1975.

STATISTICAL ANALYSIS

OF A

SOIL-SITE STUDY

by

Constance Ann Harrington

A thesis
submitted in partial fulfillment
of the requirements for the
Master of Science degree

State University of New York
College of Environmental Science
and Forestry
Syracuse, New York

August 1975

Approved:
Department of Silviculture
and Forest Influences

Approved:
Committee on Graduate
Studies

____
Major Professor

____
Chairman, Examining
Committee

____
Department Chairman

____
Dean of Graduate Studies

ACKNOWLEDGMENTS

I wish especially to thank the members of my committee, Dr. J. W. Johnson, Dr. W. M. Stiteler, and Dr. P. J. Craul, for giving me both the freedom I wanted and the advice I needed. Special thanks are due to Dr. R. D. Westfall for his assistance and encouragement in the statistical analysis.

I also wish to gratefully acknowledge Jack Karnig, manager of the Harvard Black Rock Forest, for his support in the field phase of this study.

TABLE OF CONTENTS

ACKNOWLEDGMENTS....................................ii

LIST OF TABLES..................................... v

LIST OF FIGURES....................................vi

CHAPTER 1. INTRODUCTION........................... 1

CHAPTER 2. LITERATURE REVIEW...................... 3

CHAPTER 3. THE STUDY AREA..........................18

A. General..18
B. Geology..21
C. Soil...23
D. Climate..25
E. Stand and Site Measurements......................25
F. Summary of Measurements..........................26

CHAPTER 4. INTRODUCTION TO MULTIVARIATE STATISTICAL TECHNIQUES.........................32

A. General..32
B. Graphical Representation of Probability Functions..33
C. Discriminant Analysis............................37
D. Cluster Analysis.................................40
E. Multiple Regression Analysis.....................43
F. Canonical Correlation Analysis...................46
G. Principal Components Analysis....................48
H. Factor Analysis..................................52

CHAPTER 5. RESULTS AND DISCUSSION..................55

A. Discriminant Analysis............................55
B. Cluster Analysis.................................61
C. Multiple Regression Analysis.....................63
D. Canonical Correlation Analysis...................67
E. Factor Analysis..................................69

CHAPTER 6. SUMMARY AND CONCLUSIONS....................77

A. Regeneration Variability........................77
B. Analytical Techniques..........................81

REFERENCES CITED.......................................84

APPENDICES..94

A. Appendix A....................................94
B. Appendix B....................................95

VITA..96

LIST OF TABLES

Table		Page
3-1	Number of stems per acre by species and diameter class	27
3-2	Per acre basal area distribution by species	28
3-3	Per acre basal area distribution by ten square foot classes	28
3-4	Summary of physical plot measurements	30
4-1	The basic models for principal components analysis and factor analysis	53
5-1	Classification by discriminant analysis	57
5-2	Variable factor loadings for the first four extracted factors	72
5-3	Importance of variable-factor correlations	74

LIST OF FIGURES

Figure | Page

3-1 Map of the Harvard Black Rock Forest..........19

3-2 Map of Compartment XXV, Harvard Black Rock Forest....................................20

4-1 Probability distribution for the diameters in an even-aged red maple stand...............34

4-2 Joint probability distribution for red maple diameter and tracheid length in an even-aged stand...34

4-3 Joint probability distribution for red maple diameter and height in an even-aged stand.....36

4-4 Two dimensional representation of a two variable joint probability distribution.......36

4-5 Probability distribution of discriminant scores plotted over two populations...........39

4-6 Representation of the discriminant scores plotted over two groups with reference to their probability ellipses....................41

4-7 Trivariate observations with the three principal axes drawn in...........................50

INTRODUCTION

During the winters of 1953-1954 and 1954-1955, a twenty-seven acre area was clearcut on the Harvard Black Rock Forest, Cornwall, New York. An assessment of the hardwood regeneration present on the area was undertaken during the summer of 1973. Brief analysis of the data obtained from a hundred sample plots revealed a tremendous diversity in both the type and the amount of regeneration present. Speculation arose as to whether the differences between plots were the result of the effects of various site factors, or whether the differences merely demonstrated random fluctuations about some population mean.

The following field season, 1974, half of the original plots were selected for measurement of various physiographic and edaphic factors. The field phase of the study was completed in July, 1974. Some of the multivariate statistical techniques were used to analyze the data. Analysis began in August, 1974 and was completed in March, 1975. Since multivariate statistics have not been often used in forestry research, determining the type and usefulness of the information obtained from the methods utilized was also of interest.

The objectives of this study were:

I. To determine if species density and diversity in a young, mixed oak stand can be correlated with or explained by physical site variables acting either jointly or alone.

II. To determine if multivariate statistical analysis (MSA) can yield meaningful results from soil-site data. That is,

A. What type of information does MSA give to the forest researcher; and

B. How practical are the techniques herein utilized.

LITERATURE REVIEW

Efficient forest management practices must recognize inherent productivity differences between forest sites. Identification of this need has resulted in hundreds of published forest site studies in the United States alone since the early 1900's, most of them undertaken to rate sites in terms of their growth potential. These studies have usually aimed at achieving either quantitative predictive ability or biological understanding of forest productivity. As Mader (1965) points out: "Our ultimate goal should be the capacity to identify the exact factors limiting growth and the magnitude of response to be expected from changes in these factors, including genetic variation. From this information decisions could be reached on whether it is physically or economically feasible to change the environment."

The many published studies only represent the tip of the iceberg in terms of the total number of studies or research projects that have been undertaken in this field of forest site evaluation. While this state of affairs is fairly common in any well researched field, it can result in an unrealistic view of the field when viewed by someone outside of that field. When studies are originally reviewed by an editor a natural bias usually exists to accept the "best" studies for publication, i.e., those studies in which good relationships were determined between the variables of interest. Studies of limited geographical coverage or studies that didn't report significant results are usually not published. When reviews of the literature are undertaken, abstracting of research projects tends to result in concentrating on the major factors which were found to be useful in analysis. What are considered important studies are those which, in the mind of the reviewer, provided new or useful information. Reading many of the published papers gives the impression that the techniques used in site evaluation have worked well in practice. This is not so. The paucity of studies published in the past ten years gives a good indication of the current lack of interest in these techniques which have not been found to solve the problems in forest soil site evaluation. Little concern has been evidenced to see further permutations of the ideas that just haven't worked well in most applications.

Mader (1965) has outlined five basic approaches to site classification which have been used in this country. His classification, which is based on both the underlying philosophy and the mechanics employed in the study, will be followed in this review. These approaches follow a loose chronological sequence in their order of discussion. However there is a certain amount of overlap in time between these methods of site classification.

The first method involves developing sets of site index curves by determining the height-age relationship of a large number of dominant or codominant trees from a broad range of stands. Then stands of the particular species (or group of related species) can be rated or ranked by growth rate. The height growth of the main canopy is thus used as the integrator of the environmental factors.

These studies often reported yield table data by site index classes, which is useful information for management planning. Also, this method may be suitable for short term purposes or low level extensive management. However, it is inadequate for determining site classification and evaluation. Little information is gained except concerning the measured sites. Areas without trees or lacking suitable stand conditions can not be evaluated. Comparisons with other species can not be made without either a long history involving rotations of different species on a site or unless several dominant species are present on the area. Although relatively few, such studies on comparative site indices have been published (Copeland, 1956; Curtis and Post, 1962; Deitschman and Green, 1965; Doolittle, 1958; Foster, 1959).

Most site index curves developed for an area have been anamorphic, that is having the same shape. Many authors have discussed the problems underlying the assumption that height growth patterns are uniform within a given forest region (Bull, 1931; Carmean, 1970, 1971; Heiberg and White, 1956). Hicock et al. (1931) presented a set of polymorphic site index curves for red pine; however, polymorphic curves have not become common.

Heiberg and White (1956) suggested using current site index, that is the current height growth of a given stand at any given time, to indicate current site quality rather than average site quality as measured over the life of the stand by conventional site index.

One of the assumptions underlying the preparation of site index curves is that height growth is unaffected by stand density. Thus plots from differing stand densities can be used to construct the curves. However, this assumption does not hold for some species (Gaiser and Merz, 1951; Gevorkiantz and Scholz, 1944).

Another drawback to using site index classes as ratings of productivity is pointed out by Mader (1961, 1963). Height measures and volume measures (more merchantable expressions of growth) for red pine did not correlate well with each other and did not agree in the manner in which they correlated with site variability. The best correlation was obtained with some measure of volume growth. Mader (1965) concluded that height measurements may not be sensitive enough to serve as the basis of site classification.

The second general approach to site classification has been based on the use of observed or measured soil properties and topographic features. The technique used in simplest form by Haig (1929) and Hicock et al. (1931) and extended to more complex multiple regression form by Coile and his co-workers (Coile, 1935, 1940, 1942, 1952; Coile and Schumacher, 1953; Gaiser, 1950, 1951; Ralston, 1951) was to correlate growth with a small number of soil physical characteristics such as texture or depth of soil horizons plus other variables measuring topographic features or past land history. These studies have tried to develop efficient predictive equations for site index based on trial and error combinations of variables with strong emphasis placed on the idea that only a few variables could explain the observed site differences. Relatively little concern has been given to explaining the basic biological relationships involved. Coile (1952), Ralston (1964), and Berglund (1968) reviewed the work to date quite well.

Hodgkins (1959) evaluated the soil-site index regression technique from the standpoint of statistical validity. He points out that the two conditions for validity of linear regression analysis are not always met; namely that 1) the regression be strictly linear; and 2) the error terms (the variable portion of the dependent variable not defined by the regression line) be normally distributed about the regression line in a random manner and be independent of the values of the independent variables (Wold and Jureen, 1953; Snedecor, 1965).

Linearity can be achieved by mathematical transformations. However, two additional problems arise with transformations. First, biological significance becomes increasingly difficult to interpret in a transformed system. Secondly, further assumptions should be met prior to transformation. Quoting from Lloyd and Lemmon (1970):

> "Employing a trigonometric relation, such as the sine or cosine to produce linearity in the relationship between site index and aspect, requires two assumptions: 1) The curve of site index over asimuth of aspect does in fact have the shape of a sine or cosine curve; and 2) the high and low points on the curve are built into the procedure. It follows that the assumed "high" and assumed "low" are 180° apart. These assumptions are subject to differences of opinion...(they conclude that) data rectification to provide linearity and adjustments to bring the productivity curve into juxtaposition with the cosine curve...(are) prerequisite to studying relationships between site index and aspect by regression methods."

The second condition for regression analysis, that of random and independent error terms is not met because: (Hodgkins, 1959)

> "...this can be achieved completely only through the replication and randomization procedures of the careful experimental design (Wold and Juréen, 1953)...What does this violation do to the effectiveness of the prediction equation already developed? Cochran and Cox (1957) state that dependence in statistical errors "might completely vitiate tests of significance". Wold and Juréen (1953 emphasize the damage to standard errors of the regression coefficients. They state that these, as calculated, will be in error by undeterminable amounts, and they imply that the calculated standard errors will be too low. This would mean that confidence limits, as calculated for any site index value other than the mean value for the study area, would be unreliable."

The use of stand age as an independent variable will most likely aggrevate the validity problem. Again, according to Hodgkins (1959):

> "It seems quite likely that age may be correlated with some important environmental variables, yet interaction terms are seldom included in prediction equations that use age as a variable. Thus the amount of dependence in the statistical error term, e, in the linear regression equation will be increased. The inclusion of age as a variable also has the unfortunate effect of producing large multiple correlation coefficients, causing many to conclude incorrectly that a large proportion of variation in height has been accounted for by environmental factors."

Another school of thought within this second approach to forest site classification involves using the existing national soil classification system for prediction of growth. Thus an important goal of studies of this type is to use soil series, drainage and textural classes, or horizon depths as determined by the Soil Conservation Service (SCS) to delineate site quality classes. This method has met mixed success. Some authors have reported good correlations between site quality (usally site index) and soil series or soil mapping units (Broadfoot, 1961, 1963; Einspahr and McComb, 1951; Gessel, 1949; Hill, Arnst and Bond, 1948, Stoeckeler, 1960, Westveld, 1952; Westveld, 1933). Others have found soil types or series to be of use only in determining general or approximate quality classes (Arend and Julander, 1948; Dubold, 1935; Youngberg and Scholz, 1949). Carmean (1967) and Lemmon (1970) suggest using woodland suitability groups as the basis of classification. Many authors have used specific characteristics from soil descriptions such as drainage class (Auten, 1945a), depth to tight subsoil (Auten, 1945b), textural classes (Cox, McConnell, and Matthew, 1960; Gessell and Lloyd, 1950; Zahner, 1957), color and mottling (Hansen and Mc Comb, 1958), and parent material (Minckler, 1943).

Soil series alone doesn't appear to be a good indicator of site index primarily because of the variability in characteristics important to tree growth within a series. Fertility differences based on nutrient or water availability are not used in S. C. S. soil classification per se although drainage classes and parent material as indirect indicators are used.

The third type of approach has used laboratory analysis of physical and chemical properties to gain more precise measurements of the factors believed important for tree growth. Most authors have concentrated on physical soil properties. Coile (1952), believed that chemical characteristics could be limiting under certain circumstances usually confounded with adverse soil physical properties. He believed that physical conditions were of overriding importance in site studies; thus researchers were justified in ignoring soil fertility research.

Many authors have concentrated on moisture content or equivalent (Carmean, 1954; Coile and Schumacher, 1953; Copeland, 1956; Corson, Allison, and Chegney, 1929; Gaiser, 1950; McClurkin,1963), or on measuring textural fractions (Coile, 1953; Della-Bianca and Olson, 1961; Hannah, 1968a, 1968b). Yawney and Trimble (1968) and Zinke (1958) measured soil pH. Correlations between site index and available potassium (Beaufait, 1956; Nelson and Beaufait, 1956), total nitrogen (Lunt, 1939; Hicock et al., 1931), and calcium, magnesium and nitrogen (Bowersox and Ward, 1972) have been reported. Stoeckler (1960), while finding some significant correlations between site index and chemical properties, reported much higher correlations for the mechanical properties of the soil.

Additional information is needed to assess accurately the actual availability of nutrients to trees. Tests used in the past may not have been sensitive enough to detect the differences like the agricultural quick tests, or the horizons sampled may have been too shallow to reflect accurately the nutrient status of the soil volume occupied by the tree roots. The body of information available on response to forest fertilization supports the view that fertility differences are important but inadequately measured in the past rather than the view that physical properties are most often limiting and hence of greater importance. Since soil site studies measure the factors whose mode of action is indirect, it is possible that a physical property such as stoniness, dubbed statistically significant in a study, may actually be important because it dilutes the soil volume and hence results in a reduction of available nutrients.

The fourth approach outlined by Mader uses minor vegetation or indicator plants to classify site quality. This site-type or ecological approach has been widely used in Canada and in Europe (Chikishev, 1965; Lafond, 1960; Sisam, 1938). Cajander (1926) reported good success for his system of forest quality types based on ground vegetation in Finland. Distinct ordered differences in growth were found between the site types. Heimberger (1934) and Westveld (1952, 1954) have been among those who have advocated its use in the United States. Hodgkins (1961, 1970) and Silker (1965) have reported on the successful use of indicator plants in the southern United States. Trimble and Weitzman (1956) found ground vegetation could be used to seperate three quality classes for upland oaks.

However, the complexity and diversity of temperate hardwood forests has long been a stumbling block to the acceptance and use of indicator plants. Major objections are that the plant community relationships change more rapidly than soil and topographic conditions, and the minor vegetation must be interpreted in terms of forest growth (Mader, 1965).

The fifth approach, soil-vegetation surveys, involves "broad scale forest land classification from aerial photos and ground control work using combinations of land features, stand characteristics, and soils information "(Mader, 1965). The Ontario classification system by Hills (1958) used in Ontario and in the Canadian Maritime Provinces would be an example of this method of site classification. The Ontario system is based on local climate, soil moisture, and soil nutrient regimes. In California (Gardner, 1958; Weislander and Storie, 1952) this method has been used to distinguish broad quality classes, but its accuracy is not yet sufficient for precise distinctions. Recent work in remote sensing and photogrammetry may hold future hope for increasing the accuracy and precision of soil vegetation surveys (Aldrich et al., 1970; Avery, 1966; Committee on Remote Sensing for Agricultural Purposes, 1970, Heller et al., 1964; Poulton et al., 1970; Spurr 1960).

In recent years additional studies have been made which seem to indicate the need for adding a sixth category to Mader's classification of site quality studies. This sixth approach uses multivariate statistical techniques to analyze multiple biological and physical variables and their interactions. These studies break with the past not only in their increased sophistication in analyzing complex data systems, but also in their broadened use of indicators of site quality. Almost all traditional forest site quality studies have used site index as the sole criterion of site quality. An early exception was Lunt and Baltz (1943) who related basal area to site factors.

Nuss and Borden (1968) *argue; "Since site quality investigations have included a wide variety of relatively precise measurements of independent variables, no real reason exists for doubting that the "right" independent variables have been included (in multiple regression analyses).

---

*Because this sixth type of soil site study has not been included in past reviews, these studies will be discussed in somewhat greater detail than those cited previously.

This consideration leads to the belief that a major drawback to site quality investigations could possibly be that site quality is not adequately defined by its measure in terms of site index." They used a linear combination of average sugar maple d.b.h. (diameter at breast height) number of trees per acre, average height of dominant and codominant sugar maple to reflect volume productivity. This composite variable was then related to site variables such as percent slope, bulk density, texture and field capacity, by a technique called canonical correlation. Canonical correlation is the multivariate extension of correlation analysis (see Chapter 4).

La Bastide and Van Goor (1970) used average height at age five for slash pine and at age twenty for Brazilian pine and related this measure of tree growth to soil units, and to soil and foliage nutrient analysis using factor analysis, another multivariate technique (see Chapter 4).

Lewis and Ward (1967) used log (height x basal area) of the dominant and codominant white ash as their measure of site quality. This index of quality, designated Q, was then used in three analysis; soil taxonomic classification, indicator species, and manipulation of environmental variables via multiple regression. Multiple regression with Q as the dependent variable was the most successful of the techniques utilized.

Berglund (1968) used simple product-moment correlation analysis, factor analysis, and multiple regression analysis together to predict sugar maple site productivity on the basis of physical soil properties and topographic characteristics. The original objectives of the study were to use soil taxonomic units to seperate relatively uniform areas of productivity and to develop biologically sound productivity predictor equations. However, due to numerous taxonomic changes by the S.C.S. it was concluded that the soil taxonomic units were too unstable to form the basis for a classification system. Site index, basal area periodic annual increment, and cubic foot periodic annual increment were the measures of site productivity.

Jackson and Gifford (1974) used periodic volume increment of radiata pine as the dependent variable in a multiple regression analysis with mean precipitation, seasonal rainfall distribution, seasonal departures from optimum temperature, effective soil depth, total nitrogen, and available potassium being the independent variables. The study was set up to include the maximum range of New Zealand climates. After the effects of age had been removed, 66 percent of the remaining variation in periodic volume increment was accounted for by the independent variables. Thus, including climatic variables may have real utility in extending the area of coverage of such studies.

Jeglum (1974) used another multivariate technique, principal components analysis, to analyze black spruce growth and vegetational diversity. The environmental variables were divided into two classes; moisture - aeration related measures including depth to water and ferric, ferrous, and sulfate analysis, and nutrient related measures including pH, depth of peat, conductivity, and nutrient concentrations. While site index was used to estimate site quality for black spruce, the combination of minor vegetation and the results of physical and chemical analysis in a little used statistical technique make this paper quite unusual.

While only a few studies using multivariate techniques have been published in forestry research, a number of studies utilizing MSA have been published in forestry related fields such as soils, hydrology, taxonomy, and photogrammetry. This seems to augur well for these methods of analysis. The study presented in this thesis belongs to this sixth approach to site classification. It uses a series of multivariate methods to analyze species density and diversity.

THE STUDY AREA

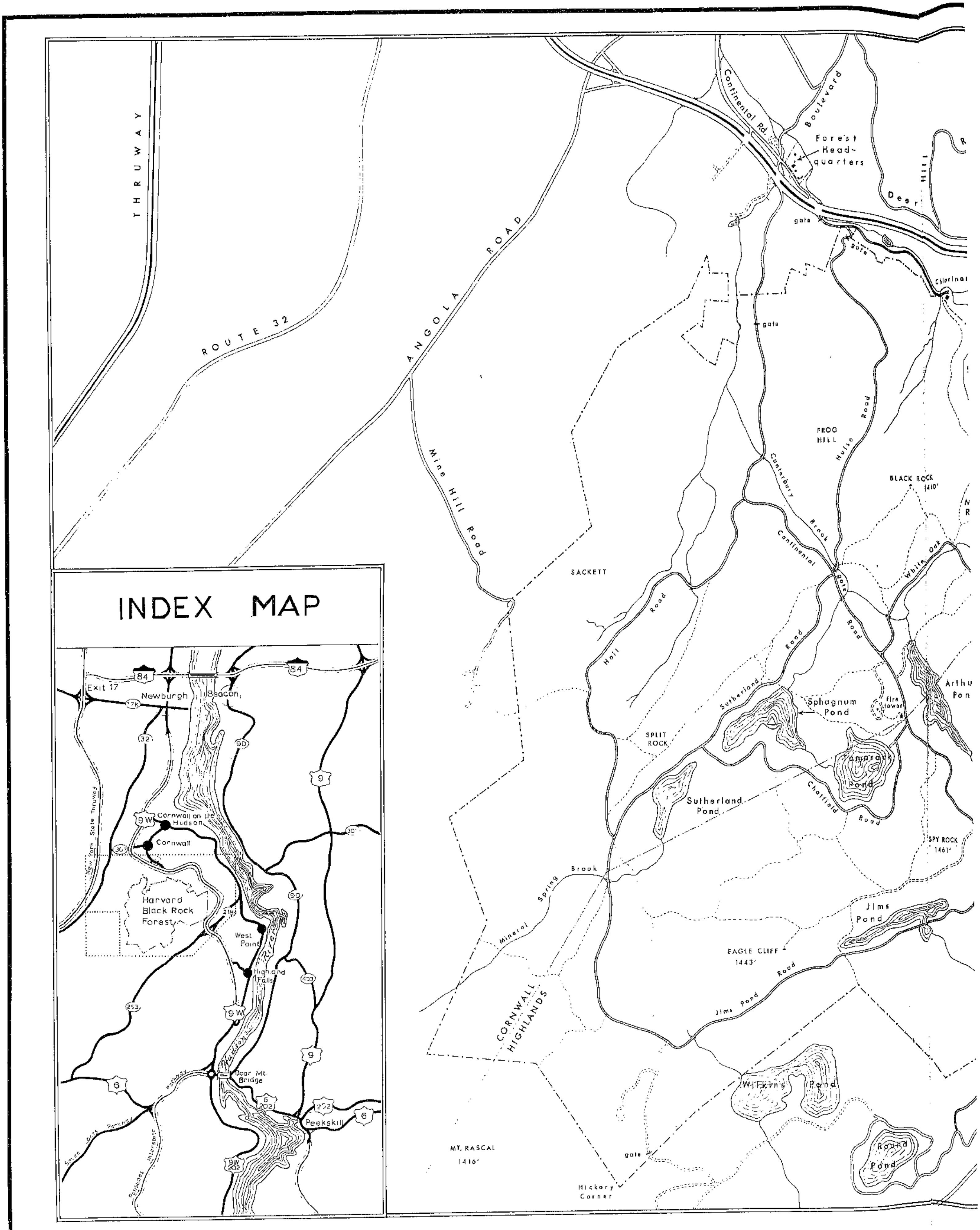
General

The Harvard Black Rock Forest is a 3600 acre tract near Cornwall, New York. (see Figure 3-1). Owned by Harvard University, the Forest has had ongoing research projects in silviculture and timber management for the past forty years. The study area is located in Compartment XXV of the Forest, which is to the west-southwest of Tamarack Pond. This area is bounded by the Chatfield Road, and the Chatfield and Secor trails. (see Figure 3-2). The lowest elevations, about 1260 feet, are found along the road. The elevation increases to 1410 feet and then levels off close to the intersection of the two trails.

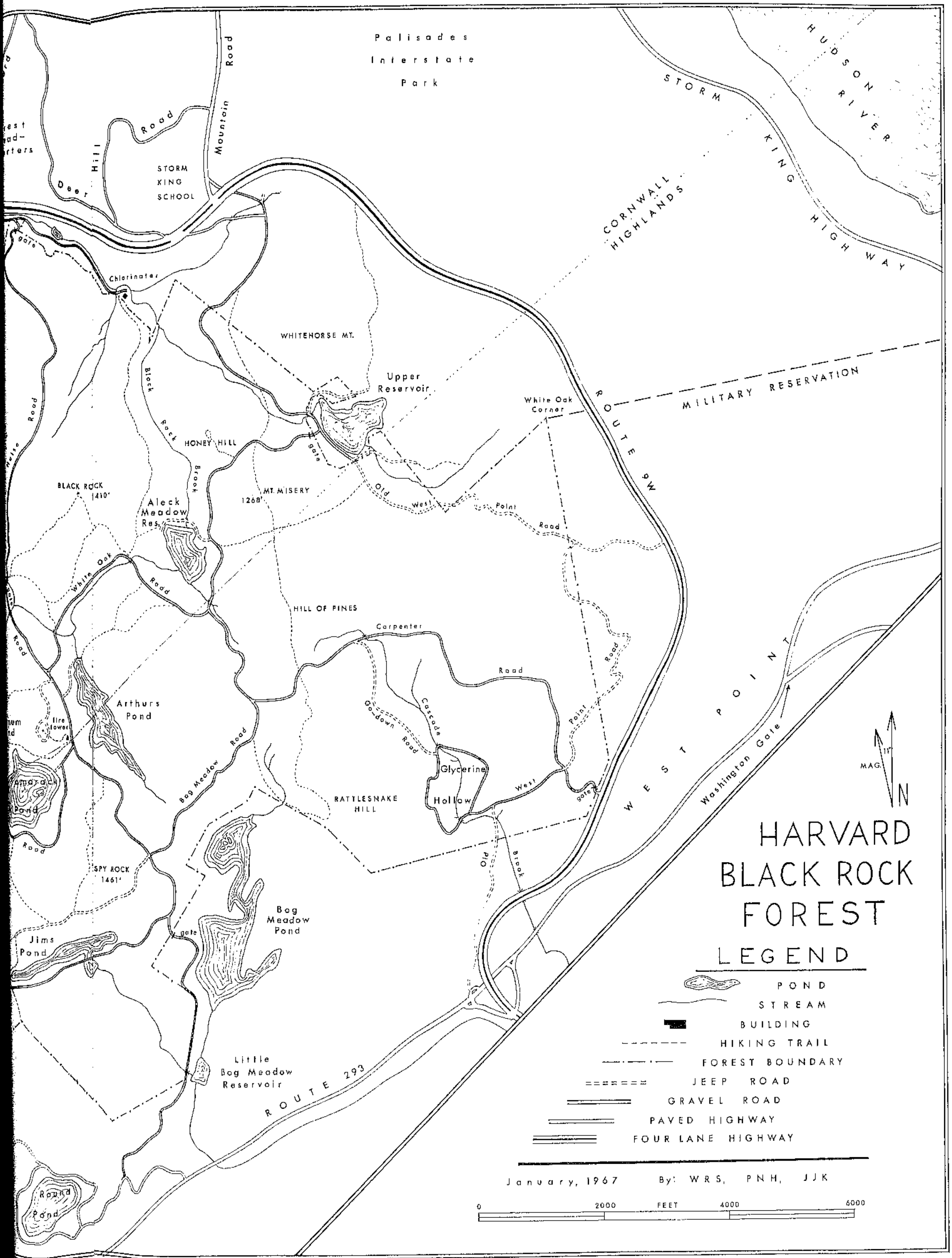
The twenty-seven acre study area was clearcut, including all trees and underbrush, during the winters of 1953-1954 and 1954-1955. The ages of the major stand at the time of cutting were 40 years in the eastern section of the area and 82 years in the western section. Diameters in the stand ranged from three to five inches at the top of the hill to ten inches along the road and in the gully in the eastern section of the area. An average of 18.3 cords per acre was removed during the logging.

After the cut, part of the area was sprayed with a solution of 2,4,5,-T in kerosene to kill as many sprouts and existing root systems as possible. However, sprouts were very

Figure 3-1. Map of the Harvard Black Rock Forest.



THRUWAY
ROUTE 32
ANGOLA ROAD
Mine Hill Road
Continental Rd.
Boulevard
Forest Head-quarters
Deer Hill
gate
Chlorinat
FROG HILL
Hulse Road
Canterbury Brook
Continental Road
BLACK ROCK 1410'
White Oak
SACKETT
Hall Road
Sutherland Road
Sphagnum Pond
fire tower
Arthur Pond
SPLIT ROCK
Tamarack Pond
Sutherland Pond
Chatfield Road
SPY ROCK 1461'
Mineral Spring Brook
Jims Pond
EAGLE CLIFF 1443'
Jims Pond Road
CORNWALL HIGHLANDS
Wilkins Pond
Round Pond
MT. RASCAL 1416'
Hickory Corner
INDEX MAP
84
Exit 17
Newburgh
Beacon
17K
32
90
9
9W
Cornwall on the Hudson
Cornwall
307
New York State Thruway
Harvard Black Rock Forest
218
West Point
Highland Falls
Hudson River
293
403
6
Bear Mt. Bridge
202
Peekskill
Palisades Interstate



HARVARD
BLACK ROCK
FOREST
LEGEND
POND
STREAM
BUILDING
HIKING TRAIL
FOREST BOUNDARY
JEEP ROAD
GRAVEL ROAD
PAVED HIGHWAY
FOUR LANE HIGHWAY
January, 1967
By: WRS, PNH, JJK
0
2000
FEET
4000
6000
MAG.
N
Palisades Interstate Park
HUDSON RIVER
STORM KING HIGHWAY
CORNWALL HIGHLANDS
Mountain Road
Deer Hill Road
STORM KING SCHOOL
Chlorinator
gate
WHITEHORSE MT.
Upper Reservoir
MILITARY RESERVATION
White Oak Corner
ROUTE 9W
Black Rock Brook
HONEY HILL
BLACK ROCK 1410'
MT. MISERY 1268'
Aleck Meadow Res.
Old West Point Road
White Oak Road
HILL OF PINES
Carpenter Road
Arthurs Pond
fire tower
Bog Meadow Road
Cascade
Goodown Road
Glycerine Hollow
West Point Road
RATTLESNAKE HILL
WEST POINT
Washington Gate
SPY ROCK 1461'
Bog Meadow Pond
Jims Pond
Old Brook
Little Bog Meadow Reservoir
ROUTE 293
Round Pond

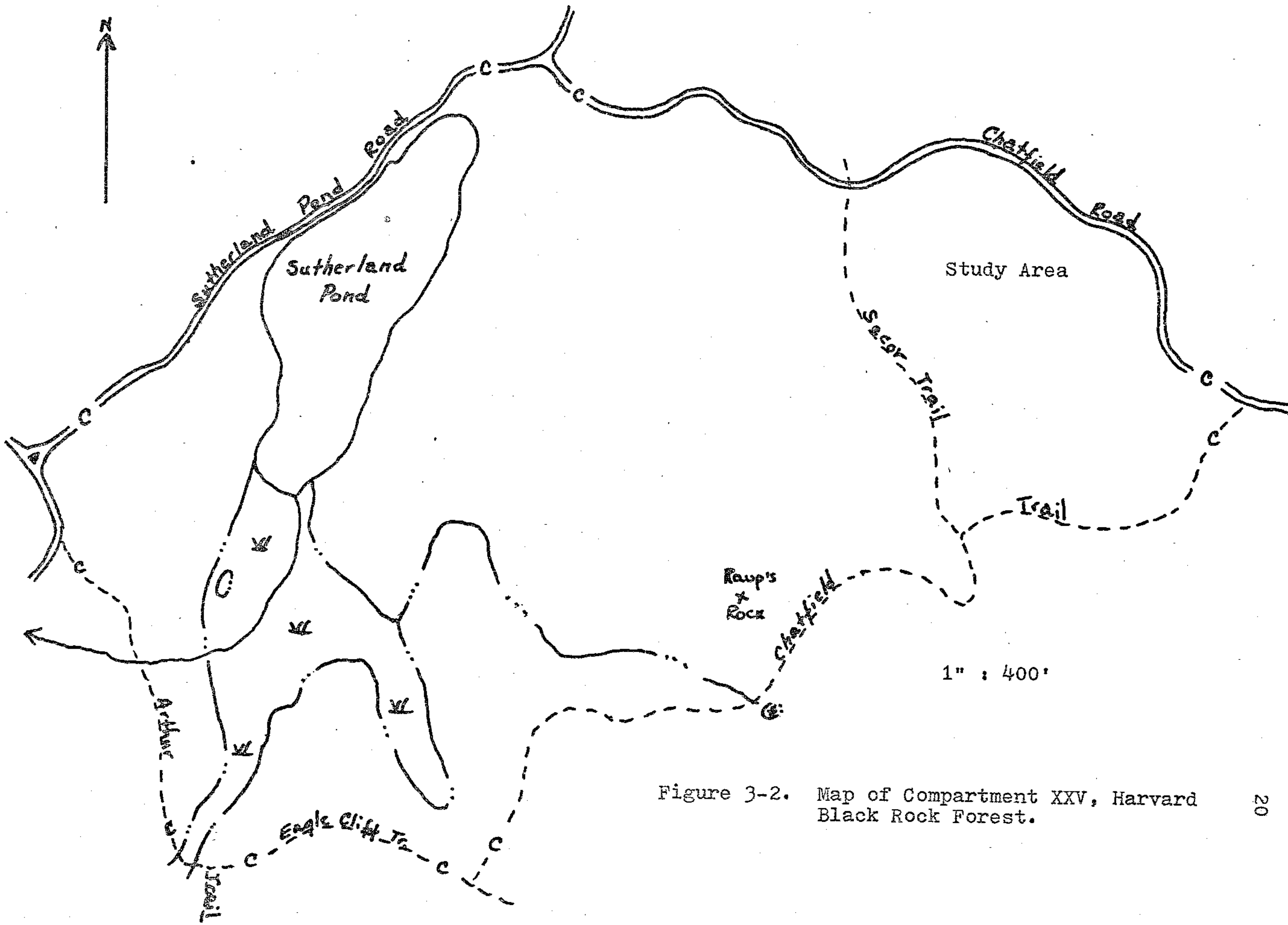


Figure 3-2. Map of Compartment XXV, Harvard Black Rock Forest.

vigorous even the first year after spraying, and they soon took over the area. Approximately 95 percent of the regeneration present in 1973 was of sprout origin.

The existing stand is mostly mixed oak. Northern red oak[1], scarlet oak, chestnut oak, white oak and red maple predominate; black and grey birch, striped maple, American chestnut, speckled alder, flowering dogwood and scrub oak are also present in smaller numbers. The major minor vegetation included blueberry, huckleberry, laurel and speckled alder.

Geology[2]

The Harvard Black Rock Forest lies on the north side of the Hudson Highlands, an area of pre-cambrian crystalline rocks. The area can be seperated into two physiographic divisions, the Highland section and the Northern Slope section. The Highland section is an area of swamps and artificial ponds seperated by low hills and rolling uplands. The relief is seldom over 200 feet with the elevation varying from about 1225 to 1463 feet. The Northern Slope section lies at altitudes of 420 to 1225 feet, and is characterized by steep, northwest slopes and deep ravines. The study area is in the Highland section of the forest.

Small rounded bedrock outcrops, ten to twenty feet high

---

[1] The scientific names according to Gleason and Cronquist, 1963 and Little, 1953 are given in Appendix A.

[2] The following discussion is adopted from Denny, 1938.

are very abundant in the Highland area. Exfoliation is removing thin plates of rock from the exposed outcrops. The surface of the rock ledges is very rough, with large feldspar crystals projecting above the surface. Weathered pits and solution grooves in the rocks are common. Scattered boulders lie on the soil and bedrock surfaces.

Till and warp are the only glacial deposits found within the Forest. The upper one to four feet of the weathered zone (the warp) is loose and contains more boulders than the underlying till. It is a jumbled mass of pebbles, rocks, and boulders in a yellowish-brown matrix of sand and clay. Largely derived from the underlying till, the warp is a frost-heaved layer produced when frost action more intense than that of today was prevalent. The original till over the glaciated bedrock has been largely removed by periglacial solifluction; this is the movement of the surface layer more or less as a unit down slope under climatic conditions found adjacent to a glacier (Flint, 1971). Wherever the glacial deposits are less than two feet thick, the weathered warp rests directly upon bedrock.

The warp on the study area is 60 to 70 percent crystalline rock (granite, gneiss, schist, etc.), 20 to 30 percent sandstone, with slate making up the remainder (Denny, 1938). The percentages of slate and sandstone usually increase with depth while the percent of crystalline rocks decreases. Some bedrock fragments, the Storm King granite, are also included.

Soil

The major soil series present on the study area is Hollis, a member of the loamy, mixed, mesic family of Entic Lithic Haplorthods. The Soil Conservation Service (S.C.S.) mapping unit for this area is 071-BC, Hollis extremely rocky soils, 3 to 15 percent slope or Rock outcrop-Hollis complex, gently sloping. The Hollis soil is a shallow, excessively drained, medium to moderately coarse textured soil. It is formed in glacial till derived from crystalline rock, mostly schist and gneiss. A representative profile of Hollis gravelly loam that occurs in a wooded area in the mapping unit and the range in characteristics for the series are included below. The Hollis series has bedrock ten to twenty inches below the surface. In an associated soil, the Charlton, depth to bedrock is greater than fifty inches. No currently named series has the characteristics of the Hollis and Charlton series[3/], with depth to bedrock between twenty and fifty inches.

[3/] Personal communication, Middleton, New York office, Soil Conservation Service, July, 1974.

Hollis Series[4/]

Location: Town of Tuxedo, 200 yards south of New York 210, ¼ of mile east of 1-87, on New York 210.

02 3 to 0 inches, dark reddish brown (5YR 3/2) loose decomposed roots, sticks, and leaves; strongly acid; clear smooth boundary.

A1 0 to 4 inches, dark brown (10YR 4/3) gravelly loam; moderate medium granular structure; very friable; many roots; few pores; 15 percent gravel; strongly acid; clear wavy boundary.

B2 4 to 14 inches, strong brown (7.5YR 5/6) gravelly loam; moderate medium subangular blocky structure; fiable; many roots; common pores; 20 percent gravel; strongly acid; abrupt smooth boundary.

R 14 inches, hard gray granitic bedrock.

Range in Characteristics:

Solum thickness and depth to bedrock ranges from 10 to 20 inches. Coarse fragments range from 5 to 25 percent. Textures range from sandy loam through loam. Reaction ranges from strongly acid to very strongly acid. The A1 horizon has

---

4/ Xeroxed copy of Hollis Series, an eight page unpublished draft report obtained from the Middletown, New York office, Soil Conservation Service, July, 1974.

color of 10YR hue, 2 through 4 value, and 2 or 3 chroma. The B horizon has color of 10YR to 7.5YR hue, 4 or 5 values, and 5 through 8 chroma. Consistance ranges from friable to very friable. Exposed rock outcrop ranges from rocky through rock-outcrop, which ranges in percentage from 2 through 90 percent.

Climate

The nearest climatological station is at the West Point Military Academy, approximately three miles from the study area. Average length of the growing season is 190 days with the number of frost free days ranging from 171 to 219 days (20 years of record). Mean annual precipitation is 46.7 inches with 25-30 inches falling during the growing season (29 years of record - United States Department of Agriculture, 1973).

Stand and Site Measurements

During the summer of 1973, a hundred 0.01 acre plots were established on the area. The plots were put in on a systematic grid on east-west lines. All vegetation 1.5" d.b.h. and larger was tallied by species. In 1974 fifty of the original plots were revisited. This time various edaphic and topographic variables were measured on the plots.

A soil pit was dug at the center of each plot until bedrock was reached or until 42 inches, whichever came first. In only two cases was digging stopped at the 42 inch level. Soil color at the top and the bottom of the pit was measured using a Munsell Color Chart. Percent rock by volume in the pit and percent surface rock on the plot were measured by ocular estimates. Aspect, percent slope, and ground cover were also tallied. Percent slope was determined using an Abney hand level. Ground cover on each plot was recorded in order of decreasing frequency.

From detailed maps of the area, slope position, distance to the perimeter of the cut, and age of the previous stand when cut were determined for each plot. Slope position classes were deliniated by 20-foot contour lines with class 1 being inside the 1400 foot contour line and class 8 being between 1260 and 1280 feet.

Summary of Measurements

Northern red oak and red maple were the most common species on the study area. Together they accounted for over half of the stems that were tallied. Scarlet, chestnut, and white oak were also present in significant numbers. About one tenth of the stems were black and grey birch, striped maple, American chestnut, flowering dogwood, scrub oak, and speckled alder. Most of these minor species, while

Table 3-1. Number of stems per acre by species and diameter class.

	Diameter Class in Inches				
Species	1.5-2.4	2.5-3.4	3.5-4.4	4.5+	Total
Northern red oak	251	94	17	3	365
Red maple	293	48	6	0	347
Scarlet oak	118	49	14	1	182
Chestnut oak	111	26	10	3	150
White oak	90	20	5	2	117
Black birch	58	11	2	0	71
Grey birch	46	4	0	0	50
Other 1/	30	3	0	0	33
Total	1008	255	54	9	1326

1/ Other includes striped maple, American chestnut, flowering dogwood, scrub oak and speckled alder.

Table 3-2. Per acre basal area distribution by species.

Species	Basal Area (Square feet)	Number of Plots Represented
Northern red oak	12.2	81
Red maple	7.2	67
Scarlet oak	6.2	48
Chestnut oak	4.8	43
White oak	3.5	50
Black birch	1.8	26
Grey birch	1.0	22
Other[1/]	.8	15

1/ Other includes striped maple, American chestnut, flowering dogwood, scrub oak, and speckled alder.

Table 3-3. Per acre basal area distribution by ten square foot classes.

Basal Area Classes (square feet)	Number of Plots Represented
0 - 10	8
10 - 20	8
20 - 30	14
30 - 40	30
40 - 50	17
50 - 60	8
60 - 70	9
70 - 80	3
80 - 90	3

unimportant in terms of the total area were important on individual plots.

The distribution of stems by diameter class showed different patterns by species. Red maple was the species with the greatest number of stems in the two inch diameter class (1.5-2.4 inches). However, in the four inch class red maple was fourth in numbers of stems. Scarlet and chestnut oak proportionately had the greatest representation in the four and the five inch and greater classes. These results are presented in Table 3-1. Total basal area by species and the number of plots on which each species was found is presented in Table 3-2. The distribution of basal area classes follows an approximate normal distribution. The thirty to forty square feet per acre class represented the mode, having 30 out of the original 100 plots. These results are shown in Table 3-3. Total basal area and average diameter at breast height ($\overline{dbh}$) were calculated for each plot.

Of the 50 plots where ground cover was tallied, only seven had tree seedlings; six of these were oak, the seventh was red maple. Forty-five plots had blueberry or huckleberry present. Mountain laurel, speckled alder, and wood fern were each represented on 10-15 plots. Other ground cover that was important on individual plots included sweet fern, honeysuckle, scrub oak, and sarsparilla.

Table 3-4. Summary of physical plot measurements.

Variable	Average	Range
Soil depth	17.1 inches	0-43 inches
Aspect	320-20° *	--
Percent slope	17.7%	0-58%
Slope position class	3.7	1-8
Distance to perimeter	176.2 feet	15-420 feet
Age of previous stand	----	40 or 82 years
Percent surface rock	19.7%	3-80%
Percent rock in pit	20.4%	0-60%

* Most common.

Soil depth to bedrock averaged 17.1 inches with a range from 0 inches where bedrock was exposed to 43 inches which occured in a level area close to the top of the slope. Percent rock in the soil pit ranged from 0%, which was rare, to 60%, with an average of about 20%. The color at the bottom of the pit was coded so that organic layers and very bright colors received the highest numbers. Drab colors indicating poor drainage received the lowest numbers. The study area, taken as a whole, faced north, but the aspect of the individual plots was quite variable. Aspect was coded with southeast considered the most favorable, and given the highest number. The physical site measurements are summarized in Table 3-4.

After the calculations of species basal area and $\overline{dbh}$ for each plot and the coding and transformation of the variables was completed, the data were punched onto computer cards. Different methods of recording the species information were tried, resulting in three data cards for each plot.

INTRODUCTION TO MULTIVARIATE STATISTICAL TECHNIQUES

General

Recent use of multivariate statistical methods in the analysis of biological data has met with mixed response. These techniques are undoubtedly attractive and useful alternatives for the analysis of complex data sets. However, interpretation of the basic models is often confused by the brief explanations in the literature, and the techniques depend heavily on the use of complicated mathematical treatments. If there is puzzlement over what a particular model is actually doing, lack of confidence in the conclusions may result. To add to the problem, multivariate statistics has long been plagued with nomenclature problems. Some techniques are known by an assortment of names. Other techniques have incorrectly been viewed as being so basically similiar as to result in interchangeable nomenclature.

To aid in an understanding of what some multivariate methods actually do and what types of information can be obtained from their use, a brief explanation will be presented for each of the techniques employed in this analysis. Instead of giving the formal mathematical models, a descriptive and graphical approach will be used to enable, hopefully, an intuitive understanding of each of the methods.

Graphical Representations of Probability Functions

Estimation of the variability or dispersion in a population is often very important in statistical analysis. Plotting the relative frequency or probability against the variable values for a population results in a graphical representation of the probability function for the population. Examination of the resulting curve will reveal the mean and the approximate variance of the population. This type of representation can be a useful aid in understanding what many statistical techniques are actually doing with reference to the original population.

The probability distribution for the diameters at breast height for an even-aged red maple stand could look like the familiar bell-shaped curve for a normal distribution seen in Figure 4-1. Plotting the probability distribution for another variable, say tracheid length, measured on the same individuals in the population, would also result in a bell-shaped curve. Graphing the joint probability distribution for the two variables would result in a three dimensional model. Assuming the variables are uncorrelated with equal variances, the model will look like the symmetric hill pictured in Figure 4-2.

However, to assume equal variances is too restrictive for most uses. Relaxing this assumption will change the

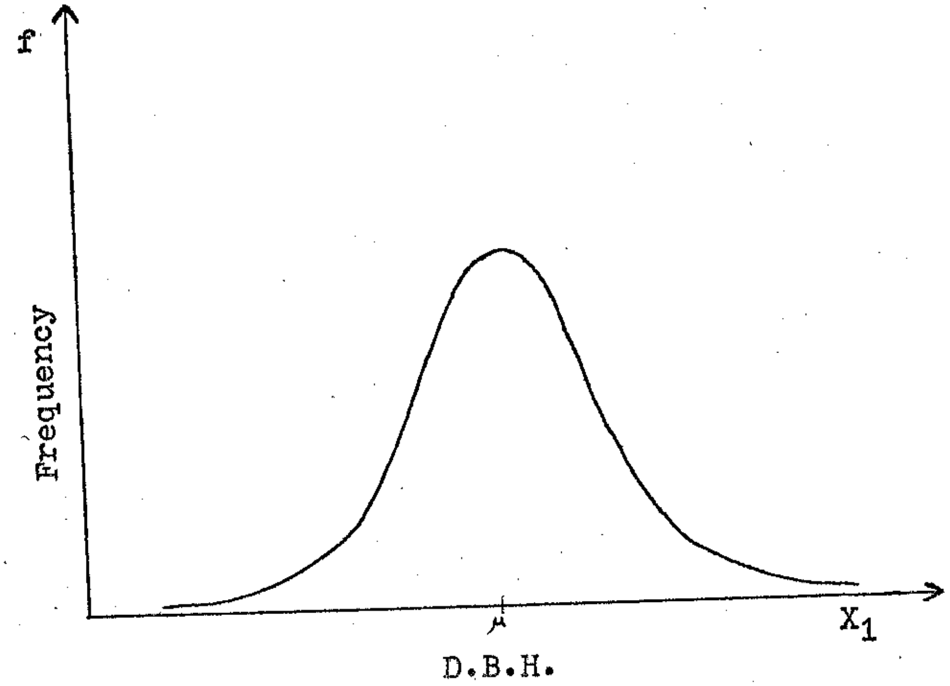


Figure 4-1. Probability distribution for the diameters in an even-aged stand.

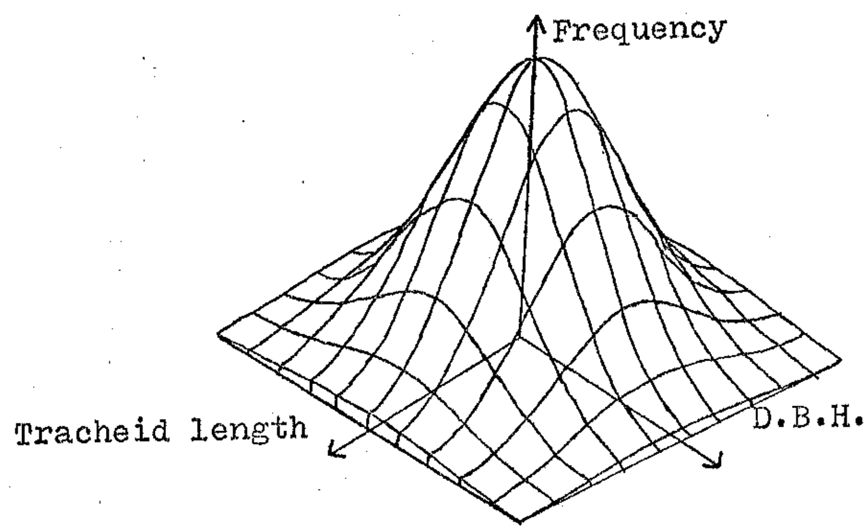


Figure 4-2. Joint probability distribution for red maple diameter and tracheid length in an even-aged stand. (Adapted from Seal, 1966)

shape of the resulting model so that the hill is no longer symmetric about a point. The model will be narrower in the direction of the variable with the smaller variance.

The orientation of this hill will also be changed if we relax the restriction that the variables be uncorrelated. For example, plotting the joint probability function for diameter and height measured on our original red maple population will result in a canted or rotated hill (see Figure 4-3). The magnitude and direction of this rotation will depend on the correlation between the variables.

Taking one of these models, say that pictured in Figure 4-3, we can treat it as a cartographer would treat an actual hill for representation in two dimensions. Starting at the top of the hill, directly above ($\overline{X}_1$, $\overline{X}_2$), we can draw a set of closed contour lines on the X_1, X_2 plane, such that 25% of the population values (the frequency distribution) lies inside the first line, 50% inside the second line and so on. The result is a set of nested concentric ellipses which can be plotted resulting in a diagram like that pictured in Figure 4-4. Additional variables can be handled by the same chain of reasoning resulting in multidimensional ellipsoids rather than ellipses.

To summarize:

I. When the variables are uncorrelated the shape of the

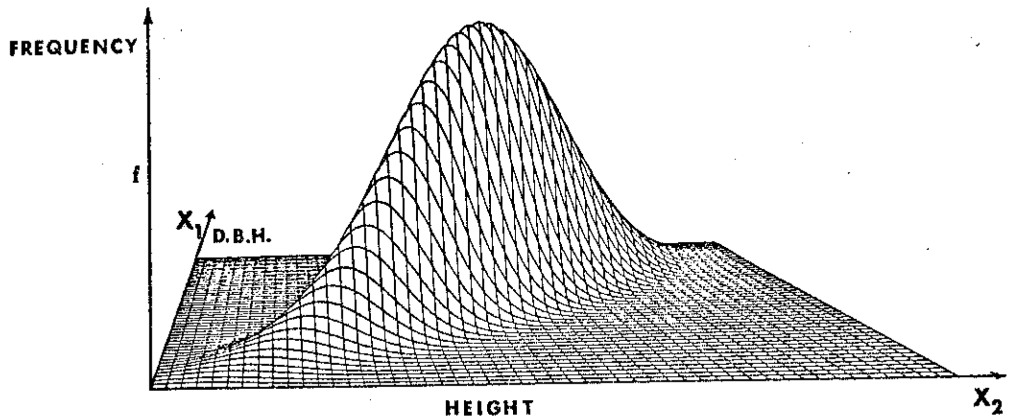


Figure 4-3. Joint probabilty distribution for red maple diameter and height in an even-aged stand. (Adapted from Sokal and Rohlf,1969)

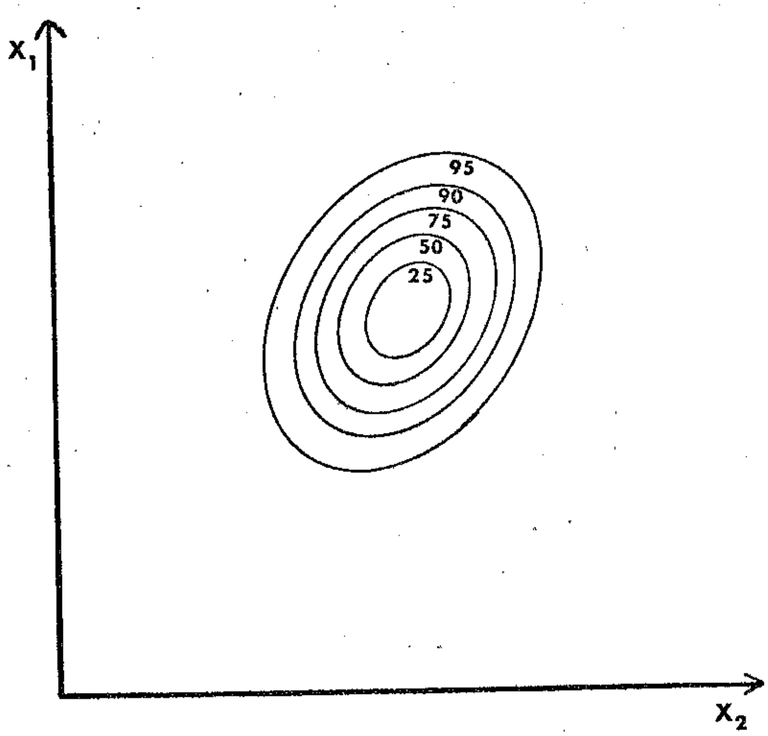


Figure 4-4. Two dimensional representation of a two variable joint probability distribution.

probability or frequency contours is determined by the relative values of the variances associated with each variable. Equal variances will result in circles, spheres, or hyperspheres depending on the number of variables. Unequal variances will result in ellipses, ellipsoids, or hyperellipsoids.

II. The orientation and the width of the probability ellipses is determined by the correlation or covariance between the variables.

Discriminant Analysis

The purpose of discriminant analysis, as the name implies is to discriminate or distinguish between two or more groups or populations. This is done by deriving a set of composite linear functions which are based on the original variables measured. Discriminant analysis was originally used to determine the membership of an individual with reference to certain populations of interest to which the individual could belong. In most current research, discriminant analysis has been used "to determine the extent and manner in which two or more previously defined groups of subjects may be differentiated by a set of dependent variables operating together" (Veldman, 1967).

In terms of graphical representation, the method reduces the number of dimensions required to represent the groups to the number of groups minus one or to the number of original variables, whichever is smaller. Thus data collected on 25 variables measured over the individuals in three groups can be presented using two reference axes. The axes are linear combinations of the variables weighted to give maximum group seperation.

To extend the example discussed in the previous section, let us assume we have measured four variables, d.b.h., height, tracheid length, and live crown ratio on individuals from two red maple stands, age 20 and age 40. Discriminant analysis will determine a function for each group composed of discriminant weights (a_1, a_2, a_3, a_4) which when multiplied by the individual's variable values will give a discriminant score. These scores can then be plotted on a line and differences between the groups can then be easily seen. When the scores are plotted against frequency, bell shaped curves like those seen in Figure 4-5 result.

To relate this discriminant axis back to the original populations, let us simplify our example to two variables measured on our two groups. Plotting an increasing series of probability ellipses will eventually result in ellipses from the group intersecting. Constructing a line perpendicular to the line marking intersection will result in a line

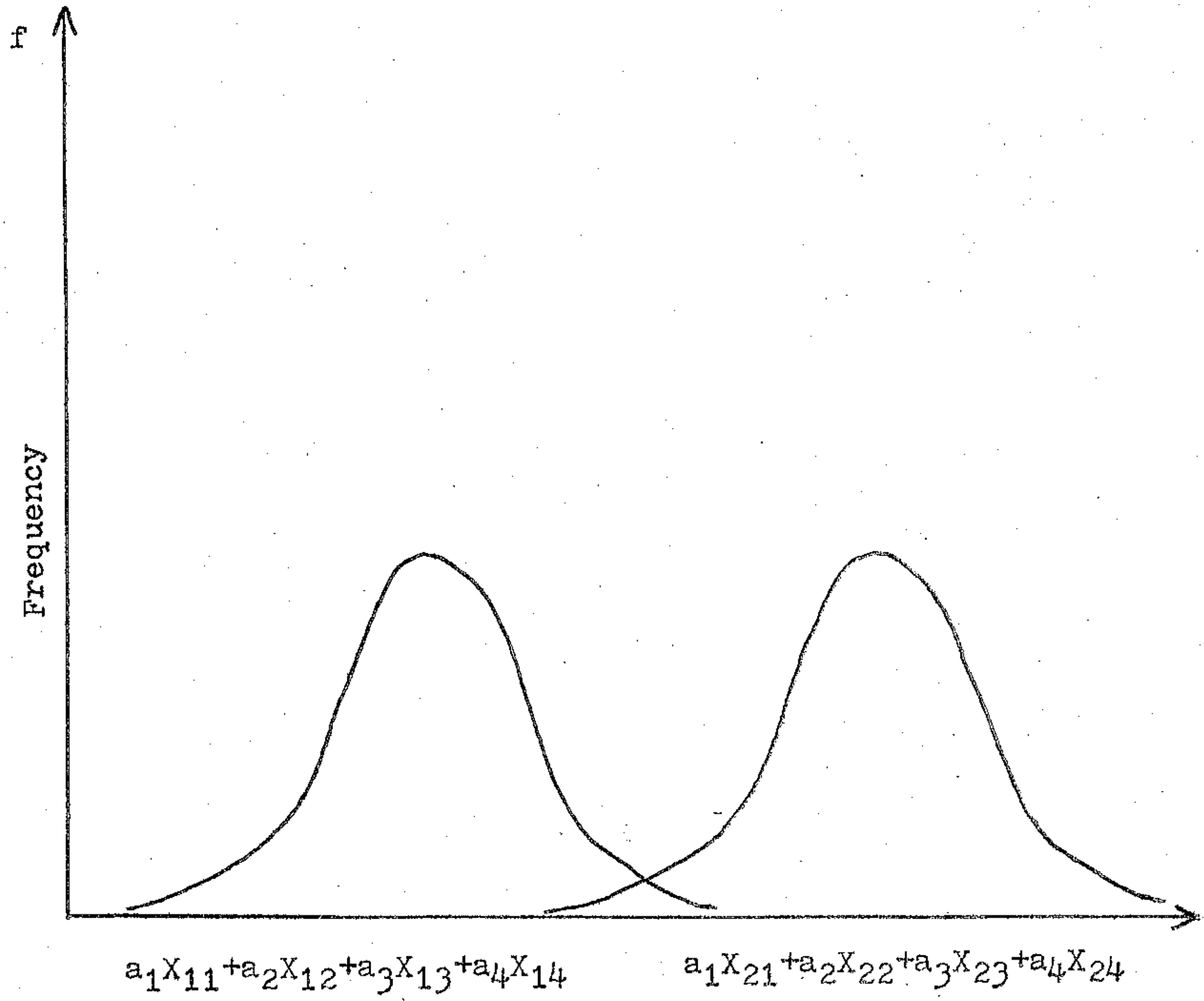


Figure 4-5. Probability distribution of discriminant scores plotted over two populations.

of maximum discrimination. This, or a line parallel to it, is the line that the individuals are projected to by the discriminant function (see Figure 4-6). Representing additional variables is handled by extending the concept to higher dimensions.

Changes in scale in which the variables are measured will not alter the final result but will change the values of the discriminant weights. Thus a variable with a low discriminant weight should not be dismissed as unimportant without performing other statistical tests of significance. If the variances of the responses are nearly equal, the weights will give the relative importance of the contribution of each variable to the discriminant function.

Cluster Analysis

Cluster analysis determines natural groupings or associations of individuals based on their similiarities. Thus cluster analysis is basically the reverse of discriminant analysis. The two techniques are often used in tandem to check the results from one method or to gain additional information to be used in further analysis.

A cluster is "a set of objects characterized by the properties of isolation and coherence" (Jardine and Sibson, 1971). The variable measurements or values for an individual

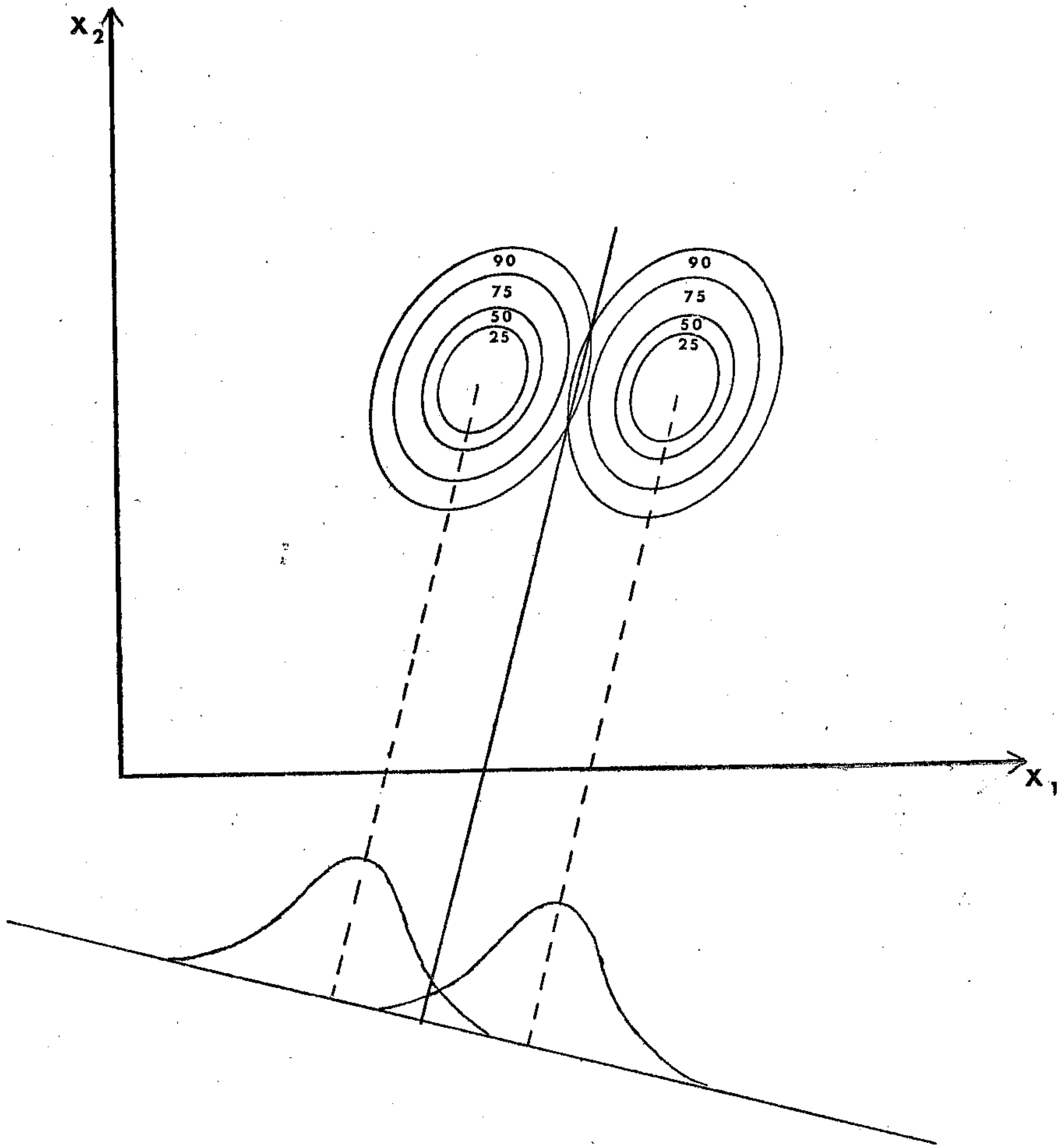


Figure 4-6. Representation of the discriminant scores plotted over two groups with reference to their probability ellipses.

determine its coordinates and hence its position in multi-dimensional space. Cluster analysis can be thought of as a technique which sweeps through this space searching for natural groupings. The end result has minimized within group variance while maximizing between group variance. It is important to recognize then, that changes in the scale in which the variables are measured will change the variance associated with that variable. Variables with large variance will be weighted most heavily in determining the end groupings. Standardization may be advantageous, expecially in interpreting the results. Many computer programs allow differential weighting of the variables to be used. This can be of considerable value if the reasons for assigning the weights are strong; if not, the value of cluster analysis can be seriously impaired.

There are two basic techniques which can be used to determine the cluster. The first technique begins with as many groups as there are individuals. The individuals and later the resulting groups having the greatest similarities are merged in a step-wise process. The process ends when either a predetermined number of groups is reached or when a pre-set probability cut off level is exceeded. There are many versions of this lumping or aggregation technique which differ basically in their determination of "similarity".

The other basic method is called splitting or disaggregation. In this method the data are originally viewed as comprising one group, then split into additional groups so as to minimize the within group variance.

Cluster analysis is one of the methods discussed in this chapter which does not require a priori grouping of individuals or of variables. For this reason cluster analysis can be used as a rational check of previously defined groups such as are encountered in taxonomic evaluation. Groupings other than those expected could result in further analysis of either the characteristics used for classification or the original basis for lumping or splitting the taxonomic groups.

The results of a cluster analysis are commonly subjected to discriminant analysis. Discriminant analysis used in this context provides information on the contribution of each variable to the grouping process. Conversely the groups used in discriminant analysis can be analyzed by cluster analysis to determine further their natural coherence.

Multiple Regression Analysis

The basic purpose of multiple regression analysis is to obtain an efficient equation relating a dependent variable to a group of independent variables. The most common use of multiple regression in forestry research has been the development of predictor equations for site index or volume. Because of its widespread use, most foresters are familiar with regression analysis. However, it will be briefly discussed here to aid in understanding the multivariate extension on the concept, as well as for comparison purposes.

Multiple regression uses the method of least squares to fit a straight or curved line to the sample data. The measurements on an individual or plot can be thought of as representing its coordinates in n-dimensional space. The method of least squares fits a line through this space to minimize the squared distances from the points to the line. The line (assuming a straight line for this case) is specified by an equation in the form $Y = B_o + B_1 X_1 + B_2 X_2 + \ldots B_n X_n$ where Y is the dependent variable, the X_i's are the independent variables and the B_i's are the regression coefficients.

A regression line can be thought of as a moving average (Freese, 1967). It gives an average of the dependent variable associated with specific values of the independent variables. Before the regression line was fitted, Y had a certain amount of variation about its mean. Fitting the line is, in effect, an attempt to explain part of this variation by the association of the independent variables with the dependent variable. Testing the correlation of Y with the association of X's in the regression equation asks whether the variation in Y explained by the fitted line is significantly greater than the variation left unexplained. It does not show that the fitted hyperplane is the best to describe the data. Neither does it show that we have found the true mathematical relationship between Y and the X_i's. "There is a dangerous tendency to ascribe more meaning to a fitted regression than is warranted" (Freese, 1967).

The method of least squares is used in the analysis of data both from planned experiments and from unplanned situations. However, the word "regression" is most often used to describe analysis of unplanned data. It is the tacit assumption that the requirements for the validity of least squares analysis are satisfied for unplanned data that can cause trouble (Box, 1966). The violation of these assumptions has been discussed in Chapter II and will not be duplicated here.

However, it might well be asked what we want regression analysis to do for us. If we want to predict Y in the future from passive observations of the X_i's, regression can be a powerful tool. Multiple regression analysis has been used to develop quite accurate predictor equations for both stand volume and site index. These equations will be most useful if their future use is to be in the population from which the original data were collected. Care should be used in extrapolating beyond the measured values.

On the other hand if our purpose is to discover how changes in X_i's will affect Y with the intention of actually modifying the system or to try to gain biological understanding of the relationships involved, regression analysis should not be used.

Canonical Correlation Analysis

Canonical correlation is the multivariate extension of multiple correlation analysis. Its objective is to study the relationships between two sets of variables, a set of p independent variables and a set of q dependent variables. Canonical analysis forms a series of new pairs of variables called canonical variates or canonical variables. The canonical variable V_1 is formed from a linear combination of the dependent variables; U_1 from a linear combination of the independent variables. V_1 and U_1 are determined so that the correlation between them is maximal. The next pair of canonical variates, V_2 and U_2, has maximum correlation between them and is uncorrelated with the original pair. This is continued until there are as many pairs of canonical variables as there were variables in smallest original variable set, i.e. p or q.

All the correlation between the sets of the original variables is channeled through the canonical correlations between the new variable pairs (Morrison, 1967). Through the analysis of the correlations between the canonical variate pairs, the dependence between the original variable sets can be concisely described. Anderson (1958) views this process as finding a new coordinate system in the space of each set of the original variables such that the new coordinates display unambiguously the system of correlation. The linear combinations of the variables that form V_1 and U_1 are the first coordinates in the new systems. Similiarly, the linear combinations that form V_2 and U_2 are the second coordinates. The process is continued until the two new coordinate systems are completely specified.

Another method of looking at canonical correlation requires that the original responses be considered as representing coordinates in one universe. Then if the canonical variates are specified to have unit variance, the correlation between each pair can be interpreted as the cosine of the angle between them when they are represented in or superimposed upon the original response system.

Hotelling (1936) originally developed canonical analysis to study the relationships between two batteries of psycological tests administered to the same subjects. The goal of canonical analysis has been to determine the ways in which two sets of measures are related, and the strengths and nature of the relationships so defined (Veldman, 1967). Anderson (1958), Horst (1961) and Roy (1957) have extended the idea of canonical correlation to three or more sets of variables. However, these extensions of canonical analysis will not be discussed here.

Principal Components Analysis

The next two analytical techniques, principal components and factor analysis, are often grouped together under the general heading of factor analysis. These methods attempt to determine the latent or hidden factors which cause the variation in the measured variable responses. Although principal components analysis (PCA) was not used in this study it is discussed here so that the reader will understand the distinctions between it and factor analysis.

PCA involves the rigid rotation of the response axes of the original coordinate system into an orientation corresponding to the directions of maximum variance in the sample scatter configuration. Orthogonal least squares solution for the best-fitting line leads immediately to the first principal axis; that is the rotated axis which passes through the direction of maximum variance of the sample points in n-dimensional space. It should be recognized that this is the major axis of the population probability ellipsoid. The second principal axis will be the first minor axis through the ellipsoid (see Figure 4-7), and so on.

This rotation results in new uncorrelated variables which are defined by linear combinations of the original variables. The p original correlated variables are transformed into p uncorrelated variables in order of decreasing variance. The total or aggregate variance of the original variable set is equal to the total variance of the new variable set; however, the partitioning of the variance is quite different in the new system. The first new variable will have the greatest variance, often a significant portion of the total. Thus the result of the rotation is to produce a linearly transformed set of variables, which are mutually independent and may be considered seperately. Because of the decreasing order of variance associated with the principal components it is possible that only the first few variables will account for the majority and possibly all of the variance and covariance of the original variables. "The objective of the analysis is thus <u>parsimonious summarization of a mass of</u> observations" (Seal, 1964). This technique can be quite useful in pinpointing which variables or combinations of variables has the greatest variability and hence warrant further study. Morrison (1967) discusses the ulility of PCA as follows:

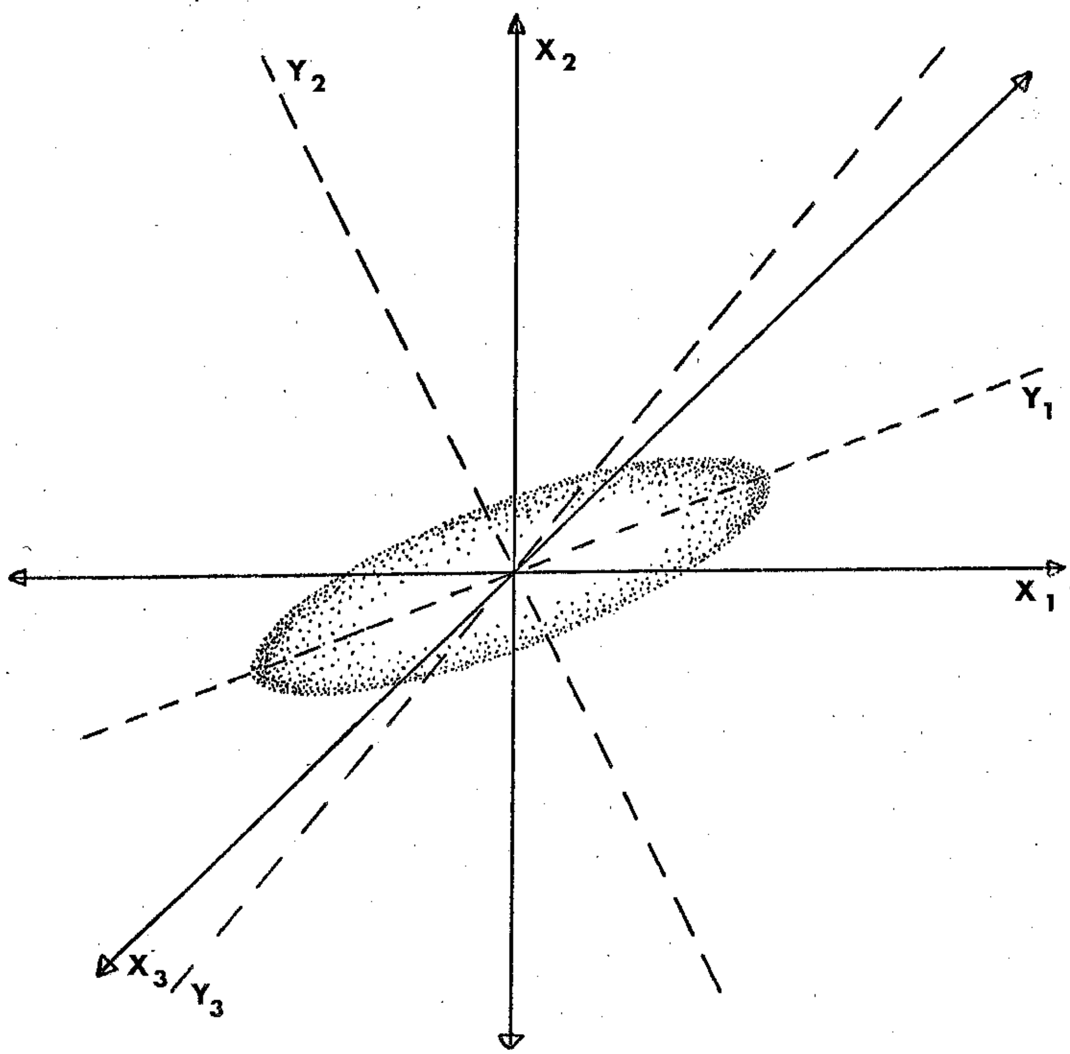


Figure 4-7. Trivariate observations with the three principal axes drawn in. (After Morrison, 1967)

> "In the extreme case of X of rank one[1/] the first principal component would explain all the variation in the multivariate system. In the more usual case of the data matrix of full rank the importance and usefulness of the component would be measured by the proportion of the total variance attributable to it. If 87 percent of the variation in a system of six responses could be accounted for by a simple weighted average of the response values, it would appear that almost all the variation could be expressed along a single continum rather than in six-dimensional space. Not only would this appeal to our sense of parsimony, but the coefficients of the six responses would indicate the relative importance of each original variate in the new derived compound."

Unfortunately, principal components analysis has many limitations which restrict its applicability. The following comments summarized from Seal (1964) should be expecially noted. First, PCA should not be done when the variables are measured on different scales. The biological interpretation of the principal components becomes quite difficult when direct variable measurement on a single scale is not possible. Standardization is one solution but both standardization and changes in variable scales completely change the results. Second, if the off diagonal elements of the correlation matrix are approximately equal the interpretive value of PCA is dubious. Morrison (1967), also points out that no provision can be made for variance components that are attributable only to the unreliability or sampling variation of the original responses.

[1/] X is the original data matrix or array. Rank one means that only one of the original variables is independent.

Factor Analysis

The goal of factor analysis is the reduction of a set of variables to a smaller group of new uncorrelated variables. This technique is used to try to extract underlying common or latent factors from the original variable set. Each of the response variables is represented or defined in terms of a linear composite function composed of the latent unobservable common variables plus an uncorrelated residual component called a latent specific variable. The common factors generate the covariances among the observable responses, while the specific terms contribute only to the variances of their particular responses (Morrison, 1967). The models for PCA and factor analysis are presented in Table 4-1 for comparison purposes.

The orthogonal axes of the common factors once determined may be rotated to new orthogonal or oblique axes to conform with theoretical ideas underlying the formulation of an explanatory model or to facilitate the interpretation of the factors. Orthogonal rotation will retain independent factors while oblique rotation will result in correlated common factors. All rotated systems are mathematically equivalent to each other and to the unrotated solution in that all of them account for the original correlation matrix equally well. "This is so because in the final analysis, different factor

Table 4-1. The basic models for principal components analysis and factor analysis. (Adapted from Morrison, 1967)

Let $X_1, X_2, X_3, \ldots, X_p$ represent the p observable random variables.

<u>Principal Components Analysis</u>

$$Y_1 = a_{11}X_1 + a_{12}X_2 + a_{13}X_3 + \ldots + a_{1p}X_p$$

$$Y_2 = a_{21}X_1 + a_{22}X_2 + a_{23}X_3 + \ldots + a_{2p}X_p$$

$$\vdots$$

$$Y_p = a_{p1}X_1 + a_{p2}X_2 + a_{p3}X_3 + \ldots + a_{pp}X_p$$

Where the Y_i's are the principal components of the sample values of the responses (i.e. the X_i's);

And the a_{ij}'s are coefficients, which when the X_i's are standardized, represent the relative importance of the X_i term in the Y_j component.

<u>Factor Analysis</u>

$$X_1 = w_{11}Y_1 + w_{12}Y_2 + w_{13}Y_3 + \ldots + w_{1m}Y_m + e_1$$

$$X_2 = w_{21}Y_1 + w_{22}Y_2 + w_{23}Y_3 + \ldots + w_{2m}Y_m + e_2$$

$$\vdots$$

$$X_p = w_{p1}Y_1 + w_{p2}Y_2 + w_{p3}Y_3 + \ldots + w_{pm}Y_m + e_p$$

Where the Y_i's are the common factor variates ($m \geq p$);
The w_{ij}'s are parameters reflecting the importance of the jth Y in the composition of the ith X, when the X_i's are standardized; And the e_i's are the specific factor variables.

solutions, rotated or otherwise, merely represent different basis systems (or sets of coordinate axes) for the same vector space" (Comrey, 1973).

The number of common factors must be specified prior to analysis. If a specific model has not been formulated many researchers will take a step-wise approach. First, as many factors will be extracted as there are response variables. At this stage some of the factors will be repetitions; that is some of the factors will overlap with the original variables and no specific factors will be present. Then the number of factors is decreased step by step until the analyst believes that further reduction will seriously reduce the amount of covariance accounted for the factors. Thus it will be realized that a unique solution does not exist for factor analysis.

Because factor analysis represents a model to account for covariance (dependence between variables) it is not subject to the restraint of measuring variables on a single scale and standardization will not change the results. However, Comrey (1973) feels that some type of scale transformation is important in cases where a few scores in a sample depart radically from the others.

RESULTS AND DISCUSSION

Discriminant Analysis

One of the primary advantages of discriminant analysis stems from its flexible use in determining populations of interest from the basic multivariate sample. The assumptions underlying the model are that the populations of the response variables have multivariate normal distributions and a common covariance matrix, but different response means. Violation of these assumptions is not usually a serious problem.

Once the groups have been determined, often based on one of the response variables, this variable is taken out of the data array before the program is run. The step-wise discriminant program utilized[1/] determines the F values for each variable. The variable with the highest F value is the first one used in constructing the discriminant functions. The program output at each step includes the number of plots in each group correctly classified by the discriminant functions, the group discriminant functions, and the F values between the determined groups. After a variable has been put into the discriminant function, the remaining variables' F values are recalculated. The variable with the next highest F value is then added to the discriminant function. The

---

[1/] The specific computer programs used are given in Appendix B.

individual variable's contribution to the function is monitored step by step; thus the variable can be dropped from the function if its contribution is deemed insignificant.

For the first two analyses the plots were divided into three groups based on total basal area. The groups were determined so as to divide the range of basal area into equal thirds. Then the first analysis was run with total basal area and the individual species' values for basal area and $\overline{dbh}$ (average diameter at breast height) omitted from the data. The first variable used in constructing the discriminant functions was percent rock in the soil pit. Using this variable alone 24 plots were correctly classified into the proper basal area groups. (see Table 5-1). The correlation between percent rock in the pit and total basal area was only-.10. Three other site variables had higher correlations with total basal area. Thus one of the interesting results from discriminant analysis is the pinpointing of the variables best able to discriminate between the groups. Percent rock in the pit is probably important because it reduces the soil volume available for water and nutrient extraction by tree roots. The second variable used, number of species on the plot, resulted in discriminant functions that correctly classified 28 plots. This variable may be

Table 5-1. Classification by discriminant analysis.

Case 1: Classification into basal area groups using percent rock in the soil pit.

Group	Number of cases classified into each group		
	Low	Medium	High
Low	5	3	8
Medium	5	13	8
High	0	2	6

Case 2: Classification into basal area groups using red maple basal area.

Group	Number of cases classified into each group		
	Low	Medium	High
Low	13	1	1
Medium	14	7	6
High	2	1	5

Case 3: Classification into groups, based on the age of the previous stand when cut, using $\overline{dbh}$ of chestnut oak.

Group	Number of cases classified into each group	
	Age 40	Age 82
Age 40	20	6
Age 82	2	22

important because few species can tolerate the harsh physical conditions which prevail on the worst sites. On this area, the presence of a large number of species on a plot probably indicates better site conditions. Using all ten variables resulted in correct classification of 2/3 of the plots (i.e., 34 plots).

Next the program was rerun this time including the individual species' values for basal area and $\overline{dbh}$. The first variable used in this analysis was basal area of red maple. This resulted in correct classification of 25 plots. The addition of the second variable, basal area of northern red oak, increased the correct classification by the discriminant functions to 35 plots. By step five, adding the basal area of scarlet oak, chestnut oak, and birch, 44 plots were correctly classified. The importance of percent rock in the pit and number of species was reversed in this analysis. Number of species was added at step six, slope position class at step seven, and percent rock in the pit at step eight. At the end of the analysis all the plots were correctly classified by the discriminant functions.

For the third analysis the original plots were divided into two groups. Group 1 contained the plots that had regenerated from stands that were 40 years old when cut; Group 2, comprised the plots that regenerated from 82 year old stands.

Univariate F tests on the physical site variables revealed no significant differences between the groups. Forest compartment records did not show any important differences in species composition on the two areas at the time of the 1953-1955 clearcut. However, analysis revealed significant species differences between the two groups at the time of measurement.

The first variable used in the analysis, $\overline{\text{dbh}}$ of chestnut oak, correctly classified 42 plots. By step three having added basal area of birch and $\overline{\text{dbh}}$ of a combined group of species called "other", the discriminant function correctly classified 45 plots. At the end of the analysis four plots were misclassified. Referring back to the location map of the study area it was determined that all four plots were near the original boundary line between the two stands. Thus the possibility exists that the plots may have been misclassified by an error in the original compartment map rather than misclassified by the discriminant functions.

The variable means and standard deviations are given in the computer output for the groups used in each discriminant analysis. Comparison of the species' basal area values for the two groups used in the third analysis revealed some interesting differences. Group 1 (regenerated from the younger stand) had three-and-a-half times the basal area of northern red oak that Group 2 had. Northern red oak is the most valuable of the species mix on the study area. Red maple, considered a weed on these poor sites, was present in significantly lower numbers in Group 1. Although further research is definitely needed, the results of this analysis appear to indicate that species composition can be regulated by cutting at different stand ages. The reason behind this response to rotation age apparently lies in the differential relationships between age and sprouting vigor by species.

However, it should be recognized that while hypotheses may be formulated based on the results of a discriminant analysis, the results can not be considered to represent a cause and effect relationship. Cause and effect relationships can only be established through careful control and manipulation under proper experimental design conditions.

In the fourth use of discriminant analysis the groups were based on the result of a cluster analysis. Because the results are most closely tied to the original analysis, they will be discussed in the following section.

From the preceeding discussion it should be evident that many other types of groupings could have been tried. For example, the plots could have been grouped based on aspect or slope position or any other variable of interest. However, given limitations on time and money, trying all possible groupings was not considered feasible.

Cluster Analysis

The cluster analysis program used in this study employs the lumping or aggregation technique described in Chapter VI. The first program run used the raw data from all variables. This resulted in two final groups which were identified as basically low and high total basal area. Since this technique is based on minimizing within group variance while maximizing between group variance, it is not surprising that the groups were primarily seperated on the basis of the variable with the largest variance.

In the second analysis total basal area was deleted from the data. This resulted in two final groups of 15 and 35 plots. Discriminant analysis was then run using these two groups. All plots were correctly classified. The three most important variables in discriminating between the groups were $\overline{\text{dbh}}$ scarlet oak, $\overline{\text{dbh}}$ chestnut oak, and basal area white oak. Using the first two variables alone the differences between the groups were significant at the 97.5% level (F test, $\alpha =$ .025). With the three variables used the significance level increased to 99.5%. The first group had no birch or striped maple, greater basal area and $\overline{\text{dbh}}$ of scarlet and white oak, lower percent surface rock and percent slope, and contained more of the plots that were in the stand that was younger when cut. The second group had significantly higher basal area and $\overline{\text{dbh}}$ of red maple and chestnut oak.

Although the usefulness of these groups may not be immediately apparent, it should be remembered that the sampling efficiency of any further research in this area can be greatly increased by using natural clusters or strata such as these groups represent. For example, the development of accurate site index prediction equations is enhanced when related groups are formed prior to multiple regression analysis. In addition, further analyses may be suggested by the type of groupings, leading to increased understanding of the relationships within the system.

Thus, future research might confirm that scarlet and white oak are often associated together under certain site conditions. On this area this association might indicate similiar species tolerances. If so, then the planting of scarlet oak would be assured a better chance of survival on sites where white oak is present. In the planning stages of a planting program, white oak sites should have a high priority.

Further computer runs using cluster analysis were done using standardized variables both with and without the inclusion of total basal area as a variable. Deleting basal area made no difference whatsoever in the plot groupings when standardization was employed. Interestingly enough the final two groups obtained using standardized variables were quite similiar to those that resulted when basal area was deleted from the raw data.

Analysis of other group numbers (i.e., when the number of groups was three or more) was not done because it was felt that the information that would be gained would be repetitions considering the analytical methods used. However, it should be emphasized that cluster analysis can be a powerful statistical tool used either alone or in conjunction with other techniques of analysis. Most computer programs give the results step by step. This means that the groups are listed at each step along with the error calculated for that grouping. Groups or individual plots that remain intact through a large number of steps can be examined to see what determines their coherence or isolation. When proportionately large jumps are seen in the calculated error, the merged groups can be carefully examined to determine their basic differences.

Multiple Regression

Linear multiple regression was run using a step-wise program. This program is similiar to the step-wise discriminant program in that variables are added to the function one at a time based on their F values. The regression function, partial and multiple correlation coefficients, standard error of estimation, and an analysis of variance table are all presented at each step. Five analyses were run; total basal area with all the other variables, total basal area with the site variables, and individually the basal area of red maple, northern red oak, and scarlet oak with the site variables, The results were basically disappointing.

The first analysis, which included the species values for basal area among the independent variables, gave the most accurate regression equation. The first six variables, all basal area of individual species, gave an equation that accounted for 99% of the variation in total basal area. The first site variable, percent rock in the pit, was added at step 10. Changes in the multiple R^2 were in the fourth or fifth decimal place from step 10 to the end of the analysis.

The second analysis used only the physical site variables and number of species as the independent variables with total basal area as the dependent variable. Using nine of the independent variables, only 17% of the variation in total basal area was accounted for. This is equivalent to a multiple correlation coefficient of .42. The regression function was not significant.

Multiple regression analysis with red maple basal area as the dependent variable also yielded nonsignificant results. Including seven of the physical site variables in the regression equation resulted in a multiple R^2 of .10. Thus the site variables acting together only accounted for 10% of the variation in red maple basal area.

The next two analyses, which used northern red oak and scarlet oak as the dependent variables, both resulted in highly significant regression equations (F test, $\alpha = .01$). In the analysis of northern red oak age at the time of cutting was the first variable used, followed by color, percent rock in the pit, distance to the perimeter of the cut, soil depth, percent surface rock, slope position class, and aspect. The multiple correlation between these variables and northern red oak was .63. Thirty-nine percent of the variation in the dependent variable was accounted for by variation in the independent variables. Percent slope was not included in the equation because the F-level was too low to warrant further computation.

The relative importance of the variables was almost completely reversed in the analysis of scarlet oak. Percent slope was first variable in the regression function followed by slope class, soil depth, distance to the perimeter of the cut, percent rock in the pit, percent surface rock, aspect, and color. Age was not included in the function. The multiple correlation between scarlet oak basal area and the physical site variables was .61. Thirty-seven percent of the variation in the dependent variable was explained by variation in the independent variables.

Regression analysis is the most practical or management oriented of the methods utilized. When high multiple R^2 values are obtained, the equations can be useful for prediction purposes. It should again be mentioned that high R^2 values do not indicate a cause and effect relationship. Care should be taken not to fall into this attractive trap. However, when low R^2 values are obtained in analysis, thought should be given to the possible causes. Were the variables measured with sufficient precision? Or should additional variables be measured to gain the required precision? If a large number of variables will need to be measured, then regression can become an expensive and time consuming method of analysis.

Multiple regression analysis can also be done using variable transformations or combinations in linear or non-linear models. These alternatives might have resulted in more accurate equations, that is, equations in which the independent variables accounted for a greater percent of the variation in the dependent variable. However, these alternatives were not explored because it was felt that additional transformations, variable combinations, and non-linear models would give results that would be difficult to interpret biologically.

Canonical Correlation Analysis

Six analyses were run using canonical correlation. With the exception of the analyses of white oak and red maple, all the correlations obtained with the first pair of canonical variates in each analysis were significant at the 99% level or better.

Using all of the productivity measures (i.e., all the basal area and $\overline{dbh}$ values) as one variable set and the physical site measures as the other variable set, the first canonical variates had a correlation of .89 significant at α = .0008. The second variate pair had a correlation of .84 significant at α = .0096. The relationships between the other seven variate pairs were not significant. Total basal area and the species' basal area values had the highest coefficients or loadings on both of the productivity variates. The site variates however showed different patterns. On the first site variate, age of the previous stand and slope position class had the highest loadings, while on the second site variate percent slope and distance to the perimeter of the cut had the highest loadings. In canonical analysis multiplying the correlation between the canonical variates by the loading of a variable on one of the variates gives the correlation between that variable and the other canonical variate. For example, slope position class on the first site variate had a loading of .64. Multiplying this loading by .89 (the correlation between the variate pair) gives a correlation of .57 between slope class and the first productivity variate. This means that 32% ($.57^2$) of the variation in the productivity variate can be explained by variation in position slope alone.

Canonical correlation analyses were also run for each of the five major species using the basal area and $\overline{\text{dbh}}$ measures for the productivity variate. The resulting correlations ranged from .49 for white oak to .71 for chestnut oak. All of the correlations were higher than those obtained from the comparable multiple regression analyses.

In the analysis of chestnut oak, age of the previous stand had the highest loading, .83, followed by percent slope at .39. On the productivity variate $\overline{\text{dbh}}$ was the most important variable having a loading of .99. This relationship between $\overline{\text{dbh}}$ chestnut oak and age also came out in discriminant analysis. With only two variables in the productivity set, the usual pattern was for one of the variables to be of major importance in one variate pair while the other productivity measure was more heavily loaded on the second variate pair.

The analysis of northern red oak showed age of the previous stand and soil depth explained 39% and 34%, respectively, of the variation in the productivity variate with basal area being most heavily weighted (.88). Age in the analysis was negatively correlated with the productivity variate. Scarlet oak productivity, with both basal area and $\overline{\text{dbh}}$ having moderate to high loadings, was most strongly related to percent slope, age, and soil depth with loadings of -.70, -.45, and .40 respectively.

The most important site variable obtained from examination of the canonical loadings was in all cases the first variable selected in step-wise multiple regression analysis. However, the relative importance of the other variables used in analysis often changed drastically between canonical and regression analysis. For example, regression analysis of scarlet oak did not include age as a variable because the probability of significance was too low; however, age had the second highest loading in canonical analysis.

Factor Analysis

The step by step method mentioned in Chapter 4 was used to determine the number of factors to be extracted for the final analysis. It wasn't felt that the measured variables were precise enough to serve as estimators of the basic factors that influence site productivity. For this reason a model to explain the variation in the observed responses was not constructed.

In the first run the maximum number of factors (i.e. 27) was extracted. Of the variation in the responses 84.6% was common. This means that the common-factor variates generate 84.6% of the variance among the responses (covariance) while 15.4% of the variation is due to variation of the specific responses. Examination of the factors revealed that over half of them individually explained less than 2% of the covariance. All of the extracted factors together accounted for 100% of the observed variation. The second computer run extracted ten factors, accounting for 87.8% of the variance. The last three factors extracted on this run individually accounted for less than 5% of the variance. A third and last factor analysis was performed, this time extracting seven factors. Together these factors accounted for or extracted 77.5% of the variation. It was felt that this was a reasonable stopping point for the extraction process.

The computer program used for this analysis employed Kaiser Varimax rotation of the computed factor scores. The

Varimax method is the most popular rotational procedure for factor analysis in use today. It maximizes the variance of the squared factor loadings by columns (i.e. maximizes the loadings on the factors rather than on the variables). Comrey (1973) discusses the interpretation of Varimax loadings as follows:

> "On any given factor, a pattern is desired such that there are some high loadings and lots of low loadings with few intermediate-sized loadings. This type of solution offers easy interpretability because variables with high loadings are similiar to the factor in character, where as variables with low loadings are not. A variable with an intermediate loading is... of little use in telling what the factor is or is not like. Varimax rotations, then, tend to push high loadings higher and low loadings lower, to the extent that this is possible within the constraints maintained by an orthogonal reference frame. Factor positions are determined, therefore, by the locations of clusters as well as by the location of hyperplanes. A Varimax factor will tend to be drawn toward a neighboring cluster because this elevates the variance of the squared factor loadings for the factor."

This interpretation of a factor being drawn toward a neighboring cluster is especially interesting when the actual factors are examined. Careful examination of the first factor in Table 5-2 will show that the variables with the highest loadings are the same variables which were most important in constructing the discriminant functions which seperated the groups determined by cluster analysis. The approximate

Table 5-2 Variable factor loadings for the first four extracted factors.

Variable	Variable loadings				Communality
	Factor 1	Factor 2	Factor 3	Factor 4	
Total basal area	-.30	-.04	.05	-.04	.934
# of species	-.42	.26	.41	.48	.765
Soil depth	-.50	.25	-.45	.21	.615
Aspect	.15	-.05	.15	.16	.117
% slope	.51	-.12	.36	.09	.591
Slope position	.37	.57	.19	.14	.563
Distance to edge	-.26	.09	-.02	.00	.156
Age	.00	.30	.64	-.20	.578
% surface rock	.41	-.34	.32	-.08	.533
% soil rock	-.12	.50	-.09	-.07	.317
Color	.11	-.08	.04	-.01	.187
Basal area					
Red maple	.17	.02	.13	-.16	.772
N. red oak	-.02	-.32	-.55	.11	.799
Scarlet oak	-.81	-.04	-.18	-.18	.745
Chestnut oak	.07	.03	.82	.02	.723
White oak	-.75	-.17	.11	.06	.656
Birch	.26	.82	.15	.05	.850
Other	-.07	.07	-.05	.94	.890
dbh					
Red maple	.08	.05	.10	.19	.684
N. red oak	-.40	-.25	-.35	.21	.525
Scarlet oak	-.78	-.09	.07	-.02	.637
Chestnut oak	.03	.00	.84	.08	.830
White oak	-.73	-.20	.14	.30	.726
Birch	.17	.70	.29	.36	.744
Other	-.03	.07	-.18	-.93	.898

relationship between the factor loading (the variable factor correlation) and its value for factor interpretation purposes is presented in Table 5-3. This first factor has high loadings for $\overline{dbh}$ and basal area for scarlet and white oak and low loadings for chestnut oak and red maple. These were the characteristics that defined the first group determined by cluster analysis. The second group defined by cluster analysis is represented on factor 3. Here the highest loadings are on the chestnut oak variables. Age is also weighted fairly heavily, no surprise considering the relationship between chestnut oak and age that was revealed in discriminant analysis.

Factor 2 is heavily weighted for basal area of birch and moderately to heavily weighted for $\overline{dbh}$ birch, slope position class and percent rock in the pit. Birch $\overline{dbh}$ and basal area were correlated most strongly with slope position in the basic correlation matrix. A weak but positive correlation was also observed between slope position and percent rock in the pit. Although factor indentification is not definite it seems probable that this factor represents water availability. Moist soils, especially those associated with protected sites, are usually considered most favorable for birch. Percent rock in the pit is probably weighted because of its relationship with slope position and water availability. Perhaps better

Table 5-3. Importance of variable-factor correlations. (Adapted from Comrey, 1973.)

Factor loading for variable	Percentage of variance common with factor	Rating for factor interpretation
.71	50	Excellent
.63	40	Very good
.55	30	Good
.45	20	Fair
.32	10	Poor

results could have been obtained if the data from grey and black birch had not been combined for analysis.

Factor 4 is apparently a measure of diversity. It is highly correlated with basal area and $\overline{\text{dbh}}$ of the less common species. Moderate loading of number of species is also present on Factor 4. However, what the common-factor variates are behind Factor 4 or Factors 1 and 3 is not clear. Perhaps they represent specific fertility factors important for the species heavily loaded on these factors. However, without additional information on nutrient availability these hypotheses can not be tested.

Factors 5 and 7 appear to be general productivity factors. Factor 5 has high negative loadings for red maple, the most common species, and a moderate-sized negative loading for total basal area. Factor 7 in contrast, has high positive loadings for total basal area and for northern red oak basal area. Northern red oak is the species with the greatest basal area on the study area. Factor 6 appears to be correlated with the absence of striped maple. These factors were judged to be too general to warrant further discussion.

Another aspect of interpreting a factor analysis deals with examining the communalities of the variables. These communalities, often represented h^2 are given in Table 5-2. The h^2 value for a variable represents the overlap between the variable and the seven factors in what they measure. If the communality for a variable was 1.0, it would indicate that the variable overlaps totally with the factors. In this case, the variable's scores could be predicted perfectly by a weighted combination of the factor scores. With this in mind it is interesting to note that the variables with the lowest communalities (below .20) are the artificial or coded variables, aspect, color, and distance to the perimeter of the cut. These variables play an insignificant role on these factors which apparently measure productivity or the underlying influences on productivity.

SUMMARY AND CONCLUSIONS

Regeneration Variability

One of the objectives of this study was to discover some of the factors which influence the variability of regeneration in some young mixed oak stands. To aid in achieving this goal, fifty plots were located on the study area and a series of edaphic, physiographic, biological, and management variables were measured. The data were analyzed using a number of multivariate statistical methods. The results were fairly good. Insights were gained about some of the factors which affect species density and diversity.

Water availability is probably the most important factor influencing the amount of basal area present on a plot. The single most important measured variable that discriminated between basal area classes was percent rock in the soil pit. This variable is probably important for a number of reasons. First, as percent of rock in the pit increases, the soil volume at any given soil depth decreases. In droughty, shallow soils, reducing the soil volume available for water or nutrient extraction by tree roots probably acts as a major factor limiting the density a plot can support. In adjacent mature stands complete canopy closure is rare, most likely indicating severe root competition.

The average amount of rock in the soil pit varies from 16% in the low basal area class to 26% in the medium class to 10% in the highest class. From the previous discussion, the expected rock percentages in the low and medium classes should be the reverse of those obtained. However, average soil depth to bedrock showed a similiar up and down pattern with the values in the low, medium and high groups being 14, 19½, and 15 inches respectively. Thus the high percentage of rock in the medium group can be partially explained by the higher associated soil depth. It was not deemed possible to determine accurately the actual soil volumes for the basal area groups because the percent rock measurements were made by occular estimation.

While it is felt that a linear relationship may exist between soil volume and basal area, this relationship is complicated by the correlation between slope position class and percent rock in the pit. Slope position was also an important variable in discriminating between basal area groups. The average values for slope position class were 3.8, 3.4, and 4.5 for the low, medium and high groups respectively. The highest position classes are at the bottom of the slope where the greatest amount of percent rock in the pit is found. The highest slope position classes are probably most favorable because of their more sheltered position and their increased water supply gained from downslope water movement. Slope position probably partially compensates for increased soil rock.

An eastern aspect predominated on the best sites while north-northwest was most common on the poor sites. This relationship lends support to Lloyd and Lemmon's (1970) remarks that optimum rectification of azimuth depends on the soil series under consideration. If further research is done in this area, aspect should be coded to result in a linear relationship with site productivity.

Rather clearly, species diversity was strongly tied to the age of the previous stand when cut. This one factor was of overriding importance in determining the presence and vigor of many of the most common species. White oak was the only species that showed no differences between the two groups. The difference in birch basal area was significant at the 96% level. Group differences in $\overline{\text{dbh}}$ and basal area of northern red oak and chestnut oak were significant at the 99% level or better. Northern red oak basal area and $\overline{\text{dbh}}$ were higher in the group of plots that regenerated from 40 year-old stands while $\overline{\text{dbh}}$ and basal area of chestnut oak and basal area of birch were significantly higher in the group that regenerated from 82 year-old stands. The probability that the differences between the groups were significant for the other species ranged from 75 to 87 percent.

If this relationship between a species' presence and vigor and the rotation age of the previous stand is confirmed in future research, the forester in these types of low productivity stands will gain an important silvicultural tool. On poor sites investments in planting or seeding with weed species control can not currently be justified in terms of financial return. However, if species manipulation can be gained via commercial cuts, the future value of these stands can be increased considerably.

Slope position class and percent slope were also important in determining species importance. The distributions of birch, red maple, and scarlet oak were strongly influenced by these two variables. Scarlet oak was more prevalent at the higher elevations with steeper slopes while red maple and birch were most common toward the bottom of the slope, associated with milder gradients.

It is felt that the use of basal area and $\overline{dbh}$ has real utility in analyzing young even-aged stands. These measurements are easy to obtain and are moderately well correlated with the measured site factors. Site index, the most widely used measure of forest productivity, is not applicable in young, sprout-origin, mixed-species stands. Measures of density, diversity, or species importance used in many ecological studies were examined in a preliminary analysis and found to have lower correlations with the site factors than the measures of $\overline{\text{dbh}}$ and basal area did. Also, the use of synthetic indices makes the interpretation of the results more difficult.

Analytical Techniques

All of the statistical techniques utilized yielded information that was useful in identifying some of the factors influencing the variability of the regeneration. There was quite a bit of overlap in the information gained from some of the analytical techniques, but this was to be expected considering the number of analyses performed. The specific information gained from each technique was given in the preceding chapter, but some general comments on each method are presented below.

First, none of the methods should be considered impratical because of their cost. Thirty-eight successful and uncounted unsuccessful computer runs were made during the course of this study. The total bill for computer time came to less than two hundred dollars. A discriminant analysis program with the maximum number of options selected and printed out cost about three dollars. All of these techniques are available in widely distributed statistical packages such as BMD biomedical computer programs. Thus, these statistical methods should not be considered either expensive or unavailable.

Discriminant analysis was the most successful technique in terms of the information gained. The results were easy to interpret and were least duplicated by the other statistical methods. Future forestry research might well include this method of analysis.

Cluster analysis was most useful when used in conjunction with discriminant analysis. Two natural groupings were determined which could easily form the basis for stratified sampling if future research is conducted on the study area. In particular, if site index classification is desired when the stands are older, sampling within these clusters will probably result in increased accuracy.

Multiple regression analysis usually has an immediate practical goal, such as the prediction of forest productivity classes. Both in terms of accurate prediction and of the basic information gained, regression analysis was fairly unsuccessful. Whether the poor results were due to the type of relationships tried (i.e., linear) or whether the measured variables were insufficient to account for the observed variation in basal area can not be determined without further analysis.

Canonical correlation analysis gave information on the types and strengths of the relationships between the productivity and the site variables. While this information was useful, it is felt that canonical analysis would be of greatest value when larger, more complex variable sets are used.

Factor analysis in this study served basically as additional confirmation of the relationships revealed in other analyses. Factor analyses are often undertaken in series to identify accurately the exact nature of the underlying factors. Future research would probably attempt to measure variables which are hypothesized to influence or measure the factors. Large sample sizes are needed for the researcher to have confidence in the results. Factor analysis alone offers the possibility of actually identifying the common causes which influence productivity, and should not be neglected because of the scope of the research projects needed.

REFERENCES CITED

Aikman, J. M. 1941. The effect of aspect of slope on climate factors. Iowa State College Journal of Science 15: 161-167.

Aldrich, R. C., W. J. Greentree, R. C. Heller, N. X. Norick, 1970. The use of space and high altitude aerial photography to classify forest land and to detect forest disturbances. Pacific Southwest Forest and Range Experiment station Annual Progress Report.

Anderson, R. W. 1958. Introduction to Multivariate Statistical Analysis, John Wiley and Sons, Inc. New York.

Arend, J. L. and L. W. Gysel. 1951. Some effects of soil and topography on oak yields in southern Michigan. Lake States Forest Experiment Station Technical Note 365. SD 11 .S15 301-450.

Arend, J. L. and Julander, O. 1948. Oak sites in the Arkansas Ozarks. Arkansas Agricultural Experiment Station Bulletin 484.

Auten, J. T. 1945a. Some soil factors associated with site quality for planted black locust and black walnut. Journal of Forestry 43: 592-598.

Auten, J. T. 1945b. Prediction of site index for yellow poplar from soil and topography. Journal of Forestry 43: 662-668.

Avery, T. E. 1966. Forester's guide to aerial photo interpretation. Agriculture Handbook No. 308. U.S.D.A. Washington, D. C.

Beaufait, W. R. 1956. Influence of soil and topography on willow oak sites. U.S.D.A. Forest Service, Southern Forest Experiment Station Occasional Paper 148.

Berglund, J. V. 1968. The use of Modal Soil Taxonomic Units for the Prediction of Sugar Maple Site Productivity in Southern New York. Unpublished doctoral thesis, State University College of Forestry, Syracuse.

Beers, T. W., P. E. Dress, and L. C. Wensel. 1966. Aspect transformation in site productivity research. Journal of Forestry 64: 691-692.

Bowersox, T. W. and W. W. Ward. 1972. Prediction of oak site index in the Ridge and Valley Region of Pennsylvannia. Forest Science 18: 192-195.

Broadfoot, W. M. 1961. Guide for evaluating cherrybark oak sites. U.S.D.A., Forest Service Southern Forest Experiment Station Occasional Paper 190.

Broadfoot, W. M. 1963. Guide for evaluating water oak sites. U.S.D.A., Forest Service Research Paper SO-1. Southern Forest Experiment Station.

Bull, H. 1931. The use of polymorphic curves in determining site quality in young Red Pine plantations. Journal of Agricultural Research 43: 1-28.

Cajander, A. K. 1926. The Theory of Forest Types. The Printing Office of Society for the Finnish Literary, Helsinki.

Carmean, W. H. 1954. Site quality for Douglas-fir in southwestern Washington and its relationship to precipitation, elevation, and physical soil properties. Soil Science Society of America Proceedings 18: 330-334.

Carmean, W. H. 1961. Soil survey refinements needed for accurate classification of black oak site quality in southeastern Ohio. Soil Science Society of America Proceedings 25: 394-397.

Carmean, W. H. 1965. Black oak site quality in relation to soil and topography in southeastern Ohio. Soil Science Society of America Proceedings 29: 308-312.

Carmean, W. H. 1967. Soil survey refinements for predicting black oak site quality in southeastern Ohio. Soil Science Society of America Proceedings 31: 805-810.

Carmean, W. H. 1970. Site quality for eastern hardwoods. The Silviculture of Oaks and Associated Species. U.S.D.A. Forest Service Research Paper NE-144.

Carmean, Willard H. 1971a. Site index curves for black, white, scarlet, and chestnut oaks in the Central States. U.S.D.A. Forest Service Research Paper NC-62.

Carmean, W. H. 1971b. Soil-site relationships of the upland oaks. Oak Symposium Proceedings, pp 23-29. U.S.D.A. Forest Service, Northeast Forest Experiment Station.

Chikishev, A. G., editor. 1965. Plant Indicators of Soils, Rocks and Subsurface Waters. Translation of the Proceedings of the Conference on Indicational Geobotany, Moscow, 1961. Consultants Bureau Enterprises, Inc., New York.

Coile, T. S. 1935. Relations of site index for shortleaf pine to certain physical properties of the soil. Journal of Forestry 33: 726-730.

Coile, T. S. 1940. Soil changes associated with loblolly pine succession on abondoned agricultural land of the Piedmont Plateau. Duke University School of Forestry Bulletin 5.

Coile, T. S. 1942. Some physical properties of the B horizons of Piedmont soils. Soil Science 54: 101-103.

Coile, T. S. 1952a. Soil productivity for southern pines. Forest Farmer X1 (7): 10, 11-13; X1 (8): 11-12.

Coile, T. S. 1952b. Soil and the growth of forests. Advances in Agronomy 16: 329-398.

Coile, T. S. and F. X. Schumacher. 1953. Relation of soil properites to site index of loblolly and shortleaf pines in the Piedmont region of the Carolinas, Georgia, and Alabama. Journal of Forestry 51: 739-744.

Committee on Remote Sensing for Agricultural purposes, Agricultural Board National Research Council. 1970. Remote sensing with special reference to agriculture and forestry. National Acadamy of Sciences, Washington, D. C.

Comrey, Andrew L. 1973. A First Course in Factor Analyses. Academic Press, New York.

Copeland Jr., O. L. 1956. Preliminary soil-site studies in the western white pine type. U.S.D.A. Forest Service, Intermountain Forest and Range Experiment Station Research Note 33.

Corson, C. W., J. H. Allison, E. G. Cheyney. 1929. Factors controlling forest types on the Cloquet Forest, Minnesota Ecology 10: 112-125.

Cox, G. S., R. C. Mc Connell, and L. M. Matthew. 1960. Ponderosa pine productivity in relation to soil and land form in the western Montana. Soil Science Society of America Proceedings 24: 139-142.

Curtis, R. O. and B. W. Post. 1962. Site-index curves for even-aged northern hardwoods in the Green Mountains of Vermont. University of Vermont and State Agriculture College Bulletin 629.

Curtis, R.O. and B.W. Post. 1964. Basal area, volume, and diameter related to site index and age in unmanaged even-aged northern hardwoods in the Green Mountains. Journal of Forestry 62: 864-870.

Deitschman, G. H. and A. W. Green. 1965. Relations between western white pine site index and tree height of several associated species. U.S.D.A., Forest Service, Intermountain and Range Experiment Station, INT-22.

Della-Bianca, L. and D. F. Olson,Jr. 1961. Soil-site studies in Piedmont hardwood and pine-hardwood upland forests. Forest Science 7: 320-329

Denny, C. S. 1938. Glacial Geology of the Black Rock Forest. The Black Rock Forest Bulletin Number 8. Cornwall, Press, Inc. Cornwall, New York 70 pp.

Diebold, C. H. 1935. Some relationships between soil type and forest site quality. Ecology 16: 640-647.

Dixon, W. J., editor. 1971. BMD Biomedical Computer Programs. University of California Press, Berkeley.

Doolittle, W. T. 1957. Site index of scarlet and black oak in relation to southern Appalachian soil and topography. Forest Science 3: 114-124.

Einspahr, D. and A. L. Mc Comb. 1951. Site index of oaks in relation to soil and topography in northeastern Iowa. Journal of Forestry 49: 719-723.

Flint, R. F. 1971. Glacial and Quaternary Geology. John Wiley and Sons, Inc., New York pp. 267-286.

Foster, R. W. 1959. Relation between site indexes of eastern white pine and red maple. Forest Science 5: 279-291.

Fowells, H., editor. 1965. Silvics of Forest Trees of the United States. Agriculture Handbook 271. U.S.D.A., Washington D. C.

Freese , Frank. 1967. Elementary Statistical Methods for Foresters. Agriculture Handbook 317. U.S.D.A., Forest Service.

Gaiser, R. N. 1950. Relation between soil characteristics and site index of loblolly pine in the coastal plain region of Virginia and the Carolinas. Journal of Forestry. 48: 271-275.

Gaiser, R. N. 1951. Relation between topography, soil characteristics, and site index of white oak in southeastern Ohio. U.S.D.A., Forest Service. Central States Forest Experiment Station Technical Paper 121.

Gaiser, R. N. and R. W. Merz. 1951. Stand density as a factor in estimating white oak site index. Journal of Forestry 49: 572-574.

Gessel, S. P. 1949. Correlation between certain soil characteristics and site for Douglas-fir in northwestern Washington. Soil Science Society of America Proceedings 14: 333-337.

Gessell, S. P., and W. J. Lloyd. 1950. Effect of some physical soil properties on Douglas-fir site quality. Journal of Forestry 48: 405-410.

Gevorkiantz, S. R. and H. F. Scholz. 1944. Determining site quality in understocked oak forests. Journal of Forestry 42: 808-811.

Gleason, H. A. and A. Cronquist. 1963. Manual of Vascular Plants of Northeastern United States and Adjacent Canada. Van Nostrand Reinhold Company, New York.

Haig, I. T. 1929. Collodial content and related soil factors as indicators of site quality. Yale University School of Forestry Bulletin 24.

Hannah, P. R. 1968a. Estimating site index for white and black oaks in Indiana from soil and topographic factors. Journal of Forestry 66: 412-417.

Hannah, P. R. 1968b. Topography and soil relations for white and black oak in southern Indiana. U.S.D.A., Forest Service Research NC-25.

Hansen, J. N. and A. L. Mc Comb. 1958. Growth of planted green ash, black walnut, and other species in relation to observable soil-site characteristics in southeastern Iowa. Journal of Forestry 56: 473-480.

Hebb, E. A. 1962. Relation of tree growth to site factors. Tennessee Agricultural Experiment Station Bulletin 349.

Heiberg, S. O. and D. P. White. 1956. A site evaluation concept. Journal of Forestry 54: 7-10.

Heimburger, C. C. 1934. Forest-type studies in the Adirondack region. Cornell University Agricultural Experiment Station Memoir 165.

Heller, R. C., G. E. Doverspike, R. C. Aldrich, 1964. Indentification of Tree Species on Large Scale Panchromatic and Color Aerial Photographs, Agriculture Handbook No. 261. U.S.D.A., Washington, D. C.

Hicock, H. W., M. F. Morgan, H. J. Lutz, H. Bull, H. A. Lunt. 1931. The relation of forest composition and rate of growth to certain soil characteristics. Connecticut Agricultural Experiment Station Bulletin 330.

Hill, W. W., A. Arnst, and R. M. Bond 1948. Method of correlating soils with Douglas-fir site quality. Journal of Forestry 46: 835-841.

Hills, G. A. 1952. The classification and evaluation of site for forestry. Research Report 24, Ontario Department of Lands and Forests, Toronto.

Hodgkins, E. J. 1959. Forest site classification in the Southeast. 8th Annual Forestry Symposium-Southern Forest Soils. Louisiana State University Press, Baton Rouge, pp. 34-48.

Hodgkins, E. J. 1961. Estimating site index for longleaf pine through quantitative evaluation of associated vegetation. Proceedings, Society of American Foresters, 1960, pp. 28-32.

Hodgkins, E. J. 1970. Productivity estimation by means of plant indicators in the longleaf pine forests of Alabama. Tree Growth and Forest Soils, Third North American Forest Soils Conference 1968. Oregon State University Press, Corvallis pp. 461-474.

Horst, P. 1961. Relations among m sets of measures. Psychometrika 26: 129-149.

Jackson, D. S. and H. H. Gifford, 1974. Environmental variables influencing the increment of radiata pine (1) periodic volume increment. New Zealand Journal of Forestry Science 4: 3-26.

Jardine, N. and R. Sibson. 1970. Mathematical Taxonomy. John Wiley and Sons, Inc., London.

Jeglem, J. K. 1974. Relative influence of moisture-aeration and nutrients on vegetation and black spruce growth in northern Ontario. Canadian Journal of Forestry Research 4: 114-126.

La Bastide, J. G. A. and Van Goor, E. P. 1970. Growth-site relationships in plantations of Pinus elliottii and Araucaria angustifolia in Brazil. Plant Soil 32: 349-366.

Lemmon, P. E. 1970. Grouping soils on the basis of woodland suitability. Tree Growth and Forest Soils. Proceedings of the Third North American Forest Soils Conference, 1968. Oregon State University Press, Corvallis.

Lewis, R. A. and W. W. Ward. 1967. Site quality evaluation for white ash. Pennsylvania State University, School of Forest Resources, Research Briefs 2: 26-29.

Little, E. L.,Jr. 1953. Check List of Native and Naturalized Trees of the United States Including Alaska). USDA Agriculture Handbook 41. USDA, Washington, D.C.

Lloyd, W. J. and P.E. Lemmon. 1970. Rectifying azimuth (of aspect) in studies of soil-site index relationships. Tree Growth and Forest Soils. Proceedings of the Third North American Forest Soils Conference, 1968, Oregon State University Press, Corvallis, pp. 435-448.

Lunt, H. A. 1939. Soil characteristics, topography, and lesser vegetation in relation to site quality of second-growth oak stands in Connecticut. Journal of Agricultural Research 59: 407-428.

Lunt, H. A. and M. C. Baltz. 1943. Basal area in relation to soil-site factors in two Connecticut forests. Soil Science Society of America, Proceedings 8: 415-419.

Mader, D. L. 1961. Relationships between soil properties and red pine growth in Massachusetts. Soil Science Society of America, Proceedings 25: 62-65.

Mader, D. L. 1963. Volume growth measurement-analysis of function and characteristics in site evaluation. Journal of Forestry 61: 193-198.

Mader, D. L. 1965. Where are we in soil-site classification? Applications of Soils Information in Forestry, Joint Meeting Society of American Foresters and Soil Conservation Society of America, Syracuse, 1964. pp. 23-32.

Mc Clurkin, D. C. 1963. Soil-site index predictions for white oak in north Mississippi and west Tennessee. Forest Science 9: 108-113.

Minckler, L. S. 1943. The effect of rainfall and site factors on the growth and survival of a young forest plantation. Journal of Forestry 41: 829-833.

Morrison, D. F. 1967. Multivariate Statistical Methods. Mc Graw-Hill Book Company, New York.

Nelson, T. C. and W. R. Beaufait. 1956. Studies in site evaluation for southern hardwoods. Society of American Foresters Annual Meeting Proceedings, pp. 67-70.

Nuss, D. E. and F. Y. Borden. 1968. Canonical correlation analysis as a method of arriving at an estimator of the quality of sugar maple in Pennsylvania. Pennsylvania State University, School of Forest Resources, Research Briefs 2: 71-74.

Peterson, G. W., R. L. Cunningham and R. P. Matelski. 1971. Moisture characteristics of Pennsylvania soils: III Parent material and drainage relationships. Soil Science Society of America Proceedings 35: 115-119.

Phillips, J. J. 1966. Site index of yellow-poplar related to soil and topography in sourthern New Jersey. U.S.D.A., Forest Service Research Paper NE-52.

Poulton, C. E., J. R. Johnson, D. A. Mouat. 1970 Inventory of native vegetation and related resources from space photography. Department of Range Management. Agricultural Experiment Station, Oregon State University.

Ralston, C. W. 1951. Some factors related to the growth of longleaf pine in the Atlantic Coastal Plain. Journal of Forestry 49: 408-412.

Ralston, C. W. 1964. Evaluation of Forest Site Productivity. International Review of Forestry Research 1: 171-201. Academic Press, New York.

Roy, S. N. 1957. Some Aspects of Multivariate Analysis. John Wiley and Sons, Inc., New York.

Seal, H. 1964. Multivariate Statistical Analysis for Biologists. Methuen and Company, Ltd., London.

Silker, T. H. 1965. Plant indicators communicate ecological relationships in North America. Second North American Forest Soils Conferences 1963. Oregon State University Press, Corvallis. pp. 317-329.

Sisam, J. W. 1938. Site as a factor is silviculture: its determination with special reference to the use of plant indicators. Canadian Department of Mines and Resources, Silvicultural Research Note 54.

Snedecor, G. W. 1956. Statistical Methods. 5th edition. The Iowa State College Press, Ames.

Sokal, R. R. and J. F. Rohlf. 1969. Biometry. W. H. Freeman and Company, San Francisco, p. 501.

Spurr, S. A., 1960. Photogrammetry and Photo-Interpretation with a Section on Applications to Forestry. The Ronald Press Company, New York.

Stoeckeler, J. H. 1960. Soil factors affecting the growth of quaking aspen forests in the Lake States. Minnesota Agricultural Experiment Station Technical Bulletin 233.

Trimble, Jr., G. R. and S. Weitzman. 1956. Site index studies of upland oaks in the northern Appalachians. Forest Science 2: 162-173.

Veldman, D. J. 1967. Fortran Programming for the Behavioral Sciences. Holt, Rinehart and Winston, New York.

Weislander, A. E. and R. E. Storie. 1952. The vegetation-soil survey in California and its use in the management of wild lands for yield of timber, forage, and water. Journal of Forestry 50: 521-526.

Westveld, M. 1952. A method of evaluating forest site quality from soil, forest cover, and indicator plants. U.S.D.A., Forest Service, Northeast Forest Experiment Station Paper 48.

Westveld, M. 1954. Use of plant indicators as an index of site quality. U.S.D.A., Forest Service, Northeastern Forest Experiment Station Paper 69.

Westveld, R. H. 1933. The relation of certain soil characteristics to forest growth and composition in the northern hardwood forest of northern Michigan. Michigan State College Agriculture Experiment Station Technical Bulletin 135.

Wold, H. and L. Jureen. 1953. Demand Analysis. John Wiley and Son, New York.

Yawney, H. W. 1964. Oak site index of Belmont limestone soils in the Alleghany Mountains of West Virginia. U.S.D.A., Forest Service Research Paper NE-30.

Yawney, H. W. and G. R. Trimble, Jr. 1968. Oak soil-site relationships in the Ridge and Valley Region of West Virginia and Maryland. U.S.D.A., Forest Service Research Paper NE-96.

Youngberg, C. T., and H. F. Scholz. 1949. Relation of soil fertility and rate of growth of mixed oak stands in the driftless area of southwestern Wisconsin, Soil Science Society of American Proceedings 14: 331-332.

Zahner, R. 1957. Field procedures for soil-site classification of pine land in south Arkansas and north Lousiana. U.S.D.A., Forest Service, Southern Forest Experiment Station Occasional Paper 155.

Zinke, P. J. 1958. Site quality for Douglas-fir and ponderosa pine in northwestern California as related to climate, topography, and soils. Proceedings, Society of American Foresters 1958, pp. 167-171.

Appendix A

Common and scientific names of plant species on the study area.

Common Name Used in Text	Scientific Name According to Little, 1953
Trees	
Striped maple	*Acer pensylvanicum* L.
Red maple	*Acer rubrum* L.
Black birch	*Betula lenta* L.
Grey birch	*Betula populifolia* Marsh.
American chestnut	*Castenea dentata* (Marsh.)Borkh.
Flowering dogwood	*Cornus florida* L.
White oak	*Quercus alba* L.
Northern red oak	*Quercus rubra* L.
Scarlet oak	*Quercus coccinea* Muenchh.
Scrub oak	*Quercus ilicifolia* Wangenh.
Chestnut oak	*Quercus prinus* L.

Common Name Used in Text	Scientific Name According to Gleason and Cronquist, 1963
Shrubs and ground vegetation	
Speckled alder	*Alnus rugosa* (DeRoi)Spreng.
Sarsparilla	*Aralia nudicaulis* L.
Bush honeysuckle	*Diervilla Lonicera* Mill.
Wood fern	*Dryopteris austrica* var. *spinulosa* (Muell.)Fiori
Huckleberry	*Gaylussacia baccata* (Wang.)K.Koch and *Vaccinium stamineum* L.
Mountain laurel	*Kalmia latifolia* L.
Sweet fern	*Myrica asplenifolia* L.
Blueberry	*Vaccinium angustifolium* Ait. and *Vaccinium Lamarckii* Camp.

Appendix B

Computer programs used in analysis.

Analysis	Program Name	Source
Discriminant Analysis	BMD 07M, Stepwise Discriminant Analysis	Dixon, 1971
Cluster Analysis	HGROUP	Veldman, 1967
Multiple Regression Analysis	BMD 02R, Stepwise Regression	Dixon, 1971
Canonical Correlation Analysis	CANONA	Veldman, 1967
Factor Analysis	FACTOR	Veldman, 1967

VITA

Name: Constance Ann Harrington

Date and place of birth: December 31, 1951
Bryn Mawr, Pennsylvania

Education:

	Name and location	Dates	Degree
High School	Jamesville Dewitt High School Dewitt, New York	1963-1969	
College	University of California at San Diego LaJolla, California	1969-1970	
	SUNY College of Environmental Science and Forestry Syracuse, New York	1971-1973	BS

Employment experience:

Employer	Dates	Position
SUNY College of Environmental Science and Forestry	1973 to present	Graduate Assistant
Harvard Black Rock Forest	Summer 1973	Research Assistant