Ordination and classification of Swiss and Canadian coniferous forests by various biometric and other methods

by H. van Groenewoud

Contents

1. Zusammenfassung/Abstract 29

2. Introduction 30

3. Nomenclature and terminology 32

4. Theoretical considerations 32

4.1 Description 32
4.2 Organization (ordination and classification) 34
4.3 Interpretation 47
4.4 General considerations 50

5. Methods 51

5.1 Vegetation 51
5.1.1 Sampling 51
5.1.2 Variance analysis 52
5.1.3 Relationship between distribution of light levels and plant species, within sample plots 52
5.1.4 Classification of sample plots according to the Zürich-Montpellier method 52
5.1.5 Principal component analysis of the covariance matrices 52
5.1.6 Computation of the D^2 matrices 52
5.1.7 Analysis of the D^2 matrices 53
5.1.8 Principal component analysis of the transformed D^2 matrices . . . 53
5.1.9 Cluster analysis 53

5.2 Habitat features 53
5.2.1 Swiss data 53
5.2.2 Canadian data 53

5.3 Interpretation 54

6. Results 54

6.1 Sampling and small scale distribution 55
6.1.1 Comparison of vegetation sampling methods 55
6.1.2 Small-scale non-random pattern within sample plots 57
6.1.3 Distribution of the vegetation along line intercepts in relation to light 57

6.2 Swiss sample plots 58
6.2.1 Classification of sample plots (Zürich-Montpellier method) 58
6.2.2 Habitat factors in relation to classification of sample plots 59
6.2.3 Principal component analysis (R-method) of the covariance matrix 59
6.2.4 Habitat factors in relation to the principal axes 62
6.2.5 Analysis of the D^2 matrix 63
6.2.6 Habitat factors in relation to the main axis (D^2 matrix) 63
6.2.7 Principal component analysis (Q-method) of the transformed D^2 matrices 65
6.2.8 Habitat features in relation to principal axes (Q-method) 66
6.2.9 Clustering of sample plots 66
6.2.1 Habitat factors in relation to the clustering of sample plots . . . 66
6.3 Canadian sample plots 68
6.3.1 Classification of sample plots (Zürich-Montpellier method) 68
6.3.2 Habitat features in relation to classification of sample plots . . . 68
6.3.3 Principal component analysis of the covariance matrix 70
6.3.4 Habitat features in relation to the principal axes 71
6.3.5 Analysis of the D^2 matrix 75
6.3.6 Habitat features in relation to the main axes 75
6.3.7 Principal component analysis (Q-method) of the transformed D^2 matrix 76
6.3.8 Habitat features in relation to the principal axes (Q-method) 77
6.3.9 Clustering of sample plots 77

7. Discussion 79

8. Conclusions 86

9. Acknowledgements 91

10. References 92

Appendix I Vegetation table, Swiss data . . . . . after page 94
Appendix II Vegetation table, Canadian data . . . . . after page 94
Appendix III Computational scheme for finding clusters (Swiss data) 95
Appendix IV Computation of Mahalanobis' generalized distance (D^2) 96
Appendix V Principal component analysis 97

1. Zusammenfassung/Abstract

Mehrere Methoden wurden angewendet, um Vegetationsaufnahmen von *Abies*-Wäldern in der Schweiz und *Picea*-Wäldern in Saskatchewan, Kanada, zu "ordinieren" (d.h. nach bestimmten Gradienten zu ordnen) und zu klassifizieren. Verwendet wurden: 1. "Principal component analysis of the covariance" (zwischen Arten) Matrix; 2. eine Ordinierungsmethode, die auf dem "verallgemeinerten Abstand" von Mahalanobis (D^2) basiert; 3. "Principal component analysis" von der transformierten D^2 Matrix; 4. die Zürich-Montpellier-Methode der Differentialarten-Gruppen und 5. eine Gruppenanalyse, die auf der Ordination beruht. Die ersten drei sind reine Ordinierungsmethoden; die vierte erstrebt vor allem eine Typisierung und Klassifizierung; auch die fünfte arbeitet gruppierend.

Die Beziehungen zwischen Standortsfaktoren und den Achsen der Ordinationen sind sehr ähnlich bei allen Ordinierungsmethoden. Alle Achsen vertreten bestimmte ökologische Faktoren oder Faktorenkomplexe. Die Achsen der Ordinationen, die aus der "Principal component analysis of the covariance" Matrix resultieren, repräsentieren

jedoch einen grösseren Anteil der Variation und sind näher korreliert mit Standortsfaktoren als die Achsen der anderen Ordinationen.

Die Gruppenanalyse (cluster analysis) beruht auf dem "verallgemeinerten Abstand" (D^2). Sie erlaubte zwar eine Klassifizierung der schweizerischen, nicht aber der kanadischen Vegetationsaufnahmen. Unterschiede der Mittelwerte von Licht- und pH-Verhältnissen in diesen Gruppen sind in den meisten Fällen statistisch gesichert.

Die Methode der Differentialarten-Gruppen ergab eine Klassifizierung sowohl der schweizerischen als auch der kanadischen Aufnahmen. Die Gruppen der schweizerischen Aufnahmen waren (statistisch gesichert) verschieden in bezug auf ihr pH, aber nicht auf die Lichtverhältnisse. Die Gruppen der kanadischen Aufnahmen unterschieden sich in keinem der untersuchten Standortsfaktoren (statistisch gesichert).

Die Kombination zweier Methoden, einer ordinierenden mit einer klassifizierenden oder einer Gruppenanalysemethode, die auf der Ordination beruht, erwies sich als wirksamste objektive Arbeitsweise für die Ordnung der Vegetationsaufnahmen.

Vegetation data collected in *Abies* forests in Switzerland and *Picea* forests in Saskatchewan, Canada, were organized by three ordinating methods, 1. principal component analysis of the covariance matrix, 2. an ordinating method based on the D^2 statistic, and 3. principal component analysis of the transformed D^2 matrix, and by two grouping methods, 1. the Zürich-Montpellier method of differential species-groups, and 2. a cluster analysis based on an ordination.

The relationships found to exist between habitat features and the axes of the ordinations follow the same pattern for all ordinations. All axes were ecologically significant. The axes of the ordination resulting from the principal component analysis of the covariance matrix, however, account for a larger portion of the variation and show a closer relationship with the habitat factors and other site features than the axes of the other ordinations.

Cluster analysis, based on the D^2 statistic, produced grouping of the Swiss vegetation samples but not of the Canadian samples. Differences between mean levels of light and soil pH in those groups were in most instances statistically significant.

The Zürich-Montpellier method distinguished groups of sample plots in both the Swiss and Canadian data. In the case of the Swiss data, the groups of plots were significantly different with reference to soil pH, but not with reference to light conditions. The groups of Canadian sample plots were ecologically not significantly different.

The combination of an ordinating technique with a classification technique or a cluster analysis, based on the ordination, was shown to be a powerful, objective method for vegetation analysis.

2. Introduction

In an earlier study (VAN GROENEWOUD 1965) of the ecological conditions associated with the occurrence of a root-rotting disease complex in white spruce *(Picea glauca)* stands in Saskatchewan, Canada (VAN GROENEWOUD 1956), an attempt was made to design a classification of plant communities containing white spruce, which would show a significant correlation with certain habitat factors and other site features.

The classification of these communities was considered of extraordinary importance as it would aid in the study of the conditions prevailing in the different community-types and it would help to define the habitat conditions at the time the trees become infected. The disease is usually discovered only

after trees have died, by which time the ecological conditions (e.g. light and water) often have changed (VAN GROENEWOUD 1965). A classification also would be valuable by describing the vegetation-types associated with the disease complex, which would be of great help in locating disease prone areas.

A tentative classification of the vegetation samples into five groups based on physiognomy, statistically significant associations between species, ratio between percentage moss and herb cover, and soil texture was drafted. The averages of the quantitative measures of several habitat features of these groups were compared by "t" tests. If the samples proved to be drawn from different populations these groups were accepted as distinct entities with regard to the factors investigated. The groups are comparable to Poore's noda (POORE 1955, a, b).

Of the five groups originally recognized only three were retained, i.e., the *Equisetum pratense* community-type, the *Equisetum arvense* community-type, and a very variable upland type, typified by a more or less developed moss-herb-shrub vegetation.

The last community-type represents at least 80% of all white-spruce stands present in the area investigated. It contains the stands that suffer most from the forementioned disease complex. It appeared essential to investigate the possibilities of improved methods to order vegetation samples, in such a manner that an ecologically significant system does arise.

The object of this study was to evaluate the potentialities of different approaches, to elucidate the principles involved, and to demonstrate the potentials of different combinations of various biometric and other methods presented in this dissertation.

No attempt was made to explain, in detail, the mathematical procedures involved, because most ecologists are not experts in mathematical statistics. The separate methods are more fully described elsewhere (HARMAN 1960, IHM 1964, RAO 1952). It is, however, of vital importance to the ecologist to know what is achieved by using these statistical procedures and to understand how the analysis is performed. Therefore, emphasis has been placed on explaining the principles involved, wherever possible elucidating these principles by geometrical representations. The methods are illustrated with an example of principal component analysis of hypothetical data in appendix V.

The data were collected in forests in Saskatchewan, Canada, and in Switzerland. The Canadian sites are located at Candle Lake ,approximately 65 miles north-east of Prince Albert at 54° latitude and between 105° and 106° longitude. The landform is the Wappewekka Hills Upland, represented by gently to strongly rolling morainic plains with an elevation of 1800 to 2500 feet

(Acton *et al.* 1960). The precipitation is approximately 15.5 inches (38.75 cm) of which approximately 6 inches (15 cm) falls during the summer. The soils are of the grey-wooded type, and generally belong to the «Waitville association» (Mitchell *et al.* 1950). The region is part of the Mixedwood Section of the Boreal Forest (Rowe 1959).

The Swiss forests studied are located in the proximity of Roggwil, Langenthal (Kanton Bern) and Murgenthal (Kanton Aargau), between 47°21′ and 47°16′ latitude and between 7°48′ and 7°54′ longitude. They are situated on moraine deposits of the Riss period. The soils are mottled podzolized brown earths, and pseudogleys of the brown earth group (Pallmann *et al.* 1943). The precipitation is approximately 116 cm of which 49% (56.8 cm)[1] occurs in the summer (Meyer 1949). These forests belong to the *Querco-Abietetum* and partly to the *Melico-Fagetum* (Frehner 1963). Some of these forests were described by Meyer as *Mastigobryeto-Piceetum abietosum* (Meyer 1949, 1954).

3. Nomenclature and terminology

The nomenclature of Binz-Becherer (1961) was followed for the *Pteridophyta* and the *Spermatophyta* in Switzerland. The nomenclature of Bertsch (1959) was used for the Swiss *Musci*.

Where possible the nomenclature of Fernald (1950) was followed for the Canadian *Pteridophyta* and *Spermatophyta*; elsewhere, Rydberg's (1954) nomenclature was followed. The nomenclature of Grout (1928–1940) was used for the Canadian *Musci*, with the exception of *Calliergonella schreberi*, which is replaced by *Pleurozium schreberi* (Willd) Mitt.

The terms principal component, principal factor, and principal axe have the same meaning. The term factor, however, can easily be mistaken in ecological work for a habitat factor, which it is not. Therefore, the term principal factor is not used in this publication. The term factor is used exclusively in the sense of habitat factor. Wherever other features of these habitats or plant communities were included in the analysis (e.g. height-growth of the white spruce trees, nitrogen content of the white spruce foliage) the term features is used.

4. Theoretical considerations

Investigations of the ecology of vegetation can be divided into three stages (Ellenberg 1954):

(1) Description;
(2) Organization (ordination and classification);
(3) Interpretation.

4.1 *Description*

A sample consists of a small portion separated from some large population, about which certain information is sought. The problem is to gather adequate

[1] 116 cm = 46 inches, 56,8 cm = 22,6 inches

information from each vegetation-type and each sample plot, to detect significant, rather than accidental, trends and differences. This raises the question of delimitation of sample areas, sampling units and number of samples to collect. If too many data are collected, this generally lowers the quality both of the important and of the unimportant data. Certain rules can and must be set up to enable the field worker to decide without much difficulty, whether a sample belongs to the population to be sampled.

In order to contain reliable information about the population, each member of the sample (sampling unit and sample plot) must be selected *at random.*

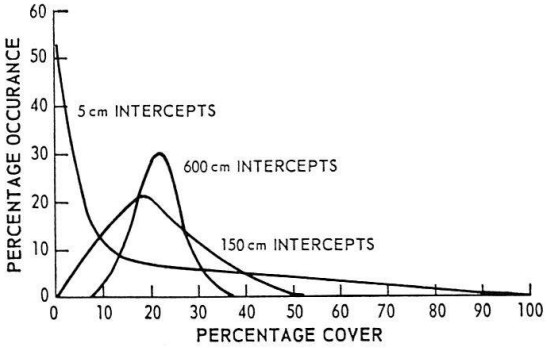

Fig. 1. Smoothed frequency curves indicating the increasing normality of the distribution of the cover percentages with increasing length of line-intercepts.

This random selection implies that each point within the sampling area has the same chance of appearing in the sample.

With a normally distributed estimate, the whole shape of the frequency distribution is known, if information is available about the mean and the variance. If the individuals or groups of individuals are not distributed at random, and if the sampling unit is approximately of the same size as the individuals or groups of individuals, the frequency distribution of the estimate of the cover percentages is not normal.

In several instances, it has been found that non-random patterns of different scales can be present simultaneously in one vegetation-type (Greig-Smith 1961, Kershaw 1958, 1959).

It is important that the vegetation data contain information about the spatial distribution of the plant species of each sample plot, for the following reasons: (1) From the statistical point of view, it is desirable that the size of the sampling unit be such that, (a) the data are normally distributed (Fig. 1), because then the shape of the frequency distribution is known, if the mean and the variance are known and, (b) the variance is relatively little affected by small deviations in the size of the sampling unit (see Figs. 8 and 9). (2) If

a small scale non-random pattern within the sample plot is present, this can be investigated and elucidated. When the material is far from uniform, as in vegetation studies, the method by which the sample is obtained is crucial and the study of techniques that insure an adequate sample becomes important. Different sampling techniques should be tested for efficiency. One way of fulfilling the combined requirements is by a method first proposed by GREIG-SMITH (1952) and later refined by him and Kershaw, the use of contiguous quadrats or other contiguous systematic samples (point-quadrat-line, line-intercept, KERSHAW 1958).

4.2 *Organization (ordination and classification)*

The basic task of organizing vegetation data is to simplify them, so that a relatively simple model of the vegetation emerges. This simplification can, of course, only be successful if the distribution of the species is governed by a few factors of overriding ecological importance, or if the quantities of certain species on the various sample plots are correlated. To attain the above goal, many different approaches were developed.

One of the most widely used methods is that of the Zürich-Montpellier School. As the result of the recognition of the limited usefulness of the "characteristic species" concept, the methods employed by the disciples of this school to differentiate between the various plant community-types have gradually developed towards the use of the differential species-group (SCHÖNHAR 1953, ELLENBERG 1956, SCHLÜTER 1957, FREHNER 1963).

Each differential species-group is made up of species which occupy identical, or nearly identical, ranges within moisture, pH and other gradients (SCHLÜTER 1957, p.48). ELLENBERG (1963 p.84) presents an elaborate scheme of the species-groups of the forest plants of Central Europe. The sample plots are then classified according to the occurrence of these species-groups. A classification based on the occurrence of species belonging to these differential groups, of course, does not need to assume a basic discontinuity in the pattern of distribution of the vegetation as a whole, as has often been implied by critics of the Zürich-Montpellier school. It devides the vegetational catena into segments, which are typified by the presence or absence of the species belonging to these differential groups.

As a result of the development of large and fast electronic computers, it has become feasible to make use of multivariate statistical methods, such as factor analysis, discriminant analysis and Mahalanobis' generalized distance, to organize phyto-sociological and ecological data.

A concept which is fundamental in considering many variables together is

the test-space. If measurements have been made of the cover (or any other attribute) of m species on N sample plots, the data can be presented in matrix form as follows:

Spec. \ Plot	1	2	.....	N
1	x_{11}	x_{21}	.....	x_{N1}
2	x_{12}	x_{22}	.....	x_{N2}
⋮	⋮	⋮		⋮
m	x_{1m}	x_{2m}	.....	x_{Nm}

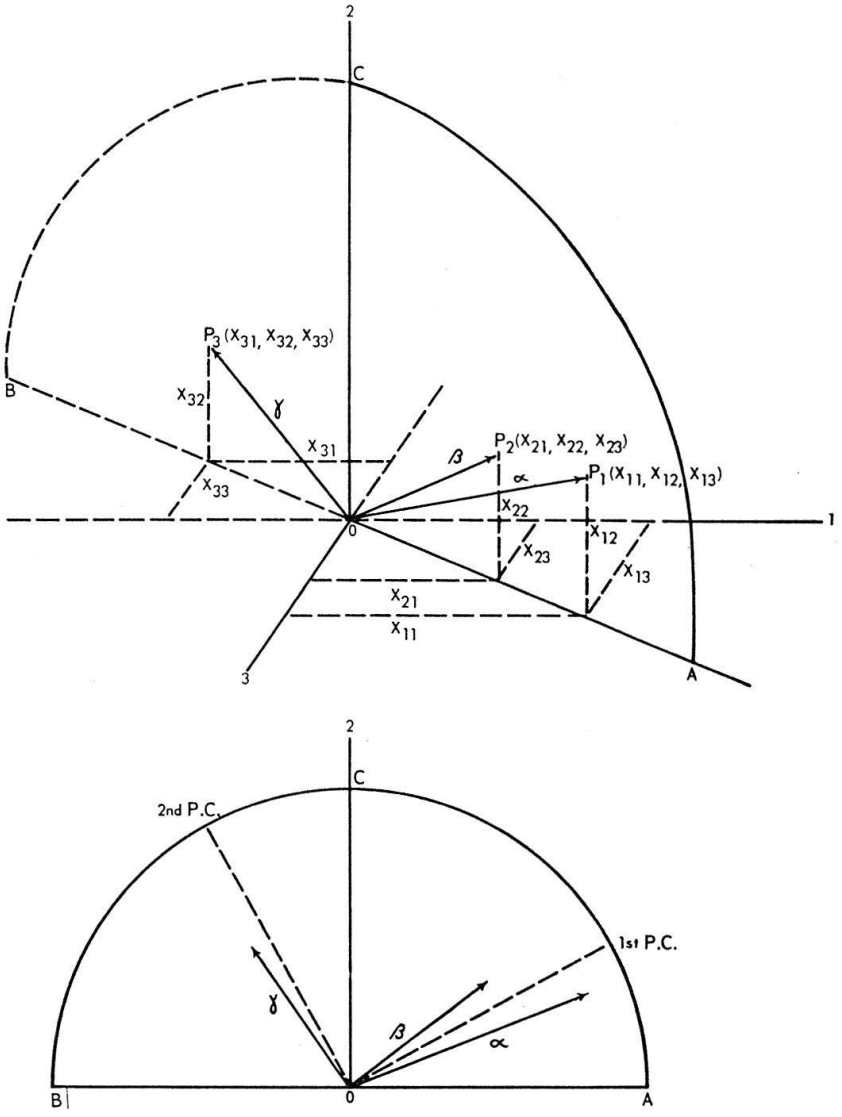

Fig. 2. Top: Vector representation of plant species (explanation in text).
Bottom: Principal components (P.C.) of above vector representation.

These data can be presented geometrically in the forementioned test-space in two different ways:

(1) The vector representation. Each species is represented by one vector (see appendix V) in an N-dimensional test-space (as many axes as there are sample plots). Each vector is determined by two points: the origin of the test-space, and the end point of the vector, which is determined by an ordered set of coordinates, e.g. for species 1 (x_{11}, x_{21}, ... x_{N1}). The cosines of the angles between the vectors are equivalent to the coefficients of correlation between the respective species. A three-dimensional example is presented in Fig. 2. For the sake of simplicity the vectors denoted by α, β and γ, representing species 1, 2 and 3, are shown as lying in one plane ABC.

The interpretation of an $m \times m$ matrix of indices, expressing relationships between species, is the vector representation. The test-space involved will be referred to later to as the R-space.

(2) The point representation. Each sample plot can be represented by one point in an m-dimensional test-space in which each axis represents one species (m species = m axes). So plot 1 is represented by the point having the following coordinates (x_{11}, x_{12}, ... x_{1m}). There are thus N points representing N sample plots in an m-dimensional test-space. Fig. 3 is an example of the point representation of two sample plots containing three species 1, 2 and 3. On plot 1 the species have the following cover percentages 20, 15 and 10, respectively; on plot 2 these percentages are 6, 7.5 and 5 respectively. The interpretation of an $N \times N$ matrix of coefficients of similarity between sample plots is the point representation. The test-space involved will be later referred to as the Q-space.

In the case of the vector representation, the first problem arising is the choice of the most efficient statistic of interspecific relationship. Many workers have discussed these coefficients but the most complete account has been presented by Dagnelie (1960). Statistics expressing interspecific relationships can be based on either qualitative or quantitative data. In both cases, the statistics are sensitive to changes in size of the sampling units. Until now, this fact has received little attention in ecological literature, but it should not be neglected (Kershaw 1961).

Further, interspecific association coefficients are sensitive to enlargement of the study area, where the added areas are devoid of one of the species under consideration (Bray 1956). They are, thus, most useful when the vegetation samples are fairly similar. Some ecologists believe that various species respond in the same way to various combinations of different levels of habitat factors, thus obscuring the habitat relationships. This possibility increases when a wider range of conditions is sampled. The narrower the range of habitats

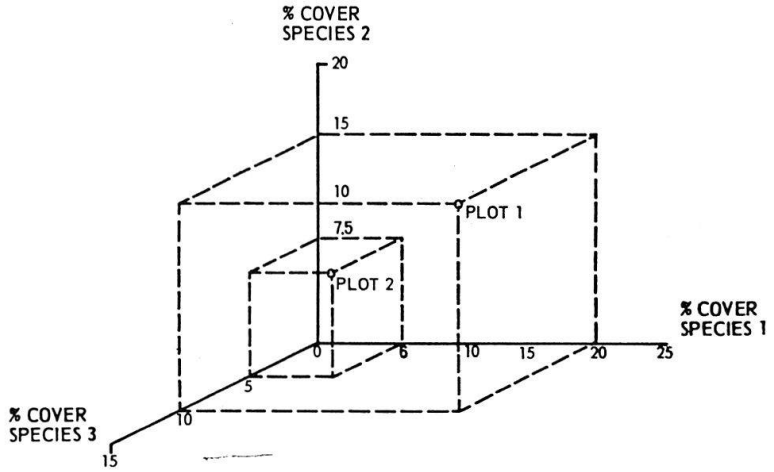

Fig. 3. Point representation of sample plots (explanation in text).

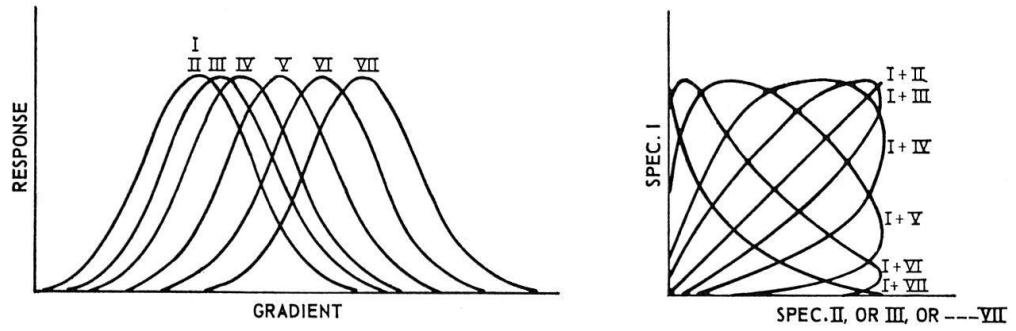

Fig. 4. Top: Hypothetical gradient response curves of seven species at six different ranges along gradient.
Bottom: Graphical representation of the quantitative relationships of species I with the other six species.

sampled, the closer the relationship between the vegetation and the habitat can be expected to be.

The quantitative relationships between species are very complex. Consider, for instance, the relationship between a number of species all of which have identical bell-shaped response curves along an ecological gradient (gradient response curves), but which are occupying different ranges along the gradient (Fig. 4, top). These are much simplified situations which are highly improbable in nature, but which can illustrate the problems under consideration. Graphically, the quantitative relationships between each pair of species can be shown by plotting the quantities of the two species occurring at each point along the gradient, on a pair of orthogonal axes, each axis representing one species (a two-dimensional test-space). The result is a very artistic family of curves (Fig. 4 bottom), each curve representing the relationship between two species. If the species occupy the same range (e.g. species I and II), the curve

representing the relationship between the species is a straight line. The curves representing the relationships between species, which occupy ranges which overlap to a lesser degree, e.g. species I and III, or I and IV, are more egg-shaped. When the ranges overlap even less, the curves become triangular finally to change to parabolic shapes. Attempts made by this author to develop a single formula which satisfactorily describes these relationships were not successful. In nature several ecological gradients can be expected to act concurrently. Some of these may be correlated; others not. Further, all shapes of response curves may be expected, especially where species affect one another (ELLENBERG 1953, 1956, 1963). Clearly the foregoing example is very much simplified.

By assuming linear relationships between species, valuable information may be lost, but as far as this author is aware, no mathematical methods based on curvilinear relationships have been developed. An examination of forementioned curves, however, reveals that, if sampling is restricted to a narrow range of conditions (a small piece of the gradient), the relationship between the species can be represented by a short segment of the curve, thus more closely approaching linearity. Thus, in order that the interspecific-association indices (or indices expressing relationships between vegetation samples) will yield as much information about habitat and interspecific relationships as possible, the range of habitats to be sampled should be kept fairly narrow.

After the choice of the statistic of interspecific relationship has been made, two basic problems arise:

(1) Is it possible to replace the initial N axes in the original test-space by a few variables (components), which account for practically all the variance? The answer depends on the correlations between the species (angles between the vectors). If the m vectors can be divided into groups of vectors, in such a manner that within a group the angles between the vectors are small, but with large angles between the vectors of different groups, the replacement of the N axes by a small number of new, mutually orthogonal, components (as small a number of components as there are groups of vectors) should be possible. This is accomplished by principal component analysis.

A simple geometrical illustration of the replacement of a number of axes by two components is presented in Fig. 2. This concept can be considered a refinement of the differential species-group concept. Whereas the species-group rests on the co-occurence of certain species, the component is determined by the quantitative relationships (both positive and negative) among species.

(2) In the event that mutually orthogonal components have been found, it still remains to be determined whether they are ecologically significant.

This, however, is a matter of interpretation, which will be dealt with in the section under that heading.

If mutually orthogonal components can be found and estimates of the amount that each species contributes to the various components are available, it is then possible to calculate the coordinates of each plot along the components by multiplying the quantity of each species by its corresponding coefficient and summing these new values. The result is an ordination of the sample plots in the new test-space as delineated by the components, or in other words, *through a vector representation of the species, a point representation of the sample plots is obtained.*

The following question now arises: Do the points in the new space form a set of clusters? There is no unique way in which a cluster can be defined. The judgement of the research worker forms the ultimate criterion in the appraisal of the value of these definitions. In this study a cluster is defined as a group of samples, whithin which the samples are more similar to one another than to vegetation samples outside the cluster. A cluster would represent a unit in a classification system.

To determine whether the points form clusters, different methods can be employed. If only a few clusters are involved, an examination of the projections of the points, representing sample plots, on the planes spanning the components is adequate. If this is not satisfactory, other methods are available, most of which require the calculation of the distances between the points in the new test-space. This is easily accomplished using the Pythagorean theorem. One method of cluster analysis, based on the distances between the points, will be discussed later.

Most ecologists, who have determined measures of quantitative interspecific relationships, have not tried to analyse these data statistically. Goodall (1954) and Dagnelie (1960) used factor analysis to analyse matrices of interspecific correlations. Goodall observed indications of clustering (a bimodal distribution) along the first principal axis in the frequency distribution of the values of the various sample plots. No cluster analysis in the strict sense, however, was made to investigate the grouping of the sample plots.

Of the different methods of factor analysis available, to determine if the original variables can adequately be replaced by a few components, mathematically the most robust method is principal component analysis, as first developed by Hotelling (1933).

Ihm (1964) developed a computer program, for principal component analysis of covariance matrices, which is particularly suited for vegetation analysis. It does not only compute the trace (= sum of all eigenvalues), the

eigenvalues, and the eigenvectors of the covariance matrix (see appendix V) but also calculates, for each sample plot, the value of each component (the coordinates of each sample plot) and plots the projections of the points, representing the sample plots, on the planes spanning the first and second axes, and the first and third axes. The eigenvalues indicate the relative importance of the species combinations (principal components), represented by the coefficients of the associated eigenvectors. From the eigenvalues it is possible to calculate the percentage of the total variation accounted for by each component (see appendix V). The species that have large coefficients, contributing to the eigenvectors, are said to be causing most of the variation represented in the eigenvalue.

In the case of the point representation, the first problem to be solved is the choice of the statistic expressing the similarity (or dissimilarity) between the vegetation samples. As DAGNELIE's review (1960) clearly shows, many different indices, expressing similarities between vegetation samples, have been proposed. It is obvious that similarity indices based on quantitative measures are more sentitive than those based only on presence and absence data[1]. Practically all quantitive similarity indices, however, suffer from some imperfections. Most of these indices increase rapidly with an increase in the number of species. It is not easy to determine from the changes in similarity index whether the newly added species supply additional information for the purpose of classification or ordination. If the quantities (percentage cover) of the species on two sample plots are used to calculate a similarity index which, it is hoped, will also express the similarity in habitat, and if the species are quantitatively correlated, a non-orthogonal comparison is being used, in other words the axes in the test-space are not mutually orthogonal. If the species are positively correlated (e.g. they react more or less similarly to differences in habitat, or the presence of the one species favors the growth of another species), then the similarity index indicates a greater similarity (a smaller distance) between the vegetation samples than where the species are not correlated. This phenomenon is geometrically illustrated in Fig. 5. Assume that points P_1 and P_2 represent two sample plots. The quantities of species A and B on plot 1 are respectively a_1 and b_1, and on plot 2 respectively a_2 and b_2. The species A and B can each be represented by one axis in a two dimensional test-space. If the species are quantitatively not correlated, the axes are mutually orthogonal. The distance (D) between P_1 and P_2 can now be calculated with the help of the Pythagorean theorem: $D^2 = (a_1 - a_2)^2 + (b_1 - b_2)^2$. If, however, the species A and B[1] are quantitatively correlated,

[1] Recent work by LAMBERT and DALE (Adv. ecol. Res. *2*, 1964) indicates that this may not always be true.

the axes are not orthogonal but oblique. The cosine of the angle between the axes, in this case, is equal to the correlation coefficient. If we assume that the correlation coefficient is. 682, then cos $\alpha = .682$ and $\alpha = 47°$, and the position of P_1^1 and P_2^1 can be constructed. Assuming the quantities for species A and B^1, as before, the distance between P_1^1 and P_2^1, D^1, is obviously smaller than D (D^1 = appr. .85 D). If the species are negatively correlated, the distance is greater than when the species are uncorrelated.

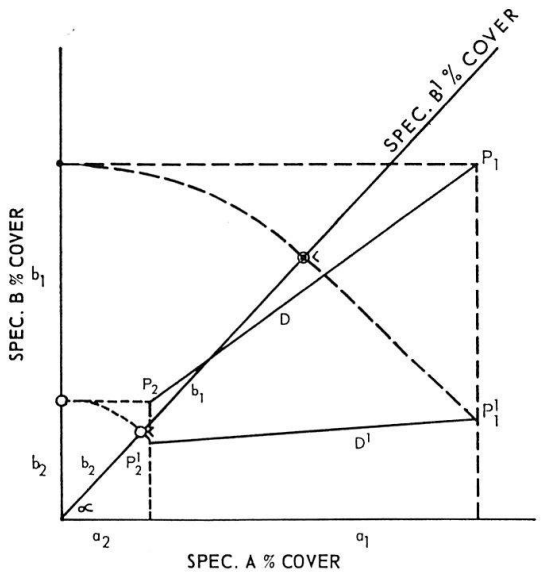

Fig. 5. Distance D, if species quantities are not correlated, compared to distance D^2 when the correlation coefficient between the species is .68 ($D = 1.18\ D^2$).

The problem created by the quantitative correlations between species is taken care of by the D^2 statistic. This statistic was first proposed by Mahalanobis (1936) and used by him and co-workers (Mahalanobis *et al.* 1949) for comparisons between ethnic groups in India. Hughes (1954), first applied it in plant ecology to test the differences between groups of sample plots, which had been established by criteria other than the vegetation. The D^2 statistic is used in this investigation as an index of dissimilarity between separate sample plots.

The D^2 statistic is a *measure of distance* (reciprocal of similarity), rather similar to the more familiar Student's "t". In fact the D^2 reduces to the "t" if comparisons are made with one variable only. The D^2 statistic, also called Mahalanobis' "generalized distance", is best illustrated by a geometrical figure (Fig. 6).

For example, suppose the dissimilarity between two sample plots (stands) is to be calculated. For the sake of simplicity, presume that only two species

are present (A and B, Fig. 6). On plot 1, species A has a mean cover, expressed in standard deviation units, of a_1; on plot 2, the cover is a_2. On plot 1, species B has a mean cover of b_1; on plot 2, the cover is b_2. The location of each plot in two-dimensional space is fixed by the coordinates, a_1 and b_1 for plot 1, and a_2 and b_2 for plot 2. The distance between plot 1 and 2 can now be calculated using the Pythagorean theorem,

$$D^2 = (a_1 - a_2)^2 + (b_1 - b_2)^2.$$

These D^2's can be tested for signifiance.

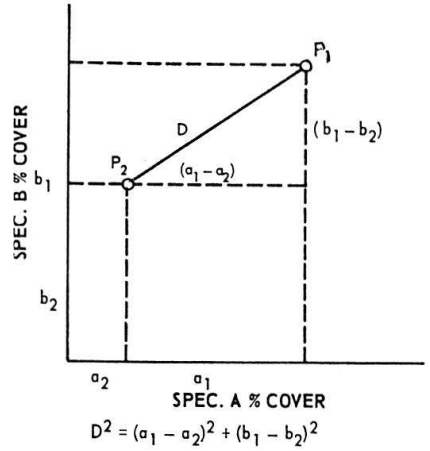

Fig. 6. Distance between two sample plots in two-dimensional space (species not correlated).

The method can be extended for 3, 4, or more species by adding mutually orthogonal axes to the two originally used, one axis for each species added. The formula for D^2 now becomes,

$$D^2 = (a_1 - a_2)^2 + (b_1 - b_2)^2 + (c_1 - c_2)^2 + (d_1 - d_2)^2 + \ldots$$

This formula is valid only if the variables (quantitative measures of species) are not correlated.

If the variables are correlated, as is usually more or less the case with different species, they can be replaced by a set of transformed variables, which are linear functions of the observed variables and are mutually uncorrelated (Mahalanobis *et al.* 1949; see also appendix IV). The location of each plot is now fixed in multidimensional space (as many axes as there are species) and the distances between these plots and their probability levels are known.

It seems opportune to discuss at this point the different concepts held by various ecologists on the structure and spatial distribution of plant communities.

Although recognizing the continuous nature of vegetation (Ehrendorfer 1954), most of the European-schooled workers, for practical purposes, assume in most cases a non-random pattern or actually a pattern closely enough resembling a non-random pattern that no large errors are made. How close this assumption approaches reality, however, is seldom proved by showing: (1) actual discontinuity in the vegetation or of the habitat factors; or (2) statistical meaningful differences between the classification units in vegetation or levels of habitat factors. The methods developed and used for grouping vegetation samples result in the division of the vegetation into segments, which may or may not coincide with significant different levels of the habitat factors or site features.

Ellenberg (1956) showed how indices of stand similarity based on species quantities, (Massen-Gemeinschaftskoeffizienten) can be used to verify the grouping of vegetation samples.

Much of the work done by ecologists of the Wisconsin school is based on the assumption that all patterns, except the largest scale patterns, are random. Different parts of the vegetational pattern form a so called continuum. A method was developed called a continuum analysis in which "adaptation numbers" (a sort of coefficient of interspecific relation, determined from a matrix of indices of amplitudinal correspondence (Bray 1956), were used to calculate a "Continuum Index" (C.I.) for each stand-sample. The C.I. is used to ordinate stands along an ecological syndrome or continuum. As such, the C.I. can be classified under the statistics expressing relationships between vegetation samples. The quantitative distribution of species along the continuum showed continuous and interchanging patterns. The conclusion was drawn that this proves the non-existence of distinct plant community-types (associations, forest-types etc.).

Maycock (1963), after grouping sample plots according to moisture classes, came to the same conclusion. He states: "When the forests were grouped together on the basis of their site moisture features and importance values for individual component trees were averaged for all the five classes, it was clearly shown that each tree has a graded specific relationship in response to the grouping or ordering of stands. When the assemblages of tree responses were considered as a whole it was evident that a forest continuum is existent (at least within the tree layer). Specific distinct forest communities are either a figment of the imagination or those segments of forest composition on particular portions of the environmental moisture series that an investigator has chosen for specific research investigations". (Maycock 1963, p. 424–425). Other ecologists working along the same lines have made similar statements (Bray *et al.* 1957, Curtis 1959).

These studies, however, elucidate only the relationships between species and between species and habitat factors, and do not take into account the spatial distribution of the different communities. If an area that is covered by a non-randomly distributed vegetative cover is sampled at random, the samples can be placed in groups in such a manner that, in each group, the samples are more similar to one another than to samples outside their group, and yet an analysis of the type mentioned before will not show discontinuity in the quantitative distribution along gradients, but unbroken, smoothly intergrading patterns.

A non-random distribution at different levels of occurrence, can have different causes: (1) morphological properties of the species; and (2) non-random distribution of the different levels of the controlling habitat factors.

Most ecologists have considered only the co-occurence of species as a basis for classification. BRAY (1956) states: "The floristic basis of community classification, both European and American, is dependant upon the fact that certain species tend to occur together". The spatial aspects of vegetation distribution, however, should definitely be a major part of the considerations.

It is quite possible that individual species have smoothly graded response curves over a wide range of several gradients, with their optima at different positions along the gradients. Yet, classification is possible, because the spatial distribution of the different levels of the habitat factors (and as a result that of the vegetation as a whole) is not random. This can be shown by studying the frequency distribution, if the sample plots can be adequately ordinated along one axis (one-dimensional space), or by cluster analysis if the plots are ordinated along several axes (multidimensional space).

If groups have been established in another manner, e.g. as was done by MAYCOCK (1963), using moisture classes, these groups can be tested for vegetational differences by multivariate methods, e.g. Hotelling's T or Mahalanobis' D^2.

The next problem to be solved, is whether the points in the forementioned multi-dimensional space form clusters. Several types of cluster analyses are possible. One of the simplest was, according to Rao, suggested by Tocher (RAO 1952). The principle is as follows: One starts with two plots which are most similar and finds a third plot which is closest to the first two, by determining which has the smallest average distance from the first two plots. The fourth plot is chosen by determining the smallest average distance from the first three plots, etc., etc. If at any stage the average distance increases substantially, (this is a subjective criterion), the last plot is considered not to belong with the group of sample plots that has been processed. The data of the group in question are removed from the general set and the rest is treated as before (Appendix III).

In cases where the vegetation is continuously variable, in other words, where all samples form one cluster, the investigator is left with a constellation of points in multi-dimensional space, which cannot be utilized in this form. The possibility, however, remains that the variation within the set of vegetation samples can be adequately explained by a few variables instead of the original large number m. These new variables may or may not be related to ecological features.

This problem can be attacked by different ordinating techniques. Two procedures can be followed. The first originated with TORGERSON (1952) and was used with apparent success by BRAY and CURTIS (1957) in their ordination of upland forest communities of southern Wisconsin. They have described this method in detail: "This technique depends upon the selection of a pair of reference stands for the determination of stand positions on any one axis. Given proximate interstand distances, the choice of reference stands is of crucial importance. In making this choice it is evident that reference stands are comparable, in part, to sighting points as used in plane table surveying and that those stands which are furthest apart will be more accurate for judging interstand distances than those in close proximity".

"It is necessary for any ordination that the sphere of fluctuation for any stand be small in relation to the space occupied by the ordination as a whole. The choice of reference stands should be, therefore, of those stands which are furthest apart, and as a consequence, have the greatest sensitivity to overall compositional changes".

"To locate stands between a pair of selected reference stands, a line connecting the reference stands is drawn to scale on a piece of blank paper and the position of each other stand is projected onto this line. The projection is accomplished by rotating two arcs representing the distance of the projected stand from each of the reference stands and then projecting the point of arc intersection perpendicularly onto the axis".

"A second axis can be constructed by the same method, using a line on the paper erected at a right angle to the X axis. Two new reference stands are selected which are in close proximity on the X axis, but which are nevertheless separated by a great interstand distance".

There are, however, some drawbacks to this procedure. Where the relationships between a great number of plots are being analysed, the method becomes rather laborious. A disadvantage of this procedure is also the burdensome method used to obtain a measure of the total variation accounted for by the axes determined. At each stage the distances between the points are compared with the distances in the original matrix, and the correlation coefficients calculated.

A mathematically more sophisticated procedure is the "Q" method of factor analysis as used by DAGNELIE (1960), in particular that method of factor analysis called "principal component analysis". In this case (see appendix V) principal component analysis, also called the method of principal axis or principal factors, investigates whether the information conveyed by all the "generalized distances" between the plots may be adequately represented by fewer variates, which may be used in place of the original variates (the species quantities).

The D^2's (a distance measure) as such, however, cannot conveniently be analysed by principal component analysis. For this analysis, the D^2's preferably should be transformed into indices of similarity (R). The simple reciprocals should not be used, because they have the same theoretical limits as the D^2's, 0 and infinity.

Two transformations, which can be used were suggested by Ihm[1]:

$$a)\ R_1 = e^{-D^2}$$

$$b)\ R_2 = (1 + D^2)^{-1}$$

Either transformation can be used, because the limits of both of the new indices vary between 0 and 1, from complete dissimilarity to complete similarity.

In geometrical terms, the plots, represented by points in a multidimensional coordinate system, form a constellation of points of an hyper-ellipsoidal form. The principal component method finds the so-called principal axes of this hyper-ellipsoid, and projects the forementioned points on these. It determines first an axis along which the variance is maximum, and second an axis, at right angles to the first axis, along which the remaining variance is maximum etc. etc. Theoretically, there are as many principal axes as there were axes in the original test-space, but usually a large proposition of the total variance is accounted for by the first few (3–5) axes. This represents a convenient simplification of the original N variables. For a geometrical representation of the reduction of a three-dimensional to a two-dimensional test-space, see Fig. 7. For the sake of simplicity, assume that all points are lying in the plane through A, B and C. The original three axes can then be replaced by two new axes, going through the centre of gravity of the group of points, P, and lying in the plane A B C. These axes are the principal components sought. The method supplies the trace of the matrix (= sum of all eigenvalues), the "eigenvalues" and the "coefficients contributing to the eigenvectors". The eigenvalues give a quantitative indication of the relative

[1] Personal communication.

importance of each axis. From the eigenvalues and from the trace, it is possible to calculate the percentage of the total variation accounted for by each axis (see appendix V). The coefficients of the eigenvectors contributing to the eigenvalues denote the coordinates of the points in this "component space".

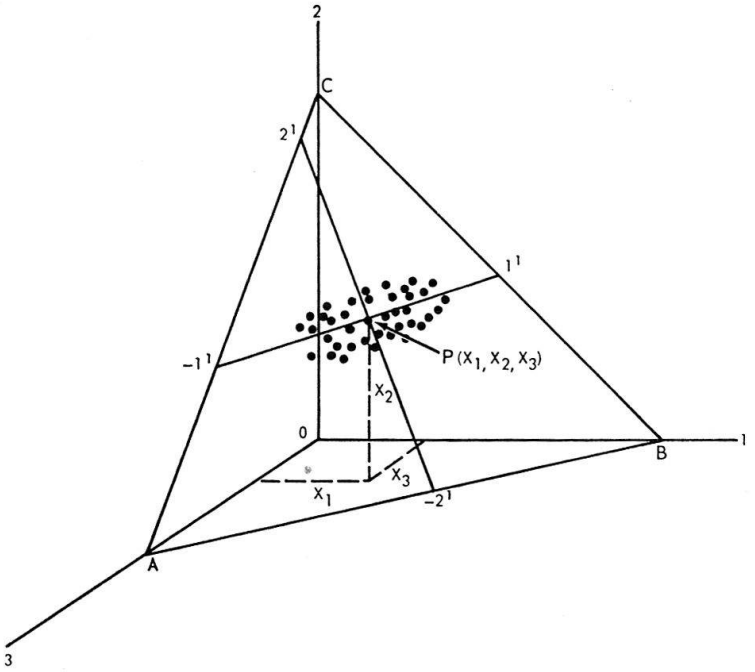

Fig. 7. Hypothetical ellipsoid shaped cluster of sample plots in three dimensional space with principal components (explanation in text).

4.3 *Interpretation*

The differential species-group concept postulates that the species belonging to such a group are present in the vegetation over a limited range of levels of certain habitat factors. It is then to be expected that a grouping of sample plots, based on this concept, results in community-types which occupy different ranges along various environmental gradients e.g. the soil moisture and pH gradients.

Associated with these different ranges along environmental gradients, one expects to find significantly different levels of one or more other site features e.g. height-growth of the various tree species, reproductive power of these species, disease susceptibility, etc.

If the sample plots have been grouped according to above-mentioned method, it remains to be decided to what extent the aim of the method has been accomplished. To this end, the habitat and other data can be subjected to statistical tests. FREHNER (1963) used the "t" test to investigate the signi-

ficance of the differences in average height, reached at an age of 100 years, of several tree species in the community-types distinguished by him. VAN GROENEWOUD (1965) used the "t" test to compare the mean levels of both habitat and other site features in various community-types. With respect to the separation of different ranges along environmental gradients, however, the "t" test is not very discriminative. It should be realized that if sufficient measurements are made on each sample plot, very small differences between the mean values of the various plots will almost always be statistically significant. The mean levels of two ranges along an environmental or other gradient can be highly significantly different ($P \leqq .001$) yet the ranges may almost completely overlap. A more satisfactory criterion for the relative separation of different ranges would be the ratio of the difference between the mean levels, Δ, to the sum of the standard deviations $\sigma_1 + \sigma_2$:

$$(\Delta/\sigma_1 + \sigma_2).$$

The outcome of the vegetation analysis, can take three possible forms:

(1) The points, representing sample plots or stands in either the R or Q space, form a cluster of hyper-spherical shape. No classification or ordination of practical value can by made. In the principal component analysis, this would result in a number of eigenvalues which vary little in size.

(2) The points do not form separate clusters, but one cluster of hyper-ellipsoid shape, the axes of which can be determined. These axes form the ordinates of an ordinating system.

(3) The points form a number of clusters. These clusters are equivalent to units in a classification system.

In the latter case, classification is a possibility, but not a necessity. The choice between classifying or ordinating depends on the requirements of the investigator.

There are no indications, so far, that the possibility mentioned under 1 is other than theoretical. It is doubtful whether it actually occurs in nature. If it does occur, a different approach to vegetation analysis could be used (e.g. gradient analysis). This will not be further considered here.

After extraction of the principal components, the ecological meaning of these components can be elucidated. The principal component analysis of the matrix of coefficients expressing relationships between plots (the point representation) can be interpreted in ecological terms as follows: the points, representing the sample plots, form a swarm (constellation or cluster) of as yet undefined shape in the original test-space. If we assume that the points form a swarm of hyper-ellipsoid form, the points (plots) farthest apart are the most dissimilar in floristic composition. It can be reasoned that this is not the

result of chance, but that this dissimilarity is caused by a concurrent dissimilarity in the habitats. It can be postulated that the axis which joins these extremes does represent a gradient of intermediate habitat conditions (usually a set of correlated habitat factors). The forementioned axis, however, represents only the variability in one direction. The reasoning in explaining the variability along the second axis is analogous to that for the first axis.

Several authors (Greig-Smith 1964, Goodall 1954) have already pointed out that the principal components (R-method) do not necessarily represent ecological factors, but are strictly expressions of relationships between the quantities of the species. It can be reasoned, however, that these relations are not just the result of chance, but are more or less caused by the fact that species are affected in the same or dissimilar manner by particular sets of habitat factors, or by the effect that one species has on the performance of the other. It should then be possible to decompose the covariances among the species into a small number of orthogonal components, each comprising a set of species, which have high covariances among them. Theoretically then, these components should be correlated with habitat factors, or better, with sets of correlated factors.

To test the hypothesis that the principal components are related to habitat features, the various levels of each habitat feature can be related to the corresponding values of the principal component (projections on the axes = coordinates). In order to avoid much unnecessary work, the level of each habitat feature for each plot can first be plotted against the corresponding values of each plot for each component, to determine if any relationship, linear or curvilinear, is present. If the graph indicates the likelihood of a relationship, this relationship can further be defined by fitting a straight line or a curve to the data.

Because the axes are perpendicular to each other, it is to be expected that, if relationships are detected, the different axes are related to habitat factors which operate independently of each other.

The habitat factors can also be included in the principal component analysis. This, however, in most cases is not desirable because these factors would also contribute to the ordination of the plots, even in cases where no actual causal relationship exists among the factors included and the distribution of the vegetation.

In the case of clustering, as in grouping according to the differential species-group method, it remains to be decided if the clusters also have ecological significance. In other words, it must be determined if the groups, besides being vegetationally different, are also significantly different in habitat.

To establish this, first the ecological meaning of the principal axes is considered. It is unlikely that significant differences could be obtained between the means of the levels of a habitat factor on various plots, if the habitat factor is not related to one of the principal axes of the ordination. The reverse is not true. This depends on the type of relationship between the principal axes and habitat factors and on the relative position of the clusters with respect to the principal axes.

Where more than two groups are involved, if the number of measurements of the habitat factor are the same for the groups to be compared and if the variances are not markedly different, the differences between the means can be tested by variance analysis. Otherwise, "t" tests with corrections for different numbers of samples and for differences in variance are indicated (Greig-Smith 1964, p.35).

In case the analysis of variance in relation to the length of line-intercept, or in relation to quadrat size, does not indicate a non-random pattern in the species distribution of the sample plot, the possibility still exists that a causal relationship between vegetation distribution and habitat factors does exist, if both are distributed at random. Within the sample plots, the variation in the levels of most habitat factors is small. There are, however, a few factors which can form random patterns in small areas, within which the levels can vary considerably. One such factor is light. If a non-random vegetational pattern can not be shown to exist, the forementioned case would be analogous to that in the D^2 analysis, where no grouping can be discovered, and yet the distribution of the species is affected by factors as indicated by the principal component analysis. This problem could be solved by the use of principal component analysis. The labor that would be involved makes this an impractical approach.

A quicker method is to plot the level of the habitat factor concerned as measured along the line-intercepts, directly against the percentage cover of each species, as measured in each section of the line intercept (Fig.12). Relationships, if any are present, will be evident in the graphs.

4.4 *General considerations*

In judging the relative merits of different methods for ordinating and classifying vegetation samples, the following should be considered:

(1) The method must result in the *simplest possible* ordering of the vegetation samples which should account for the *largest possible* portion of the variation within these samples.

(2) Both the ordination and the classification should preferably be related to habitat factors;

(3) Ideally, the method should be based on a statistic which forms both an objective criterion to determine whether classification is possible, and a basis for an ordination, which can be used as an alternative in describing the vegetation, in case it is continuously variable,

(4) The method should furnish a means of placing newly measured vegetation samples in a previously derived ordination or classification system, without going through the whole analysis each time; and

(5) If the objectives of 1 and 4 can be accomplished, it should then be possible to devise a system that will allow the mapping of vegetation samples which tend to form a continuum. This would greatly increase the usefulness of the method.

5. Methods

5.1 *Vegetation*

5.1.1 Sampling

Location of sample plots.

As mentioned before, ideally the samples should be located at random. In practice, however, this is subject to limitations. In locating the sample plots, the following conditions were adhered to: (1) Each sample plot with surrounding area should be undisturbed; (2) it should be representative of a sizeable part of the stand in which it is located; and (3) it should not cross any obvious transition zones or boundaries in the vegetation.

This set of rules greatly limited the number of sites available, especially in Switzerland, where the forest was severely damaged by a heavy snow-fall early in this study. It is thought that the sample plots chosen represent a fairly random sample of the forests in both localities (Switzerland and Canada).

The samples (line intercepts, quadrats) were located at random, within the boundaries of each plot (10×10 m in Switzerland and 50×50 ft in Canada).

Four methods were employed in measuring the vegetation:

(1) The Braun-Blanquet method

In each sample plot a complete list was made of all species present, with an ocular estimate of their cover and abundance (Appendices I and II).

(2) The contiguous quadrat method

Ten randomly distributed, 1-square-meter quadrats each divided into 16 equal squares with sides of 25 cm, were used. The percentage cover was estimated for four different sizes of quadrats; 160 quadrats of 25×25 cm, 40 quadrats of 50×50 cm, 10 quadrats of 75×75 cm and 10 quadrats of 100×100 cm.

(3) The contiguous point-quadrat-line method

This is the method developed and used by Kershaw and called by him, the line-interception method (Kershaw 1958). The name used here was created to differentiate it from the line-interception method as devised originally by Canfield (1941).

A frame, consisting of two parallel sheets of plexiglass (24×10 cm) bolted together, approximately 8 cm apart, with a series of 20 holes, 1 cm apart, along either edge of the plexiglass, was used (Kershaw 1958). By pivoting this frame around a pin through either end of the frame, it was possible to take contiguous readings of the vegetation along

straight lines. Adjacent readings were grouped into units of 5 cm. The length of the lines was 6 m, for a total of 60,000 points per plot.

(4) The line-interception method (CANFIELD 1941)

A tape-measure was stretched tight between two pins, 6 meters apart. For each species, the beginning and the end where the line crossed the species, were recorded to the nearest mm. The data were tape-recorded and later reconstructed on paper. The line was divided into contiguous pieces, 5 cm long. Then, the percentage cover for each species over each 5 cm of the transect was determined.

5.1.2 Variance analysis

To determine that size of quadrat, or length of line (point-quadrat-line or line-intercept), that was most efficient in decreasing the variance, the variance was calculated for each size of quadrat (respectively 25×25 cm, 50×50 cm, 75×75 cm and 100×100 cm) and for different lengths of line-intercept (respectively 5, 10, 20, 40, 80, 160, 300, and 600 cm long). The variance was then plotted against quadrat size or length of line.

This variance analysis technique also makes it possible to check whether the species form a small scale non-random pattern (GREIG-SMITH 1952), because the variance rises to a peak in the graph at a sampling-unit size equivalent to the size of the pattern.

In some cases, the frequency distributions of the cover percentages were plotted for different sizes of line-intercept to show the increased normality of distribution with increased length of line-intercept (Fig. 1).

By plotting the increase in number of species encountered through increases in the number of quadrats or lines, the efficiency of these methods for producing estimates of the cover of as many species as possible was estimated.

5.1.3 Relationship between distribution of light levels and plant species, within sample plots

For each plot, the levels of light as measured by two different methods along five line-intercepts was plotted against the percentage cover of the more abundant species as measured on contiguous 10 cm sections along these lines.

5.1.4 Classification of sample plots according to the Zürich-Montpellier method

The species and sample plots were arranged and grouped in such a manner that groups of species (differential species-groups) which would occur only in a certain group of sample plots (see Appendices I and II) became evident. The sample plots were classified according to the presence or absence of the species belonging to these differential species-groups.

5.1.5 Principal component analysis of the covariance matrices

For reasons later explained, only the line-intercept data were used for computing the principal components and the D^2's.

The actual analysis was performed by IBM 7090. The program was developed and written by Dr. Ihm of the Euratom Research Centre at Ispra, Italy (1964).

5.1.6 Computation of the D^2 matrices

The D^2's (Mahalanobis' generalized distances) between the plots and their significance levels, were computed by IBM 7090. The program was developed and written by Dr. P. Ihm.

5.1.7 Analysis of the D^2 matrices

A two or three dimensional ordination was constructed from the D^2 data according to Torgerson's method. The correlation coefficients and their significance levels were calculated among the distances in the two dimensional projections and the D^2's. The square power of these correlation coefficients indicate the percentages of the total variation accounted for by the ordinations.

5.1.8 Principal component analysis of the transformed D^2 matrices

As mentioned before a D^2 matrix can not conveniently be analysed by principal component analysis. Two transformations were performed:

$$R_1 = (1 + D^2)^{-1}; \text{ and } R_2 = e^{-D^2},$$

with the help of a G-20 computer.

The transformed D^2 data were analysed by the Statistical Research Service of the Canada Department of Forestry in Ottawa, using a G-20 computer.

As stated before, the method is explained in appendix V and will not be treated here.

5.1.9 Cluster analysis

The projections of the points on the planes spanning the three principal axes were examined for indications of clustering. The projections on the planes spanning the three main axes of the constellation of points described by the D^2 matrix (according to Torgerson) were examined likewise.

The D^2 matrix was analysed for clustering by the method due to Tocher.

5.2 *Habitat features*

Different features were measured and sometimes different methods were followed for the Swiss *Abies* and Canadian *Picea* forests.

5.2.1 Swiss data

(1) The soil pH was determined on 50 random samples within each sample plot, according to the method described by DOUGHTY (1941).

(2) The light conditions in each plot were determined on sunny days, along 5 random, 6-meter lines (identical in location to the line-intercepts), with a method developed by the author during his stay at the Geobotanical Institute at Zürich (method 1, VAN GROENEWOUD 1963) and also with a portable electric light meter (Lange), on completely overcast days (method 2).

5.2.2 Canadian data

(1) Basal area, defined as the sum of the cross-section areas of all boles at breast height (4 ½ feet), is expressed in terms of square feet per acre. The trees on all plots were tallied for diameter and the data converted to area of cross-section.

(2) Height-growth curves were constructed for four of five dominant white spruce trees on each sample plot. The trees were felled and sectioned at 5-ft intervals; the growth rings were counted at each interval, and from these data the growth curves were constructed. The height-growth (ft/year) during the "intermediate" stage (BAKER 1950) was determined and this index was used as an index of site quality.

(3) The soil profiles on all sample plots were described in detail as to depth and thickness of all soil horizons, distribution of feeder roots, estimate of soil texture and structure, and soil colors on air-dry samples (Munsel color charts).

(4) The permanent wilting point (P.W.P.) was determined with a pressure membrane apparatus on 4 samples of the A_2 horizon of each sample plot. The results are expressed on a per cent dry-weight basis.

(5) The field capacity (F.C.) was determined as follows: Five small isolation plots were established at random within each sample plot (roots were cut, top of vegetation removed). Each isolation plot was drenched and covered with plastic sheeting to reduce evaporation. The soil was allowed to drain for 4 days, after which two samples were taken from the A_2 horizon of each isolation plot. The results are expressed on a per cent dry-weight basis.

(6) "Available moisture" was determined by calculating the difference between P.W.P. and F.C.

(7) The pH of the soil was determined in the field or in the field laboratory within a few hours after sampling, using a Beckman pH meter (Beckman Instruments, Inc., 2500 Harbor Blvd., Fullerton, California) with a combination glass electrode. The soil was prepared as a soil paste according to DOUGHTY (1941). The "measured mean" and the range of pH on each sample plot were determined (VAN GROENEWOUD 1961).

(8) Samples of foliage were taken in midwinter and always from the tops of the trees, to minimize the effects of seasonal fluctuation and of position on the tree. The tree tops were shot down with a .22 caliber rifle with telescopic sight. The foliage was kept at –18°C until it could be further processed. The spruce needles were dried at 60°C. After drying, scales and other contaminations were removed by hand. Dust adhering to the foliage was removed with an air-jet of 60 p.s.i. The samples were ground in a Wiley mill and stored at –18°C until the analysis could be performed. The nitrogen content was determined by the micro-Kjeldahl method. The results were corrected for moisture content of the samples at the time of analysis. Ten samples were analysed from each of the 43 plots. This number was considered necessary, because a preliminary study revealed considerable variation in foliage composition within each plot.

5.3 *Interpretation*

The habitat features were all tested for their relationship with the principal axes (covariance matrix), with the main axes of the constellation of points described by the D^2 matrix, and with the principal axes of the transformed D^2 matrices.

The levels of the habitat feature to be tested, were plotted against the corresponding projection of each plot on each axis. If a relationship was evident, a line or curve was fitted to the points. Where a straight line relationship was found, a correlation coefficient was calculated.

When groups of points could be recognized in the ordination, the differences among the mean levels of the habitat features in these groups were tested by a modified "t" test only for those features that had shown a relationship with the principal or main axes.

The differences among the mean levels of the habitat features in the groups of sample plots, as distinguished by the differential species-group method, were tested by "t" tests. To obtain a measure of the separation of the ranges occupied by these groups along the various gradients, the sum of the standard deviations were compared with the differences among the means.

6. Results

Only a relatively small number of the several hundreds of graphs prepared to test all possible relationships is presented. The number of figures has been limited by applying the following rules:

(a) only statistically significant relationships are shown;

(b) further, only those relationships are shown that convey information not already contained in other graphs, unless they are used to prove a certain point, such as showing the differences in the results of relationships among the various methods.

6.1 *Sampling and small scale distribution*

6.1.1 Comparison of vegetation sampling methods

Several factors have to be considered in judging the relative merits of different sampling methods.

The vegetation data collected by the line-interception method were practically identical to those taken by the point-quadrat-line method. The correlation coefficient was .9994. To facilitate the presentation, the results of both methods were regarded as being identical.

(1) The first factor to be considered is the accuracy of the estimation of the mean cover per plot for each species.

The variation is dependent on the dimensions of each sampling unit (length of line-intercept, area of sampling quadrat), the spatial distribution of the species, and the number of samples. The variance was calculated for different sizes of sampling unit and plotted for both line-intercepts and quadrats. Some of the results are shown in Fig.8 and 9.

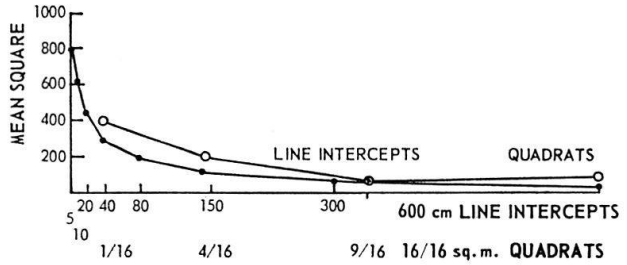

Fig. 8. Mean square/length of line-intercepts and mean square/quadrat size (*Oxalis acetosella*; plot 1, Ziegelwald, Roggwil).

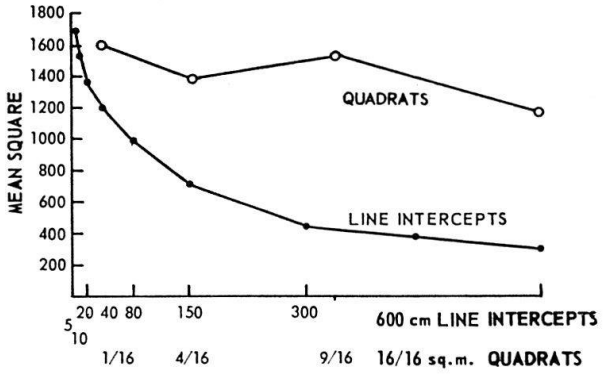

Fig. 9. Mean square/length of line-intercepts and mean square/quadratsize (*Polytrichum formosum*, Plot 4, Ziegelwald, Roggwil).

In general, the line-intercepts are more efficient in decreasing the variance with increasing size of sampling unit than the quadrats. This is probably due to the increased error in estimating the cover of the species with larger quadrats, opposed to the constant error by the line-interception method. For the same reason, the graphs of variance plotted against sample-unit size of the quadrat method also showed more irregularities than those of the line-interception method. The graphs of the line-intercept variance data were very regular.

These graphs also served to study the pattern of distribution of the species on the sample plots (see next paragraph)

(2) Another factor is the efficiency of the different sampling units in estimating the cover percentage of a high percentage of the total number of species present on the plot.

The results varied with the distribution of the species on the different plots. On some, the quadrat method was more effective; on others, the line-intercept (Figs. 10 and 11).

The overall assessment indicates that the line-interception method has the most advantages of the three methods.

The line-interception and the point-quadrat-line method have the added advantage that the vegetative cover can be rather easily related to habitat factors as measured along the lines.

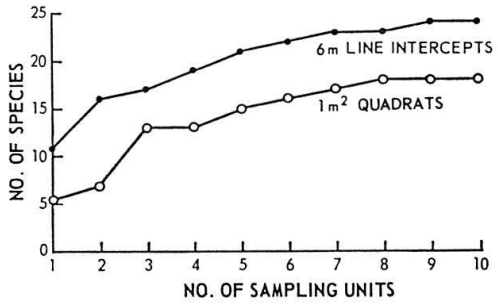

Fig. 10. Increase in number of species sampled, with increase in the number of sampling units (line-intercepts and quadrats, plot 4, Ziegelwald, Roggwil).

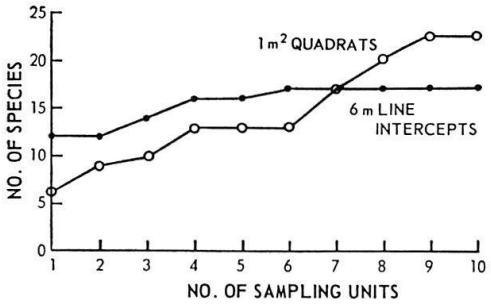

Fig. 11. Increase in number of species sampled, with increase in the number of sampling units (line-intercepts and quadrats, Plot 3, Ziegelwald, Roggwil).

In this study, only the line-intercept data was analysed using the methods mentioned before.

6.1.2 Small scale non-random pattern within sample plots

Seventy variance analysis graphs were prepared, one of which (*Polytrichum formosum*, plot 2, Roggwil) showed a peak at a length of 40 cm line-intercept. This pattern did not interfere with the planned analysis and was too unimportant to induce further investigation. All other graphs showed decreasing

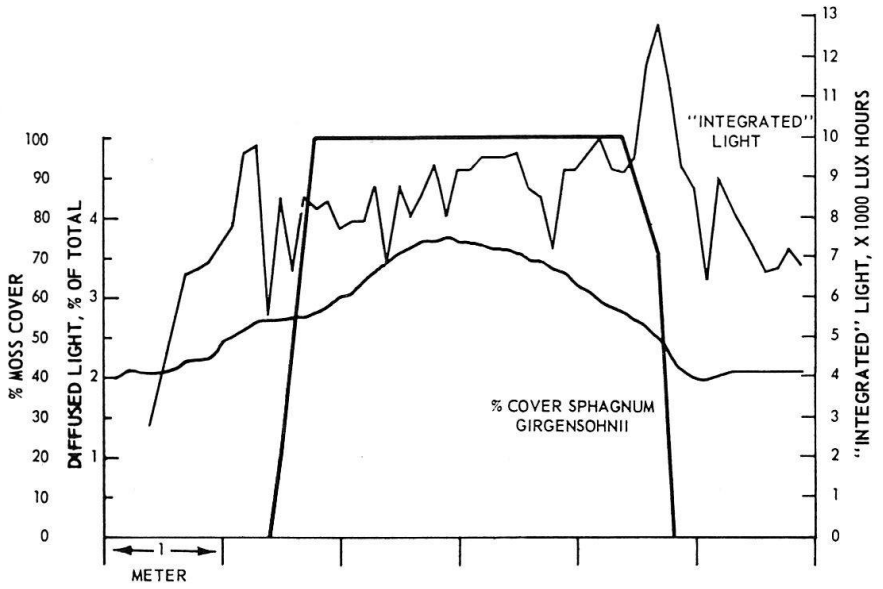

Fig. 12. Distribution of "integrated" and diffused light in relation to percentage cover of *Sphagnum quinquefarium* (not *girgensohnii*) along line-intercept (plot 10, line 3, Murgenthal).

variance with increasing lengths of line-intercepts, with an almost constant variance around a length of 600 cm (Figs. 8 and 9). Further increase of length of line intercepts would have been useless in this study. Based on this evidence, it must be assumed that, for the species investigated, *small scale non-random patterns were not present within these sample plots.*

6.1.3 Distribution of the vegetation along line-intercepts in relation to light.

The cover percentage of several species were plotted, together with light conditions along several line-intercepts of each plot. The data were collected in Swiss forests. The following results were noted:

(1) Only in dense forest did the measurements of the light conditions along the lines coincide (Fig. 12);

(2) In more open forest, the light as measured by the two methods shows an entirely different distribution pattern (Fig. 13);

(3) Only where the light, measured by the two methods, had the same

distribution, was this distribution related to the distribution of the vegetation, in particular with *Sphagnum quinguefarium* (Fig. 12);

(4) A total of 50 light and vegetation graphs were plotted but the patterns coincided only in the case of dense forest.

At this point, results of the Swiss data will be presented first, to be followed by the results of the Canadian data. Both were subjected to *identical* procedures.

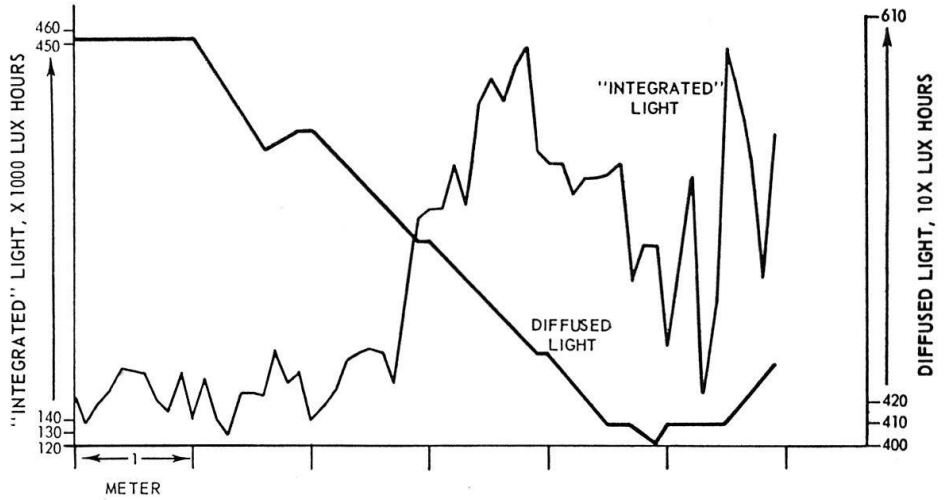

Fig. 13. Distribution of "integrated" and diffused light along line intercept (plot 12, line 5, Unterwald, Roggwil).

6.2 *Swiss sample plots*

6.2.1 Classification of sample plots (Zürich-Montpellier method)

Five groups of differential species were distinguished (Appendix I).

Group A comprises species which are present on all plots: *Vaccinium myrtillus*, the seedlings of *Abies alba* and *Picea abies*, *Hylocomium splendens*, *Rhytidiadelphus triquetrus*, *Polytrichum formosum* and *Thuidium tamariscifolium*.

Group B contains species which have fairly wide ecological amplitudes but which are particularly suited to delimit mull from mor soils (Ellenberg 1963, p. 86). To this group belong the herbs; *Anemone nemorosa*, *Fragaria vesca*, *Hedera helix*, *Lysimachia nemorum*, *Viola silvatica*, *Galium rotundifolium*, and the mosses: *Catharinea undulata*, *Mnium undulatum* and *Mnium affine*. The species of this group are present on sample plots no. 1, 2, 3, 6, and 12 (Vegetation unit I) and are absent (2 exceptions) in the other plots.

Group C comprises species which obviously have ecological requirements close to those of group B, but are differentiated by a somewhat wider amplitude. The species of group C are: *Athyrium filix-femina*, *Dryopteris austriaca*, *Luzula pilosa*, *Maianthemum bifolium*, *Oxalis acetosella*, *Rubus* spec., and *Eurhynchium striatum*.

Group D embraces species which flourish on moderately dry to moderately moist sols with a lower soil pH than foregoing groups. The species of this group have ecological amplitudes, which partly overlap those of the species of group C, but which almost completely differentiate this group from group B. The species of this group are: *Pleurozium schreberi*, *Rhytidiadelphus loreus*, *Dicranum scoparium*, *Hypnum cupressiforme*, and *Plagiothecium undulatum*.

Group E comprises species which have their optima on moderately moist to moist, and very acid soils. The species belonging to this group are: *Bazzania trilobata* and *Sphagnum quinquefarium*. These species occurred only in sample plots no. 4, 5, 8, 9, and 10 (Vegetation unit III).

Group C and D help differentiate a group of sample plots, containing no. 7, 11, 13, 14, and 15, (Vegetation unit II), which is intermediate between the plots differentiated, respectively, by the *Anemone nemorosa* and the *Bazzania trilobata* groups.

6.2.2 Habitat factors in relation to classification of sample plots

The groups of sample plots established in foregoing paragraph received the following average amounts of light with their respective standard deviations (all expressed in kilo-Lux-hours per day): 14.72 ± 7.35, 19.22 ± 16.12 and 15.56 ± 6.72. The differences among these average levels are not statistically significant.

The average soil pH of these groups of sample plots, with their standard deviations are $4.04 \pm .38$, $3.83 \pm .29$, and $3.61 \pm .22$, respectively. The differences among the means are all statistically significant. If, however, the ranges along the pH gradient occupied by these Vegetation units are considered, it is obvious that only the pH ranges of unit I and III are separated to some extent; the differences between the means is .43 and the sum of the standard deviations is .60.

6.2.3 Principal component analysis of the covariance matrix

The principal component analysis resulted in principal axes with the following eigenvalues:

trace of matrix = sum of all eigenvalues = 31875.0

Axes Eigenvalues

I 12402.0 accounting for 38.91 % ($= 12402.0 / 31875.0 \times 100$) of the total variation.

II 8016.5 accounting for 25.15 % of the total variation.

III 3833.3 accounting for 12.03 % of the total variation.

IV 2335.0 accounting for 7.33 % of the total variation.

V 1798.2 accounting for 5.64 % of the total variation.

The first two axes account for 64.06%, the first three axes for 76.09, and the first five axes for 89.06% of the total variation.

The 38 coefficients of the eigenvectors are listed in table 1. The coefficients which contribute most are in italics.

It is noteworthy that the variance of many species can only be satisfactorily described by more than one principal component, e.g. *Oxalis acetosella* by

Table 1. Eigenvalues and eigenvectors of the Swiss covariance matrix.

	I	II	III	IV	V
Eigenvalues	12402.	8016.5	3833.4	2335.0	1798.2
Species			Eigenvectors		
1. *Oxalis acetosella*	–.02317	*–.17303*	*.43073*	*.44055*	*.20080*
2. *Carex brizoides*	–.00782	–.00447	.01609	.02238	.02894
3. *Rubus* spec.	–.00543	–.00143	.02031	.03411	.02734
4. *Polytrichum formosum*	*–.56862*	*–.74735*	–.06078	–.15409	–.15084
5. *Thuidium tamariscifolium*	.02274	.03175	.05320	*–.49471*	*.30432*
6. *Hylocomium splendens*	–.02362	.10902	–.13904	–.13432	–.10456
7. *Rhytidiadelphus triquetrus*	.01274	.01593	–.04597	.07050	.03977
8. *Abies alba* (seedling)	–.04874	.02277	–.04145	.03446	.03192
9. *Luzula luzuloides*	–.00270	–.00275	.00078	.00139	.00041
10. *Maianthemum bifolium*	–.01351	.00184	–.00046	.00083	.01423
11. *Vaccinium myrtillus*	–.11249	.16698	*–.27644*	*.26820*	.04920
12. *Hypnum cupressiforme*	–.00965	.03602	–.04853	–.01272	.02898
13. *Plagiochila asplenioides*	–.02236	–.00742	*.20060*	*.20261*	*.26078*
14. *Sphagnum quinquefarium*	*.12412*	–.07078	*–.33668*	.07327	–.06556
15. *Ptilidium ciliare*	–.00718	–.01639	.00460	–.00086	–.00604
16. *Picea abies* (seedling)	–.08383	.06267	–.10661	.11761	.05315
17. *Catharinea undulata*	.00201	–.00089	.01515	.01717	.00664
18. *Eurhynchium striatum*	.03392	*.23404*	*.33803*	–.03699	*–.58078*
19. *Mnium affine*	.02879	.00318	.13410	.17271	.13835
20. *Hedera helix*	.00087	.00034	.00446	.00519	.00208
21. *Viola silvatica*	.00007	.00035	.00088	.00031	–.00253
22. *Rhytidiadelphus loreus*	–.03664	.01352	–.00817	–.00246	.01330
23. *Lophocolea bidentata*	.00170	–.00654	–.03063	–.00306	.00739
24. *Dicranum scoparium*	.00937	–.00273	–.05384	.01676	–.00941
25. *Pleurozium schreberi*	–.06675	.12852	–.19156	*.20894*	.01655
26. *Bazzania trilobata*	*.23596*	–.10909	*–.56851*	.15899	–.14877
27. *Plagiothecium undulatum*	.00052	–.00125	.00378	–.00267	.00194
28. *Chiloscyphus polyanthemus*	.00147	–.00083	–.00278	–.00370	.00105
29. *Athyrium filix-femina*	–.00874	–.00791	.00104	–.00265	.00178
30. *Dicranella heteromalla*	.00046	.00021	–.00122	–.00107	.00104
31. *Agrostis tenuis*	–.00186	.00018	.00243	.00557	.00658
32. *Galium rotundifolium*	.00128	–.00081	.00066	–.00426	.00296
33. *Lophocolea heterophylla*	.00215	.00001	–.00358	–.00179	–.00340
34. *Mnium undulatum*	–.00234	–.00266	.00031	–.00098	.00064
35. *Abies alba*	*.58045*	*–.36666*	–.02491	*–.31329*	*.38070*
36. *Picea abies*	*–.48324*	*.34412*	–.17508	*–.27516*	*.42201*
37. *Fagus silvatica*	–.0237	–.15145	.08981	.06435	–.01536
38. *Quercus robur*	.01288	.01035	.06829	.06512	–.02768

component 3 and 4, *Polytrichum formosum* by component 1 and 2, *Eurhynchium striatum* by component 3 and 5, *Picea abies* by component 1, 2, and 5. The variation of *Bazzania trilobata* and *Sphagnum quinquefarium* can be described almost completely by component 3.

If the ecological requirements of the species are known, this can be an aid in explaining the possible ecological meaning of the axes to which these species are important contributors.

The projections of the points representing sample plots, on the planes spanning the first and second, and the first and the third principal axes, are shown in Fig. 14.

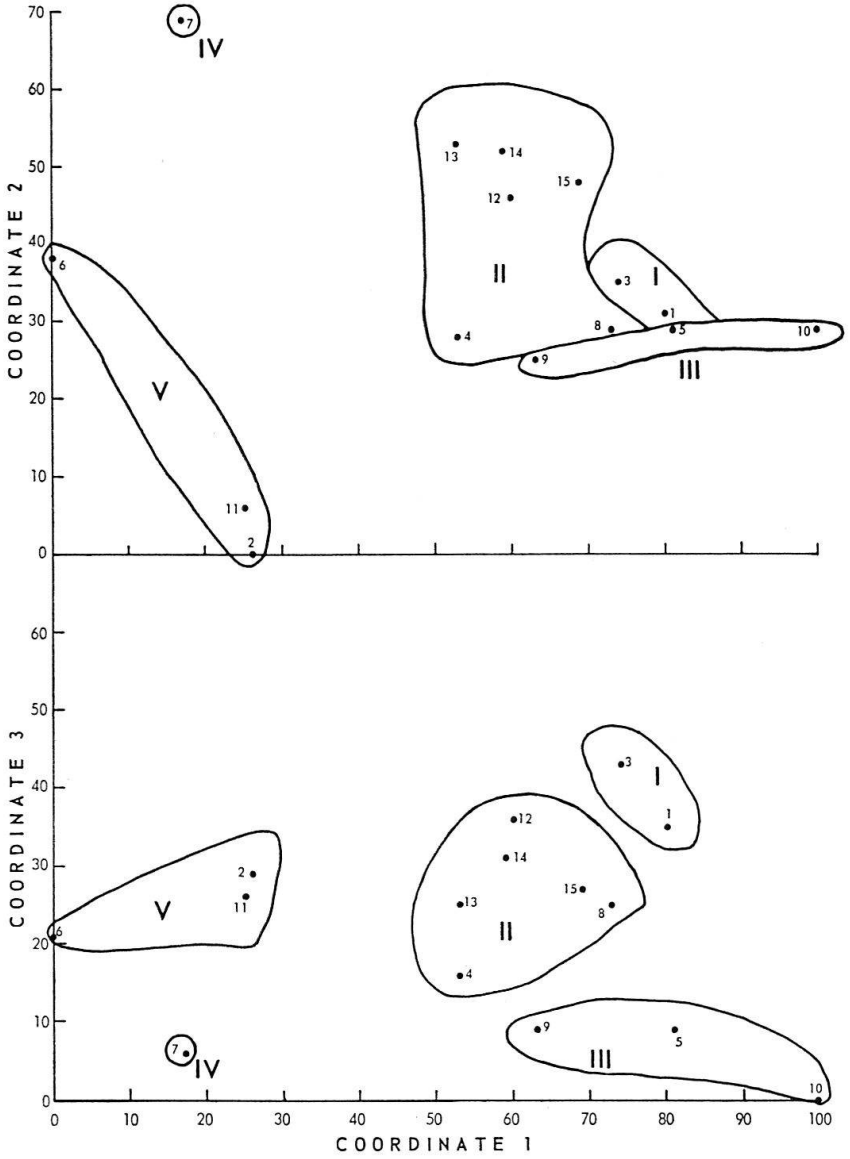

Fig. 14. Projection of sample plots on the planes spanning the first and second, and the first and third principal axes (covariance matrix, Swiss data) with clustering of plots indicated.

The relationships between the cover percentages of the more important species with the principal axes are shown in Fig. 15.

6.2.4 Habitat factors in relation to the principal axes

The habitat factors measured, light conditions and soil pH, were related to the principal axes. The "integrated light" levels and the soil pH on the sample

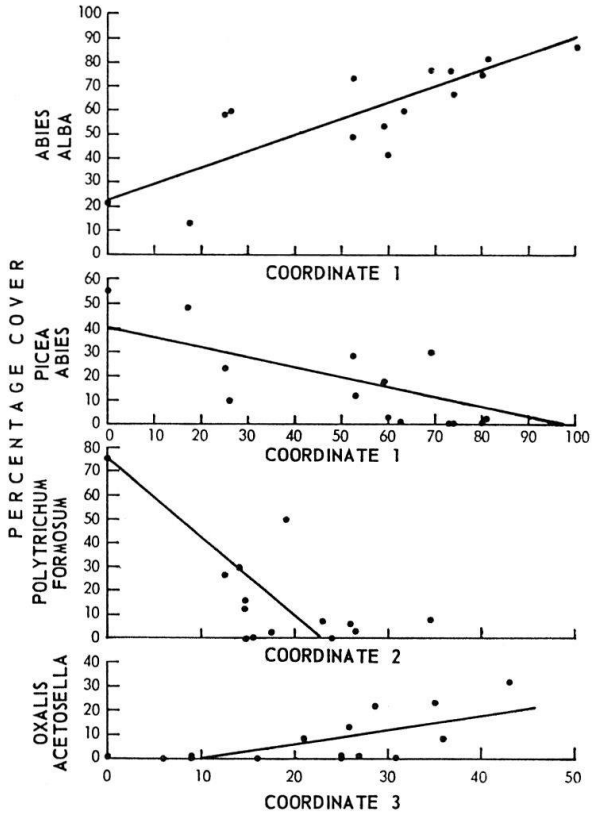

Fig. 15. Quantitative distribution of *Abies alba* and *Picea abies* along the first principal axis, *Polytrichum formosum* along the second principal axis and *Oxalis acetosella* along the third principal axis (cov. matrix, Swiss data).

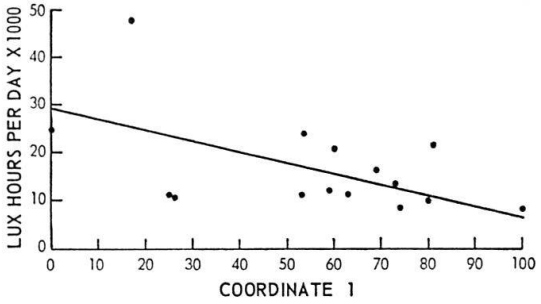

Fig. 16. Relationship between light conditions and the first principal axis (cov. matrix, Swiss data).

plots were plotted against the corresponding values (coordinates) of each plot, for each of the three principal axes. If any indication of a relationship existed, straight lines or regression curves were fitted.

Figs. 16 and 17 show the relationships of the first axis to light conditions and the third axis to soil pH. Both were statistically significant. No relationship was found with the second axis.

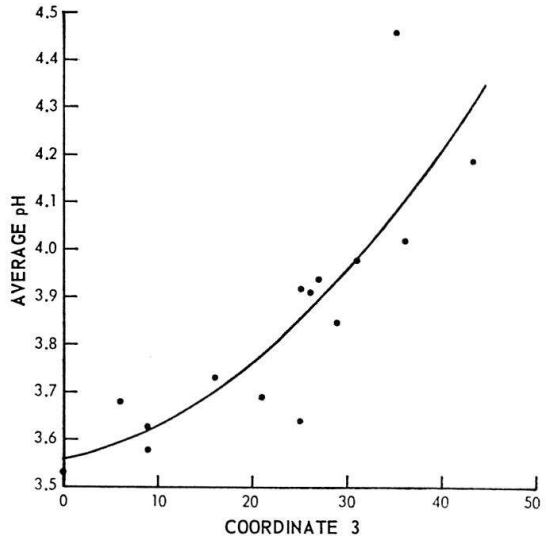

Fig. 17. Relationship between soil pH and the third principal axis (cov. matrix, Swiss data).

6.2.5 Analysis of the D^2 matrix

All D^2's were statistically significant ($P \leqq .05$).

The D^2 matrix was investigated according to the method developed by Torgerson. Three axes were constructed and the points representing sample plots were projected on the planes spanning these axes (Fig. 18).

The distances in two-dimensional space were compared with the D^2's, by calculating the correlation coefficient between these distances and the corresponding D^2's. The correlation coefficient is .939, which is significant at the .001 level. Thus 88.7% of the variation is accounted for by the two dimensional ordination.

6.2.6 Habitat factors in relation to the main axes (D^2 matrix)

The relationship between each axis and the forementioned habitat factors was tested. As under paragraph 6.2.4, the first axis was significantly related only to light conditions (method 1) and the third axis to soil pH. Although the relationships were largely identical to those mentioned under paragraph 6.2.4, they were statistically less significant (Figs. 19 and 20).

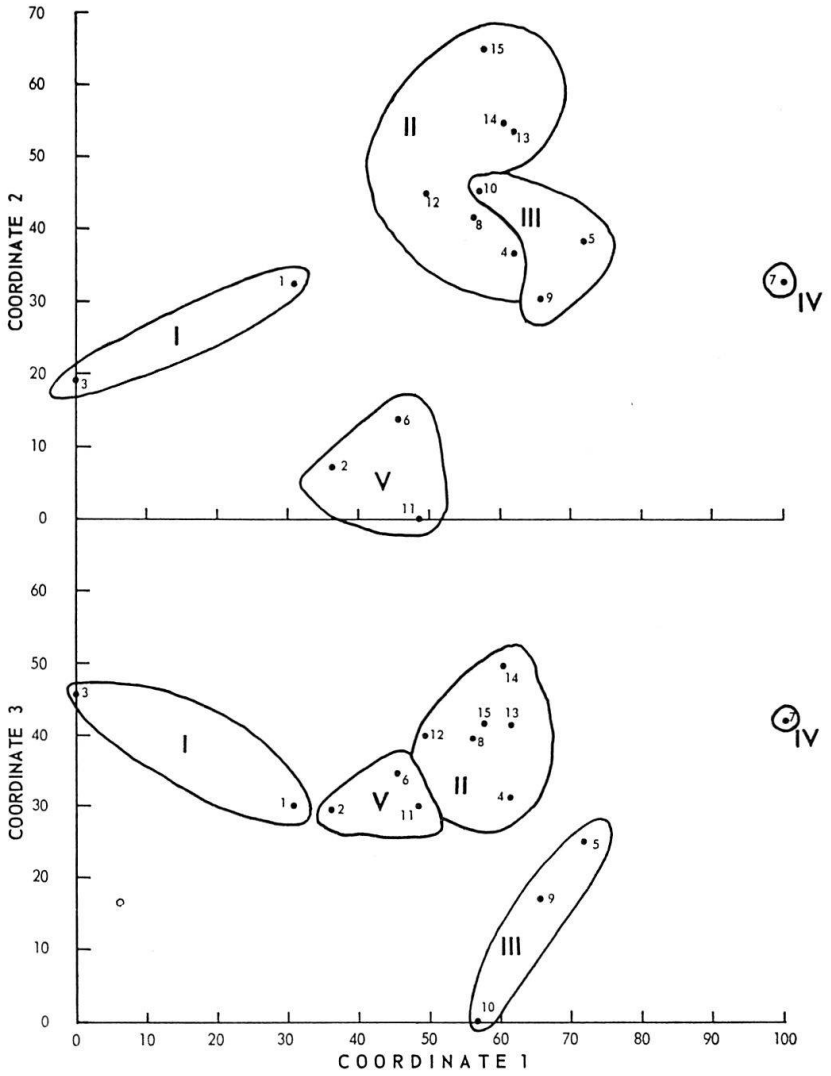

Fig. 18. Projection of sample plots on the planes spanning the first and second, and the first and third main axes (D^2 ordination, Swiss data) with clustering of plots indicated.

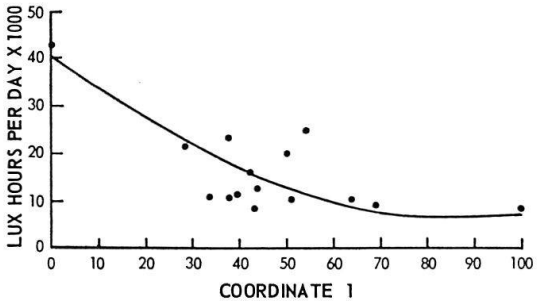

Fig. 19. Relationship between light conditions and the first main axis (D^2 ordination, Swiss data).

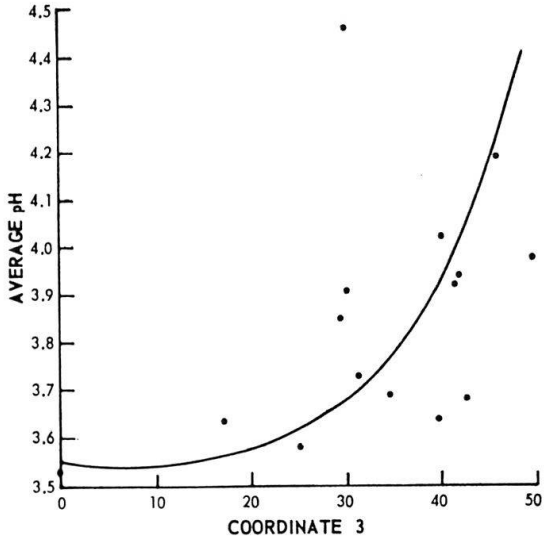

Fig. 20. Relationship between soil pH and the third main axis (D^2 ordination, Swiss data).

6.2.7 Principal component analysis (Q-method) of the transformed D^2 matrices.

Two transformations were used: $R = (1 + D^2)^{-1}$, and $R = e^{-D^2}$. The exponential subroutine used in Fortran on the G-20 computer, the $R = e^{-D^2}$ transformation works only with exponents between —63 and +63. Any exponent of which the absolute value was larger than 63 was automatically given the value 0. Since most of the off diagonal elements in the D^2 matrix were larger than 63 they were replaced by zeros in the transformed matrix. The resulting eigenvalues and eigenvectors were almost all zeros and ones. The results of this transformation were not analysed.

The results of the principal component analysis of the $R = (1 + D^2)^{-1}$ matrix are listed in table 2. Only the first three eigenvalues and the coefficients of their eigenvectors are listed.

The eigenvalues of the first three axes were as follows:

Trace of the matrix = sum of all eigenvalues = 15.0

Axes Eigenvalues

I 1.1686522 accounting for 7.79% of the total variation.
II 1.0373909 accounting for 6.92% of the total variation.
III 1.0140313 accounting for 6.76% of the total variation.

The first two axes thus account for 14.71%, and the first three axes for 21.47% of the total variation.

Table 2. Eigenvalues and eigenvectors of the Swiss $(1 + D^2)^{-1}$ matrix.

	Axes				Axes		
	I	II	III		I	II	III
		Eigenvalues				Eigenvalues	
Plot number	1.1686522	1.0373909	1.0140313	number	1.1686522	1.0373909	1.0140313
		Eigenvectors				Eigenvectors	
1	.14222321	.17624994	.21756172	9	.25679214	.30200167	−.45595084
2	.12656551	.38099206	.35592580	10	.16825722	.21976787	−.42181803
3	.06633672	.15495798	.199353751	11	.13708974	.40578184	.31920601
4	.30028464	.19683786	−.27587316	12	.30902949	−.06469361	.12709984
5	.15062401	.19914597	.31330388	13	.45123823	−.28100184	.08031121
6	.14939449	.37723863	.28399513	14	.37905062	−.25283243	.07444108
7	.08432882	.12967207	−.08409874	15	.35159907	−.27186514	.09786224
8	.37135248	−.19260563	.03315040				

No projections of the plots on the planes spanning the first three axes are presented because they do not supply information not already contained in other figures and because these axes only account for a total of 21.47% of the total variation.

6.2.8 Habitat features in relation to the principal axes (Q-method)

The relationship between each axis and the habitat factors was tested. As before, the first axis was only significantly related to light conditions (method 1) and the third axis to soil pH. The relationships are expressed in Figs. 21 and 22.

6.2.9 Clustering of sample plots

The projection of the points, representing sampling plots, on the planes spanning the principal axes is shown in Fig. 14. There is a tendency to cluster, which coincides with that shown by the projection of the plots on the planes spanning the main axes of the hyper-space described by the D^2 matrix (Fig. 18). The clusters contain the following plots: Cluster I, plot 1 and 3; Cluster II, plot 4, 8, 12, 13, 14, and 15; cluster III, plot 5, 9, and 10; cluster IV, plot 7, and cluster V, plot 2, 6, and 11.

The D^2 matrix was analysed by the method developed by Tocher (Appendix III). This analysis showed the existence of five groups which were identical to the clusters mentioned before.

6.2.10 Habitat factors in relation to clustering of sample plots

The mean values of the habitat factors for each cluster were compared by "t" test, following an approximate method due to Cochran and Cox (Greig-

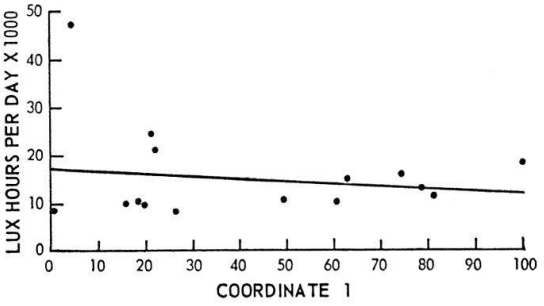

Fig. 21. Relationship between light conditions and the first principal axis ($(1 + D^2)^{-1}$ matrix, Swiss data).

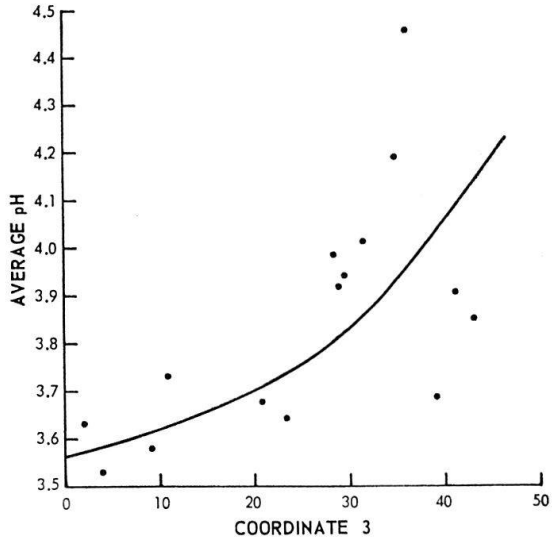

Fig. 22. Relationship between soil pH and the third principal axis ($(1 + D^2)^{-1}$ matrix Swiss data).

Smith 1964). The frequency distributions are close to normal, and no transformation was deemed necessary.

Most of the clusters have significantly different average light conditions. Not significant are the differences between the means of the levels of light received, of cluster I and III and of II and V. The average levels of light received by these sites with their standard deviations are 9.00 ± 1.56; 15.20 ± 12.43; 13.92 ± 10.45; 47.89 ± 24.51; and 16.92 ± 12.97 kilo-Lux hours per day respectively, at the time these measurements were made. The clusters are, with the exception of two, significantly different for soil pH's. The mean pH's with their standard deviations are 4.31 ± .26; 3.89 ± .28; 3.58 ± .20; 3.68 ± 1.17; and 3.81 ± .31 respectively. Clusters II and V with pH 3.89 and pH 3.81 respectively, which are not significantly different, occupy completely overlapping ranges along the third principal axis (covari-

ance matrix). Also cluster IV is not significantly different from clusters II, III, and V.

If the differences among the means are considered in relation to the sum of the standard deviations, only the differences among cluster I, II, and III and between cluster I and V carry weight.

6.3 *Canadian sample plots*

6.3.1 Classification of sample plots (Zürich-Montpellier method)

As is obvious from the plant tables (Appendix II), the vegetation of the white spruce forests in Saskatchewan is very homogeneous, with many species occurring in most of the plots. Nevertheless, it was possible to recognize, according to the differential species-group method, two groups of species which are predominantly present in a limited group of plots. These groups of species were used to group the plots into three units.

The first group (A) of species comprises: *Lonicera involucrata*, *Lonicera dioica*, *Ribes hirtellum*, *Ribes triste*, *Shepherdia canadensis*, *Amelanchier alnifolia*, *Lathyrus ochroleucus*, *Habenaria obtusata*, *Geocaulon lividum*, *Actaea rubra*, *Galium boreale* and the moss *Eurhynchium pulchellum*. These species were, with few exceptions, not present in the following plots: 1, 3, 5, 6, 7, 8, 13, 16, 19, and 38 (Vegetation unit III).

The second group (B) of species contains *Galium triflorum*, *Elymus innovatus*, *Equisetum scirpoides*, *Equisetum pratense*, *Hieracium canadense*, *Carex capillaris* and the lichen *Peltigera* spec. These species occurred predominantly in the following sample plots 9, 10, 11, 14, 22, 23, 27, 29, 31, 32, 35, 36, 37, and 43 (Vegetation unit I).

The species of group A did, and those of group B did not occur, in the following sample plots: 2, 4, 12, 15, 17, 18, 20, 21, 24, 25, 26, 28, 30, 33, 34, 39, 40, 41, and 42 (Vegetation unit II).

6.3.2 Habitat features in relation to classification of sample plots

The average level of each habitat feature for each vegetation unit, established in the foregoing paragraph, was compared statistically with those of the other two units with the following results. No significant differences were found among the average values of permanent wilting point (mean value with standard deviations in vegetation unit I, II, and III, respectively, 3.80 ± 1.96, 3.09 ± 1.16 and $2.86 \pm .99$), nitrogen content of the white spruce foliage ($1.26 \pm .04$, $1.25 \pm .06$, $1.24 \pm .09$), basal area of white spruce (139.9 ± 40.9, 137.9 ± 50.3, 139.3 ± 33.9), "measured mean" pH of the mineral soil (4.95 ± 1.15, 4.94 ± 1.14, 4.71 ± 1.26), "measured mean" pH of

humus layer (5.63 ± .53, 5.45 ± .53, 5.09 ± .7), "measured mean" pH of the fermentation layer (5.92 ± .48, 5.99 ± .28, 5.56 ± .47).

Significant differences (at P = .05) were found among the average levels of the field capacity (unit II and III), "available moisture" (unit I and III) and height growth (I.H.G.I.) (unit I and III). The average levels for the three units are as follows: field capacity: 15.20 ± 2.05, 15.36 ± 1.30, and 16.85 ± 2.16; "available moisture": 11.52 ± 2.04, 12.06 ± .95 and 13.93 ± 2.03; I.H.G.I.: 1.075 ± .19, 1.10 ± .15 and 1.18 ± .24.

The significant differences between the average levels of "available moisture" and height growth in units I and III suggest that height growth is correlated with the available moisture. To test this relationship further, all levels of available moisture were plotted against the corresponding levels of height-growth and a correlation coefficient was calculated. No statistically significant relationship was found to exist (r = .20, P > .10). A relationship did exist, if only the sample plots contained in unit I and III were used (r = .40, P = .05).

Table 3. Eigenvalues and eigenvectors of the Canadian covariance matrix.

	Axes				
	I	II	III	IV	V
Eigenvalues	34419	7106.4	3483.3	2945.5	1405.6
Species	Eigenvectors				
1. *Rosa acicularis*	.02997	−.00142	−.04535	−.02109	−.04783
2. *Linnaea bor.* var. *amer.*	.08841	.09679	*−.57951*	*−.23789*	*−.42702*
3. *Petasites palmatus*	.05866	−.01554	*−.32896*	.03517	*.22979*
4. *Cornus canadensis*	.18676	−.07685	*−.53373*	.14256	*.31060*
5. *Fragaria vesca*	.00366	.00123	−.00638	−.01131	.01474
6. *Fragaria virginiana*	.06284	−.00548	−.10823	.03739	.12304
7. *Mitella nuda*	.01005	.01577	−.10223	−.02328	.13508
8. *Mertensia paniculata*	.10699	−.04036	*−.22644*	.02635	*.37971*
9. *Maianthemum canadense*	.02280	−.00705	−.11046	−.01176	.01194
10. *Pyrola secunda*	.00624	−.00321	−.00928	−.00692	−.00755
11. *Aralia nudicaulis*	.03371	−.01406	−.07949	.04377	−.02303
12. *Vaccinium v.id.v.m.*	.00004	.00017	−.00065	.00014	.00059
13. *Pyrola virens*	−.00300	−.00301	−.01792	−.00354	.00075
14. *Trientalis borealis*	.00634	−.00097	−.04825	.00843	*−.02151*
15. *Rubus pubescens*	.05164	−.02624	−.18379	.03222	.16242
16. *Hylocomium splendens*	*−.78547*	*−.39106*	*−.21676*	*.40044*	−.10261
17. *Pleurozium schreberi*	*−.28288*	*.87947*	−.07136	*.32474*	.15732
18. *Cornus stolonifera*	.00677	.00454	.00277	−.00122	.02461
19. *Symphoricarpus alba*	.00277	.00404	.00297	−.01089	.01708
20. *Picea glauca*	*−.33991*	−.06533	.13128	*.54098*	*.60846*
21. *Populus tremuloides*	*.34709*	*−.22530*	*.21387*	*.59620*	*.23259*
22. *Populus balsamifera*	.00076	−.01140	.06081	−.03706	−.0159

6.3.3 Principal component analysis of the covariance matrix

The principal component analysis resulted in principal axes with the following eigenvalues.

Trace of the matrix = sum of all eigenvalues = 51859.0

Axes Eigenvalues

I 34419.0 accounting for 66.3% of the total variation.
II 7106.4 accounting for 13.7% of the total variation.
III 3483.3 accounting for 6.7% of the total variation.
IV 2945.5 accounting for 2.7% of the total variation.

The first two axes account for 80.0%, the first three axes for 86.7%, the first five axes for 95,1% on the total variation.

The eigenvalues and coefficients of the eigenvectors are listed in table 3. The coefficients which contribute the most are in italics.

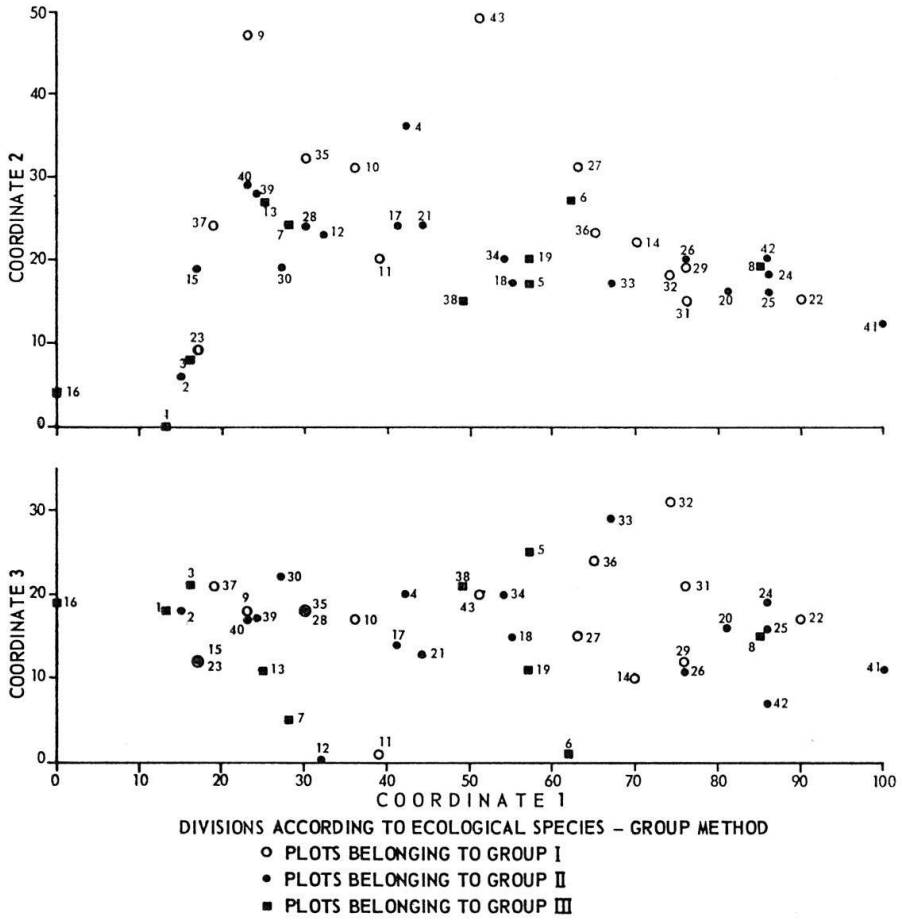

Fig. 23. Projection of sample plots on the planes spanning the first and second, and the first and third principal axes (cov. matrix, Canadian data) with classification of sample plots according to the differential species-group method, indicated.

As in table 1, it is noticeable that the variance of several species can only be explained by more than one component. The fourth and fifth components, however, have such low eigenvalues that for all practical purposes they can be omitted. On this basis the variance of *Linnaea borealis* var. *americana*, *Petasites palmatus*, *Cornus canadensis*, *Mertensia paniculata* and perhaps *Rubus pubescens* can be explained by component 3 and *Picea glauca* by component 1.

The variance of *Hylocomium splendens* and *Pleurozium schreberi* can only be explained by components 1 and 2. To explain the variance of *Populus tremuloides* all of the first three components are needed.

The projections of the points representing sample plots, on the planes spanning the first and second, and the first and third principal axes are shown in Fig. 23.

The relationships between the cover percentages for the more important species and the principal axes are shown in Figs. 24 to 27 inclusive.

6.3.4 Habitat features in relation to the principal axes

The habitat features measured, mentioned under Methods, were related to the principal axes. When the graphs indicated a possible significant relationship either a straight line or a curve was fitted to the data.

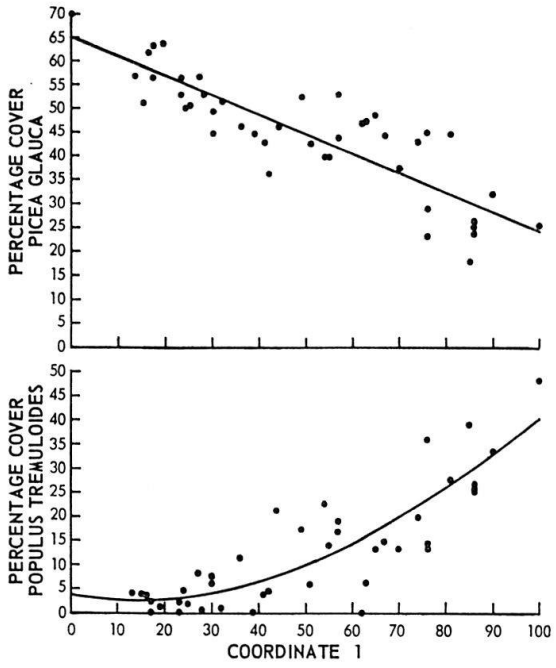

Fig. 24. Quantitative distribution of *Picea glauca* and *Populus tremuloides* along the first principal axis (cov. matrix, Canadian data).

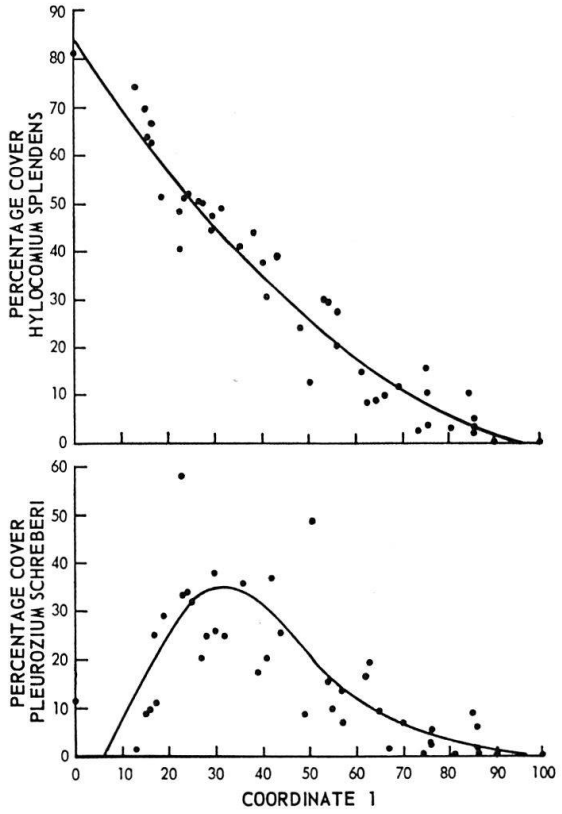

Fig. 25. Quantitative distribution of *Hylocomium splendens* and *Pleurozium schreberi* along the first principal axis (cov. matrix, Canadian data).

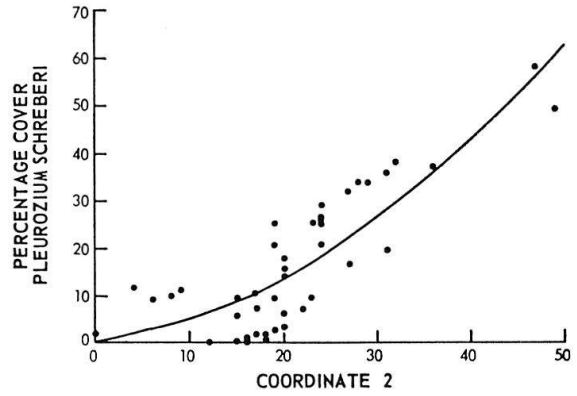

Fig. 26. Quantitative distribution of *Pleurozium schreberi* along the second principa axis (cov. matrix, Canadian data).

Basal area of white spruce, maximum pH of the humus layer, maximum pH of the top mineral soil, and the "measured mean" pH of the fermentation layer were all significantly linearly related to the first principal axis, with correlation coefficients, $r_{BA} = -.745$, $P = .001$; $r_{pH,H.} = -.56$, $P = .001$;

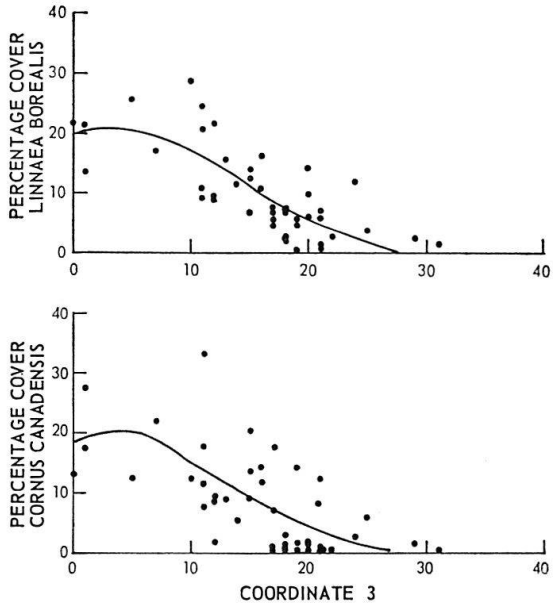

Fig. 27. Quantitative distribution of *Linnaea borealis* and *Cornus canadensis* along the third principal axis (cov. matrix, Canadian data).

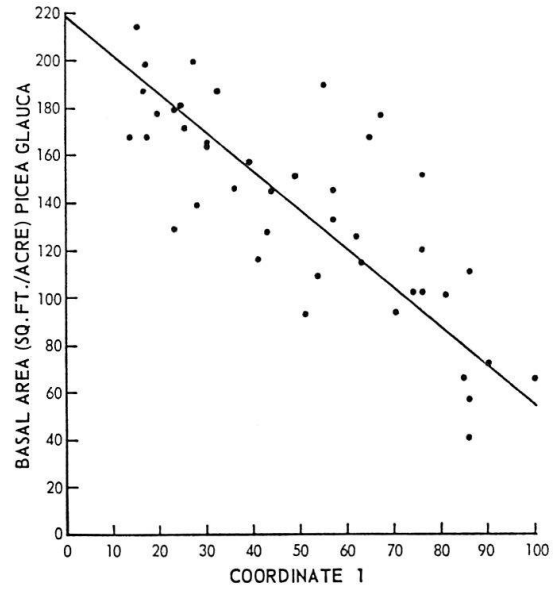

Fig. 28. Relationship between the basal area (sq. ft./acre) of *Picea glauca* and the first principal axis (cov. matrix, Canadian data).

$r_{pH,S.} = -.30$, $P = .05$; and $r_{pH,F.} = .62$, $P = .001$ (Figs. 28, 29, 30, and 31). No other features showed a significant relationship with the first axis.

Only "available moisture" (% dry weight) was linearly related to the second principal axis, $r = -.50$, $P = .001$ (Fig. 32). No other features showed a significant relationship to the second axis.

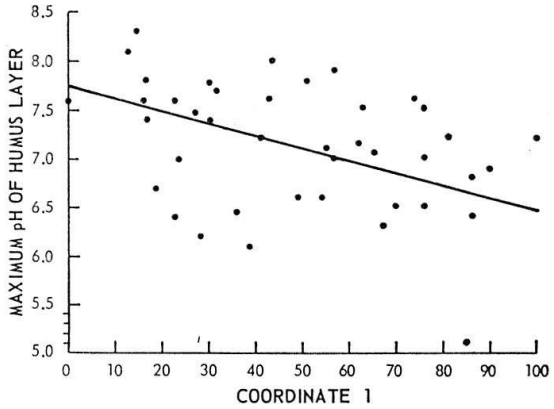

Fig. 29. Relationship between the maximum pH of the humus layer and the first principal axis (cov. matrix, Canadian data).

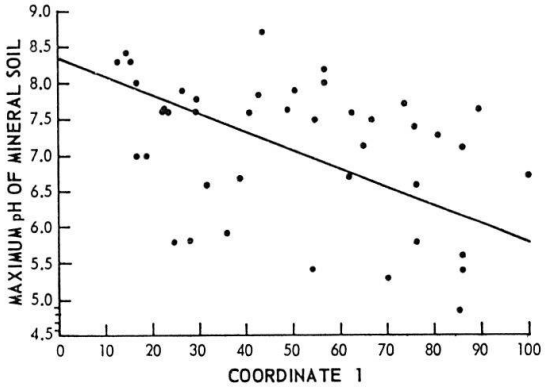

Fig. 30. Relationship between the maximum pH of the top of the mineral soil and the first principal axis (cov. matrix, Canadian data).

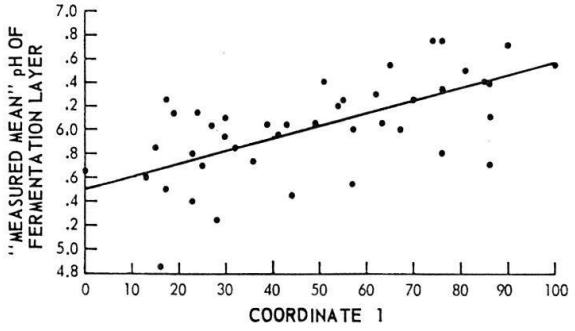

Fig. 31. Relationship between the "measured mean" pH of the fermentation layer and the first principal axis (cov. matrix, Canadian data).

The "measured mean" pH of the humus layer was significantly related to the third principal axis, $r_{pH,H.} = .32$, $P = .05$ (Fig. 33).

Contrary to expectations, no relationships were found with either the height-growth of white spruce, the nitrogen content of the white spruce foliage, field capacity, or permanent wilting point.

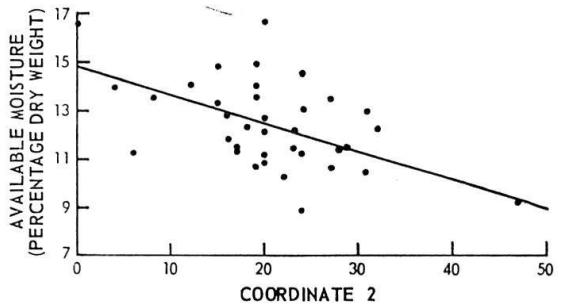

Fig. 32. Relationship between the "available moisture" and the second principal axis (cov. matrix, Canadian data).

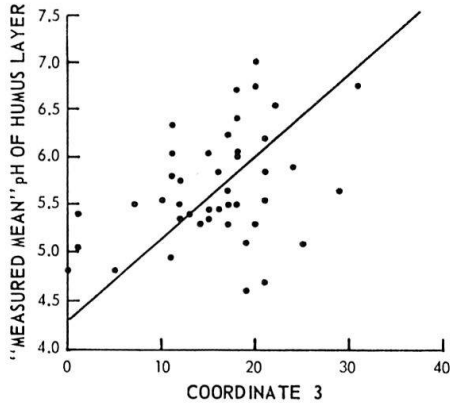

Fig. 33. Relationship between the "measured mean" pH of the humus layer and the third principal axis (cov. matrix, Canadian data).

6.3.5 Analysis of the D^2 matrix

The D^2 matrix was investigated according to a method developed by Torgerson.

The first two axes were constructed and the points representing sample plots were projected on the plane spanning these axes (Fig. 34).

The distances between the points in the one and two dimensional projections, respectively, were correlated with the corresponding D^2's. The correlation coefficients were .777 and .854 respectively. Both correlation coefficients were significant of the 0.1% level. This means that the first axis accounts for 60.37% of the variation present in the D^2 matrix. The two axes account for 72.93% of the variation.

6.3.6 Habitat features in relation to the main axes

The relationships between the above mentioned axes and habitat features were investigated. The graphs closely resembled those found with the principal axes (R method) but the relationships were slightly less significant, e.g. the correlation coefficient expressing the relationship between the first main

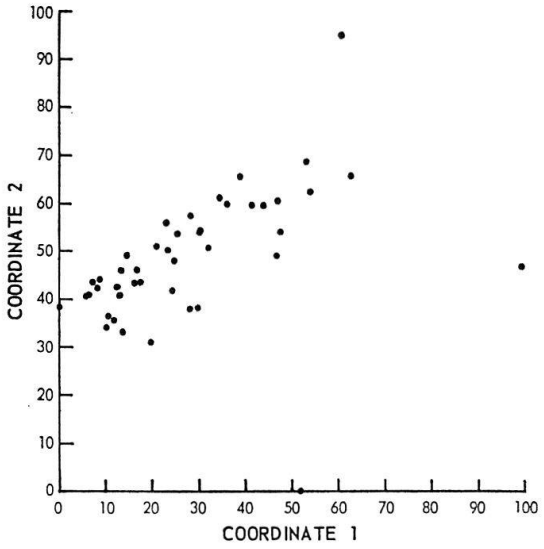

Fig. 34. Projection of the sample plots on the plane spanning the first and second main axes (D^2 ordination, Canadian data).

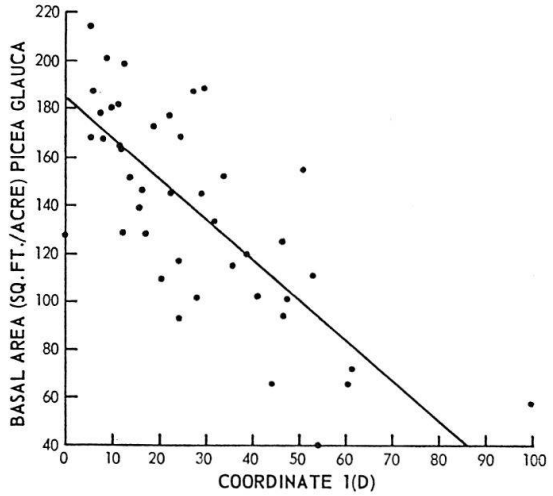

Fig. 35. Relationship between the basal area (sq. ft./acre) of *Picea glauca* and the first main axis (D^2 ordination, Canadian data).

axis and the basal area of the white spruce was –.725 (Fig. 35). No figures (except Fig. 35) are shown here because the graphs do not convey any information not already contained in Figs. 24 to 27 inclusive.

6.3.7 Principal component analysis (Q-method) of transformed D^2 matrices

The transformation, $R = e^{-D^2}$, was not successful for the reasons mentioned before (paragraph 6.2.7).

The first three eigenvalues and the coefficients of their eigenvectors resulting from the principal component analysis of the $(1 + D^2)^{-1}$ matrix are listed in Table 4.

Table 4. Eigenvalues and eigenvectors of the Canadian $(1 + D^2)^{-1}$ matrix.

	Axes				Axes	
	I	II	III	I	II	III
		Eigenvalues			Eigenvalues	
Plot number	4.0806686	1.8414362	1.6545757	4.0806686	1.8414362	1.6545757
		Eigenvectors			Eigenvectors	
1	.20518162	.32619265	−.32774792	.15197713	.02036635	−.06348675
2	.24044327	.32953310	−.23929013	.04943476	−.124469254	−.12125866
3	.25562373	.34163588	−.22040681	.05120825	−.13622218	−.12974217
4	.16712484	−.12401396	.14859947	.06658242	−.16450270	−.14332642
5	.13033962	−.19461947	−.18080734	.06418442	−.11944390	−.07672237
6	.03592506	−.06887253	−.05870816	.28425387	.02207873	.21035904
7	.11047054	−.09594629	−.03622257	.06327470	−.14348925	−.12240309
8	.05520553	−.13105740	−.12363875	.29817677	1.17514492	.07243274
9	.14213920	−.07366455	.19829468	.08458285	−.17758661	−.16518164
10	.22174998	−.09770317	.20514791	.08849949	−.18908332	−.17954269
11	.02789219	−.04236062	−.03669437	.05794918	−.09023190	−.10923668
12	.05505172	−.07654838	−.04998491	.15478582	−.18599630	−.10503147
13	.14705452	−.06792552	.03131438	.25702541	−.06008116	.33602007
14	.05373110	−.11997895	−.10965319	.11933932	−.23806116	−.19634325
15	.19527573	.04616300	.03421793	.27149510	.12739532	.17038237
16	.18256004	.28617043	−.23673914	.13846497	−.13885203	−.13263136
17	.16980892	−.13020093	.00813848	.19590994	−.02729770	.16600462
18	.07005728	−.11537413	−.10632461	.25495578	−.02206047	.28823524
19	.05662565	−.18391777	−.11756376	.02713084	−.06485623	−.06370964
20	.04151320	−.09466050	−.08648553	.02023899	−.04192575	−.03992062
21	.11655024	−.13487681	−.05092089	.10970964	−.13128951	.07590625
22	.03776468	−.09555740	−.09051711			

Trace of the matrix = sum of all eigenvalues = 43. Eigenvalue I is 4.0806686 and thus accounts for 9.49% of the total variation,

Eigenvalue II is 1.8414362 and accounts for 4.28% of the total variation,

Eigenvalue III is 1.6545757 and accounts for 3.85% of the total variation.

The first two axes thus account for 13.77%, and the first three axes for 17.62% of the total variation.

For comparison only, the relationships of *Picea glauca* and *Hylocomium splendens* with the first principal axis are shown (Figs. 36 and 37).

Projections of the plots on the planes spanning the principal axes are not shown for the same reasons mentioned under paragraph 6.2.7.

6.3.8 Habitat features in relation to principal axes (Q-method)

The habitat features were related to the principal axes mentioned above. They were found to be similar to the relationships described before. For comparison, the relationships of the basal area of *Picea glauca* with the first axis is shown in Fig. 38. The correlation coefficient was calculated to be .592 which is significant at the 0.1% level.

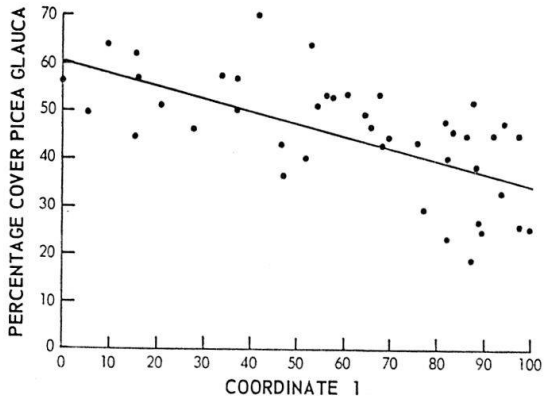

Fig. 36. Quantitative distribution of *Picea glauca* along the first principal axis ($(1 + D^2)^{-1}$ matrix, Canadian data).

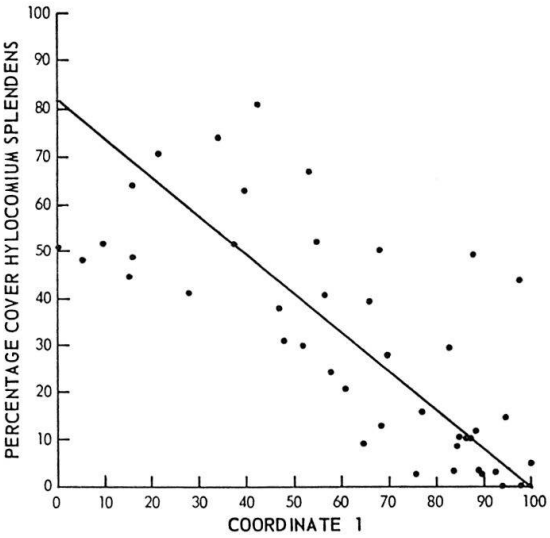

Fig. 37. Quantitative distribution of *Hylocomium splendens* along the first principal axis ($(1 + D^2)^{-1}$ matrix, Canadian data).

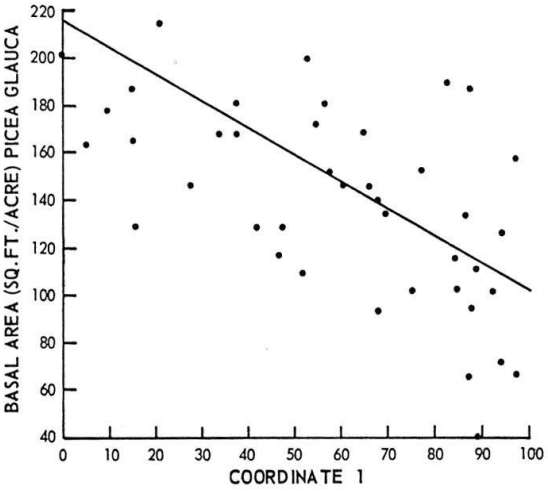

Fig. 38. Relationship between the basal area (sq. ft./acre) of *Picea glauca* and the first principal axis ($(1 + D^2)^{-1}$ matrix, Canadian data).

6.3.9 Clustering of sample plots

Neither the projection of the points representing plots on the planes spanning the principal axes, nor the projection on the planes spanning the main axes of the constellation described by the D^2 matrix, showed any tendency of clustering.

Practically all D^2's were significant ($P \leqq .05$). A group of thirteen sample plots, all with a high cover percentage for *Picea glauca* and *Hylocomium splendens*, had very few significant D^2's among them. Because the significance of the D^2's, however, is greatly dependant on the number and size of the samples, it is not a satisfactory criterion for grouping sample plots.

The D^2 matrix was also analysed by the method developed by Tocher. This analysis showed that no clustering of the points occurred.

7. Discussion

A comparison of the data obtained by three different methods of sampling indicates that there is no single procedure which is superior in all respects. An intensive investigation of this problem should involve a time-study and include different types of vegetation. This was outside the scope and interest of this study.

Of the three methods tested, the quadrat method was the quickest, but it was not generally superior either in the relative number of species sampled or in minimizing the variance. In fact, increasing the size of the sampling quadrat did not markedly decrease the variance. This is probably due to an increased error in the estimates of the cover percentages with increased size of sampling unit.

The point-quadrat-line method as used by Kershaw, was quite as efficient in the relative number of species sampled and it was much more efficient in minimizing the variance than the quadrat method. This method was, however, more time consuming than the line-interception method, due to the particular type of distribution of the species (mostly mosses).

The line-interception method was equally as efficient in the relative number of species sampled and in minimizing the variance as the point-quadrat-line method. The agreement between the data by the last two methods was very close ($r = .9994$, $P = .001$).

Summarizing, it can be stated that, in the particular vegetation types investigated, considering the accuracy of the estimates required for this study, the line-interception method proved to be the most efficient way of sampling.

Greig-Smith developed a method of continuous sampling, later improved by Kershaw, to study non-random pattern in the distribution of single species. The theory is that if species are not distributed at random, this indicates the heterogeneity of the habitat with respect to one or several factors, or groups of correlated factors, which determine the occurrence and performance of the species. This is analogous to classification, but on a small scale.

The possibility still remains that the species distribution is determined by one or several factors, without showing a non-random pattern, if the different levels of these factors are also distributed at random. Other methods should then be used.

In this study, non-random patterns were not evident within the area of the sample plots. It was suspected, however, that light played an important role in determining the distribution of the species. This aspect was studied in the Swiss forests. Only in very few cases could a relationship be demonstrated, e.g. with *Sphagnum quinquefarium* (Fig. 12). It is, however, suspected that other factors of which no measurements were taken are also involved.

Two methods were used to measure the light conditions along the same line-interceptions. The first method measured the integrated effect of the light over a 24-hour period during sunny weather; the second method made use of a photoelectric measuring device on days with a continuous cloud cover. Comparisons show completely different distributions of light on the forest floor as measured under these conditions (Fig. 13). The only exceptions occurred when the canopy was very dense, with only a few openings. Under these conditions, the distributions of light levels show closely similar patterns (Fig. 12). The distribution of different levels of precipitation, however, would also show a closely similar pattern. The question remains regarding which of these factors is the most effective in determining the distribution of *Sphagnum girgensohnii*, because no measurements of soil moisture availability were made.

The grouping of the forest vegetation sampled in Switzerland, which is evident in the projections of the points representing sample plots on the planes spanning the principal axes of the covariance matrix and on the main axes (D^2 matrix), is further objectively demonstrated by the cluster analysis due to Tocher (Appendix III). In addition, it is shown that among groups, significant differences exist in the levels of light received at the forest floor. Four of the five groups also have significantly different soil pH's.

The grouping of the Swiss sample plots, according to the Braun-Blanquet method of differential species-groups, resulted in groups of sample plots which had significant differences in average soil pH's, but with no significant

differences in the average light levels received. The grouping of the sample plots, according to the Zürich-Montpellier method, runs to some extent parallel to the grouping due to the cluster analysis. The separation of the groups along the pH-gradient is a little better in the grouping due to cluster analysis than in the Zürich-Montpellier grouping.

In the grouping due to the cluster analysis, there are significant as well as non-significant differences among the groups. The differences in the mean light levels, however, are practically all significant. In other words, some of the groups were not differentiated from others with respect to soil pH, but were differentiated from these groups by significant differences in light levels. Because the grouping of the various sample plots in the cluster analysis is directly related to the position of each sample plot in the ordination, it is more interesting to discuss this grouping after the discussion of the ecological meaning of the ordination.

Although classification or grouping of sample plots in most cases is adequate and convenient as far as simplification of the description of the vegetation and habitat conditions is concerned, it is not so satisfactory if the object of study is the underlying pattern of quantitative relationships among species, and between species and habitat factors. As far as information on the ecological behaviour of the species and of the whole vegetation is concerned, the ordinations are far more revealing than the groupings.

Groups of sample plots may have significantly different means for certain features as a result of a particular spatial distribution of these features, without having any direct relationship with the distribution of the vegetation. This will be later illustrated on the Canadian data. The possibility that the wrong conclusions are drawn, can be obviated if significant relationships are sought first between the levels of the features and the ordinations. Only the means of those features that show a relationship with the principal or main axes have to be tested for significant differences. It depends on the relative position of the groups involved, in relation to the principal or main axes, whether a significant difference will be found.

The results of the three ordinating methods are in good agreement as far as the relationships with the habitat features are concerned. The P.C.A. of the covariance matrix is, however, generally the most satisfactory method (see later discussion). Consequently, only the result of this method will be discussed in relation to the habitat features.

Since the first principal axis is related to light conditions and the third axis to soil pH (Swiss data), it is interesting to consider the extent to which each species contributes to the various principal axes.

The species that have high coefficients contributing to the eigenvector of

the first eigenvalue (first principal axis, light conditions and related factors), *Polytrichum formosum*, *Abies alba*, and *Picea abies*, are also important in the second principal axis. The second axis, however, is not related to the two habitat factors measured. Because no other measures of the habitat were taken, it was impossible to establish the ecological meaning of this axis.

It was observed that *Polytrichum f.* forms large carpets under small openings in the *Abies a.* canopy, where there is an increase in light. This agrees well with the relationships as indicated by the coefficients of the pertinent eigenvector. *Polytrichum f.* increases with decreasing *Abies a.* and increasing light.

The relationship between *Polytrichum f.* and *Picea a.* also agrees well with the observations. As *Picea a.* increases in relative importance in the stands, the canopy tends to be more open (note that *Abies a.* has a positive and *Picea a.* has a negative coefficient for the pertinent eigenvector).

The main species that contribute to the third principal axis are: *Oxalis acetosella*, *Eurhynchium striatum*, *Sphagnum quinquefarium*, *Bazzania trilobata*, and *Vaccinium myrtillus*. *Oxalis a.* and *Eurhynchium s.* tend to increase in importance with increasing soil pH. *Sphagnum q.*, *Bazzania t.* and *Vaccinium m.* increase in importance with decreasing soil pH (Fig. 39).

It is evident that the coefficient for *Sphagnum q.* in the first eigenvector is positive and rather small, indicating a small positive correlation between the quantitative measure of *Sphagnum q.* and *Abies a.* on these sample plots. This seems to be in direct contradiction to the results of the study of the relationships betwen light and species along the line-intercepts. The reason for this apparent contradiction can be found in the special ecological niche that *Sphagnum q.* occupies in these forest stands. In dense forest, it only occurs under small openings in the canopy. In a study using measuring units larger in size than these openings, this special relationship is obscured. The above shows the importance of the study of distinct ecological niches within the general habitat, in relation to the occurrence of certain species.

Bazzania trilobata also has a positive coefficient of the first eigenvector, likewise indicating a positive correlation between the occurrence of *Bazzania t.* and *Abies a.* In contrast to *Sphagnum q.*, however, this reflects a true relationship.

After foregoing observations on the relationships among the principal axes, habitat factors, and species occurrence, it is conducive to a better understanding of the significance of the grouping of the sample plots, resulting from the cluster analysis, to consider these groups in the light of the ecological and sociological affinities between the plots, expressed by the relative position of each sample plot in the ordination. An inspection of the ordination (Fig. 14)

shows that each cluster is differentiated along one or more axes. Clusters which are more or less distinct along all three axes, e.g. cluster I and V, occupy different ranges along the light, soil pH gradients (and possibly a third gradient), because these gradients are correlated to a certain degree with the principal axes. Besides being ecologically differentiated, these clusters are also distinct with regard to the level of occurrence of the species associated with these principal axes, e.g. clusters I and V respectively, have relative high and low cover-percentages for *Abies a.* (65 and 20%) and low and high cover-percentages for *Polytrichum f.* (3 and 63%). Both clusters have relative high cover-percentages for *Oxalis a.* (app. 15%).

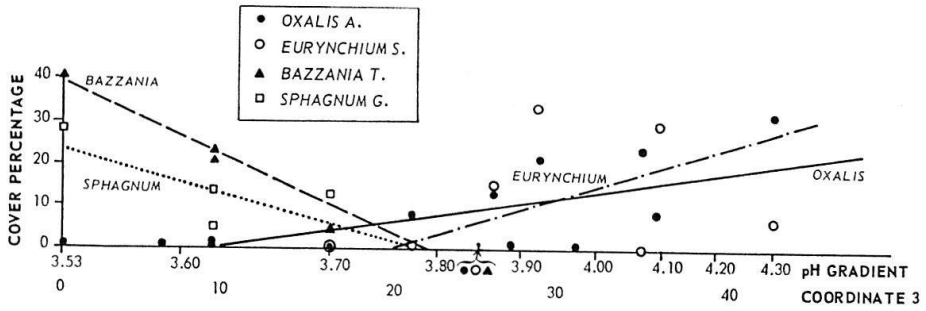

Fig. 39. Levels of occurrence (percentage cover) of *Oxalis acetosella, Eurhynchium striatum, Bazzania trilobata* and *Sphagnum quinquefarium* along the pH gradient (coordinate 3, covariance matrix, Swiss data).

Other clusters are clearly differentiated only along one axis, e.g. clusters I and III along axis 3. These two clusters occupy different ranges along the soil pH gradient (correlated with the 3rd axis), but are not significantly different with regard to light conditions. Vegetationally, these clusters are differentiated by having high and low cover-percentages for the following species: *Oxalis a.* (30 and 0%), *Eurhynchium s.* (7 and 0%); or low and high cover-percentages for *Sphagnum q.* (0 and 20%), and *Bazzania t.* (0 to 30%).

If the correlation between axis and habitat factor is not too close the relationship between the relative positions of the clusters with regard to the axis, and the level of the habitat factor, as represented by their position along the axis, breaks down. An example of this is clusters IV and V, which occupy the same range along axis 1, but which have significantly different mean levels of light, 47.89 and 15.02 kilo-Lux-hours per day, respectively.

The eigenvalues for the fourth and fifth axes are of such magnitude, that they have no practical importance.

Both the projections of the points, representing the Canadian sample plots, on the planes spanning the principal axes and the projections of the planes spanning the main axes of the constellation of points, as described by

the D^2 matrix, gave no indication of clustering. Attempts to find local concentrations of points according to Tocher's method, also failed. The conclusion must be drawn, then, that grouping according to these methods is not possible. It was found, however, that is was possible to divide the vegetation samples into three units on the basis of presence or absence of species belonging to certain groups (the differential species-group method).

A statistical analysis of the mean levels of the habitat and other features in these units revealed that division into these three groups was of limited ecological significance. Of the fifteen factors measured, only units II and III are significantly different, with regard to the field capacity of the mineral soil (15.20% and 15.36% respectively). Units I and III are significantly different with regard to the "available moisture" of the mineral soil (11.52% and 13.93% respectively) and also with regard to the height-growth of the spruce trees (1.075 and 1.182 ft/yr respectively). The significant differences between units I and III with regard both to "available moisture" and to height-growth suggests a significant relationship between the habitat and the biotic factor. If the data were taken only from the sample plots belonging to units I and III, the correlation was .43 ($P = .05$). When the data of unit II were included, however, the correlation decreased to .20 which was not statistically significant. The implication is that another habitat factor (or set of factors), about equal in importance to the growth of white spruce as the "available moisture, reaches such levels (both high and low) in the sample plots belonging to unit II, that an equally high and low level of growth is reached in this type as under the influence of the high or low levels of "moisture availability" in the sample plots of units I and III. The relationship between height-growth and "available moisture", is, then, not very stringent.

The interpretation of the ecological meaning of the principal and main axes is thought to be as follows.

Both *Picea glauca* and *Hylocomium splendens* have high coefficients, contributing to the eigenvector corresponding to the first eigenvalue. In other words, their variation is concentrated to the first principal axis. Their quantitative relationship with this axis is shown in Figs. 24 and 25. With increasing cover of *Picea g.*, *Hylocomium spl.* also increases.

The precipitation in the area investigated, is low; approximately 6 inches (15 cm) during the growing season. With increasing density of *Picea g.*, the interception by the canopy increases in proportion. Other workers have found up to 45% interception (Noirfalise 1959, Delfs 1954, Law 1956). Besides decreasing the amount of precipitation reaching the forest floor, *Picea g.* with its superficial root system is a strong competitor for moisture in

the top four inches (10 cm) of the soil. At the same time, the increasing density of the canopy results in decreasing quantities of light reaching the forest floor.

According to STÅLFELT (1937), *Hylocomium spl.* is a species which can withstand complete air drying without being killed. This explains the increasing cover of *Hylocomium spl.* under these conditions, which are too severe for other species to survive.

Pleurozium schreberi also can withstand dry conditions, but needs more light to survive. This explains why *Pleurozium schr.* increases up to a certain level with increase in density of *Picea g.*, but after reaching this peak drops off sharply (Fig. 25).

As expected, the basal area of *Picea g.* (as it is related to the cover-percentage) is strongly associated with the first principal axis and the first main axis, with correlation coefficients of –.745 and –.725, respectively (Figs. 27 and 35). P is in both cases less than .001. The basal area was primarily taken as an index of the competition between the trees. STENEKER and JARVIS (1963) found a strong correlation (r = .83) between the ten-year radial increment and the basal area of *Picea g.* within a radius of 15 feet (app. 4.5 m). It can be deduced that in the communities, typefied by a high cover-percentage of *Hylocomium spl.*, competition is extremely high. Diameter-growth slows down, but height-growth and the nitrogen content of the foliage do not seem to be affected.

Also interesting is the relationships of the "measured mean" pH of the fermentation layer, the maximum pH of the mineral soil, and the humus layer with the first principal axis (or first main axis). As the density of *Picea g.* increases (Fig. 25) and that of *Populus tremuloides* decreases (Fig. 25), the "measured mean pH" of the fermentation layer becomes gradually lower (Fig. 31). This is in agreement with the generally accepted concept that the pH of the organic layer is lower under coniferous than under deciduous trees (AALTONEN 1940, OVINGTON 1953). It is noticeable, however, that the maximum pH's of the humus layer and the top of the mineral soil show the opposite tendency. The maximum pH's become higher (Figs. 29 and 30) with increases in the density of *Picea g.* and decreases in *Populus trem.* (Fig. 25). The author believes that this can be attributed to the lesser degree of leaching which occurs under dense spruce stands and the higher degree of litter decomposition under mixed-wood stands.

The first principal axis (or main axis) could easily be mistaken to represent a developmental gradient, for it is strongly correlated with the basal area of the spruce trees which, up to a point, tends to increase with age. There is, however, very little difference in age of the trees in the different samples; all

varied between 75 and 95 years. The greater basal area on some sample plots are solely the result of a greater number of trees per unit area. The differences in the numbers of trees per unit area are thought to be the result of the action of a number of factors. All stands in the Candle Lake area originated after forest fires. The severity of the fires is an important factor determining the degree to which the original ground cover is modified, the degree to which the humus is burned, and the extent to which the roots of the original species were killed, etc. Another important factor is the availability of seed, which is governed by the distance to the seed source and the abundance of the seed (good or bad seed years). Even if all above factors are favorable, the establishment of seedlings is not ensured unless the weather cooperates with a series of moist growing-seasons after germination takes place. The differences between the densities of the stands, therefore, are primarily the result of differences in stand history and not so much the result of habitat differences or differences in the stages of development.

A large part of the variation of *Pleurozium schr.* is concentrated to the second principal axis. This second axis was found to be related only to the amount of "available moisture". It is, however, suspected that other factors of which no measurements were taken are also involved. The percentage cover of *Pleurozium schr.* increases with decreasing "available moisture" (Figs. 26 and 32).

The third principal axis, and also the third main axis, were found to be related to the "measured mean pH" of the humus layer. All the species mainly associated with these axes *(Linnaea borealis* var. *americana*, *Petasites palmatus*, *Cornus canadensis*, *Mertensia paniculata*, and to a lesser degree *Rubus pubescens*, *Fragaria virginiana*, *Mitella nuda*, and *Maianthemum canadense)* have their roots largely in this soil horizon. The pH and associated factors of this soil layer, then, are the most plausible factors in explaining the variation in distribution of these species. All the response curves along this axis had the same shape, with optima occurring between pH 4.5 and 5.5 (Fig. 27).

8. Conclusions

The purpose of organizing vegetation data is to symplify them in such a manner that a simple expression of the abundance, spacing, and other attributes of the plants emerges. Depending first on the spatial distribution of the different communities, this simplification can take two forms: an ordination and, if possible, a classification.

The foregoing study is primarily an example of the combination of two different techniques for organizing vegetation data: ordinating, and classify-

ing. A number of ordinating techniques were employed: Principal component analysis of the covariance matrix; Torgerson's method for analysing the D^2 matrix; and principal component analysis of the $(1 + D^2)^{-1}$ matrix.

The possibilities of classification were investigated by cluster analysis, based on an ordination of the sample plots, and by the differential species-group method.

Under theoretical considerations several requirements were advanced which should be satisfied by an efficient method for vegetation analysis (see chapter 4.4). The two methods for classifying and the three methods for ordinating vegetation samples will now be evaluated in the light of these requirements.

First, the ordinations will be considered separately from the cluster analysis, because theoretically the cluster analysis can be used with any of the ordinating methods. The differential species-group method is basically different from the cluster analysis and will be treated separately.

The principal component analysis of the covariance matrix and the D^2 ordination (Torgerson's method) were the most successful in ordering the vegetation samples in the simplest possible manner and account for the largest possible part of the variation within the samples. In the analysis of the Swiss data, the percentages of the variations accounted for by the first two principal axes of the covariance matrix, the first two main axes (D^2 ordination), and the first two principal axes ($(1 + D^2)^{-1}$ matrix) are 64.06, 88.17, and 14.71, respectively. In the analysis of the Canadian data these percentages are 80.00, 72.93, and 13.79, respectively.

Considering the labour involved in the construction of the main axes (D^2 ordination) and in the calculation of the percentages of the variation accounted for, the principal component analysis of the covariance matrix is preferable to the D^2 ordination. The principal component analysis of the $(1 + D^2)^{-1}$ matrix is unsatisfactory, in view of the low percentages of variation accounted for.

With regard to the second requirement, the principal axes of the covariance matrix are more closely related to the habitat features than either the D^2 ordination or the principal axes of the $(1 + D^2)^{-1}$ matrix. This is evident when the graphs of these relationships are compared. The points in the scattergrams are much more closely distributed around the fitted curves in graphs of the relationships among the principal axes of the covariance matrix and the habitat features (Figs. 16, 17, 28, up to and with 33) than in the graphs of the other relationships (Figs. 19, 20, 21, 22, 35, and 38). The correlation coefficients between the basal area of *Picea gl.* and the first principal axis of the covariance matrix, the first main axis, and the first principal axis of the $(1 + D^2)^{-1}$ matrix, are –.745, –.725, and –.592, respectively.

Requirement 3 is best satisfied by the D^2 ordination and the principal component analysis of the $(1 + D^2)^{-1}$ matrix, because both the cluster analysis and the two ordinating methods are based on the D^2 statistic. It is possible, however, using the Pythagorean theorem, to calculate the distances between the plots in the space described by the principal axes of the covariance matrix. These distances then, can be used in the cluster analysis, as were the D^2's, to determine if (and what) grouping can be recognized. Any one of the three methods, thus, easily satisfies the requirement of a common basis both for the cluster analysis and for the ordination. The use of the principal component analysis of the covariance matrix reduced computing time, which in most cases, however, is not a major consideration.

The requirement that the method should furnish a means of placing newly measured vegetation samples in a previously derived ordination or classification (4) can be satisfied in two different manners: by a quick provisional, or a slow, more accurate method. To be able to place a newly measured sample plot provisionally in an ordination or group, the distribution of some species should quantitatively be closely related to each of the axis of the ordination. This is a variation of the concept of characteristic species. Here, the quantities of the species instead of their presence are used to ordinate or classify vegetation samples. These species could conveniently be called "ordinator species". The principal component analysis of the covariance matrix fulfilled best the condition mentioned previously (see Figs. 24, 25, 26, and 27). Some of the relationships are not rectilinear. The coefficient of the eigenvector, thus, is not the best expression for the closeness of the relationship. A better statistic is the variance ratio. The variance ratio is the ratio between the mean square of the total variation due to the regression and the mean square of the residual variation. It is a measure of the goodness of fit of the regression line to the quantitative data. The goodness of fit of a number of curves can be compared directly, if the number of degrees of freedom is the same for all curves compared. The variance ratio calculated for the relationship between *Hylocomium spl.* and the first principal axis of the covariance matrix (Fig. 24) was 329.9. The variance ratio for the regression between *Pleurozium schr.* and the second principal axis, where the relationship is obviously not as close (Fig. 26), is 30.87. Both curves are highly significant (41 degrees of freedom, $P \leqq .001$).

The placing of newly measured vegetation plots in an ordination can more accurately be accomplished by multiplying the quantity of each species with the pertinent coefficient of the eigenvector and by summing the products. This will result in more precise coordinates for the sample plots.

Many ecologists and other workers (e.g. foresters) tend to feel uneasy as soon as the words "continuum" or "ordination" are mentioned. This is

mainly the result of unfamiliarity with the use that can be made of the results of an analysis based on this concept. The chief objections are the difficulty of mapping ordinated vegetation samples or of using them in management plans. This problem, however, is not always as serious as it appears to be. Even if it is impossible to classify vegetation samples, in the field it is usually quite easy to recognize stands which are rather homogeneous over extensive areas and which have sharp boundaries where they border on other stands. The stand then becomes the basic unit to be mapped and to be used in management plans. If the structure of the vegetation can be adequately described by three components (axes), each stand can be denoted by a set of three coordinates. These three coordinates describe its location in the ordination. This satisfies the fifth requirement.

In summary, the following can be suggested. Principal component analysis of the covariance matrix is superior in most of the aspects considered. The D^2 ordination developed by Torgerson ranks as the second best. The principal component analysis of the $(1+D^2)^{-1}$ matrix is unsatisfactory. It is possible, however, that other transformations of the D^2 matrices will result in a more satisfactory analysis.

Most of the foregoing considerations are also valid for the cluster analysis, because this is based on the ordinations. Since the principal component analysis of the covariance matrix results in an ordination of which the axes are most closely related to habitat features, it is to be expected that a cluster analysis based on this ordination would result in a grouping of the vegetation samples which is ecologically also most significant.

All the projections of the plots on the planes spanning the various axes indicated the same groupings, which were also supported by the cluster analysis. These groups were shown to be ecologically significantly different. Hence, no attempt was made to perform a cluster analysis on the distances in the space described by the principal components of the covariance matrix.

Both the grouping of sample plots according to the differential species-group and the cluster analysis satisfy, as most classifications do, the condition that the ordering of the vegetation samples should be simple. It is very difficult, however, to determine what part of the total variation is accounted for by the grouping according to the differential species-groups. In the grouping according to cluster analysis the variation accounted for is the same as that accounted for by the principal axes.

The cluster analysis, in the case of the Swiss data, resulted in five groups which are differentiated along one or more principal axes. A relationship among two axes and soil pH and light conditions was established. This indicates that the groups were differentiated by more than one habitat factor.

The differential species-group method resulted in three groups of sample plots, which were significantly differentiated with regard to soil pH but not with regard to light conditions.

The condition that it should be possible to place a newly measured sample plot into a previously derived classification (4) is best met by the differential species-group method. No accurate measurements of species-cover or computations are necessary. The presence or absence of the species belonging to these species-groups determines the vegetation-type in which the sample plot should be classed.

The decision to group sample plots together is, in the case of cluster analysis, based on the quantities of the species belonging to several groups, not just their presence or absence as in the differential species-group method. Within each group, the species which have a high coefficient for the eigenvectors representing that group have strong positive or negative quantitative relationships. The species belonging to such a group represented by one axis, respond highly to the level of the habitat factor (or factors) to which the axis is related. Species belonging to all groups may be present on a sample plot in different quantities, thus facilitating the quantitative differentiation from other sample plots along several habitat factor gradients. The cluster analysis thus results in a finer and ecologically more sensitive division of the vegetation than is possible by the differential species-group method. Also finer-grained classification systems, as result from cluster analysis, are less likely to have sample plots classified in the "remainder class".

As prooved since long, the vegetation units established by means of the differential species-group method are easy to map. It also satisfies the requirement of simplicity quite well and is least time consuming. Therefore, the classification according to the differential species-group method should be preferred for initially *describing and mapping* the vegetation.

The ordinating methods besides forming the basis for cluster analysis, should preferably be used to elucidate relationships within vegetation units (community-types), because the requirement of linear relationships is closest met in such a case. The data of each unit are again analysed separately. In this study, the number of Swiss sample plots in each unit was too small to warrant such a procedure. The Canadian sample plots, however, can, for practical purposes, be considered as representing one vegetation type, and, as such, the ordination of these plots can be considered an example of such an ordination within a vegetation-type.

The combination of ordination and classification into one method has several advantages:

(1) It offers an *objective* method for classifying vegetation samples.

(2) If classification is not possible, it accords the investigator an alternative in the ordination; and

(3) if classification is possible, one often would like to know if the levels of the factors in which one is interested are significantly different in the community-types recognized. This could be investigated by a variance analysis or by a "t" test. If the relationships among the factors and the axes are known, however, it can easily be shown that, in most cases, only those habitat features, that show a significant relationship with the axes, have significantly different means for the different community-types. This does not mean that they have to be significantly different. This depends on the relative position of the clusters in relation to the axes.

It was hypothesized that the correlations or covariances between the quantitative measures of the different species were not due to chance, but were reflections of the reaction of the species to their environment, including the interactions between the species. The relationships which were found to exist between the principal axes and certain habitat factors are an indication of the correctness of this concept.

The fact that the ordinations of the Canadian sample plots are not related either to the height-growth or to the nitrogen content of the white spruce foliage attracts attention. To obtain an ordination which also would be related to these factors, certain soil factors, which were not measured in this study, should be included in the analysis, or better yet they should be analyzed separately for their relationships with the height-growth and nitrogen content of the foliage.

9. Acknowledgements

During the execution of the work published here, many people have readily extended to me their cooperation, for which I wish to express my sincere gratitude.

In particular I wish to thank Professor Dr. H. Ellenberg, Director of the Geobotanical Institute of the Swiss Federal Institute of Technology in Zürich for his help in the application of the differential species-group method to my data, his constructive criticism regarding my work, and his general support, and Professor Dr. H. LeRoy, of the Departement of Animal Husbandry, of the Swiss Federal Institute of Technology for his support and consideration in his capacity as co-referendary.

I am grateful for the financial assistance provided by the Zentenar Fund of the Swiss Federal Institute of Technology. The Departement of Forestry

of the Federal Government of Canada granted me leave of absence and financial aid, for which I am much obliged.

I am much indebted to Dr. P. Ihm of the Euratom Research Centre at Ispra, Italy, whose invaluable help came at a moment when it was critically needed, for allowing me to use his computer programs and for giving most valuable advice.

I take pleasure in expressing many thanks to Dr. N. Sklov, Professor of Statistics at the University of Saskatchewan and to Dr. D. Brown of the Statistical Research Service of the Federal Department of Forestry at Ottawa for their advice and the use of certain facilities.

I wish to thank Dr. C. G. Riley, recently retired Officer-in-Charge of the Forest Pathology Laboratory at Saskatoon, for his valued support on different occasions and for his editorial comments.

Further, I am happy to name Mr. H. H. Weegar who has most efficiently and faithfully assisted me during the different phases of the work over a number of years and Mr. J. Chamberlain, who ably assisted with the field work and whose intimate knowledge of the countryside surrounding Candle Lake was of considerable benefit.

Also, I wish to express my appreciation to Mr. Siegl of the Geobotanical Institute for his help with the light measurements.

I like to recall the enjoyable visits and discussions with Mr. Peter Meyer, Chief Forester of Langenthal and with Mrs. Meyer.

Finally, I wish to express my very special thanks to Marguerite and Rein Kuin for their wonderful hospitality extended to me during my stay in Switzerland.

10. References

Aaltonen, V. T., 1940: Metsämaa. Metsämaatieteen opi-ja kasikirja, Porvoo, Helsinki.

Acton, D. F., J. S. Clayton, J. G. Ellis, E. A. Christiansen, and W. O. Kupsch, 1960: Physiographic divisions of Saskatchewan. Saskatchewan Soil Survey in co-operation with Geol. Div. Sask. Res. Council and Geol. Dept. Univ. of Saskatchewan.

Baker, F. S., 1950: Principles of silviculture. McGraw-Hill Co. Inc., New York.

Bertsch, K., 1959: Moosflora von Südwestdeutschland. E. Ulmer, Stuttgart.

Binz, A., 1961: Schul- und Exkursionsflora für die Schweiz. 10th ed., rewritten by A. Becherer.

Bray, J. R., 1956: A study of mutual occurrence of plant species. Ecology *37*, 21–28.

Bray, J. R., and J. T. Curtis, 1957: An ordination of the upland forest communities of southern Wisconsin. Ecol. Monogr. *27*, 325–349.

Canfield, R. H., 1941: Application of the line interception method in sampling range vegetation. J. For. *39*, 388–394.

Curtis, J.T., 1959: The vegetation of Wisconsin: an ordination of plant communities. Univ.Wisconsin Press.

Dagnelie, P., 1960: Contribution a l'étude de communautés vegetales par l'analyse factorielle. Bull.Service de la Carte Phytogéographique, Sér. B, tome V, fasc. 1, 7–71, and fasc. 2, 93–195.

Delfs, J., 1954: Niederschlagszurückhaltung in verschiedenen alten Fichtenbeständen. Mitt.Arbeitskreis Wald u. Wasser *1.*

Doughty, J.L., 1941: The advantages of a soil paste for routine pH determinations. Sci. Agr. *22*, 135–138.

Ehrendorfer, F., 1954: Gedanken zur Frage der Struktur und Anordnung der Lebensgemeinschaften. Angew.Pflanzensoziol. (Wien), Festschr.E.Aichinger. *1*, 151–167.

Ellenberg, H., 1953: Physiologisches und ökologisches Verhalten derselben Pflanzenarten. Ber.dtsch.bot.Ges. *65*, 350–350.
– 1954: Zur Entwicklung der Vegetationssystematik in Mitteleuropa. Angew.Pflanzensoziol. (Wien), Festschr. E.Aichinger. *1*, 134–143.
– 1956: Aufgaben und Methoden der Vegetationskunde. In: Einführung in die Phytologie. Vol. IV. Grundlagen der Vegetationsgliederung, part 1. E. Ulmer, Stuttgart.
– 1963: Vegetation Mitteleuropas mit den Alpen. In: Einführung in die Phytologie. Vol. IV, Grundlagen der Vegetationsgliederung, part 2. E. Ulmer, Stuttgart.

Fernald, M.L., 1950: Gray's Manual of Botany. 8th ed. American Book Co., New York.

Frehner, H.K., 1963: Waldgesellschaften im westlichen Aargauer Mittelland. Beitr. geobot.Landesaufn. Schweiz *44.*

Goodall, D.W., 1954: Objective methods for the classification of vegetation. III. An essay in the use of factor analysis. Austr.J.Bot. *2*, 304–324.

Greig-Smith, 1952: The use of random and contiguous quadrats in the study of the structure of plant communities. Ann.Bot., London, N.S. *16*, 293–316.
– 1961: Data on pattern within plant communities. II. Ammophila arenaria (L) Link.J. Ecol. *49*, 703–708.
– 1964: Quantitative plant ecology. Butterworth Co., London.

Grout, A.J., 1928–1940: Moss flora of North America north of Mexico. Publ. by the author, Newfane, Vermont.

Harman, H.H., 1960: Modern factor analysis. Univ. Chicago Press.

Hotelling, H., 1933: Analysis of a complex of statistical variables into principal components.J.Educ.Psych. *24.*

Hughes, R.E., 1954: VIII Congress International de Botanique, Paris. Rapports et Communications aux Section 7 et 8, 16–18.

Ihm, P., 1964: Automatic classification in anthropology; in "The use of computers in anthropology", ed. D.Hymer, Mouton & Co., 's-Gravenhage, 373–411.

Kershaw, K.A., 1958: An investigation of the structure of a grassland community. I. The pattern of Agrostis tenuis. J.Ecol. *46*, 571–592.
– 1959: An investigation of the structure of a grassland community. II. The pattern of Dactylis glomerata, Lolium perenne and Trifolium repens. III. Discussion and conclusions. J.Ecol. *47*, 31–53.
– 1961: Association and covariance analysis of plant communities. J.Ecol. *49*, 643–654.

Law, F., 1956: The effect of afforestation upon the yield of water catchment areas. J.Br. Waterworks Ass. *10*, 489–494.

Mahalanobis, P.C., 1936: On the generalised distance in statistics. Proc.Nat.Inst.Sc. (India) *12*, 49.

Mahalanobis, P.C., D.N.Majumbar and C.R.Rao, 1949: Anthropometric survey of the United Provinces (1941): A statistical study. Sankhya (India) *9*, 90–253.

Maycock, P.F., 1963: The phytosociology of the deciduous forests of extreme southern Ontario. Can.J.Bot. *41*, 379–438.

Meyer, P., 1949: Das Mastigobryeto-Piceetum abietosum im Schweizerischen Mittelland und seine forstlich-waldbauliche Bedeutung. Vegetatio *1*, 203–216.

– 1954: Das Piceeto-Abietetum praealpinum Oberdorfer 1950 asperuletosum subass. nova im Schweizerischen Mittelland. Vegetatio *5–6*, 302–308.

Mitchell, J., H.C.Moss, and J.S.Clayton, 1950: Soil Survey of Saskatchewan. Saskatchewan Soil Survey Rep. *13*.

Noirfalise, A., 1959: Sur l'interception de la pluie par le couvert dans quelques forêts belges. Bull.Soc.For.Belg. *66*, 433–439.

Ovington, J.D., 1953: Studies of the development of woodland conditions under different trees. I. Soils pH. J.Ecol. *41*, 13–34.

Pallmann, H., E.Frei, and H.Hamdi, 1943: Die Filtrationsverlagerungen hochdisperser Verwitterungsprodukte im Bodenprofil einiger Glieder der schweizerischen Braunerdenserie. Ber.schweiz.bot.Ges., B. 53 A.

Poore, M.E.D., 1955 a: The use of phytosociological methods in ecological investigations. I. The Braun-Blanquet system. J.Ecol. *43*, 226–245.

– 1955 b: The use of phytosociological methods in ecological investigations. II. Practical issues involved in an attempt to apply the Braun-Blanquet system. J.Ecol. *43*, 245–269.

Rao, C.R., 1952: Advanced statistical methods in biometric research. J. Wiley & Sons, Inc., New York.

Rowe, J.S., 1959: Forest Regions of Canada. Can.Dept.north. Aff., Nat.Res., For. Branch, Bull. *123*.

Rydberg, P.A., 1954: Flora on the Rocky Mountains and adjacent plains. 2nd ed., Hafner Pub. Co., New York.

Schlüter, H., 1957: Ein Beitrag zur Frage ökologischer und soziologischer Artengruppen. Arch.Forstwes. *6*, 44–59.

Schönhar, S., 1953: Die ökologischen Artengruppen. Mitt. Ver. forstl. Standortskartierung *3*, 26.

Stålfelt, M.G., 1937: Der Gasaustausch der Moose. Planta *27*, 30–60.

Steneker, G.A., and J.M.Jarvis, 1963: A preliminary study to assess competition in a white spruce-trembling aspen stand. For.Chron.*39*, 334–336.

Torgerson, W.S., 1952: Multidimensional scaling: I. Theory and method. Psychometrika *17*, 401–419.

van Groenewoud, H., 1956: A root disease complex in Saskatchewan white spruce. For.Chron. *32*, 11–13.

– 1961: The variation of pH and buffering capacity of the organic layer of grey wooded soils. Soil Sci. *92*, 100–105.

– 1963: A simple method for "integrated light" measurements. Ber.geobot.Inst. ETH Stiftg. Rübel, Zürich, *34*, 133–134.

– 1965: An analysis and classification of white spruce communities in relation to certain habitat features. Can. J. Bot. *43*, 1025–1036.

Table 1.

Systematische Einheit	Orno-Quercetum ilicis pinosum dalmaticae					Erico-Rosmarinetum pinetosum dalmaticae					Genisto-Ericetum verticillatae pinetosum dalmaticae				Seslerio-Ostryetum pinosum dalmaticae						Stipo-Salvietum officinalis pinetosum dalmaticae								Junipero-Pinetum dalmaticae												
Nr. der Aufnahme	1	2	3	4	5	6	7	8	9	1o	11	12	13	14	15	16	17	18	19	2o	21	22	23	24	25	26	27	28	29	3o	31	32	33	34	35	36	37	38	39	4o	41
Gebiet	H	H	B	H	B	H	H	B	H	H'	P	P	P	P	E'	B	B	B1	B1	B1	B	B	B	B	B	B	E'	B	B1	B1	B1	B1	B1	B1	B1	B1	B1	B1	B1	B1	B1
Höhe über dem Meer	48o	45o	48o	40o	42o	45o	40o	42o	50o	30o	80o	80o	50o	75o	70o	46o	64o	56o	90o	87o	65o	72o	66o	64o	74o	60o		60o	115o	118o	128o	123o	11oo	137o	135o	139o	137o	14oo	14oo	13oo	13oo
Exposition	NW	N	NW	N	S	O	S	O	N	N						O	SO	N	S	W	N	-	-	N	NW	-		-	SW	SW	S	NO	NO	SW	SW	SW	NO	SW	SW	O	SO
Inklination (∡)	4o	4o	4o	5o	3o	1o	2o	25	2o							5o	4o	25	2o	25	1o	-	-	15	1o	-		-	4o	4o	3o	45	25	3o	3o	35	25	4o	35	25	3o
Deckungsgrad in %	9o	9o	9o	9o	8o	9o	9o	8o	9o							8o	6o	85	8o	9o	8o	5o	7o	5o	5o	6o		7o	7o	75	8o	65	8o	65	7o	7o	9o	7o	9o	9o	8o
Grösse der Aufnahme in m^2	3oo	3oo	2oo	3oo	2oo	3oo	2oo	3oo	2oo	3oo	3oo	4oo	15o	2oo		3oo	2oo	3oo	5oo	4oo	2oo	3oo	3oo	3oo	3oo	2oo	3oo	3oo	2oo	2oo	2oo	5oo	4oo	1oo	1oo	1oo	1oo	1oo	2oo	1oo	1oo

Klassen-, Ordnung-, Verband-, Associations-Zugehörigkeit	Art
VO	Fraxinus ornus (Strauch)
VO	Fraxinus ornus (Baum)
AI	Quercus ilex (Strauch)
AI	Quercus ilex (Baum)
VCB	Brachypodium ramosum
VO	Coronilla emeroides
AI	Lonicera implexa
AP	Pistacia terebinthus
AI	Rhamnus alaternus
VCB	Convolvulus elegantissimus
VE	Juniperus phoenicea
AI	Rosa sempervirens
KI	Arbutus unedo
VE	Dorycnium hirsutum
KI	Phillyrea media
KQ	Viola silvestris
KQ	Festuca heterophylla
AI	Carex distachya
AE	Erica multiflora
AE	Rosmarinus officinalis
VE	Erica verticillata
	Helianthemum nummularium
VE	Hieracium stupposum
OSC	Linum tenuifolium
AG	Asperula scutellaris
AG	Genista sericea var. rigida
AG	Linum elegans
VI	Lonicera etrusca
AG	Salvia brachyodon
AG	Genista silvestris subsp. michelii
	Micromeria croatica
ASO	Ostrya carpinifolia
VO	Acer monspessulanum
VO	Rhamnus rupestris
VO	Sesleria autumnalis
AO	Astragalus illyricus
AP	Rhamnus intermedia
	Ornithogalum tenuifolium
KB	Tunica saxifraga
VF	Helianthemum nitidum
	Cerastium trichogynum
KB	Teucrium polium
OSC	Festuca vallesiaca
	Paronychia kapela
	Trifolium aureum
	Allyssum petraeum
	Rosa rubiginosa
	Lappula echinata
	Viola arvensis
	Carduus micropterus
	Satureia acinos
VCS	Euphorbia myrsinites
OSC	Bromus erectus
	Hieracium cymosum var. laxiflorum
	Myosotis collina
	Sedum ochroleucum
	Festuca duriuscula
KA	Ceterach officinarum
	Carduus candicans subsp. cylindricus
	Arabis muralis
AC	Campanula portenschlagiana
	Trifolium pratense var. nivale
VPM	Juniperus sibirica
AJ	Sesleria nitida (robusta)
	Polygala vulgaris
	Muscari botryoides
AC	Cerastium grandiflorum
AD	Cynanchum adriaticum
KA	Asplenium ruta muraria
KA	Asplenium trichomanes
	Arenaria serpyllifolia
	Festuca rubra subsp. fallax
VSV	Centaurea triumfetti
VPM	Juniperus sabina
	Senecio rupester
	Luzula campestris subsp. multiflora
	Poa pumila
OS	Ranunculus montanus subsp. carinthiacus
	Lathyrus pratensis
OF	Lilium martagon var. cattaniae
	Hieracium psammogenes
AA	Linaria cymbalaria
KT	Poa alpina var. vivipara
OQ	Trifolium alpestre
VE	Juniperus oxycedrus (Strauch)
VE	Juniperus oxycedrus (Baum)
VO	Genista silvestris subsp. dalmatica
OSC	Teucrium chamaedrys
OSC	Eryngium amethystinum
	Koeleria splendens
	Rubus dalmatinus
	Echinops ritro
VCC	Picris hispidissima
VCS	Edraianthus tenuifolius
AJ	Pinus nigra subsp. dalmatica s.l. (Baum)
AJ	Pinus nigra subsp. dalmatica s.l. (Strauch)
OQ	Galium lucidum
VPA	Bunium montanum
KB	Sanguisorba muricata
ASS	Salvia officinalis
VCS	Aethionema saxatile
KB	Lotus corniculatus var. hirsutus
VO	Prunus mahaleb (Strauch)
VO	Prunus mahaleb (Baum)
VST	Sesleria tenuifolia
VCS	Teucrium montanum
KQ	Hedera helix
VCS	Carex humilis
	Geranium purpureum
	Cerastium viscosum
VCC	Inula candida
VCS	Globularia cordifolia
VCS	Satureia montana
VCS	Stachys subcrenata
AC	Moltkea petraea
	Cephalaria leucantha
	Hypericum perforatum
OSC	Brachypodium pinnatum
OSC	Asperula longiflora
	Leontodon crispus
	Potentilla hirta
KI	Asparagus acutifolius
KB	Helichrysum italicum
	Leucanthemum cinerariaefolium
	Muscari comosum
KC	Centaurea calcitrapa
VO	Cnidium silaifolium
	Thymus balcanus
	Tragopogon balcanicus
	Hieracium biokovoënse
VF	Helianthemum grandiflorum
OQ	Dorycnium germanicum
VCS	Euphorbia spinosa
VI	Clematis flammula
AG	Veronica orbiculata
OSC	Helianthemum italicum
VCS	Onosma javorkae
	Hieracium florentinum
	Stachys italica
	Polygala nicaeensis
	Plantago lanceolata
	Veronica dentata subsp. austriaca
	Medicago orbicularis
	Clematis vitalba
KQ	Prunus spinosa
KQ	Sorbus domestica
KE	Fragaria vesca
OSC	Thymus longicaulis
OSC	Hippocrepis comosa
	Valeriana tuberosa
	Armeria canescens
	Cerastium campanulatum
VSV	Dorycnium herbaceum
KT	Rumex scutatus
ASH	Thalictrum aquilegifolium
VO	Colutea arborescens
VO	Rhamnus saxatilis
KQ	Rhamnus fallax
VSV	Ononis spinosa
KQ	Crataegus oxyacantha
KQ	Crataegus monogyna
KQ	Geranium robertianum
VE	Pinus halepensis (Baum)
VE	Pinus halepensis (Strauch)
VE	Spartium junceum
VE	Cistus salviaefolius
OF	Fagus silvatica (Baum)
	Spiraea ulmifolia
OQ	Sorbus aria

Appendix I

Vegetation table, Swiss data

Differential species-groups		Vegetation unit	I					II					III				
		Plot number	1	2	3	6	12	7	11	13	14	15	4	5	8	9	10
		Tree cover (%)	85	70	75	75	50	55	80	78	90	45	75	80	80	80	85
		Shrub cover (%)	1/10		3	20	2	20	40	0	0	70	1	13	0	0	0
		Herb cover (%)	60	30	35	25	40	40	20	1	1/10	1/10	2	15	1/100	30	3
		Moss cover (%)	1/10	85	60	90	50	55	90	80	95	70	90	95	90	95	80
		Trees															
		Abies alba	4	2	4	2	3	2	+	3	4	2	3	4	4	3	5
		Picea abies		1	1	3	1	3	4	3	2	3	1	3	1	2	
		Fagus silvatica	2	1	1	2	1			+			+			1	
		Quercus robur			+	1			+								
		Shrubs (1 - 5 m)															
		Abies alba	+		+	+	+	2	3			4	+	2			
		Picea abies				2							+				
		Fagus silvatica				1	+	+	+				+				
A		Differential species															
	H	Vaccinium myrtillus		+		1	+	3	2	1	+	+	+	3	+	+	+
		Abies alba (seedlings)	+	+		+	+	+	+	1	+	+	1	2		+	+
		Picea abies (seedlings)		+		2	+	+	+	1	+	+	+	1		+	+
	M	Hylocomium splendens		+		+	+	2	+	1	1	+	2	3	4	+	+
		Rhytidiadelphus triquetrus	+	1	1	2	2	+	+						+	+	
		Polytrichum formosum	+	4	1	5	2	2	4	3	+		3	3	2	2	1
		Thuidium tamariscifolium	+	1	1	1	1	+	+	1	+	4	1	1	3	1	+
B	H	Anemone nemorosa	+		+		+										
		Fragaria vesca	+	+	+		+										
		Hedera helix			+	+	+										
		Lysimachia nemorum	+	+	+	+	+										
		Viola silvatica					+										
		Galium rotundifolium	+			+	+										
	M	Catharina undulata				+	+										
		Mnium undulatum	+		+		+		+								
		Mnium affine			2	+	+		+								
C	H	Athyrium filix-femina	+	+	+	+	+		+	+			+				
		Dryopteris austriaca	+	1	+	1	+		+		+	+		+			
		Luzula pilosa	+	+		+	+	+	+	+	+		+				
		Majanthemum bifolium	+	+	+	+	+	+	1					+			
		Oxalis acetosella	3	2	3	2	3		2	+	+	+	+				
		Rubus spec.	+	+	+	1	+	+	+								
	M	Eurhynchium striatum	+	+	2	1	3		1	2	3	2					
D	M	Pleurozium schreberi				+		2		1				1	+	+	+
		Rhytidiadelphus loreus							2					3	+	1	
		Dicranum scoparium				+		+	+	+		+	1	1	1	+	+
		Hypnum cupressiforme				+	+	+	+	+	+	+	1	1		1	+
		Plagiothecium undulatum								+		+			+		
E	M	Bazzania trilobata											2	3	+	3	3
		Spagnum girgensohnii											2	2		1	2
		Other species															
		Herb layer															
		Melampyrum pratense		+		+		+	+							+	+
		Blechnum spicant						+					+	+			
		Carex brizoides				+			1								
		Carex pilulifera				+											
		Dryopteris filix-mas							+				+				
		Solidago virgaurea			+		+									+	
		Fagus silvatica (2 cm < seedlings)	+	+	+		+	+		+							
		Quercus robur							+								
		Moss layer															
		Plagiochila asplenoides		+	3	1	+		1		+		1		2		
		Leucobryum glaucum	+							+							
		Lophocolea bidentata				+			+		+						
		Lophocolea heterophylla								+	+						
		Calypogeia trichomanis				+			+								
		Dicranella heteromalla				+				+	+			+	+		
		Lepidozia reptans												+			
		Picea abies (seedlings < 2 cm)													+		

Note: Instead of *Spagnum girgensohnii* read: *Sp. quinquefarium*, instead of *Catharina Catharinea*, instead of *asplenoides asplenioides*. Appendix II: Instead of *Actea Actaea*.

Appendix II

Vegetation table, Canadian data

Vegetation unit	I														II																		III.								
Plot number	37	27	31	22	11	35	43	10	36	9	32	14	29	23	39	20	25	40	26	30	4	28	41	33	24	42	34	12	15	2	18	21	17	38	3	1	7	19	13	8	6	5	16
Tree cover (%)	64	50	65	65	45	65	47	57	65	53	65	50	36	60	55	70	50	57	60	70	40	55	72	60	50	50	60	51	60	59	53	70	50	70	65	60	53	62	52	56	45	70	70
Shrub cover (%)	1	5	1	8	2	1	1	1	1	1	1	5	1	1	1	1	2	1	2	1	2	1	3	1	1	8	1	1	1	1	2	1	1	1	1		1	1	1	2	1	1	1
Herb cover (%)	5	45	30	70	75	10	20	20	20	10	5	50	45	20	10	65	65	13	50	10	7	15	80	7	40	90	17	62	25	4	40	40	30	25	1	4	37	50	25	36	60	12	1
Moss cover (%)	80	30	31	1	74	83	70	76	25	98	5	22	15	85	85	7	4	85	7	75	67	75	5	12	4	12	47	78	90	80	40	65	60	35	73	85	75	45	85	30	35	30	95
Trees																																											
Picea glauca	4	3	3	3	3	3	3	3	4	4	3	3	2	4	3	3	3	4	3	4	3	3	2	3	2	3	3	4	4	4	3	3	3	4	4	4	4	3	4	2	3	4	4
Populus tremuloides		1	3	+	+			2	2	+	2	2	2			3	3		2		+	2	3		2	3	2			+				2	+	+	+	2	+	3	+	1	+
Populus balsamifera	+			3		1	1		+		+			+	+	+		+		1	+				+					+	+	+	+	+			+				+	+
Abies balsamea								+							+														1		+		+								+
Picea mariana																									+										+							
Pinus banksiana																												+									+			+		
Salix spec.								+										+																				+			
Betula papyrifera																																									+	
Shrubs																																										
Viburnum trilobum	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+			+	+	+	+	+		+	+	+	+	1	+	+			+		+	+	+	+
Symphoricarpus albus	+	+		+		+	+		+	+	+		+	+	+	+	+	+	+	+	+	+	+		+	1			+	+		+		+	+			+			+		
Cornus stolonifera	+	+		1		+			+	+	+				+	+	+				+					+	+			1	+			+						+	+	
Alnus crispa				+	+							1						+					+					+										+		1	3	+	
Ribes lacustre		+		+		+		+		+			+		+			+			+			+	+				+	+						+					
Rubus idaeus			+	+				+	+				+		+	+							+			+					+			+	+			+			
Ribes americanum										+				+		+		+							+						+								+			
Differential species																																										
S Lonicera involucrata	+	+		+		+	+	+	+	+	+				+	+		+		+	+	+	+	+				+		1		+		+							
Lonicera dioica	+	+	+	+		+	1	+	+				+		+	+			+		+	+	+			+			+	+			+								
Ribes hirtellum	+	+	+	1		+	+	+		+	+				+		+			+	+	+	+	+									+								
Ribes triste		+	+	+	+	+	+	+	+		+						+				+		+	+		+	+					+									
Shepherdia canadensis		+		+			1											+			+	+		+	+	+	+			+											
Amelanchier alnifolia	+				+					+													+			+				+						+					
H Lathyrus orchroleucus	+	+	+	1	+	+		+	+			+	+		+	+	+	+			+		+	+	+	1	+				+	+	+				+	+	+		
Habenaria obtusata	+		+		+					+	+	+		+	+	+					+	+			+	+	+														
Geocaulon lividum	+	+				+	+	+			+	+	+	+	+	+		+		+					+		+			+	+										
Actea rubra	+	+			+		+	+		+				+	+		+	+	+		+		+	+			+														
Galium boreale	+			+	+							+				+									+			+													
M Eurynchium pulchellum		1	+		+				+		+	1	+			+	+		+	+		+		+	+			+	+		+										
H Galium triflorum	+	+		+			+	+			+	+							+	+																						
Elymus innovatus	+		+	+	+				+		+		+				+																									
Equisetum scirpoides	+			+		+	+								+	+																									
Equisetum pratense	+		+	+						+																																
Hieracium canadensis		+		+									+	+																												
Carex capillaris						+								+																												
M Peltigera spec.	+					+		+	+	+								+																								
Other species																																										
Herbs																																										
Linnaea borealis v. americana	+	2	+	+	+	+	+	2	2	2	+	3	2	1	1	2	2	1	2	1	+	1	2	+	1	2	3	2	1	+	1	2	2	1	+	+	2	2	2	1	3	+	+
Cornus canadensis	+	2	2	2	2	+	+	1	+	1	+	2	2	+	+	2	2	+	2	1	+	1	3	1	2	2	+	2	1	+	2	2	1	1	+	2	1	1	2	3	+	+
Rosa acicularis	+	+	+	1	+	+	+	+	+	+	+	1	+	+	+	+	1	+	1	+	+	+	+	+	+	+	+	+	+	+	1	1	1	1	+	+	2	1	1	2	+	+	
Petasites palmatus	+	1	1	2	2	+	+	1	+	1	+	+	2	1	+	1	1	+	1	1	+	+	+	2	+	1	2	+	2	1	+	1	+	1	+			2	1	+	1	+	
Mertensia paniculata	+	2	1	2	2		+	1	+	+	+	+	+	+	+	2	2	+	2	+	+	+	1	+	1	1	+	2	1	+	2	+	1	+	+	+		1	+	+	+	+	
Fragaria vesca	+	1	+	2	+	+	+	1	+	+	1	1	1	+	+	1	1	+	1	+	+	+	+	+	2	1	+	1	+	1	+	1	1	1	+			1	+	1	+	+	
Fragaria virginniana	+	+	+	2	+	+	+	+	+	+	+	+	1	+	+	1	+	+	+	+	+	+	1	+	1	1	+	+			+	+	+	+	+			1	+	+	+	1	+	
Mitella nuda	+	1	+	1	2	+	+	1	+	+	1	+		1	+	1	1	+	+	+	+	+	+	+	+	+	+	2	1	+	1	+	+	+	+	+	+	1	1		+	+	
Pyrola virens	+	+	+		+	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	1	+					+	+	+	+	+	
Pyrola secunda	+	+	+	+	+			+	+	+	+	+	1	+	+	+	+	+	+	+	+	+	+	+	+	+		+	+	+	+	+	+	+	+	+			+	+	+	+	
Rubus pubescens	+	1	+	2	2	1	+	1		+		1	+	1	+	2	1		1	+	+	+	1	+	1		+	2	1	+	1	1	1			+	+	+	1	+	1	+		
Maianthemum canadense	+	+	+	+	+	+		+	+	+		1	+	+	+	+	+		+			+	2	+	+	2	+	1	+		+	+	+	+	+			+	+	+	+	
Aster puniceus	+	+	+	1	+	+	+	+	+	+	+	+	+	+	+	1	1	+	+	+	+	+	+	+	+		+	+	+	+	+	1	1				+	1	1	+		
Viola renifolia	+	+			+	+	+			+	+	+	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		+				1	+			
Pyrola asarifolia	+		+	+	+			+			+	+	+			1	1	+	+		+	+	+	+	1	+	+	+		+	+	1				+	+	+	+	+	+	
Achillea millefolium	+	+	+	+	+	+	+	+	+	+	+		+		+	+	+	+			+				+	+					+	+							+		
Aralia nudicaulis	+			+	+	+	+					+											+		2	2		+	+		1	+	+		+		+	+	+			
Carex disperma	+	+	+	+		1	+	+	+		+			1	+	+	+		+	+	+	+		+			+		+			+					1					
Solidago hispida	+	+	+		+	+	+		+	+			+	+	+			+	+	+			+		+		+	+		+		+						+	+	+	
Trientalis borealis	+		+	+	+	+						+		+	+	+			+					+		+	+		+		+	1	+					+	1	+	+	
Vicia americana		+					+			+	+			+	+		+	+	+		+				+		+							+						+		
Arctostaphylos uva-ursi						+	+	+	+			+	1							1					1		+			+		+										+
Epilobium angustifolium		+			+							+	+				+	+					+		+		+				+				+			+	+			
Taraxacum spec.						+	+				+				+			+	+								+									+			+			
Vaccinium vitis-idaea v. minus					+		+					+														+	+					+							+			
Ledum groenlandicum		+																		+		+	+			+	+											+	+			
Moneses uniflora														+		+				+							+	+		1			+							+		
Rubus acaulis	+			+	+								+					+																			+			+		
Goodyera repens						+				+						+			+					+			+										+	+				
Lycopodium complanatum					+							1	+														+										+			+		
Equisetum arvense								+	+		+										+																			+		
Lycopodium clavatum					+							+																											+	+		
Actea alba			+								+																		+										+	+		
Equisetum silvaticum	+		+														+																									
Viola rugulosa												+				+																										
Habenaria orbiculata																				+						+																
Mosses																																										
Hylocomium splendens	4	2	2	+	4	3	1	3	2	2	+	2	1	4	4	1	+	4	1	3	3	2	+	1	+	1	3	4	4	5	3	3	3	2	3	5	4	2	3	3	2	2	5
Pleurozium schreberi	3	1	1		2	3	+	2	2	4	+	+	1	2	2	+	+	2	+	2	2	2		+		1	2	2	3		1	3	+	3	2	3	1	2	1	+		
Dicranum rugosum	+	+	+		+	1	4	+	+	+			+	+	+		+	+	+	+	+	+	+			+	+	+	+		-	1		+	+		+	+	+	+	+	+	
Hypnum crista-castrensis			+		+			+				+	+	+				+	+		+					+	+	+			+		+			+	+	+			+
Mnium cuspidatum	+			+				+	+	+							+												+		+					+	+			+		
Thuidium recognitum		1			+							1	+		+	+		+							+						1								+			
Rhytidiadelphus triquetrus													+																		+											+
Polytrichum juniperinum																											+													+			

Other species: Alnus rugosa 17(+), 22(+); Corallorhiza maculata 3(+), 1(+); Empetrum nigrum 29(+), 43(+); Smilacina stellata 32(+), 39(+); Hedysarum alpinum 12(+); Habenaria dilatata 42(+); Lycopodium obscurnum 11(+); Deschampsia caespitosa 22(+); Lilium philadelphicum v. andinum 5(+).

Appendix III

Computational scheme for finding clusters (Swiss data)

Plot No.	D^2	Mean D^2	Increase
13–14	18.367907	18.367907	
15–13	24.239020		
15–14	26.371720		
	60.610740 : 2 =	30.305370	11.937463
78–15	25.544942		
8–13	26.386967		
8–14	36.276877		
	88.208786 : 3 =	29.402929	–0.902441
12– 8	50.735104		
12–15	58.855423		
12–13	27.257727		
12–14	45.991048		
	182.839302 : 4 =	45.709826	16.30689
4–12	72.138194		
4– 8	46.041220		
4–15	74.711975		
4–13	42.746613		
4–14	53.580815		
	289.218817 : 5 =	57.843763	12.133937
1– 4	151.77657		
1–12	76.46608		
1– 8	132.04625		
1–15	147.99643		
1–13	116.10162		
1–14	160.93687		
	785.32382 : 6 =	130.887303	73.043540
9–10	47.815447	= 47.815447	
5– 9	72.679807		
5–10	119.59655		
	192.276357 : 2 =	96.138179	48.322732
1– 5	254.41571		
1– 9	147.25925		
1–10	189.28631		
	590.96127 : 3 =	196.987090	100.848911
2–11	61.445963	61.445963	
6– 2	73.735058		
6–11	60.437793		
	134.172851 : 2 =	67.086426	5.640463

Plot No.	D^2	Mean D^2	Increase
7– 6	255.19504		
7– 2	350.06796		
7–11	277.88278		
	883.14578 : 3	= 294.381927	227.295501
1– 3	169.41938	= 169.41938	
7– 1	371.67753		
7– 3	527.89705		
	899.57458 : 2	= 449.787290	280.367910

Appendix IV

Computation of Mahalanobis' generalized distance (D^2)

If the species are not quantitatively correlated the generalized distance, D^2, between the sample plots simply is the sum of the squares of the differences in average percentage cover for the various species. When the species are correlated, however, the cover percentages can be replaced by a set of transformed variates which are linear functions of the original cover percentages and which are mutually uncorrelated (Rao 1952). The following method of transforming correlated into uncorrelated variables is due to Rao (Mahalanobis *et al.* 1941, Appendix 5, p. 251; and Rao 1952).

If $x_1, x_2 \ldots\ldots x_m$ represent the original variables (indices for plots are not given, N measurements for each species), transformed to unit standard deviation and r_{ij} is the correlation between the i^{th} and j^{th} species, a system of new variables $y_1, y_2 \ldots\ldots y_m$ can be defined by

$$\begin{aligned} y_1 &= x_1 \\ y_2 &= x_2 - a_{21}y_1 \\ y_3 &= x_3 - a_{32}y_2 - a_{31}y_1 \\ &\vdots \\ y_m &= x_m - a_{mm-1}\, y_{m-1} \ldots\ldots - a_{m1}\, y_1 \end{aligned}$$

in which the constants $a_{21}, \ldots\ldots a_{mm-1}$ are selected in such a manner that the correlation between the y's become zero (the y's are then said to be linearly independant).

In order to find the first coefficient a_{21} the covariance between y_1 and y_2 must be made equal to zero.

$$\text{cov}\,(y_1y_2) = \text{cov}\,(x_1x_2) - a_{21}V\,(y_1) = 0$$

in which (cov y_1y_2) denotes the covariance of y_1 and y_2, V (y_1) denotes the variance of y_1. Since Vx_1, Vx_2, and Vy_1 all have standard deviation = 1.

$$\text{cov}\,(y_1y_2) = r_{21} - a_{21} = 0$$

thus $a_{21} = r_{21}$

and $V(y_2) = 1 - r_{21}^2$

To facilitate the computation of the constants a, it is advantageous to introduce at this stage a new constant b, as will be defined below. The computational steps to obtain the constants a_{31} and a_{32} are now as follows:

$$b_{31} = \text{cov}\,(x_3x_1) = r_{31} \qquad a_{31} = b_{31}$$
$$b_{32} = \text{cov}\,(x_3x_2) - a_{21}b_{31} \qquad a_{32} = b_{32}/V(y_2)$$
$$= r_{32} - a_{21}b_{31},$$
$$V\,(y_3) = 1 - b_{31}a_{31} - b_{32}a_{32}$$

The computational steps to find y_4 are

$$b_{41} = r_{41} \qquad a_{41} = b_{41}$$
$$b_{42} = r_{42} - a_{21}b_{41} \qquad a_{42} = b_{42}/V(y_2)$$
$$b_{43} = r_{43} - a_{32}b_{42} - a_{31}b_{41}, \qquad a_{43} = b_{43}/V(y_3)$$
$$V\,(y_4) = 1 - b_{41}a_{41} - b_{42}a_{42} - b_{43}a_{43}$$

In general if $y_1, y_2 \ldots\ldots y_{i-1}$ are known, the steps to calculate y_i are

$$b_{i1} = r_{i1} \qquad a_{i1} = b_{i1}$$
$$b_{i1} = r_{i2} - a_{21}b_{i1} \qquad a_{i2} = b_{i2}/V(y_2)$$
$$b_{i3} = r_{i3} - a_{32}b_{i2} - a_{31}b_{i1} \qquad a_{i3} = b_{i3}/V(y_3)$$
$$\vdots$$
$$b_{ij} = r_{ij} - \sum_{t=i-1}^{1} a_{jt}b_{it}, \qquad a_{ij} = b_{ij}/V(y_j),\ (j \leqq i-1)$$

$$V\,(y_i) = 1 - \sum_{j=1}^{i-1} a_{ij}b_{ij}$$

Since the calculation of each new constant depends on the constants derived previously it is necessary to keep checking for errors.

At the end, the new variables are transformed to unit standard deviation by deviding each variate by its standard deviation, $\sqrt{V(y)}$.

Now the calculation of the "generalized distances" can progress with very little difficulty by applying the Pythagorean theorem. If $d_1, d_2 \ldots\ldots d_m$ are the differences, in transformed data of the species, between two sample plots, the distance (D^2) is calculated as follows

$$D^2 = d_1^2 + d_2^2 + \ldots\ldots + d_m^2$$

It is clear that when the number of species is large, the computation of the D^2's becomes very time consuming. The use of large electronic computers, however, has made it possible to apply the method to almost any number of species (up to 200).

Appendix V

Principal component analysis

Principal component analysis of the two types of matrices, vector or point representation, is the same.

Because the mathematics of principal component analysis involves the manipulation of matrices, a short review of basic matrix operations is presented here, before proceeding to the explanation of the analysis.

(1) A vector is a directed line segment in test space which is described by a one-dimensional array of ordered numbers, which is arranged either as a column or as a row. The numbers are the coordinates of the end point of the vector. The beginning of all vectors is in the origin zero. Every vector is thus adequately described by the coordinates of its endpoint. The numbers are ordered because a vector with the coordinates (x_1, x_2, x_3) is obviously not the same as the vector with the coordinates (x_2, x_3, x_1) or (x_3, x_2, x_1).

Example:

$$\alpha = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$ (a vector in two dimensional test space).

$\beta = (x_1, x_2, x_3)$ (a vector in three dimensional test space).

(2) A matrix is a two-dimensional array of numbers
Example:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

a_{23} denotes that element of the matrix A which is found in the second row and the third column.

(3) Matrices are added or subtracted by adding or subtracting the corresponding elements of the matrices.

Example:

if $$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$ and $$B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

then $$A + B = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{pmatrix}$$

and $$A - B = \begin{pmatrix} a_{11} - b_{11} & a_{12} - b_{12} \\ a_{21} - b_{21} & a_{22} - b_{22} \end{pmatrix}$$

(4) In matrix multiplication the sum of the products of the elements from a row of the first matrix and a column of the second matrix are computed for each combination of rows and columns.
Example, using the same matrices A and B:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

$$A \cdot B = \begin{pmatrix} a_{11} \cdot b_{11} + a_{12} \cdot b_{21} & a_{11} \cdot b_{12} + a_{12} \cdot b_{22} \\ a_{21} \cdot b_{11} + a_{22} \cdot b_{21} & a_{21} \cdot b_{12} + a_{22} \cdot b_{22} \end{pmatrix}$$

In the foregoing example A is called the premultiplier and B the postmultiplier. If A has more rows than B has columns it is not possible to form the product.

(5) To multiply a matrix with a scalar (ordinary algebraic quantity, real or complex number) each element of the matrix is multiplied with the scalar.

(6) Post multiplying of a matrix with m columns by a column vector with m elements produces a column vector with as many elements as the matrix has rows.

(7) The matrix I, in which all the elements of the main diagonal are unity (1) and all the off-diagonal elements are zero, is called the identity matrix.

An example is:

$$I = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

In matrix algebra it plays the same role as unity (the number 1) does in ordinary algebra. It is important in the inversion of matrices.

(8) Matrix division involves an operation called matrix inversion. If a matrix A is divided by a matrix B then this is not denoted by A/B but by $A \cdot B^{-1}$ (B^{-1} is called B inverse). The matrix B^{-1} has the property that $B^{-1} \cdot B = I$ (identity matrix).

(9) The determinant of a square matrix (number of columns = number of rows) is a single number which represents a unique function of the numbers in the matrix. The determinant is calculated from the elements of the matrix. By definition a determinant of the n^{th} order (n rows, n columns) stands for the sum of n! (n factorial) terms ($n! = 1 \times 2 \times 3 \ldots\ldots n$) each of which is the product of the n elements, only one from each column, and only one from each row. The signs attached to these products must satisfy the following rule. If the number of interchanges in each permutation is even, the sign is +, if the number is odd the sign is —.

example $$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

The six terms entering in the expansion of the determinant are:

	order of rows			number of interchanges	sign
$a_{11} \cdot a_{22} \cdot a_{33}$	1	2	3	0	+
$a_{11} \cdot a_{32} \cdot a_{23}$	1	3	2	1	—
$a_{21} \cdot a_{12} \cdot a_{33}$	2	1	3	1	—
$a_{21} \cdot a_{32} \cdot a_{13}$	2	3	1	2	+
$a_{31} \cdot a_{12} \cdot a_{23}$	3	1	2	2	+
$a_{31} \cdot a_{22} \cdot a_{13}$	3	2	1	3	—

The determinant in expanded form thus is as follows:

$$A = a_{11} \cdot a_{22} \cdot a_{33} - a_{11} \cdot a_{32} \cdot a_{23} - a_{21} \cdot a_{12} \cdot a_{33} + a_{21} \cdot a_{32} \cdot a_{13} + a_{31} \cdot a_{12} \cdot a_{23} - a_{31} \cdot a_{22} \cdot a_{13}$$

(10) The characteristic equation of a square matrix is formed by subtracting from the diagonal elements of the matrix some value, which is chosen so that the determinant of the new matrix is equal to zero.

example $$|A - \lambda I| = \begin{vmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ a_{21} & a_{22} - \lambda & a_{23} \\ a_{31} & a_{32} & a_{33} - \lambda \end{vmatrix} = 0$$

In expanded form the determinant may be written as follows:

$$(a_{11} - \lambda)\ (a_{22} - \lambda)\ (a_{33} - \lambda) - (a_{11} - \lambda)\ (a_{32})\ (a_{23}) - (a_{21})\ (a_{12})\ (a_{33} - \lambda) + (a_{21})\ (a_{32})\ (a_{13}) + (a_{31})\ (a_{12})\ (a_{23}) - (a_{31})\ (a_{22} - \lambda)\ (a_{13}) = 0$$

The various non-zero λ's, are usually called roots, eigenvalues or latent roots. In plant ecology the R matrix (matrix of correlation coefficients or covariances between species) is symmetric and in the case of the matrix of correlation coefficients the diagonal elements are all unity (1's) which simplifies the determinant considerably

$$|R| = \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{vmatrix}$$

the characteristic equation now becomes

$$\begin{vmatrix} 1-\lambda & r_{12} & r_{13} \\ r_{12} & 1-\lambda & r_{23} \\ r_{13} & r_{23} & 1-\lambda \end{vmatrix} = 0$$

or $$(1 - \lambda)^3 - (1 - \lambda)\ r_{23}^2 - r_{12}^2\ (1 - \lambda) + 2\ (r_{12})\ (r_{23})\ (r_{13}) - r_{13}^2\ (1 - \lambda) = 0$$

(11) Several methods are available for solving for the various nonzero roots or eigenvalues (λ's) of a polynomial of the third order. Consider the polynomial $\lambda^3 - a\lambda^2 + b\lambda - c$.

a, b, and c are real coefficients. The 3 roots (λ) for which the polynomial equals 0, are desired. First the upper and lower limits of this polynomial are determined.

$$f(\lambda) = \lambda^3 - a\lambda^2 + b\lambda - c = 0$$

$$f^1(\lambda) = 3\lambda^2 - 2a\lambda + b = \lambda^2 - {}^2/_3 a\lambda + {}^b/_3 = 0$$

$$\lambda = \frac{2a \pm \sqrt{(-2a)^2 - 4 \cdot 3 \cdot b}}{2a} = 1 \pm \sqrt{\frac{4a^2 - 12b}{4a^2}} = 1 \pm \sqrt{1 - \frac{3b}{a^2}}$$

thus the equation reaches a maximum value for

$$\lambda_1 = 1 + \sqrt{1 - \frac{3b}{a^2}}$$

and $$\lambda_2 = 1 - \sqrt{1 - \frac{3b}{a^2}}$$

thus λ_1 for which $\lambda^3 - a\lambda^2 + b\lambda - c$ becomes 0, lays between $-\infty$

and $1 - \sqrt{1 - \frac{3b}{a^2}}$

λ_2 lays between $1 - \sqrt{1 - \frac{3b}{a^2}}$ and $1 + \sqrt{1 - \frac{3b}{a^2}}$

and λ_3 between $1 + \sqrt{1 - \frac{3b}{a^2}}$ and $+\infty$.

Now trial values for λ can be inserted in the equation and the roots found by interpolation.

(12) For each eigenvalue found, there will be an associated vector, called the eigenvector, V which satisfies the matrix equation $A \cdot V = \lambda V$. If all the eigenvalues of the matrix A are placed in the diagonal elements of a matrix $\wedge$ in which the off diagonal elements all are zero, and the corresponding set of eigenvectors of the matrix A are placed as columns in a matrix X the following matrix equation holds true $A \cdot X = X \cdot \wedge$

(13) The trace of a matrix is the sum of all the diagonal elements. The trace of matrix A is identical to the trace of the matrix $\wedge$. In other words the sum of the eigenvalues is identical to the trace of the original matrix A.

As mentioned before the basic task of principal component analysis firstly is to find an axis (component, factor) in the original test space along which the variance is maximum, then successively a number of axes along which the remaining variance each time is maximum. This task can be accomplished by solving for a set of equations which are the result of partial differentiation of the function to be maximized, subject to a restriction due to the use of Lagrange multipliers.

This set of equations may be written as follows:

$$\begin{aligned}
&V_1(1-\lambda) + V_2 r_{12} + \ldots\ldots V_n r_{1n} = 0\\
&V_1 r_{21} + V_2(1-\lambda) + \ldots\ldots V_2 r_{2n} = 0\\
&\vdots\\
&V_1 r_{n1} + V_2 r_n + \ldots\ldots V_n(1-\lambda) = 0
\end{aligned}$$

In matrix form these equations can be written as follows:

$$\begin{vmatrix} (1-\lambda) & r_{12} & \cdots & r_{1n} \\ r_{21} & (1-\lambda) & \cdots & r_{2n} \\ \vdots & \vdots & & \\ r_{n1} & r_{n2} & \cdots & (1-\lambda) \end{vmatrix} \cdot \begin{vmatrix} V_{1i} \\ V_{2i} \\ \vdots \\ V_{ni} \end{vmatrix} = \begin{vmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{vmatrix}$$

or in matrix notation $(R - \lambda_i I) \cdot V_i = 0$ $(i = 1, 2 \ldots\ldots n)$ in which R is the matrix of the correlation coefficients. There are n nontrivial solutions in which each V_i represents the coefficients (transformation vector) for converting the original data to one of the uncorrelated scores of the new components (factors). For the nontrivial solutions the determinant of the coefficient of V is zero. In matrix notation this can be written as follows: $R - \lambda I = 0$. This is the characteristic equation for which there are n possible solutions for λ (if R is a correlation matrix with n column and n rows).

The vector V_i can be calculated for any corresponding eigenvalue by substituting the eigenvalue in the set of equations mentioned before and solving for V_i. When the vectors are normalized, the variance of each set of component scores is λ_i. The eigenvector V_1 produces the principal component scores with the maximum variance, this variance is the value of the largest eigenvalue. The sum of all the eigenvalues is equal to the sum of the diagonal elements of R matrix, which is called the trace. Because the trace of the R matrix is the total variance to be accounted for by the principal components (factors) the sum of the eigenvalues, associated with the principal components retained, divided by the trace of the matrix R is the proportion of the variance accounted for. Numerical example of the principal component analysis of the matrix of correlation coefficients (hypothetical).

Let the matrix of correlation coefficients be

$$R = \begin{pmatrix} 1 & 0.10 & 0.20 \\ 0.10 & 1 & 0.90 \\ 0.20 & 0.90 & 1 \end{pmatrix}$$

The task is then to find the eigenvalues and the associated eigenvectors. The eigenvalues are the roots of the characteristic equation which may be written in the form of a determinant as follows.

$$\begin{vmatrix} 1-\lambda & 0.10 & 0.20 \\ 0.10 & 1-\lambda & 0.90 \\ 0.20 & 0.90 & 1-\lambda \end{vmatrix} = 0$$

or in expanded form (a polynomial of the third order)

$$(1-\lambda)^3 - (1-\lambda)\,(0.90^2) - (0.10^2)\,(1-\lambda) + (0.10)\,(0.90)\,(0.20) + (0.20)\,(0.10)\,(0.90) - (0.20)\,(1-\lambda)\,(0.20) = 0$$

$$1 - 3\lambda + 3\lambda^2 - \lambda^3 - 0.81\,(1-\lambda) - 0.01\,(1-\lambda) + 0.018 + 0.018 - 0.04\,(1-\lambda) = -\lambda^3 + 3\lambda^2 - 2.14\lambda + .176 = \lambda^3 - 3\lambda^2 + 2.14\lambda - .176 = 0$$

We now determine the upper and lower limit of the roots

$$f(\lambda) = \lambda^3 - 3\lambda^2 + 2.14\lambda - .176$$

$$f'(\lambda) = 3\lambda^2 - 6\lambda + 2.14 = \lambda^2 - 2\lambda + .7133$$

$$\lambda_{12} = 1 \pm \sqrt{1 - .7133} = 1 \pm .535$$

$$\lambda_1 = .465$$

$$\lambda_2 = 1.535$$

Inserting these values in the characteristic equation we find that $\lambda^3 - 3\lambda^2 + 2.14\lambda - .176$ reaches a positive maximum for $\lambda = .465$ and a negative maximum for $\lambda' = 1.535$

Also for $\lambda = - \ldots, \lambda^3 - 3\lambda^2 + 2.14\lambda - .176$ is negative
and for $\lambda = + \ldots, \lambda^3 - 3\lambda^2 + 2.14\lambda - .176$ is positive

By inserting trial values and interpolating we now find:

$$\lambda^1 = 1.948$$
$$\lambda^2 = .958$$
$$\lambda_3 = .0943$$

The trace of the R-matrix is 3 and the sum of the calculated eigenvalues is 3.0003. The accuracy of the computations is thus very satisfactory. The first principal component accounts for $1.948/3 \times 100\% = 64.93\%$; the second component for $0.958/3 \times 100\% = 31.93\%$ and the third component for $.0943/3 \times 100\% = 3.14\%$ of the total variance. The first and second components together thus account for 96.86% of the total variance. The third component has such a low eigenvalue that for practical purpose it does not need to be considered.

To find the eigenvectors, corresponding to these eigenvalues, the eigenvalues are substituted in the set of equations.

$$V_1 (1 - \lambda) + V_2 r_{12} + V_3 r_{13} = 0$$
$$V_1 r_{12} + V_2 (1 - \lambda) + V_3 r_{23} = 0$$
$$V_1 r_{13} + V_2 r_{23} + V_3 (1 - \lambda) = 0$$

If λ_1 is substituted, this results in the following:

$$V_1 (1 - 1.948) + V_2 (0.1) + V_3 (0.2) = 9$$
$$V_1 (0.1) + V_2 (1 - 1.948) + V_3 (0.3) = 0$$
$$V_1 (0.2) + V_2 (0.3) + V_3 (1 - 1.948) = 0$$

These equations can be solved for V_1, V_2, and V_3 with the use of determinants.

$$\frac{V_1}{\begin{vmatrix} 0.1 & 0.2 \\ (1 - 1.948) & 0.3 \end{vmatrix}} = \frac{V_2}{\begin{vmatrix} 0.2 & (1 - 1.948) \\ 0.3 & 0.1 \end{vmatrix}} = \frac{V_3}{\begin{vmatrix} (1 - 1.948) & 0.1 \\ 0.1 & (1 - 1.948) \end{vmatrix}}$$

Thus $\lambda_1 = 1.948$ has the following eigenvectors

$$V_1 = \{(0.1)\ (0.3) - (-0.948)\ (0.2)\}\ \varkappa = .1926\ \varkappa$$
For $\varkappa = 1$, $V_1 = 0.1926$ and
$$V_2 = \{(0.1)\ (0.2) - (0.3)\ (-0.948)\}\ \varkappa = .3044\ \varkappa$$
For $\varkappa = 1$ $V_2 = 0.3044$
$$V_3 = \{(-0.948)\ (-0.948) - (0.1)\ (0.1)\}\ \varkappa = 0.889\ \varkappa$$
For $\varkappa = 1$ $V_3 = 0.889$

If λ_2 is substituted the corresponding eigenvectors are:

$$\frac{V_1}{\begin{vmatrix} 0.1 & 0.2 \\ 0.042 & 0.3 \end{vmatrix}} = \frac{V_2}{\begin{vmatrix} 0.2 & 0.042 \\ 0.3 & 0.1 \end{vmatrix}} = \frac{V_3}{\begin{vmatrix} 0.042 & 0.1 \\ 0.1 & 0.042 \end{vmatrix}}$$

$$V_1 = (0.03 - 0.0084)\ \varkappa = 0.0216\ \varkappa$$
For $\varkappa = 10$, $V_1 = 0.216$
$$V_2 = (0.02 - 0.0126)\ \varkappa = 0.0074\ \varkappa$$
For $\varkappa = 10$, $V_2 = 0.074$
$$V_3 = (0.001764 - 0.01)\ \varkappa = -0.0082\ \varkappa$$
For $\varkappa = 10$, $V_3 = -.0082$

With this the actual principal component analysis is complete.

It is easy to see that, with more than three variables, it is difficult to perform the analysis with a desk calculator and it is very time consuming. With the help of an electronic computer of sufficient capacity, however, the complete analysis takes only minutes. The time required for an analysis depends on the size of the computer used, the number of variables, and the accuracy required. In many cases, large institutions make computertime available without cost for research purposes.