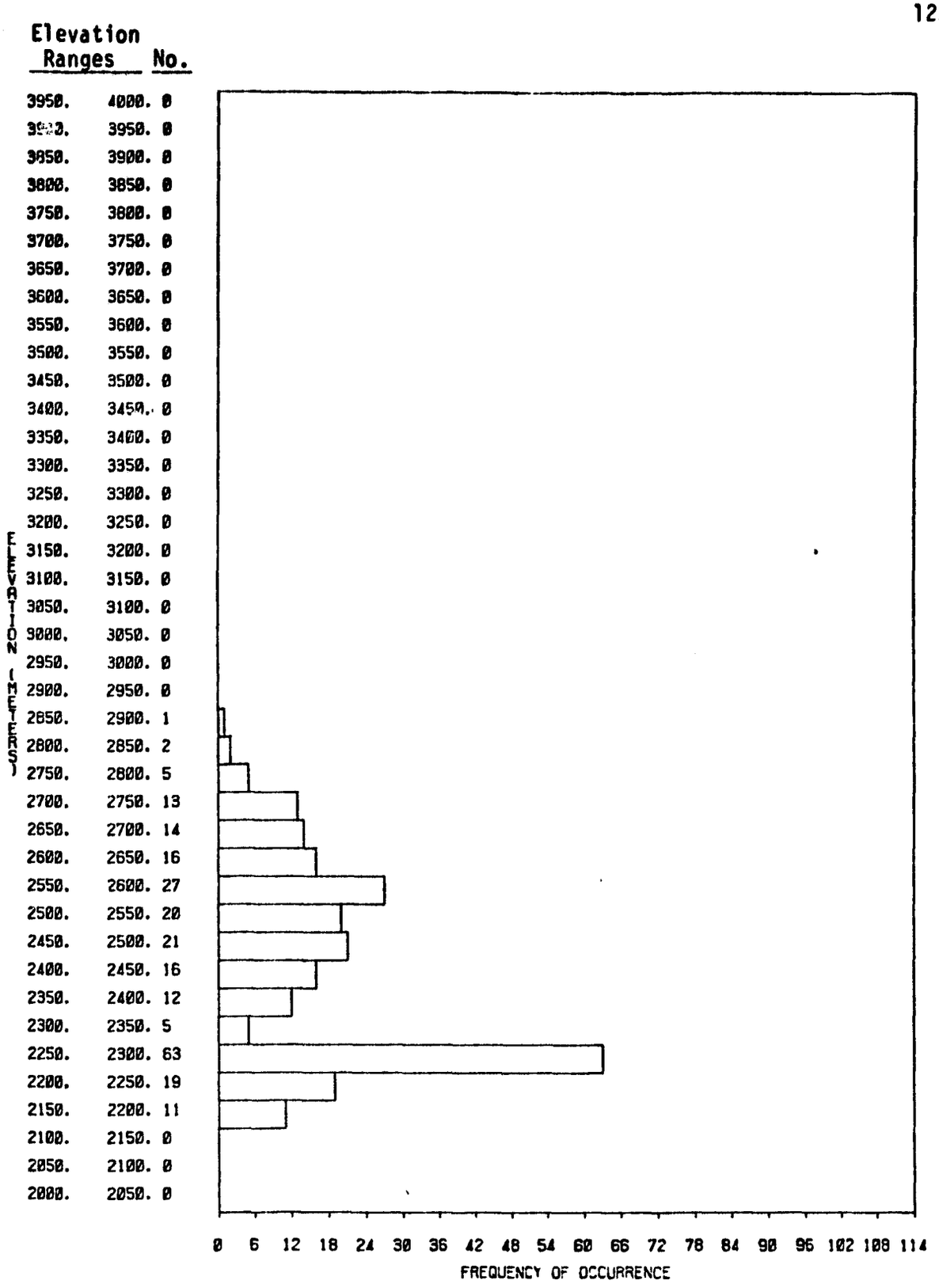


Figure C.10 Mixture of Douglas/white fir and ponderosa pine as a function of elevation.

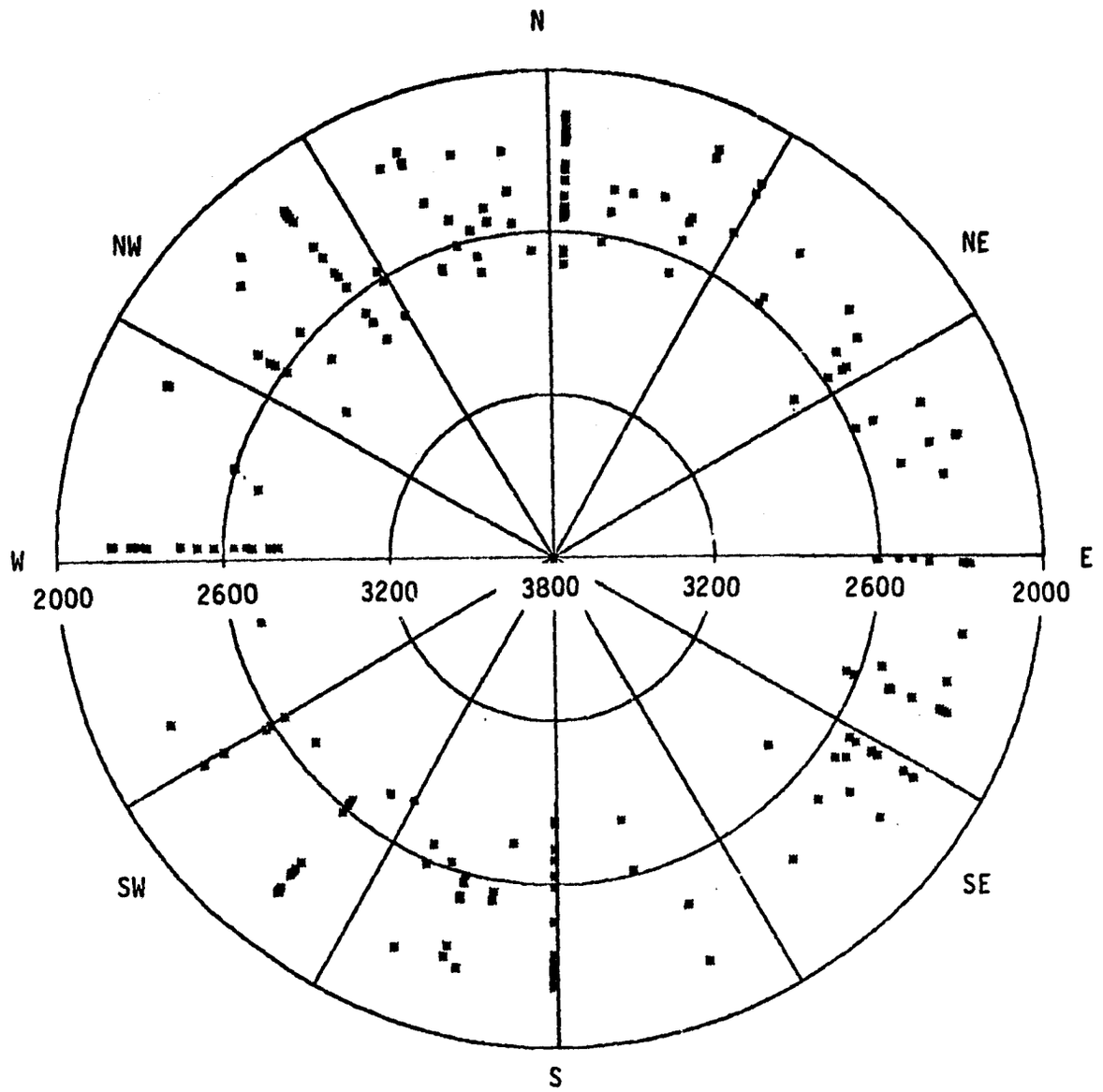


Figure C.11 Mixture of Douglas/white fir and ponderosa pine as a function of elevation (in meters) and aspect.

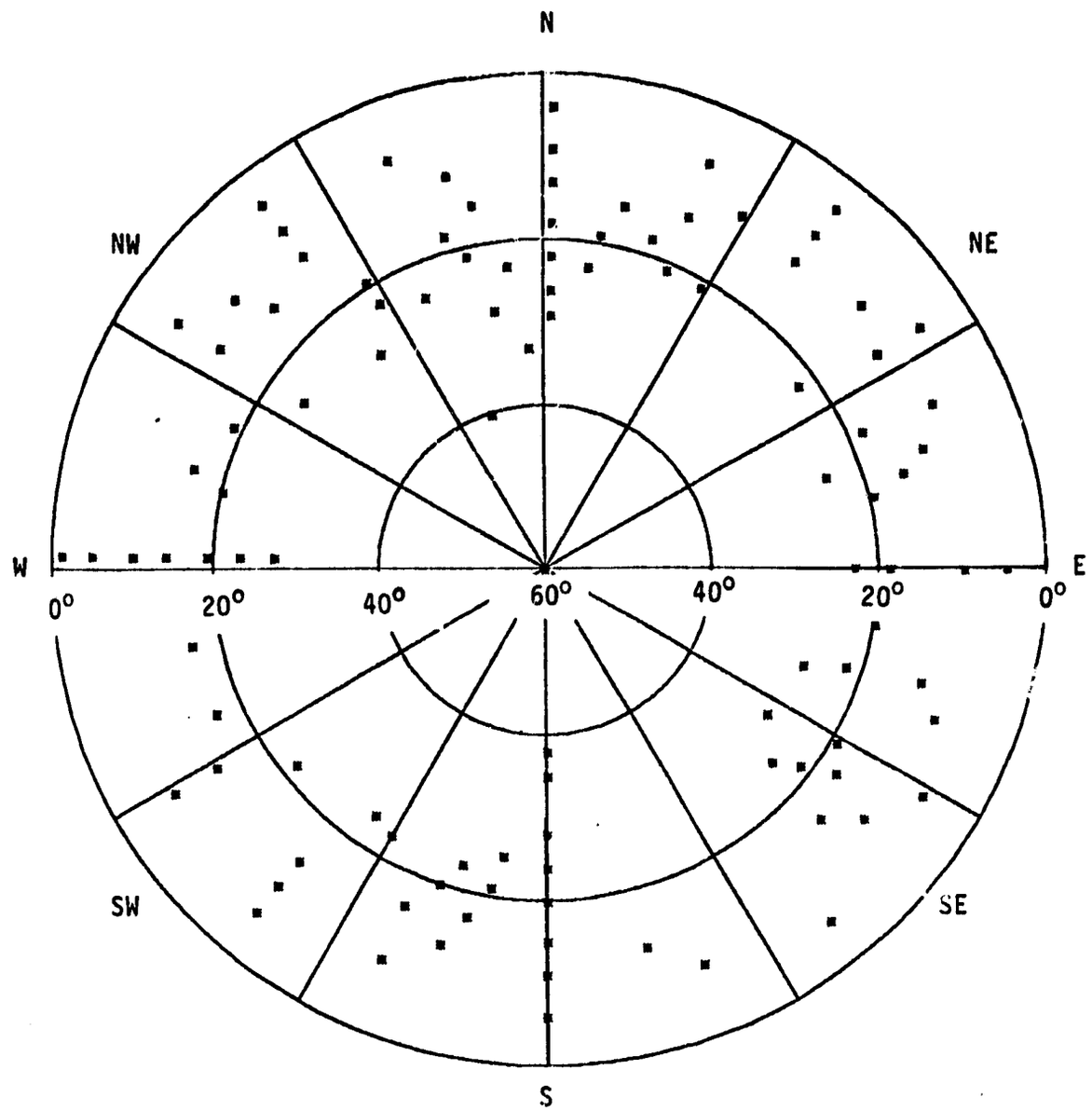


Figure C.12 Mixture of Douglas/white fir and ponderosa pine as a function of slope and aspect.

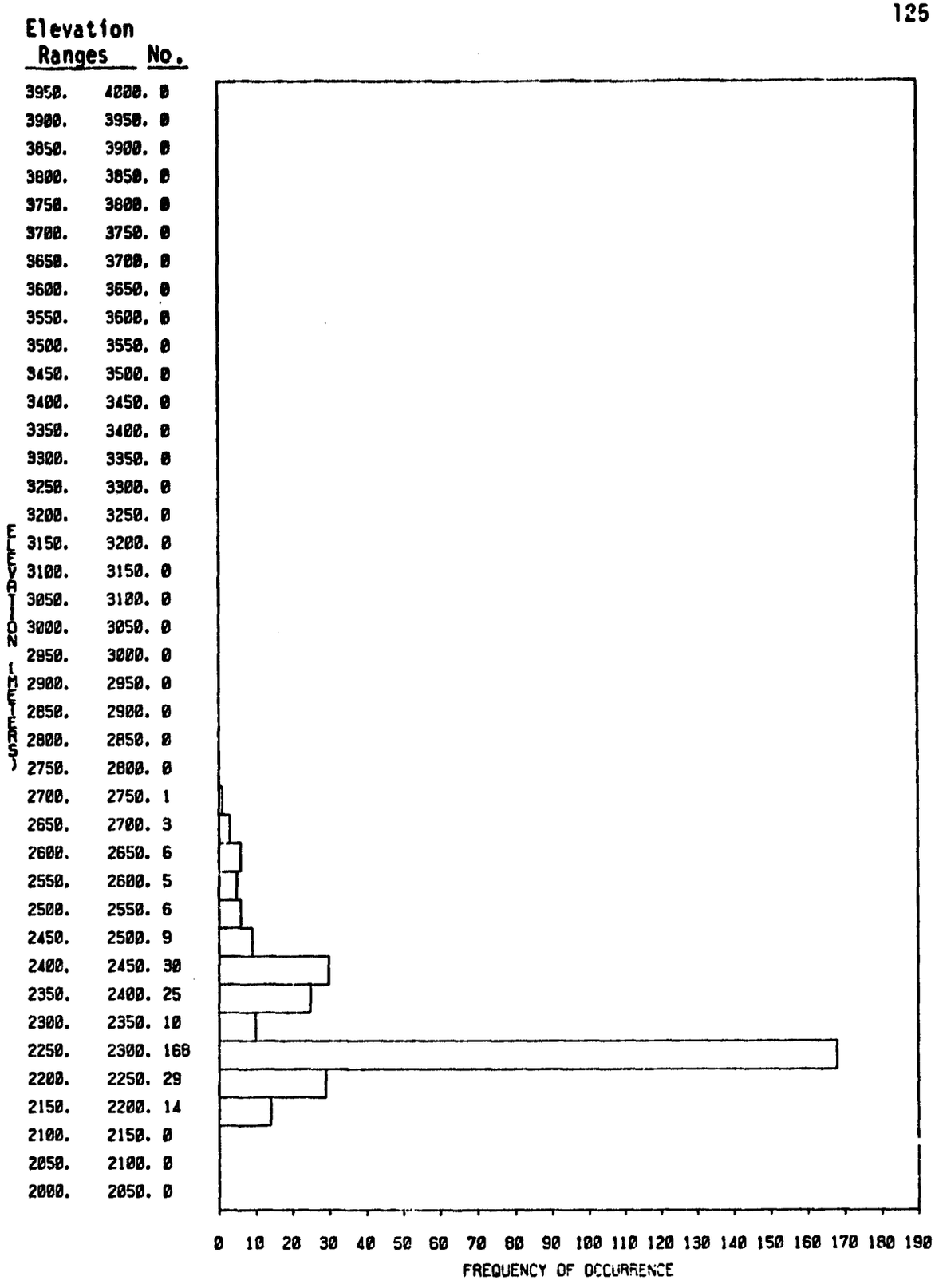


Figure C.13 Distribution of ponderosa pine as a function of elevation.

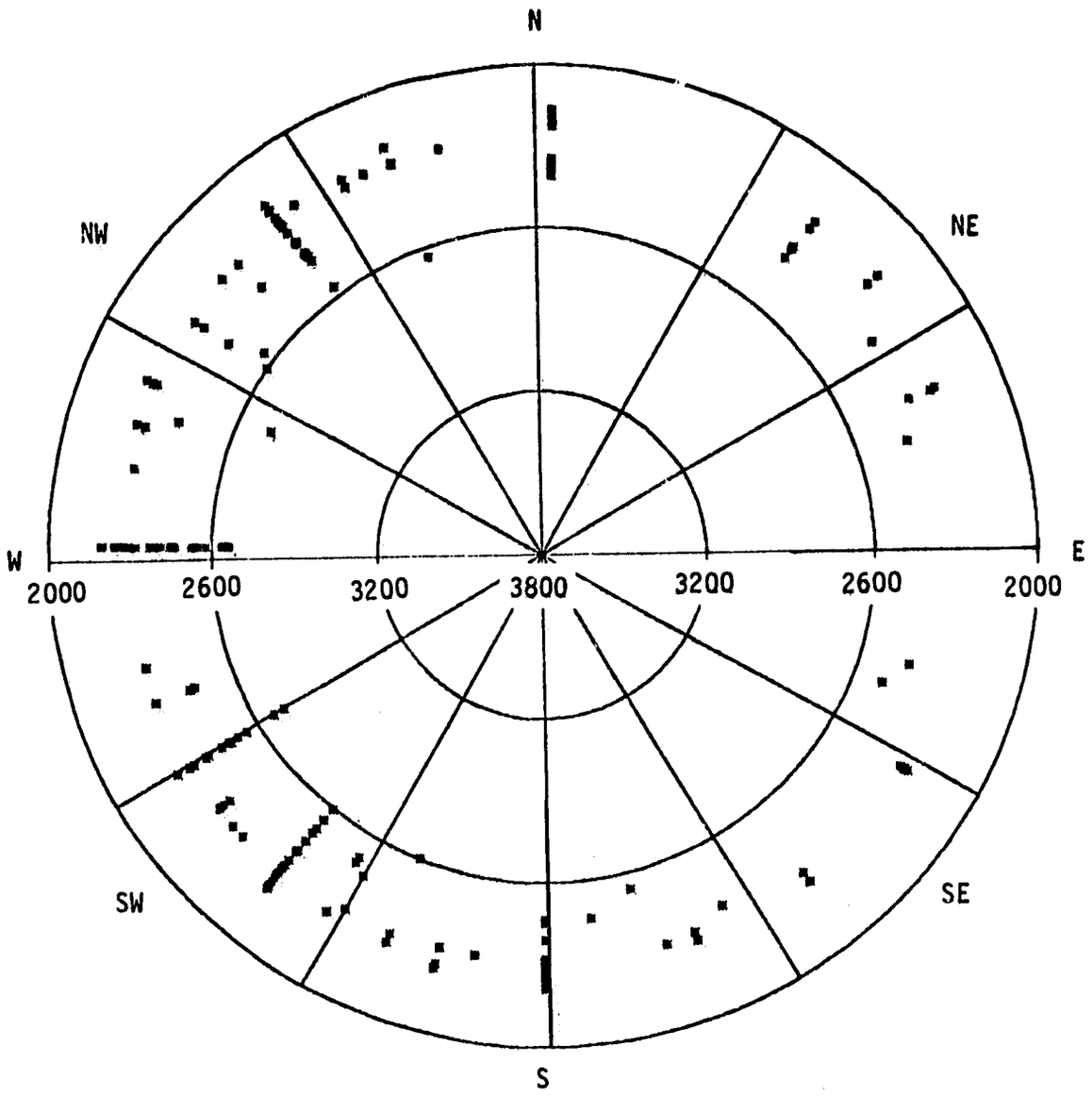


Figure C.14 Distribution of ponderosa pine as a function of elevation (in meters) and aspect.

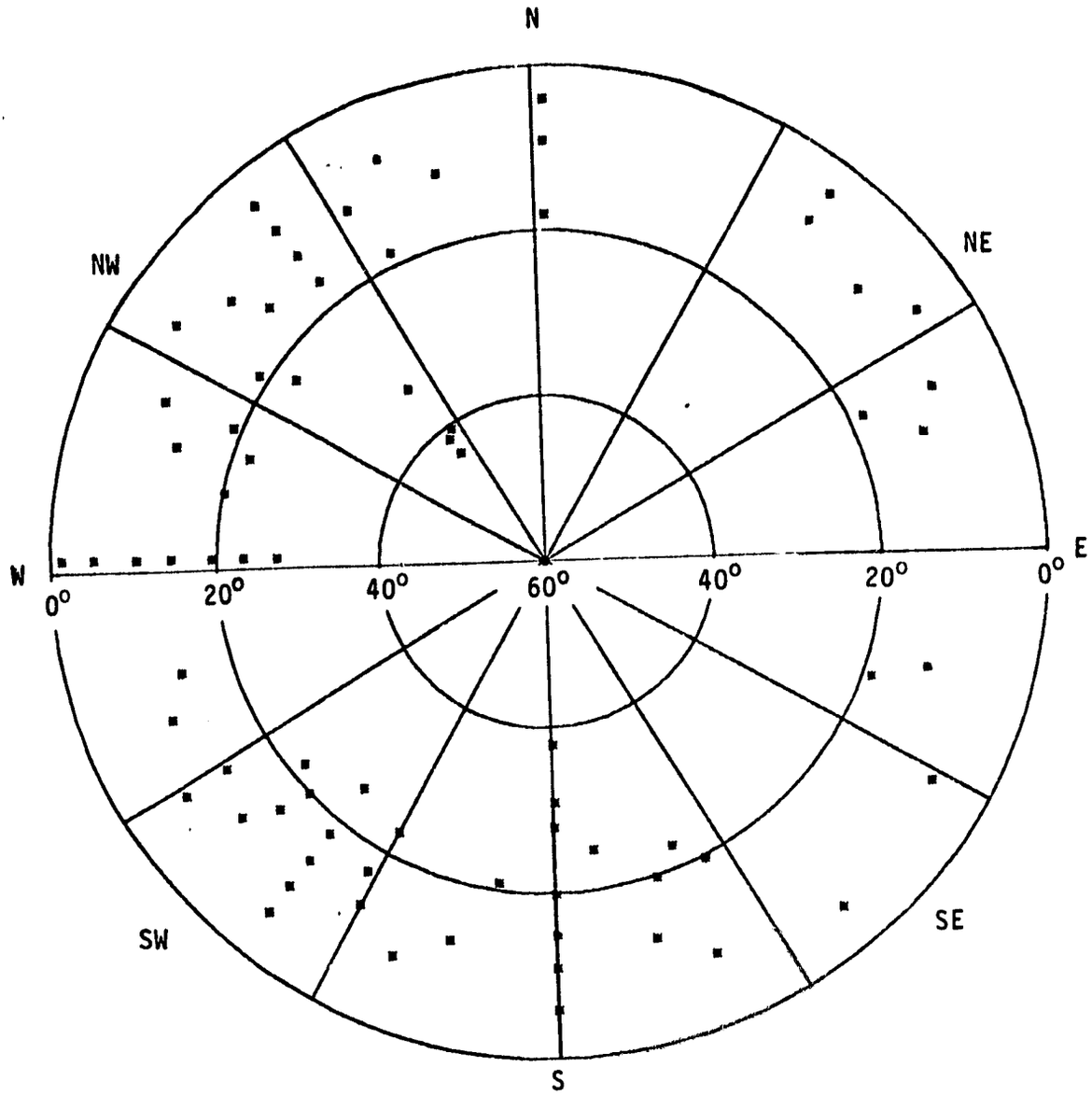


Figure C.15 Distribution of ponderosa pine as a function of slope and aspect.

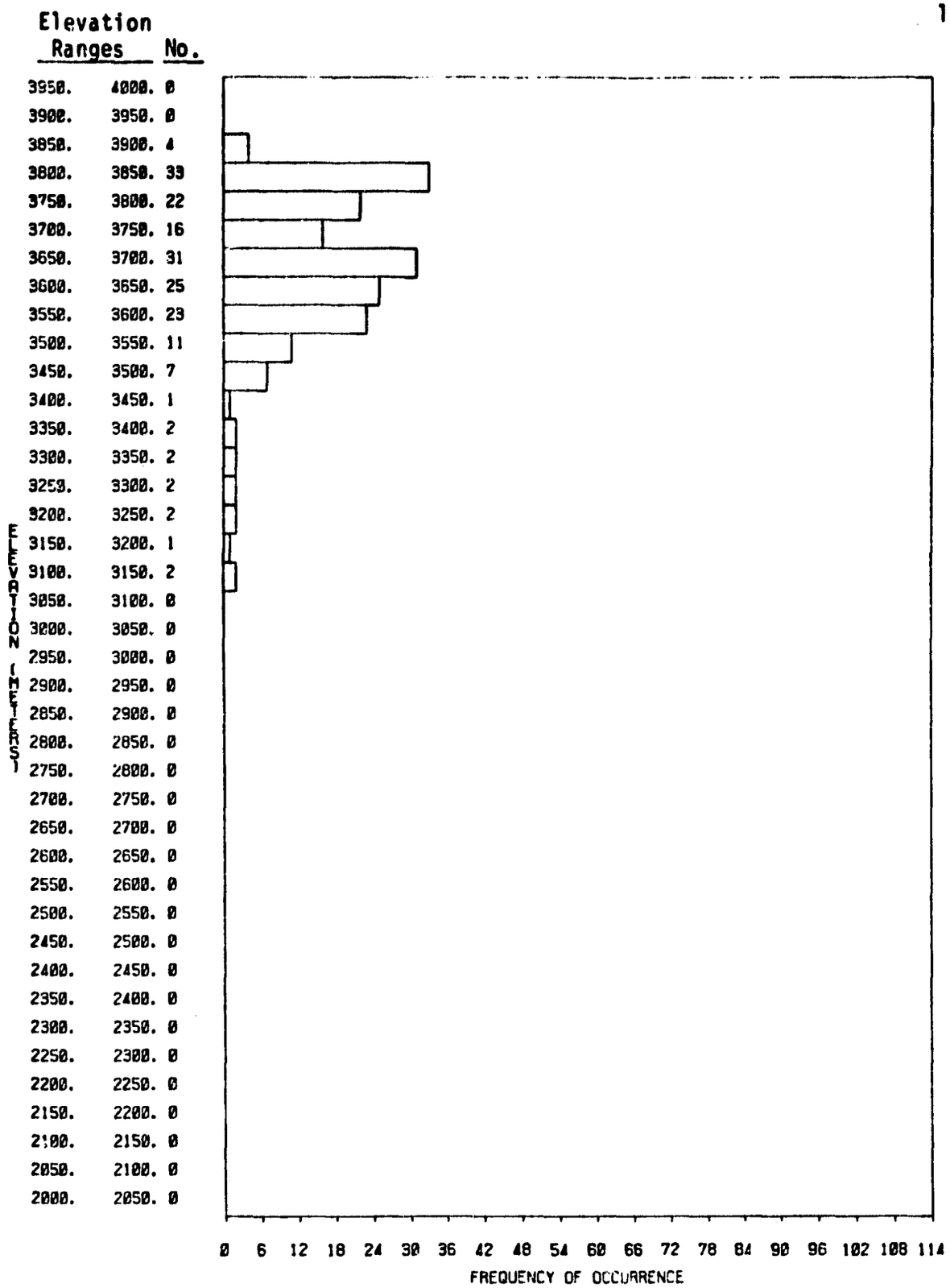


Figure C.16 Distribution of alpine/willow as a function of elevation.

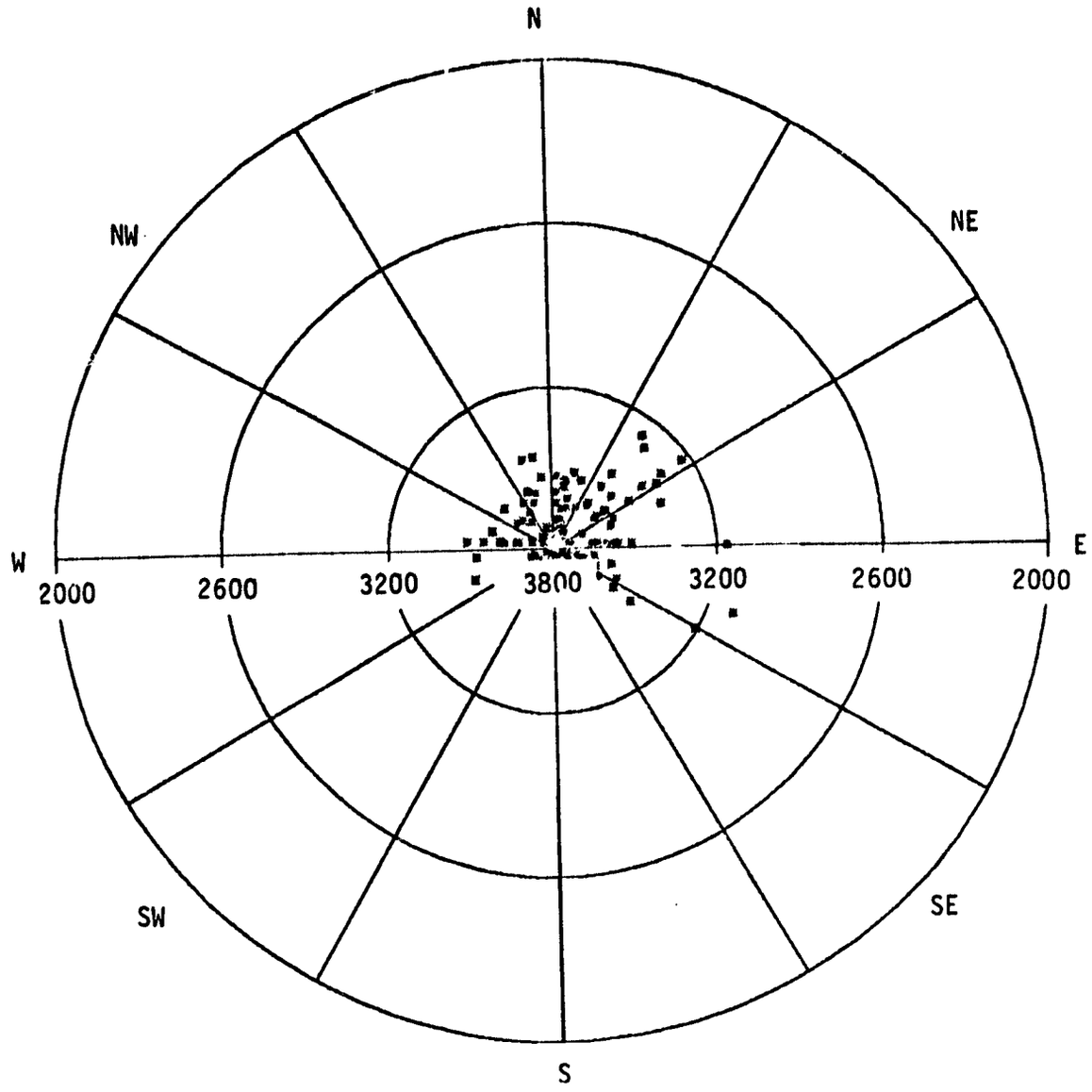


Figure C.17 Distribution of alpine/willow as a function of elevation (in meters) and aspect.

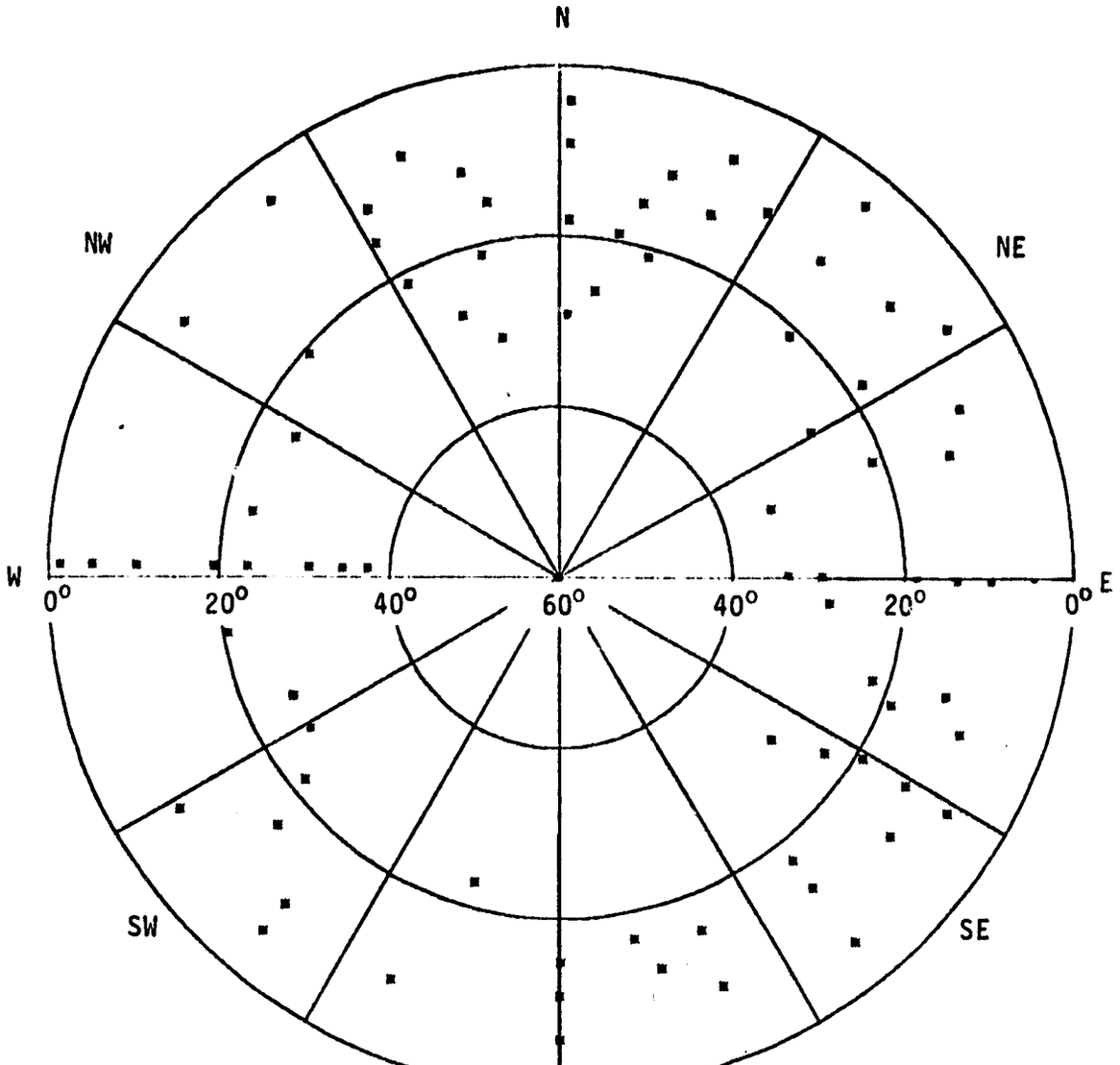


Figure C.18 Distribution of alpine/willow as a function of slope and aspect.

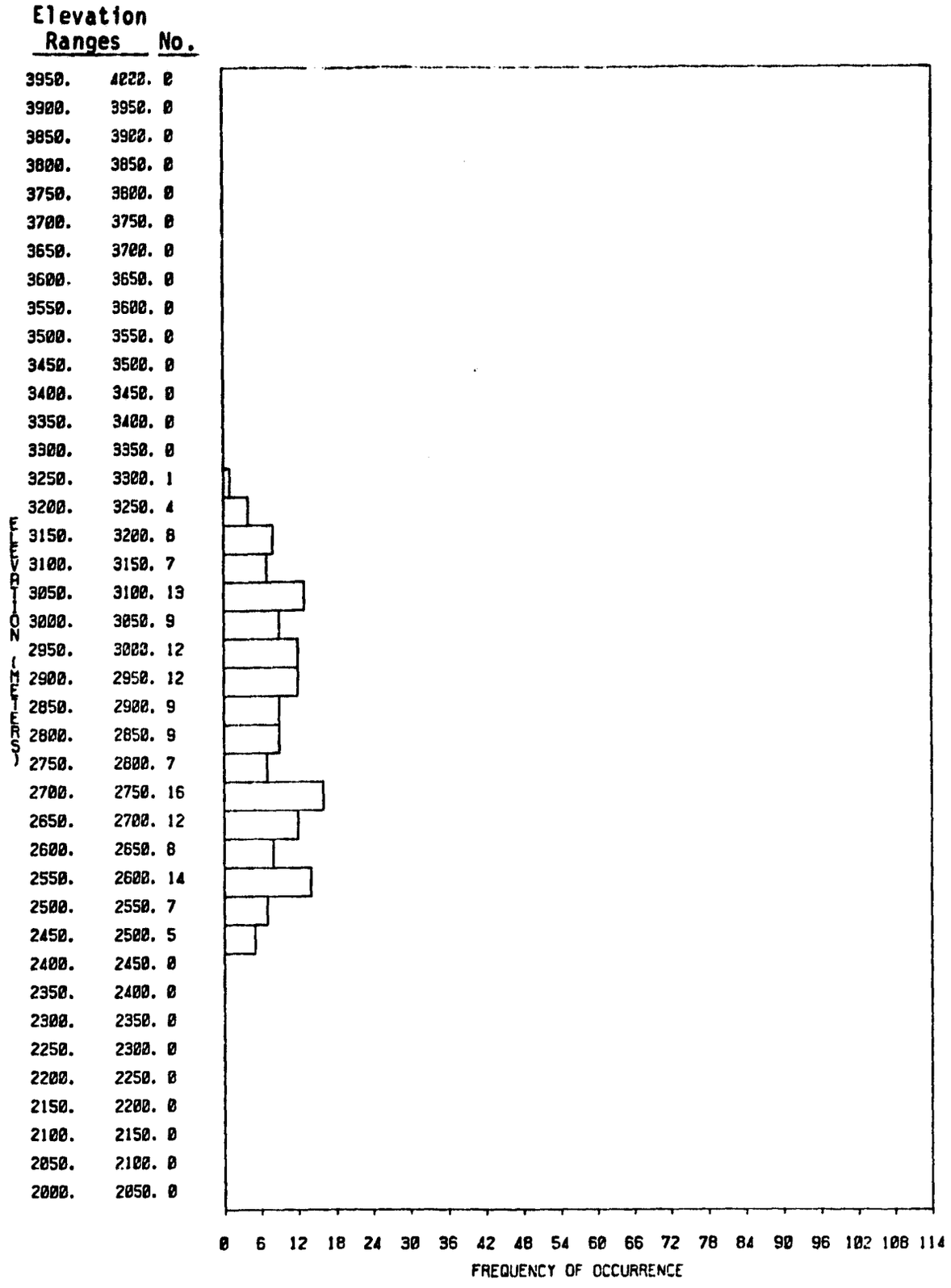


Figure C.19 Distribution of aspen as a function of elevation.

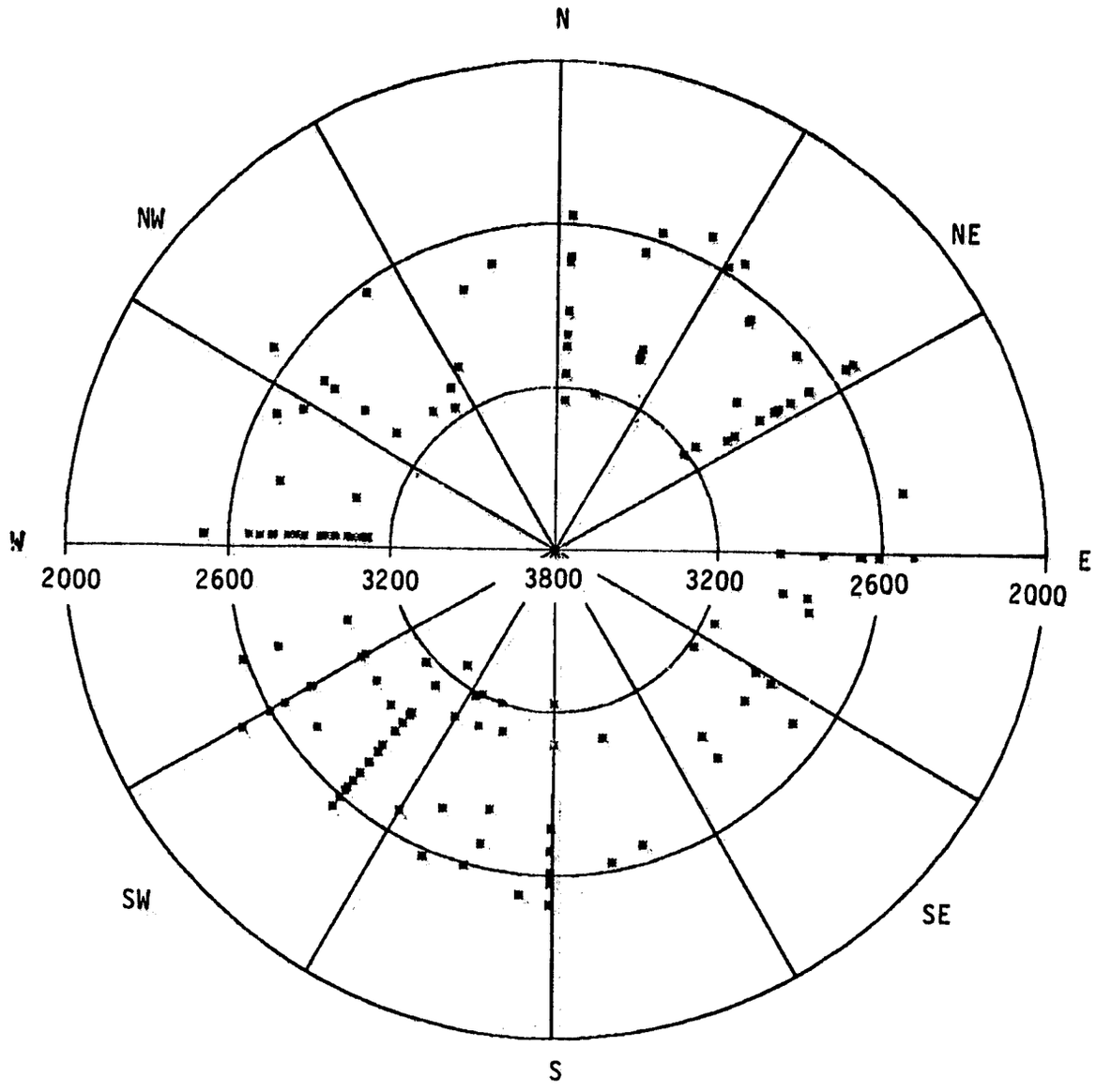


Figure C.20 Distribution of aspen as a function of elevation (in meters) and aspect.

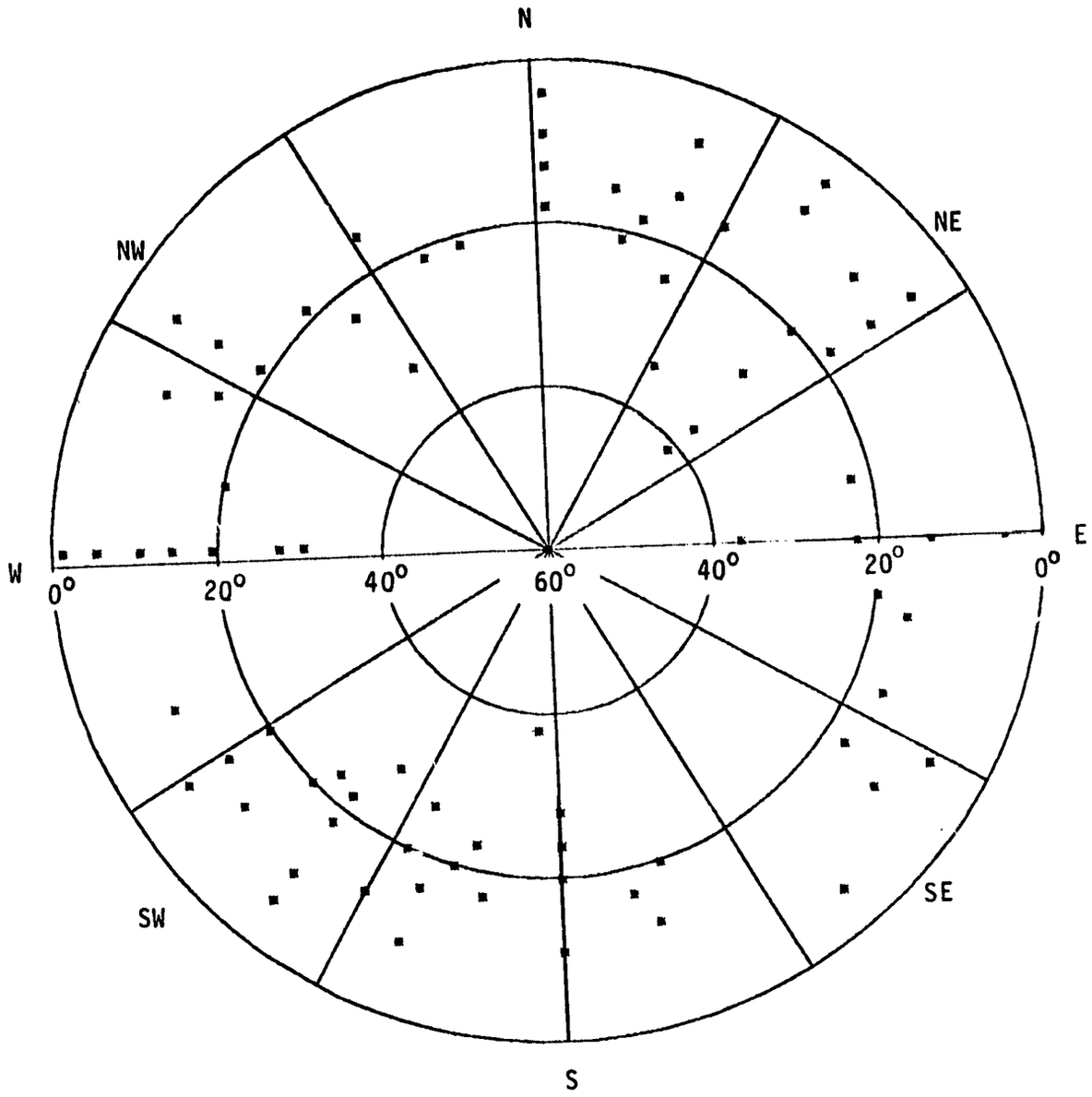


Figure C.21 Distribution of aspen as a function of slope and aspect.

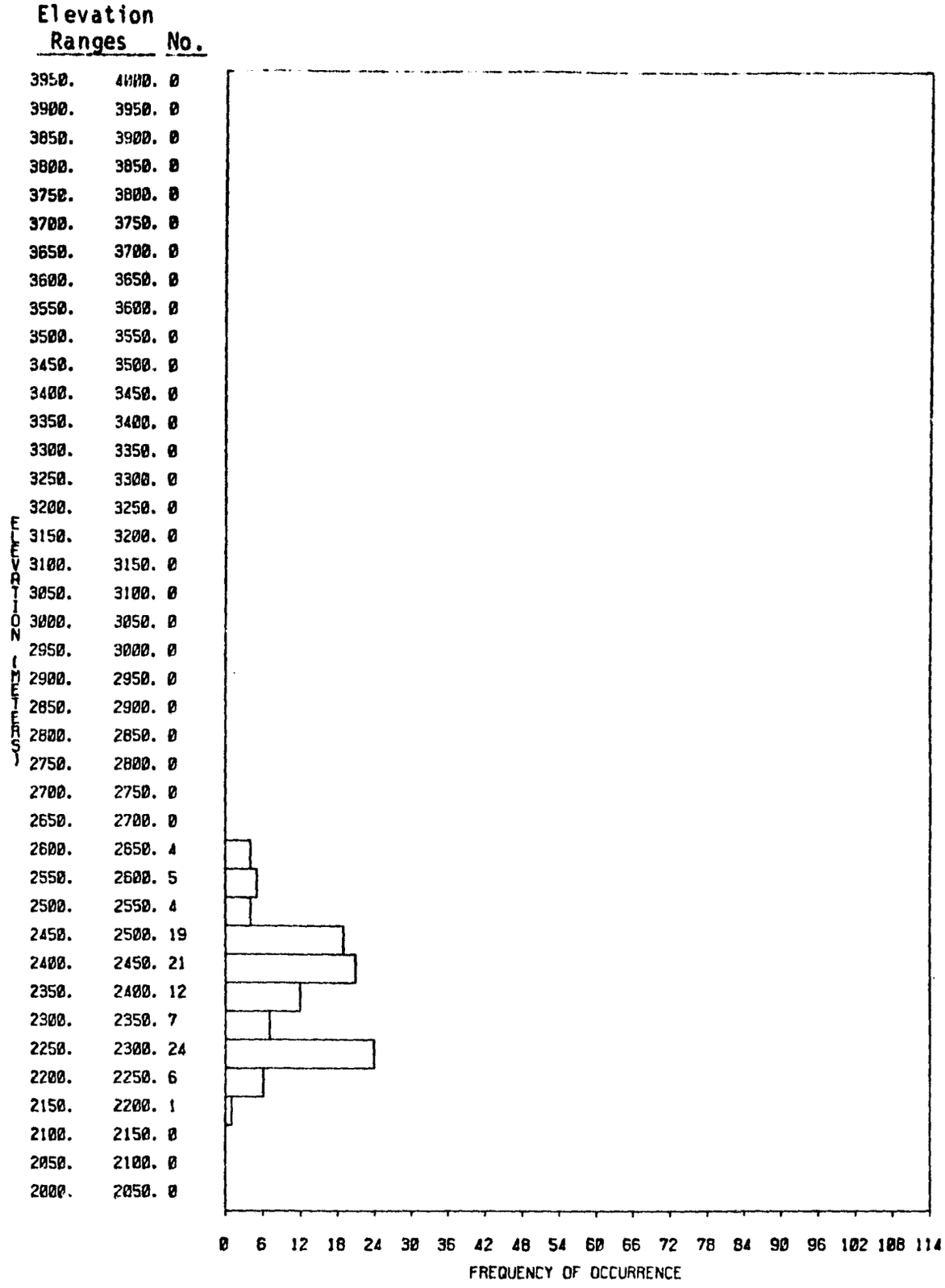


Figure C.22 Distribution of oak as a function of elevation.

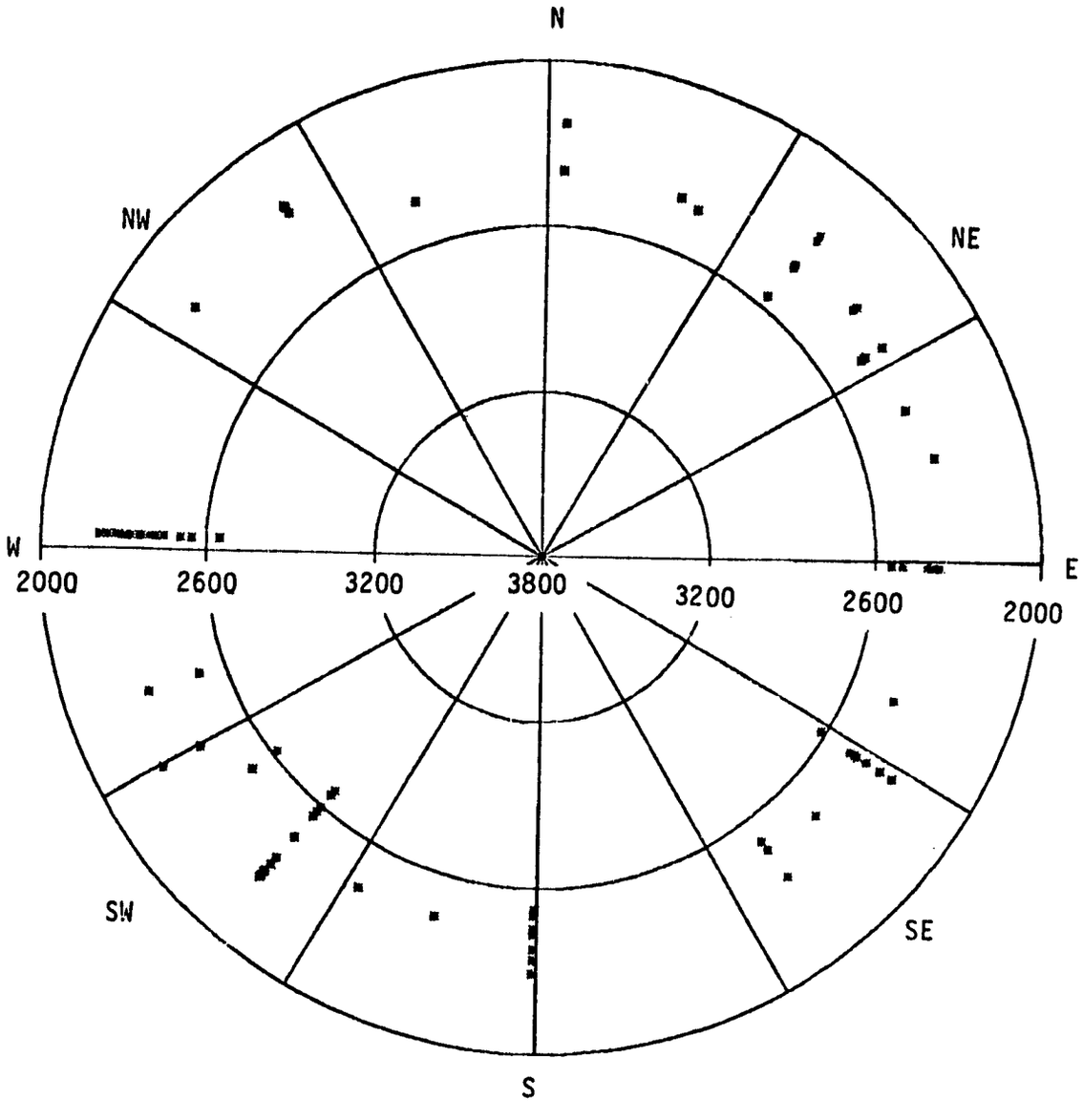


Figure C.23 Distribution of oak as a function of elevation (in meters) and aspect.

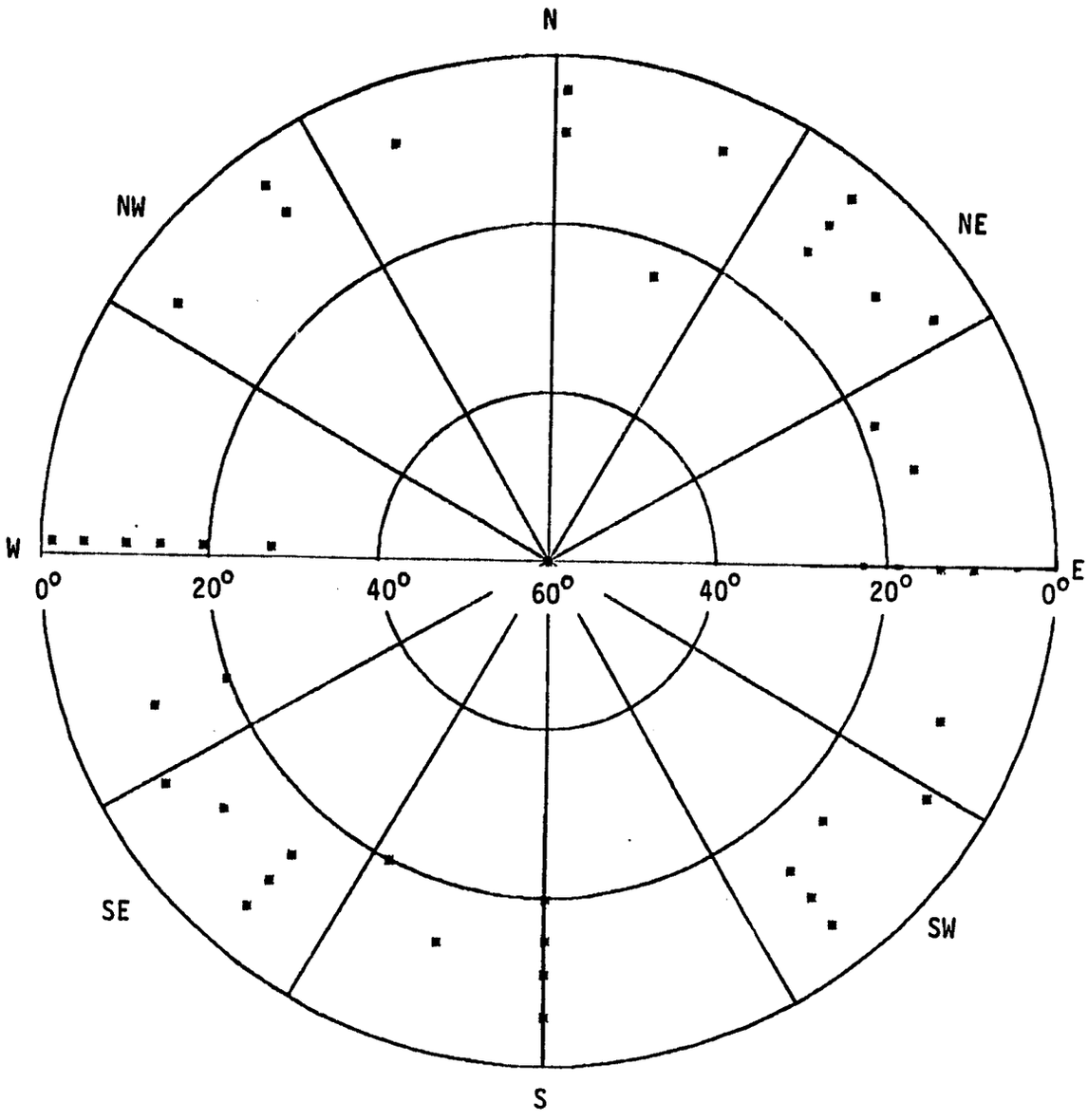


Figure C.24 Distribution of oak as a function of slope and aspect.

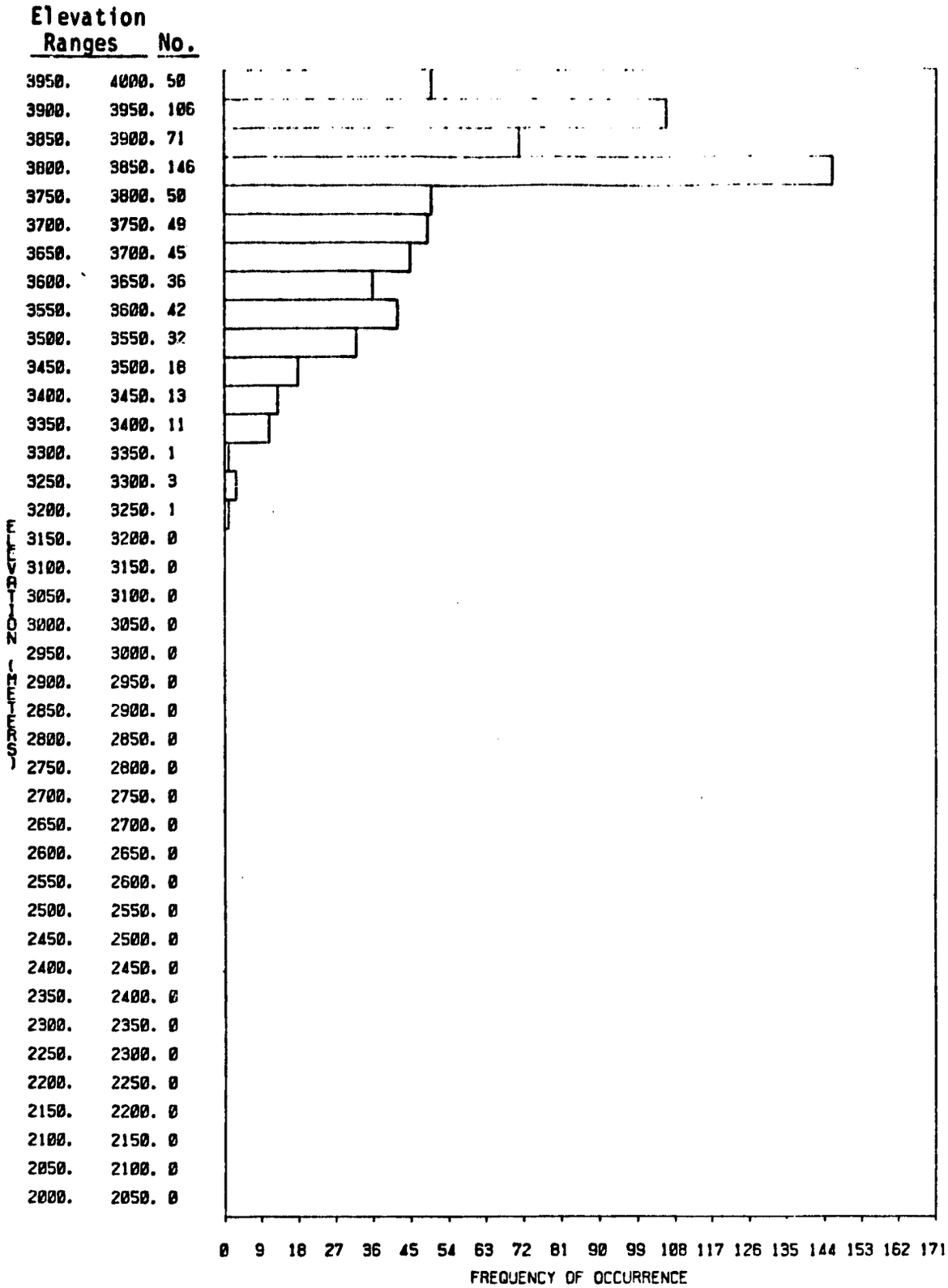


Figure C.25 Distribution of tundra as a function of elevation.

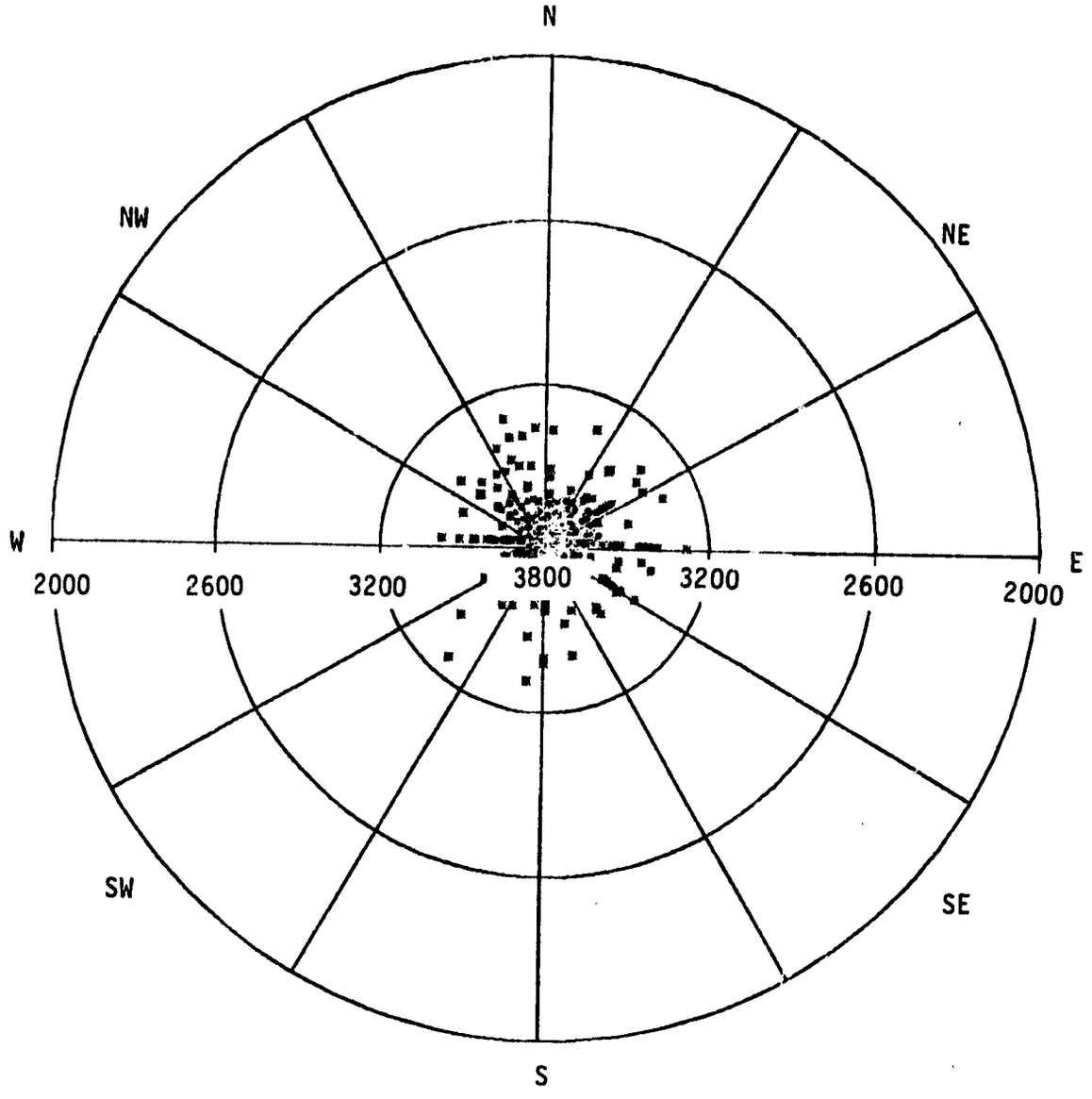


Figure C.26 Distribution of tundra as a function of elevation (in meters) and aspect.

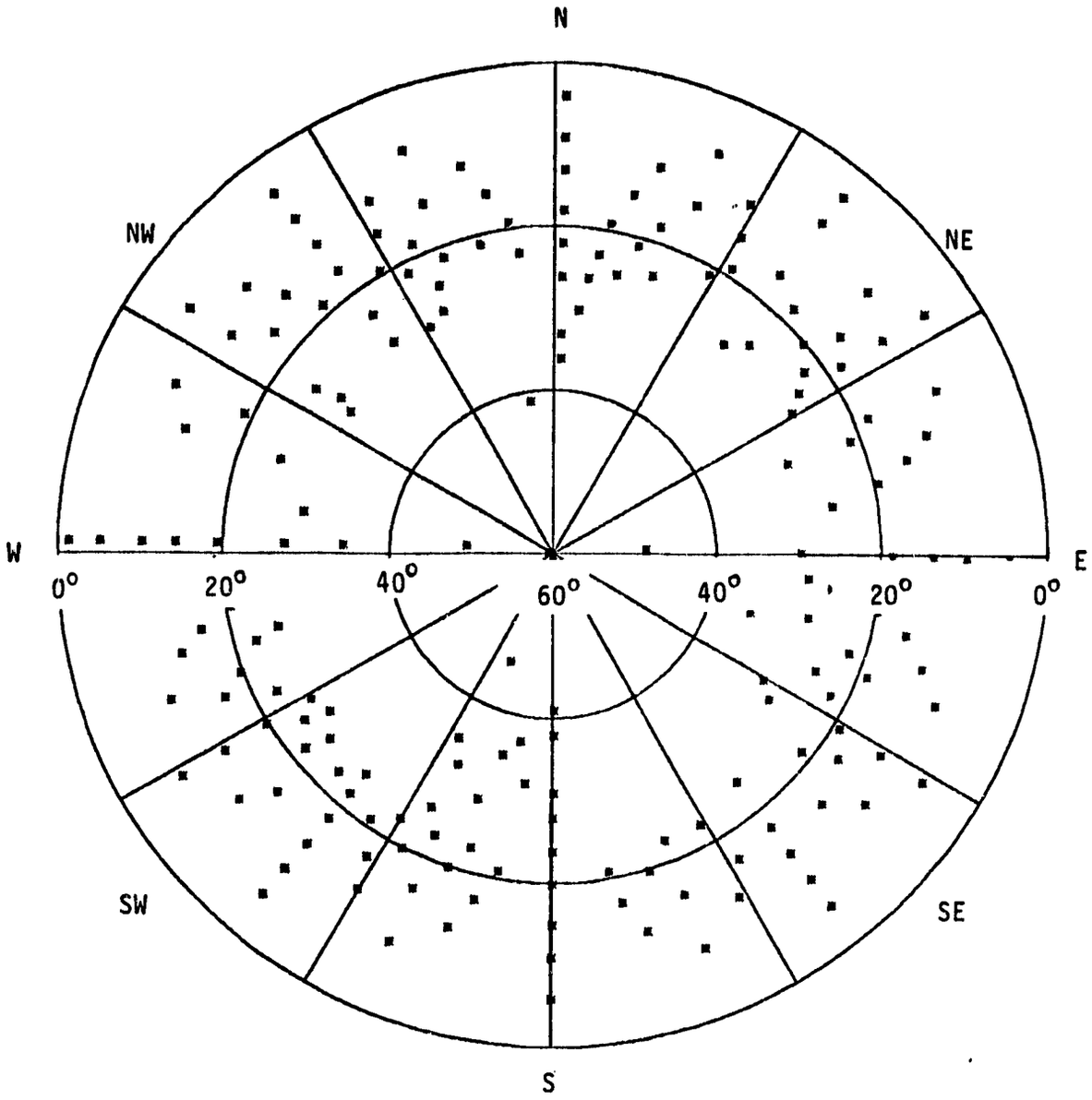


Figure C.27 Distribution of tundra as a function of slope and aspect.

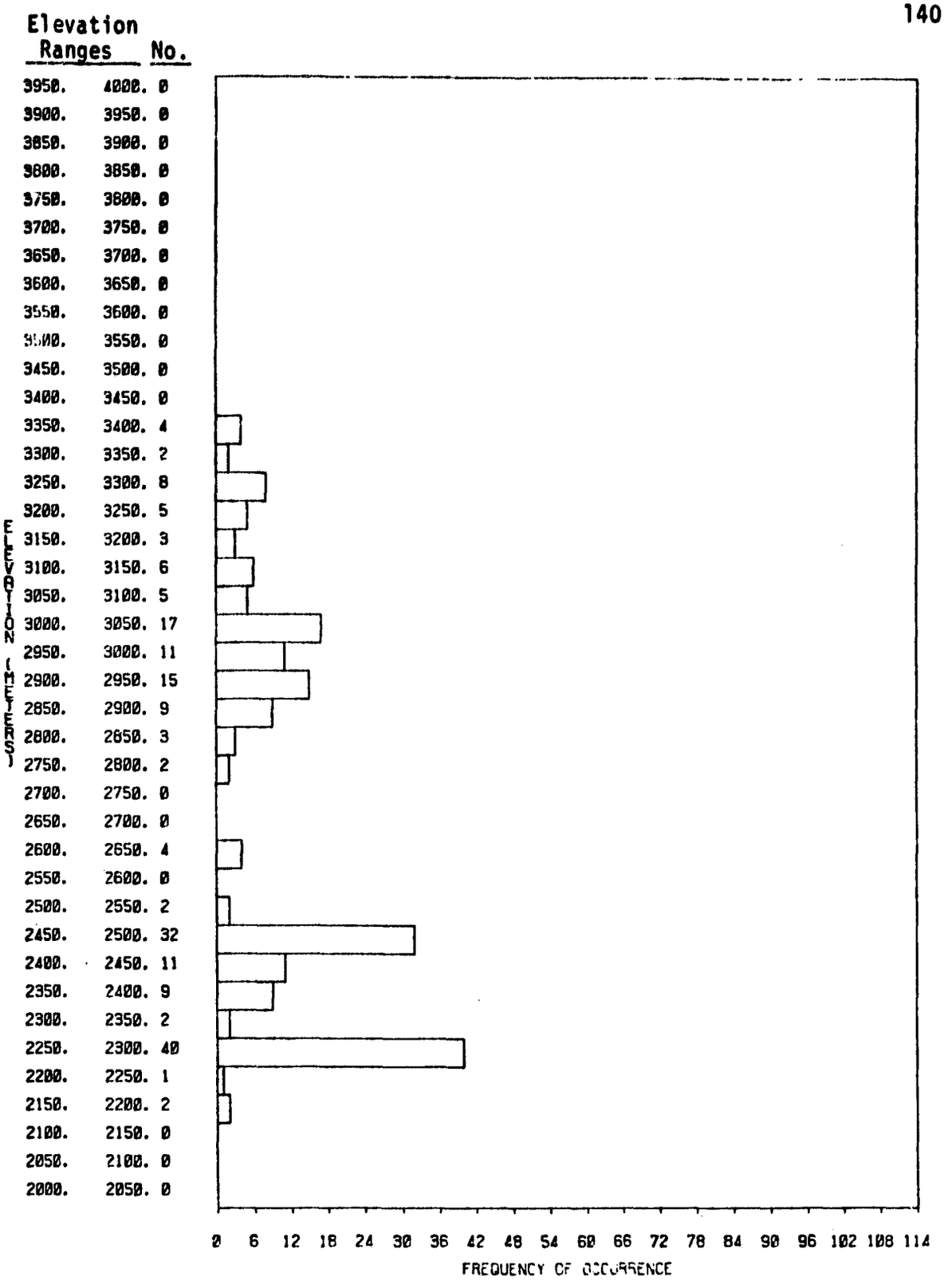


Figure C.28 Distribution of grassland as a function of elevation.

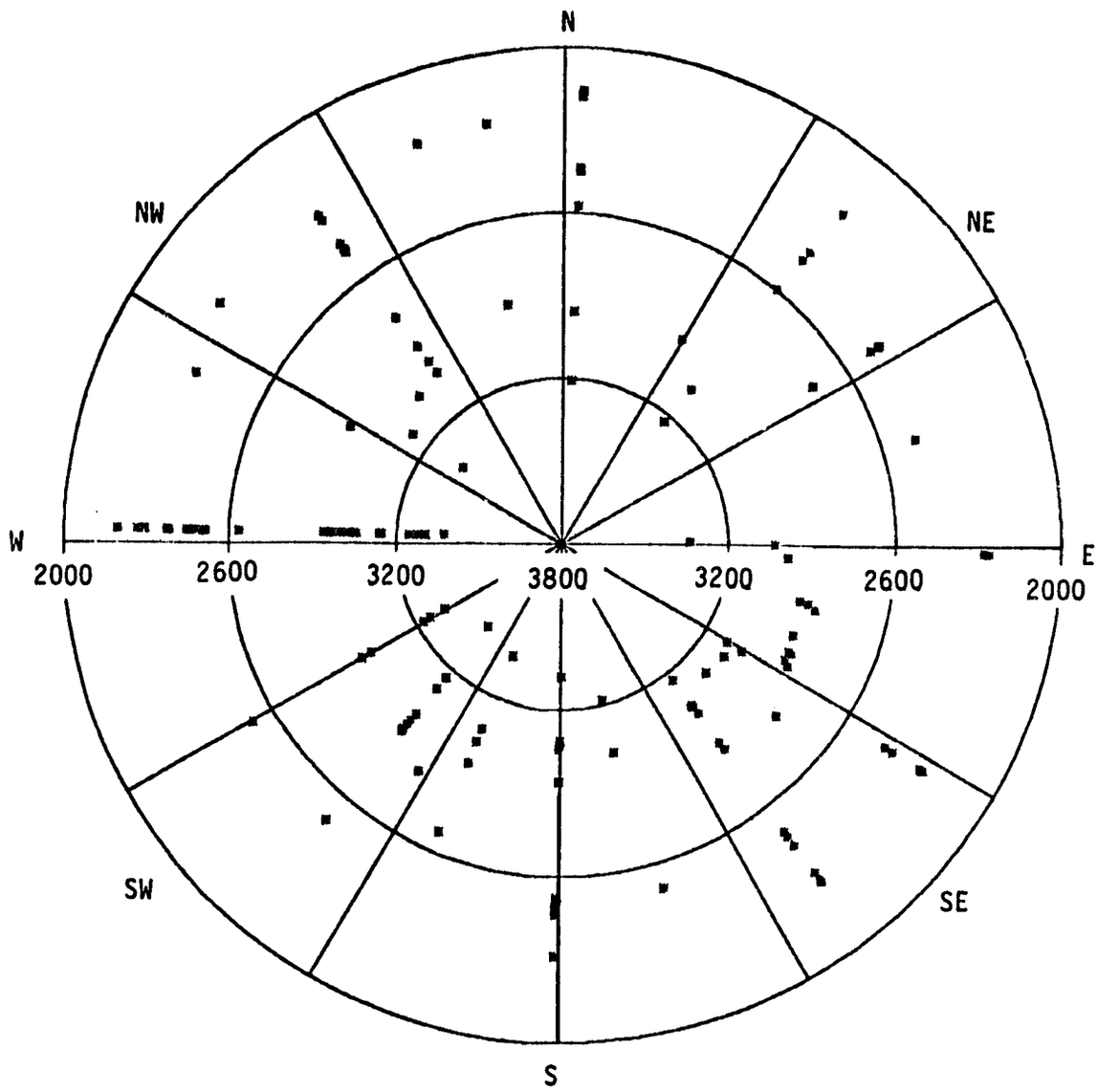


Figure C.29 Distribution of grassland as a function of elevation (in meters) and aspect.

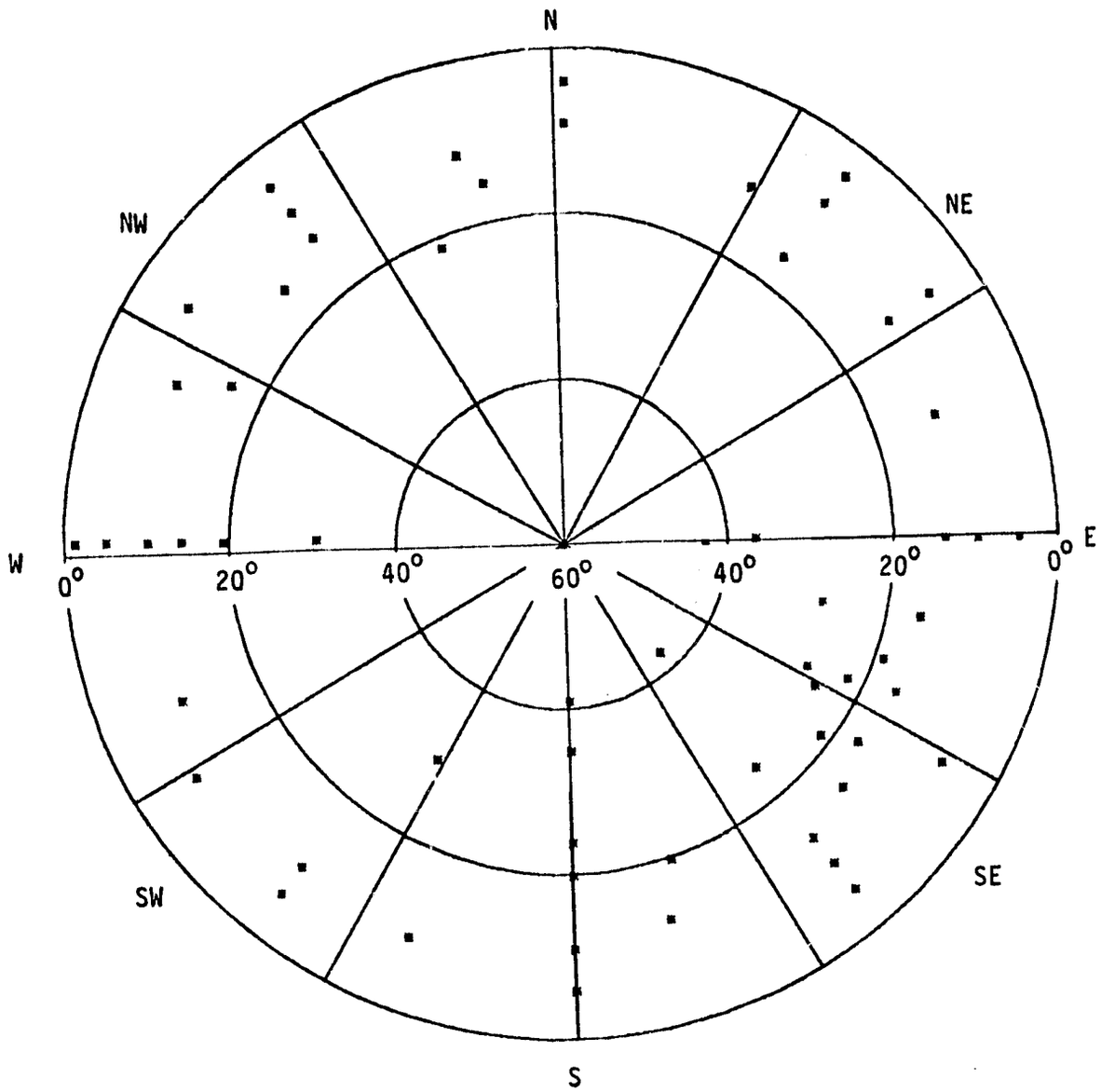


Figure C.30 Distribution of grassland as a function of slope and aspect.

Appendix D

SOFTWARE DEVELOPMENT AND MODIFICATION

To carry out the work in this project, programming activities were completed which resulted in the creation of new software and modifications in existing software. All programs were implemented on the IBM 370/148 at LARS.

1. New Programs

Three programs were written in support of the development of the topographic distribution model: EXTRACT, RANDOM, and SELECT.

Program EXTRACT. This program reads standard LARS results tapes on which classifications are stored. In this case, it read the tapes containing the classification of the data into 91 topographic strata. The input parameters to this program are: 1) the class or group of classes of interest, 2) the area location (block description card), 3) the "minpoints" parameter (see LARSYS Ver. 3.1 User's Manual for description of "minpoints"), and 4) the tape and file number in which the result classification is stored. The output of this program is a listing of the coordinates of all the points belonging to each one of the requested strata or classes. This output can be in a disk file, punched card file, and/or printer file format.

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

Program SELECT. This program basically selects the desired number of points (which in this study was 50) from all the points belonging to each one of the topographic classes. The input parameters for this program are: a list of random numbers and the list of points in the stratum. The output is a list of coordinates for the 50 selected random points. A standard LARS

field description format (LARS-12 format) is used to represent these coordinates. This format is described in detail in the LARSYS Ver. 3.1 User's Manual.

2. Program Modifications

Modifications were made in several LARSYS processors to accommodate the requirements of this project. Another major programming effort involved modification of the *LAYERED CLASSIFY function to accept multiple statistics decks, i.e., containing definitions of both spectral and topographic distributions. A second modification in the same processor was made to permit classification with either equal or weighted probabilities.

The *TRANSFERDATA was modified to provide a more satisfactory output format for meeting the requirements of this project, and the *PRINTRESULTS processor was modified to allow printing (displaying) of a symbol to locate each of the randomly selected points on a quadrangle-by-quadrangle basis. The "@" symbol is used to represent single points and a "$" to designate points that were selected randomly more than once (caused by the selection with replacement).

Appendix E

DETAILED TOPOGRAPHIC DISTRIBUTION AND CLASSIFICATION RESULTS TABLES

This Appendix contains a number of tables that were not needed for specific comparisons in the main body of the report, but which contain results that should be included in the report in support of the evaluations and conclusions.

Table E-1. Classification test pixel results and error matrix for Level II cover types.

(Classification No. 2[1/]: Spectral data only; Spectral training by Multi-Cluster Blocks; With weights; Only the first stage of the Layered classification sequence was involved, resulting in a Level-II degree of detail.)

			No. Samples Classified As:				
Cover Type	Sample Size	Percent Correctly Classified	Coniferous	Deciduous	Herbaceous	Barren	Water
Coniferous	917	82.0	752	101	50	10	4
Deciduous	252	52.0	78	131	40	3	0
Herbaceous	279	55.2	35	72	154	18	0
Barren	86	41.9	10	5	35	36	0
Water	5	60.0	2	0	0	0	3
Total	1539						
Overall Performance		69.9%					

[1/] As indicated on Table 7.

Table E-2. Classification test field results and error matrix for Level II cover types.

(Classification No. 1: Spectral data only; Spectral training by Multi-Cluster Blocks; Equal weights; Only the first stage of the Layered classification sequence was involved, resulting in a Level-II degree of detail; Evaluation based on manually selected test fields.)

			No. Samples Classified As:				
Cover Type	Sample Size	Percent Correctly Classified	Conif-erous	Decid-uous	Herba-ceous	Barren	Water
Coniferous	7920	80.6	6382	711	94	81	652
Deciduous	1953	66.8	294	1304	350	4	1
Herbaceous	2568	59.5	174	649	1528	217	0
Barren	573	66.7	21	15	155	382	0
Water	248	91.1	7	0	2	13	226
Total	13262						
Overall Performance		74.1%					

Table E-3. Classification results of training data based on Stratified Random Sample.

(Classification No. 11 Training Data: Spectral + Topographic Data; Training by Stratified Random Sample; Equal weights; Single Stage Classification.)

Forest Cover Types	Sample Size	Percent Correct Classification of Training Pixels
SF	682	90.3
SF/DWF	427	77.3
DWF	149	83.9
DWF/PP	283	75.3
PP	412	78.6
Aspen	162	37.0
Oak	135	57.0
Alpine	950	86.1
Grassland	214	57.5
Barren	364	53.0
Water	6	83.3
Total	3784	
Overall Performance		76.2%

Table E-4. Classification test results showing impact of topographic data for Level III forest cover types, summarized over all quadrangles.

(Classification Nos. 5 and 9: Spectral training by Multi-Cluster Blocks; Topographic training by Stratified Random Sample; Equal weights; Layered classifier.)

Forest Cover Types	Sample Size	Percent Correct Classification of Test Pixels	
		Spectral + Elevation Data	Spectral + Topographic Data
SF	313	88.5	89.1
SF/DWF	156	60.3	58.3
DWF	39	23.1	46.2
DWF/PP	144	83.3	81.6
PP	265	60.8	60.0
Aspen	110	42.7	43.6
Oak	97	48.5	46.4
Alpine	79	72.2	70.9
Grassland	245	50.6	50.6
Barren	86	46.5	46.5
Water	5	60.0	60.0
Total	1539		
Overall Performance		63.6%	63.6%

Table E-5. Classification test results showing impact of topographic data for Level III forest cover types, by quadrangle, and using weights.

(Classification Nos. 4, 8, and 12: Training by Stratified Random Sample; Weighted a priori probabilities of occurrence; Singe-Stage classifier.)

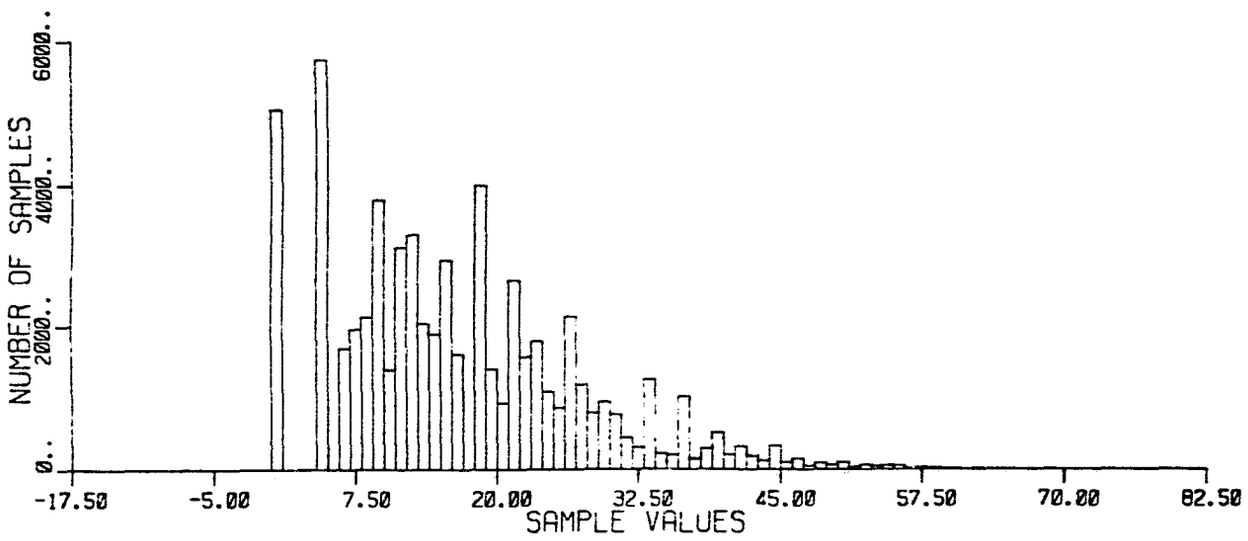
		Percent Correct Classification of Test Pixels		
Quadrangle	Sample Size	Spectral Data Only	Spectral + Elevation Data	Spectral + Topographic Data
Oakbrush	199	37.7	54.3	59.3
Finger Mesa	214	46.5	66.5	63.3
Granite Peaks	202	72.3	75.2	77.7
Pagosa Springs	237	25.8	64.7	63.1
Devil Mountain	233	39.5	57.1	61.8
Weminuche	212	64.2	75.5	74.1
Ludwig Mountain	242	21.1	58.7	53.7
Total	1539			
Overall Performance		42.9%	64.5%	64.4%

Table E-6. Classification test results showing impact of topographic data for Level III forest cover types, summarized over all quadrangles, and using weights.

(Classification Nos. 4, 8, and 12: Training by Stratified Random Sample; Weighted a priori probabilities of occurrence; Single-stage classifier.)

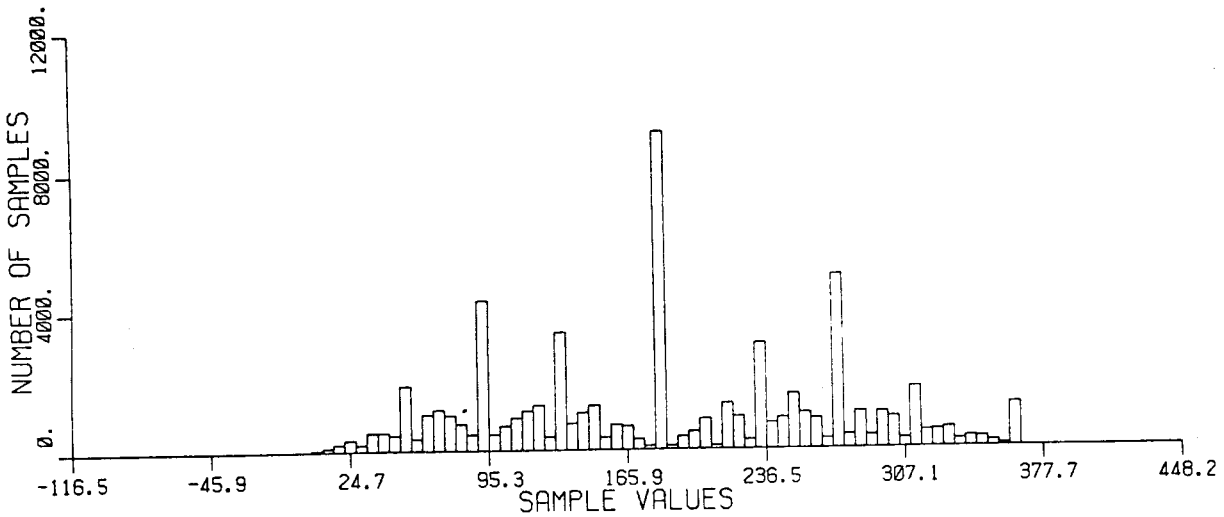
Forest Cover Types	Sample Size	Percent Correct Classification of Test Pixels		
		Spectral Data Only	Spectral + Elevation Data	Spectral + Topographic Data
SF	313	89.8	91.7	93.0
SF/DWF	156	82.7	81.4	82.1
DWF	39	38.5	61.5	51.3
DWF/PP	144	11.8	41.7	51.4
PP	265	19.6	69.4	69.1
Aspen	110	26.4	32.7	25.5
Oak	97	8.2	36.1	28.9
Alpine	79	36.7	78.5	75.9
Grassland	245	33.5	57.1	60.0
Barren	86	19.8	39.5	33.7
Water	5	40.0	60.0	60.0
Total	1539			
Overall Performance		42.9%	64.5%	64.4%

Table E-7. Tabular summary and histogram of slope data (Channel 6).



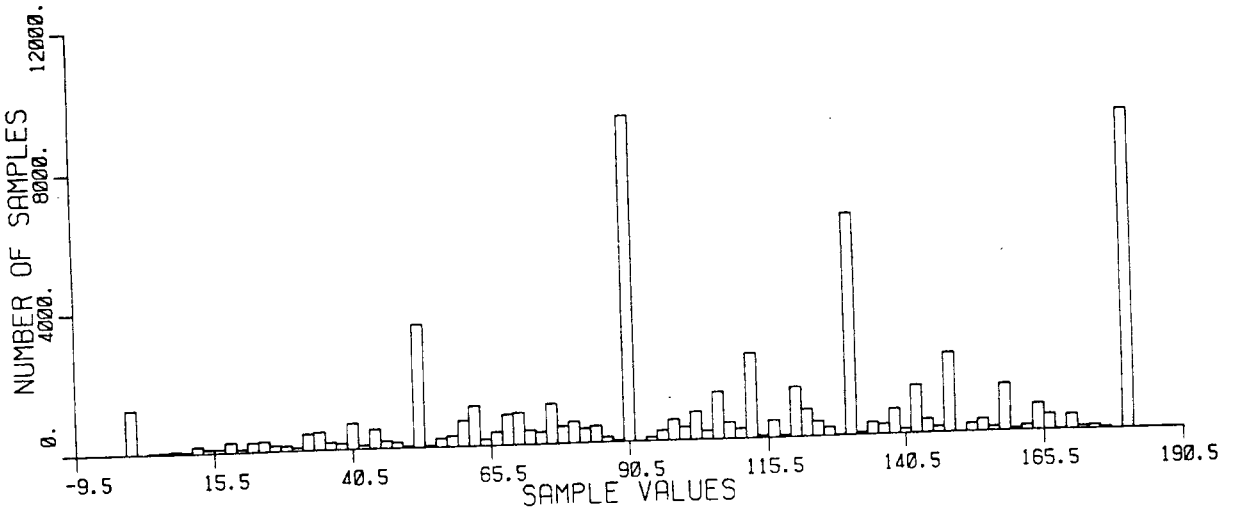
RANGE		SAMPLES
-0.5 -	0.5	5069
3.5 -	4.5	5738
5.5 -	6.5	1675
6.5 -	7.5	1947
7.5 -	8.5	2120
8.5 -	9.5	3781
9.5 -	10.5	1382
10.5 -	11.5	3114
11.5 -	12.5	3291
12.5 -	13.5	2026
13.5 -	14.5	1877
14.5 -	15.5	2933
15.5 -	16.5	1602
17.5 -	18.5	3991
18.5 -	19.5	1391
19.5 -	20.5	910
20.5 -	21.5	2651
21.5 -	22.5	1556
22.5 -	23.5	1784
23.5 -	24.5	1074
24.5 -	25.5	852
25.5 -	26.5	2127
26.5 -	27.5	1167
27.5 -	28.5	778
28.5 -	29.5	930
29.5 -	30.5	761
30.5 -	31.5	444
31.5 -	32.5	304
32.5 -	33.5	1242
33.5 -	34.5	219
34.5 -	35.5	205
35.5 -	36.5	1003
36.5 -	37.5	145
37.5 -	38.5	293
38.5 -	39.5	513
39.5 -	40.5	208
40.5 -	41.5	315
41.5 -	42.5	194
42.5 -	43.5	121
43.5 -	44.5	326
44.5 -	45.5	95
45.5 -	46.5	146
46.5 -	47.5	54
47.5 -	48.5	91
48.5 -	49.5	60
49.5 -	50.5	95
50.5 -	51.5	25
51.5 -	52.5	60
52.5 -	53.5	43
53.5 -	54.5	47
54.5 -	55.5	47
55.5 -	56.5	10
56.5 -	57.5	27
57.5 -	58.5	10
58.5 -	59.5	5
59.5 -	60.5	6
60.5 -	61.5	4
61.5 -	62.5	10
62.5 -	63.5	3
63.5 -	64.5	1
64.5 -	65.5	3

Table E-8. Tabular summary and histogram of aspect data in Channel 9 (0-360^{0}).



RANGE		SAMPLES	RANGE		SAMPLES	RANGE		SAMPLES
2.1 -	7.8	31	126.4 -	132.0	3359	244.9 -	250.6	1553
7.8 -	13.4	100	132.0 -	137.7	760	250.6 -	256.2	1024
13.4 -	19.1	198	137.7 -	143.3	1054	256.2 -	261.9	864
19.1 -	24.7	312	143.3 -	148.9	1266	261.9 -	267.5	274
24.7 -	30.4	179	148.9 -	154.6	356	267.5 -	273.2	4979
30.4 -	36.0	512	154.6 -	160.2	726	273.2 -	278.8	379
36.0 -	41.7	524	160.2 -	165.9	672	278.8 -	284.5	1046
41.7 -	47.3	446	165.9 -	171.5	304	284.5 -	290.1	357
47.3 -	52.9	1835	171.5 -	177.2	93	290.1 -	295.8	1021
52.9 -	58.6	347	177.2 -	182.8	9102	295.8 -	301.4	871
58.6 -	64.2	1026	182.8 -	188.5	96	301.4 -	307.1	253
64.2 -	69.9	1169	188.5 -	194.1	353	307.1 -	312.7	1722
69.9 -	75.5	1009	194.1 -	199.8	492	312.7 -	318.4	471
75.5 -	81.2	753	199.8 -	205.4	851	318.4 -	324.0	509
81.2 -	86.8	453	205.4 -	211.1	92	324.0 -	329.7	558
86.8 -	92.5	4307	211.1 -	216.7	1308	329.7 -	335.3	227
92.5 -	98.1	465	216.7 -	222.4	919	335.3 -	340.9	290
98.1 -	103.8	695	222.4 -	228.0	267	340.9 -	346.6	280
103.8 -	109.4	924	228.0 -	233.7	3028	346.6 -	352.2	153
109.4 -	115.1	1113	233.7 -	239.3	730	352.2 -	357.9	69
115.1 -	120.7	1274	239.3 -	244.9	889	357.9 -	363.5	1235
120.7 -	126.4	377						

Table E-9. Tabular summary and histogram of aspect data in Channel 7 (0-180°).



RANGE		SAMPLES
-1.5 -	0.5	1235
0.5 -	2.5	2
2.5 -	4.5	16
4.5 -	6.5	36
6.5 -	8.5	64
8.5 -	10.5	48
10.5 -	12.5	183
12.5 -	14.5	104
14.5 -	16.5	93
16.5 -	18.5	279
18.5 -	20.5	80
20.5 -	22.5	263
22.5 -	24.5	275
24.5 -	26.5	165
26.5 -	28.5	144
28.5 -	30.5	89
30.5 -	32.5	468
32.5 -	34.5	496
34.5 -	36.5	204
36.5 -	38.5	167
38.5 -	40.5	723
40.5 -	42.5	105
42.5 -	44.5	541
44.5 -	46.5	201
46.5 -	48.5	150
48.5 -	50.5	41
50.5 -	52.5	3478
52.5 -	54.5	44
54.5 -	56.5	235
56.5 -	58.5	309
58.5 -	60.5	715
60.5 -	62.5	1113
62.5 -	64.5	177
64.5 -	66.5	382
66.5 -	68.5	850
68.5 -	70.5	902
70.5 -	72.5	395
72.5 -	74.5	333
74.5 -	76.5	1120
76.5 -	78.5	471
78.5 -	80.5	602
80.5 -	82.5	410
82.5 -	84.5	452
84.5 -	86.5	144
86.5 -	88.5	39
88.5 -	90.5	9267
90.5 -	92.5	2
92.5 -	94.5	100
94.5 -	96.5	284
96.5 -	98.5	574
98.5 -	100.5	364
100.5 -	102.5	775
102.5 -	104.5	236
104.5 -	106.5	1318
106.5 -	108.5	464
108.5 -	110.5	280
110.5 -	112.5	2393
112.5 -	114.5	62
114.5 -	116.5	482
116.5 -	118.5	60
118.5 -	120.5	1401
120.5 -	122.5	751
122.5 -	124.5	415
124.5 -	126.5	233
126.5 -	128.5	16
128.5 -	130.5	6329
130.5 -	132.5	57
132.5 -	134.5	339
134.5 -	136.5	280
136.5 -	138.5	690
138.5 -	140.5	121
140.5 -	142.5	1338
142.5 -	144.5	385
144.5 -	146.5	165
146.5 -	148.5	2240
148.5 -	150.5	6
150.5 -	152.5	211
152.5 -	154.5	336
154.5 -	156.5	125
156.5 -	158.5	1311
158.5 -	160.5	50
160.5 -	162.5	136
162.5 -	164.5	750
164.5 -	166.5	431
166.5 -	168.5	17
168.5 -	170.5	429
170.5 -	172.5	79
172.5 -	174.5	102
174.5 -	176.5	49
176.5 -	178.5	4
178.5 -	180.5	9101

Appendix F

STATISTICAL EVALUATION OF CLASSIFICATION ACCURACIES

The tables included in this appendix summarize the 2-factor analysis-of-variance tests for significant differences among classifications, and at the same time, among quadrangles or cover types. Thus, two factors (classifications and quadrangles or classifications and cover types) are tested simultaneously. The different classifications evaluated include: (1) the type of data Channels Utilized (spectral only, spectral + elevation, and spectral + topographic), (2) the Analysis Approaches used to train and classify the data (Techniques A and B), (3) the utilization of a priori weights, and (4) the spectral Data Source used (original and "spectrally corrected"). The appropriate procedure followed to develop the ANOVA's for this application is described in the 1976 SRT final report (Landgrebe, 1976).

DATA SOURCE

ANOVA

Source	df	SS	MS	F	α=.10 Fcrit
DATA SOURCE	1	0.93	0.93	0.25	2.71
QUADRANGLES	6	394.21	65.70	17.50	1.77
Interaction/error	∞	11.83	3.754*		
Total	13	406.97			

ANOVA

Source	df	SS	MS	F	α=.10 Fcrit
DATA SOURCE	1	14.08	14.08	1.45	3.29
COVER TYPES	10	2606.01	260.60	26.84	2.32
Interaction/error	10	97.13	9.71		
Total	21	2717.22			

* residual mean square non-significant

ANALYSIS APPROACHES

ANOVA

Source	df	SS	MS	F	α=.10 Fcrit
APPROACHES	1	7.87	7.87	2.10	3.78
QUADRANGLES	6	369.92	61.65	16.42	3.05
Interaction/error	6	21.16	3.53		
Total	13	398.95			

ANOVA

Source	df	SS	MS	F	α=.10 Fcrit
APPROACHES	1	12.53	12.53	0.61	3.29
COVER TYPES	10	2139.77	213.98	10.49	2.32
Interaction/error	10	203.94	20.39		
Total	21	2356.24			

WEIGHTS

ANOVA

Source	df	SS	MS	F	α=.10 Fcrit
WEIGHTS	1	1.26	1.26	0.34	2.71
QUADRANGLES	6	339.41	56.57	15.07	1.77
Interaction/error	∞	5.53	3.754*		
Total	13	346.197			

ANOVA

Source	df	SS	MS	F	α=.10 Fcrit
WEIGHTS	1	34.12	34.125	1.64	3.29
COVER TYPES	10	3143.49	314.35	15.13	2.32
Interaction/error	10	207.77	20.78		
Total	21	3385.39			

* residual mean square non significant

CHANNELS UTILIZED

ANOVA

Source	df	SS	MS	F	α=.10 Fcrit
CHANNELS	2	426.27	213.1	42.37	2.81
QUADRANGLES	6	436.61	72.8	14.47	2.33
Interaction/error	12	60.42	5.03		
Total	20	923.30			

ANOVA

Source	df	SS	MS	F	α=.10 Fcrit
CHANNELS	2	611.17	305.58	7.42	2.59
COVER TYPES	10	2906.07	290.61	7.06	1.94
Interaction/error	20	823.14	41.16		
Total	32	4340.38			