LARS Contract Report 101579

Quarterly Progress Report

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

Contract No. NAS 9-15508

Reporting Period: July 16, 1979 - October 15, 1979

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

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

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

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

October 1979

TABLE OF CONTENTS

I. OVERALL STATUS AND PROGRESS TO DATE

A. Training of Washington DNR Personnel

During the period August 27 - September 3, 1979, Eric Barthmaier, Tim Gregg, and Larry Sugarbaker, of the Washington Department of Natural Resources participated in a two-day intensive training course at LARS. The primary aim of this course was to provide an introduction to the LARSYS data processing capabilities and the Landsat MSS analysis procedures developed at LARS. After this introductory short course, the DNR representatives were actively involved in the analysis of a Landsat MSS data set from the state of Washington.

B. Spectral Analysis of the Washington Landsat MSS Data

The region designated as the primary study site is a 18-mile by 48-mile area in the Okanogan valley of north-central Washington. The area is defined by 18 adjacent townships. To locate the study area in the Landsat MSS data, a dot-matrix grayscale image of Landsat band 6 was generated at a scale of 1:60,000. The coordinates (lines and columns) of the site were then determined to be lines 280 to 1320 and columns 450 to 950, a total of 521,541 pixels (approximately 600,000 acres).

The analysis approach for developing the spectral training statistics was a Multi-Cluster Blocks (Fleming, 1977) procedure. The five steps are: 1) Selection of blocks, 2) Individual Cluster of each block, 3) Identification of each cluster class, 4) Pooling of the spectral classes into spectral/informational classes, and 5) Test classification and evaluation of the classes.

A total of 15 blocks, each approximately 40 lines by 40 columns were selected throughout the entire study site. Fourteen of the blocks were grouped into similar pairs and each pair of blocks were then clustered into

15 spectral classes. The extra (15th) block was clustered by itself into 15 spectral classes. One pair of blocks had to be re-clustered into more classes (20) to obtain an adequate amount of detail. Thus, a total of 125 cluster classes were defined. All cluster classes were identified using the available 1:24,000 black and white orthophotos of each township and the quality of each class rated. A 4-step iterative approach was used to pool the 125 cluster classes into 28 spectrally separable spectral/informational classes. First, mixture classes and other poor quality classes were deleted. Next all class pairs with a separability below 500 were pooled, then those below 1000, and finally those below 1500 were also pooled.

Once the spectral/informational training classes were defined, six representative townships were selected for a trial classification and evaluation. An evaluation of all 28 classes indicated that most of the classes were accurately defined and identified, with only minor problems in a couple of classes.

C. Analysis of the GRIDS Data and Development of the Forest Topographic Distribution Model

In order to obtain a set of topographic training statistics, a forest topographic distribution model was developed following the methodology defined during the first year of this investigation (Colorado test site). To accomplish this task using the GRIDS data, three new computer programs had to be developed.

1. Software Development

The three computer programs developed during this phase of the project were:

PREP: This program reads the GRIDS data tape and verifies if each sample a) is within the study site, b) is duplicated on the GRIDS tape due to the updating procedure, c) contains topographic information. It also determines the topographic position of the samples and writes the data in a random access format and assigns a sorting index for the various topographic positions.

SELECT: This program selects a stratified random sample (without replacement) of 50 points from each of the 65 topographic strata defined for the Washington study area [5 elev. x ({3 slope x 4 asp.} + 1 flat)]. In this case, the resulting sample included 3201 points (rather than 3250) because two strata did not have the required 50 points. This program also determines the species codes for each sample and generates a sorting index for the species codes.

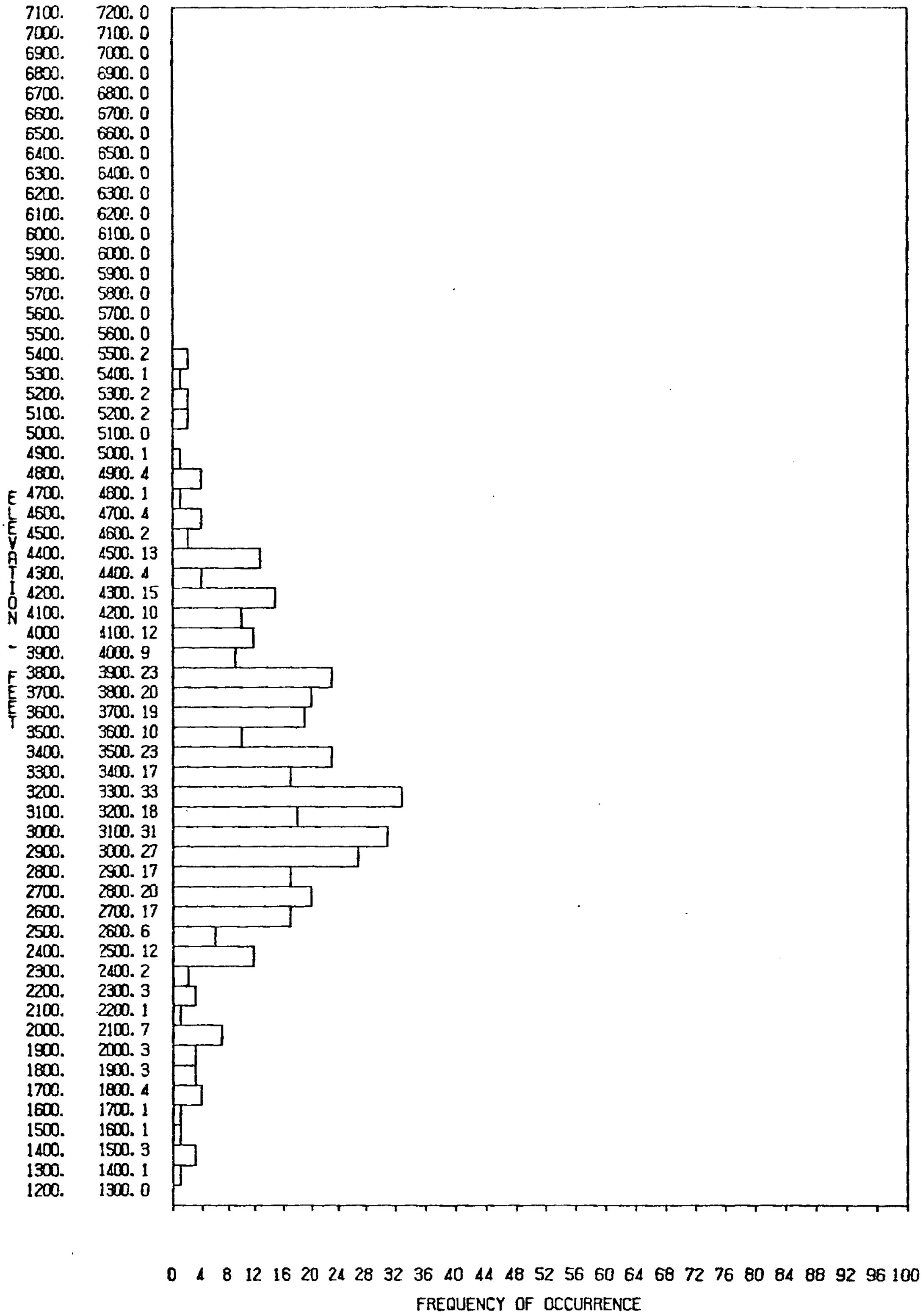
SUMMARY: This program summarizes the topographic characteristics for each species code as illustrated in Table 1, and it groups the data into data files for each species.

In addition to the development of these new computer programs, several plotting and statistical sub-routines were also modified.

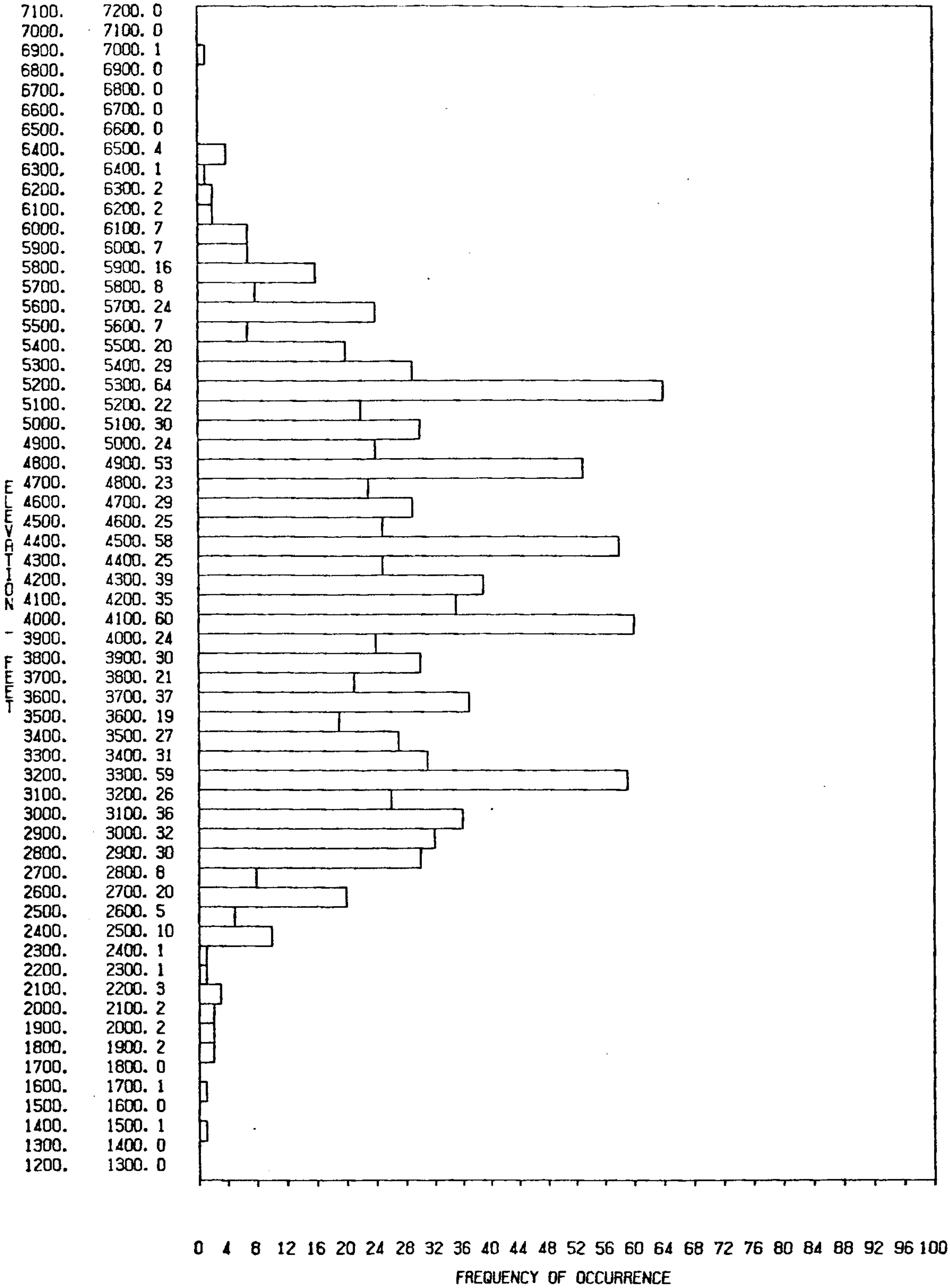
2. The results of the forest topographic distribution model for the Washington test site are illustrated by examples of histogram graphs, normalized distribution curves, and polar plots shown on the following pages of this report.

D. DMA - Digital Terrain Model Tape Reformatting and Geometric Correction

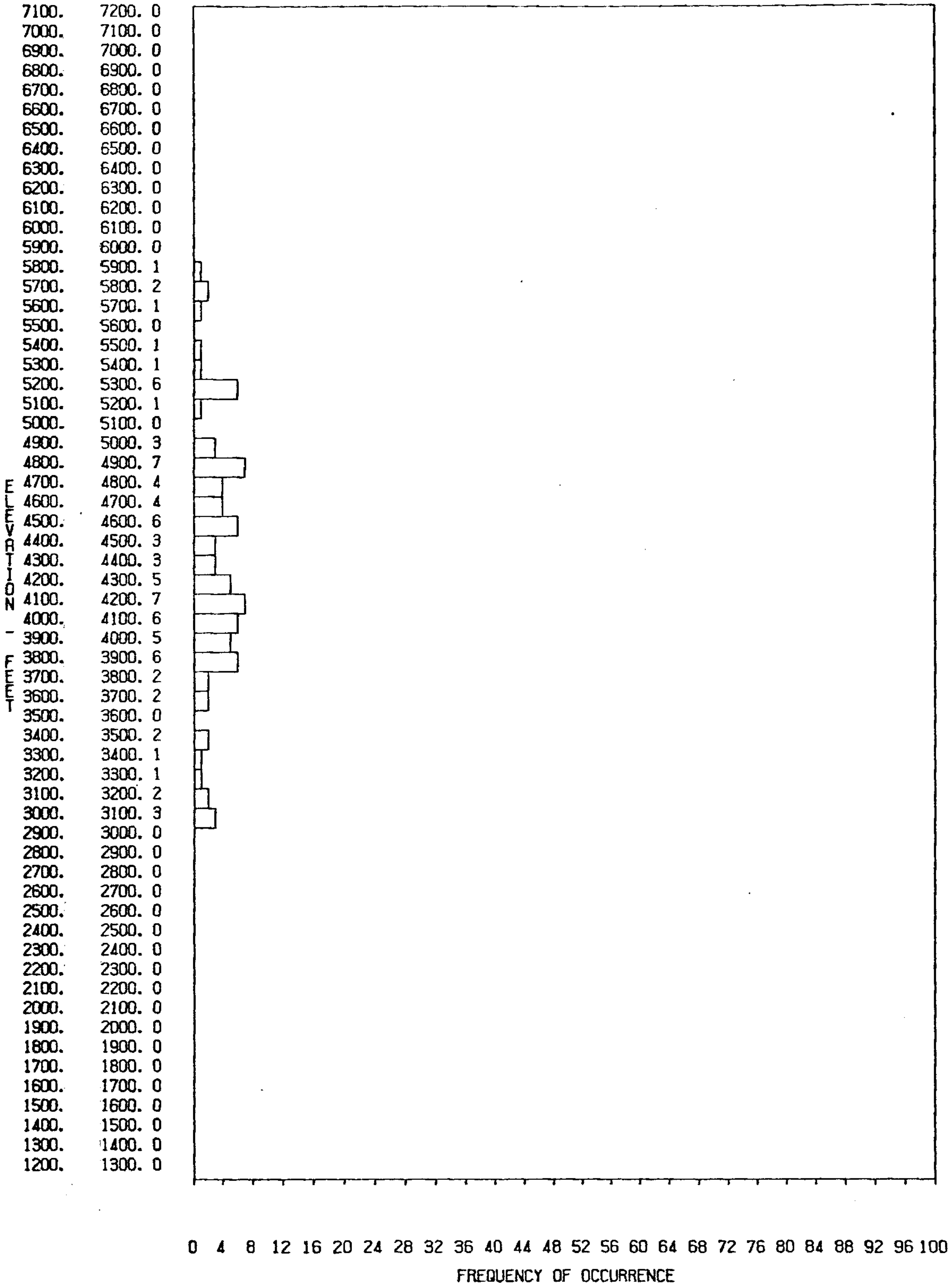
A digital terrain model tape containing the topographic data for the Okanogan 1:250,000 quadrangle was ordered from the Defense Mapping Agency. Upon arrival, it was discovered that the tape format had been modified, as compared to the previous DMA tape format. Since the format of this tape was not compatible with the existing LARSYS data storage tape format and due to changes within the LARSYS system, a new reformatting and geometric correction software package had to be developed. This geometric correction refers to



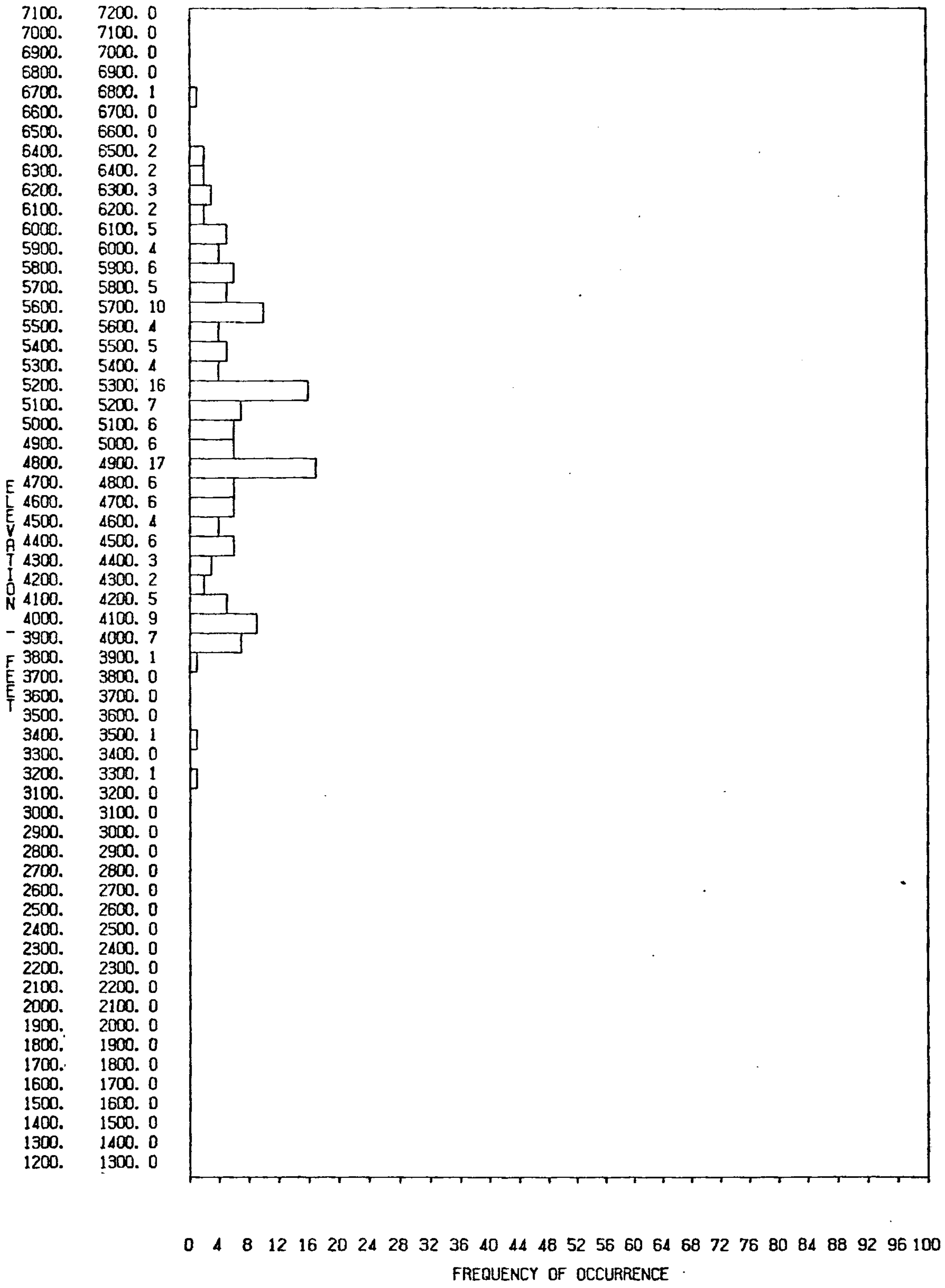
Frequency of Ponderosa Pine as a function of elevation.



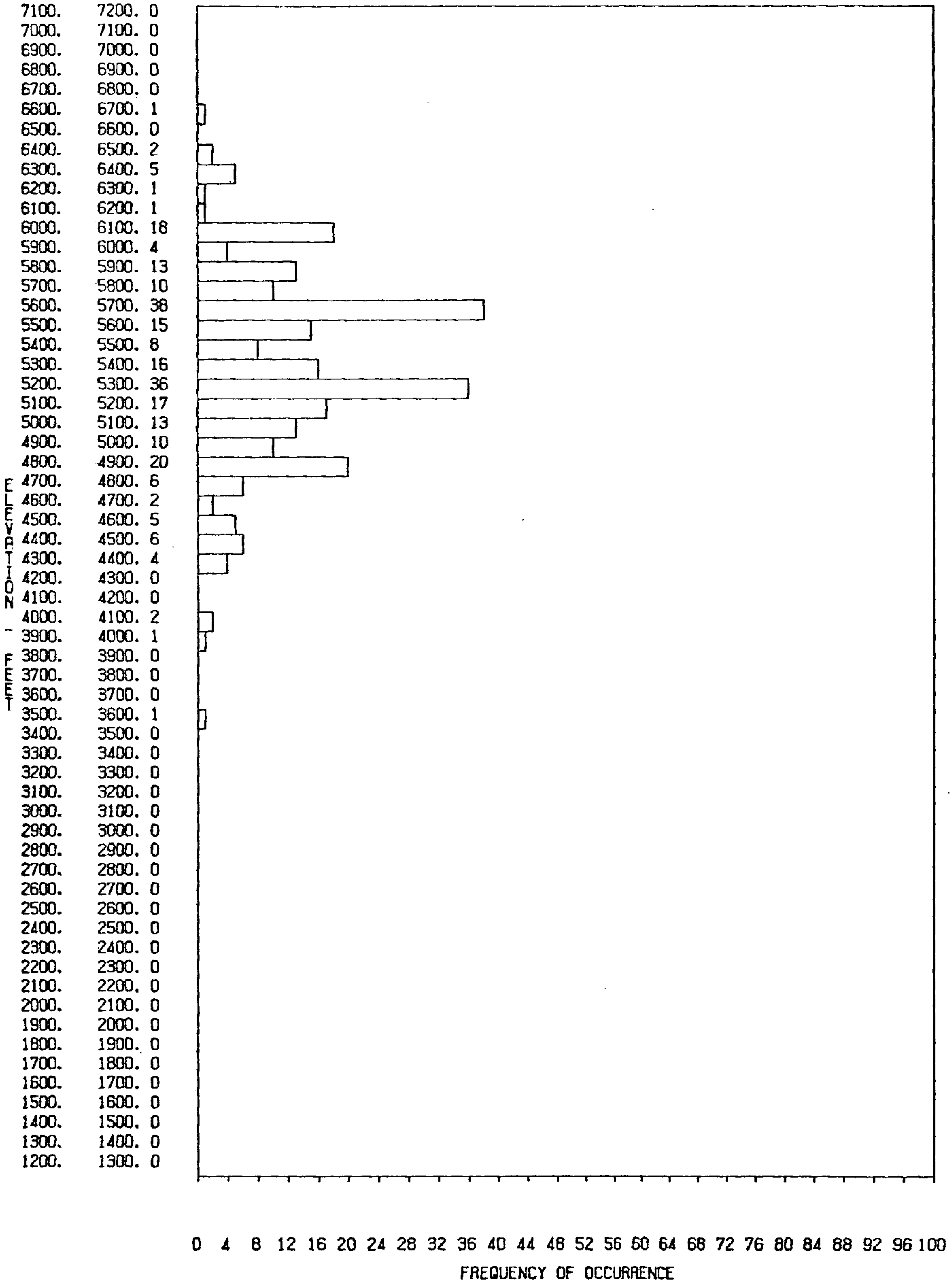
Frequency of Douglas Fir as a function of elevation.



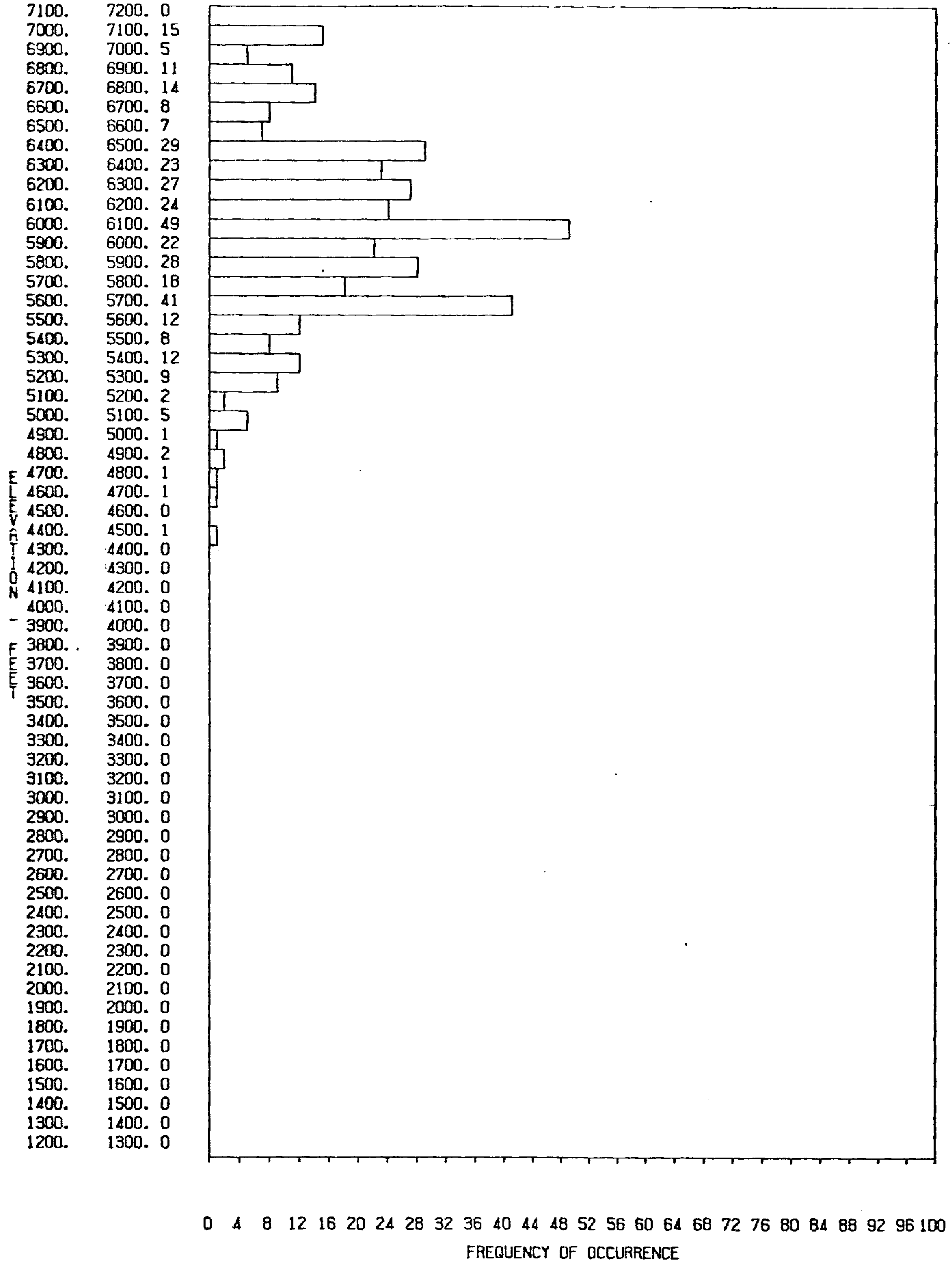
Frequency of Larch as a function of elevation.



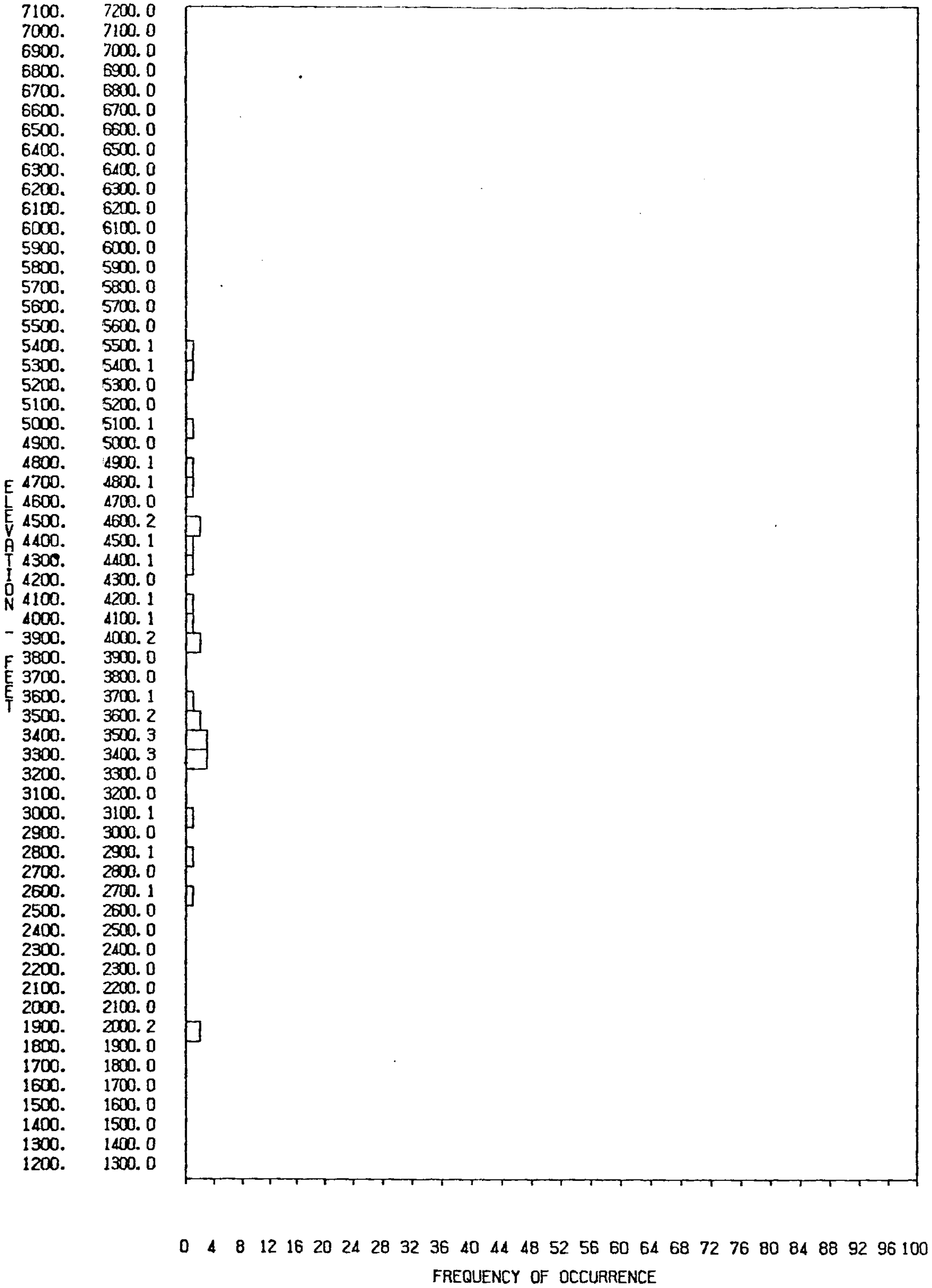
Frequency of Spruce/Fir as a function of elevation.



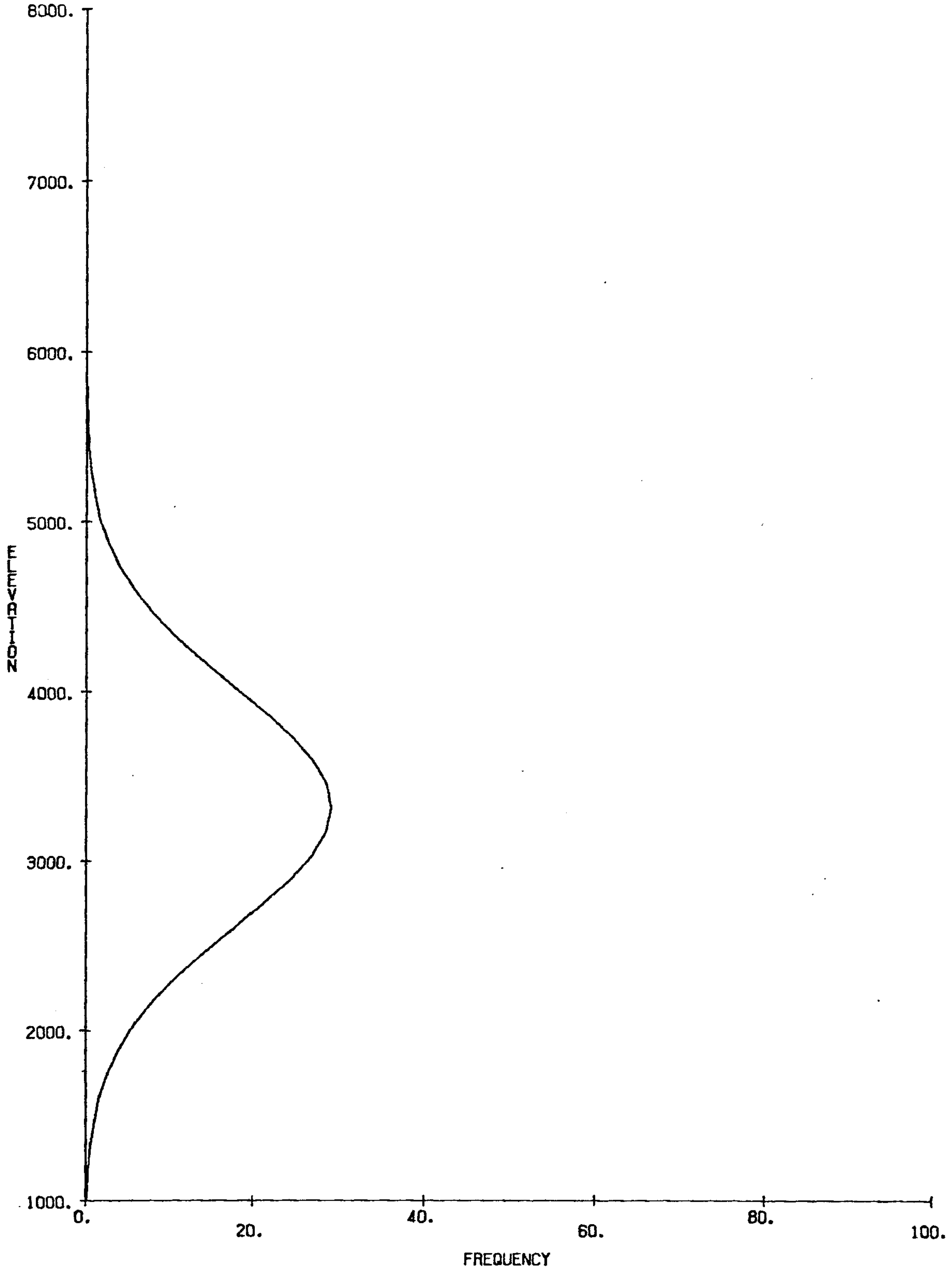
Frequency of Lodgepole Pine as a function of elevation.



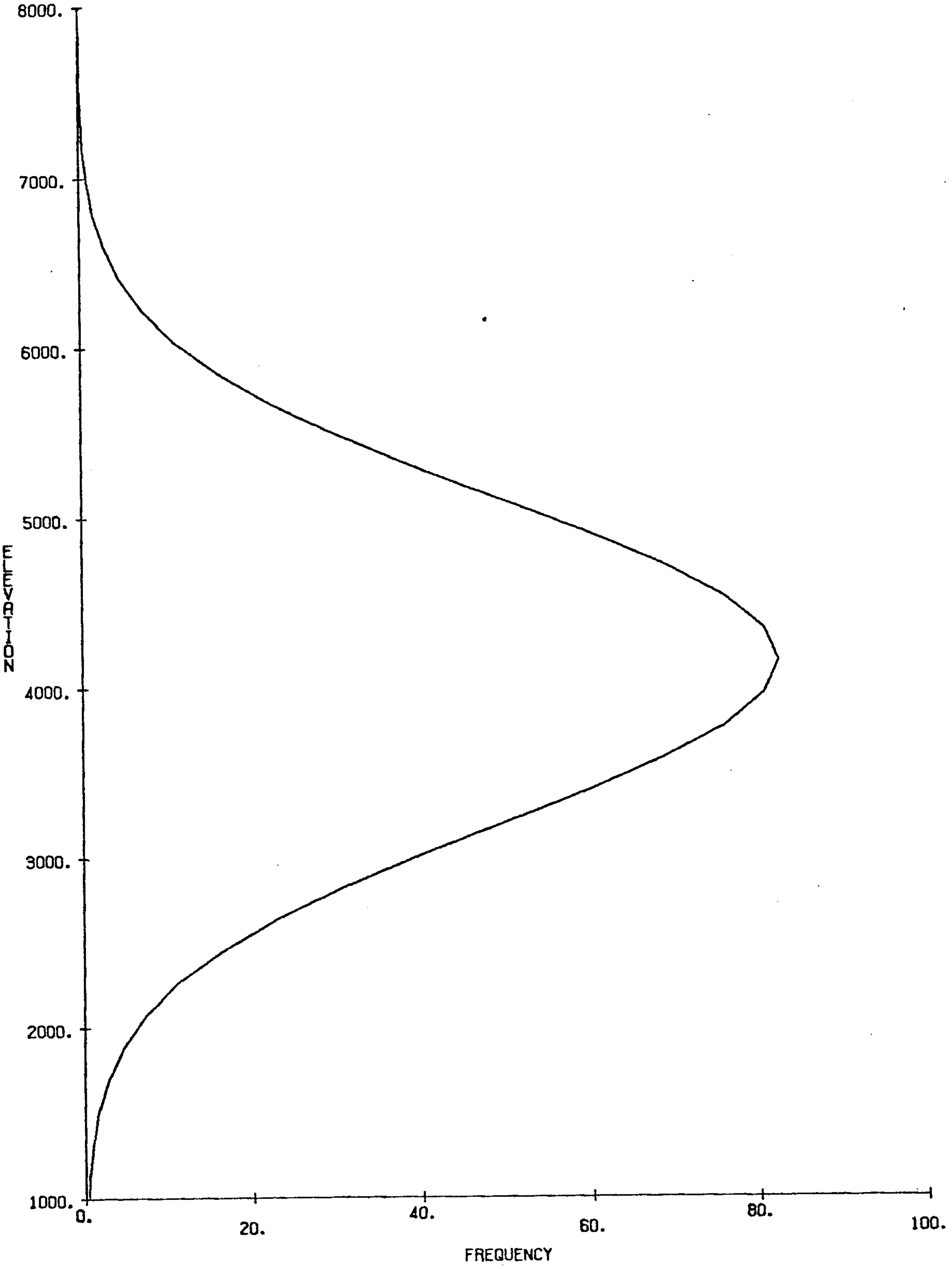
Frequency of Subalpine as a function of elevation.



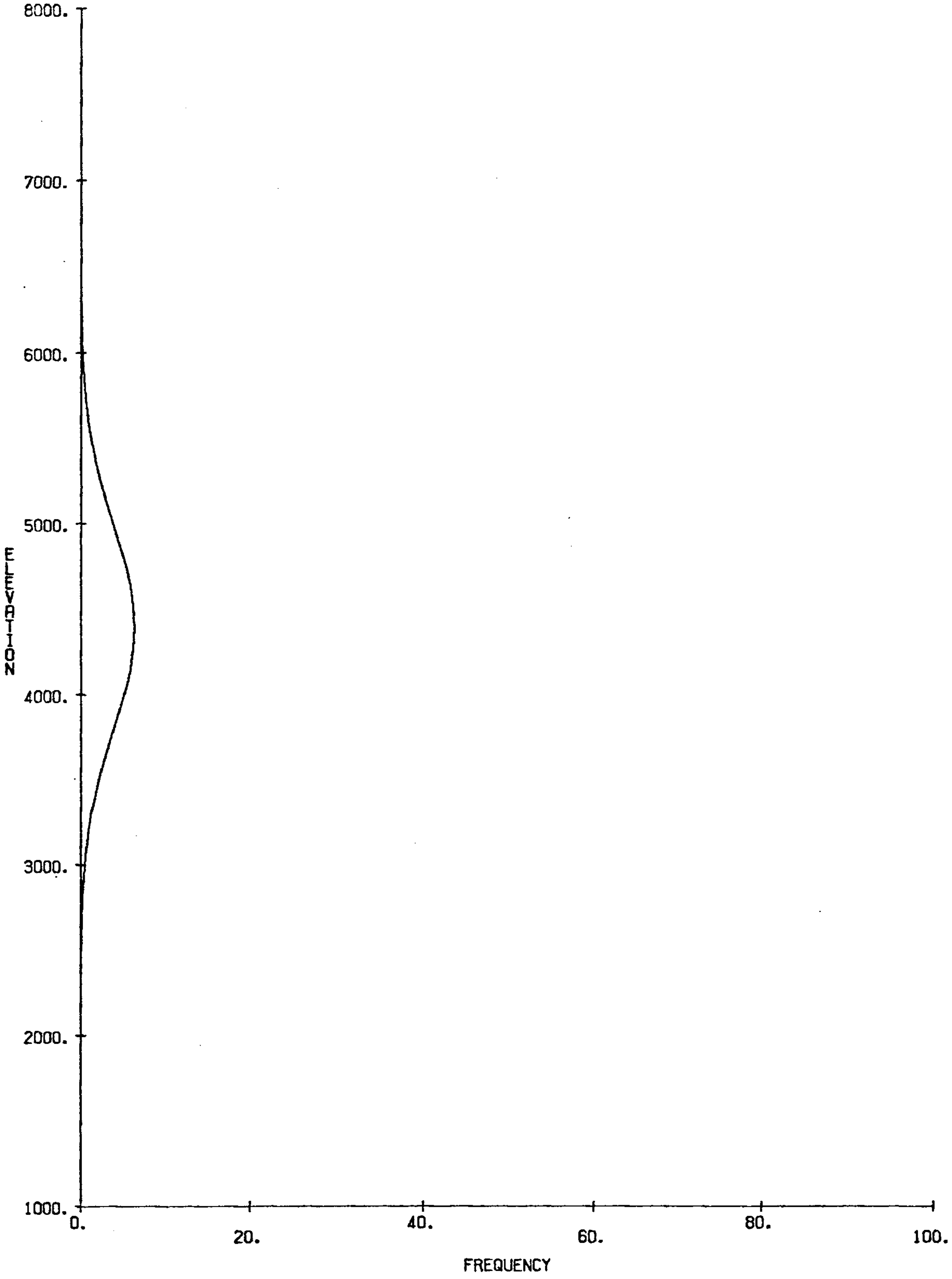
Frequency of Aspen as a function of elevation.



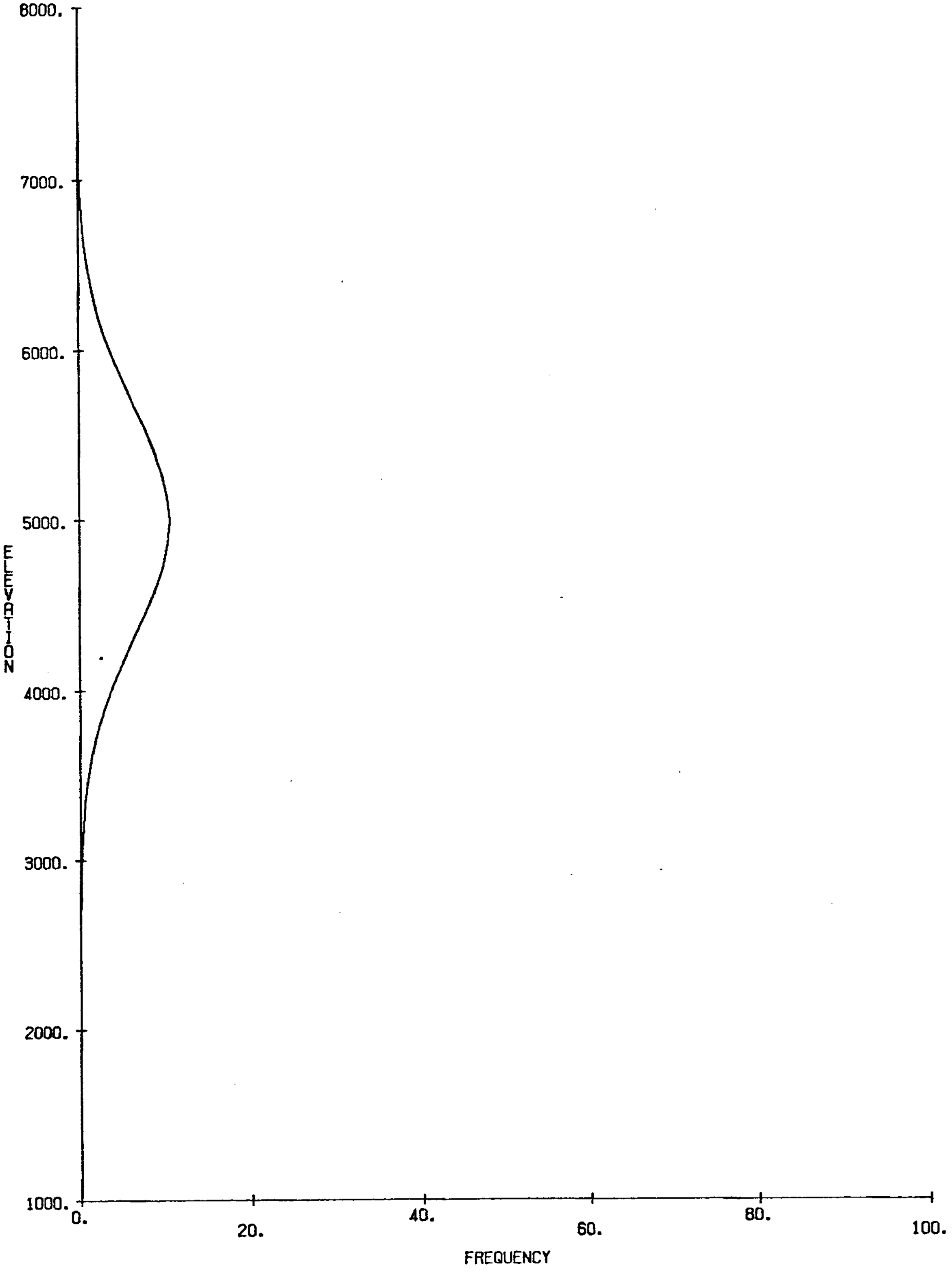
Normalized frequency distribution of Ponderosa Pine as a function of elevation.



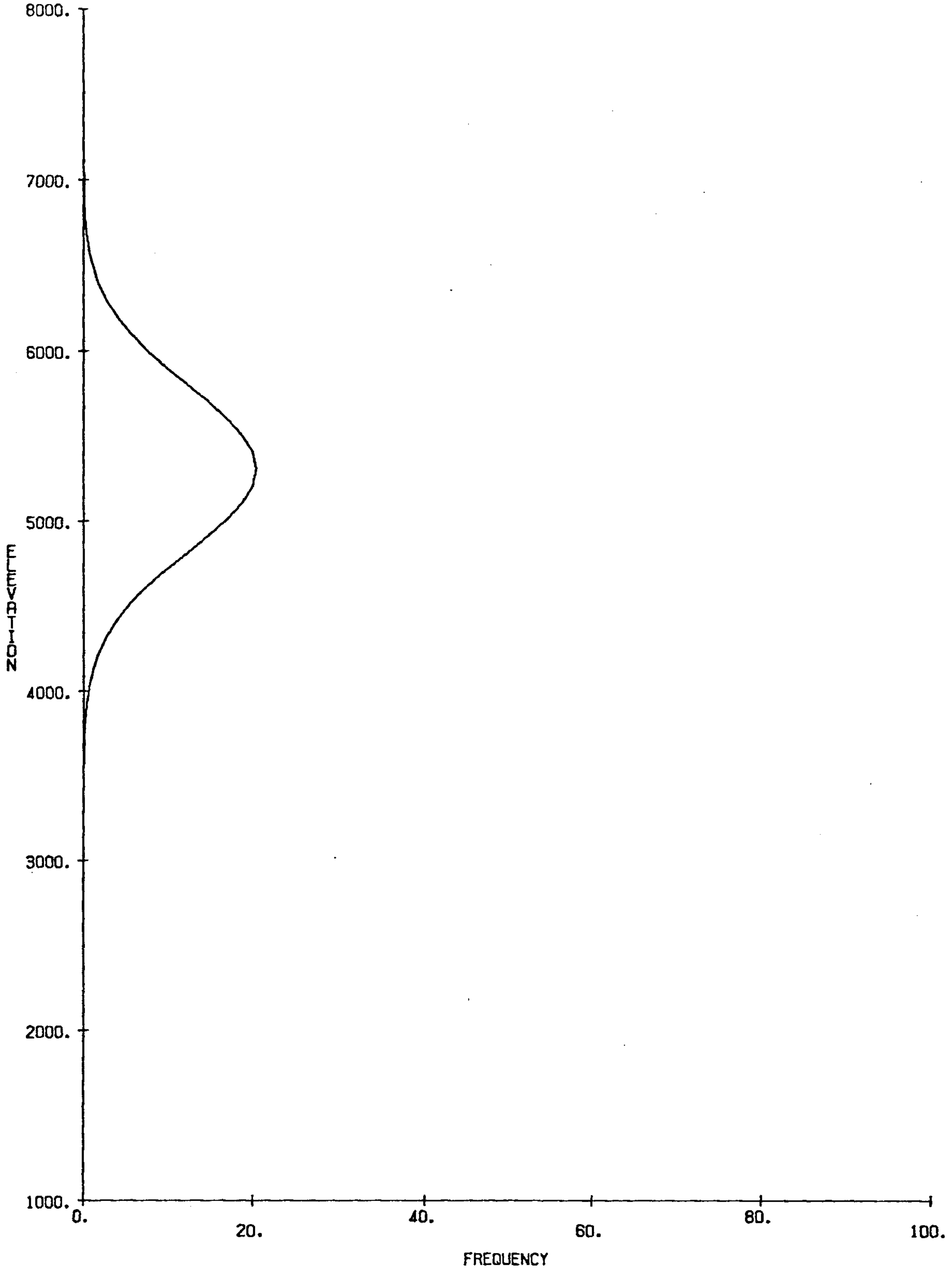
Normalized frequency distribution of Douglas Fir as a function of elevation.



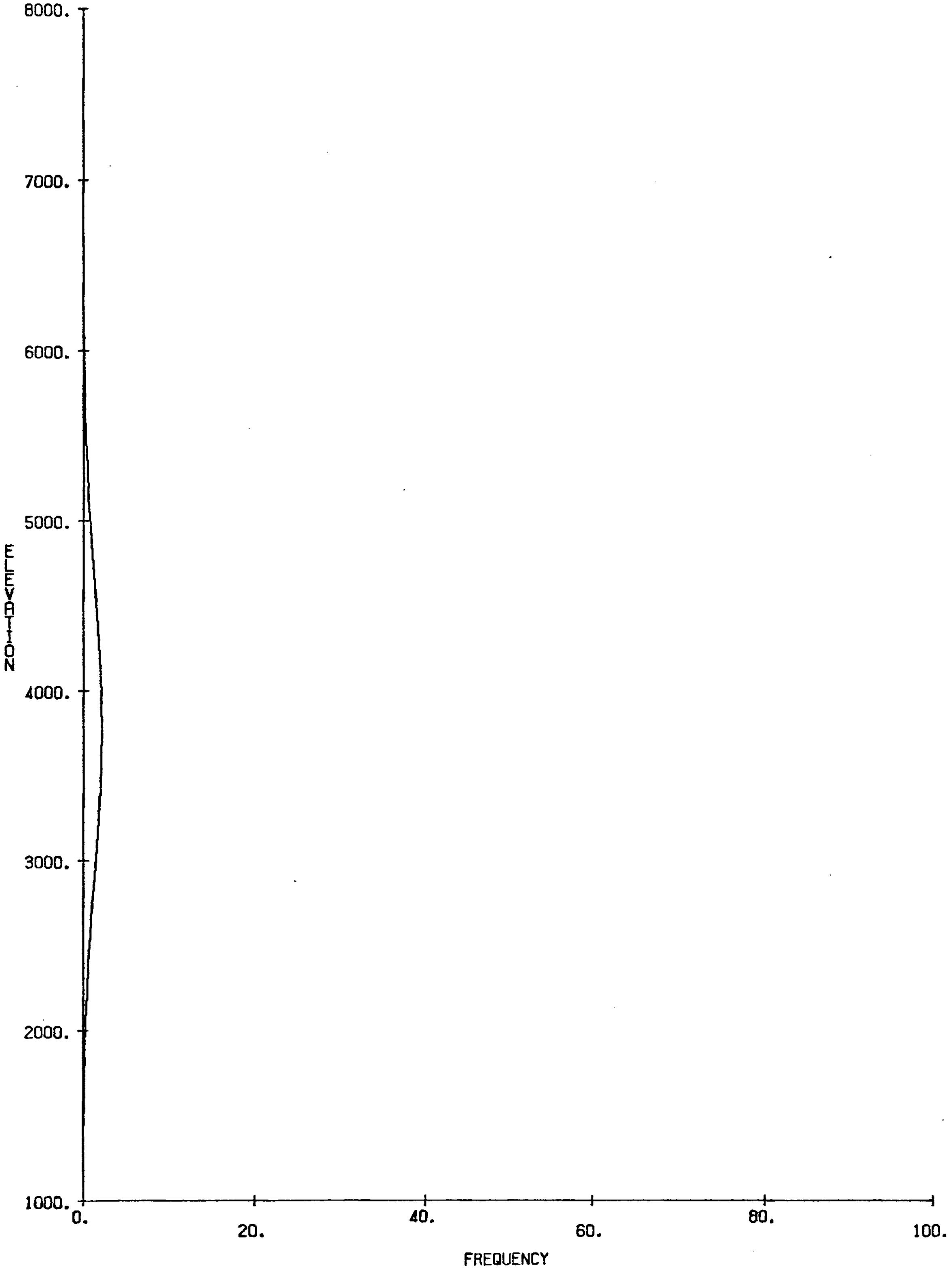
Normalized frequency distribution of Larch as a function of elevation.



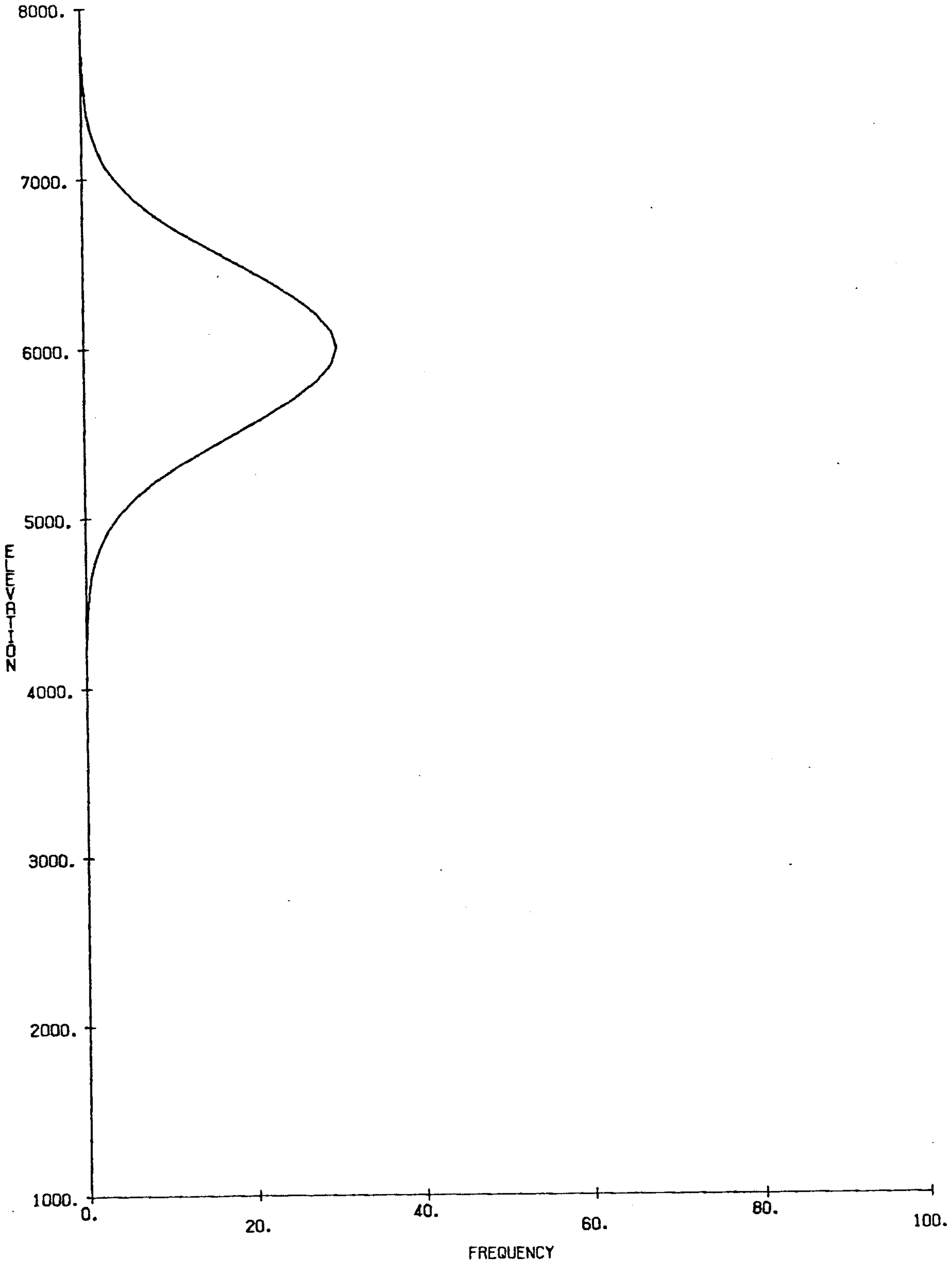
Normalized frequency distribution of Engelmann Spruce as a function of elevation.



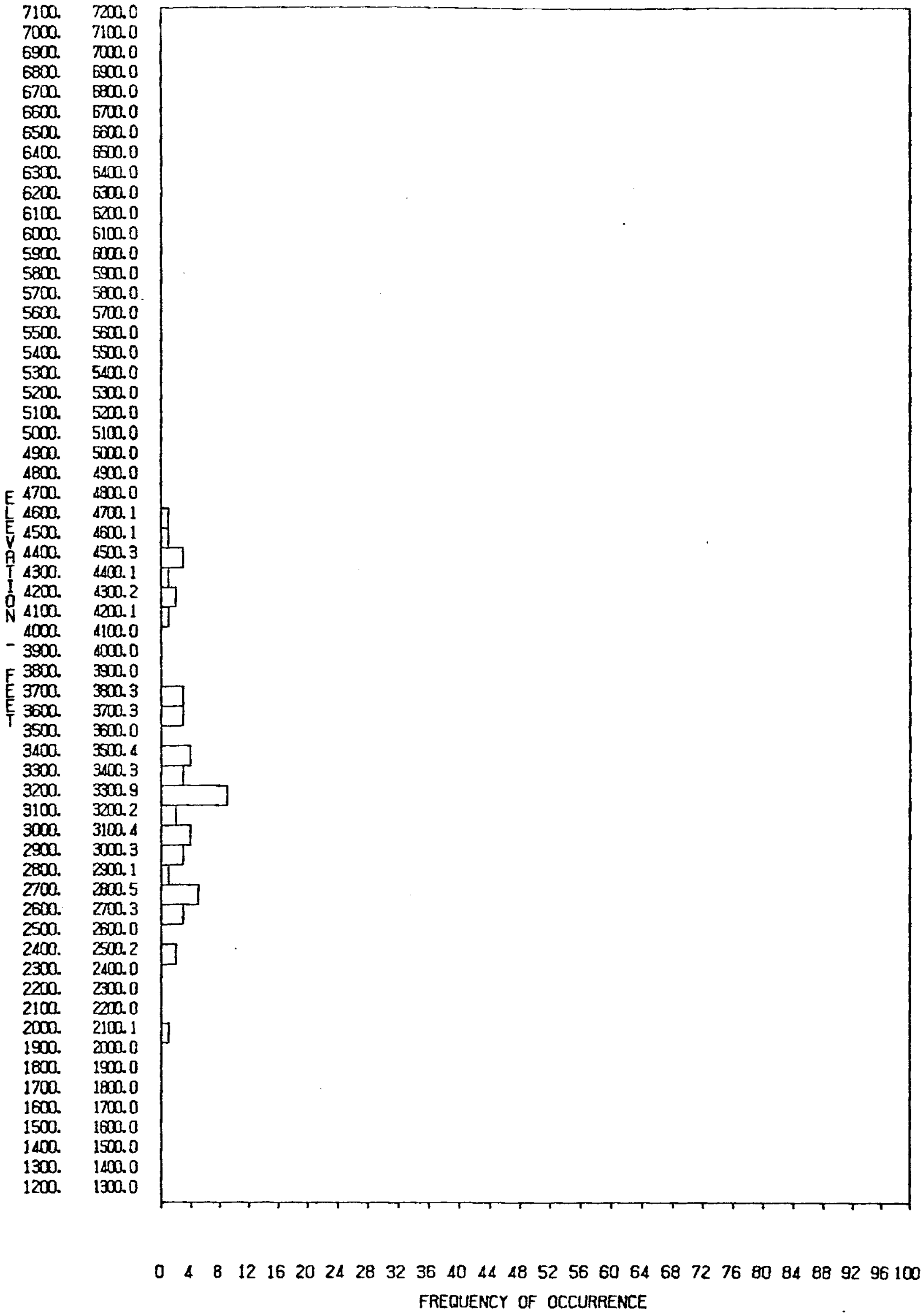
Normalized frequency distribution of Lodgepole Pine as a function of elevation.



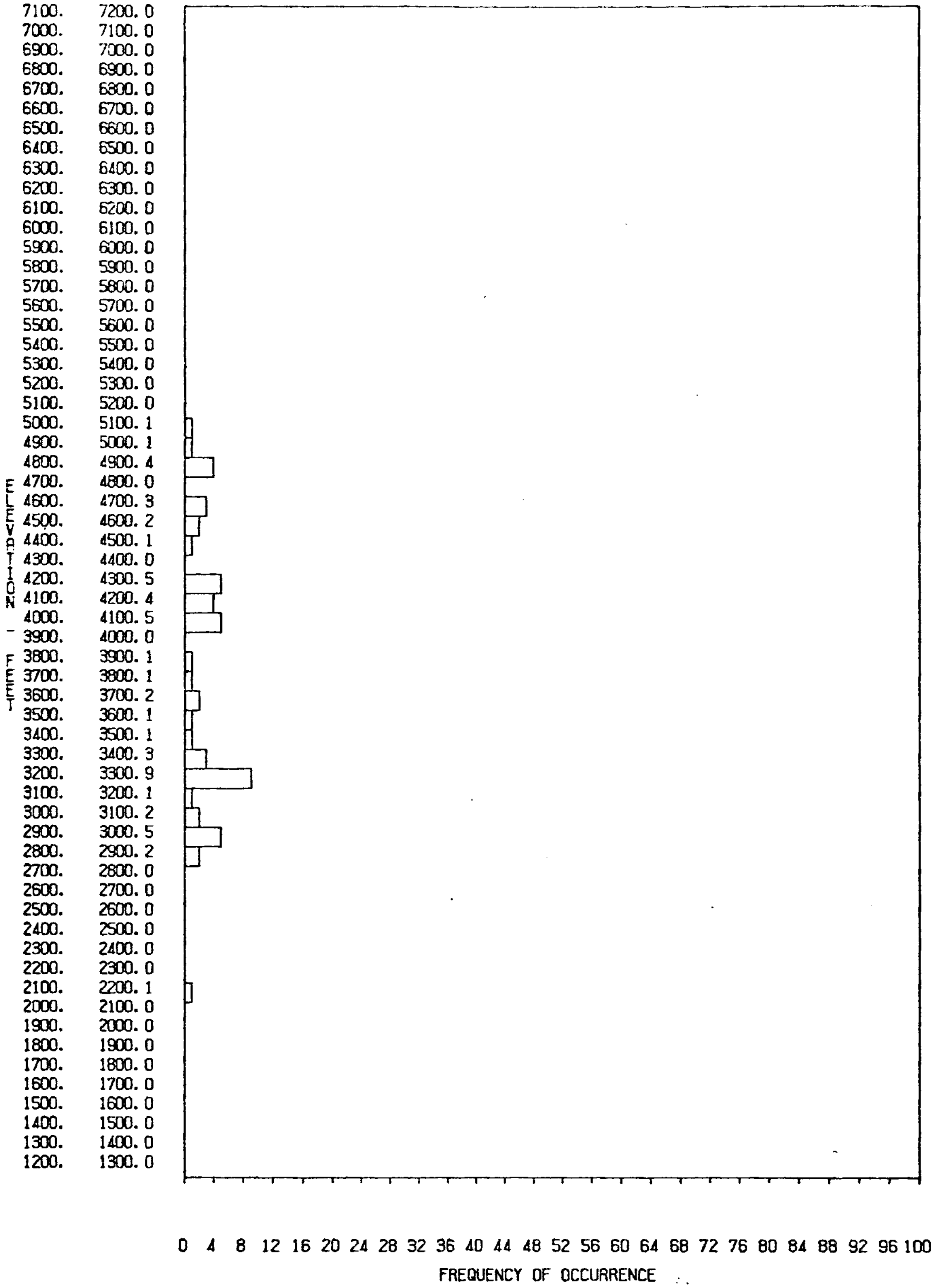
Normalized frequency distribution of Alpine Fir as a function of elevation.



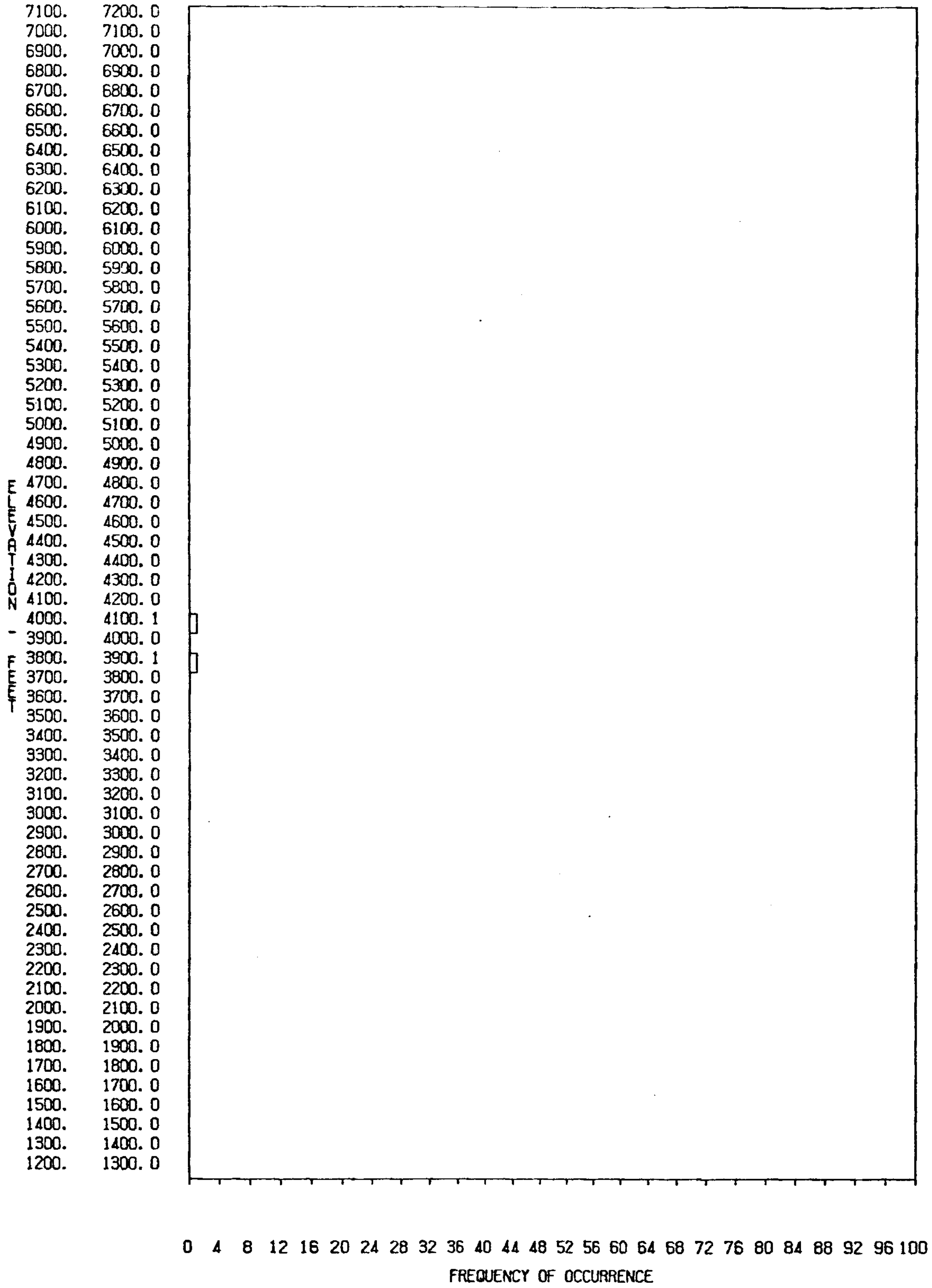
Normalized frequency distribution of Subalpine as a function of elevation.



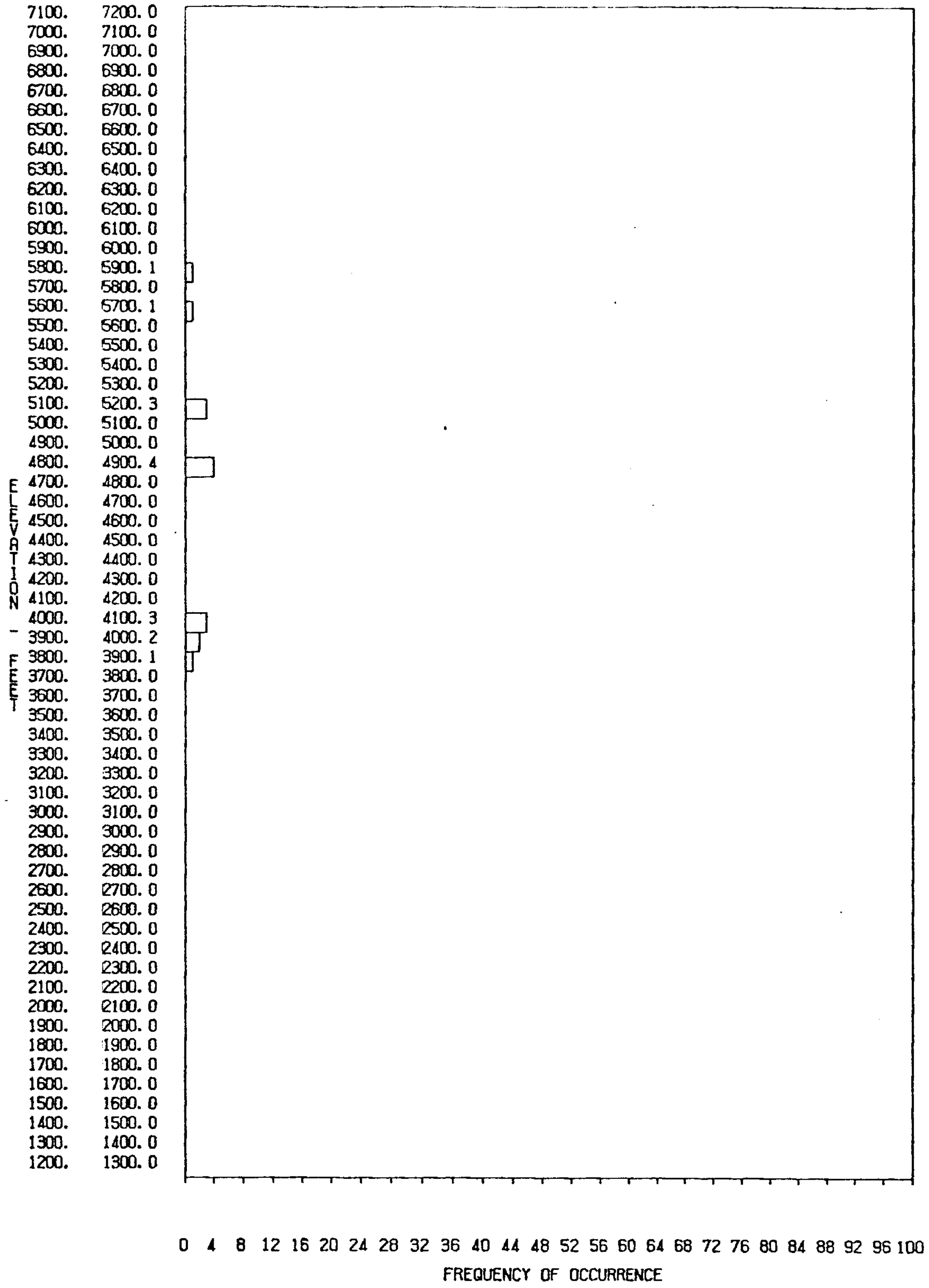
Frequency of Ponderosa Pine on a horizontal surface as a function of elevation.



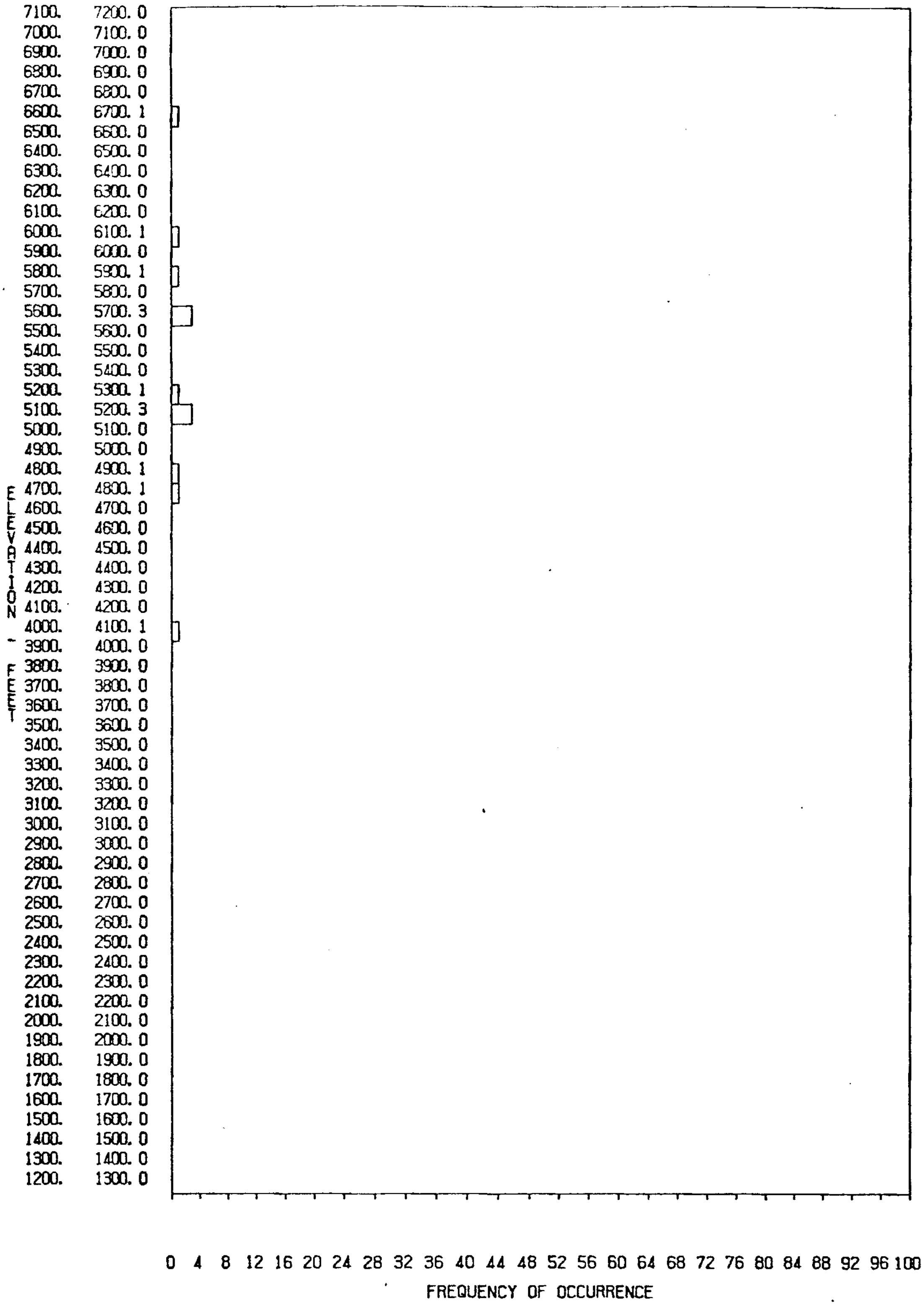
Frequency of Douglas Fir on a horizontal surface as a function of elevation.



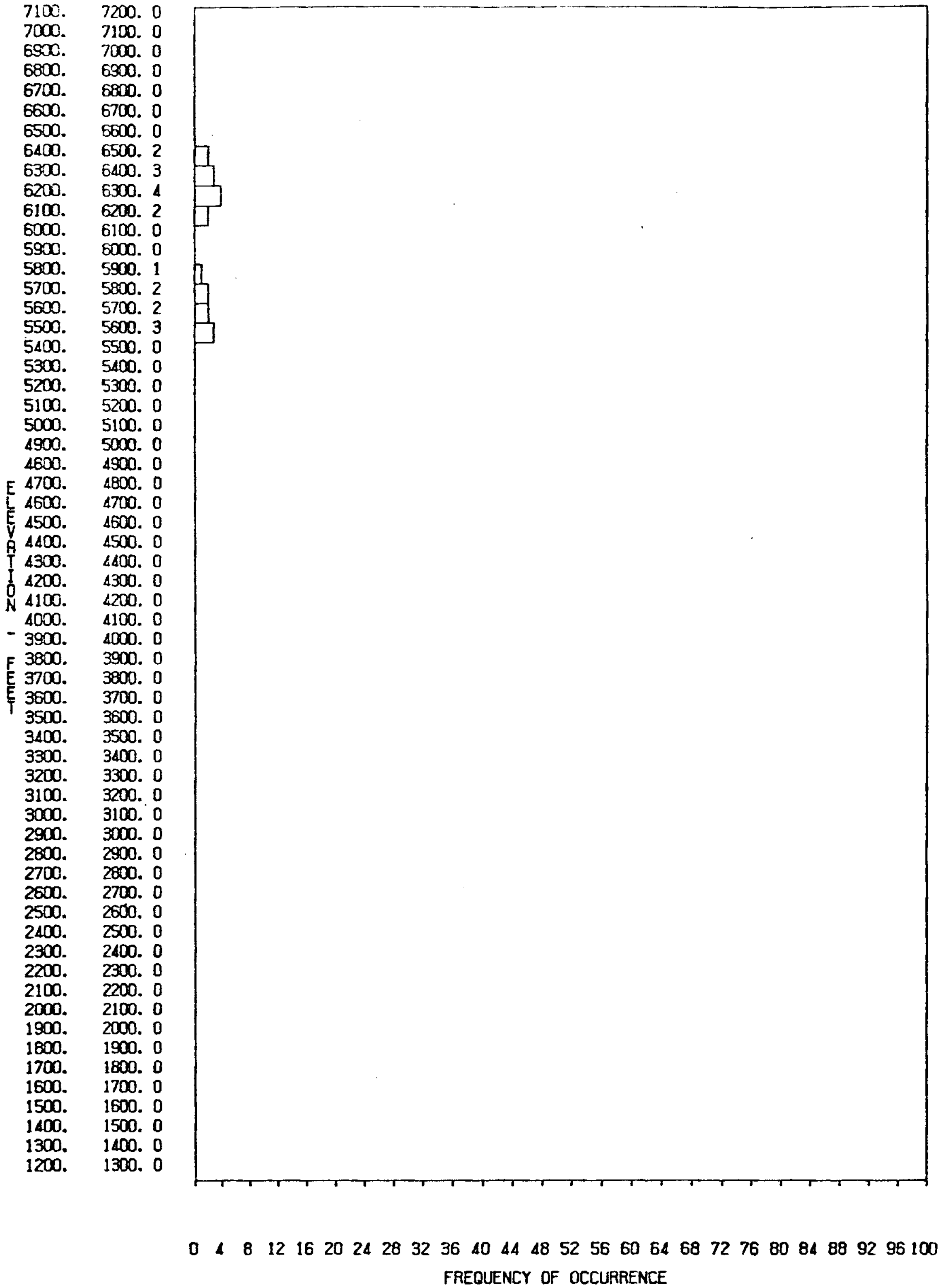
Frequency of Larch on a horizontal surface as a function of elevation.



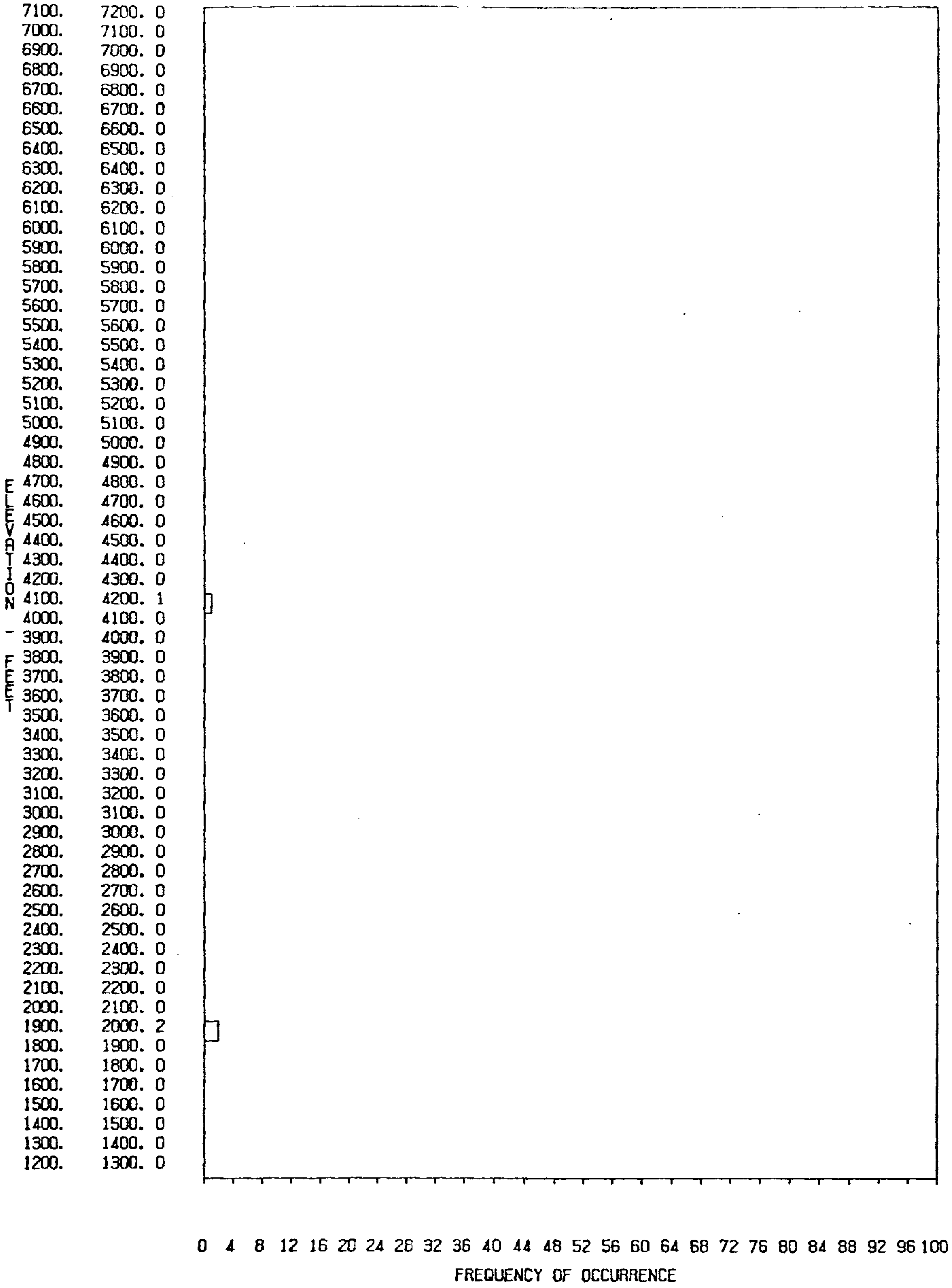
Frequency of Spruce/Fir on a horizontal surface as a function of elevation.



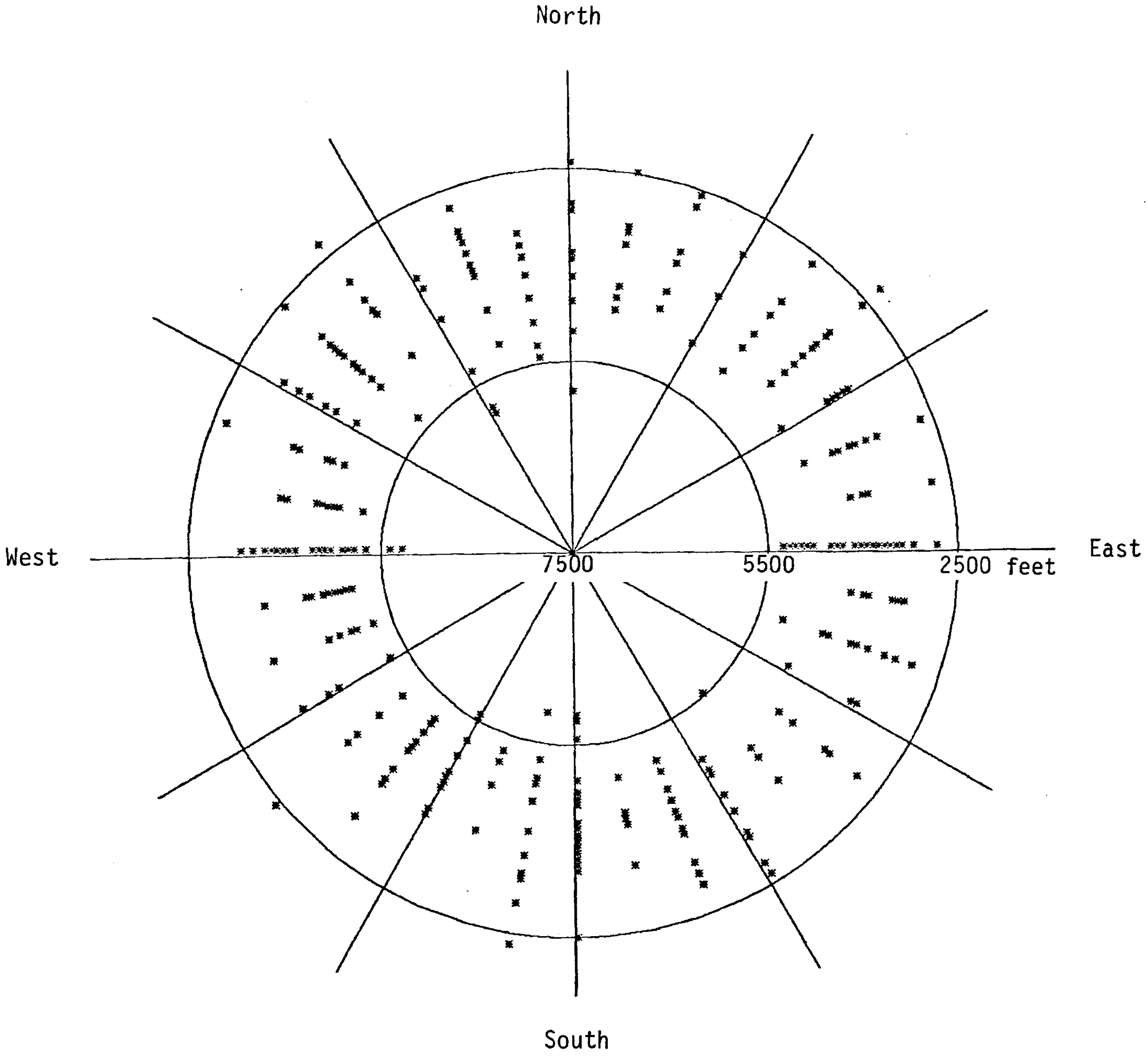
Frequency of Lodgepole Pine on a horizontal surface as a function of elevation.



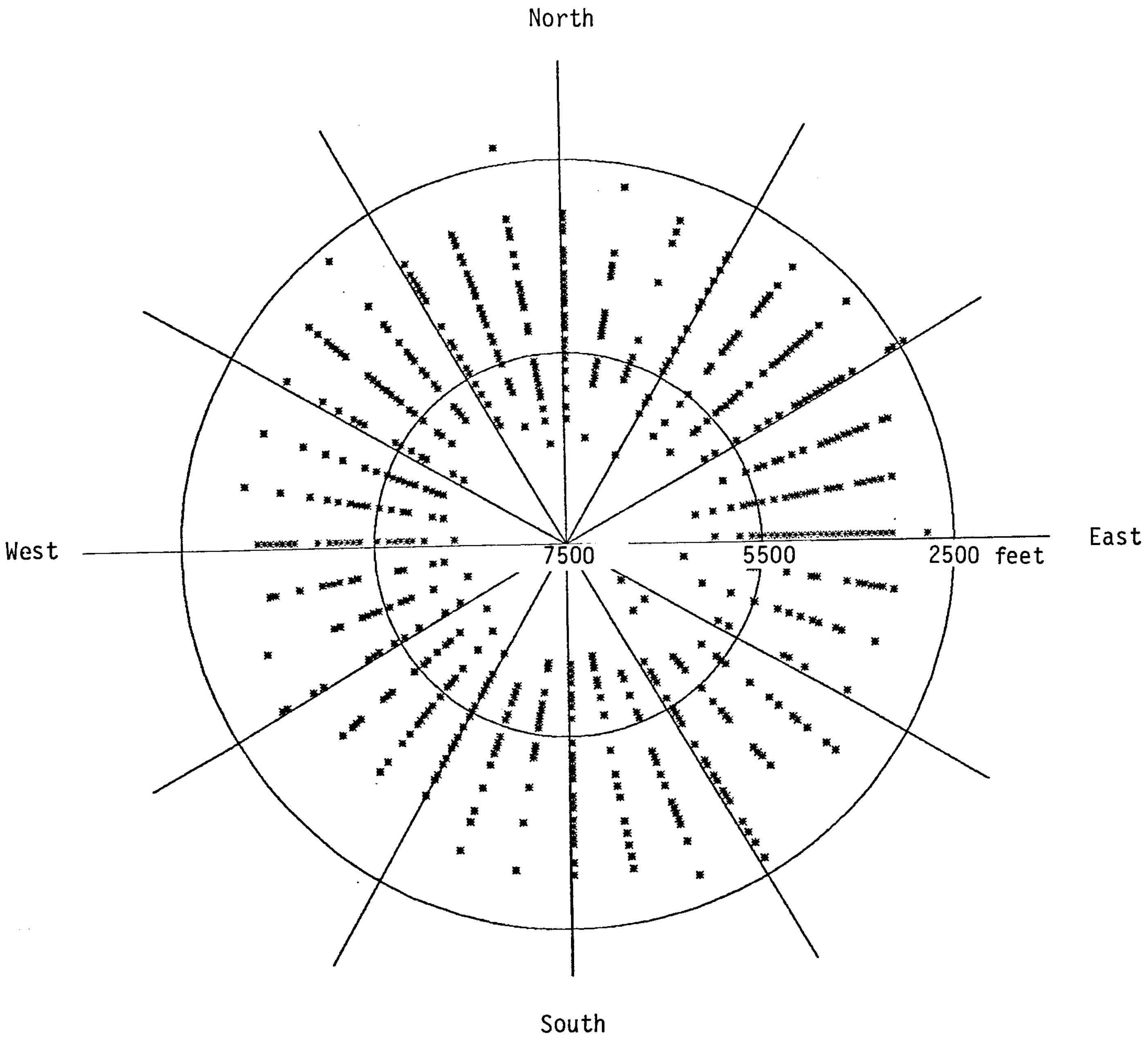
Frequency of Subalpine on a horizontal surface as a function of elevation.



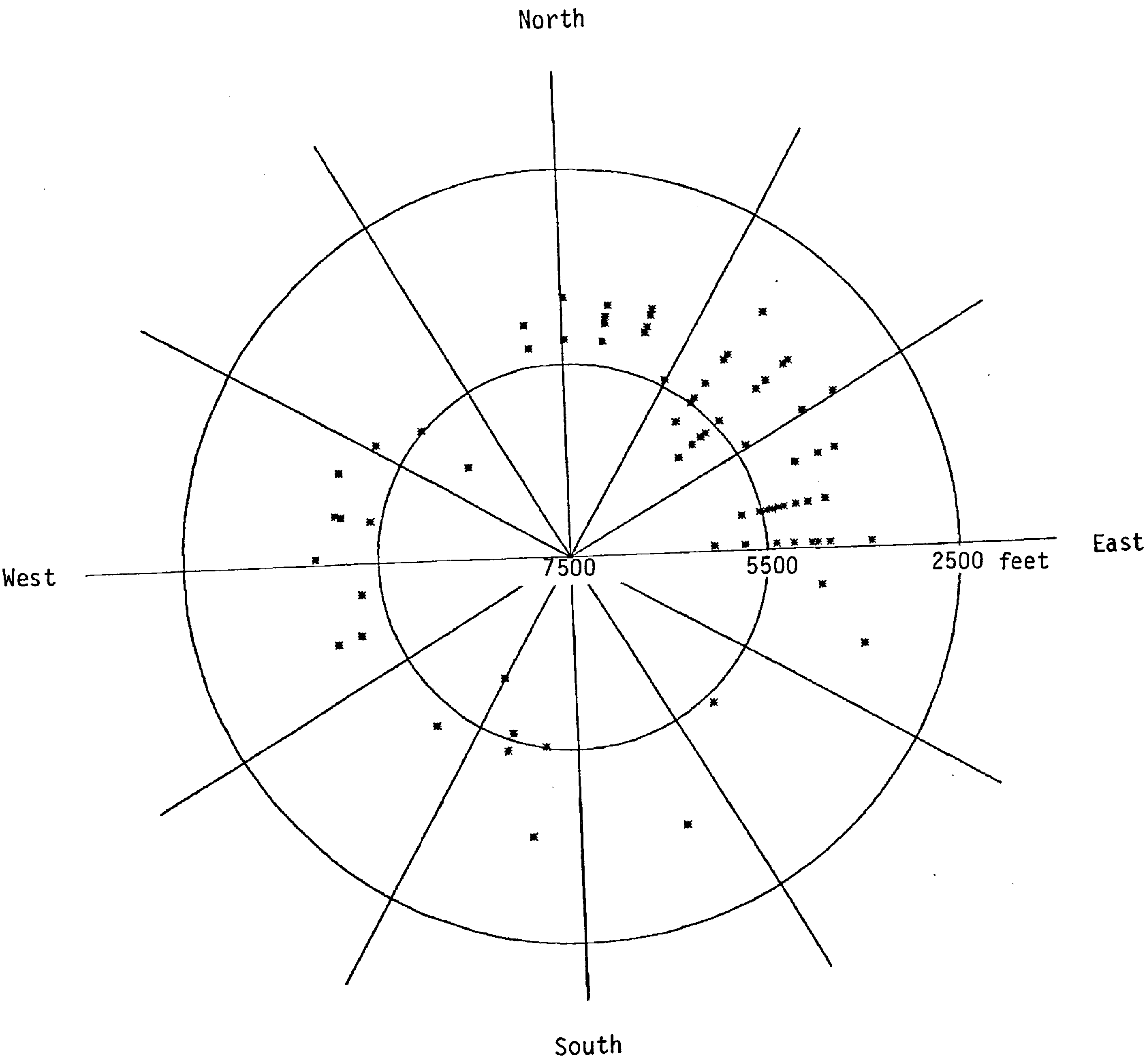
Frequency of Aspen on a horizontal surface as a function of elevation.



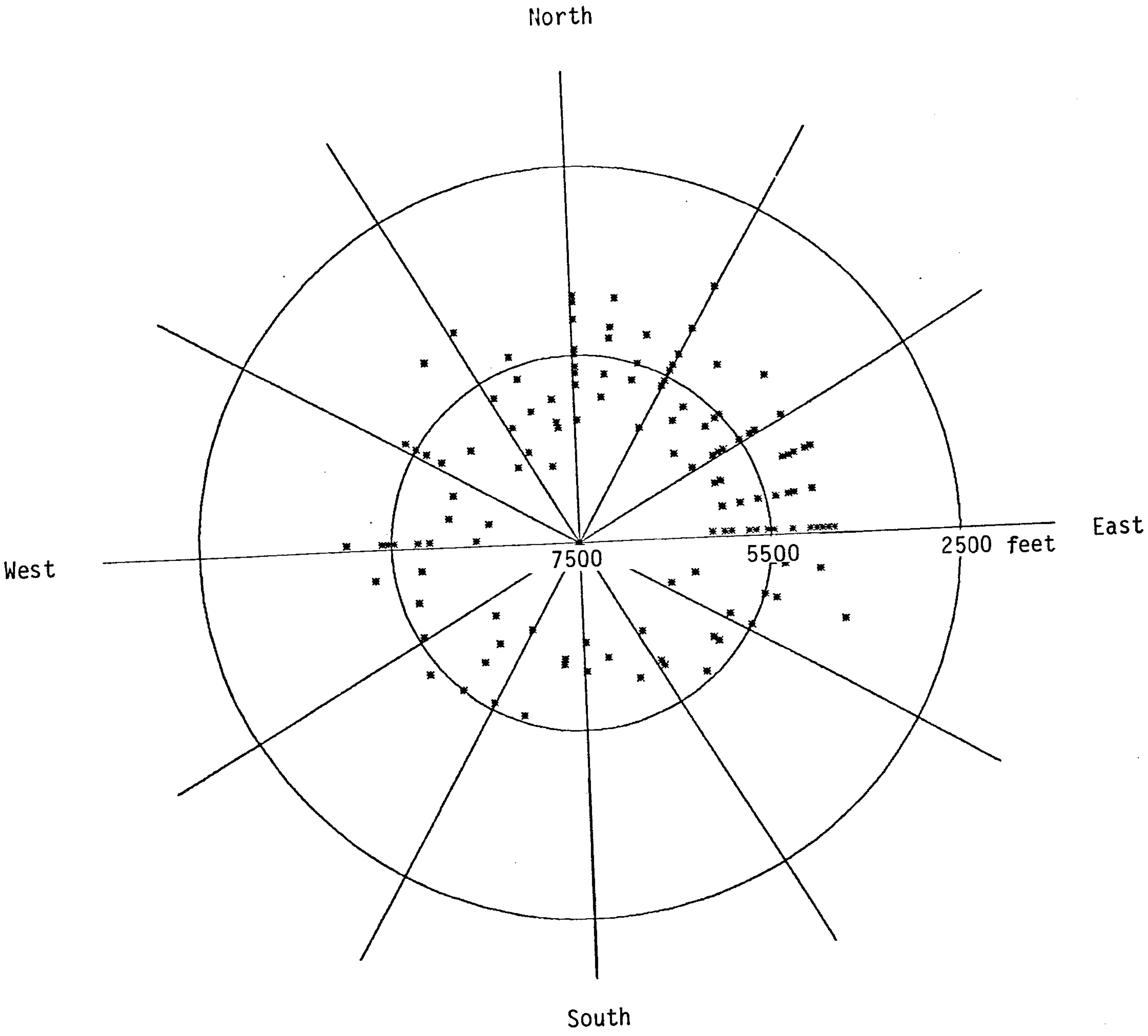
Distribution of Ponderosa Pine as a function of elevation and aspect.



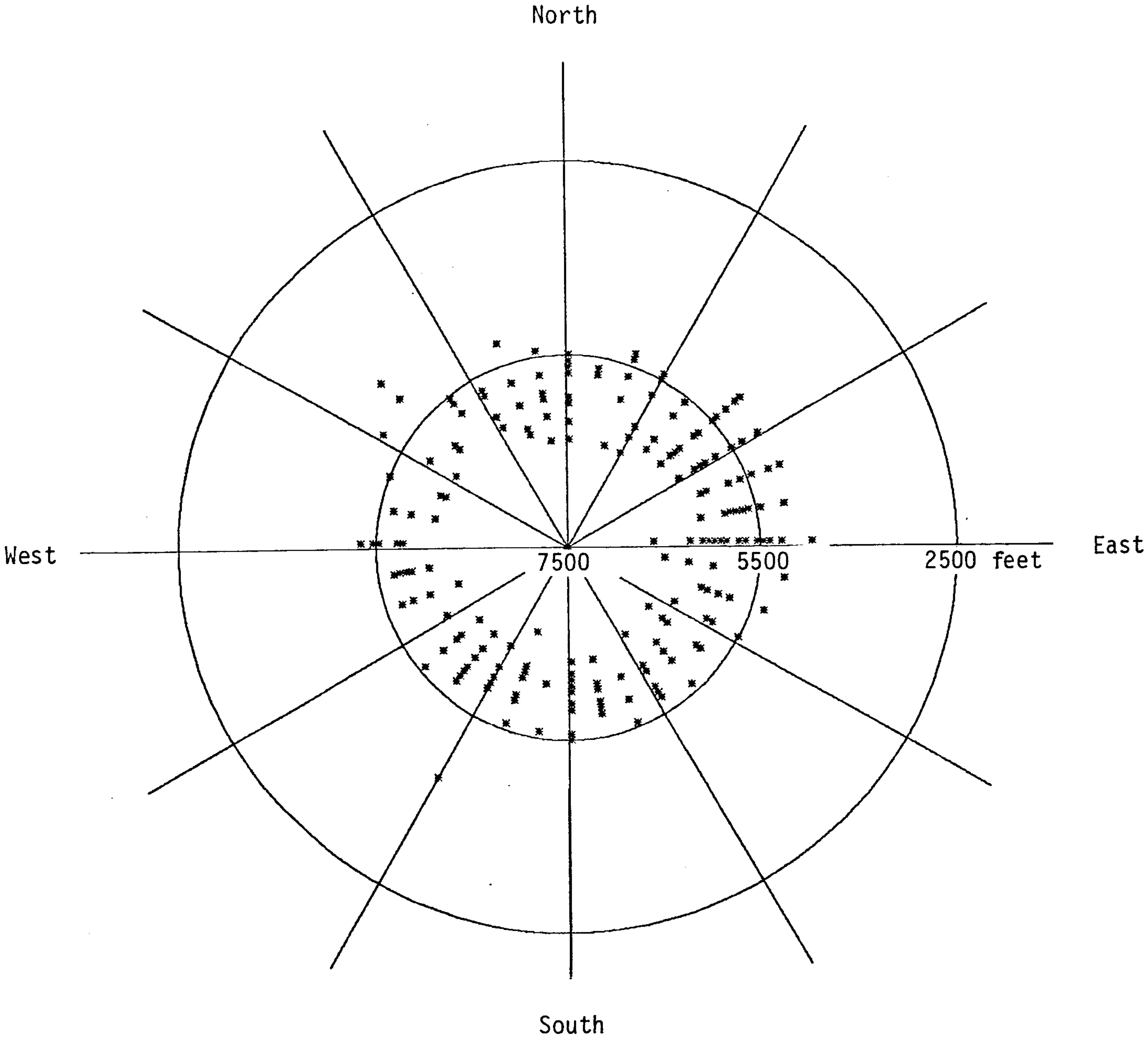
Distribution of Douglas Fir as a function of elevation and aspect.



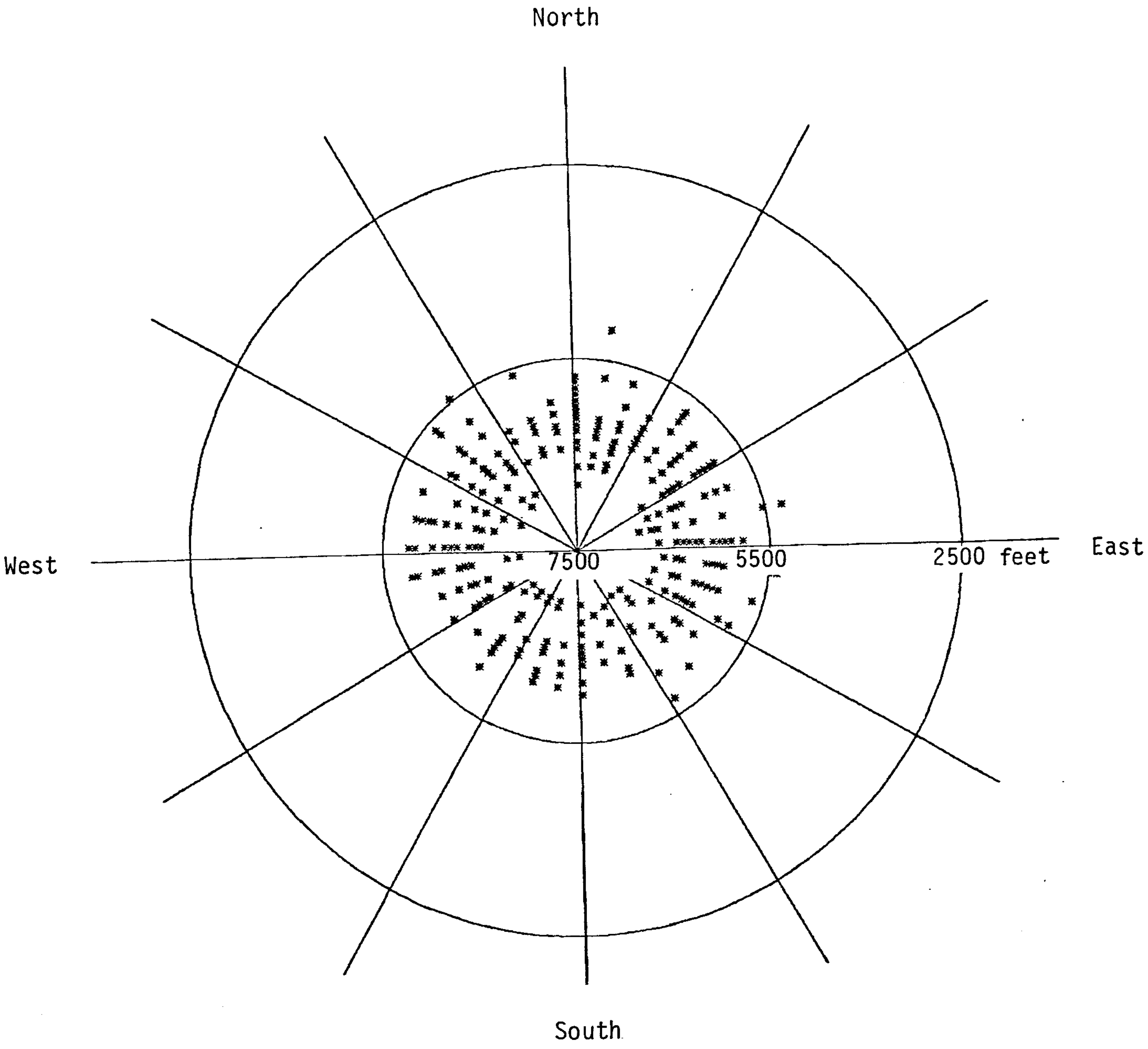
Distribution of Larch as a function of elevation and aspect.



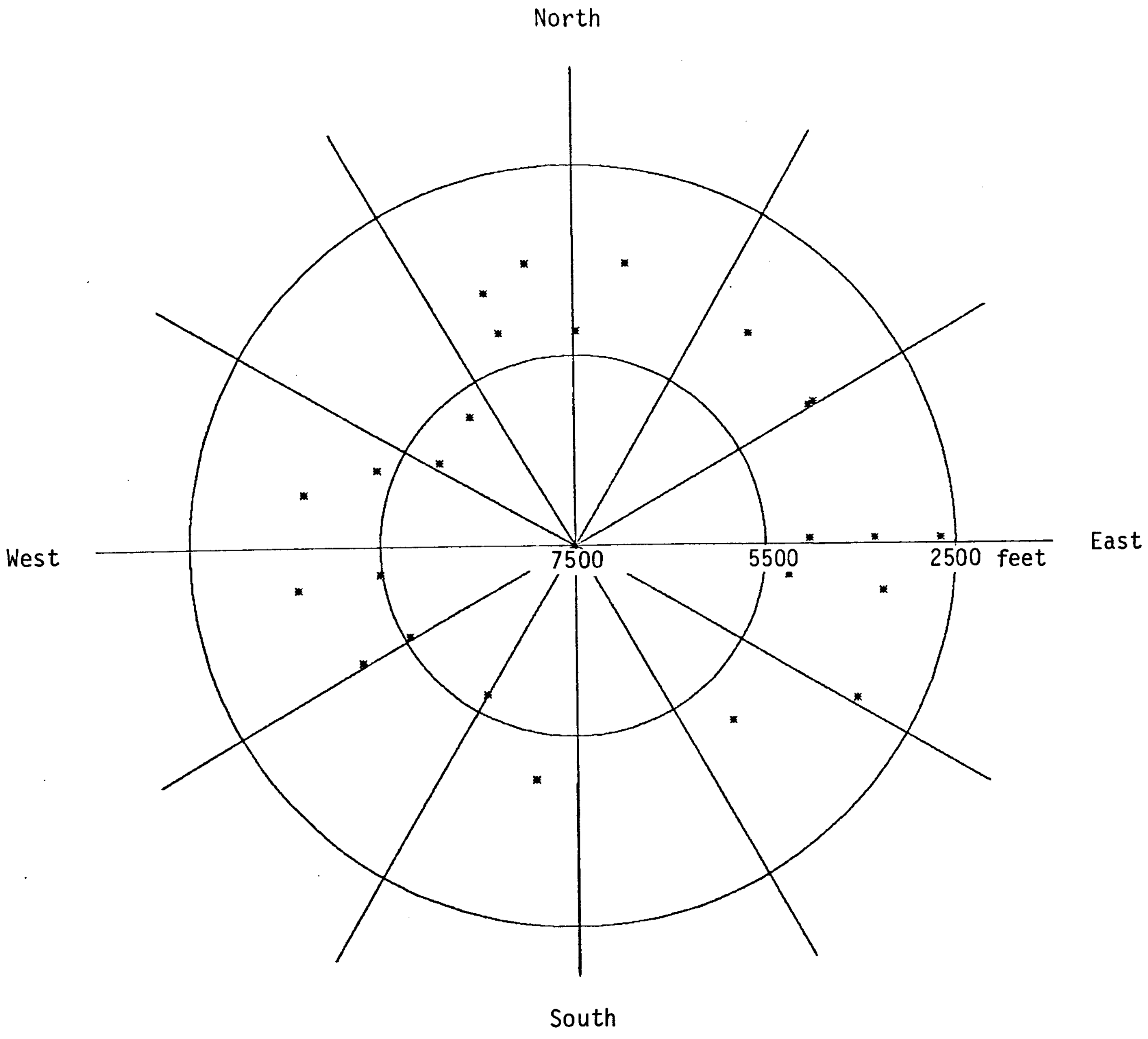
Distribution of Spruce/Fir as a function of elevation and aspect.



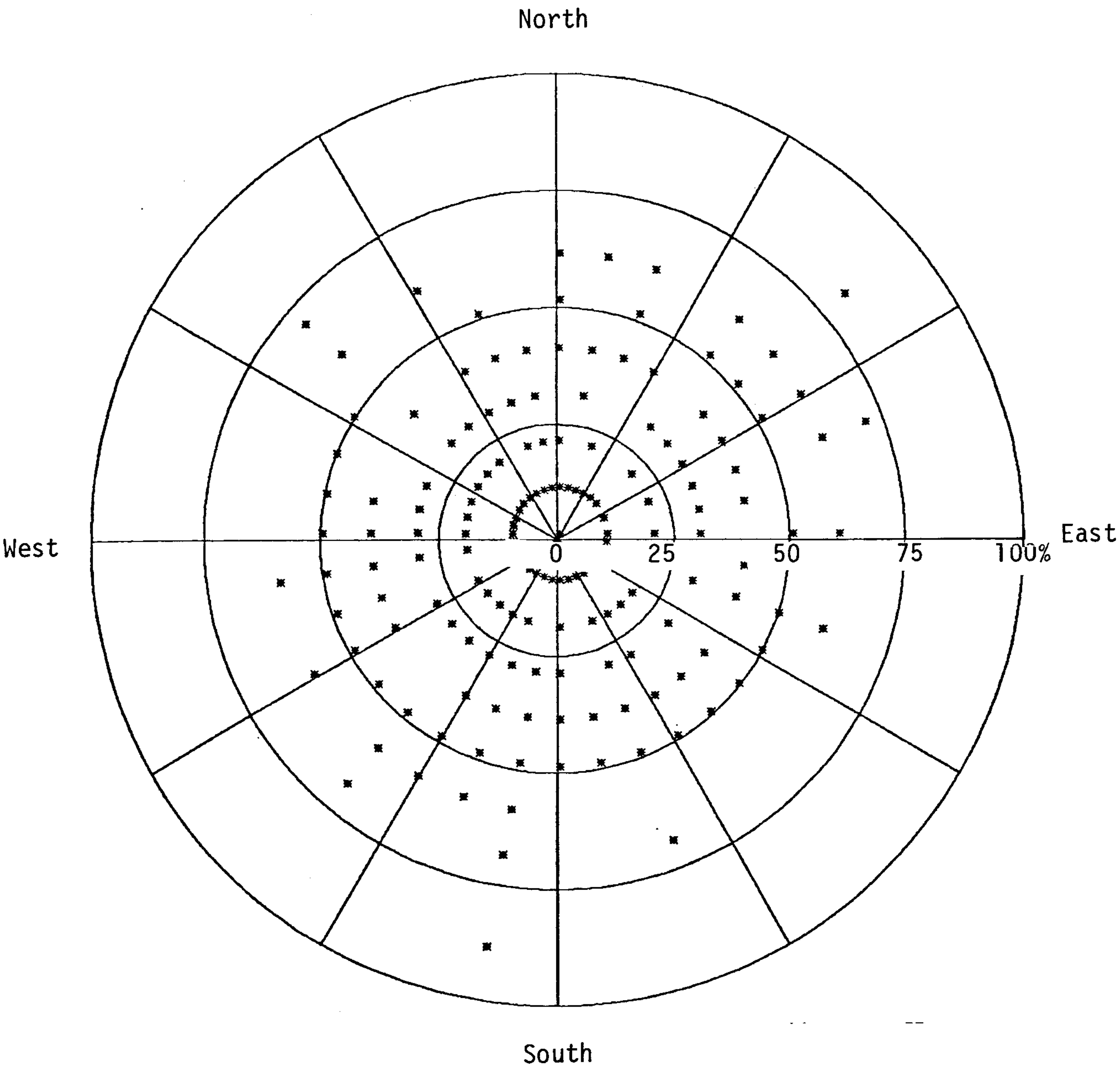
Distribution of Lodgepole Pine as a function of elevation and aspect.



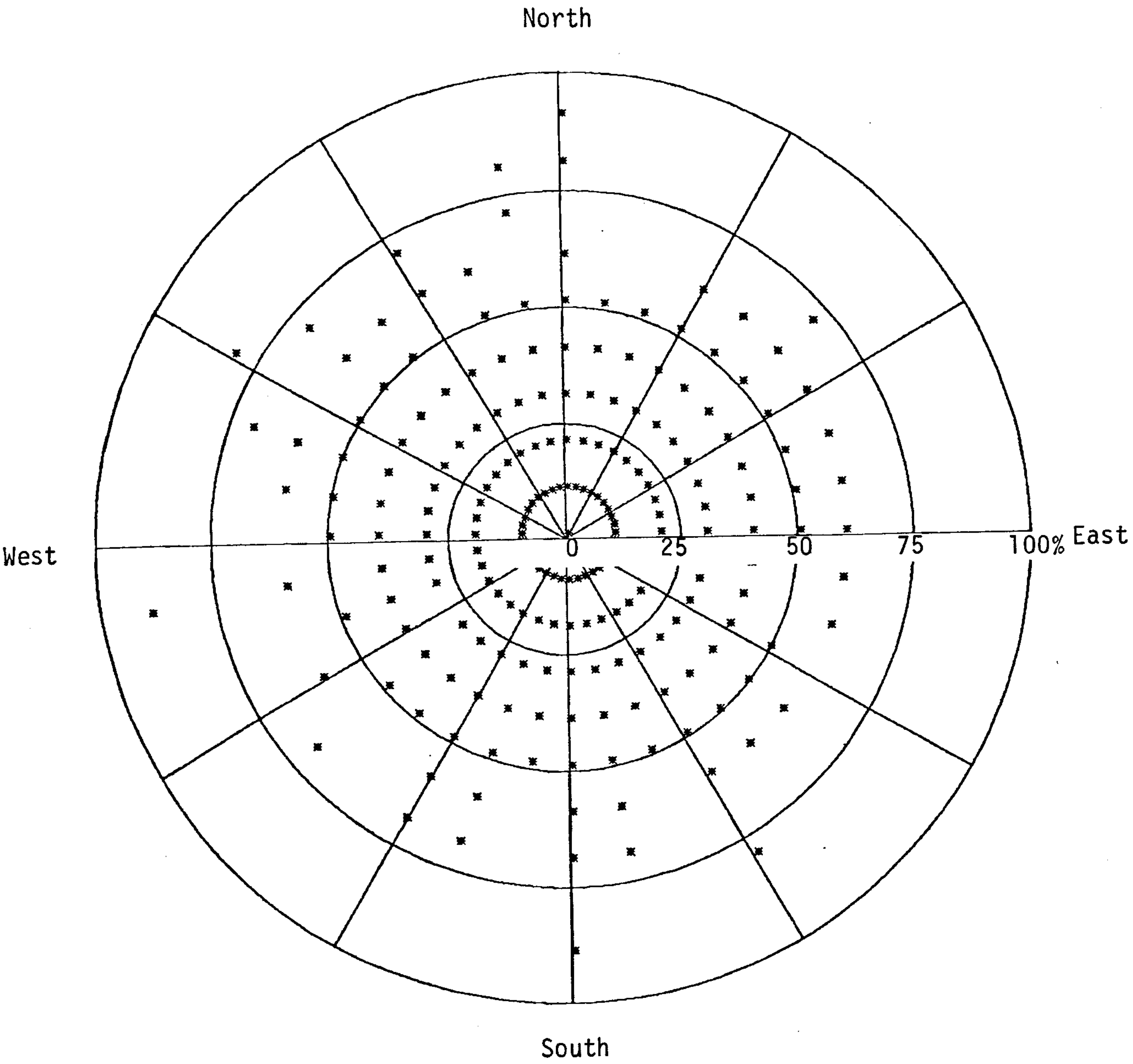
Distribution of Subalpine as a function of elevation and aspect.



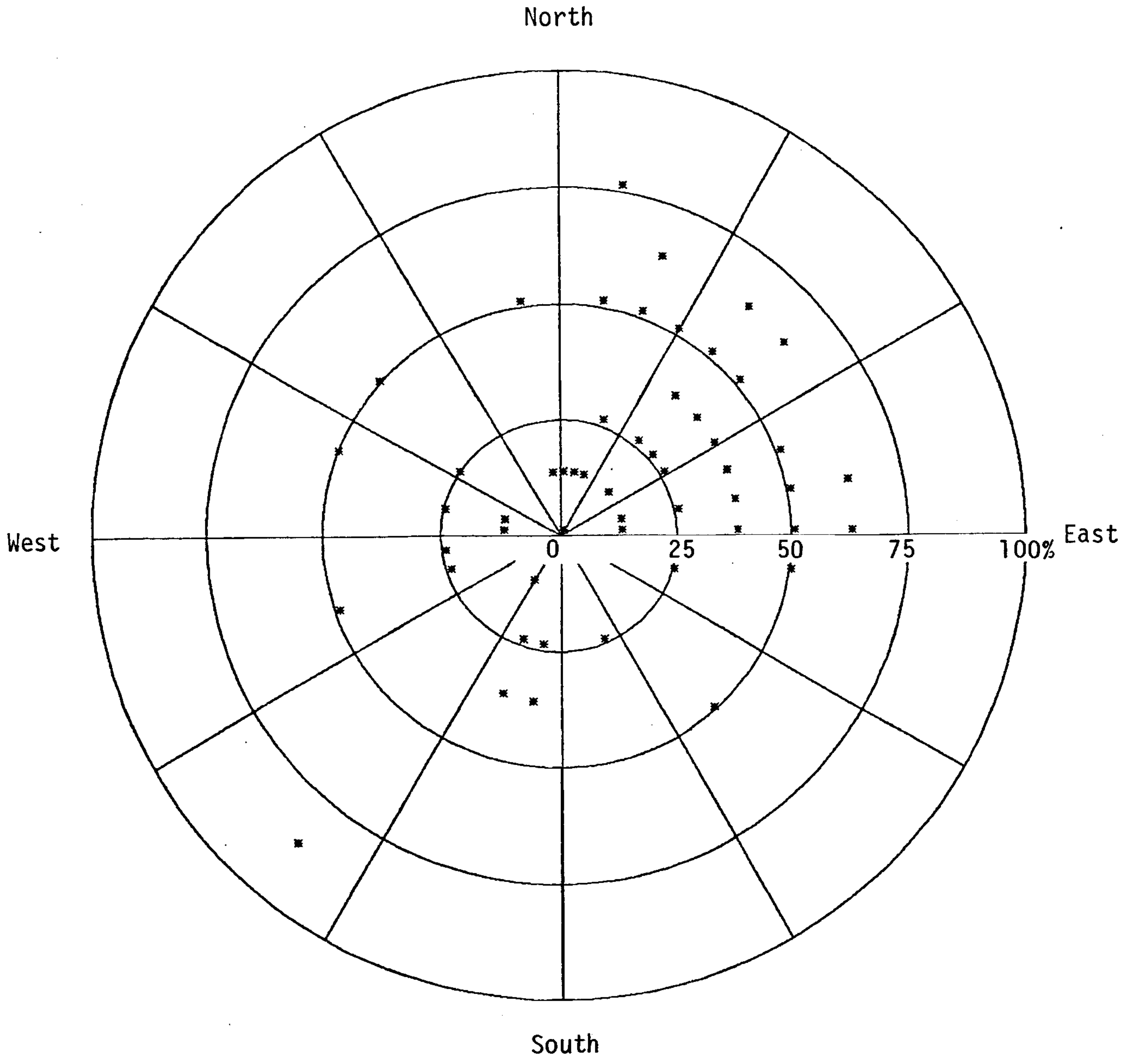
Distribution of Aspen as a function of elevation and aspect.



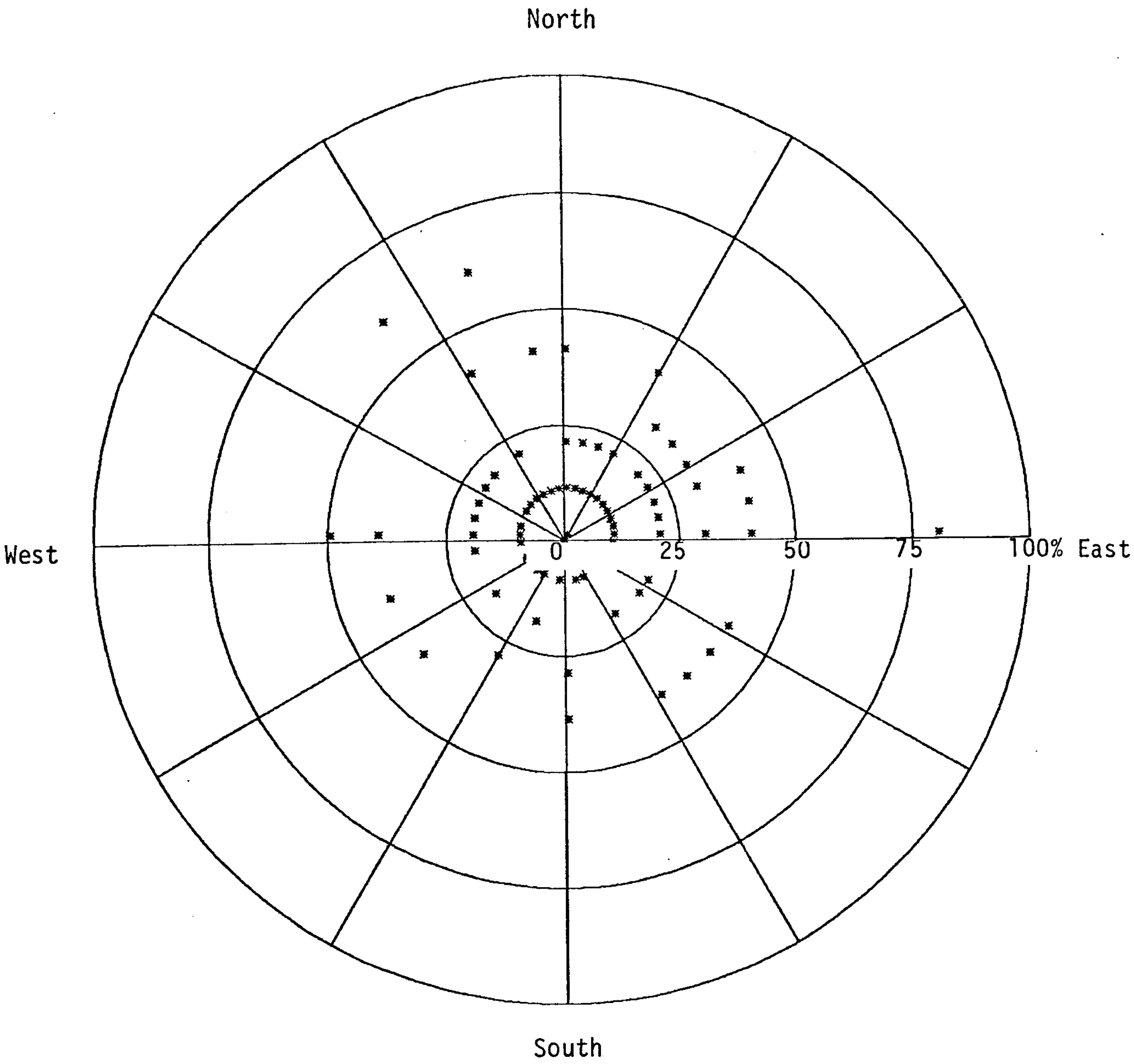
Distribution of Ponderosa Pine as a function of slope and aspect.



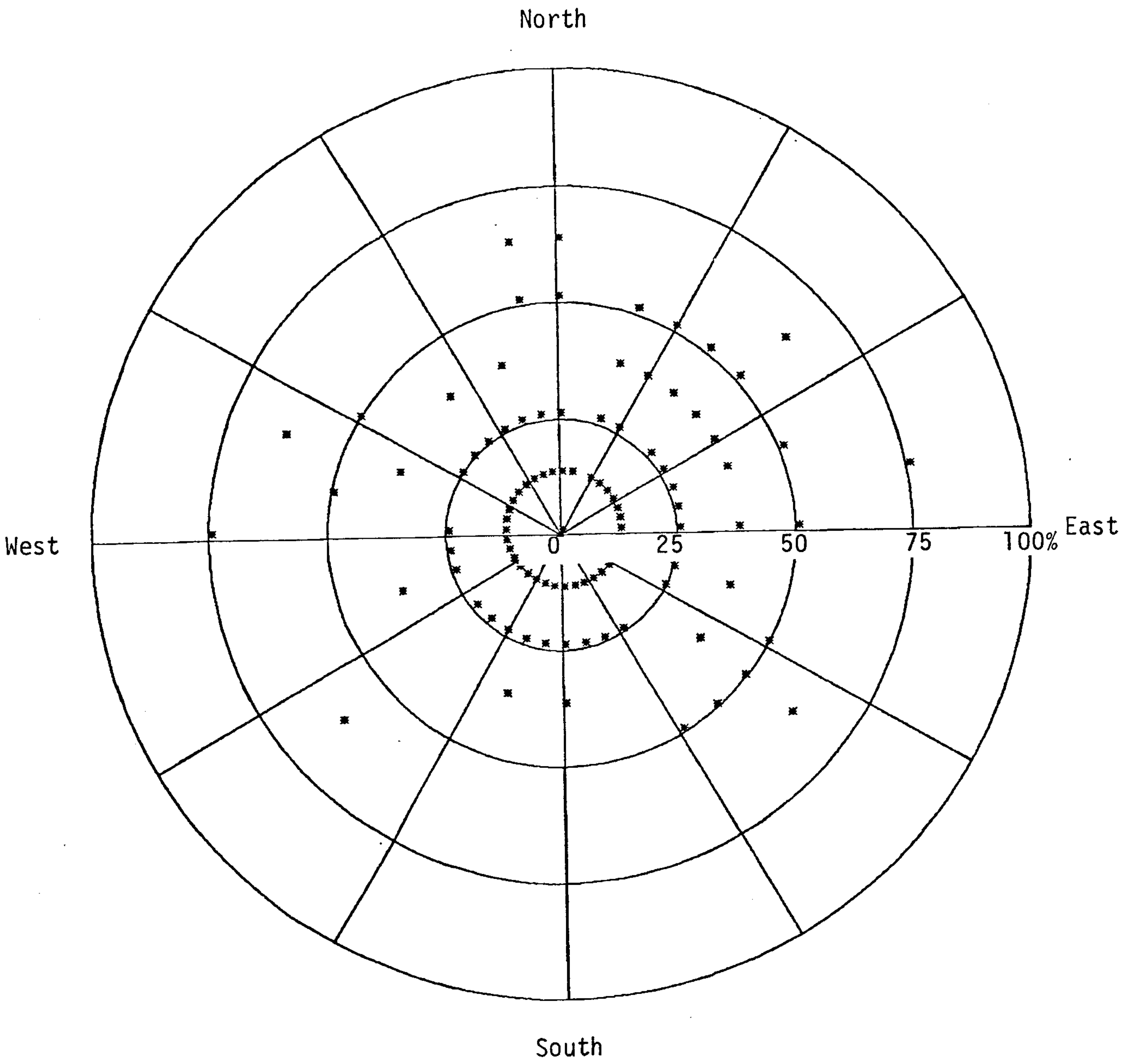
Distribution of Douglas Fir as a function of slope and aspect.



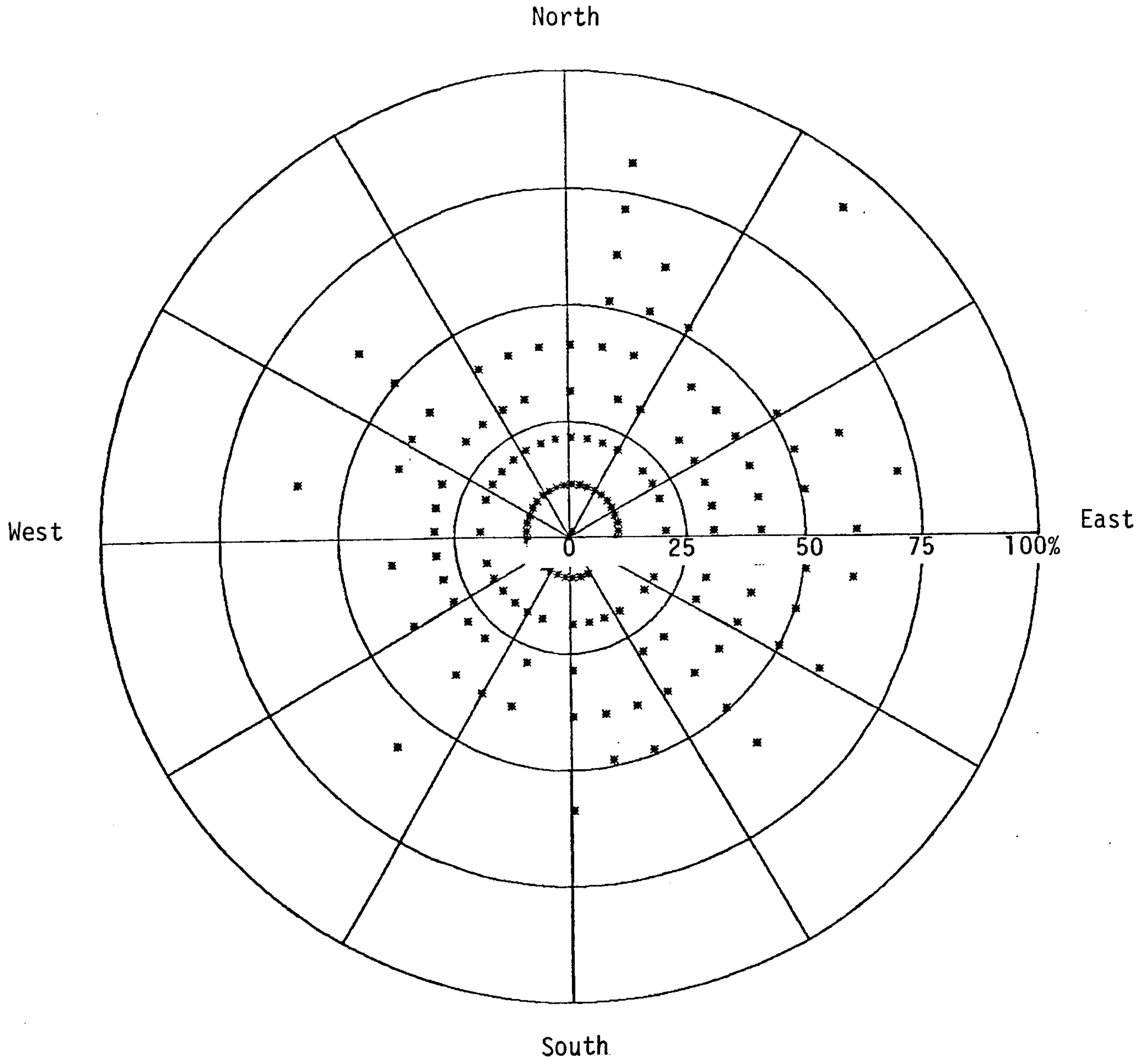
Distribution of Larch as a function of slope and aspect.



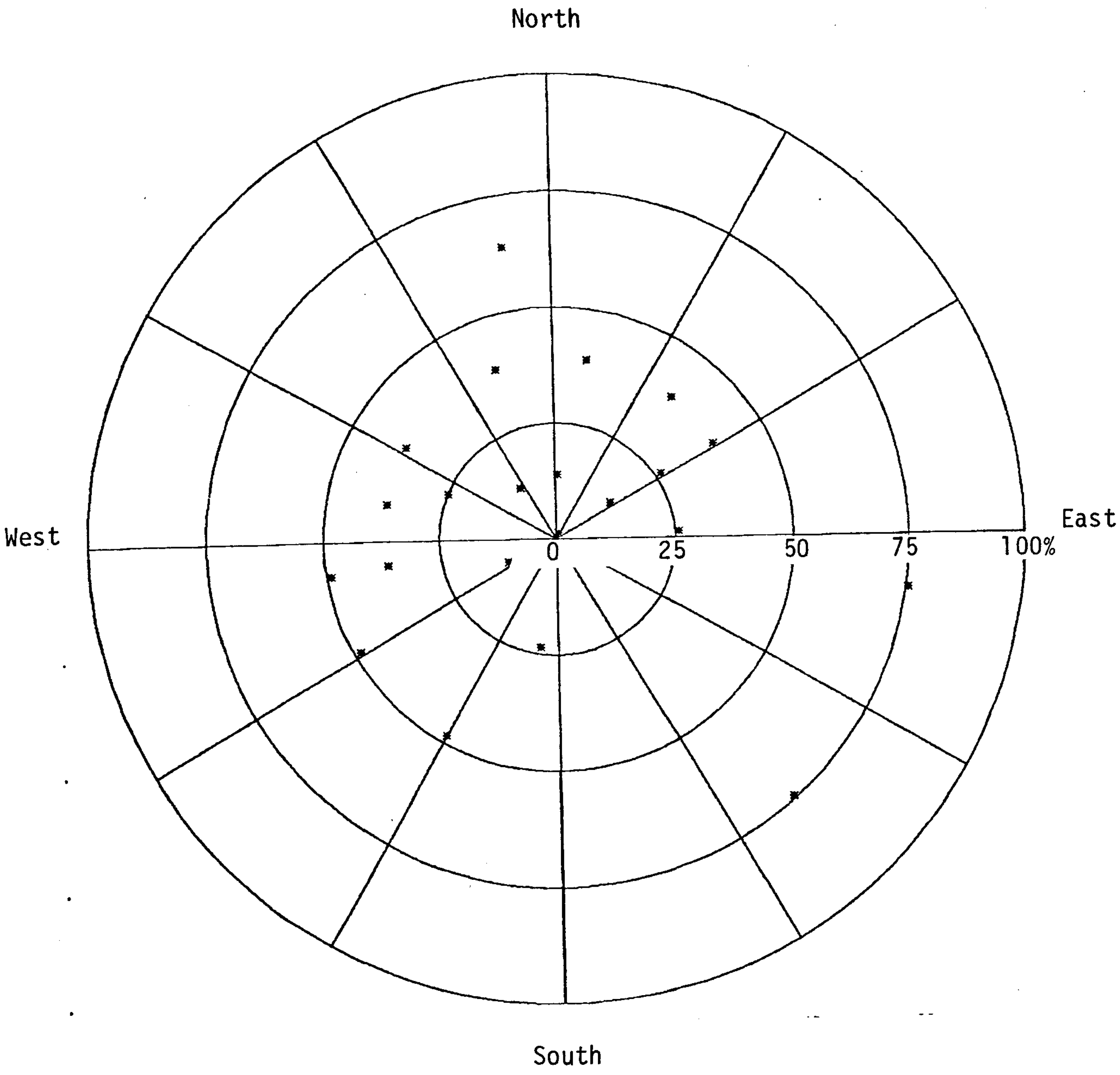
Distribution of Spruce/Fir as a function of slope and aspect.



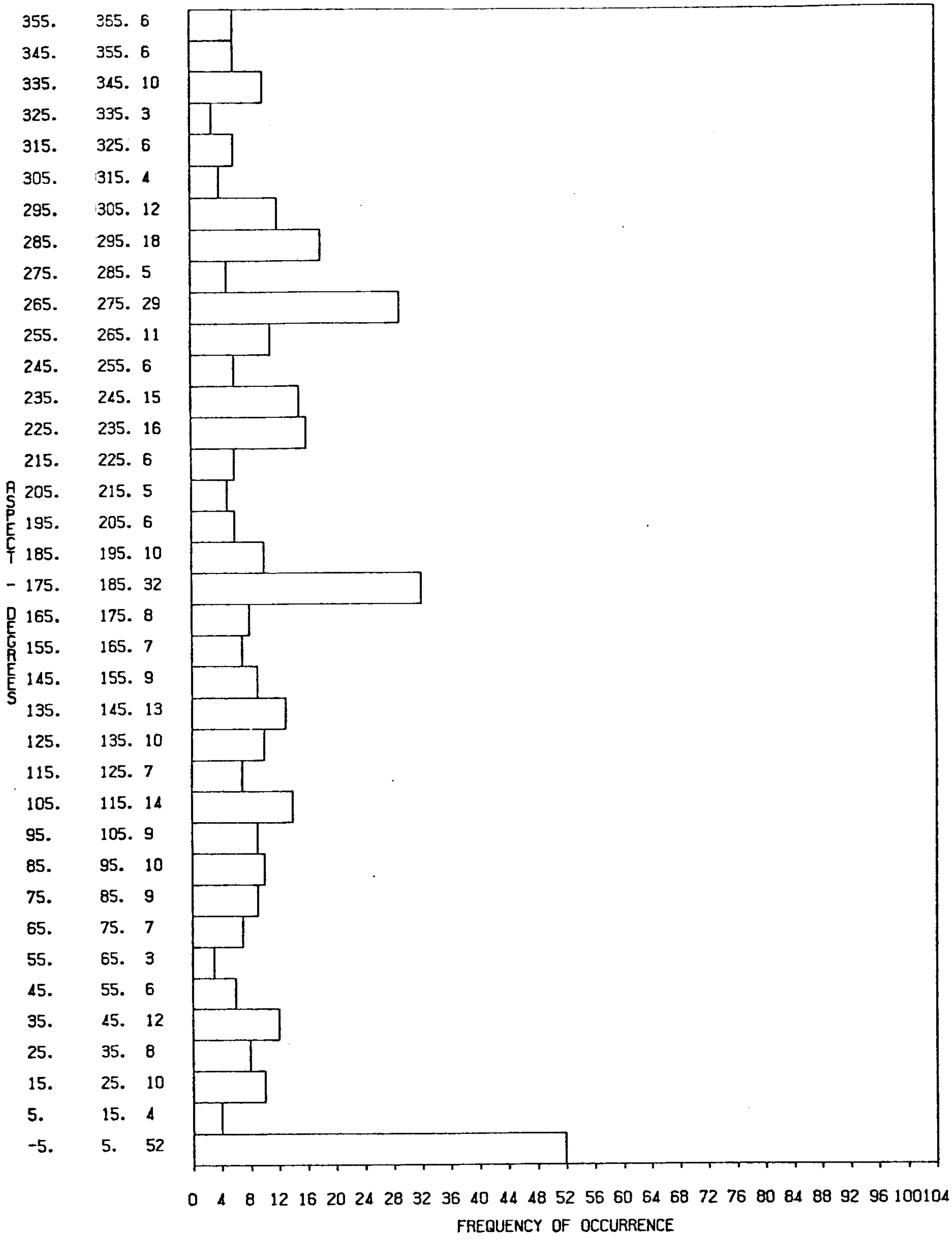
Distribution of Lodgepole as a function of slope and aspect.



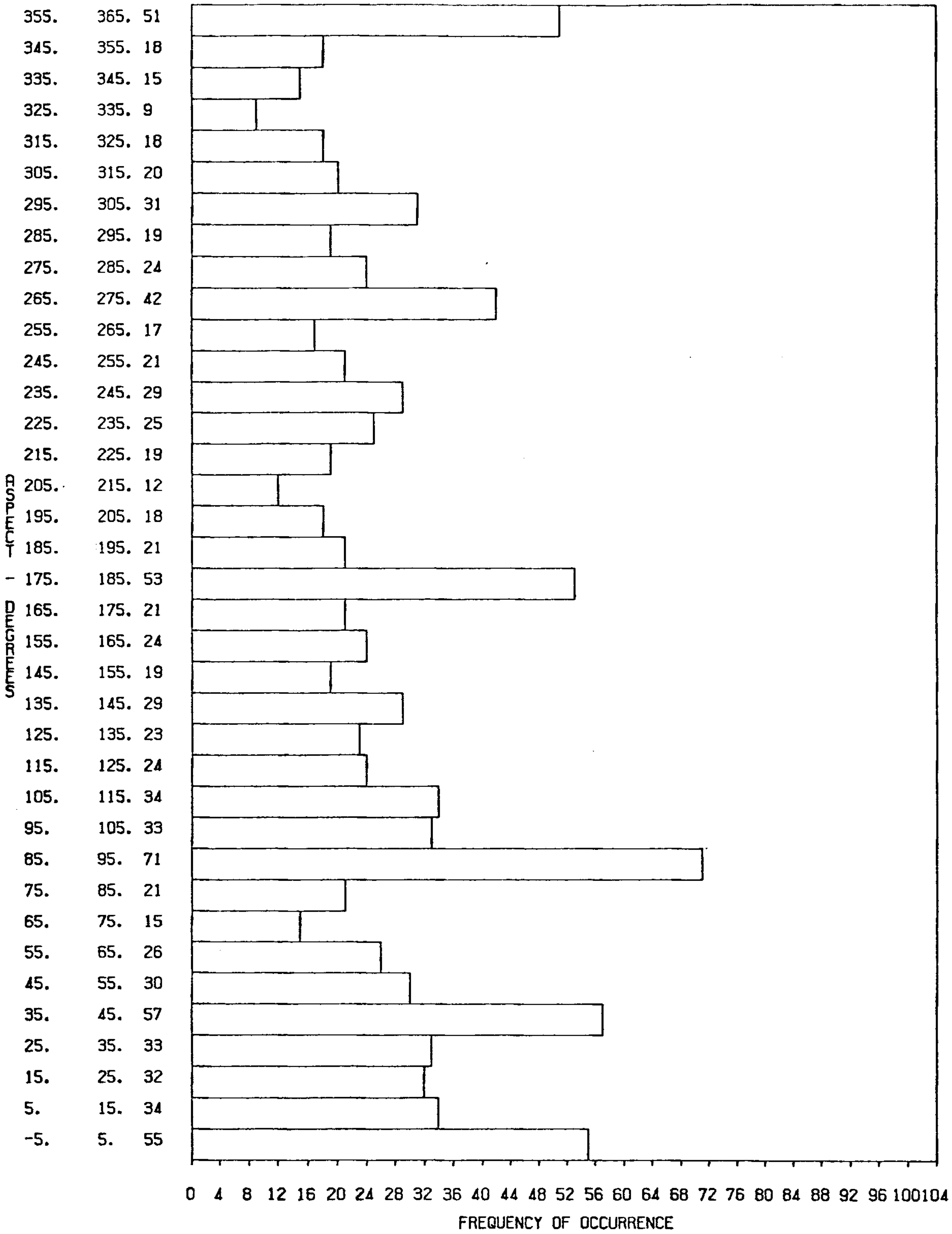
Distribution of Subalpine as a function of slope and aspect.



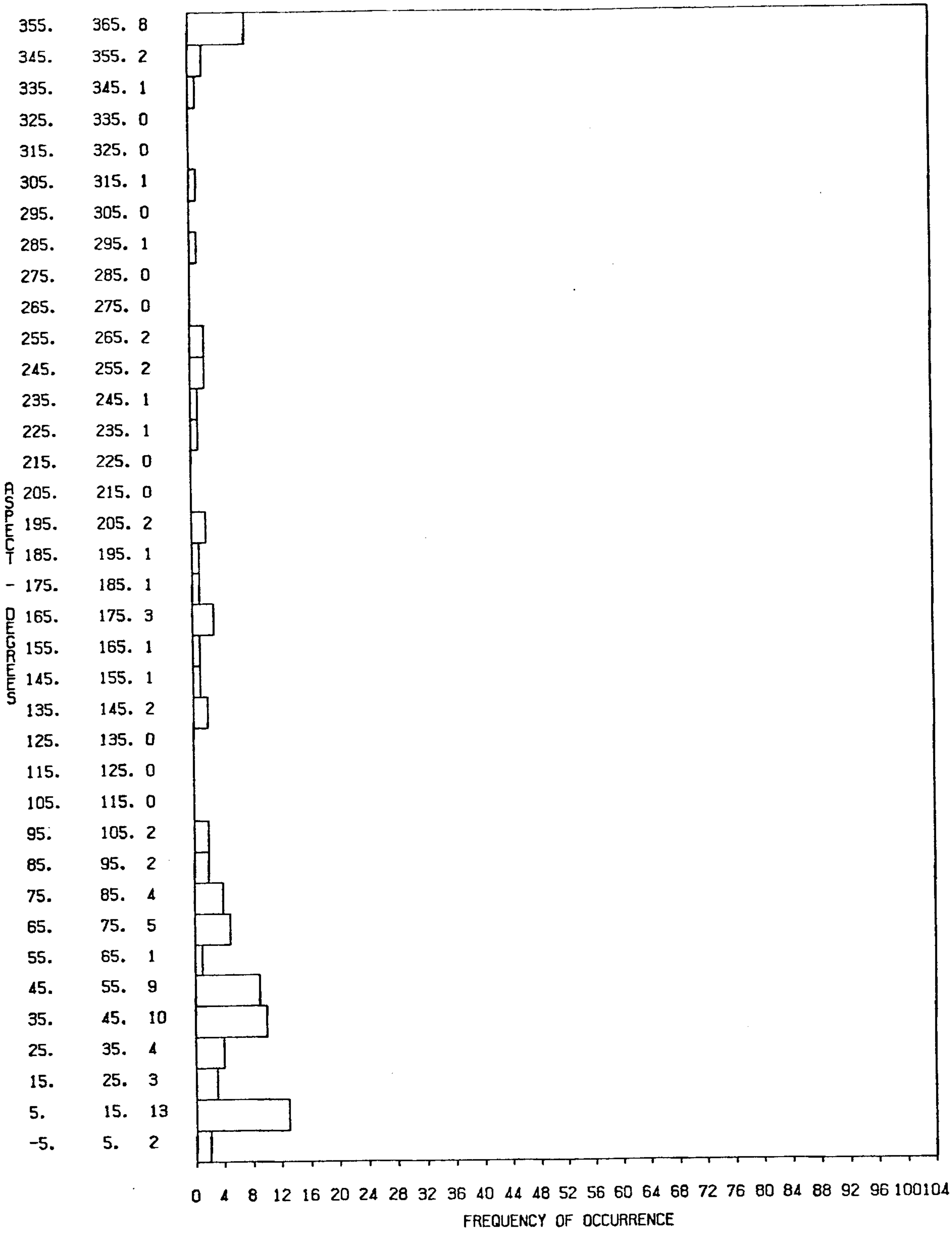
Distribution of Aspen as a function of slope and aspect.



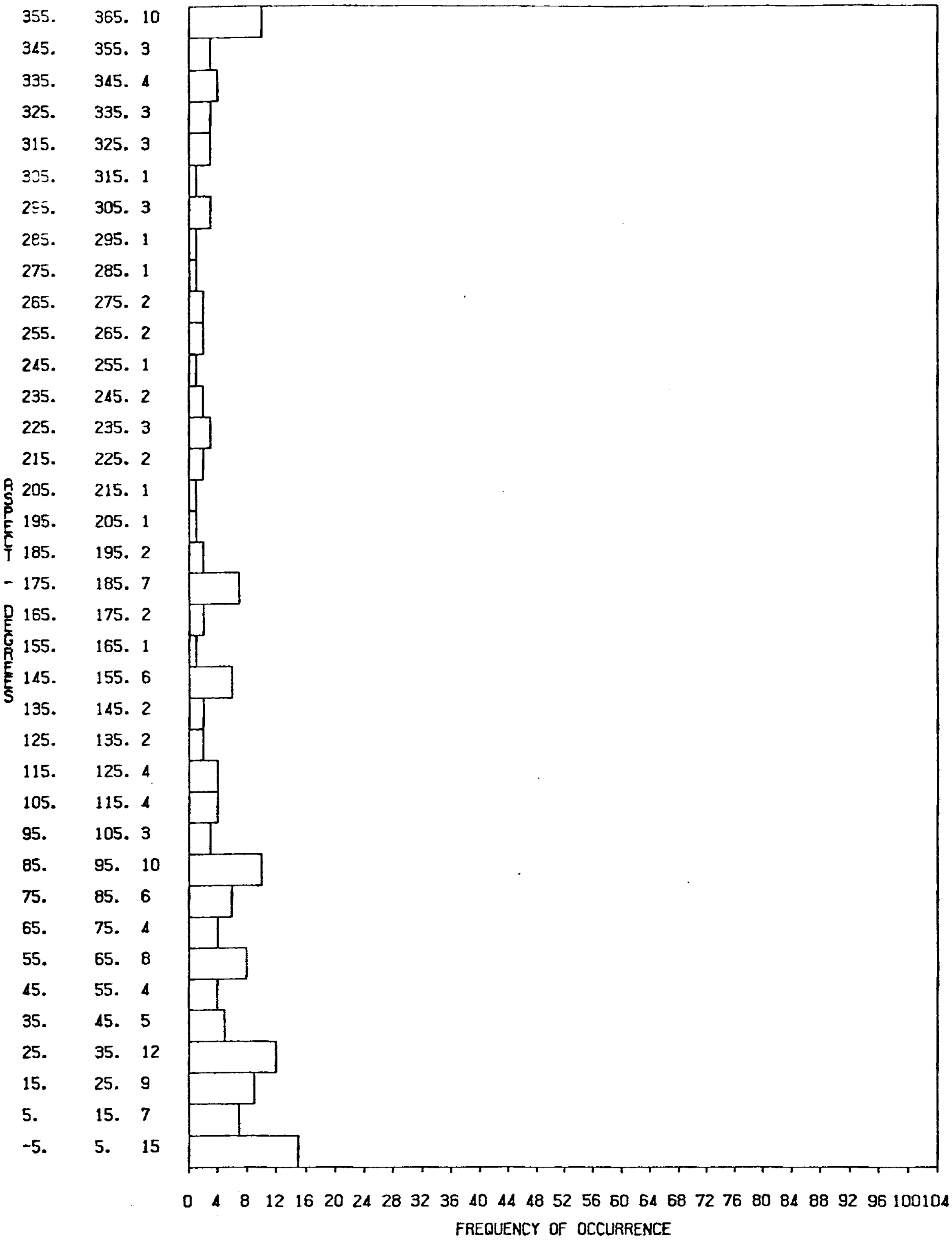
Frequency of Ponderosa Pine as a function of aspect.



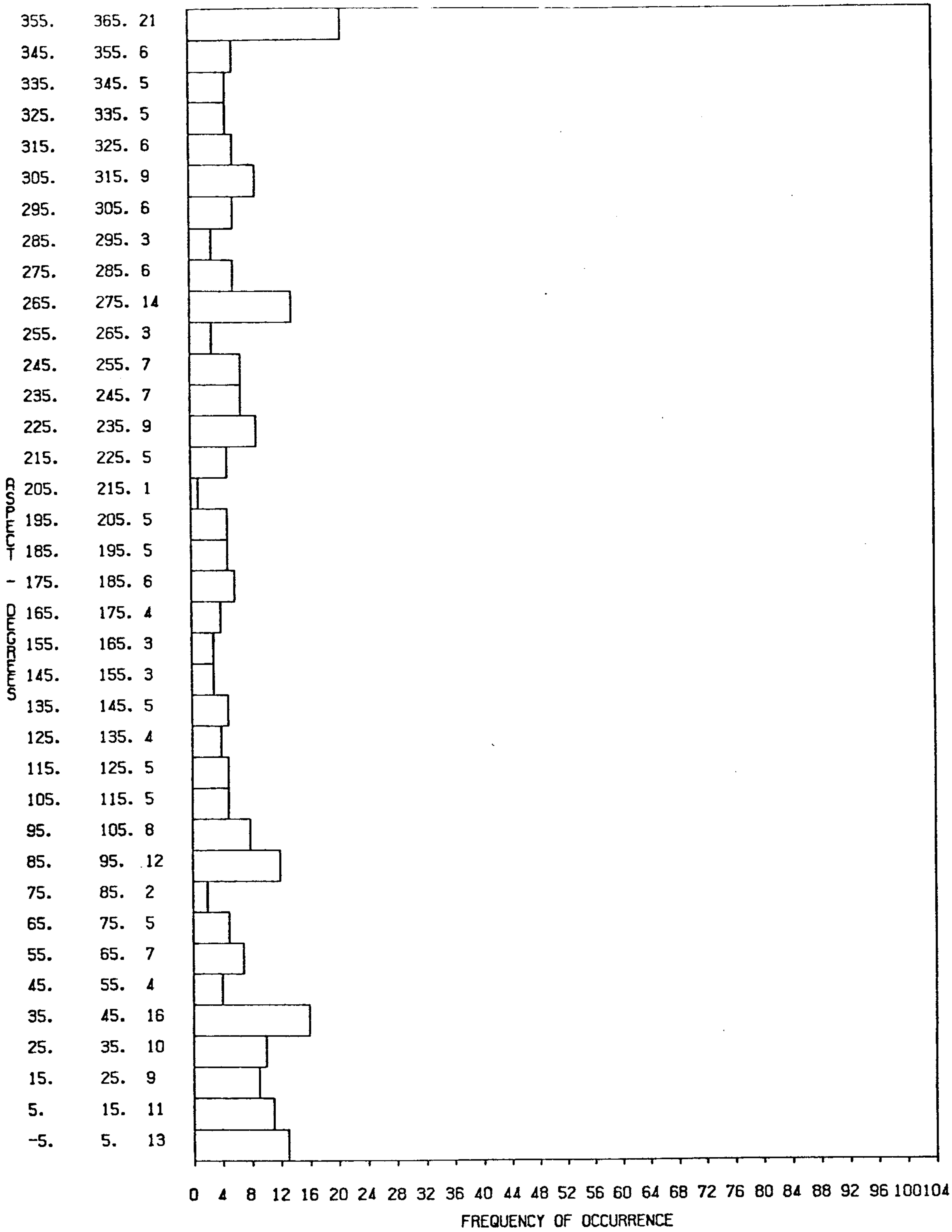
Frequency of Douglas Fir as a function of aspect.



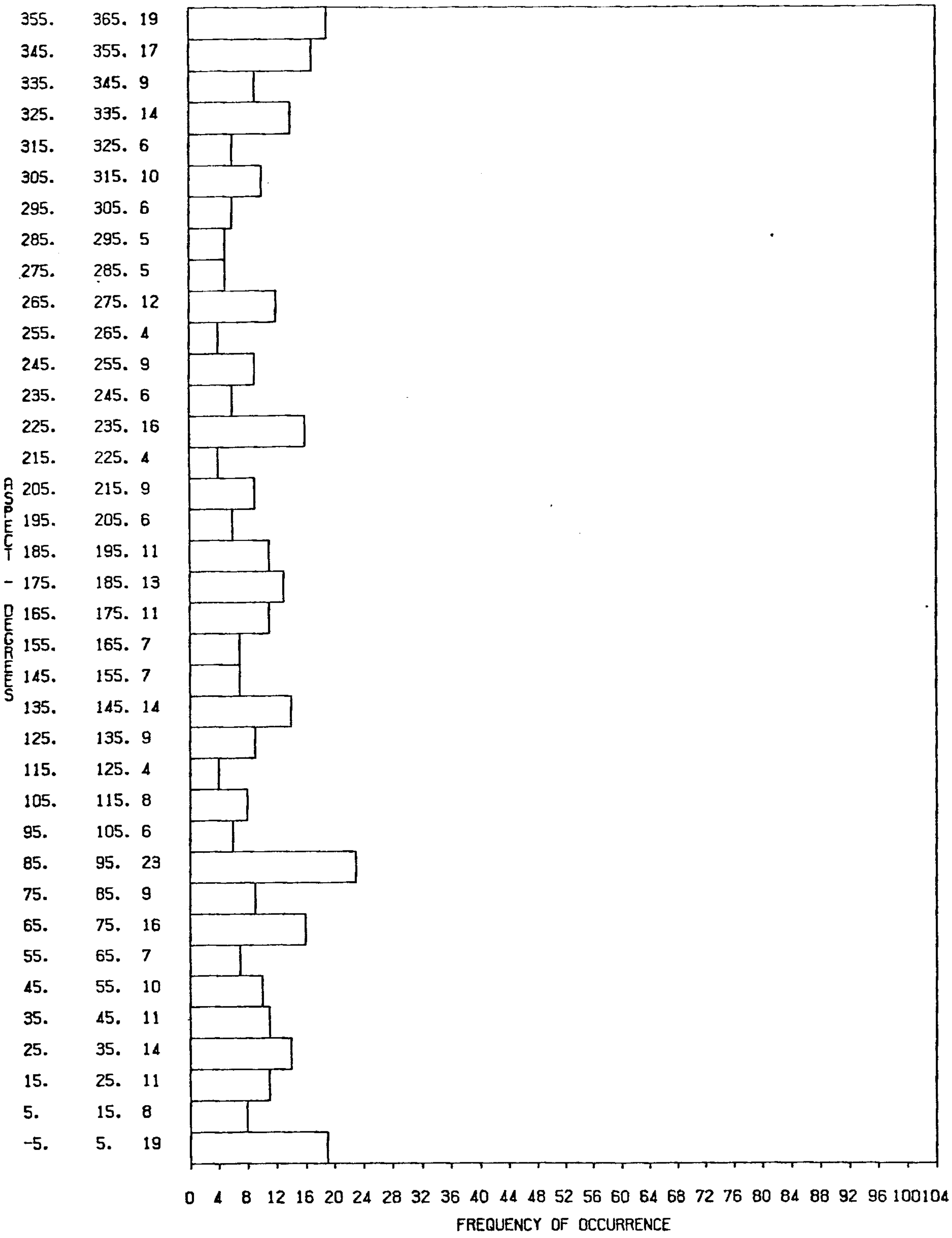
Frequency of Larch as a function of aspect.



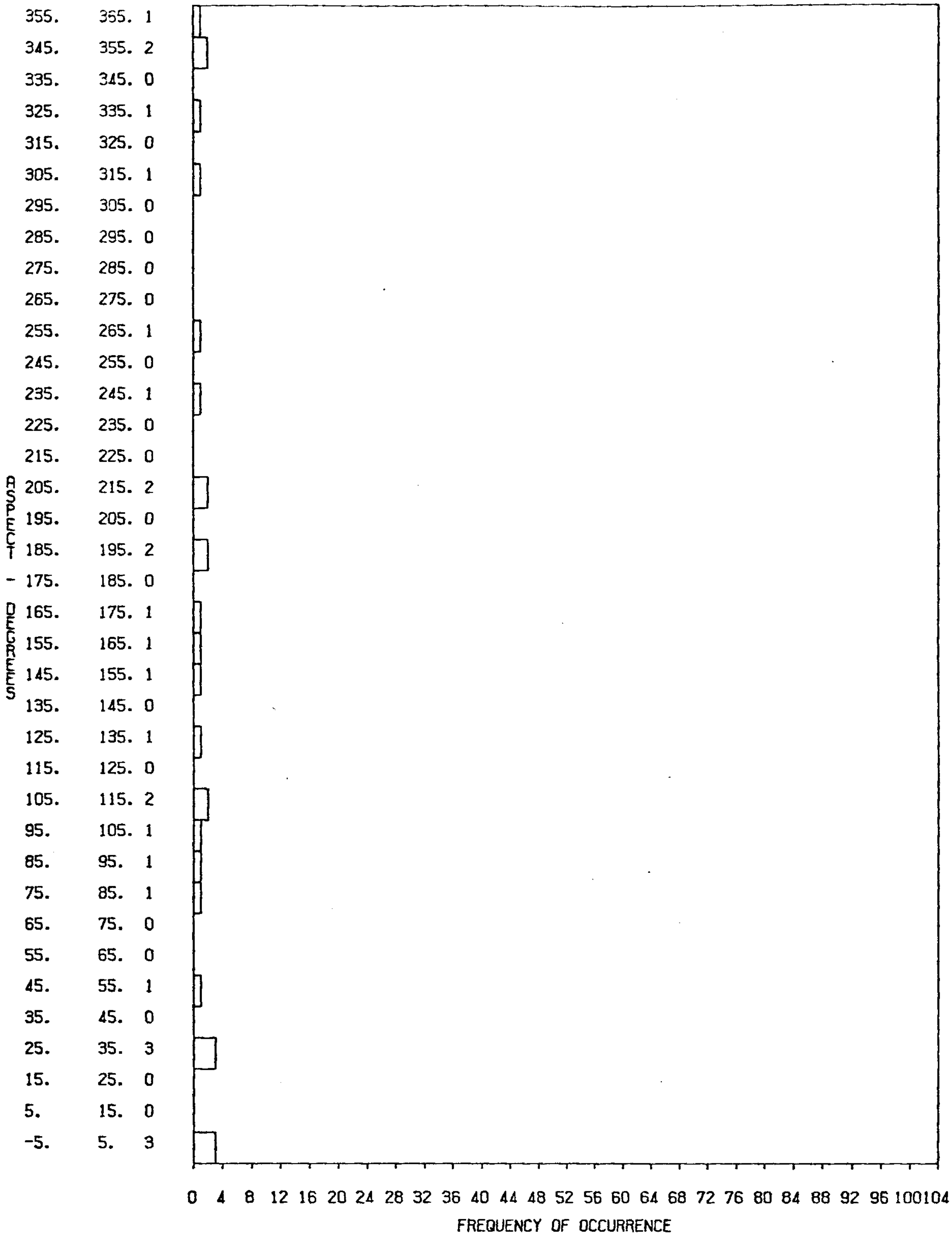
Frequency of Spruce/Fir as a function of aspect.



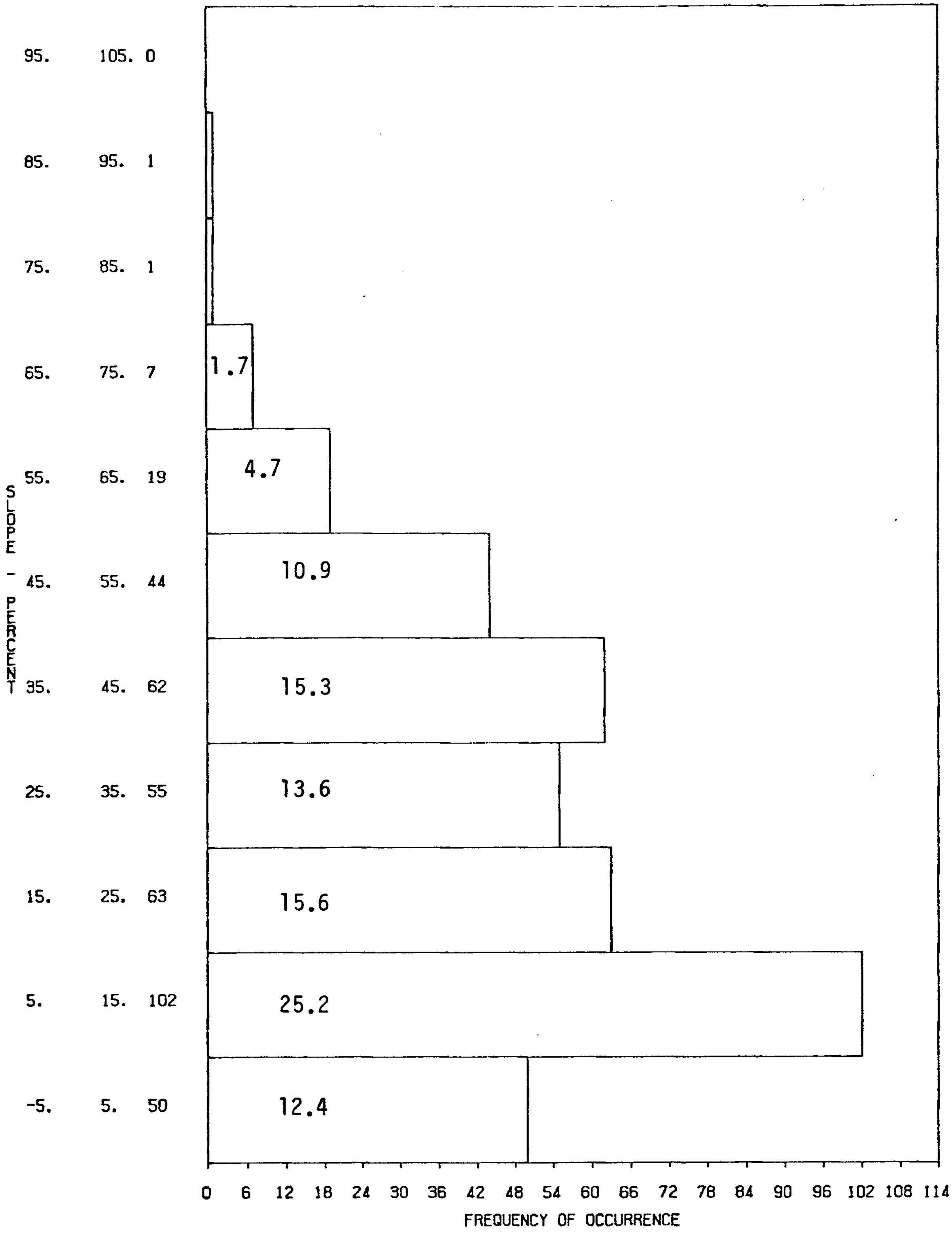
Frequency of Lodgepole Pine as a function of aspect.



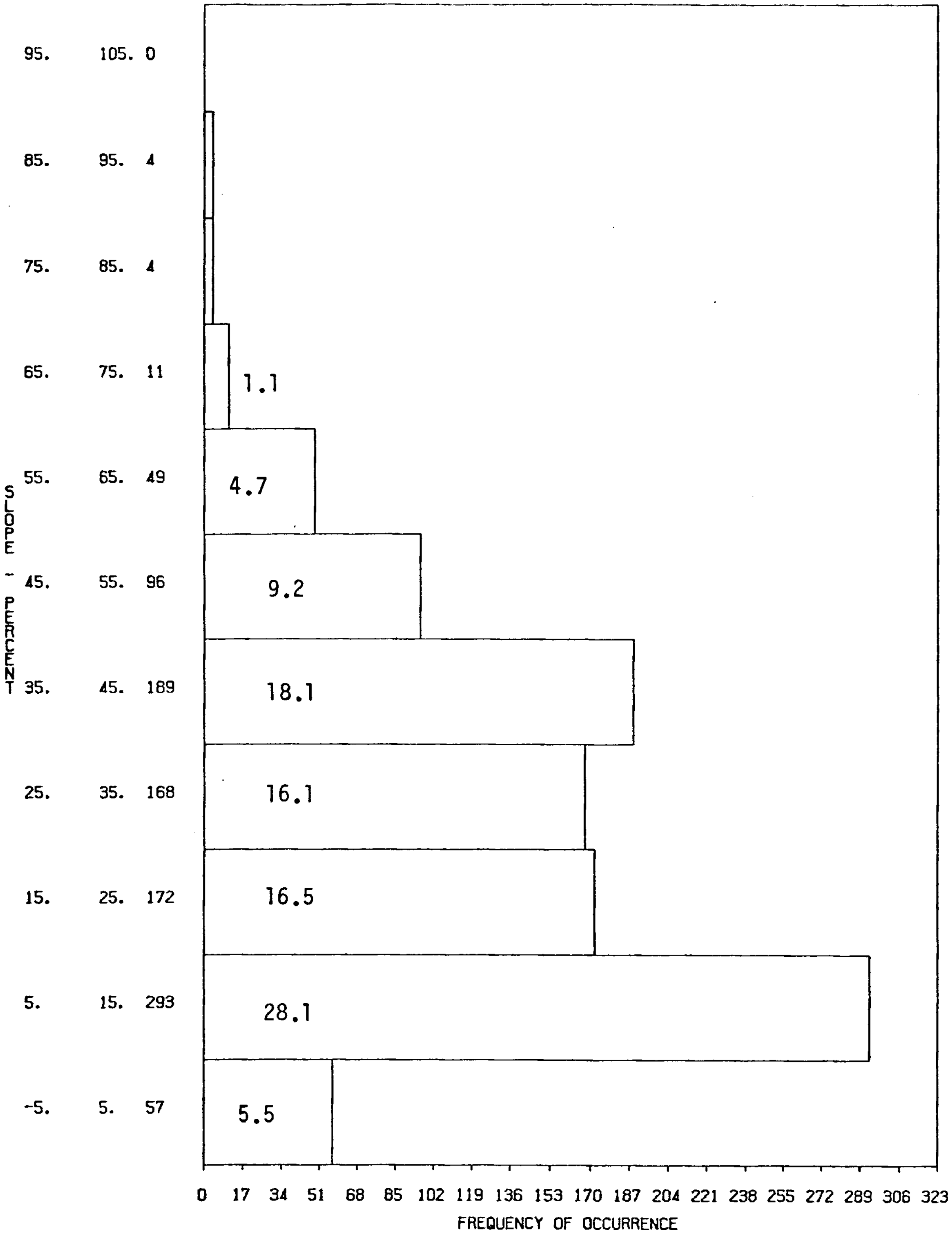
Frequency of Subalpine as a function of aspect.



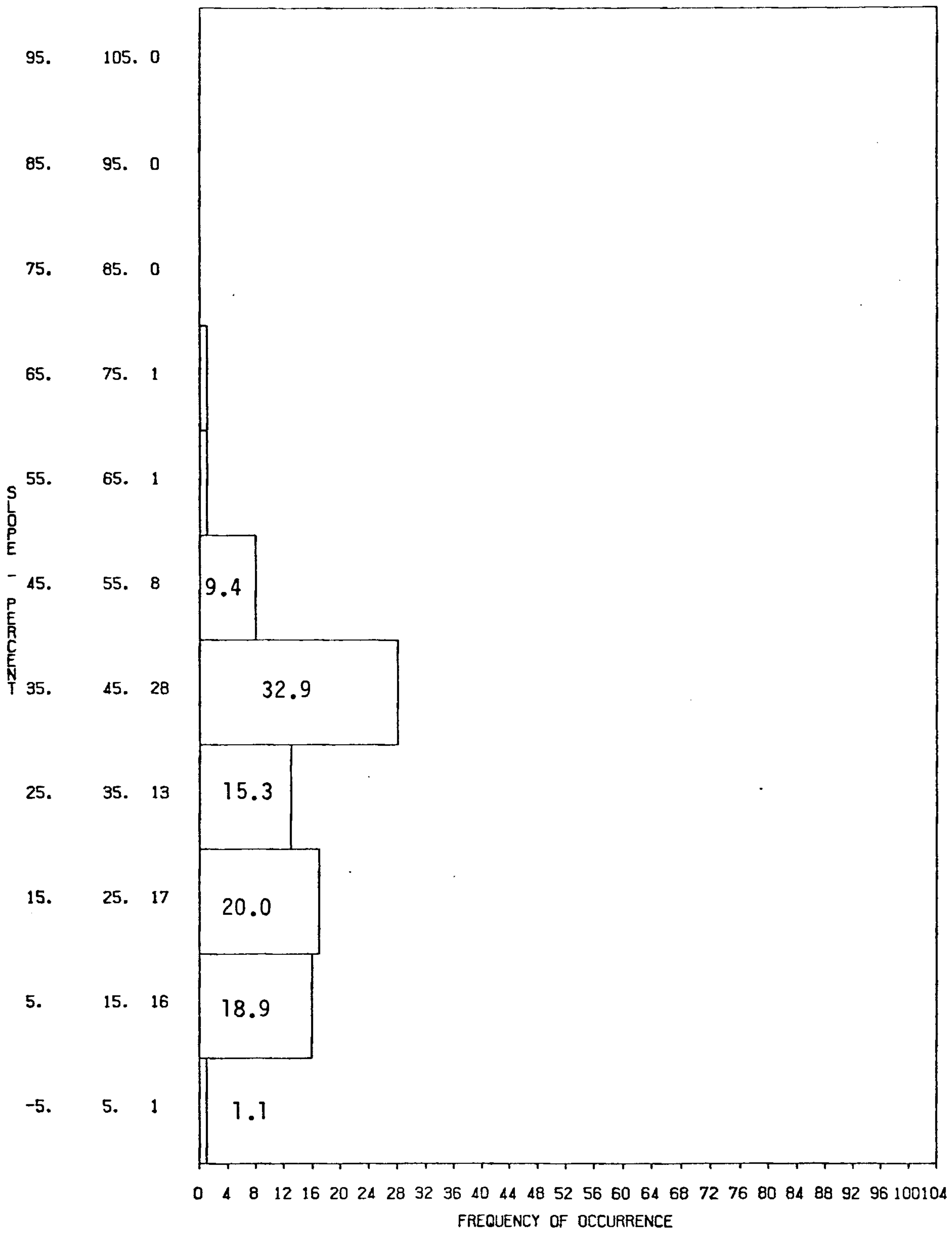
Frequency of Aspen as a function of aspect.



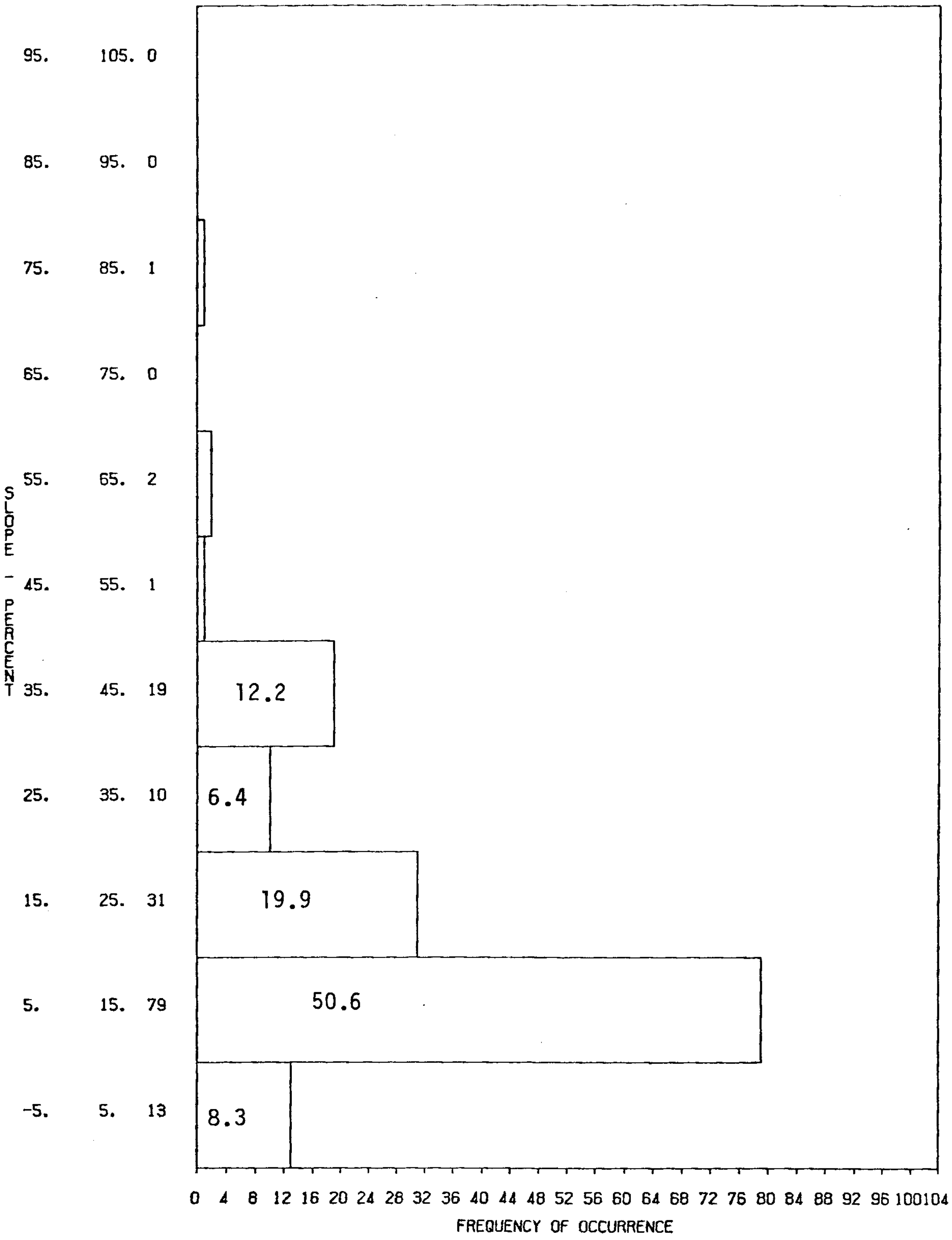
Frequency of Ponderosa Pine as a function of slope.



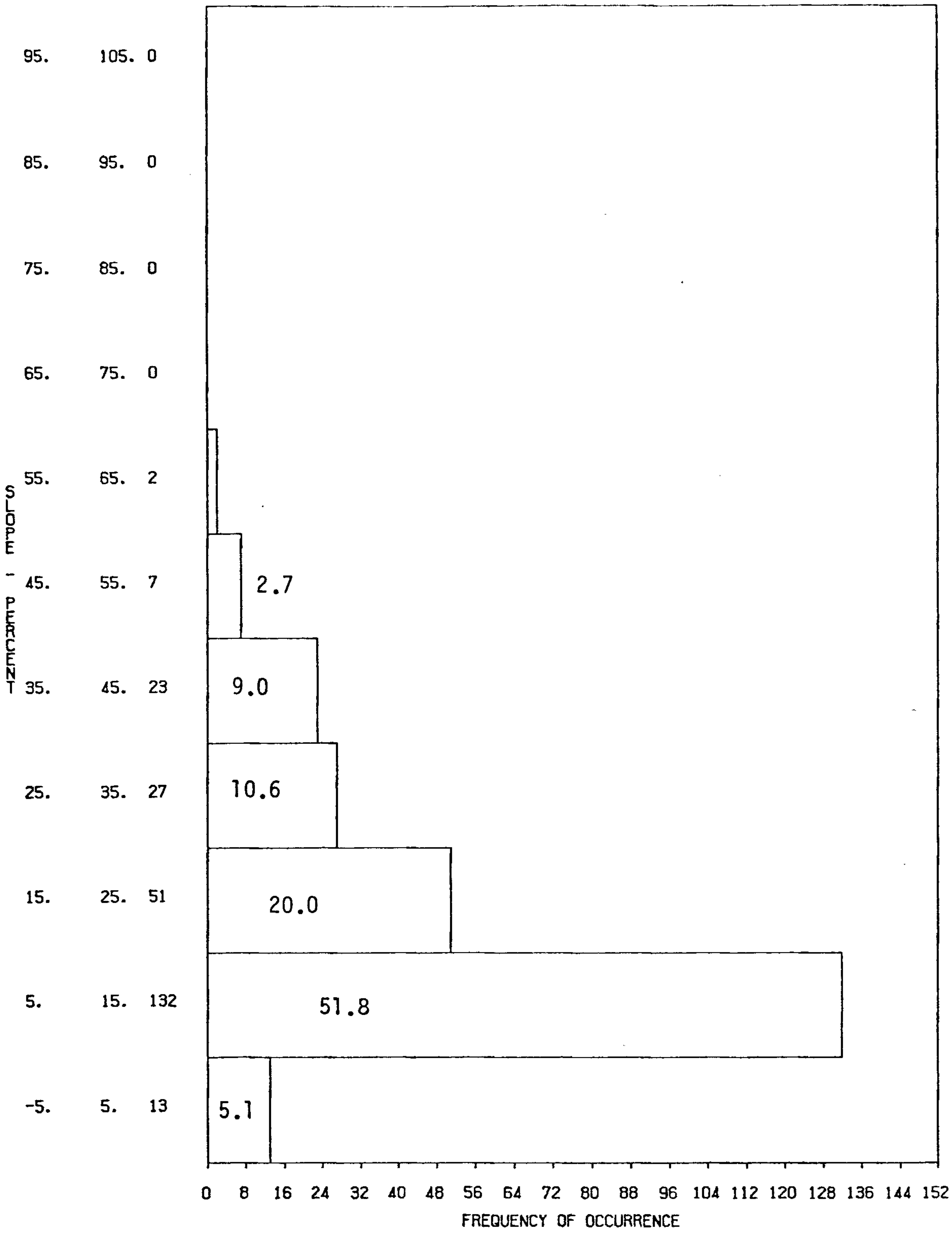
Frequency of Douglas Fir as a function of slope.



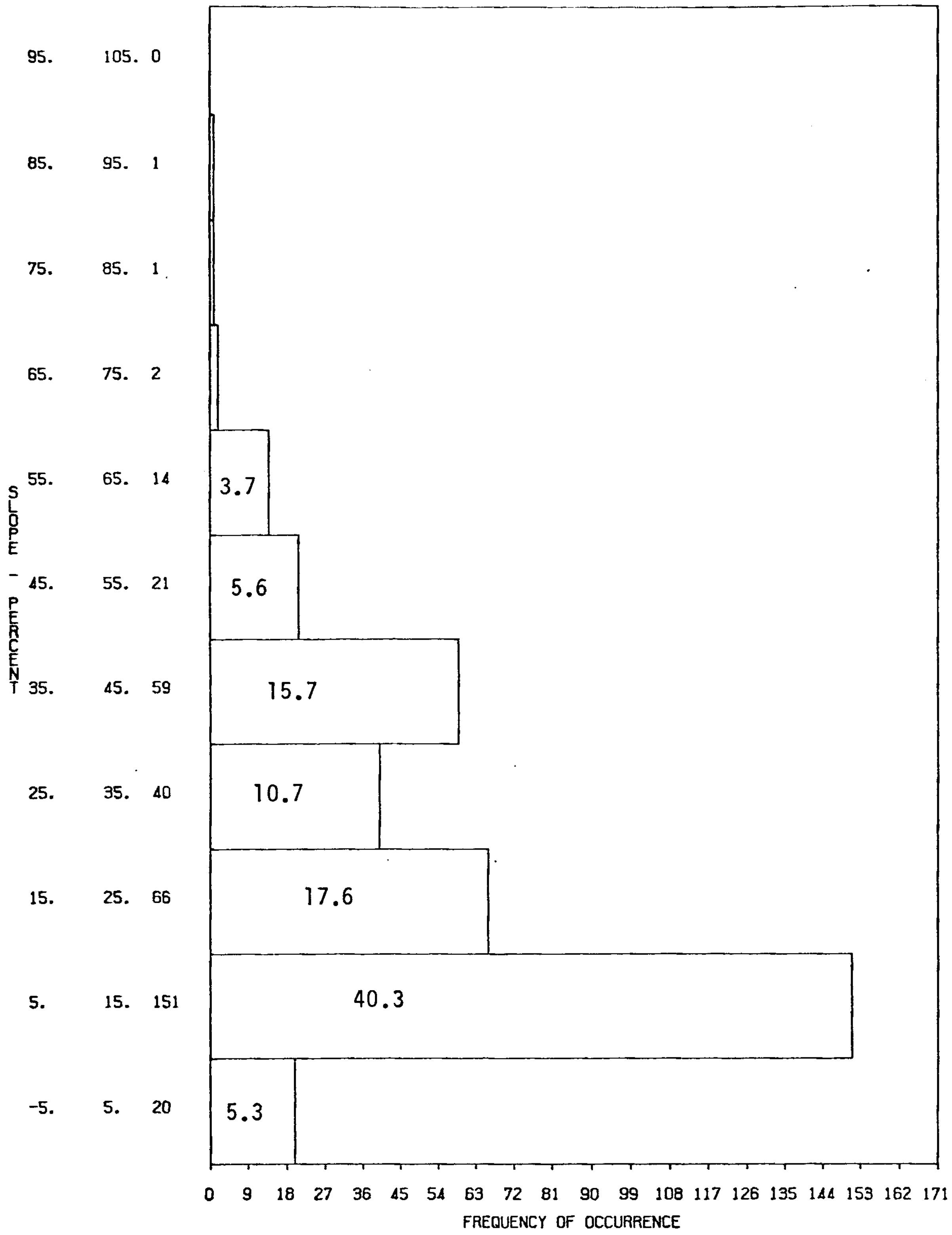
Frequency of Larch as a function of slope.



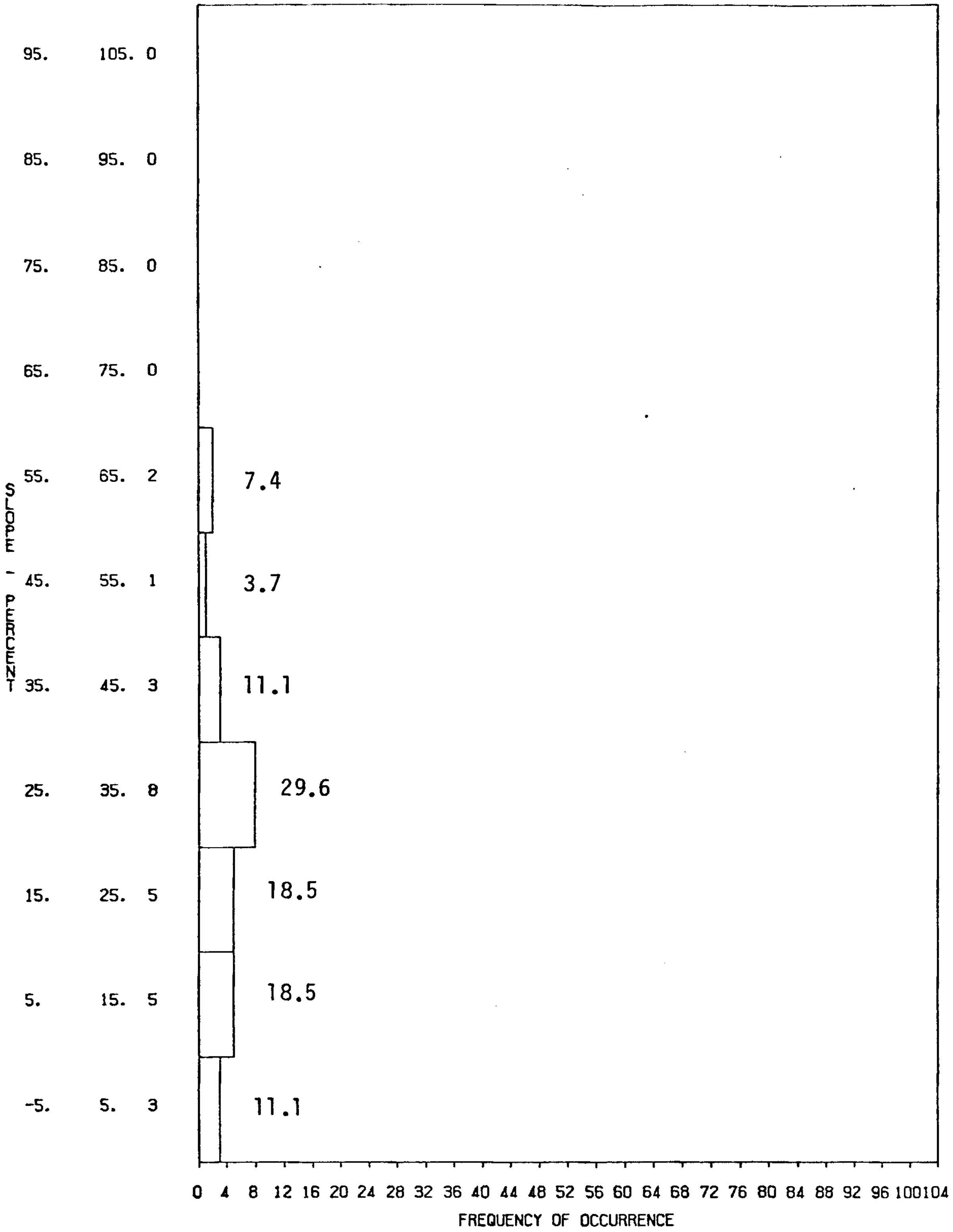
Frequency of Spruce/Fir as a function of slope.



Frequency of Lodgepole Pine as a function of slope.



Frequency of Subalpine as a function of slope.



Frequency of Aspen as a function of slope.

the 90^0 rotation of the topographic data sets that is necessary to allow the registration (overlay) of this data set with the Landsat MSS data.

E. Discussions With Washington DNR Personnel

During the period of September 10-19, Roger Hoffer traveled to Moscow, Idaho, where he presented an invited paper at the International Symposium on Remote Sensing for Natural Resources. He also had the opportunity to discuss the progress and expected accomplishments of the Washington spectral/topographic/GRIDS data analysis with Roger Harding of the Washington DNR. From Moscow, Dr. Hoffer went to Seattle, Washington, to discuss various aspects of the project with Bob Scott of the Washington DNR. In the discussions with both of these gentlemen, key points that were brought out included:

(1) The support of the Washington DNR thoughout the current portion of this project has been outstanding, and has been critical to the success of the project. Washington DNR has put several thousand dollars of effort into this project in order to better understand and assess the LARS approach to analyzing Landsat and Landsat + topographic data obtained over a forested area in rugged terrain, and to determine the value and limitations of the results obtained on this area in central Washington in relation to the needs of the DNR.

(2) Washington DNR personnel believe that the approach developed in Colorado for combining spectral and topographic data has potential, but there is a need to more fully evaluate the use of Washington GRIDS data in developing training statistics.

(3) Washington DNR would like to have some of the LARS software transferred to the Washington computer system so that DNR personnel would have the capability to pursue various analysis activities using the LARSYS software and some aspects of the LARS approach to data analysis.

During these discussions, it was clear that not all of these desires could be met during the current contract period. It was therefore agreed that a request would be made to NASA to modify and continue the existing contract to LARS in order to allow these additional objectives to be pursued. Such a request is in the process of being generated.

During the discussions, Dr. Hoffer also pointed out that the contract objectives can be met within the existing time-frame of the contract, but a complete spectral/topographic classification of the test site area (which is desired by DNR officials) will not be possible due to the amount of computer time required. Therefore, such a classification will also have to be carried out as part of a contract extension.

F. Publication of the Annual Report

Following the review by NASA personnel and publication approval, 100 copies of LARS Technical Report 011579 (covering the period 12/16/77 - 1/15/79) are in the process of being reproduced, fifty copies of which will be submitted to NASA/JSC as required by the contract.

G. Completion of Report on the P-1 Study

The P-1 Study has been completed and the final report is in the process of being reproduced as LARS Technical Report 102679. This study indicated that the use of GRIDS data to seed the ISOCLAS cluster algorithm in the process of developing training statistics would have considerable potential, but that ISOCLAS is a difficult program to use effectively due to the large number of interrelated control parameters. The specific conclusions reached during this study include the following:

1. The P-1 approach using ISOCLAS in an unseeded, iterative mode and the Multicluster Blocks approach using CLUSTER were the two best approaches to developing training statistics. The use of either

method would basically depend upon the type and amount of ancillary information available to the analyst.

2. The P-1 approach makes maximum use of current forest inventory information. Availability of this information minimizes the analyst involvement in the classification procedure. The reduction in analyst involvement is critical because (a) it reduces the largest source of bias, and (b) in an operational mode, analyst time would be one of the biggest cost factors.
3. LACIE parameters should not be used to classify forested areas. The essentially noniterative ISOCLAS clustering processor and K-Nearest Neighbor (K=1) method of labelling the ISOCLAS statistics do not produce consistent results in a heterogeneous area.
4. Seeding ISOCLAS with Type 1 dots reduces CPU time, and shows promise for classifying forested areas. Preliminary research indicates that each cover type should have at least two seed dots to better insure that the smaller cover types are spectrally represented.
5. When K-Nearest Neighbor is used to label the spectral classes, K should equal that number of dots in the covertype with the smallest dot representation in the dotfile.

It was found that in addition to the possibilities, P-1 does have it's restrictions. First the dotfile must contain a sufficient number of pixels in each cover type of interest and should cover the range of variability within a cover type GRIDS data or forest inventory data (which provide only information on forested plots) may be used, but additional dots must be identified and located in all of the nonforest cover types (urban, grassland, barren, and water). The accurate location and identification of all dots is critical, for mislabelled or poorly located dots may lead to spectral classes incorrectly identified.

The Multicluster Blocks approach to developing training statistics produced classification results on par (statistically indistinguishable) with the best P-1 results. The McB approach requires a great deal of analyst interaction which, under certain circumstances, is desireable. Experience has shown that P-1's automated labelling processor often misses critical cover type that make up only a small percentage of the study area. Also, naturally, LABEL cannot abide by arbitrary definitions unless those definitions are exemplified in the dotfile. Crown density differences (for instance, between barren and conifer, where crown densities less than 30% are considered nonforested) may lead to mislabelling. These problems are overcome using the McB approach since spectral class labelling is done by the analyst. The McB approach is also efficient when developing training statistics for large study areas. P-1 requires the entire study area be clustered, and this can become expensive when the area involved reaches into the hundreds of thousands of acres, unless the area is clustered on an interval of 2, 3, or 4 (i.e., sample one-fourth, one-ninth, or one-sixteenth of the pixels). The clustering of representative training blocks (Multicluster Blocks approach) or clustering of subsets of the entire data set may be more realistic. Further work on this aspect of efficient use of P-1 should be pursued.

Because of the questions raised concerning the effective use of P-1 in different geographic areas, and because of very limited personnel and computer resources, it has been decided that the P-1 approach to developing training statistics will not be used on the Washington test site, unless there is a contract extension. However, this preliminary study of P-1 has provided a basis for future work with the P-1 approach and valuable insights were obtained concerning the strengths and weaknesses of the P-1 approach.

II. PROBLEMS ENCOUNTERED

No major problems were encountered during this reporting period, except for some unforseen delays in developing the DMA/DTM reformatting and geometric correction software.

III. PERSONNEL STATUS

The following personnel were actively involved in this investigation during the present reporting period (7/16/79 - 10/15/79):

Name	Level of Effort
Dr. L. Bartolucci	30%
M. Fleming	88%
Dr. R. Hoffer	28%
R. Nelson	25%
B. Prather	20%

IV. EXPECTED ACCOMPLISHMENTS

During the next reporting period it is expected that the DMA/DTM data will be registered onto the Landsat MSS data and that the analysis (classification) of the combined topographic and spectral data set will be completed and evaluated. The final products required by NASA/JSC will be generated and incorporated into the final report for this contract.