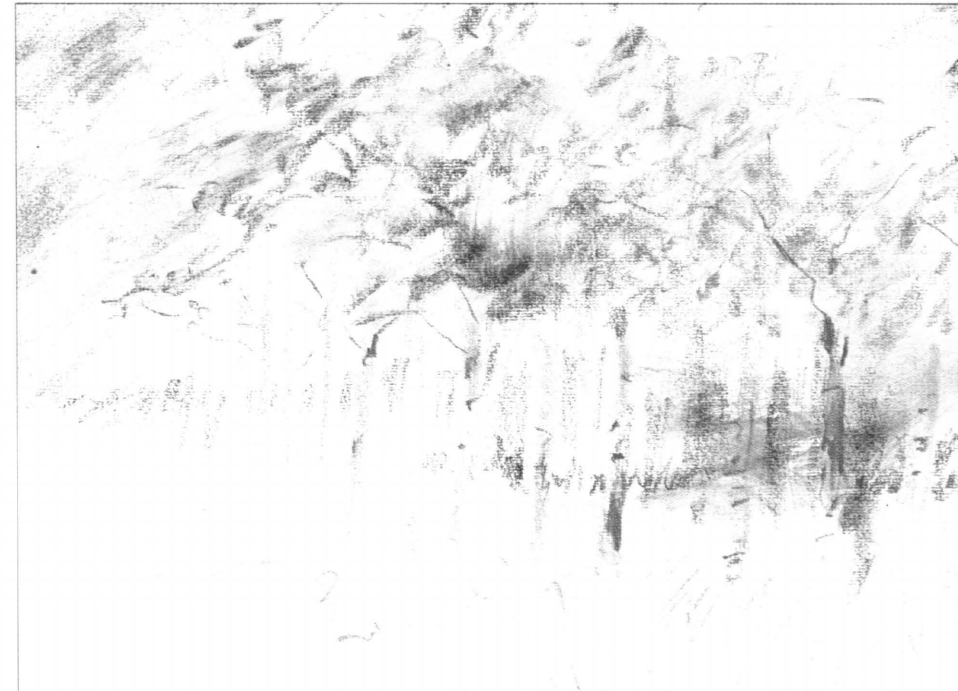


ACTA FORESTALIA FENNICA



Tapani Lahti

Understorey Vegetation as an Indicator of
Forest Site Potential in Southern Finland

246 • 1995

Publishers The Finnish Society of Forest Science
The Finnish Forest Research Institute

Editors Editor-in-chief Eeva Korpilahti
Production editors Tommi Salonen, Seppo Oja

Editorial Office Unioninkatu 40 A, FIN-00170 Helsinki, Finland
Phone +358 0 857 051, Fax +358 0 625 308, E-mail silva.fennica@metla.fi

Managing Board Erkki Annala (The Finnish Forest Research Institute), Jari Kuuluvainen (The Finnish Forest Research Institute), Esko Mikkonen (University of Helsinki), Lauri Valsta (The Finnish Forest Research Institute), Harri Vasander (University of Helsinki), and Seppo Vehkamäki (University of Helsinki).

Editorial Board Per Angelstam (Grimsö Wildlife Research Station, Sweden)
Julius Boutelje (Swedish University of Agricultural Sciences, Sweden)
Finn H. Brække (Norwegian Forest Research Institute, Norway)
J. Douglas Brodie (Oregon State University, USA)
Raymond L. Czaplewski (USDA Forest Service, USA)
David Ford (Lancaster University, United Kingdom)
George Gertner (University of Illinois, USA)
Martin Hubbes (University of Toronto, Canada)
William F. Hyde (Virginia Polytechnic Institute and State University, USA)
Jochen Kleinschmit (Lower Saxony Forest Research Institute, Germany)
Michael Köhl (Swiss Federal Institute for Forest, Snow and Landscape Research, Switzerland)
Noel Lust (University of Gent, Belgium)
Bo Långström (Swedish University of Agricultural Sciences, Sweden)
William J. Mattson (USDA Forest Service, USA)
Robert Mendelsohn (Yale University, USA)
Hugh G. Miller (University of Aberdeen, United Kingdom)
John Pastor (University of Minnesota, USA)
John Sessions (Oregon State University, USA)
Jadwiga Sienkiewicz (Environment Protection Institute, Poland)
Richard Stephan (Federal Research Centre for Forestry and Forest Products, Germany)
Elon S. Verry (USDA Forest Service, USA)
S.E. Vompersky (Russian Academy of Sciences, Russia)
A.G.D. Whyte (University of Canterbury, New Zealand)
Claire G. Williams (North Carolina State University, USA)

Aim and Scope Acta Forestalia Fennica publishes monographs in forest science. These may be original research articles or comprehensive reviews aiming at a synthesis of knowledge in a particular topic. The series covers all aspects of forest science, both basic and applied subjects. Manuscripts are subject to peer review.

ACTA FORESTALIA FENNICA

246 · 1995

To be presented, with the permission of the Faculty of Agriculture and Forestry of the University of Helsinki, for public criticism in the Auditorium of the Forestry Field Station of Hyytiälä, Korkeakoski, on 1 July 1995, at 12 o'clock noon

Tapani Lahti

Understorey Vegetation as an Indicator of Forest Site Potential in Southern Finland

Contents

Lahti, T. 1995. Understorey vegetation as an indicator of forest site potential in southern Finland. *Acta Forestalia Fennica* 246. 68 p.

The relationship between site characteristics and understorey vegetation composition was analysed with quantitative methods, especially from the viewpoint of site quality estimation. Theoretical models were applied to an empirical data set collected from the upland forests of southern Finland comprising 104 sites dominated by Scots pine (*Pinus sylvestris* L.), and 165 sites dominated by Norway spruce (*Picea abies* (L.) Karsten). Site index H_{100} was used as an independent measure of site quality.

A new model for the estimation of site quality at sites with a known understorey vegetation composition was introduced. It is based on the application of Bayes' theorem to the density function of site quality within the study area combined with the species-specific presence-absence response curves. The resulting posterior probability density function may be used for calculating an estimate for the site variable.

Using this method, a jackknife estimate of site index H_{100} was calculated separately for pine- and spruce-dominated sites. The results indicated that the cross-validation root mean squared error ($RMSE_{cv}$) of the estimates improved from 2.98 m down to 2.34 m relative to the "null" model (standard deviation of the sample distribution) in pine-dominated forests. In spruce-dominated forests $RMSE_{cv}$ decreased from 3.94 m down to 3.16 m.

In order to assess these results, four other estimation methods based on understorey vegetation composition were applied to the same data set. The results showed that none of the methods was clearly superior to the others. In pine-dominated forests, $RMSE_{cv}$ varied between 2.34 and 2.47 m, and the corresponding range for spruce-dominated forests was from 3.13 to 3.57 m.

Keywords classification, estimation, forest vegetation, Norway spruce, Scots pine, site quality

Author's address Muroleenkatu 14 C 37, FIN-33720 Tampere, Finland

Accepted April 19, 1995

ISBN 951-40-1435-9

ISSN 0001-5636

Tampere 1995, Tammer-Paino Oy

1	INTRODUCTION	6
1.1	Goals and methods of site characterization	6
1.2	Understorey vegetation and site characteristics	6
1.3	Classification vs. estimation	7
1.4	Purpose of the present study	7
2	MATERIAL AND METHODS	9
2.1	Field material	9
2.2	Methods of data analysis	10
2.2.1	Site index H_{100}	10
2.2.2	Estimation of probability density functions	10
2.2.3	Estimation of presence-absence response curves	12
2.2.4	Estimation of site quality from vegetation data	14
2.2.5	Evaluation and comparison of the estimation methods	15
3	ANALYSIS OF SPECIES RESPONSES TO SITE CHARACTERISTICS	16
3.1	The concept of "site"	16
3.2	What is an indicator species?	18
3.2.1	Analysis of incidence data	18
3.2.2	Analysis of abundance data	20
3.2.3	Empirical results	21
3.3	Estimation of presence-absence response curves	23
3.3.1	Theoretical background	23
3.3.2	Empirical results	24
3.3.3	The effect of quadrat size	24
4	ESTIMATION OF SITE CHARACTERISTICS FROM UNDERSTOREY VEGETATION COMPOSITION	26
4.1	Theoretical background	26
4.2	Empirical results	27
5	ALTERNATIVE ESTIMATION METHODS	29
5.1	Overview	29
5.2	Forest site types	29
5.2.1	Theoretical background	29
5.2.2	Empirical results	32
5.3	Hierarchical clustering	34
5.3.1	Theoretical background	34
5.3.2	Empirical results	37
5.4	Multivariate ordination and regression	37
5.4.1	Theoretical background	37
5.4.2	Empirical results	40
5.5	Weighted averaging	42
5.5.1	Theoretical background	42
5.5.2	Empirical results	45
5.6	Comparison of the methods	46

6	DISCUSSION	48
6.1	Evaluation of the results	48
6.2	Classification vs. estimation	49
6.3	General principles for the characterization of forest sites	50
	REFERENCES	52
	APPENDICES	56

List of Symbols

Symbol	Explanation		
α	Critical value for statistical significance	$k_i(s)$	Status of species i at site s
λ_j	Local bandwidth factor for adaptive kernel estimates	$k_{ia}(s)$	Status of pseudospecies i with limit abundance a at site s
π_i	Differentiation index (weighting parameter) for species i	$K(s)$	Vector of site-specific additional information
σ	Standard deviation	$p(a b)$	Conditional probability of a , given b
A	Study area	$p(x)$	Density function of variable x
$a_i(s)$	Abundance of species i at site s	$P(x)$	Distribution function of variable x
c	Number of comparisons in repeated tests	q_i	Indicator value of species i
CER	Comparisonwise error rate	$q(s)$	Site variable
D_i	Kolmogorov statistic for species i	R	Interquartile range
DI_i	Differentiation index for species i	r_i	Number of the occurrences of species i
$E(x)$	Mean of x	RMSE	Root mean squared error
EER	Experimentwise error rate	RMSE _{cv}	Cross-validation estimate of RMSE
g	Geometric mean	s	Site
$G(x)$	Kernel function	t_i	Tolerance of species i
H_{100}	Site index	$\text{var}(x)$	Variance of x
		w	Window width (smoothing parameter) of the kernel function

Acknowledgements

I would like to express my sincere thanks to Emeritus Professor Matti Leikola, who gave me the original inspiration and continuous support during the long gestation period of the present study. I have had a great opportunity to enjoy his excellent expertise on both the scientific and cultural aspects of forest site classification in Finland.

Discussions with Professor Pertti Hari, Mr. Juha-Pekka Hotanen, Lic.Phil., Professor Seppo Kellomäki, Dr. Juha Lappi, Mr. Henrik Lindberg, M.Sc., Dr. Heikki Smolander, Dr. Pauline Stenberg and Ms. Tiina Tonteri, M.Sc. have also been essential for clarifying my thoughts on this topic. Especially Juha Lappi and Pauline Stenberg gave invaluable help in my attempts to find a suitable mathematical and statistical approach to the problems encountered.

I want to thank Dr. Heikki Smolander and the staff of the Suonenjoki Research Station of the Finnish Forest Research Institute for providing

me with a peaceful environment during the preparation of the first version of the manuscript. I am also very grateful to Dr. Jussi Kuusipalo and Dr. Pekka Tamminen, who gave me a permission to use their field data for testing the estimation methods. Technical help from Dr. Helena Henttonen is also gratefully acknowledged.

For valuable comments on the various versions of the manuscript, I want to thank Professor John Birks, Dr. Mark Hill, Dr. Juha Lappi, Professor Matti Leikola, Dr. Jari Oksanen, Dr. Heikki Smolander, Dr. Pauline Stenberg, Ms. Tiina Tonteri, Dr. Harri Vasander and two anonymous referees. I am also grateful to Mark Hill for suggesting some improvements to the English of the final draft.

Finally, I want to express my sincere gratitude to my wife Irene, for love and patience.

Part of this study was financed by the Foundation for Research of Natural Resources in Finland.

1 Introduction

1.1 Goals and Methods of Site Characterization

One of the principal goals of forest site classification has been to develop a reliable procedure for identifying and describing those characteristics of the site that are important to the production of harvestable biomass. In Finland, this requirement originated from the agricultural sector, in the need to recognize forest stands suitable for slash-and-burn cultivation (see Heikinheimo 1915), and resulted in the practical classification system of forest lands already in the 18th century (Lehto and Leikola 1987). Classification of forest sites for purely silvicultural purposes was the goal of Blomqvist (1872) in his growth and yield tables for pine, spruce and birch. It was succeeded by the introduction of the forest site type concept by Cajander (1909).

In the present study, the concept of "forest site quality" is used to mean "the timber production potential of a site for a particular species or forest type" (Clutter et al. 1983). In the evaluation of site quality, a number of different methods can be applied. Usually these methods are based on various stand parameters, composition of understorey vegetation, or soil characteristics. Since there is no absolute measure of site quality, the comparison of alternative indirect estimation methods can only be done with respect to each other. However, tree stand parameters may be considered more "direct" in the sense that they refer to the harvestable timber volume available at the site.

In addition to site quality estimation, there are also a number of other goals for the operational typification of different forest sites. Forest management practices are selected according to the expected responses of the tree stand to the treatment, which may be very characteristic for certain types of forest sites. Biological research, on the other hand, attempts to relate the distribution and abundance of organisms to measurable site variables. Therefore, characterization of forest

sites is always goal-oriented. It is unlikely that there will be a single universal classification scheme that could satisfactorily serve all the diverse needs of forest research.

1.2 Understorey Vegetation and Site Characteristics

Understorey vegetation has played an important role in the site classification systems of Finnish forests (Mikola 1982, Oksanen 1990). Paucity of tree species in the boreal forests of Northern Europe implies that site classification on the basis of the floristic composition of arboreal vegetation alone is crude. The number of plant species occurring in the understorey vegetation of Finnish forests is not very high either but does offer a more suitable basis for site characterization.

There are two principal approaches in the utilization of understorey vegetation as an indicator of site characteristics. One of these is the analysis of vegetation composition as a whole ("community") by constructing a species-by-sites matrix where each matrix cell [i,s] indicates the "importance" value (e.g. biomass) of species *i* at site *s*. Patterns in the species-by-sites matrix are then analysed, usually with multivariate ordination methods. The relationship between observed pattern in the vegetation and various site factors can also be studied with a number of methods (e.g. Gauch 1982, Jongman et al. 1987).

An alternative approach is to draw attention to a number of indicator species that are assumed to be of primary importance for revealing interesting site characteristics. The procedure for studying species-environment interactions comprises two separate steps. First, the responses of species to environmental gradients are modelled using statistical techniques (e.g. ter Braak and Looman 1986, Austin 1987). This phase of the analysis is called "regression" (ter Braak and Prentice 1988).

There are two frequently used concepts for describing the functional response of indicator species to ecological factors. "Optimum" refers to that portion of the ecological gradient where the species reaches its maximum importance, whereas "tolerance" indicates the steepness of species decline when moving away from the optimal region along the gradient (ter Braak and Looman 1986). For several indicator species occurring at a given site, some kind of averaging of their indicator values is used for estimating the site-specific parameter value. This process is known as "calibration" (ter Braak and Prentice 1988).

1.3 Classification vs. Estimation

When studying the relationship between site characteristics and the composition of understorey vegetation within a formal statistical framework, it becomes necessary to draw attention to the nature of variation in the variable of interest. Traditionally, classification of forest sites into a practical number of site types has been performed by grouping together all those sites that are similar enough to each other. More recently, a number of statistical classification systems and clustering algorithms have been applied in order to find a more solid theoretical basis for this procedure (e.g. Gordon 1981, Birks 1987, Hill 1989).

However, there are both theoretical and empirical reasons to assume that variation in site characteristics from one site to another is at least in some (maybe most) cases continuous without distinct classes (for a review of historical ideas and opinions, see Crawley 1986). This hypothesis implies that any classification scheme, irrespective of the number of site classes used, is based on more or less arbitrary criteria for determining class boundaries in relation to site quality. The Cajanderian forest site type system that is widely used in Finland is a typical classification scheme. In Cajander's system, understorey vegetation is used as an indicator of the biological value of the site, in pursuit of determining the "natural quality classes of the sites" (Cajander 1926).

So far there have not been very many alternative methods for forest site characterization on the basis of understorey vegetation in Finland. Kuusipalo's (1985) classification scheme is based

on a hierarchical approach that enables a more objective and finer tuning of the class boundaries and the number of site type classes than Cajander's system. The flexibility of the hierarchical classification scheme remained largely unused in his attempt to find correspondence between the established Cajanderian forest site types and the new system.

Recent studies by Nieppola (1992, 1993a, 1993b, Nieppola and Carleton 1991) have also addressed the problem of describing the relationship between site characteristics (especially site quality) and understorey vegetation composition in the pine-dominated forests of southern Finland. Although largely based on statistical methods that utilize quantitative data (such as multivariate regression), his approach seems to follow the classification paradigm. This is somewhat surprising, especially when Nieppola and Carleton (1991) explicitly mention that "the notion of discrete types is not supported [by the data]".

1.4 Purpose of the Present Study

The purpose of the present study is to apply modern quantitative methods to the analysis of Finnish forest vegetation, especially from the viewpoint of elucidating the indicator value of understorey vegetation. The structure of the study follows the distinction between "regression" and "calibration" sensu ter Braak and Prentice (1988). I shall first describe and evaluate methods for the analysis of species responses to ecological gradients. The concept of "indicator species", as well as quantitative measures of the "indicative power" of individual species, are analysed.

In the second ("calibration") part of the study I shall define and apply a probabilistic estimation method that utilizes the quantitative description of species responses to environmental gradients obtained in the "regression" phase. The theoretical model, referred to as "posterior probability estimation" (PPE), is then applied to an empirical data set collected from the upland forests of southern Finland.

In the third part of the study I shall compare this method with four alternatives that can be applied for the estimation of site characteristics. Of these, Cajander's (1909) forest site type system (FST)

and Kuusipalo's (1985) hierarchical clustering method (HC) are based on the classification scheme. For them, class averages are used as estimates of site parameters among the members of the class. Third, an estimation method based on multivariate ordination and subsequent regression (MO) offers a quantitative, "community"-type analysis of covariation between vegetation composition and site characteristics. The fourth method, known as weighted averaging (WA), is a simple method for obtaining estimates of site variables from a list of species with previously determined indicator values (e.g. ter Braak and Looman 1986, Birks et al. 1990). Nieppola (1993a) has recently used this method (referred to as "Ellenberg's calibration method") for the analysis of Finnish forest vegetation.

The variable used for comparing these methods is "site quality" described by an independent measure of site index based on the height-to-age

relationship of dominant trees at each sample site (referred to as H_{100} , with an index age of 100 years). Site index itself has a number of theoretical and methodological problems, but for the purposes of the present study it offers a satisfactory reference variable. However, due to these problems the empirical results shown in the present study should be interpreted with caution.

Throughout the text, the concept of "site variable" is used as a more comprehensive and general term than "site quality". Theoretically, any of the methods discussed in the present study could be used for the estimation of several different site variables by utilizing the empirical regression-calibration scheme with a suitable data set (known as a "training set"). Site quality is only one of the site variables that in this study is used for the comparison of different estimation methods with an empirical data set.

2 Material and Methods

2.1 Field Material

The empirical field data used in the present study were collected by Dr. Jussi Kuusipalo and Dr. Pekka Tamminen from the upland forests of southern Finland in 1982–84. Of the 410 sample plots in the original data set, only the 269 ones with tree stand data available were used here. Of these, 104 sample plots were dominated by Scots pine (*Pinus sylvestris* L.) and 165 plots by Norway spruce (*Picea abies* (L.) Karsten). All the sites were located within the southern boreal

vegetation zone (Ahti et al. 1968; see Fig. 1). This is the same data set as used by Kuusipalo (1985), who has also given a more detailed description of the sampling methods used in the field.

Only the sample plots located on mineral soil having a pine- or spruce-dominated stand at least 20 years of age were included in the data. Additionally, sites classified as low-yielding were omitted (for a definition, see Kuusela and Salminen 1969, p. 16). Age distributions of the data sets for pine- and spruce-dominated forests are given in Fig. 2. Basic statistics for the age distribution of the stands are given in Table 1.

Within each sample plot of $16 \times 16 \text{ m}^2$, six vegetation sample quadrats, each $2 \times 2 \text{ m}^2$ in size, were located subjectively in order to represent the

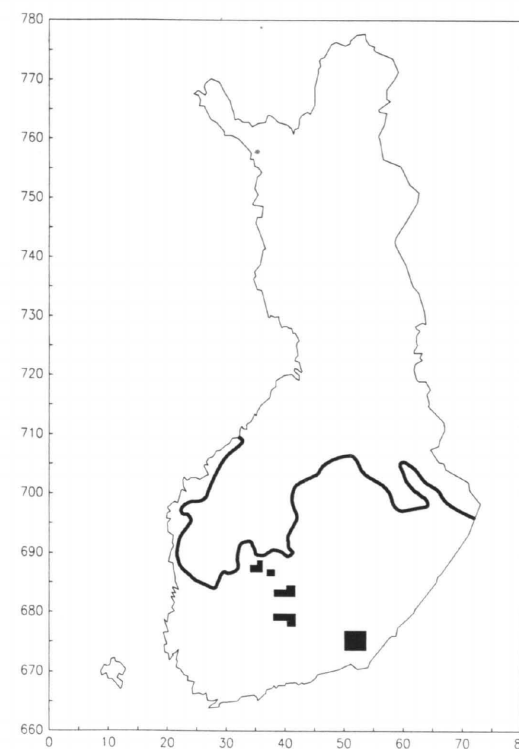


Fig. 1. Geographical locations of the study sites in southern Finland. Thick line shows the northern border of the southern boreal vegetation zone of Ahti et al. (1968).

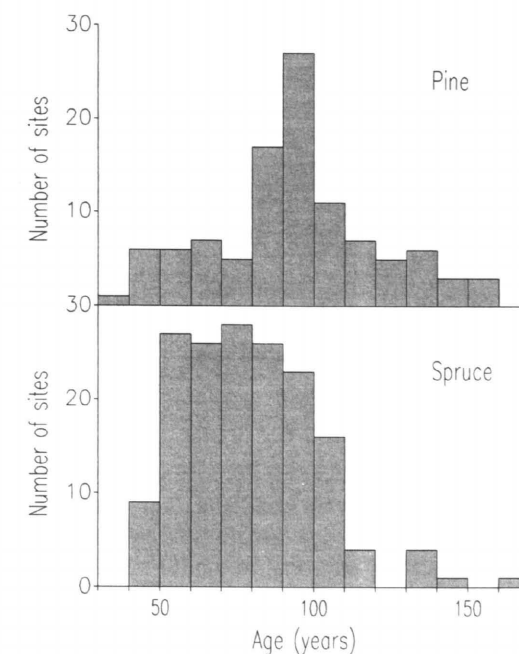


Fig. 2. Age distributions of the study sites dominated by Scots pine and Norway spruce in the data set.

Table 1. Characteristics of the age and site index (H_{100}) distribution of the data set.

Tree species	n	Mean	Age (yrs)			Mean	H_{100} (m)		
			s.d.	Min	Max		s.d.	Min	Max
Pine	104	93.0	26.9	34	157	22.0	2.96	12.7	27.9
Spruce	165	78.7	21.8	40	161	26.9	3.93	14.8	34.7

compositional variation of vegetation within the sample plot. The abundance of all species of vascular plants, mosses and lichens were recorded by estimating their percentage cover within the quadrat. The vegetation occurring on larger stones, stumps, partially decayed logs, etc. was disregarded. In the present study, the vegetation descriptions of the six quadrats within each sample plot have been combined to obtain a collective vegetation description of the sample plot. For compatibility, species nomenclature in the present study is the same as in Kuusipalo (1985).

Each site was assigned into one of the established Cajanderian forest site types already in the field. This site type classification was used in the present study to represent the FST model in the comparison of different estimation methods.

Tree stand measurements were made by a field group of Finnish Forest Research Institute following standard procedures (Kuusela and Salminen 1969). Dominant height of the stand, referring to the average height of the 100 thickest trees per hectare, was calculated on the basis of 3–10 stems per sample plot. Dominant age of the stand was computed as the average of the ages of the dominant trees.

2.2 Methods of data analysis

2.2.1 Site index H_{100}

Site index H_{100} , referring to the dominant height (in meters) of the tree stand at a reference age of 100 years, was used in this study as an independent measure of site quality whose values were estimated by a number of methods based on the composition of understorey vegetation. Empirical results have indicated that there is a rather strong correlation between H_{100} and a number of stand parameters referring to the productivity of

the site (Vuokila and Väliaho 1980, Oikarinen 1983). These observations have made it possible to use H_{100} as an alternative to the Cajanderian forest site type system in the estimation of site quality for forest management.

However, there are also some methodological problems in the application of site index as an estimate of site quality. For example, Kilkki and Ojansuu (1981) have drawn attention to the fact that the stem form of trees varies in different parts of Finland, and therefore growth curves may be significantly biased when applied outside the geographical regions where the original stand data were collected. In the data set used in this study, these problems are alleviated by the fact that all the data have been collected from the same (southern boreal) vegetation zone (see Fig. 1), where variations in the developmental patterns of the stand are presumably small.

From the viewpoint of the present study, methodological problems associated with the application of site index as a measure of true site quality are not very serious in the sense that all estimation methods were applied to the same data set. Therefore, they all respond to identical biases in the data. The principal criterion of performance in the present study is predictive accuracy relative to the other methods that can thus be reliably assessed.

Scatter plots of the dominant height vs. age values of the sample plots are shown in Fig. 3. Using the height-over-age curves of Gustavsen (1981), site index H_{100} was calculated from these values (Fig. 4 and Table 1).

2.2.2 Estimation of Probability Density Functions

For the estimation of a density function, $p(q)$, of site quality q , I utilized kernel estimators (Sil-

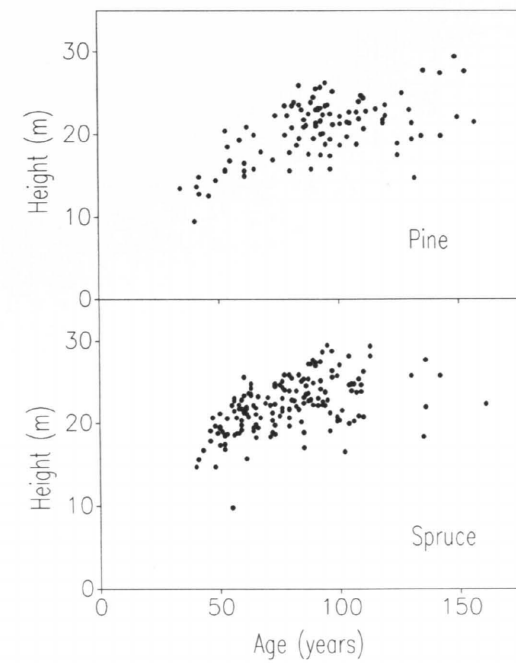


Fig. 3. Dominant height vs. stand age in the study sites.

verman 1986, Izenman 1991) defined as follows. Let G be a symmetric, non-negative kernel function, which satisfies the condition

$$\int_{-\infty}^{\infty} G(x) dx = 1 \quad (1)$$

With a sample of n observations X_j from the unknown distribution $f(x)$, the kernel estimator $\hat{f}(x)$ with kernel G is defined by

$$\hat{f}(x) = \frac{1}{nw} \sum_{j=1}^n G\left[\frac{x - X_j}{w}\right] \quad (2)$$

where w is the window width (smoothing parameter) of the kernel function G .

Silverman (1986) examined the properties of most commonly used kernel functions and concluded that, in terms of efficiency (measured as the mean integrated square error (MISE) of $\hat{f}(x)$ relative to $f(x)$), all frequently used functions are close to each other and the choice of the kernel function G can be based on other criteria, such as the degree of differentiability required. In the present study, I have used the triangular kernel defined as

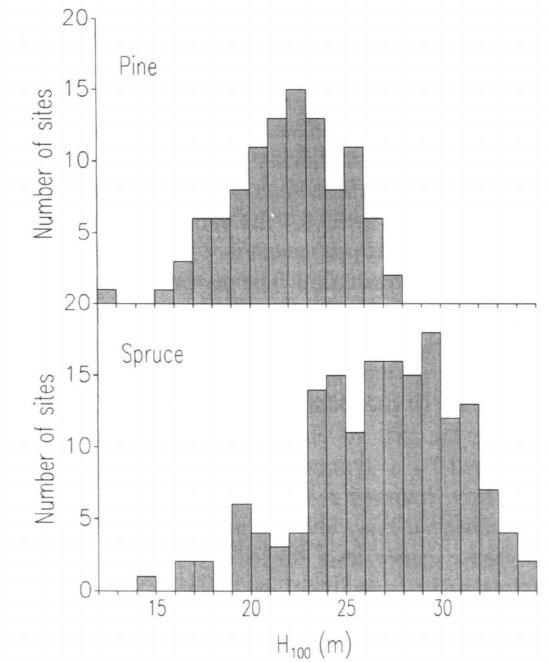


Fig. 4. Distribution of site index H_{100} in the empirical data set.

$$G(x) = 1 - |x| \text{ for } |x| < 1, = 0 \text{ otherwise} \quad (3)$$

Considerably more important for density estimation is the choice of window width w , which is a compromise between the risks of oversmoothing and undersmoothing. Silverman (1986) and Izenman (1991) discussed the various principles that can be applied to the choice of the value of w . In the case of a Gaussian kernel function G , Silverman (1986) concluded that, in terms of MISE, a good choice of w is

$$w = 0.9 \min[\sigma, R/1.34] n^{-1/5} \quad (4)$$

where n is the number of observations, σ is the standard deviation, and R is the interquartile range of the sample. Taylor (1989) and Hall (1990) showed that a bootstrap procedure can also be used for selecting the value of w in large samples.

2.2.3 Estimation of Presence-Absence Response Curves

Kernel estimators defined in the previous chapter were used for estimating the density functions of site quality. With slight modifications they were also used for the estimation of presence-absence response curves of each species of understorey vegetation. The starting point was a data set where the quantitative value of the target variable H_{100} was shown together with the presence-absence data of each (pseudo)species (for a definition of pseudospecies, see Chapter 3.2.2.). Along the axis of H_{100} , each sample plot was replaced by a symmetric, non-negative kernel function, centered at the value of the H_{100} at the site (Fig. 5). The value of the presence-absence response curve as a function of H_{100} was obtained as a proportion of the "weights" (from the kernel function) of sites with the species present divided by the sum of the "weights" of all sites.

Estimation of the presence-absence response curve is also sensitive to the overall frequency of occurrence of each species. A general rule of thumb is that if the species is very common or very rare, its presence-absence response curve is difficult to estimate satisfactorily. Apart from some exceptions (e.g. Fryer 1976), research emphasis in nonparametric density estimators has

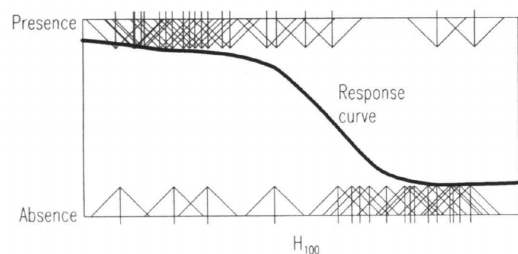


Fig. 5. A schematic view of the estimation of species-specific presence-absence response curves. Depending on the presence or absence of the species at each site, a symmetric kernel function, centered at the value of H_{100} at the site, was located along the site quality axis. The presence-absence response curve was then calculated as a mean of the "weights" of presences and absences for each value of H_{100} . See text for details.

been in developing large-sample properties (Izenman 1991). Therefore, the methods of nonparametric density estimation should be used with caution in such cases.

Some of the most difficult problems in nonparametric smoothing of the presence-absence curves occur at the tails of the response, where the number of observations is very low. In order to obtain smoother behaviour at these regions, adaptive kernel estimators can be applied. Adaptive kernel estimators are an extension of the fixed kernel method defined as follows (see Silverman 1986). First, a pilot estimate $\tilde{f}(x)$ that satisfies $\tilde{f}(X_j) > 0$ is calculated for all j . Then, local bandwidth factors λ_j are calculated as

$$\lambda_j = \left[\frac{\tilde{f}(X_j)}{g} \right]^{-\beta} \quad (5)$$

where g is the geometric mean of the $\tilde{f}(X_j)$:

$$\log(g) = \frac{1}{n} \sum_{j=1}^n \log[\tilde{f}(X_j)] \quad (6)$$

and β is the sensitivity parameter satisfying $0 \leq \beta \leq 1$. Silverman (1986) concluded that there are good reasons for setting $\beta = 1/2$.

The adaptive kernel estimator is then defined as

$$\hat{f}(x) = \frac{1}{nw} \sum_{j=1}^n \frac{1}{\lambda_j} G \left[\frac{x - X_j}{\lambda_j w} \right] \quad (7)$$

The advantage of adaptive kernel estimator is adjustment to the local density of data, which is useful in smoothing the tails of the response. A natural pilot estimate in the application of the adaptive kernel method is a corresponding fixed kernel estimate (Silverman 1986). Breiman et al. (1977), Abramson (1982) and Silverman (1986) have noticed that the adaptive kernel method is insensitive to the fine detail of the pilot estimate.

It is evident that the value of the window width in the kernel function is important for optimal smoothing of the data. The formulas given for the selection of window width in the case of density estimation (cf. Eq. (4)) were not applicable here. Therefore it was necessary to find an alternative method for choosing optimal window width in order to avoid both oversmoothing and under-

smoothing. A feasible solution was to use some computer-intensive technique based on simulated resampling, such as jackknife or bootstrap methods (see, e.g., Efron 1979, Efron and Gong 1983, Taylor 1989, Hall 1990).

The presence-absence response curves were estimated separately for each taxon with a jackknife method by seeking the window width that minimized the cross-validation root mean squared error (RMSE_{cv}) between the curve and actual ob-

servations. The principle of the jackknife is to remove from the data set one sample site at a time. For each removed site, the parameter of interest is estimated with the remaining sites. This procedure is repeated for each site, and the cross-validation sum of squared residuals is calculated from the differences between the observed values and their jackknifed estimates. An example of this process is shown in Fig. 6, where the estimated response curve and its RMSE_{cv} for a number of

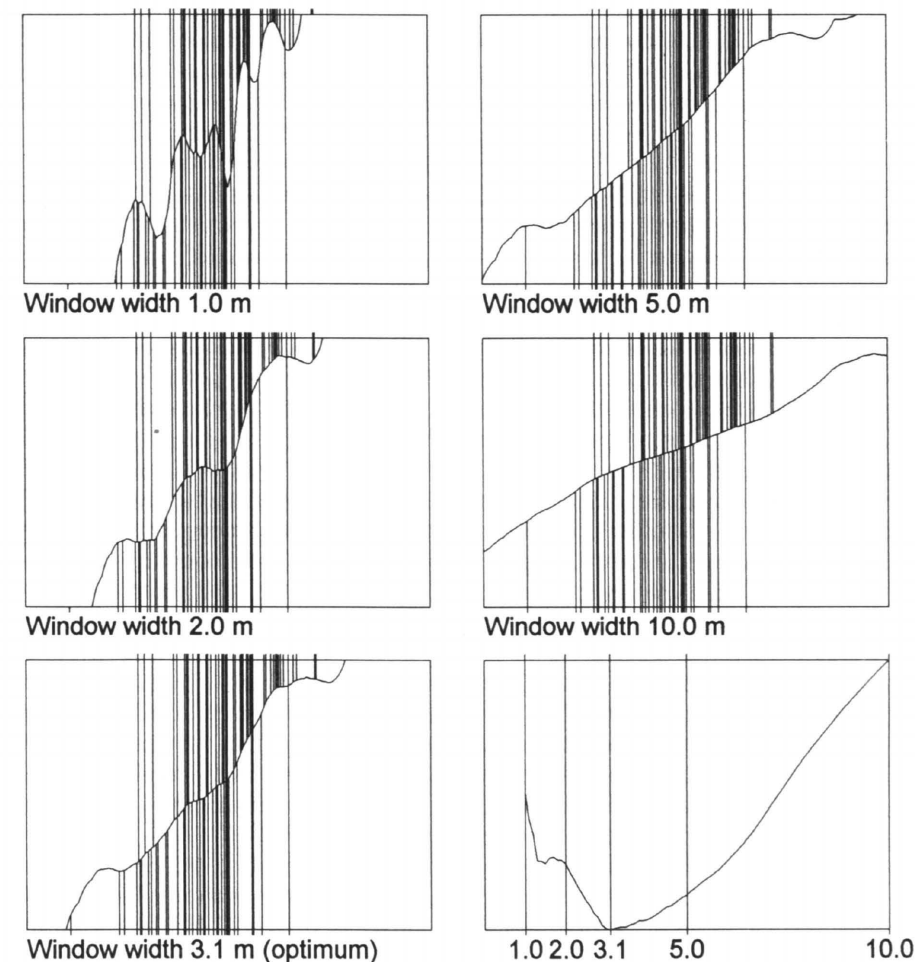


Fig. 6. An example of the estimation of the presence-absence response curve for *Calamagrostis arundinacea* in pine-dominated forests. The H_{100} range of the x-axis is 10–35 m. Between undersmoothing (window widths 1.0 and 2.0 m) and oversmoothing (window widths 5.0 and 10.0 m) there is an optimal window width (3.1 m) that minimizes the cross-validation RMSE. The behaviour of the RMSE is shown in the last figure. Vertical lines above the curve show the location of sites with the species present, those below the curve sites with the species absent.

different values of window width are shown for *Calamagrostis arundinacea* in pine-dominated forests.

For most of the common or rare species, where there were either very few absences or very few occurrences, the cross-validation estimate of window width did not have a definite optimum, implying that there were not enough observations for the estimation of the presence-absence curve. In some cases the few occurrences were so close to each other along the H_{100} axis that the jackknife estimate of the optimal window width became very small reflecting only random fluctuation in the data. This problem was avoided by setting the lower limit of acceptable window width to three meters.

2.2.4 Estimation of Site Quality from Vegetation Data

For the estimation of site quality on the basis of the vegetation composition of the sites, all the methods examined in the present study require a training set for the determination of model parameters. A jackknife method was used for assessing the accuracy of the estimates. This procedure ensured that no sampling unit could affect its own estimate. On the other hand, it may be rather demanding in terms of computing power needed. In the present study, however, the structure of the primary model (PPE), as well as its alternatives, made it possible to obtain jackknifed estimates of prediction error rather easily. Three out of the four alternative methods (FST, HC and MO) did not require knowledge of the actual site quality values in the first phase of the analysis. WA, on the other hand, related site quality directly with the occurrence of plant species, but it was computationally so simple that removal of one site at a time for calculating the cross-validation estimates could be done without difficulties.

In PPE, four different versions of the estimation method were applied. First, species-specific presence-absence response curves were estimated with a nonparametric method based on fixed kernel estimates (see Silverman 1986, Izenman 1991). Optimal window width for the kernel function was estimated with a jackknife method sep-

arately for each understory plant species. A discrete data vector with a resolution of 0.1 m and covering the H_{100} range of 0–50 m was used in calculations.

Next, an estimate of H_{100} for each site was calculated using the species composition of the understory vegetation together with the species-specific presence-absence response curves. Theoretical principles of the method are described in Chapter 4.1. From the resulting posterior probability density function, mean and variance were calculated using Eq. (10). Due to simplifying assumptions (e.g. mutually independent species responses) in the estimation method, it is possible that direct PPE estimates are biased. This bias was corrected by calculating a jackknifed linear regression equation between the estimated and measured values of H_{100} in the training set. Separate estimates were calculated for the partial (only presence data included) and complete (also absence data included) versions of the model.

Next, the behaviour of the model with more complex data-processing options were examined by estimating the species-specific presence-absence response curves with an adaptive kernel method. Also in this case, optimal window width was selected with a jackknife method. These curves were then applied to the abundance data by utilizing the pseudospecies concept. Also here, both a partial and a complete version of the estimates were calculated separately.

In FST, classification of the sites into Cajanderian forest site types was done already in the field, and calculation of site quality estimates was straightforward. In HC, clustering of the sites was taken directly from the appendix of Kuusipalo (1985). In both of these cases, class averages were used as estimates of site quality among the sites within each class.

In the MO model, DCA ordination of the vegetation data was performed with the CANOCO software (ter Braak 1988) with logarithmic transformation ($y = \log_e(100x + 1)$) of the abundance (percentage cover) values and no downweighting of rare species. Ordination of presence-absence data was obtained by replacing each occurrence in the data matrix with the value 1, and by performing an ordinary DCA ordination with no transformations. Linear regression for site index H_{100} was calculated separately for the first axis (DCA1)

and a combination of the first four axes (DCA1–4) by removing one of the sites at a time. Parameters of the regression equation were then used for obtaining an estimate of site index for the removed site. Thus, four different estimates based on the DCA ordination were obtained for both canopy tree species.

WA estimates of site quality were calculated according to the same principles, so that the indicator values and tolerances of the species were calculated by removing one site at a time from the data set. For the removed site, four estimates based on WA were calculated (abundance data vs. presence-absence data, both with and without tolerance downweighting; see Birks et al. 1990 for details). As with PPE, a linear regression equation was calculated to correct for the bias in the direct WA estimates of site quality.

2.2.5 Evaluation and Comparison of the Estimation Methods

Assessment of the predictive accuracy of the methods was based on the cross-validation root mean squared error (denoted by $RMSE_{cv}$) that was obtained from the same data set for all models. Therefore, $RMSE_{cv}$ could be used for ranking the methods according to accuracy: the smaller the $RMSE_{cv}$, the better was the method. For the classification methods FST and HC, as well as the null model (sample mean), $RMSE_{cv}$ was obtained from the standard deviation (σ) of the distribution using the equality

$$RMSE_{cv} = \sqrt{\frac{n}{n-1}} \sigma \quad (8)$$

Pairwise tests of means (with adjustments of the p values due to repeated tests) were performed for each pair of methods in order to find out which of the differences in $RMSE_{cv}$ were statistically significant. Due to deviations from normality in the sample distributions, a nonparametric signed rank test was applied. The UNIVARIATE routine of the SAS software package (SAS 1989) was used in the calculations.

3 Analysis of Species Responses to Site Characteristics

3.1 The Concept of "Site"

The number of different site variables is potentially infinite. However, ecologists usually refer to a rather limited number of site factors that are considered relevant for the understanding of various patterns and processes in forest ecosystems. The concept of "site" in these contexts is often used quite implicitly without giving any unambiguous definition for it. For example, analysis of site-specific variation in tree stand productivity is usually based on a number of fixed-sized sample plots located according to a predefined plan. From each sample plot, a number of variables is measured and subjected to statistical analysis in order to obtain results that are relevant to the problem in question. Comparison of several such studies often reveals that there is considerable variation in the sizes of the sample plots, as well as in the schemes for locating the plots. There seem to be even national preferences towards certain sample plot sizes and sampling procedures, although these preferences may also change over time.

In these studies, the concept of "site" usually refers to a single sample plot. As such it can often be used fairly consistently without any noticeable ambiguity in its interpretation. Sometimes the interpretation is not self-evident. For example, the kind of empirical data most relevant for the purposes of the present study would be sample sets where both tree stand-specific data and data on understorey vegetation composition have been collected from the same locations.

It is worth noticing that Finnish data sets of this kind are surprisingly rare. The reason for this may be that reliable analysis of the tree stand on one hand, and understorey vegetation on the other, require fairly different technical skills that are not

often fulfilled by one person. Sample plots for measuring tree stand parameters tend to be two to four degrees of magnitude bigger than those used for vegetation analysis. Therefore, a common procedure is to position – systematically, randomly, or even subjectively – a reasonable number of sampling quadrats for the analysis of understorey vegetation within the larger sample plot used for describing the tree stand. This sampling scheme was also applied in the collection of the data set analysed in the present study (Kuusipalo 1985).

This kind of sampling setup immediately raises a number of questions. How well do the vegetation quadrats, covering only a few percent of the total area of the sample plot, represent the actual variation of understorey vegetation? What is the degree of spatial heterogeneity in the structure of the tree stand? Is the sample plot located on a border between two or more sites of different "site types"? These questions draw attention to the possible spatial inhomogeneity within the sample plots. Traditionally, difficulties resulting from heterogeneity in the sample plots have been avoided by using subjective selection of sampling sites.

For the purposes of the present study, an exact definition of the "site" is needed in order to make the application of the analytic methods feasible. I shall now define a conceptual framework that will be used throughout the rest of the study. Let us examine a spatially bounded forest area A in a two-dimensional orthogonal coordinate system, with axes labeled as x and y . The location of each point ("site") within this area is given by vector $[x,y]$. Let $q(s) = q(x,y)$ denote a continuous variable that is defined for each point $s = [x,y]$, and $p(q)$ is a density function of q within area A , so that

$$\int_{-\infty}^{\infty} p(q) dq = 1 \quad (9)$$

From these definitions it follows that the mean, $E(q)$, and variance, $\text{var}(q)$, of q within area A are given by

$$\begin{cases} E(q) = \int_{-\infty}^{\infty} q p(q) dq \\ \text{var}(q) = \int_{-\infty}^{\infty} (q - E(q))^2 p(q) dq \end{cases} \quad (10)$$

In the present study we are primarily interested in the estimation of $q(s)$, the value of site quality q at a given site s . Some site characteristics, such as the diurnal and annual variation in the amount of solar radiation, may be calculated to a certain level of accuracy from general theoretical principles. However, the majority of site variables are estimated empirically by using previously collected sample sets and observed covariation between different variables. For example, there is no direct measure of the timber production potential of a given site s . This implies that the estimates of site quality can only be compared with each other, not with the "true" value of site quality. Therefore, the problem of site quality estimation is tackled mainly by analysing the coherence between different indirect measures of site quality and their applicability in different conditions.

Estimation of the density function (Eq. (9)) for site quality q within area A from empirical data requires that the sample of sites is unbiased with respect to q . This assumption can be satisfied by following established statistical sampling procedures. If there are good reasons to assume that $p(q)$ belongs to some parametric family of distributions (such as the normal distribution), then the estimation of $p(q)$ is reduced to the estimation of the parameters of the underlying density function. In the present study, however, I have applied non-parametric kernel methods for the estimation of density functions.

For the estimation of the density function of site index H_{100} in the empirical data set used in the present study, a fixed kernel method (Eq. (2)) was used with a window width chosen according to Eq. (4). The resulting density functions of site index for pine- and spruce-dominated forests deviate significantly from normality, but they are still clearly unimodal (Fig. 7, cf. Fig. 4). According to Kuusipalo (1985), the data set is virtually a random sample of the forest sites within the sam-

pling region (see Fig. 1), and therefore the results may be considered representative of the region in general.

Unimodality in the distribution of site quality (here represented by the site index H_{100}) has also been observed within the framework of the Cajanderian forest site type system. In southern Finland, there are only three important forest site types representing the average values of site quality (the *Vaccinium* site type covers 27.8 %, the *Myrtillus* site type 35.3 %, and the *Oxalis-Myrtillus* site type 13.5 %), whereas site types indicating either the poor or productive end of the site quality gradient are less common (Ilvessalo 1956).

From the viewpoint of site quality estimation, the density function of site quality offers a natural "null" model. Without additional site-specific information available, the best estimate of site quality q , in terms of RMSE, is $E(q)$ obtained from Eq. (10) for all sites within area A . In this case, the square root of $\text{var}(q)$ is an estimate of RMSE. Any estimation method based on site-specific information must perform better than this (i.e. result in lower RMSE) in order to be useful.

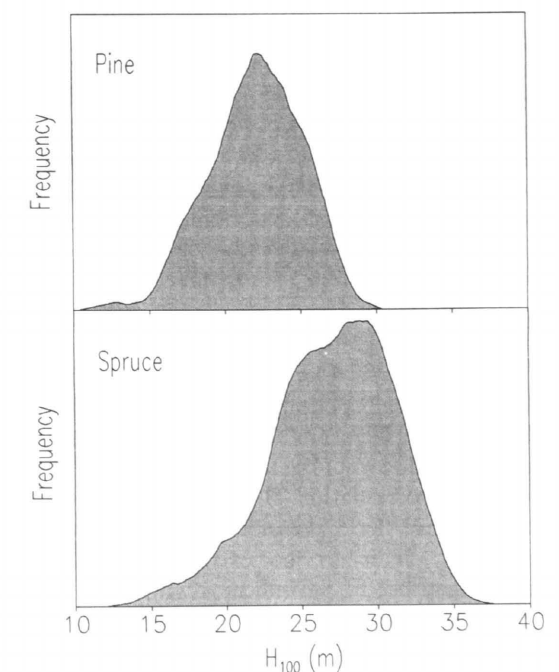


Fig. 7. Density functions for site index H_{100} in the data set, estimated with the fixed kernel method.

3.2 What is an Indicator Species?

3.2.1 Analysis of Incidence Data

The concept of "indicator species" occurs frequently in ecological literature (e.g. Hill et al. 1975, Persson 1981). Recent interest towards environmental monitoring (e.g. Goldsmith 1991) has further increased the need to find suitable population and community parameters that can be used as components in a monitoring scheme. Ellenberg's (1979, 1988) studies of the ecological "indicator values" of species for various site variables have also been the inspiration for a number of studies (e.g. ter Braak and Gremmen 1987). For the purposes of the present study, the theoretical background of the concepts of "indicator species" and "indicator value" need to be elucidated, because none of the previous studies completely satisfies the requirements of the model developed here.

Let us begin with the definition of "indicator species". In the present study I shall examine only a simple case where the responses of species occurrence and abundance are described as a function of single quantitative environmental gradient. Graphically these kinds of relationships may be shown on a scatterplot with the site variable on an x-axis and the response of the species on a y-axis. Traditionally, a distinction has been made between qualitative data, referring to the presence or absence of the species at a given site, and quantitative data, referring to the abundance (number of individuals, percentage cover, biomass, etc.) of the species (e.g. Gauch 1982).

An indicator species for site variable q is a species whose occurrence and/or abundance within the study region is dependent on the values of q . If the theoretically possible extreme values of any site variable q are examined, it is difficult to find any species that would be totally indifferent to the whole scale of variation. Therefore it is necessary to emphasize the expression "within the study region": an indicator species within one study region may not be an indicator within another region. It is also evident that the "indicator species" concept itself refers to a continuum rather than to a clear dichotomy.

"Indicator value" as used by, e.g., Ellenberg (1979, 1988) and ter Braak and Barendregt (1986)

refers to the location of species "optimum" along the environmental gradient, not to the degree of the differentiation itself. Therefore, an alternative concept is needed for this purpose. In this study I shall use the term "differentiation index" for referring to this relationship between species i and site variable q .

How should the response of species i to the site variable q be quantified? The methods of direct gradient analysis that originated from the works of Gause (1930), Ramensky (1930) and Whittaker (1956, 1967) (for a review, see Gauch 1982) are clearly of special interest in this context. Closely associated with the later evolution of direct gradient analysis, as well as various ordination methods, is the model of Gaussian species responses (see Swan 1970, Gauch and Whittaker 1972, Austin 1980, Gauch 1982, Økland 1990). It has also been criticized, for example Austin (1987) noticed that "bell-shaped response curves [of species] to environmental gradients are not universal".

Ter Braak and Looman (1986) studied the usability of weighted averaging and logistic regression for describing Gaussian-like species responses to environmental gradients. They defined the "presence-absence response curve" of a species as the probability, $p(k_i = 1|q)$, that the species i occurs in a quadrat of fixed size as a function of an environmental variable q . I shall use this probabilistic definition of species responses as a starting point for the quantification of the differentiation index discussed above. I shall first examine the case of presence-absence, or incidence, data, where only the occurrence of each species is used in the analysis.

The probability of occurrence of species i within the whole study region, $p(k_i = 1)$, is obtained by combining the presence-absence response curve with the probability density function of the site variable q :

$$p(k_i = 1) = \int_{-\infty}^{\infty} p(k_i = 1|q) p(q) dq \quad (11)$$

Probability $p(k_i = 0|q)$ is defined in an analogous way, and it integrates in Eq. (11) to $p(k_i = 0) = 1 - p(k_i = 1)$.

If species i is *not* an indicator species, then $p(k_i = 1|q)$ is independent of the value of q , and thus equal to $p(k_i = 1)$. For testing this (null) hypothe-

sis for each species i and site variable q it is easier to turn the problem inside out. Instead of $p(k_i = 1|q)$, let us examine the density function $p(q|k_i = 1)$, i.e. the distribution of variable q at sites where species i occurs. If species i is not an indicator species and $p(k_i = 1|q)$ is constant, then the null hypothesis H_0 states that $p(q|k_i = 1) = p(q)$.

Since we are interested in all differences between $p(q|k_i = 1)$ and $p(q)$ (not only differences in their means and/or variances), the test statistic should be chosen accordingly. The most important tests for comparing continuous distributions are based on the comparison of the distribution functions

$$P(q) = \int_{-\infty}^q p(u) du \quad (12)$$

of the two distributions (see Kendall and Stuart 1979, p. 467).

For detecting differences between $p(q|k_i = 1)$ and $p(q)$ in the present study, I utilized the Kolmogorov statistic D , which is defined as

$$D_i = \sup | \hat{P}(q|k_i = 1) - \hat{P}(q) | \quad (13)$$

where $\hat{P}(q|k_i = 1)$ and $\hat{P}(q)$ are the empirical distribution functions of $q|k_i = 1$ and q , respectively (see Kendall and Stuart 1979, p. 476–485, Sokal and Rohlf 1981, p. 440–445). In the discrete case of a sample of size n , $\hat{P}(q)$ is calculated as the proportion of the sampling units with a value less than q . Since $\hat{P}(q|k_i = 1)$ and $\hat{P}(q)$ lie within the unit interval $[0,1]$, so also does their maximum difference; i.e. $0 \leq D_i \leq 1$.

There are two similar tests based on the D statistic: the Kolmogorov-Smirnov one-sample test for testing the fit of a sample to a known theoretical distribution, and the corresponding two-sample test for testing the hypothesis that two independent samples have been derived from the same theoretical distribution. When using the test statistic D_i for analysing the response of species i , a theoretically correct way to obtain two independent samples would be to divide the sample of n sites into two groups on the basis of the occurrence of species i . These two subsets would then be analysed with the Kolmogorov-Smirnov two-sample test.

However, even though the cases $k_i = 1$ and $k_i = 0$ are symmetric in all formulations given above,

there is a fundamental difference in the biological interpretation of them. In theory it is possible that a given species is a noteworthy indicator of site variable q *when it is absent*, but not when it occurs. Performing a Kolmogorov-Smirnov two-sample test does not tell us which of the cases (or both) is the reason for a significant test statistic. Therefore, there are biological reason for testing the estimated probability density functions $\hat{p}(k_i = 1|q)$ and $\hat{p}(k_i = 0|q)$ against $\hat{p}(q)$, although the two samples used for estimating the density functions are not independent of each other. This approach is warranted by the following arguments:

With large sample sizes, $\hat{p}(q)$ converges towards $p(q)$ and the differences between the one-sample and two-sample tests become small. As $p(k_i = 1) \rightarrow 1$, $p(q|k_i = 1) \rightarrow p(q)$ and $D_i \rightarrow 0$. However, at the same time $p(k_i = 0) = 1 - p(k_i = 1) \rightarrow 0$ and the possible indicator value of species i is based on the cases $k_i = 0$, i.e. absence of species i at sample sites. Thus, by testing two separate null hypotheses, $H_0: p(q|k_i = 1) = p(q)$ and $H_0: p(q|k_i = 0) = p(q)$, we obtain two separate estimates for the differentiation index of species i , one for cases $k_i = 1$ (species presence) and the other for cases $k_i = 0$ (species absence).

In order to obtain a useful measure for the magnitude of differentiation of each species, I shall develop this analysis a little further. In addition to using the D statistic for testing $p(H_0)$ in different cases, it is possible to calculate the confidence interval for the values of $\hat{P}(q|k_i = 1)$ so that the true distribution function $P(q|k_i = 1)$ lies within $\hat{P}(q|k_i = 1) \pm D_\alpha$ with probability $1 - \alpha$. With a small number of the occurrences of species i , r_i , the confidence interval is very wide, but converges in approximate proportion to the square root of r_i towards zero (Kendall and Stuart 1979, p. 482).

Olkin and Spiegelman (1987) introduced a semi-parametric model for estimating density function $f(x, \pi)$ as a combination of a parametric density function (with weight $1 - \pi$) and a nonparametric density function (with weight π), where π is an unknown parameter ($0 \leq \pi \leq 1$) to be estimated from the data. In an analogous approach, let us define probability density function $\tilde{p}(q|k_i = 1, \pi_i)$ as

$$\tilde{p}(q|k_i = 1, \pi_i) = \pi_i \hat{p}(q|k_i = 1) + (1 - \pi_i) \hat{p}(q) \quad (14)$$

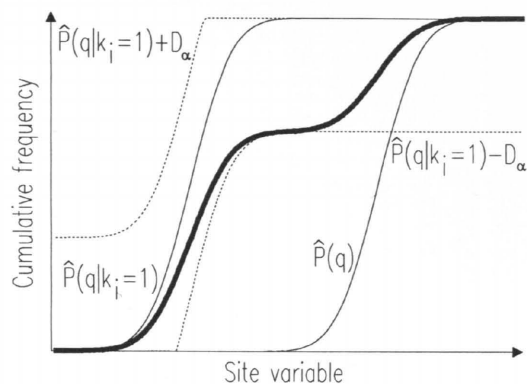


Fig. 8. A schematic view of applying the Kolmogorov-Smirnov test statistic together with confidence intervals for estimating the differentiation index of the indicator species. Parameter π_i is calculated as the minimum value, for which the compound function (thick line) of Eq. (14) lies completely within the confidence interval of $\hat{P}(q|k_i=1)$. See text for details.

The purpose of applying Eq. (14) is somewhat different from that of Olkin and Spiegelman (1987). Incorporation of $\hat{p}(q)$ into the equation is not done in order to obtain “better” estimates of $p(q|k_i=1)$, but rather to obtain “conservative” estimates (i.e. estimates where $\hat{p}(q|k_i=1, \pi_i) \approx \hat{p}(q)$, when r_i is small. This is obtained by calculating the minimum value of π_i , for which the distribution function $\hat{P}(q|k_i=1, \pi_i)$ lies completely within the $1 - \alpha$ confidence interval $\hat{P}(q|k_i=1) \pm D_\alpha$ for a given value of α (Fig. 8).

The estimate of π_i thus obtained can be used as a quantitative measure of the differentiation of species i with respect to site variable q . It incorporates satisfactorily the empirical information on the occurrence of species i at the study sites so that both sample size and deviations from the null model are taken into account. It can be seen that with small values of r_i , π_i is equal to 0 due to wide confidence limits D_α for a given value of α . On the other hand, when $r_i \rightarrow n$, the two component functions on the right side of Eq. (14) converge to a common value, and $\pi_i \rightarrow 0$. We expect to find the most important indicator species with the highest values of π_i when $0 \ll r_i \ll n$. In order to obtain a more convenient scale, differentiation index DI_i for species i is defined as

$$DI_i = 100\pi_i \quad (15)$$

With this kind of analysis, we obtain an estimate for the DI_i for cases $k_i = 1$ (species presence). An analogous procedure can be used for estimating the DI_i for cases $k_i = 0$ (species absence) with obvious modifications to the equations given above.

Further information from the type of response of species i to the environmental gradient q can be obtained by examining more closely the values of the Kolmogorov statistic (Eq. (13)). When only the overall statistical significance of the difference is examined, the absolute value of $\hat{P}(q|k_i=1) - \hat{P}(q)$ is needed. Whenever the result is statistically significant, there are biological reasons to analyse separately the magnitude of the positive difference D_i^+ defined as

$$D_i^+ = \sup(\hat{P}(q|k_i=1) - \hat{P}(q)) \quad (16)$$

and the negative difference D_i^- defined as

$$D_i^- = \sup(\hat{P}(q) - \hat{P}(q|k_i=1)) \quad (17)$$

If the value of D_i is statistically significant (i.e. $DI_i > 0$) and $D_i^+ > D_i^-$, it implies that species i is primarily an indicator of the lower end of the gradient when present. Conversely, indicators of the higher end of the gradient appear with $D_i^+ < D_i^-$. In the present study, this difference between the indicators of the higher and lower end of the gradient is incorporated into the values of DI_i by showing the indicators of the lower end of the gradient as negative values.

3.2.2 Analysis of Abundance Data

After examining species incidence data there is still another interesting question left for analysis: is there any additional information available in the quantitative data on species abundance at different sites? The probabilistic model defined for the analysis of incidence data in the previous chapters is not directly applicable to species abundance data. Therefore it is necessary to examine first some of the important issues associated with the analysis of abundance data.

In studies of vegetation-environment relation-

ships based on, e.g., Gaussian-type (parametric) species response models both abundance and probability of occurrence are used as variables. It is not easy to find any discussion referring to the relative merits of one or another as a target of modelling, although from a theoretical viewpoint they are very different entities. Regional frequency of occurrence and local abundance are not necessarily correlated (cf. Rabinowitz 1981), although due to certain biological processes it is reasonable to expect this correlation to occur (Hanski 1982, Brown 1984).

In line with this ambiguity between incidence and abundance data, the concept of “species optimum” may also refer to that portion of the environmental gradient where either the species reaches its maximum frequency of occurrence or where its abundance is at maximum. Examination of scatterplots showing species abundance values as a function of a certain environmental variable may indicate that there are a number of sites with very low species abundance values even in the optimal region of the variable (e.g. Nieppola 1993a: Fig. 3). This phenomenon is easy to understand on the basis of multidimensional niche theory (e.g. Crawley 1986). If species responses are even approximately Gaussian in several ecological dimensions, then the transect along any single axis contains the whole range of variation below the upper limit of abundance determined by the variable in question. High abundance values are obtained only when all significant environmental factors are simultaneously advantageous.

This phenomenon has important implications for the use of individual species as indicators of site variables. It is easy to see that whenever the species occurs with a high abundance value, we are close to its optimum region. However, the opposite is not true: from a low abundance value we often cannot deduce that the site is not suitable for the species *with respect to a certain site variable* q . There may be a number of other factors limiting the abundance of the species even at the optimal level of q . This asymmetry should be taken into account when utilizing species abundance data for the estimation of site characteristics.

The asymmetry between high and low values of species abundance can be conveniently taken into account within the probabilistic modelling framework of the present study by using the con-

cept of “pseudospecies” introduced by Hill et al. (1975). The main idea behind the method of pseudospecies is to reduce quantitative data to presence-absence data without undue loss of information. This is accomplished by converting the frequency data into classes, each class being treated as though it were a species in its own right. For each pseudospecies determined by the limit abundance a (where, in the case of percentage cover, $0 \leq a \leq 100$), occurrence of a pseudospecies (denoted by $k_{ia} = 1$) is recorded when the abundance of the corresponding “real” species i is higher than a . Correspondingly, the notation $k_{ia} = 0$ is used to indicate that the abundance of species i is less than or equal to a .

The pseudospecies concept reduces the analysis of abundance data into the framework defined previously for the analysis of incidence data. The presence-absence response curve $p(k_i=1|q)$ is extended into $p(k_{ia}=1|q)$ which describes the probability of occurrence of pseudospecies i_a as a continuous function of q and a (notice that $k_{ia} = k_i$ when $a = 0$). A new dimension is also added into the calculation of the differentiation index DI_i . In addition to calculating separate values for the presence and absence data, it is also possible to calculate the index as a function of limit abundance a (denoted by DI_{ia}). If any of the pseudospecies' values is higher than that calculated for the presence or absence data, we may conclude that there is additional information available in species abundance data beyond mere incidence data. If necessary, it is also possible to test the differences between the density functions $\hat{p}(q|k_{ia}=1)$ and $\hat{p}(q|k_i=1)$ with the Kolmogorov statistic, as described previously.

3.2.3 Empirical Results

The total numbers of taxa recorded in pine- and spruce-dominated forests were 97 and 109, respectively. Of these, 35 in pine-dominated forests and 52 in spruce-dominated forests had non-null index values in at least one type of category (Table 2). There were considerably more taxa indicating the productive end of the site index gradient than those occurring primarily at poor sites. This result is in agreement with the previously published data on the Finnish forest flora

Table 2. Numbers of taxa in different categories of the differentiation index DI in the empirical data set. Columns "low" and "high" show the numbers of taxa indicating the poor and productive end of the site index gradient, respectively. Notice that the values in the "Total" columns are not obtained by summing up the values in other columns, because some taxa occur in more than one category.

	Incidence			Abundance			Absence			Total
	Low	High	Total	Low	High	Total	Low	High	Total	
Pine	4	21	25	8	27	35	2	3	5	35
Spruce	4	25	29	15	35	50	8	1	9	52
Total	6	36	42	20	48	64	9	4	13	65

(e.g. Kalliola 1973, Kujala 1979). When abundance is taken into account via the application of the pseudospecies concept, increase in the number of useful taxa is most notable among the indicators of sites with low values of site index H_{100} . In pine-dominated forests the number of taxa doubles (from four to eight), and in spruce-dominated forests the number almost quadruples (from four to 15).

Species absence was interesting among very few taxa (Table 2). This result is a natural outcome of the uneven distribution of taxon frequency in the data set: the majority of taxa are very rare. The most notable exception are the eight common taxa whose absence indicates poor productivity in spruce-dominated forests.

Taxon-specific indices (Appendix 1) do not reveal any surprising patterns in the types of response. The taxa obtaining highest DI scores are mostly indicator species also in the Cajanderian forest site type system (cf. Kalliola 1973, Kujala 1979). In pine-dominated forests, the most reliable indicators of high site index are (index value given in parentheses) *Rubus saxatilis* (39), *Solidago virgaurea* (35), *Melampyrum sylvaticum* (31), *Pteridium aquilinum* (27), *Brachythecium* spp. (27), *Viola riviniana* (26), *Dryopteris carthusiana* (24) and *Deschampsia cespitosa* (20). The poor end of the gradient is characterized by the lichen species *Cladonia arbuscula* (-19), *C. rangiferina* (-16) and *Cetraria islandica* (-15), with *Calluna vulgaris* (-13) as the only vascular plant species.

Differentiation of pseudospecies based on abundance data increases the number of indicator taxa from 21 to 27 at the productive end, and from four to eight at the poor end of the gradient. The taxa

with the greatest increase in the index value in abundance data were *Calamagrostis arundinacea* (from 10 to 36), *Maianthemum bifolium* (0→25), *Luzula pilosa* (0→19), *Pteridium aquilinum* (27→40) and *Oxalis acetosella* (7→20) at the productive end of the gradient. At the poor end of the gradient, the indicator taxa mentioned above do also have additional information available in abundance data: *Cladonia rangiferina* (-16→-39), *Cladonia arbuscula* (-19→-40) and *Calluna vulgaris* (-13→-34). *Arctostaphylos uva-ursi* (0→-22) is the most important addition to the list.

The absence of taxa is naturally indicative of the opposite end of the site index gradient in relation to their presence. Thus, absence of *Calluna vulgaris* (37), *Cladonia rangiferina* (19) or *C. arbuscula* (2) indicates high site quality, whereas absence of *Calamagrostis arundinacea* (-25) or *Trientalis europaea* (-6) indicates low productivity.

In spruce-dominated forests the list of taxa is somewhat different. The most reliable indicators of high productivity are *Rubus idaeus* (43), *Dryopteris carthusiana* (43), *Plagiothecium* spp. (32), *Veronica officinalis* (26), *Carex digitata* (26), *Melica nutans* (26) and *Poa nemoralis* (26), whereas *Cladonia rangiferina* (-29), *Calluna vulgaris* (-27), *Goodyera repens* (-6) and *Diphasiastrum complanatum* (-2) indicate low productivity. The number of indicator pseudospecies increases from 25 to 35 at the productive end of the gradient. The highest increases in index values occur in *Oxalis acetosella* (5→47), *Viola riviniana* (0→32), *Brachythecium* spp. (15→40), *Fragaria vesca* (17→39), *Pteridium aquilinum* (0→21), *Luzula pilosa* (0→19) and *Hepatica nobilis* (21→34). At the sites with low productivity,

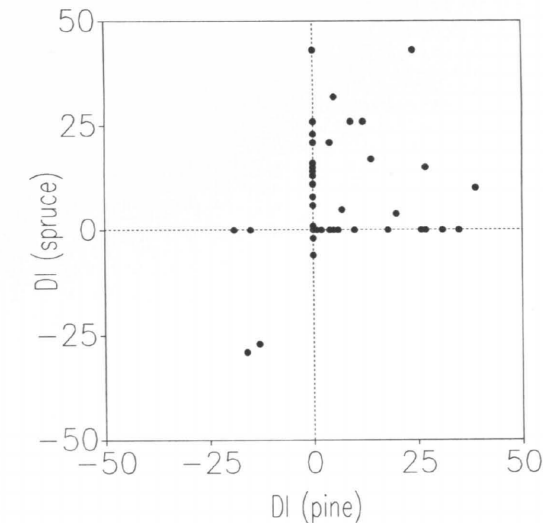


Fig. 9. Value of the differentiation index DI in pine-dominated forests (x-axis) vs. the index value in spruce-dominated forests (y-axis). Each dot represents one taxon.

the number of indicator pseudospecies increases from four to 15, with *Vaccinium vitis-idaea* (0→-54), *Hylocomium splendens* (0→-38), *Pleurozium schreberi* (0→-32), *Melampyrum sylvaticum* (0→-25) and *Vaccinium myrtillus* (0→-23) showing greatest increase in the index value.

In terms of species absence in the spruce-dominated forests, *Deschampsia flexuosa* (18) is the only indicator taxon at the productive end of the gradient. At the less productive sites, the absence of eight taxa provides interesting information: *Maianthemum bifolium* (-36), *Dryopteris carthusiana* (-34), *Brachythecium* spp. (-27), *Trientalis europaea* (-24), *Luzula pilosa* (-23), *Oxalis acetosella* (-19), *Plagiothecium* spp. (-10) and *Carex digitata* (-1).

Comparison of the taxon-specific index values in pine-dominated forests vs. spruce-dominated forests (Fig. 9) indicates that there are a few taxa that are indicators in both types of forests, whereas a considerable number of taxa are indicators in association with only one of the canopy tree species. The upper left and lower right quarters of the scatterplot in Fig. 9 are empty, showing that no taxa have opposite responses in pine- vs. spruce-dominated forests.

Note that the index value of zero does not nec-

essarily imply that the occurrence of the taxon is indifferent with respect to site quality. Because the index value depends on the confidence limits of the Kolmogorov-Smirnov index as well as its absolute value (see Chapter 3.2.1.), it is possible that the number of observations is too small to make reliable conclusions of the ecological response of the taxon in question.

3.3 Estimation of Presence-Absence Response Curves

3.3.1 Theoretical Background

The methods described in the previous chapter make it possible to identify potentially interesting indicator species on the basis of a single parameter (differentiation index DI). It can be used in the preliminary selection of taxa for further analysis from the available species pool. When necessary, both species absence and abundance data can be examined separately. The next logical step is to analyse in greater detail the actual presence-absence response curves of the species that obtain nonzero DI values in at least one of the categories.

Ter Braak and Looman (1986) used logistic regression and Gaussian response models for describing species occurrence patterns along environmental gradients. The key concepts in this approach were "optimum" or indicator value (the value of q with the highest probability of occurrence) and "tolerance" (a measure of ecological amplitude). This method has also been used by Niippola (1993a) for describing the probability of occurrence of understorey plant species as a function of site quality in the forests of southern Finland.

In this study I applied nonparametric methods for the estimation of presence-absence response curves. The advantage of nonparametric methods over their parametric alternatives is that it is not necessary to make any prior assumptions on the shape of the response. It is possible to view nonparametric methods as a generalization of parametric methods. If the response modelled by nonparametric methods seems to be close to the corresponding parametric model, the selection between the two may be based on statistical criteria

(Hastie and Tibshirani 1990, Yee and Mitchell 1991). The actual details of the estimation procedure have been described already in Chapter 2.2.3.

3.3.2 Empirical Results

The presence-absence response curves were calculated for all the taxa that had enough presences (rare taxa) or absences (common taxa) so that the jackknife estimation of the optimal window width was successful (Appendix 2). The most suitable published data for comparing the results are those of Nieppola (1993a: Fig. 2), who shows presence-absence response curves estimated with parametric methods for a number of understorey species in pine-dominated forests. The general patterns in Nieppola's results are very similar to those of the present study. Due to a larger data set (222 sample plots), a number of rare species could also be modelled better than in this study.

No taxa modelled in these studies show contradictory responses. However, there are many differences in the fine details of the responses. The most notable general pattern is that Nieppola's results show response curves that run more clearly from (near) zero probability to (near) unity probability of occurrence, whereas the curves in the present study indicate less steep responses. Most of the difference may be explained by differences in the sizes of sample quadrats, but more data are obviously needed in order to obtain a better view of the actual response patterns.

As described in the previous chapters, analysis of species abundance data with the PPE method is possible by applying the concept of pseudospecies originally defined by Hill et al. (1975). The presence-absence response curve can be given as a continuous function of site quality q and limit abundance a . The definition of the pseudospecies concept implies that as the limit value a increases, the probability of occurrence decreases. A contour plot with a number of limit values was calculated for those taxa whose differentiation indices (Appendix 1) indicated that they contain significant information in abundance data (Appendix 3). Notice that for the estimation procedure described in Chapter 4, the presence-absence response curves of pseudospecies may be used instead of the curves of the "real" species without any mod-

ifications to the estimation procedure.

It is worth noticing that the differentiation indices and the presence-absence response curves are empirical results obtained from the data set available for analysis. With the accumulation of new data these values may change. Therefore, given enough time it is possible to obtain a gradually improving insight into the specific response of each species, even those at the rarest and commonest end of the spectrum.

3.3.3 The Effect of Quadrat Size

Ter Braak and Looman (1986) defined the presence-absence response curve of a species as "the probability $p(q)$ that the species occurs in a quadrat of fixed size as a function of an environmental variable q ". What happens if quadrat size is changed? This question has important implications for the present model, because the quadrat size may vary among different data sets. If species-specific presence-absence response curves are also sensitive to the quadrat size used in the collection of field data, there is one more factor added to increase the complexity of the model.

It appears that there is no general theoretical solution to the problem. In order to examine the patterns to be expected in empirical data, we can simplify the problem without losing generality. Let us assume that, instead of continuously varying quadrat size, we have from each sample plot a varying number, m , of quadrats of fixed size. If $p(k_i = 1|q, 1)$ is the probability that species i occurs within one quadrat, given q , what is the corresponding probability $p(k_i = 1|q, m)$ that species i occurs within a set of m quadrats?

If the m quadrats are independent of each other in terms of the occurrence of species i , we can apply the multiplication rule of independent probabilities to obtain

$$p(k_i = 1|q, m) = 1 - [1 - p(k_i = 1|q, 1)]^m \quad (18)$$

It is easy to see that $p(k_i = 1|q, m) \geq p(k_i = 1|q, 1)$ for all $m \geq 1$.

If there is spatial covariance between the quadrats, we can assume that species occurrence is aggregated, as is usual when sampling biological populations. This phenomenon can be described

by dividing the variance of the occurrence of species i into fine-scale (within one sample plot) and broad-scale (between sample plots) components. If the quadrats are totally independent of each other, the broad-scale component of variance is zero and Eq. (18) gives an accurate estimate for $p(k_i = 1|q, m)$. At the other extreme, if the fine-scale component of variance is equal to zero, then all variation in the occurrence of species i is between different sample plots. This implies that $p(k_i = 1|q, m) = p(k_i = 1|q, 1)$ for all values of m .

Given these assumptions, we see that the expected lower limit for $p(k_i = 1|q, m)$ is $p(k_i = 1|q, 1)$ and the expected upper limit is obtained from Eq. (18). The actual value of $p(k_i = 1|q, m)$ within this range must be determined empirically. Furthermore, it is probable that the value is species-specific and also dependent on the value of q . This approach is easily generalized to analyse continuous variation in quadrat size. It is worth mentioning that, in the estimation of presence-absence response curves from empirical data, the aggrega-

tion of vegetation quadrats (i.e. several quadrats within one sample plot) does not affect the estimation procedure. For an alternative approach for analysing the same problem, see Crawley (1986).

In conclusion, the presence-absence response curves estimated in the previous chapters are dependent on the size of the sampling quadrat. For example, part of the differences in the species-specific responses obtained in this study and by Nieppola (1993a) may result from different reference areas. The data used by Nieppola (1993a) were collected from twenty 1×1 m² quadrats that were located within a sample plot 50×50 m². However, all species occurring outside the vegetation quadrats but inside the sample plot were also recorded. Therefore it remains somewhat ambiguous whether the proper quadrat size for the presence-absence response curves shown in his study was 20 m² or 2500 m². From the viewpoint of the theoretical analysis given above this is not a trivial problem.

4 Estimation of Site Characteristics from Understorey Vegetation Composition

4.1 Theoretical Background

The theoretical background for the estimation of site characteristics from the composition of understorey vegetation is developed by expanding the framework defined in Chapter 3.1. Estimation of a given site variable $q(s)$ at site s within area A may be viewed as a problem of determining a posterior probability density function $p(q|K(s))$, given $p(q)$ and some additional site-specific information $K(s)$. If there is no additional information available for the estimation procedure, we have $p(q|K(s)) = p(q)$ for all q .

For the purposes of the present study, additional information $K(s)$ is defined to be a vector of n random binary variables, $[k_i(s)]$, where $i = 1, \dots, n$. For each site s within area A , each of the vector elements $k_i(s)$ is either 0 or 1. Thus, in the analysis of empirical vegetation data the vector $[k_i(s)]$ gives the presence (value 1) or absence (value 0) of each (pseudo)species at site s . For utilizing this additional information $K(s)$ for the estimation of site quality, we apply Bayes' formula and write

$$p(q|K(s)) = \frac{p(K(s)|q)p(q)}{\int_{-\infty}^{\infty} p(K(s)|q)p(q) dq} \quad (19)$$

Notice that the integral in the denominator is equal to $p(K(s))$.

A verbal description of this estimation procedure is as follows: In order to calculate an estimate for the value of variable q at site s , given the observed species composition of understorey vegetation shown as the elements of the vector $[k_i(s)]$, we shall first try to estimate, for all possible values of q , the probability of having such a species composition. This probability is then weighted by the relative frequency of the sites with the given value of q . These probabilities are then summed

together (integral in the denominator) in order to obtain a scaling factor for individual probability values.

Apart from unrealistically simple cases, estimation of $p(K(s)|q)$ is a demanding task. Therefore, in reality it is necessary to make some simplifying assumptions in order to make Eq. (19) computationally feasible. Assuming that the vector elements $k_i(s)$ are independent of each other for given q , probability $p(K(s)|q)$ can be written, by applying the multiplication rule of independent probabilities, as

$$p(K(s)|q) = \prod_{i=1}^n p(k_i = k_i(s)|q) \quad (20)$$

where $k_i(s)$ is the value (0 or 1) of the i th element in vector $[k_i(s)]$ at site s . The values $p(k_i = k_i(s)|q)$ in the product on the right side of Eq. (19) can now be obtained from the species-specific presence-absence response curves estimated in the previous chapters.

With these assumptions we have now a model that can be applied for solving Eq. (19) with empirical data. The result of this estimation procedure is a probability density function, showing the so-called posterior probabilities of variable q , given the prior distribution $p(q)$ and additional information $K(s)$. When a single value is needed as an estimate of $q(s)$, a natural unbiased choice is the mean, $E(q|K(s))$, of the distribution (Eq. (10)). Correspondingly, variance of the posterior probability density function may be compared with that of the prior distribution: the smaller is the ratio $\text{var}(q|K(s)) / \text{var}(q)$, the better is the species composition available at the site as an indicator of q . It is worth noticing that this ratio should always be less than one. If $\text{var}(q|K(s))$ is greater than $\text{var}(q)$, it implies that incorporation of additional information $K(s)$ does not improve the estimates.

In nature, this kind of situation may occur when indicator species of both ends of the gradient (of variable q) occur simultaneously at the same site. If this is the case, some kind of habitat heterogeneity within the sample plot is the most probable explanation.

It is evident that some of the assumptions needed to make the model operational (e.g. the independence of species responses) are violated in nature. However, the ultimate criterion for the goodness of the model should be its performance (in terms of its accuracy and consistency) in the analysis of actual empirical data.

4.2 Empirical Results

The model described in the previous chapter was applied to the empirical data set in order to obtain estimates of site index H_{100} for each sample plot. Four alternatives of the basic model were applied. A simple version of the estimation procedure comprised species-specific presence-absence response curves estimated with a fixed-kernel method, and only incidence data were used in the estimation procedure. A partial model where only species presences were taken into account, and a complete model where both presence and absence data were incorporated into the estimation of the posterior probabilities, were applied separately. A more sophisticated version of the model used the adaptive kernel method for the estimation of presence-absence response curves and utilized quantitative abundance data via the pseudospecies concept. Also in this case, separate estimates were calculated for the complete and partial model. In order to correct for possible systematic biases in the estimates resulting from unrealistic model assumptions, a linear regression equation was calculated between measured and predicted values of H_{100} before computing the estimates of the error values. This procedure may be criticized for theoretical reasons, but in practical applications of the model it is always possible to make corrections to esti-

mates if better results are to be expected. A similar approach was adopted also in association with the weighted averaging (WA) estimates (see Chapter 6.5.).

The results (Table 3, Fig. 10) indicate that the model produces satisfactory estimates of the site quality in the sense that estimation error is reduced relative to the "null" model. For both tree species, application of a more sophisticated version of the model (with adaptive kernel estimates and abundance data) did not improve the accuracy of the estimates over the simple version. In pine-dominated forests, a complete model (with absence data included) gave somewhat better results, whereas in spruce-dominated forests the partial model was slightly better. Overall, the reduction in RMSE_{cv} relative to the "null" model (the background distribution, see Table 1) was not very impressive: from 2.98 m down to 2.32 m in pine-dominated forests, and from 3.94 m down to 3.13 m in spruce-dominated forests.

Evaluation of these results in isolation is fairly difficult. Therefore the next chapter describes four different methods based on the analysis of understorey vegetation composition that can be used as tools for the estimation of site quality. Each of these methods is applied to the same data set in order to obtain comparable assessment of their relative accuracy.

Table 3. Accuracy of the PPE estimates applied to the empirical data set. Values of RMSE_{cv} (in meters) are given for four different versions of the model: incidence data with fixed kernel estimates of the presence-absence response curves, and abundance data (pseudospecies) with adaptive kernel estimates of the response curves.

	Incidence data		Abundance data	
	Complete	Partial	Complete	Partial
Pine	2.34	2.51	2.74	2.78
Spruce	3.30	3.16	3.25	3.44

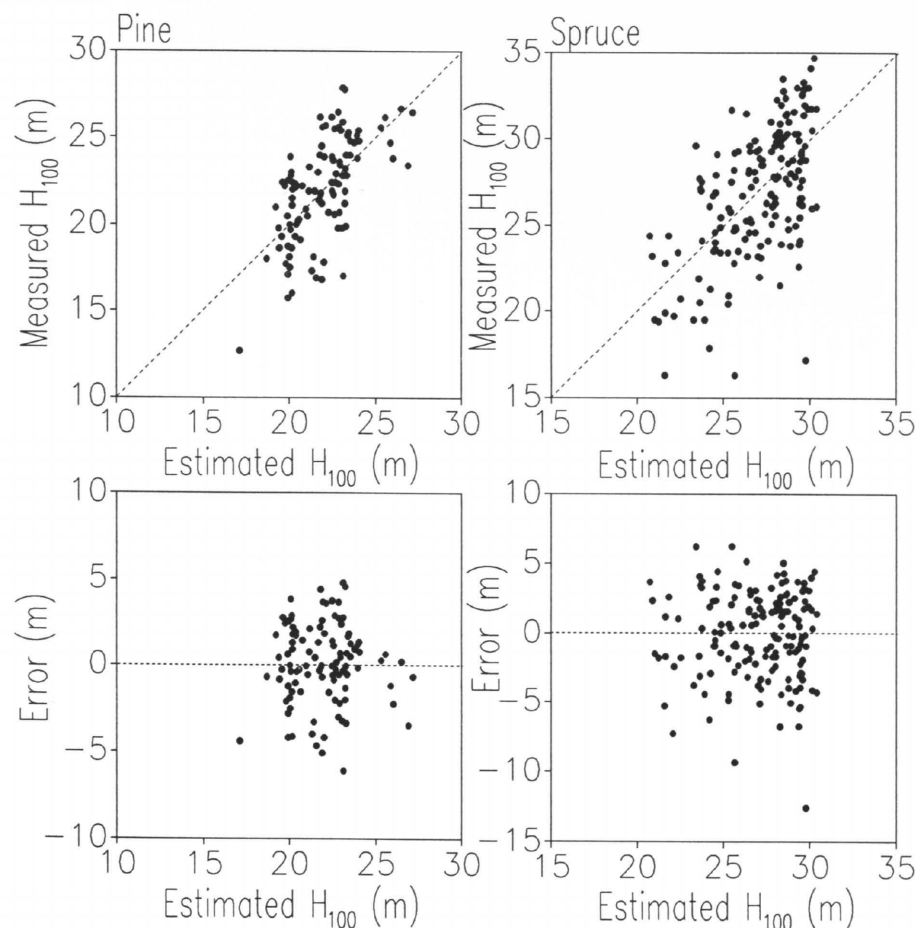


Fig. 10. Measured vs. estimated values of site index H_{100} in the version of the PPE model that performed best in terms of $RMSE_{cv}$. For pine-dominated forests, it was the fixed-kernel version with incidence data and a complete model (absence data included). For spruce-dominated forests, the partial model with incidence data is shown.

5 Alternative Estimation Methods

5.1 Overview

In order to obtain a more realistic view of the performance of the PPE model in the estimation of site quality in the upland forests of southern Finland, four alternative estimation methods were chosen to be applied to the same empirical data set as the PPE. The basic requirement for all the methods that were selected for the comparison was that their principles and computational algorithms have been satisfactorily published in scientific literature. Previous usage in practical forest management in Finland was not required; it is well known that the Cajanderian forest site type system has been by far the dominant scheme applied in the forestry in Finland (e.g. Mikola 1982).

The alternative estimation methods described below may be grouped according to several different criteria. Two of the methods, Cajander's (1909) forest site type system and Kuusipalo's (1985) hierarchical clustering scheme are typical classification methods, where the sample sites are grouped into a reasonable number of site types or classes. The main difference between these two methods is that HC is based on more explicitly defined criteria for assigning a site into a certain class, whereas FST depends more on the expertise of the field workers. Estimation of the site quality with these methods is based on calculating average values within each class.

The other two methods, multivariate ordination and regression and weighted averaging are quantitative methods that can be used for estimating the values of site quality in a continuous scale. The fundamental methodological difference between these methods is in their treatment of the vegetation data. MO is a typical "community"-type approach: its starting point is a species-by-sites matrix, from which the principal gradients of variation are extracted with multivariate methods. WA, on the other hand, is a typical indicator

species approach: the response of each species to the environmental gradient is analysed independently of the other taxa.

In the following chapters, the basic theoretical principles of each method are described and their previous usage in Finnish forest research and management are shortly reviewed. Each method is then used for obtaining cross-validation estimates of H_{100} and their RMSE in the same data set that was used for testing the PPE method above. This approach enables a side-by-side comparison of the relative performance of the alternative methods as tools in practical site quality estimation.

5.2 Forest Site Types

5.2.1 Theoretical Background

The theory of forest site types was introduced by Cajander (1909). Since then it was developed further in a number of other publications, both by Cajander and his colleagues and successors. Among the more recent review articles published in English, those by Frey (1973) and Mikola (1982) are worth noticing. In the present study it is possible to give only a very superficial treatment to the diverse field of scientific and practical questions associated with the Cajanderian theory of forest site types.

If the frequency distribution of site quality is a natural "null" model in a situation where there are no data on the understorey vegetation available, Cajander's classification of forest site types has a similar status among the actual estimation methods. Since its publication, FST has been widely used as a tool in silviculture and forest management in Finland, and it has also had some influence on site quality estimation in other countries.

FST is a non-hierarchical classification system. Each site is classified into one of the predefined site types according to the principles summarized

by Cajander (1949): "All those stands are to be classed in the same forest type, the vegetation of which, when the stand is exploitable or nearly exploitable, and of normal density, is characterized both by mainly identical floristic composition and a similar ecologico-biological nature, as well as all those stands the vegetation of which differs from that defined above only in those respects which – resulting from the difference in age of the stand, fellings, change in species of stand, etc. – have to be considered as merely accidental or ephemeral or at any rate not permanent. Permanent differences call forth a new forest type where they are sufficiently well-marked, or a sub-type where the differences are less essential but nevertheless noticeable." One of the key concepts in Cajander's description of the theory of forest site types was interspecific competition in undisturbed forest vegetation that was believed to result in distinguishable community types. Cajander derived the basic ideas for this view of vegetation pattern as a result of competitive interactions from his studies of the shore vegetation along the Lena river (e.g. Nieppola 1986, Oksanen 1991). However, in later studies Cajander noticed that in the forest vegetation community borders are more blurred than along the shores of rivers (e.g. Cajander 1949).

Whether the site type classes thus formed are natural or more or less artificial entities is not easy to resolve. In vegetation science, the problem of continuity has inspired a considerable amount of both theoretical and empirical research (Austin 1985, Shipley and Keddy 1987, Austin and Smith 1989). As a proponent of the community view of vegetation, Cajander himself proposed that the goal of site classification should be to find "natural" quality classes which "combine sites biologically near together but keep apart those which differ biologically" (Cajander 1949).

The value of forest site types as a basis for the estimation of growth and yield of tree stands was empirically demonstrated by Ilvessalo (1920; see also Ilvessalo and Ilvessalo 1975). He showed that the site types can be arranged into a natural series where average stand productivity differs from one site to another. These results, originally published in Finnish, were made accessible to a wider audience when reviewed in German by Cajander and Ilvessalo (1921). Tamm and Petrini (1922a), in a

review of the "Ueber Waldtypen II", criticized some of their conclusions. The discussion that followed (Cajander and Ilvessalo 1922, Tamm and Petrini 1922b) further elucidated the fundamental differences in the study approach between the Finnish and Swedish schools of forest research. The assertion of Cajander and Ilvessalo (1921), based on the earlier work of Ilvessalo (1920), was that forest site types form rather homogeneous classes that differ from each other with respect to all significant measures of site quality. Tamm and Petrini (1922a) questioned this conclusion and suggested that, given the empirical data available, sample sites are distributed continuously over the whole range of variation in site quality.

As can be seen from the results of later studies (e.g. Kuusipalo 1985, Nieppola and Carleton 1991), the arguments of Tamm and Petrini have strong support in empirical data. Unfortunately the debate between Cajander and Ilvessalo vs. Tamm and Petrini faded to become a frequently cited curiosity in the historical evolution of the Cajanderian system. However, the issue is still as topical as before. Scientific interest into the problems of site classification in Finnish forests has been revived recently (Hotanen and Kuusipalo 1988), but considerable effort is still directed toward the problems of improving the classification scheme (e.g. Kuusipalo 1985, Nieppola 1993b) or identifying different forest site types under the influence of modern silvicultural practices (e.g. Vanha-Majamaa 1993).

Cajander's preference for classification methods over quantitative estimation methods is expressed clearly in several of his papers. Even in his last article on forest site types, Cajander (1943; English translation in Cajander 1949) criticized the method of deriving site quality classes from the (at that time graphical) analysis of stand parameters as a function of stand age on the basis that "the site quality classes obtained in this way are but graphic constructions arbitrarily limited – the number of quality classes distinguished may be 3, 5, 10, or even more – and that as such they have no equivalent in nature". The critique continued elsewhere in the same article: "These auxiliary methods, each of them individually, and especially all of them together, naturally give a fairly good guidance when drawing the normal curves, and their direction can thus be drawn with consid-

erable accuracy. But the main thing, the quality classification, remains unsettled. These methods will, it is true, help us to draw the growth curves with relative accuracy, but *there will be an infinite number of them* [italics added]. All these auxiliary methods necessarily require that the quality classes are already determined in advance in some other way."

These examples suggest that Cajander was persistently reluctant to the adoption of the idea of continuous variation in site characteristics and understorey vegetation, and the implications of this observation. Instead, his goal was to establish the classification of forest sites on "natural quality classes which would not be based on graphic or mathematical constructions of any kind" (Cajander 1949).

With modern analytical methods, the question of the existence of distinguishable community types among the forest vegetation should be easier to approach empirically (cf. Shipley and Keddy 1987). Empirical data sets have often a methodological bias that cannot be ignored: in many studies, sample plots have been located at sites with homogeneous vegetation patterns based on subjective evaluation. This kind of sampling pattern implies that gradual variation in natural vegetation becomes underrepresented in the sample, and consequently the analysis of the data set is expected to reveal community patterns that seem to be more distinct than they actually are in nature.

When taking into account the direction of this bias, it is surprising to notice how consistent different studies based on multivariate ordination of vegetation data are in their inability to support the notion of distinct community types in empirical data sets (e.g. Kuusipalo 1985, Lahti and Väisänen 1987, Tonteri et al. 1990a, 1990b, Nieppola and Carleton 1991). The ability to identify distinct forest site types on the basis of the understorey vegetation alone seems to be in the eyes of the beholder. The hypothesis of the continuity of the vegetation gradient along the site type sequence receives further support from the application of the FST in the field. In addition to the actual forest site types, there are a number of established intermediate types that are used to make the transition from one "true" site type into another more gradual. For example, transition from the subdry

vegetation of the *Vaccinium* site type (VT) into moister vegetation of the *Myrtillus* site type (MT) is smoothed with site types denoted by VT+ and MT- (see also Keltikangas 1959).

From the viewpoint of applicability in the field, FST is superior to the other methods. Identification of the forest site type can be done directly on location, and it is in fact more difficult to determine the "correct" site type on the basis of the vegetation description alone. However, the price of this straightforward applicability of the FST in the field is considerable subjectivity in practice. There are a number of field guides (e.g. Kujala 1979, Lehto and Leikola 1987) for helping the determination of the site types on the basis of the presence and absence of certain indicator species, but there is no guarantee that two independent observers obtain identical classification of the same sites. This problem becomes more pronounced when the same site is classified at different times with changes in environmental conditions and stand structure.

Determination of forest site types is based on the overall pattern in vegetation. Particular attention is drawn to a number of indicator species whose presence or absence and vitality are known to differ from one site type to another (e.g. Kalliola 1973, Kujala 1979, Mikola 1982, Lehto and Leikola 1987). It is worth noticing that most of the indicator species are indicators of either very poor or productive sites, whereas the number of species with an optimum at intermediate values of site quality is quite low. Characteristic indicators of poor sites are lichens, whereas a number of herb species can be used for distinguishing different site types at the most productive end of the gradient.

There is always a certain degree of subjectivity in the application of FST in the field. However, later studies with multivariate ordination methods (e.g. Lahti and Väisänen 1987, Tonteri et al. 1990a, 1990b) have indicated that the forest site type sequence can also be detected with more objective methods.

One of the empirical observations noticed in several studies is that the frequency distribution of Finnish forest sites into different site types is very uneven (e.g. Ilvessalo 1956). The three most common site types cover up to 70–80 % of the land area, whereas the rarest site types occupy

less than one percent. The most common site types tend to be those with average site quality, and sites with extremely low or high productivity are less common.

Estimation of site quality according to the FST system is straightforward. Each forest site is assigned into one of the site types. An unbiased estimate of site quality within each site type is the average site quality among the sites assigned to that type within the training set. Overall performance of FST depends on the accuracy of the estimates within each site type and the relative proportions of different site types within a given region.

When comparing the Cajanderian forest site type system with the other estimation methods according to the criteria defined above, it must be taken into account that the basic principles of the FST were defined more than 80 years ago. After its introduction in 1909, the forest site type approach was refined during the following decades (e.g. Leikola 1984, Nieppola 1986). One of the significant extensions over the original model was the introduction of climatic parallel types (Kujala 1936, 1938, Kalela 1961).

On the other hand, the fact that the FST is still in widespread use in practical forestry and even in biological and forest research (see Mikola 1982, Lehto and Leikola 1987) is a sufficient reason for evaluating its basic approach and performance according to modern standards. The only viable alternative for the Cajanderian site classification that is in general use in the forestry in Finland today is the site index based on the height growth of dominant trees, and it has also a somewhat limited scope of application (see Vuokila and Väliäho 1980, Gustavsen 1981).

5.2.2 Empirical Results

An unbiased estimate of the quality of the site is the average quality of the site type. Table 4 shows the parameters of the distributions of site quality among different site types in the sample data. The results indicate that, on the average, the site types form a natural sequence of increasing site quality, although the overlap of distributions between neighboring site types is considerable (Fig. 11; cf. Lahti and Väisänen 1987). Due

Table 4. Statistics of the site index estimates for different site types in the Cajanderian system (FST). All values are in meters.

Type	n	Mean	s.d.	Min	Max	
Pine	OMaT	1	23.8	.	23.8	23.8
	OMT	5	25.7	1.34	23.4	26.7
	MT	40	23.4	1.94	19.8	27.8
	VT	49	21.1	2.95	12.7	27.9
	CT	9	18.5	1.87	15.7	21.5
	Total	104	22.0	2.45	12.7	27.9
Spruce	OMaT	5	30.4	1.26	29.3	32.4
	OMT	59	28.6	3.28	17.2	34.7
	MT	97	25.9	3.83	14.8	33.5
	VT	4	20.8	2.44	19.4	24.4
	Total	165	26.9	3.56	14.8	34.7

Table 5. Pairwise comparisons of the average site index H_{100} among the site types classified according to the Cajanderian forest site type system. The p values of the Mann-Whitney U test are shown for each pair of site types. Values smaller than the critical p values obtained from the Bonferroni additive inequality equation (for $\alpha = 0.05$ and six comparisons, $0.05/6 \approx 0.008$) are shown in bold-face.

		OMaT	OMT	MT	VT
Pine	OMT
	MT	.	0.011	.	.
	VT	.	0.002	0.000	.
	CT	.	0.003	0.000	0.008
Spruce	OMaT
	OMT	0.193	.	.	.
	MT	0.005	0.000	.	.
	VT	0.014	0.002	0.011	.

to differences in sample sizes and variances of site quality, a nonparametric method (Mann-Whitney U test) was used for comparing site quality between different site types (Table 5). To control the experimentwise error rate (EER) with repeated tests, the critical p values for the comparisonwise error rate (CER) can be calculated, for example, from the Bonferroni additive inequality equation (see Miller 1981)

$$CER = \alpha/c \quad (21)$$

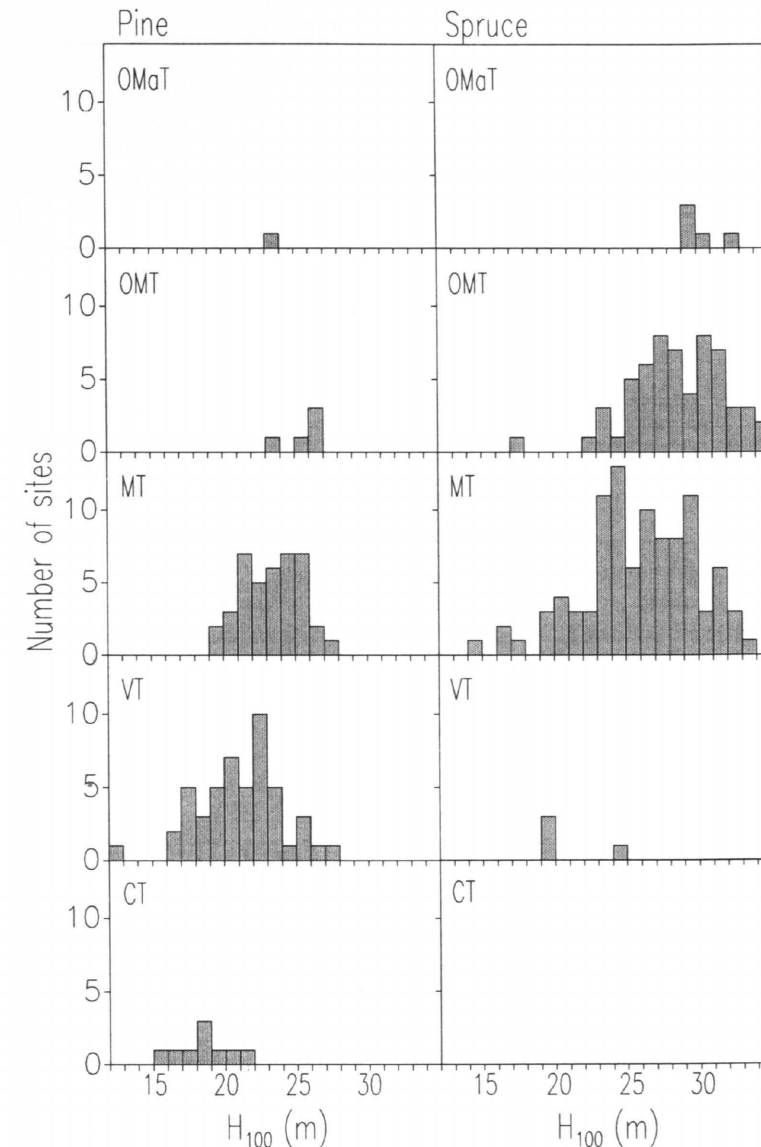


Fig. 11. Frequency distributions of site index H_{100} among different forest site types.

where c is the number of comparisons. For $\alpha = 0.05$ and six comparisons, it gives $0.05/6 = 0.008$. Except for the OMaT vs. OMT pair in spruce forests, average site quality differs among all pairs of sites types. The low number of cases ($n = 5$) in the OMaT site type probably accounts for the only exception. Comparison of the OMaT vs. VT pair in spruce forests is disturbed by low sample sizes in both site type classes.

The pattern of very uneven frequency distribution of sites among different site types is apparent. The two most common site types represent 86 % of the sites in pine-dominated forests, and 95 % of the sites in spruce-dominated forests. Variance in the values of H_{100} seems to be highest in the most common site types of intermediate frequency, whereas the low and high end of the site type sequence are less variable. On the aver-

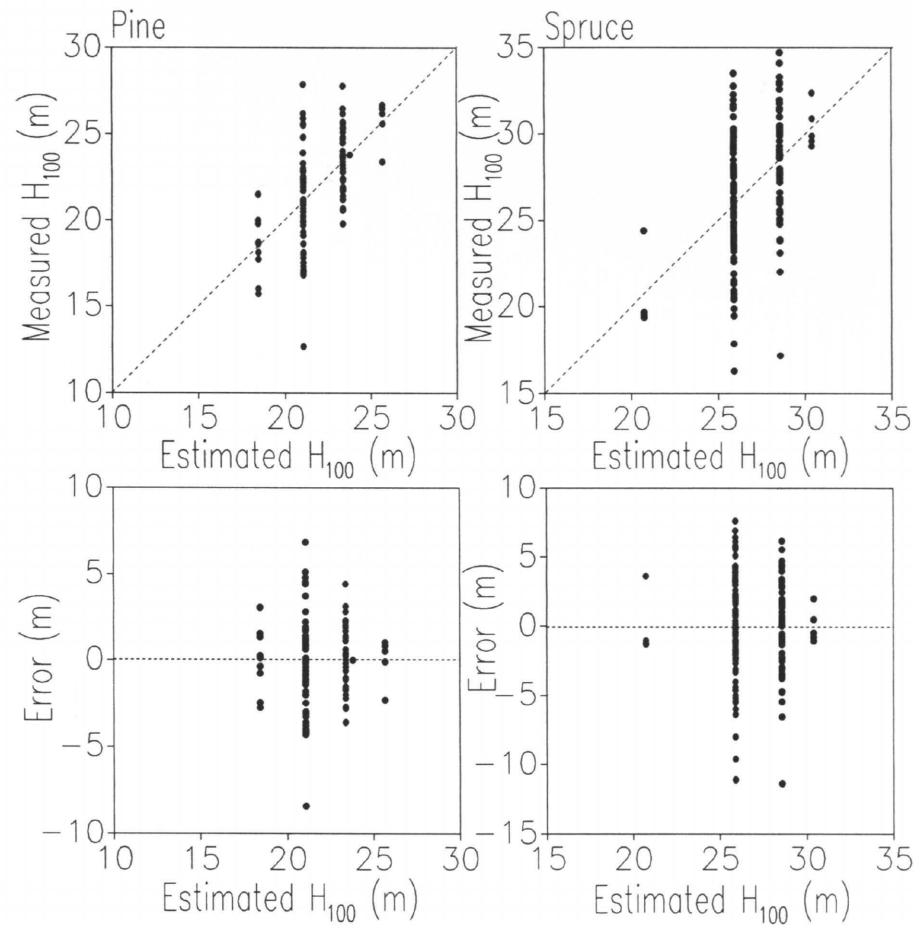


Fig. 12. Measured vs. estimated values of site index H_{100} in the FST model.

age, variance of H_{100} on spruce-dominated sites is higher than that of pine-dominated sites. Fig. 12 shows the scatterplots of measured values of H_{100} and residuals as a function of estimated values.

5.3 Hierarchical Clustering

5.3.1 Theoretical Background

Kuusipalo (1985) used a method for the hierarchical clustering of upland forest sites in southern Finland. Although it is a classification method, HC differs from FST in the sense that it is less subjective. Determination of Cajanderian forest site types depends on the examination of

the general pattern of vegetation at a given site without any strict rules to follow, whereas Kuusipalo (1985) offers rather unambiguous criteria for assigning a given site into one of the predetermined clusters. A diagnostic key, based on the occurrence and abundance of a number of indicator species, gives HC a level of objectivity far beyond that of FST.

The basic method of vegetation analysis in HC is a two-way indicator species analysis (TWINSPAN) developed by Hill (1979a), which results in a dichotomous hierarchical clustering of both samples and species. This clustering is based on the ordination method known as reciprocal averaging (Hill 1973). TWINSPAN classification in its basic form uses qualitative data. The concept

Table 6. The relationship of the clusters of Kuusipalo's (1985) hierarchical classification scheme (columns) to the Cajanderian forest site types (rows) in the data set.

Tree species		1	2	3	4	5	6	Total
Pine	OMaT	1	1
	OMT	4	.	1	.	.	.	5
	MT	1	5	17	16	1	.	40
	VT	.	.	2	10	21	16	49
	CT	1	8	9
Total		6	5	20	26	23	24	104
Spruce	OMaT	5	5
	OMT	46	12	1	.	.	.	59
	MT	9	29	57	2	.	.	97
	VT	.	.	2	.	2	.	4
	Total	60	41	60	2	2	.	165

of "pseudospecies" (Hill et al. 1975), referring to the occurrence of a certain "real" species with abundance above a given threshold value, can be used for entering quantitative information into the analysis.

Due to the hierarchical nature of TWINSPAN classifications, there is a choice of the level of division that is considered practical. Kuusipalo (1985) adopted a rather conservative approach by choosing six site type classes for further analysis. The clustering procedure was performed in two steps. In the first step, those sites that differed most from the rest of the material in the TWINSPAN classification of the whole data set were removed to form a separate cluster. In the second step, the remaining data set was reclassified with TWINSPAN, and the resulting hierarchical classification was taken as a basis for the identification of clusters. The first two levels of division were used to obtain four more clusters, and then the cluster representing the most productive sites in the truncated data set was divided into two clusters at the third hierarchical level. Further analysis of the results indicated that there was correspondence between the clusters obtained with HC and the site type classes of FST that have a more heuristic basis. Table 6 shows the relationship of Kuusipalo's clusters to the Cajanderian forest site types in the data set of the present study.

From a theoretical viewpoint, the hierarchical clustering method of Kuusipalo (1985) is on a more solid basis than Cajander's forest site type

system. It is based on a formal classification method (TWINSPAN) that produces, given the same data set, identical results independent of the user. The choice of the six clusters in the final classification system presented by Kuusipalo (1985) is somewhat more subjective, but in the present context it can be seen as an attempt to produce a classification system that is to some degree comparable to the Cajanderian forest site type system. This viewpoint gains support from Kuusipalo's attempt to find parallel classes between the site types of FST and the clusters in his own system.

The ability of the two-way indicator species analysis and reciprocal averaging, the ordination method behind TWINSPAN (Hill 1973, Hill et al. 1975), to describe the compositional variation of forest vegetation is naturally a crucial requirement for the validity of the HC model. Studies with simulated data sets have shown that in certain conditions the performance of TWINSPAN is less than satisfactory. For example, van Groenewoud (1992) criticized the robustness of TWINSPAN, especially in conditions where there are more than one principal axis of compositional variation in the data set.

The failure of the two-way indicator species method to recover compositional patterns in certain situations with simulated data sets implies that it should be used with a reasonable level of caution in the analysis of real data sets. In the context of the forest vegetation of southern Finland, several studies (e.g. Kuusipalo 1985, Lahti

and Väisänen 1987, Tonteri et al. 1990b) have shown that there is one very strong compositional gradient of understorey vegetation that is almost parallel to the site quality gradient. This implies that one of the main reasons for the failure of TWINSpan with simulated data sets, i.e. presence of several equally strong ecological gradients, is probably not a serious concern in the analysis of the upland forest vegetation of southern Finland.

Empirical results presented by Kuusipalo (1985) showed that TWINSpan produced a satisfactory and easily interpretable classification of the study material. The need to perform the analysis in two steps resulted mainly from the very high floristic diversity in the lush deciduous forests with very high site quality that represent only a very small portion of the total forested land area. In the second step, the more typical forest vegetation of the truncated data set was reanalysed with TWINSpan to produce the hierarchical system for the final identification of operational clusters.

Fig. 13 shows the diagnostic key of Kuusipalo (1985) that can be used for assigning sample sites with known vegetation composition into the pre-defined clusters. It is no surprise that several of the species that appear in the key are also listed as the indicator species for the FST (cf. Kalliola 1973, Mikola 1982). However, use of the diagnostic key of Fig. 13 is less subjective than application of the indicator species list of FST in the sense that with an identical description of the species composition of the site, two independent observers should result in the same cluster assignment. Within the classification scheme of the FST, this result cannot be guaranteed. For use in the field, the diagnostic key makes the HC model almost as operational as the FST. The extra work that results from the need to use the key for finding the correct cluster from the vegetation data is compensated by the degree of accuracy and objectivity over the FST approach. However, abundance data of the vegetation needed to determine the presence of the pseudospecies in the diagnostic key have a subjective component due to the inaccuracy of visual estimation of percentage cover of plant species. There are also systematic biases in the cover estimates among observers (Tonteri 1990).

Kuusipalo (1985) analysed also the variation of site index (obtained from the sample tree data)

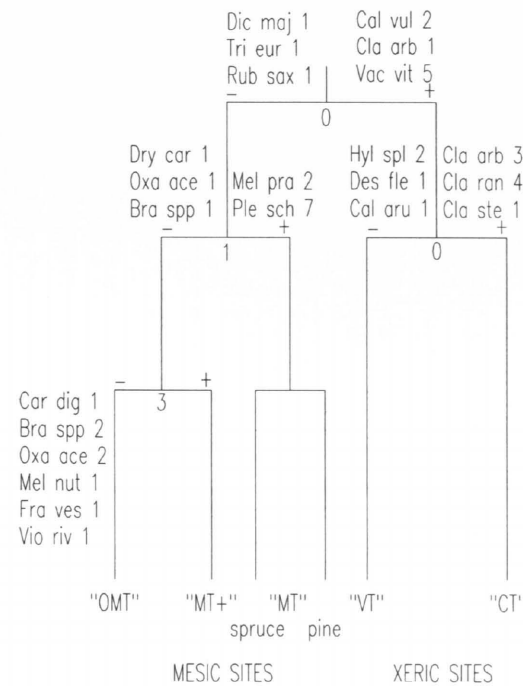


Fig. 13. The diagnostic key of Kuusipalo (1985) for assigning a site with a known species composition into one of the predetermined clusters. The number after each indicator taxon refers to the limit abundance as follows: 1 = present here and there occasionally, 2 = present more than occasionally, but with less than 1 % total coverage, 3 = total coverage 1–5 %, 4 = 5–10 %, 5 = 10–16 %, 6 = 16–25 %, 7 = 25–50 %, 8 = over 50 %. Key to taxa: *Bra spp* = *Brachythecium* spp., *Cal aru* = *Calamagrostis arundinacea*, *Cal vul* = *Calluna vulgaris*, *Car dig* = *Carex digitata*, *Cla arb* = *Cladonia arbuscula*, *Cla ran* = *C. rangiferina*, *Cla ste* = *C. stellaris*, *Des fle* = *Deschampsia flexuosa*, *Dic maj* = *Dicranum majus*, *Dry car* = *Dryopteris carthusiana*, *Fra ves* = *Fragaria vesca*, *Hyl spl* = *Hylocomium splendens*, *Mel nut* = *Melica nutans*, *Mel pra* = *Melampyrum pratense*, *Oxa ace* = *Oxalis acetosella*, *Ple sch* = *Pleurozium schreberi*, *Rub sax* = *Rubus saxatilis*, *Tri eur* = *Trientalis europaea*, *Vac vit* = *Vaccinium vitis-idaea*, *Vio riv* = *Viola riviniana*.

among the clusters, separately for pine- and spruce-dominated sites. The results indicated that some of the clusters were not significantly differ-

ent from each other with respect to the average site quality.

Although TWINSpan classification, the basis of HC, depends on reciprocal averaging, a quantitative ordination method, the resulting clustering of forest sites on the basis of vegetational similarity is still clearly a classification method. In fact, Kuusipalo (1985) suspected that attempts to develop a continuous index of site quality on the basis of vegetation characteristics are probably bound to fail for several reasons.

HC is based on a hierarchical clustering procedure, but the principles for using it as a method for site quality estimation are very similar to those of FST. The clusters selected for the operational classification of sites may represent different levels of binary divisions in the original TWINSpan analysis. As mentioned above, this is also the case in the diagnostic key presented by Kuusipalo (1985). However, from the viewpoint of site quality estimation the average site quality within each cluster in the training set is again an unbiased estimator of q . Overall performance of the model depends on the same factors as in the case of FST.

From the viewpoint of the present study it is worth noticing that the data set used for comparing the different estimation methods is exactly the same one that was used by Kuusipalo (1985) for developing the hierarchical clustering method. This fact offered some convenient shortcuts for the analysis of HC itself. For example, the assignment of sample sites into different clusters was obtained directly from the Appendix of Kuusipalo (1985), and some of the test statistics of interest in the present context were also available already in his study.

5.3.2 Empirical Results

Results from the application of the HC model for the estimation of site quality in the empirical data set are given in Table 7. The frequency distribution of sample sites among the clusters is more even than in the case of the FST model. As with the FST, the clusters form a sequence of increasing average site quality (Fig. 14). Statistical tests for the differences in average site quality between the clusters have been already published by Kuusipalo (1985: Table 10). His re-

Table 7. Statistics of the site index estimates according to Kuusipalo's (1985) hierarchical clustering method (HC). All values are in meters.

Tree species	Cluster	n	Mean	s.d.	Min	Max
Pine	1	6	25.6	1.13	23.8	26.7
	2	5	23.5	1.87	20.6	25.4
	3	20	23.4	1.93	19.8	26.5
	4	26	23.1	2.79	16.8	27.9
	5	23	20.5	2.56	15.7	26.2
	6	24	19.9	2.64	12.7	23.9
	Total	104	22.0	2.44	12.7	27.9
Spruce	1	60	28.9	3.32	17.2	34.7
	2	41	27.5	3.12	20.9	32.8
	3	60	24.8	3.83	14.8	31.7
	4	2	24.2	0.99	23.5	24.9
	5	2	19.5	0.07	19.4	19.5
	Total	165	26.9	3.43	14.8	34.7

sults showed that all clusters among spruce-dominated stands were different at the 5 % risk level, whereas among pine-dominated stands the distributions of site index were more blurred. The clusters formed two groups (1–4 vs. 5–6) that differed from each other, but within each group the site index between clusters was not significantly different (except cluster 1 vs. cluster 4). Scatterplots of measured vs. predicted values (Fig. 15) and the estimates of standard deviation (Table 7) indicate that the values of standard deviation for the whole data set are approximately the same magnitude as in the case of the FST model. Variance of the site quality has a tendency to increase with the number of sample sites in the cluster, with some deviations from the general trend (e.g. between the pine-dominated clusters 2 and 3).

5.4 Multivariate Ordination and Regression

5.4.1 Theoretical Background

In the hierarchical clustering (HC) model based on TWINSpan classification, the underlying method is reciprocal averaging (RA), a quantita-

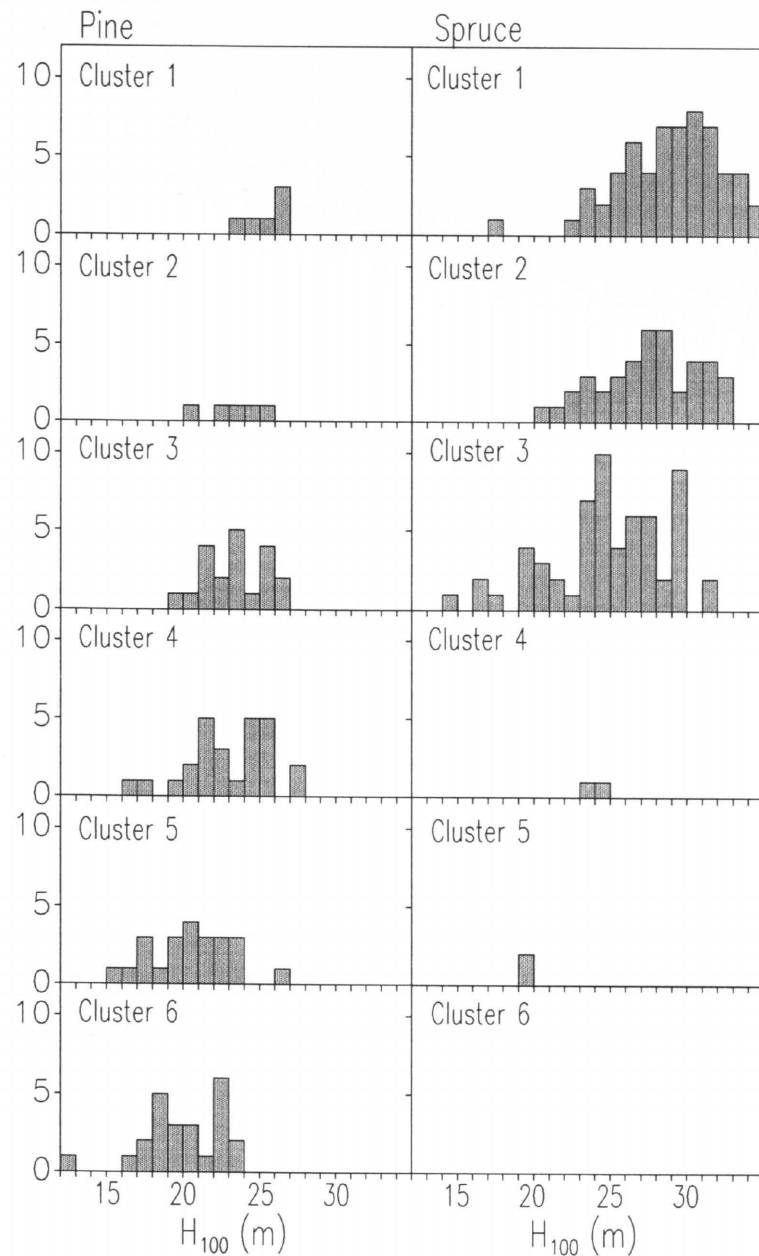


Fig. 14. Frequency distributions of site index H_{100} among different clusters in Kuusipalo's (1985) hierarchical clustering method.

tive ordination procedure. RA, like a number of other ordination methods, is able to describe continuous variation in vegetation composition in a multidimensional space, where the most important ordination axes can be used for detecting

ecological patterns (for a review of different methods, see Orłóci 1978, Gauch 1982, Pielou 1984). Since site quality q , according to the definitions given before, is also assumed to be a continuous variable, it is possible to analyse the

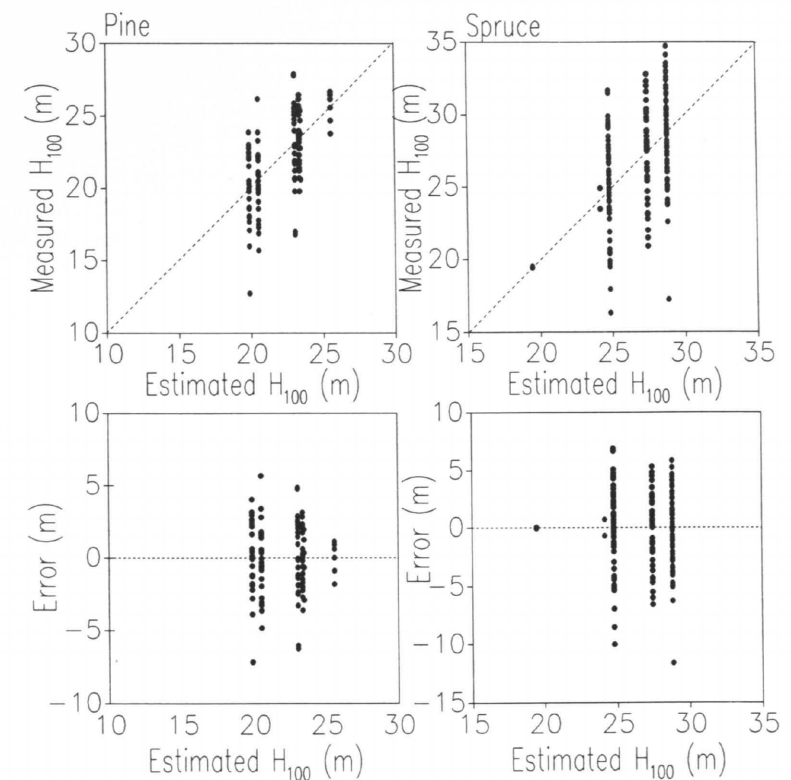


Fig. 15. Measured vs. estimated values of site index H_{100} in the HC model.

covariation between site quality and vegetation composition with quantitative methods, such as regression analysis.

There are a choice of different ordination methods available for this purpose. Some general-purpose multivariate methods, such as principal components analysis (PCA), are generally considered to be problematic in the analysis of ecological data with strong nonlinearities (Gauch 1982). The methods developed for ecological purposes have tried to overcome these problems with ingenious computational procedures, but so far there has not been any comprehensive solution to result in a general consensus among the ecologists.

One of the difficult nonlinearities in ecological data sets results from the unimodal, "bell-shaped" responses of species to ecological gradients. In theoretical models this response is often assumed to be symmetrical, although there is empirical evidence to refute this assumption as a universal

rule (e.g. Austin 1987). Recent developments in the field of nonparametric methods (Yee and Mitchell 1991) make it possible to use statistical procedures for detecting also the form of species responses to the gradients.

In the present study I have used detrended correspondence analysis (DCA: Hill 1979b, Hill and Gauch 1980), an ordination method that has become very popular in the recent years, although it has also received a fair amount of criticism. The most problematic feature of DCA is known as "detrending", manipulation of the ordination axes for the purpose of correcting distortions resulting from nonlinearities in the data. The critics (e.g. Minchin 1987, Wartenberg et al. 1987, Oksanen 1987, 1988, Jackson and Somers 1991, van Groenewoud 1992) have often drawn attention to simulated data sets that make DCA to fail in the analysis, whereas proponents of the method put more emphasis on good results obtained with real

data sets (Peet et al. 1988). The original implementation of DCA in a computer program known as DECORANA (Hill 1979b) has recently been largely replaced by a more flexible and comprehensive implementation known as CANOCO (ter Braak 1988).

Ordination of vegetation data is based on the species-by-sites matrix, where each matrix cell refers to the abundance of a given species at a given site. Before the ordination analysis, it is possible to use various transformations to the abundance data (Maarel 1979). It is also possible to perform an ordination analysis to mere presence-absence data, where occurrences are denoted by ones and absences by zeros in the data matrix. In this case, no transformations are necessary.

Application of the resulting ordination space for the estimation of site quality is based on a training set, in which there are a number of sample sites with known site quality and composition of understorey vegetation. Since the axes of the ordination space are orthogonal, model parameters can be estimated by calculating a regression line for site quality as a function of one or more ordination axes. In the present study I fitted only a planar regression surface into the data set, but in principle it is also possible to fit more complex response surfaces of site quality into the ordination space. This can be done, for example, with generalized linear models (McCullagh and Nelder 1989) or generalized additive models (Hastie and Tibshirani 1990).

The resulting regression equation can then be used for estimating site quality at new sites with known species composition of understorey vegetation. In order to find the estimate, we must first find the location of the new sites in the ordination space. There are methods for estimating the location of a new site within a predetermined ordination space, but if there are several sites to be processed simultaneously, it is a straightforward solution to perform a new ordination of the study set together with the training set. In this way both sets are drawn together into a common ordination space, and a regression line fitted for the sites of the training set can be used for estimating the unknown values of site quality in the study set.

For studying the relationship of vegetation composition to environmental variables, multivariate ordination and linear regression together have been

applied in several previous studies of Finnish forest vegetation (e.g. Kuusipalo 1985, Tonteri et al. 1990a, 1990b, Nieppola and Carleton 1991). However, none of them has treated the material from the viewpoint of accuracy in the estimation of site quality. Theoretically, there are no reasons to assume that the axes of the ordination space are parallel with environmental gradients, but there is plenty of evidence that the principal axis of ordination in the forest vegetation of Finland is usually almost parallel with the site quality gradient. Therefore it is possible to build a regression model as a function of the first ordination axis only without losing much information.

5.4.2 Empirical Results

DCA ordination was performed both for the quantitative data and the presence-absence data. For both of these cases, two regression models were fitted, one for the main axis (DCA1), and another for the combination of the first four axes (DCA1–4), in order to obtain four different estimates of site quality. Since all these methods require computerized processing of the data, there is not very much difference between them in terms of computational efficiency. Results of the DCA ordinations with different options are given in Table 8. The resulting ordination spaces with DCA axes 1 and 2 are shown in Fig. 16. As expected, the first axis is considerably stronger

Table 8. Results of the DCA ordination of the sample data. The upper value in each cell is the eigenvalue of the axis, and the lower value is the length of the compositional gradient.

Tree species	Model data	DCA1	DCA2	DCA3	DCA4
Pine	Abundance	0.385 3.188	0.136 2.022	0.103 1.876	0.077 1.343
	Incidence	0.395 3.336	0.151 2.056	0.110 1.834	0.094 1.968
Spruce	Abundance	0.235 2.720	0.101 1.677	0.077 1.914	0.056 1.308
	Incidence	0.229 2.825	0.108 1.717	0.092 1.764	0.068 1.335

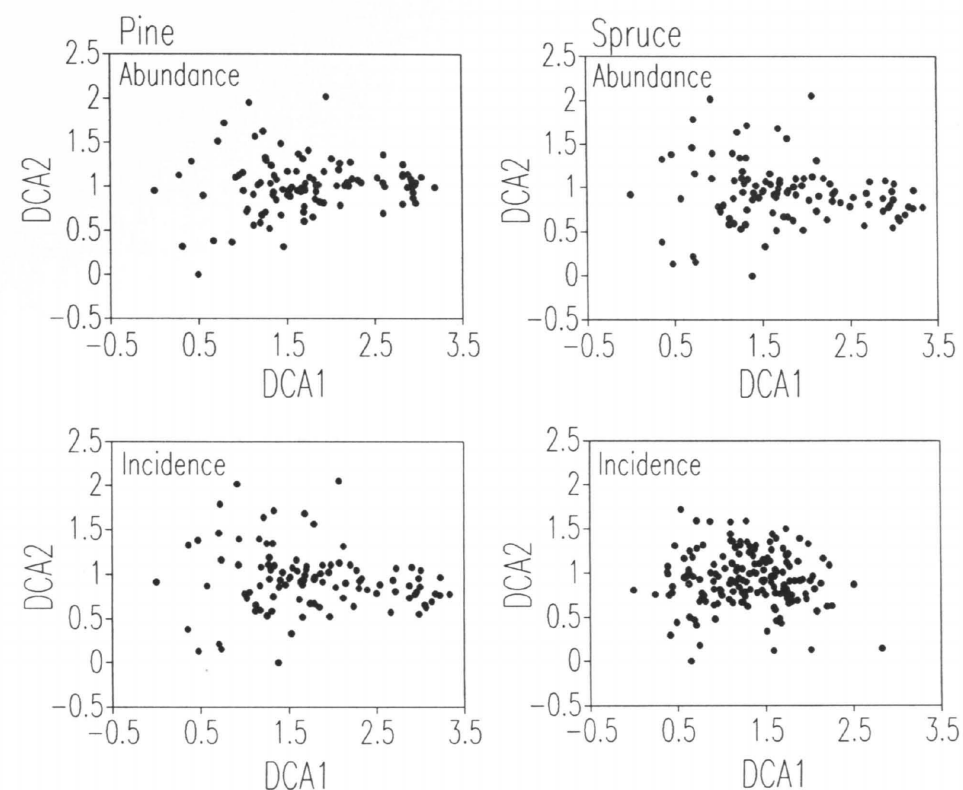


Fig. 16. The first two axes of the DCA ordination of the data set.

than the subsequent axes 2–4 (especially in pine-dominated forests), indicating the strong unidimensionality of the compositional variation in vegetation (cf. Kuusipalo 1985).

Site index H_{100} as a function of the first DCA axis is shown in Fig. 17. Notice that there may be either positive or negative correlation between site index and the ordination axes depending on the final orientation of the compositional space. For the purposes of site quality estimation this phenomenon can be ignored, and only the linearity of the relationship is interesting.

Parameters describing the goodness-of-fit between site index and the DCA axes are given in Table 9. R^2 denotes the proportion of variance in site index “explained” by the regression equation. RMSE (without the subscript) denotes the square root of MSE, the mean squared error calculated as

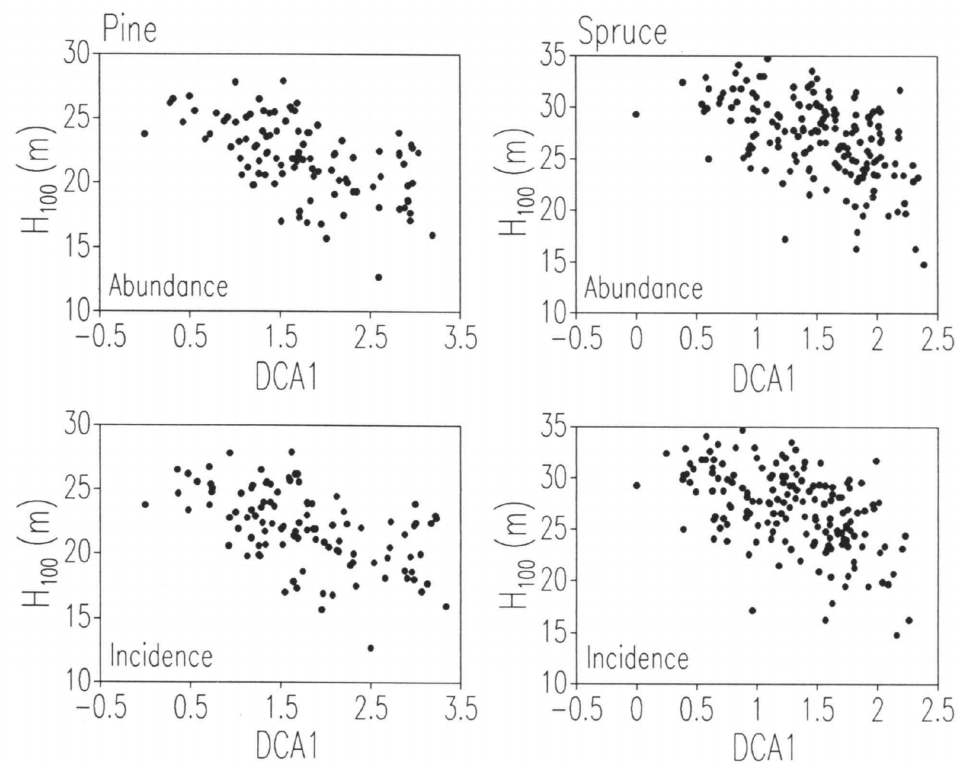
$$MSE = SSE/(n - p) \quad (22)$$

where SSE is the error sum of squares, n is the number of observations, and p is the number of parameters including the intercept. A cross-validation estimate of the same parameter, $RMSE_{cv}$, is also given in the table. Fig. 18 shows scatterplots of the observed vs. estimated values and residuals for the versions that performed best (DCA1 with abundance data for pine-dominated forests, DCA1–4 with abundance data for spruce-dominated forests).

As can be seen from the values of Table 9, all alternatives of the method produce estimates that are of the same magnitude in terms of $RMSE_{cv}$. There does not even seem to be any unambiguous difference between abundance data and presence-absence data, although the estimates based on abundance data give slightly better estimates in the majority of cases.

Table 9. Statistics of the site index estimates in the multivariate ordination and regression method (MO). Error values are given in meters.

Tree species	Model	Model data	n	R ²	RMSE	RMSE _{cv}
Pine	DCA1	Abundance	104	0.30	2.49	2.51
		Incidence	104	0.28	2.53	2.55
	DCA1-4	Abundance	104	0.31	2.50	2.56
		Incidence	104	0.30	2.53	2.61
Spruce	DCA1	Abundance	165	0.29	3.33	3.35
		Incidence	165	0.29	3.32	3.33
	DCA1-4	Abundance	165	0.41	3.06	3.13
		Incidence	165	0.37	3.15	3.20

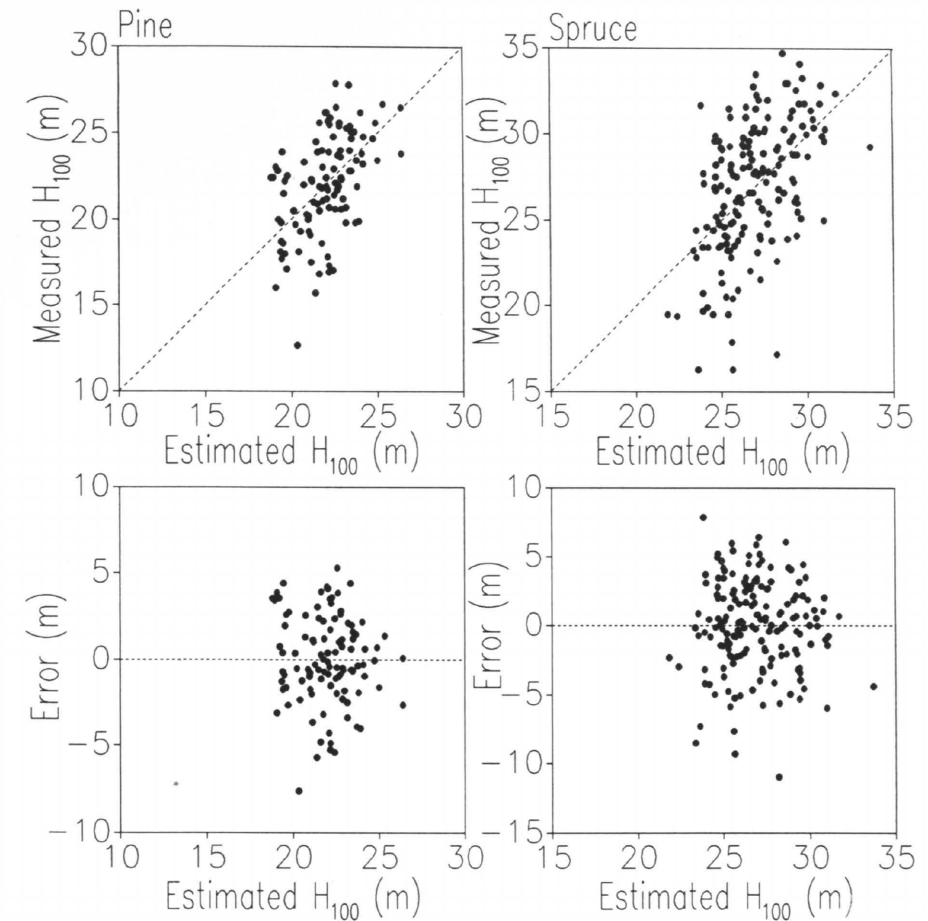
**Fig. 17.** Site index H_{100} as a function of the first ordination axis.

5.5 Weighted Averaging

5.5.1 Theoretical Background

All the other alternative estimation methods described so far, the classification methods FST and HC, as well as the ordination method MO,

process the vegetation data independently of the values of the site variables. This implies that none of them is optimized for the purpose of estimating site quality on the basis of understorey vegetation. The fourth model, known as weighted averaging (WA), is a simple method for calculating estimates of site variables on the

**Fig. 18.** Measured vs. estimated values of site index H_{100} in the version of the MO model that performed best in terms of RMSE_{cv}. For pine-dominated forests it was the regression based on the first axis with abundance data, whereas for spruce-dominated forests the regression based on DCA axes 1-4 with abundance data was selected.

basis of the indicator values of species occurring at the study site. The principles of WA have been used in several studies (Goff and Cottam 1967, Hill 1973, Ellenberg 1979, 1982, 1988, Gasse and Tekaiia 1983), but especially the recent studies by ter Braak have increased interest towards the method (e.g. ter Braak and Looman 1986, ter Braak and Barendregt 1986, ter Braak and Gremmen 1987, ter Braak and Prentice 1988, ter Braak and van Dam 1989, Birks et al. 1990).

WA as an estimation method is based on the calculation of species-specific indicator values and ecological amplitudes (also known as tolerances)

from empirical data (the "training set"). For presence-absence data, the indicator value of a given species i with respect to a certain environmental variable is simply an average of the values of the variable at those sites where species i occurs. If quantitative data are available, the average can be weighted by abundance at each site. Tolerance, on the other hand, is analogous to standard deviation of the values of the environmental variable at the sites of the occurrence of species i (see ter Braak and Looman 1986). The actual estimate of an unknown variable is obtained by averaging the indicator values of the species occurring at the

site. Tolerance downweighting (see Birks et al. 1990) can be used for putting more emphasis on the species with narrow ecological amplitudes.

The equations needed for the calculation of WA estimates are as follows (cf. Birks et al. 1990). Let \hat{q}_i and \hat{t}_i denote the estimated indicator value and tolerance of species i , respectively. Furthermore, let $a_i(s)$ denote the abundance of species i at site s . Then, the indicator value of species i is

$$\hat{q}_i = \frac{\sum_{s=1}^m a_i(s) q(s)}{\sum_{s=1}^m a_i(s)} \quad (23)$$

and its tolerance is

$$\hat{t}_i = \sqrt{\frac{\sum_{s=1}^m a_i(s) (q(s) - \hat{q}_i)^2}{\sum_{s=1}^m a_i(s)}} \quad (24)$$

where m is the number of sites. The estimate of $q(s)$, site quality at site s , is

$$\hat{q}(s) = \frac{\sum_{i=1}^n a_i(s) \hat{q}_i}{\sum_{i=1}^n a_i(s) \hat{t}_i^2} \quad (25)$$

with tolerance downweighting, and

$$\hat{q}(s) = \frac{\sum_{i=1}^n a_i(s) \hat{q}_i}{\sum_{i=1}^n a_i(s)} \quad (26)$$

without it.

The averaging procedure in WA, performed once in calibration and once in regression, results in shrinkage of the range of the estimates. A simple method to correct for this phenomenon is to calculate a linear regression between the estimated and observed values in the training set (Birks et al. 1990).

Weighted averaging is a heuristic method, and it can be shown to produce seriously biased results in certain conditions (ter Braak and Looman 1986). Its performance has been compared with a more formal statistical approach, based on the description of species responses to ecological gradients with Gaussian logistic regression (ter Braak and Looman 1986, ter Braak and van Dam 1989) and estimation of site variables with maximum

likelihood estimation (MLE). This approach has been used for example in limnology for the reconstruction of water pH from the diatom assemblages (e.g. ter Braak and van Dam 1989, Oksanen et al. 1988, 1990, Birks et al. 1990). Although it is a more sophisticated method than WA from a purely theoretical viewpoint, empirical results are somewhat contradictory. For example, ter Braak and Looman (1986) referred to empirical data sets used for predicting soil moisture where WA was unreliable for estimating indicator values of the species. On the other hand, Birks et al. (1990) compared WA and MLE as methods for reconstructing water pH from the diatom assemblages. Their conclusion was that WA gave better results than MLE in terms of lowest root mean squared error of prediction in cross-validation.

These results suggest that WA should be used with caution. The advantages of WA are computational simplicity and good performance with suitable data sets. The reasons for the failure of WA are rather well understood, thanks to the simplicity of the approach. If WA fails, a more formal statistical approach can be adopted.

In comparison to the other alternative estimation methods in the present study, WA has one essential property that makes it, at least in theory, superior to its competitors. It is the only method, in addition to PPE, that directly utilizes the covariation between vegetation composition and site quality for the calculation of the estimates. Whereas the other methods can perform the first stage of analysis without any information on site quality at the sites in the training set, WA obtains its predictive power directly from the site quality estimates. In this sense, WA is theoretically superior to the other methods. Whether this theoretical advantage has any practical significance must be determined empirically.

In terms of applicability in the field, WA requires more work than FST or HC but considerably less than PPE or DCA. The simplest variants of WA estimates can be calculated easily, but if tolerance downweighting is required, computational requirements are somewhat more demanding.

In the present study, only WA was used for evaluating its performance in the estimation of site quality in the forests. A total of four variants of WA were applied: pure presence-absence data

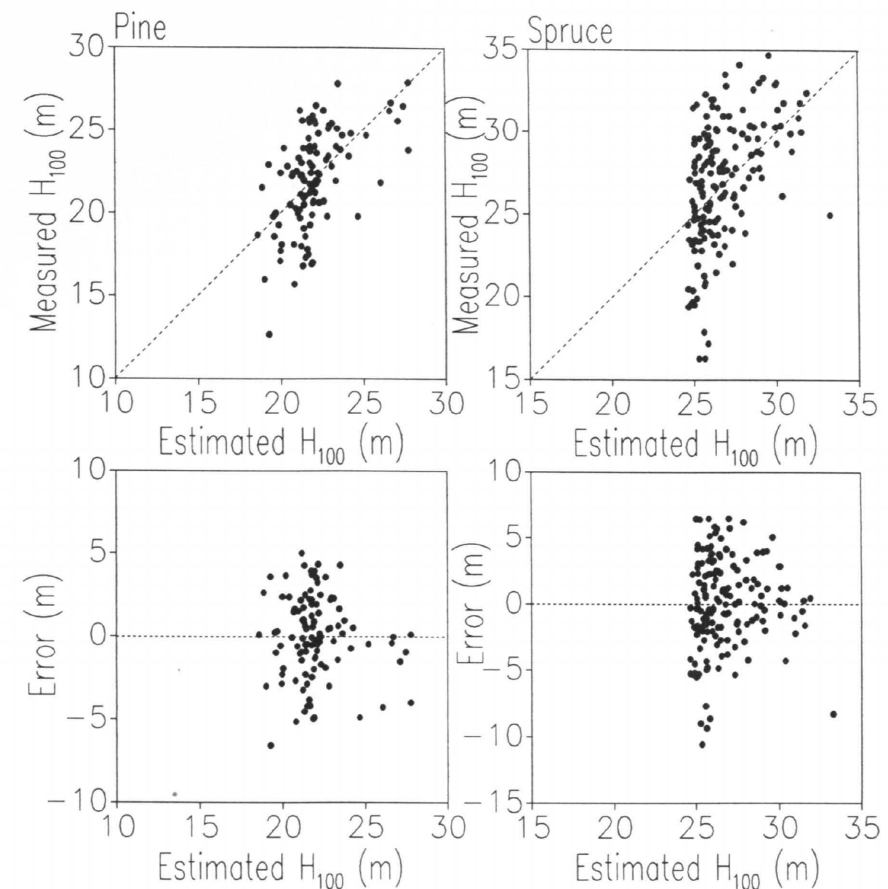


Fig. 19. Measured vs. estimated values of site index H_{100} using the version of the WA model that performed best in terms of $RMSE_{cv}$. For both tree species, it was the one with abundance data and tolerance downweighting.

Table 10. Statistics of the site index estimates in the weighted averaging method (WA). Error values are given in meters.

Tree species	Model data	Tol dw	n	r^2	RMSE	$RMSE_{cv}$
Pine	Abundance	No	104	0.39	2.33	2.48
		Yes	104	0.41	2.29	2.47
	Incidence	No	104	0.31	2.47	2.85
		Yes	104	0.43	2.26	2.82
Spruce	Abundance	No	165	0.33	3.24	3.46
		Yes	165	0.32	3.25	3.43
	Incidence	No	165	0.32	3.25	4.04
		Yes	165	0.37	3.13	3.45

with and without tolerance downweighting, and abundance data with the same options.

5.5.2 Empirical Results

Performance of the different versions of WA in the analysis of the empirical data set are shown in Table 10. There seems to be a consistent trend that estimates with abundance data and tolerance downweighting are slightly better than those with incidence data without tolerance downweighting. Fig. 19 shows the scatterplots of measured vs. predicted values and residuals for the WA estimates with abundance data and tolerance downweighting.

5.6 Comparison of the Methods

In the previous chapters, each of the estimation methods have been described and applied to the empirical data set independently of the others. Next I performed a side-by-side comparison of the methods, in order to assess the relative merits of each estimation procedure. In order to keep the number of different methods at a reasonable level, only one version of PPE, MO and WA were included into this comparison.

It was observed that there were no significant, order-of-magnitude differences between the various versions of these methods. Therefore, it was considered reasonable to select the version that performed best in the original comparison to be put against the other estimation methods. For PPE, a complete model (both presence and absence data included) with incidence data was chosen for pine-dominated forests, whereas for spruce-dominated forests the corresponding partial model was selected. In the case of MO, the selected version was the regression based on the first DCA axis with abundance data for pine-dominated forests, and a multiple regression based on DCA axes 1–4 with abundance data for spruce-dominated forests. For WA, the version that uses abundance data and tolerance downweighting was the choice for both tree species. For FST and HC, the value of $RMSE_{cv}$ used in the comparison was calculated from the standard deviation according to Eq. (8).

In this way, only five estimation methods were chosen for the final comparison. Against the actual estimation methods, a suitable null model for the use of understorey vegetation as an indicator of site quality was to use the sample mean as an estimate of site quality. If there is any information on site quality available in the composition of understorey vegetation, the estimation methods must give significantly better estimates than the null model that ignores vegetation information. Also in this case, Eq. (8) was applied to obtain a comparable value for $RMSE_{cv}$.

The theoretical basis of the separate estimation methods has already been assessed. Although there is no objective way to rank the methods on the basis of theoretical pros and cons, some general trends are apparent. As models for processing vegetation data, PPE, HC and DCA utilize well-defined statistical methods. WA operates at a con-

siderably less advanced level of mathematics, but on the other hand it is the only method, in addition to PPE, that optimizes its estimates directly for the target. FST, having its historical burden, is the most subjective method; it does not even require sample plot data to be used in the field.

It is not surprising that applicability of the different methods in the field is approximately inversely related to the consistency of the theoretical basis. The most notable exception to this pattern is HC that is based on statistical multivariate methods, but is nevertheless directly applicable in the field due to the published diagnostic key for the identification of the clusters (Kuusipalo 1985).

The most interesting feature in the comparison of the methods is their accuracy for the estimation of site quality on the basis of understorey vegetation composition. Statistics for this comparison are given in Tables 11 and 12. Table 11 shows the cross-validation RMSE of estimates of site quality for the five methods, as well as the null model of sample mean. For comparing the significance of the differences in the values of $RMSE_{cv}$, pairwise tests were performed. Due to deviations from normality in the sample distributions, a nonparametric signed rank test was applied. Table 12 shows the p values for the differences in accuracy between different methods. To control the EER in this case, Eq. (21) gives $CER = 0.05/15 = 0.0033$ for the 5 % risk level.

The first observation from the examination of Tables 11 and 12 is that all methods gave better estimates of site quality than the null model, i.e.

Table 11. Accuracy (measured by $RMSE_{cv}$ in meters) of different estimation methods in comparison to each other and the null model derived from the sample mean. For PPE, MO and WA, only the version that performed best is shown here. For sample mean, FST and HC, the value of $RMSE_{cv}$ was obtained from the standard deviation by using Eq. (8). This table is a summary of the values in Tables 1,3,4,7,9 and 10.

	Sample	PPE	FST	HC	MO	WA
Pine	2.98	2.34	2.46	2.45	2.51	2.47
Spruce	3.94	3.16	3.57	3.44	3.13	3.43

Table 12. Pairwise comparisons of the predictive accuracy of different estimation methods. The p values for the signed rank test are given for pine-dominated forests above the diagonal, and for spruce-dominated forests below the diagonal. Values smaller than the corrected critical value for the 5 % risk level (0.0033) are shown in boldface.

	Sample	PPE	FST	HC	MO	WA
Sample	.	0.0042	0.0236	0.0138	0.0078	0.0138
PPE	0.0007	.	0.2022	0.0039	0.0138	0.3776
FST	0.0614	0.1377	.	0.0371	0.7688	0.6241
HC	0.0018	0.3100	0.0865	.	0.3776	0.7688
MO	0.0018	0.4364	0.0013	0.0003	.	0.4926
WA	0.0125	0.7556	0.0037	0.8763	0.0004	.

Table 13. Correlations of site-specific residuals between different estimation methods. The values above the diagonal refer to pine-dominated forests, those below the diagonal to spruce-dominated forests.

	Sample	PPE	FST	HC	MO	WA
Sample	.	0.78	0.81	0.81	0.85	0.84
PPE	0.79	.	0.87	0.92	0.94	0.89
FST	0.90	0.86	.	0.89	0.91	0.87
HC	0.87	0.93	0.92	.	0.95	0.88
MO	0.80	0.90	0.87	0.90	.	0.91
WA	0.88	0.84	0.93	0.89	0.91	.

they can all be considered useful for the purpose of estimating site quality. However, the differences were not great enough to be statistically significant in all cases in a data set of this size. Beyond this, the pattern was very homogeneous.

In pine-dominated forests, none of the estimation methods performed significantly better than the other ones. In spruce-dominated forests, the quantitative estimation methods PPE and MO performed slightly better than FST, HC and WA, although only for MO the differences were statistically significant.

In terms of the $RMSE_{cv}$ of site index estimates, improvement obtained by using understorey vegetation as an indicator was not very impressive (Table 11). In pine-dominated forests, $RMSE_{cv}$ dropped from 2.98 m down to 2.3–2.5 m, i.e. the improvement was about 15–20 %. In spruce-dominated forests, the best estimate (with MO) dropped $RMSE_{cv}$ from 3.94 m down to 3.13 m, giving 21 % improvement. With FST, the improvement in $RMSE_{cv}$ was about 9 % (from 3.94 m to 3.57 m). Correlations of residuals between different estimation methods (Table 13) show that accuracy in the estimation of H_{100} was rather site-specific and consistent among the methods.

6 Discussion

6.1 Evaluation of the results

There are several factors that must be taken into account when assessing the general applicability of the results obtained in the present study. "Site quality", defined by Clutter et al. (1983) as a site-specific potential of biomass production, implies that there is no "true" measure of site quality. The actual biomass production of the site is a result of the interaction between site characteristics and properties of the actual tree stand growing at the site.

Site index H_{100} calculated from the height-to-age relationships of dominant trees as an indicator of site quality makes the situation even more complicated. The problems of H_{100} mentioned by Kilkki and Ojansuu (1981), as well as the geographical species turnover of the flora, imply that the results of the present study cannot be generalized directly to other phytogeographical regions.

It is rather surprising to notice that so far the accuracy of the Cajanderian forest site type system as a tool for the estimation of site quality has been analysed quite superficially. The results of Ilvessalo (1920; see also Ilvessalo and Ilvessalo 1975) and Lönnroth (1925) drew attention to the average differences in site quality between different site types, but the problem of estimating site quality at each individual site was largely ignored. The hypothesis of the continuity of spatial variation in site quality implies that all clusters of site types are more or less heterogeneous assemblages, and therefore also the site-specific error in the estimates may vary. The reasons and possible improvements of the site-specific errors in the estimates of site quality are quite difficult to analyse within the scheme of forest site types, where both vegetation composition and stand structure are described as examples of the "ideal" site type.

One of the essential results of the present study was that none of the five estimation methods was considerably better than the others as a tool for

the estimation of site quality on the basis of the understorey vegetation composition. If this observation has any general significance, it implies that the system that is chosen as a basis for the characterization of site quality variation can be determined by other criteria.

A conservative interpretation of the observation mentioned above is that there is no need to find alternatives to the well-established Cajanderian forest site type system that has evolved over several decades to become an integral part of forest research and management in Finland. However, exactly the same result can be used as an argument for a more radical viewpoint: since there are no essential differences between the Cajanderian system and its alternatives in terms of accuracy in site quality estimation, the choice of the most suitable method can be based on other criteria, such as consistency of the theoretical background and the requirements of the field work. It is worth noticing that, from a purely scientific viewpoint, the long tradition of the Cajanderian forest site type system is not a valid argument as such for its use also in the future.

Beyond the goal of site quality estimation, the methods developed in this study for describing the ecological responses of individual plant species to environmental gradients may have some scientific value as such. The most significant difference in relation to previous studies (e.g. ter Braak and Looman 1986) is the application of nonparametric methods for the estimation procedure. This approach has been enabled by recent advancements in the field of nonparametric statistics (see, e.g., Silverman 1986, Hastie and Tibshirani 1990, Izenman 1991, Yee and Mitchell 1991). The concept of "indicator species" deserves a more formal definition so that analytical methods can be applied to empirical data to elucidate differences between species. The approach described in the present study is a step in this direction.

6.2 Classification vs. Estimation

Classification of real world objects into distinguishable entities (classes) is one of the fundamental methods of organization applied by people in their thinking. Classification of forest sites may thus be viewed as a "natural" approach for describing empirically observed variation in site characteristics. Each site is assigned to one of the site type classes on the basis of the correspondence of local site features with the description of (ideal) class characteristics. Thus, although no two sites are identical, there is a degree of similarity between the members of the same class that is assumed to be higher than the degree of similarity between the members of different classes.

When using a classificatory scheme for the estimation of site quality, class means must be used as site quality estimates for each class. This limitation results from the fact that, according to theory of classification, all members within each class are, by definition, considered identical. For assessment of the accuracy of estimates, the most interesting parameter is within-class variance in the variable of interest.

Quantitative estimation methods, on the other hand, do not have this kind of limitation. A quantitative method produces estimates on a continuous scale. Hence a measure of accuracy is the difference between the true value and its estimate for the target variable. The principles that are used for obtaining estimates may vary considerably, but as soon as an estimate is available, we can utilize a well-established set of statistical tools for describing its behaviour.

The continuum view of vegetation forms the theoretical basis of modern ordination methods. Nowadays, classification and ordination methods are usually considered to be complementary approaches to the analysis of vegetation data (cf. Gauch 1982). Some modern classification methods, for example the two-way indicator species analysis (TWINSPAN; Hill 1979a), are closely associated with a corresponding ordination method.

During the last few years, a number of analyses of the forest vegetation of Finland based on multivariate methods have been published (for a review, see Oksanen 1990). Oksanen (1984), Kuusi-

palo (1985) and Sepponen (1985) analysed different approaches to ecological site classification on the basis of vegetation and soil characteristics and discussed also the principles of Cajander's forest site type system. A direct evaluation of the Cajanderian forest site type system was performed by Lahti and Väisänen (1987), and further contributions to this topic include the works of Tonteri et al. (1990a, 1990b), Mäkirinta (1990) and Nieppola and Carleton (1991).

The continuous nature of vegetation variation is noted in several of these studies. However, the proposed alternatives to the classical Cajanderian site type system are various classification procedures. For example, Kuusipalo (1985) concluded that "it seems that many advantages are gained by applying ordination-based hierarchical classification instead of direct classification such as the traditional Cajanderian one". He also drew attention to the problems of classification: "One disadvantage of all classification methods is that they can only be used to assign an individual forest stand to one of the few rough categories of vegetation response to environmental gradients. It is not possible to define the actual location of an individual stand along a gradient". However, one of his concluding remarks was that "in forestry, site types are primarily used for the estimation of productive capacity. In this, a continuous index would, of course, be highly desirable. However, attempts to develop such an index on the basis of vegetation characteristics are probably bound to fail for several reasons". What these reasons might be remained rather obscure.

Nieppola and Carleton (1991) paid attention to similar arguments: "Another approach is to correlate tree growth directly with continuous vegetation data as expressed by ordination axes. An advantage of this method is that no prior site type classification is required. In addition to avoiding the pitfalls of misclassification and differences in the homogeneity of site classes, the continuous approach offers greater accuracy in assessing the predictive capabilities of understorey vegetation." Rather surprisingly, Nieppola and Carleton mentioned that their analysis of the relationship between site quality and vegetation composition serves the purpose of forest site classification, although they explicitly mention that "the notion of discrete types is not supported".

6.3 General Principles for the Characterization of Forest Sites

Given the criticisms of the Cajanderian forest site type system expressed in this and several other studies, it is fair to assume that there are some concrete alternatives to offer. The basic tenet of the Cajanderian system, covariation between site quality and the composition of understorey vegetation, has been empirically corroborated beyond dispute. Therefore, alternative approaches to the characterization of forest sites differ from each other above this level that has gained mutual agreement.

As the results of this study indicated, PPE and the other quantitative estimation methods (MO and WA) seem to be at least as good as the classification methods for the description of spatial variation of site quality in the upland forests of southern Finland (the applicability of this conclusion in other kinds of forests remains to be studied). The opposite side of the coin, however, was that there did not seem to be any significant, order-of-magnitude advantage to be obtained from the application of quantitative estimation methods (see Table 11) that would make the adoption of an alternative paradigm more attractive. This result may have some general significance as a measure of the maximum amount of covariation between vegetation composition and site quality.

The key concept in all systems for the characterization of different forest sites is similarity. Classification methods, as well as estimation methods, are attempts to describe similarity between different sites using techniques that have been derived from the basic philosophy of the method. Similarity as a concept, however, does not have any semantic content in isolation. Two sites may be more or less similar to each other with respect to species composition of the vegetation, age and size of the tree stand, spatial and temporal pattern of abiotic factors, as well as several other characteristics of interest.

The fundamental problem of the Cajanderian forest site type system is that it compresses this multidimensional variation of site characteristics into a unidimensional site type sequence (for further discussions of the dimensionality of the Cajanderian system, see Kuusipalo 1985). The goal of modern exploratory multivariate analysis meth-

ods is to detect patterns of covariation in data sets that can be used to reduce the dimensionality of the description of the research object, but the number of axes needed for a satisfactory description of the data is determined after the analysis, not before it as happens by requiring unidimensionality.

It is worth noticing that this multidimensional view of variation in the site characteristics includes as a special case the unidimensional site type sequence expressed in the Cajanderian system. Kuusipalo's (1985) hierarchical clustering scheme is not very different from the Cajanderian system in this sense: hierarchical classification based on two-way indicator species analysis offers more flexibility for the determination of class boundaries, but otherwise the resulting system is still unidimensional.

The observation that the Cajanderian unidimensional site type system is a special case of the description of multidimensional variation in site characteristics offers some interesting insights. It seems to be more than a coincidence that the forest site type system was first developed and applied to the upland forests of southern Finland. Several studies (including this one) have indicated that the variation in site quality, as well as vegetation composition, within this region is rather unidimensional in nature (cf. Kuusipalo 1985, Lahti and Väisänen 1987, Tonteri et al. 1990b). However, the unidimensionality of ecological variation, i.e. strong covariation between the principal ecological factors, is not generally the case in nature and breaks down when extended to other biotopes. Against this background, it is not surprising to notice the problems met with the application of the Cajanderian site type system to the classification of forests in other parts of Finland, as well as in other countries.

The arguments presented in this study for the need to extend the traditional Cajanderian site classification system into a statistically consistent multivariate model of environmental variation in forests are parallel to those presented by Kilkki (1987). Contrary to the conclusions of some previous studies (e.g. Cajander 1949, Kuusipalo 1985), estimation of site conditions in a continuous scale of variation seems to be not only possible, but also preferable when seeking a common theoretical framework for the description of site-

specific variation in Finnish forests. However, the fact that the adoption of the new paradigm would not offer any significant improvements over current practices in forest management may be a sufficient reason for continuing the Cajanderian tradition.

References

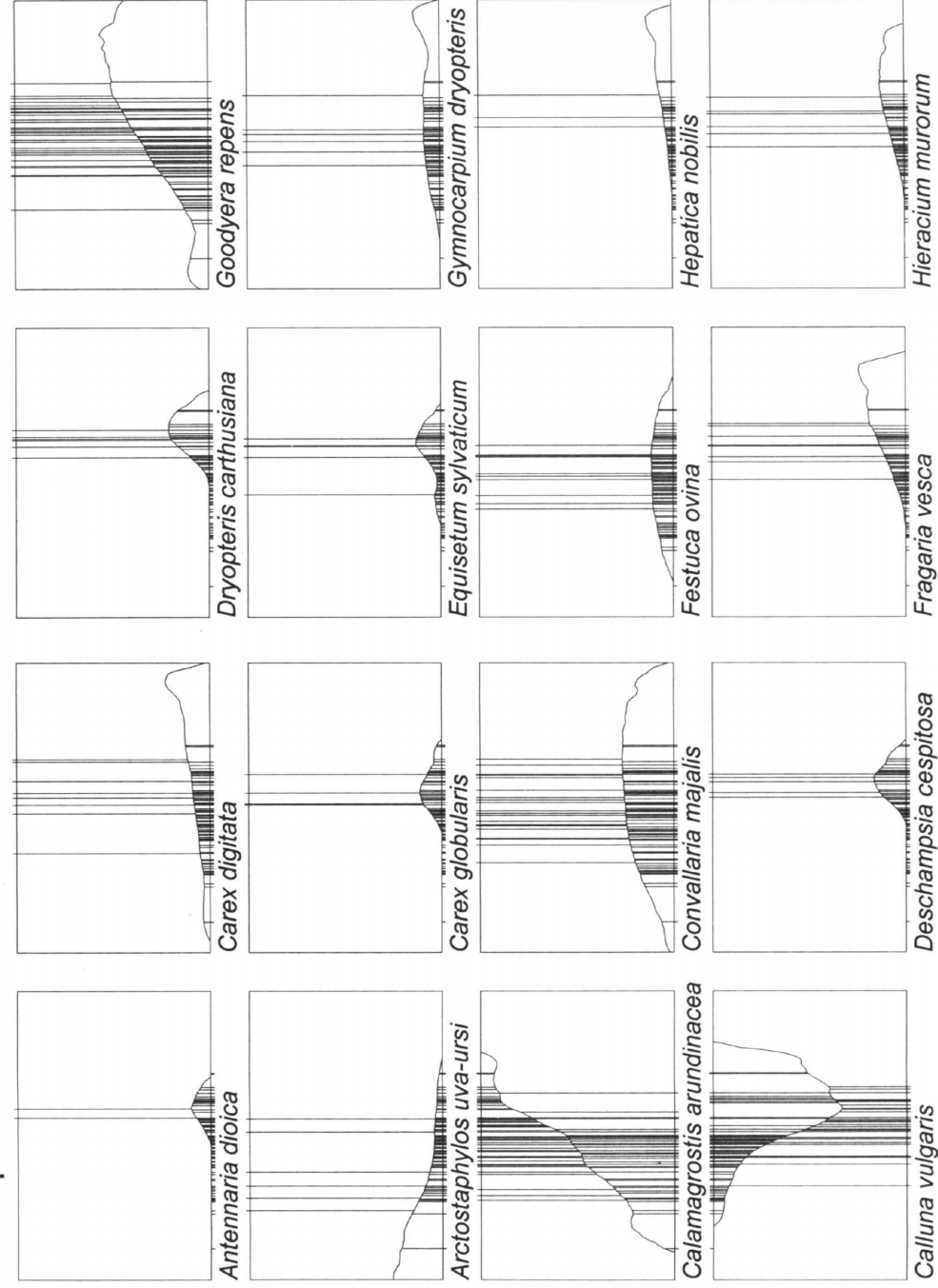
- Abramson, I.S. 1982. On bandwidth variation in kernel estimates – a square root law. *The Annals of Statistics* 10: 1217–1223.
- Ahti, T., Hämet-Ahti, L. & Jalas, J. 1968. Vegetation zones and their sections in northwestern Europe. *Annales Botanici Fennici* 5: 169–211.
- Austin, M.P. 1980. Searching for a model for use in vegetation analysis. *Vegetatio* 42: 11–21.
- 1985. Continuum concept, ordination methods and niche theory. *Annual Review of Ecology and Systematics* 16: 39–61.
- 1987. Models for the analysis of species' response to environmental gradients. *Vegetatio* 69: 35–45.
- & Smith, T.M. 1989. A new model for the continuum concept. *Vegetatio* 83: 35–47.
- Birks, H.J.B. 1987. Recent methodological developments in quantitative descriptive biology. *Annales Zoologici Fennici* 24: 165–178.
- , Line, J.M., Juggins, S., Stevenson, A.C. & ter Braak, C.J.F. 1990. Diatoms and pH reconstruction. *Philosophical Transactions of the Royal Society of London B* 327: 263–278.
- Blomqvist, A.G. 1872. Tabeller framställande utvecklingen af jernåriga och slutna skogsbestånd af tall, gran och björk. Helsinki. 30 p.
- Breiman, L., Meisel, W. & Purcell, E. 1977. Variable kernel estimates of multivariate densities. *Technometrics* 19: 135–144.
- Brown, J.H. 1984. On the relationship between abundance and distribution of species. *The American Naturalist* 124: 255–279.
- Cajander, A.K. 1909. Ueber Waldtypen. *Fennia* 28(2). 176 p.
- 1926. The theory of forest types. *Acta Forestalia Fennica* 29(3). 108 p.
- 1943. Wesen und Bedeutung der Waldtypen. *Inter-sylva* 3: 169–209.
- 1949. Forest types and their significance. *Acta Forestalia Fennica* 56(5). 71 p.
- & Ilvessalo, Y. 1921. Ueber Waldtypen II. *Acta Forestalia Fennica* 20(1). 77 p.
- & Ilvessalo, Y. 1922. I skogstypsfrågan. *Svenska Skogsvårdsföreningens Tidskrift* 21: 125–129.
- Clutter, J.L., Fortson, J.C., Pienaar, L.V., Bristley, G.H. & Bailey, R.L. 1983. *Timber management: a quantitative approach*. Wiley, New York. 333 p.
- Crawley, M.J. 1986. The structure of plant communities. In: Crawley, M.J. (ed.). *Plant ecology*. Blackwell Scientific Publications, Oxford. p. 1–50.
- Efron, B. 1979. Bootstrap methods: another look at the jackknife. *The Annals of Statistics* 7: 1–26.
- & Gong, G. 1983. A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician* 37: 36–48.
- Ellenberg, H. 1979. *Zeigerwerte der Gefäßpflanzen Mitteleuropas*. 2nd ed. *Scripta Geobotanica* 9. 122 p.
- 1982. *Vegetation Mitteleuropas mit den Alpen in ökologischer Sicht*. 3rd ed. Ulmer, Stuttgart. 989 p.
- 1988. *Vegetation ecology of Central Europe*. 4th ed. Cambridge University Press, Cambridge. 731 p.
- Frey, T.E.A. 1973. The Finnish school and forest site-types. In: Whittaker, R.H. (ed.). *Ordination and classification of communities*. Junk, The Hague. p. 404–433.
- Fryer, M.J. 1976. Some errors associated with the nonparametric estimation of density functions. *Journal of the Institute of Mathematics and its Applications* 18: 371–380.
- Gasse, F. & Tekaia, F. 1983. Transfer functions for estimating paleoecological conditions (pH) from East African diatoms. *Hydrobiologia* 103: 85–90.
- Gauch, H.G., Jr. 1982. *Multivariate analysis in community ecology*. Cambridge University Press, Cambridge. 298 p.
- & Whittaker, R.H. 1972. Coenocline simulation. *Ecology* 53: 446–451.
- Gause, G.F. 1930. Studies on the ecology of the Orthoptera. *Ecology* 11: 307–325.
- Goff, F.G. & Cottam, G. 1967. Gradient analysis: the use of species and synthetic indices. *Ecology* 48: 783–806.
- Goldsmith, F.B. (ed.) 1991. *Monitoring for conservation and ecology*. Chapman and Hall, London. 275 p.
- Gordon, A.D. 1981. Classification methods for the exploratory analysis of multivariate data. Chapman and Hall, London. 193 p.
- Gustavsen, H. 1981. Talousmetsien kasvupaikkaluokittelu valtapituuden avulla. Abstract: Site index curves for conifer stands in Finland. *Folia Forestalia* 454. 31 p.
- Hall, P. 1990. Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis* 32: 177–203.
- Hanski, I. 1982. Dynamics of regional distribution: the core and satellite species hypothesis. *Oikos* 38: 210–221.
- Hastie, T.J. & Tibshirani, R.J. 1990. *Generalized additive models*. Chapman and Hall, London. 335 p.
- Heikinheimo, O. 1915. Kaskiviljelyksen vaikutus Suomen metsiin. Referat: Der Einfluss der Brandwirtschaft auf die Wälder Finnlands. *Acta Forestalia Fennica* 4(2). 264 p.
- Hill, M.O. 1973. Reciprocal averaging: an eigenvector method of ordination. *Journal of Ecology* 61: 237–249.
- 1979a. TWINSPAN. A FORTRAN program for arranging multivariate data in an ordered two-way table by classification of individuals and attributes. *Dept. of Ecology and Systematics, Cornell University, Ithaca, New York*. iv + 90 p.
- 1979b. DECORANA: a FORTRAN program for detrended correspondence analysis and reciprocal averaging. *Dept. of Ecology and Systematics, Cornell University, Ithaca, New York*. iii + 52 p.
- 1989. Computerized matching of relevés and association tables, with an application to the British National Vegetation Classification. *Vegetatio* 83: 187–194.
- , Bunce, R.G.H. & Shaw, M.W. 1975. Indicator species analysis, a divisive polythetic method of classification, and its application to a survey of native pinewoods in Scotland. *Journal of Ecology* 63: 597–613.
- & Gauch, H.G. 1980. Detrended correspondence analysis: an improved ordination technique. *Vegetatio* 42: 47–58.
- Hotanen, J.-P. & Kuusipalo, J. 1988. Metsätyypit kiinnostavat jälleen. Summary: The Finnish forest site type approach: revival of classical influence. *Luonnon Tutkija* 92: 64–74.
- Ilvessalo, Y. 1920. Tutkimuksia metsätyyppien taksoo-
toorisesta merkityksestä. Referat: Untersuchungen über die taxatorische Bedeutung der Waldtypen. *Acta Forestalia Fennica* 15(2). 157 p.
- 1956. Suomen metsät vuosista 1921–24 vuosiin 1951–53. Kolmeen valtakunnan metsien inventointiin perustuva tutkimus. Summary: The forests of Finland from 1921–24 to 1951–53. *Communicationes Instituti Forestalis Fenniae* 47(1). 227 p.
- & Ilvessalo, M. 1975. Suomen metsätyypit metsiköiden luontaisen kehitys- ja puuntuottokyvyn valossa. Summary: The forest types of Finland in the light of natural development and yield capacity of forest stands. *Acta Forestalia Fennica* 144. 101 p.
- Izenman, A.J. 1991. Recent developments in nonparametric density estimation. *Journal of the American Statistical Association* 86: 205–224.
- Jackson, D.A. & Somers, K.M. 1991. Putting things in order: the ups and downs of detrended correspondence analysis. *The American Naturalist* 137: 704–712.
- Jongman, R.H.G., ter Braak, C.J.F. & van Tongeren, O.F.R. (eds.) 1987. *Data analysis in community and landscape ecology*. Pudoc, Wageningen. 299 p.
- Kalela, A. 1961. Waldvegetationszonen Finnlands und ihre klimatischen Paralleltypen. *Archivum Societatis Vanamo* 16: 65–83.
- Kalliola, R. 1973. Suomen kasvimaantiede. WSOY, Helsinki. 308 p.
- Keltikangas, V. 1959. Suomalaisista seinäsammaltyypeistä ja niiden asemasta Cajanderin luokitusjärjestelmässä. Summary: Finnish feather-moss types and their position in Cajander's forest site classification. *Acta Forestalia Fennica* 69(2). 266 p.
- Kendall, M. & Stuart, A. 1979. *The advanced theory of statistics*. Vol. 2. Inference and relationship. Charles Griffin, London. 748 p.
- Kilikki, P. 1987. Tutkimusmenetelmät tiedon tuotannon välineenä. In: Itä-Suomen metsätiedepäivä 9.12.1986. "Metsäntutkimus käytännön toiminnan perustana?" Metsäntutkimuslaitoksen tiedonantaja 258: 20–26.
- & Ojansuu, R. 1981. Pituusbonitoinnin ongelmia. *Metsä ja Puu* 1981(3): 26–28.
- Kujala, V. 1936. Tutkimuksia Keski- ja Pohjois-Suomen välisestä kasvillisuusrajasta. Referat: Über die Vegetationsgrenze von Mittel- und Nord-Finnland. *Communicationes Instituti Forestalis Fenniae* 22(4). 95 p.

- 1938. Metsätyyppien paralleelisuudesta. Referat: Über die Parallelität der Waldtypen. *Communicationes Instituti Forestalis Fenniae* 27(1). 17 p.
- 1979. Suomen metsätyypit. Abstract: Forest types of Finland. *Communicationes Instituti Forestalis Fenniae* 92(8). 45 p.
- Kuusela, K. & Salminen, S. 1969. The 5th national forest inventory in Finland. General design, instructions for field work and data processing. *Communicationes Instituti Forestalis Fenniae* 69(4). 72 p.
- Kuusipalo, J. 1985. An ecological study of upland forest site classification in southern Finland. *Acta Forestalia Fennica* 192. 77 p.
- Lahti, T. & Väisänen, R.A. 1987. Ecological gradients of boreal forests in South Finland: an ordination test of Cajander's forest site type theory. *Vegetatio* 68: 145–156.
- Lehto, J. & Leikola, M. 1987. Käytännön metsätyypit. Kirjayhtymä, Helsinki. 96 p.
- Leikola, M. 1984. Miten metsätyyppiteoria kehittyi v. 1909–1925. In: Sepponen, P., Pitkänen, V. & Poikajärvi, H. (eds.). *Metsien kasvupaikkaluokitus. Metsäntutkimuspäivät Rovaniemellä 1984. Metsäntutkimuslaitoksen tiedonantoja* 148: 6–20.
- Lönnroth, E. 1925. Untersuchungen über die innere Struktur und Entwicklung gleichaltriger naturnormaler Kiefernbestände, basiert auf Material aus der Südhälfte Finnlands. *Acta Forestalia Fennica* 30(1). 269 p.
- Maarel, E. van der. 1979. Transformation of cover-abundance values in phytosociology and its effects on community similarity. *Vegetatio* 39: 97–114.
- Mäkirinta, U. 1990. Validation of the Finnish forest types by numerical methods illustrated by means of two sets of relevé data. *Vegetatio* 88: 143–150.
- McCullagh, P. & Nelder, J.A. 1989. *Generalized linear models*. 2nd ed. Chapman and Hall, London. 511 p.
- Mikola, P. 1982. Application of vegetation science to forestry. In: Jahn, G. (ed.). *Vegetation science in forestry*. Junk, The Hague. p. 199–224.
- Miller, R.G. 1981. *Simultaneous statistical inference*. Springer, New York. 298 p.
- Minchin, P.R. 1987. An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio* 69: 89–107.
- Nieppola, J. 1986. Cajanderin metsätyyppiteoria. Kirjallisuuteen perustuva tarkastelu. Summary: Cajander's theory of forest site types. Literature review. *Folia Forestalia* 654. 42 p.
- 1992. Long-term vegetation changes in stands of *Pinus sylvestris* in southern Finland. *Journal of Vegetation Science* 3: 475–484.
- 1993a. Understorey plants as indicators of site productivity in *Pinus sylvestris* L. stands. *Scandinavian Journal of Forest Research* 8: 49–65.
- 1993b. Site classification in *Pinus sylvestris* L. forests in southern Finland. *Silva Fennica* 27: 9–20.
- & Carleton, T.J. 1991. Relations between understorey vegetation, site productivity, and environmental factors in *Pinus sylvestris* L. stands in southern Finland. *Vegetatio* 93: 57–72.
- Oikarinen, M. 1983. Etelä-Suomen viljeltyjen rauduskoivikoiden kasvatusmallit. Summary: Growth and yield models for silver birch (*Betula pendula*) plantations in southern Finland. *Communicationes Instituti Forestalis Fenniae* 113. 75 p.
- Økland, R.H. 1990. *Vegetation ecology: theory, methods and applications with reference to Fennoscandia*. *Sommerfeltia Supplement* 1. 233 p.
- Oksanen, J. 1984. Lichen-rich forests and related communities in Finland: ordination and classification studies. *University of Joensuu, Publications in Sciences* 1. 35 p.
- 1987. Problems of joint display of species and site scores in correspondence analysis. *Vegetatio* 72: 51–57.
- 1988. A note on the occasional instability of detrending in correspondence analysis. *Vegetatio* 74: 29–32.
- 1990. *Vegetation science in Finland*. *Journal of Vegetation Science* 1: 277–282.
- , Läärä, E., Huttunen, P. & Meriläinen, J. 1988. Estimation of pH optima and tolerances of diatoms in lake sediments by the methods of weighted averaging, least squares and maximum likelihood, and their use for the prediction of lake acidity. *Journal of Paleolimnology* 1: 39–49.
- , Läärä, E., Huttunen, P. & Meriläinen, J. 1990. Maximum likelihood prediction of lake acidity based on sedimented diatoms. *Journal of Vegetation Science* 1: 49–56.
- Oksanen, L. 1991. A century of community ecology: how much progress? *Trends in Ecology and Evolution* 6: 294–296.
- Olkin, I. & Spiegelman, C.H. 1987. A semiparametric approach to density estimation. *Journal of the American Statistical Association* 82: 858–865.
- Orlóci, L. 1978. *Multivariate analysis in vegetation research*. 2nd ed. Junk, The Hague. x + 451 p.
- Peet, R.K., Knox, R.G., Case, J.S. & Allen, R.B. 1988. Putting things in order. The advantages of detrended correspondence analysis. *The American Naturalist* 131: 924–934.
- Persson, S. 1981. Ecological indicator values as an aid in the interpretation of ordination diagrams. *Journal of Ecology* 69: 71–84.
- Pielou, E.C. 1984. *The interpretation of ecological data*. Wiley, New York. 263 p.
- Rabinowitz, D. 1981. Seven forms of rarity. In: Syngé, H. (ed.). *The biological aspects of rare plant conservation*. Wiley, New York. p. 205–217.
- Ramensky, L.G. 1930. Zur Methodik der vergleichenden Bearbeitung und Ordnung von Pflanzenlisten und anderen Objekten, die durch mehrere, verschiedenartig wirkende Faktoren bestimmt werden. *Beiträge zur Biologie der Pflanzen* 18: 269–304.
- SAS 1989. *SAS/STAT user's guide*. Version 6. Volume 2. SAS Institute, Cary, North Carolina, USA. 846 p.
- Sepponen, P. 1985. The ecological classification of sorted forest soils of varying genesis in northern Finland. *Communicationes Instituti Forestalis Fenniae* 129. 77 p.
- Shipley, B. & Keddy, P.A. 1987. The individualistic and community-unit concepts as falsifiable hypotheses. *Vegetatio* 69: 47–55.
- Silverman, B. 1986. *Density estimation for statistics and data analysis*. Chapman and Hall, London. 175 p.
- Sokal, R.R. & Rohlf, F.J. 1981. *Biometry*. 2nd ed. Freeman, New York. 859 p.
- Swan, J.M.A. 1970. An examination of some ordination problems by use of simulated vegetational data. *Ecology* 51: 89–102.
- Tamm, O. & Petrini, S. 1922a. Review of "Cajander, A.K. and Ilvessalo, Y.: Ueber Waldtypen II" (in Swedish). *Svenska Skogsvårdsföreningens Tidskrift* 20: 21X–28X.
- & Petrini, S. 1922b. I skogstypsfrågan. *Svenska Skogsvårdsföreningens Tidskrift* 20: 169X–170X.
- Taylor, C.C. 1989. Bootstrap choice of the smoothing parameter in kernel density estimation. *Biometrika* 76: 705–712.
- ter Braak, C.J.F. 1988. CANOCO – an extension of DECORANA to analyze species-environment relationships. *Vegetatio* 75: 159–160.
- & Barendregt, L.G. 1986. Weighted averaging of species indicator values: its efficiency in environmental calibration. *Mathematical Biosciences* 78: 57–72.
- & Gremmen, N.J.M. 1987. Ecological amplitudes of plant species and the internal consistency of Ellenberg's indicator values for moisture. *Vegetatio* 69: 79–87.
- & Looman, C.W.N. 1986. Weighted averaging, logistic regression and the Gaussian response model. *Vegetatio* 65: 3–11.
- & Prentice, I.C. 1988. A theory of gradient analysis. *Advances in Ecological Research* 18: 271–317.
- & van Dam, H. 1989. Inferring pH from diatoms: a comparison of old and new calibration methods. *Hydrobiologia* 178: 209–223.
- Tonteri, T. 1990. Inter-observer variation in forest vegetation cover assessments. *Silva Fennica* 24: 189–196.
- , Hotanen, J.-P. & Kuusipalo, J. 1990a. The Finnish forest site type approach: ordination and classification studies of mesic forest sites in southern Finland. *Vegetatio* 87: 85–98.
- , Mikkola, K. & Lahti, T. 1990b. Compositional gradients in the forest vegetation of Finland. *Journal of Vegetation Science* 1: 691–698.
- van Groenewoud, H. 1992. The robustness of correspondence, detrended correspondence, and TWINSPAN analysis. *Journal of Vegetation Science* 3: 239–246.
- Vanha-Majamaa, I. 1993. Metsätyypin määrittämistä vaikeudet. *Luonnon Tutkija* 97: 20–27.
- Vuokila, Y. & Väliäho, H. 1980. Viljeltyjen havumetsiköiden kasvatusmallit. Summary: Growth and yield models for conifer cultures in Finland. *Communicationes Instituti Forestalis Fenniae* 99(2). 271 p.
- Wartenberg, D., Ferson, S. & Rohlf, F.J. 1987. Putting things in order: a critique of detrended correspondence analysis. *The American Naturalist* 129: 434–448.
- Whittaker, R.H. 1956. *Vegetation of the Great Smoky Mountains*. *Ecological Monographs* 26: 1–80.
- 1967. Gradient analysis of vegetation. *Biological Reviews* 42: 207–264.
- Yee, T.W. & Mitchell, N.D. 1991. Generalized additive models in plant ecology. *Journal of Vegetation Science* 2: 587–602.

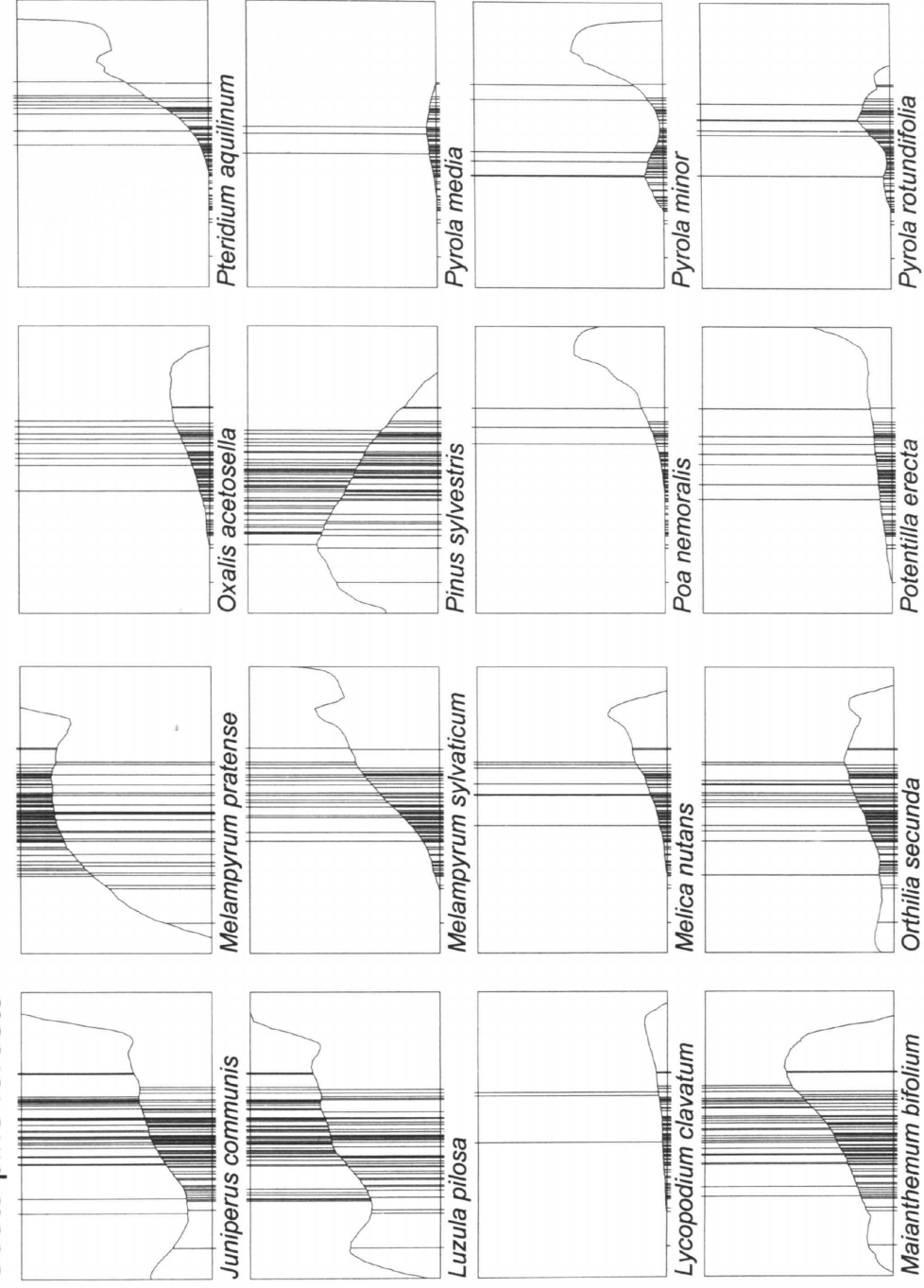
Total of 118 references

Appendix 2. Presence-absence response curves for a selection of taxa in pine- and spruce-dominated forests, estimated with an adaptive kernel method. Optimal window width was estimated separately for each taxon with a jackknife method. Vertical lines above the response curve show the locations of sample plots with the taxon present, those below the response curve show the locations with the taxon absent. The range of the H_{100} values on the x-axis is 10–35 meters. The y-axis shows the probability of occurrence (range from zero to one).

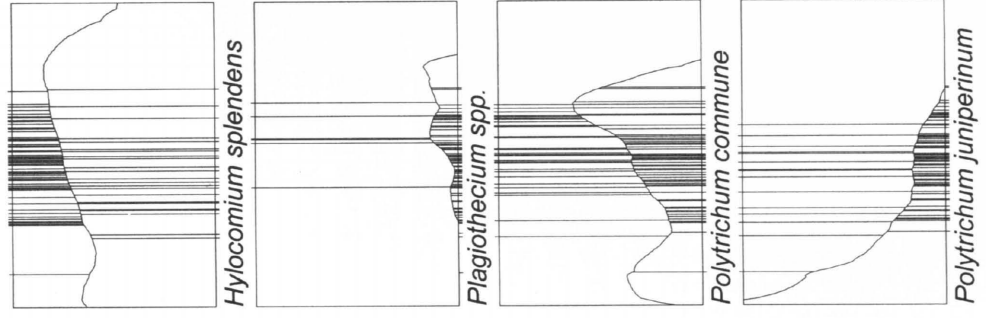
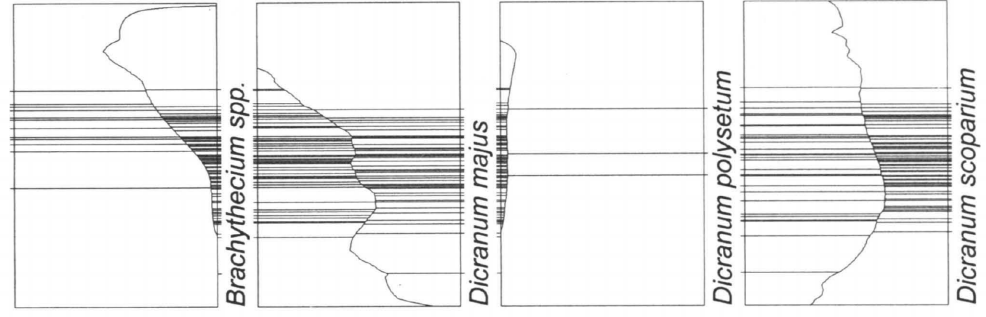
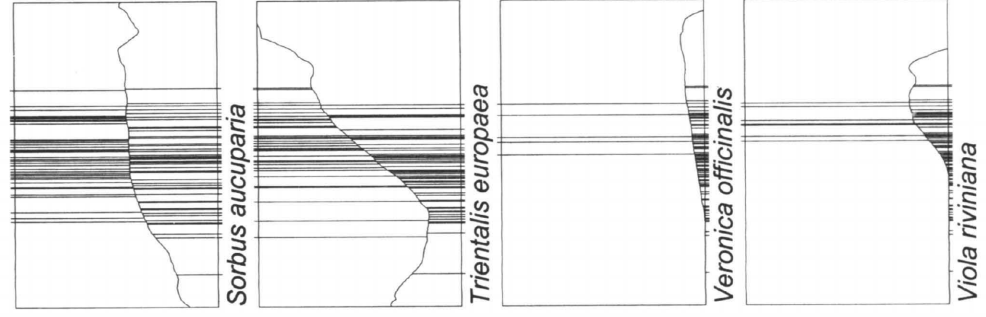
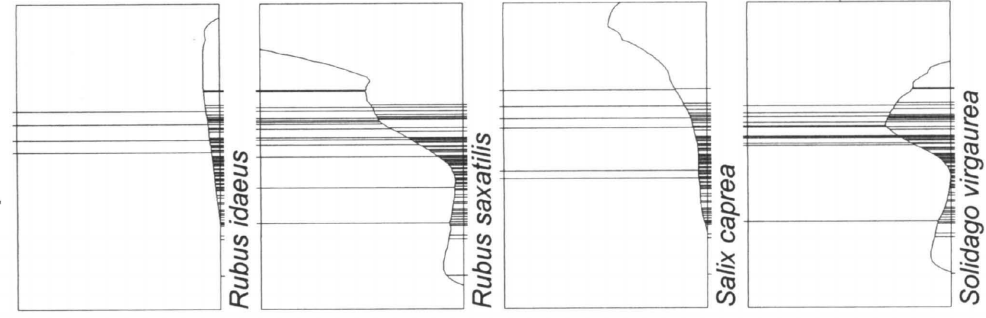
Scots pine forests



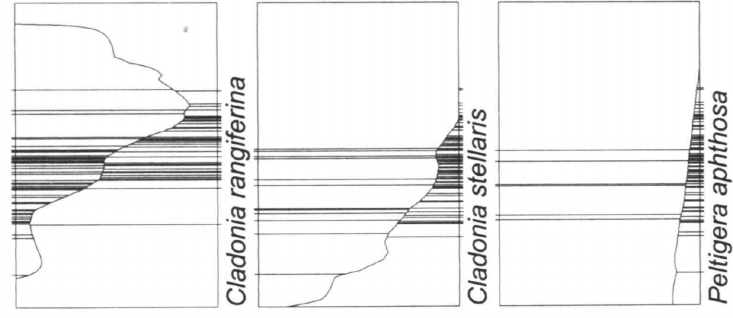
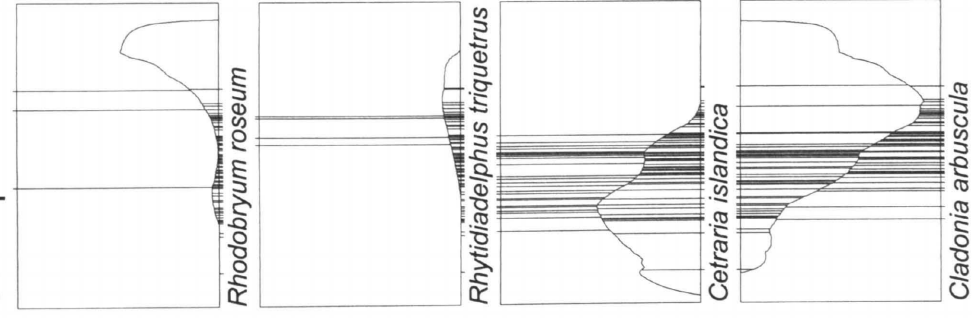
Scots pine forests



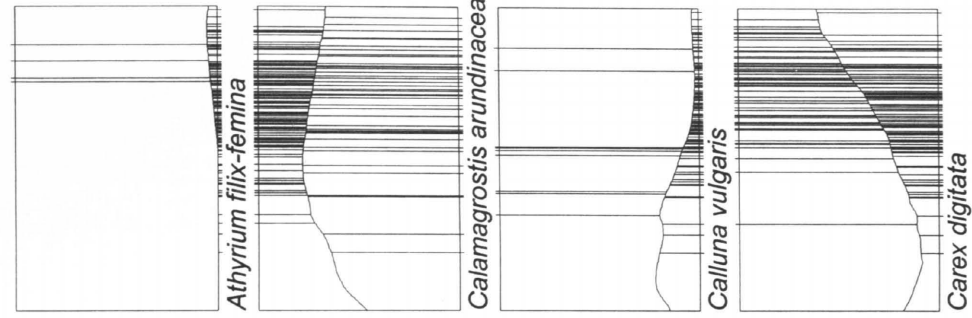
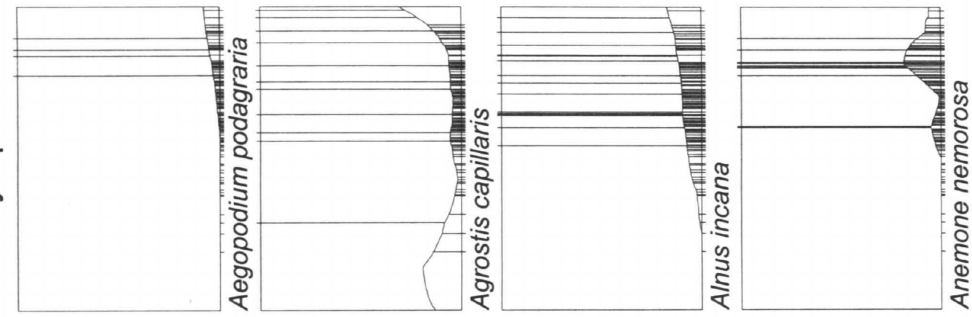
Scots pine forests



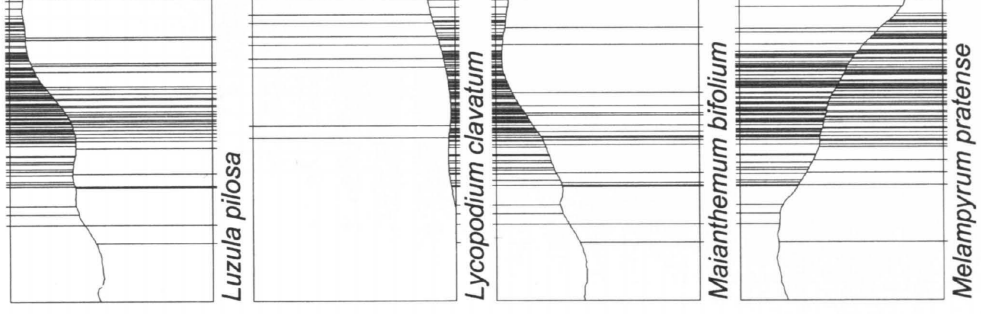
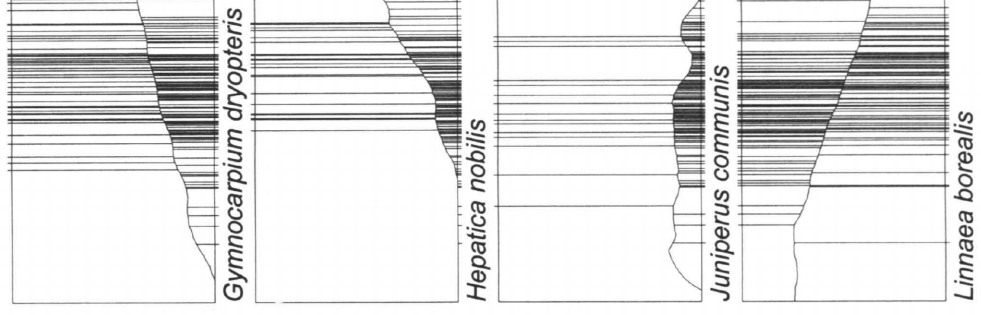
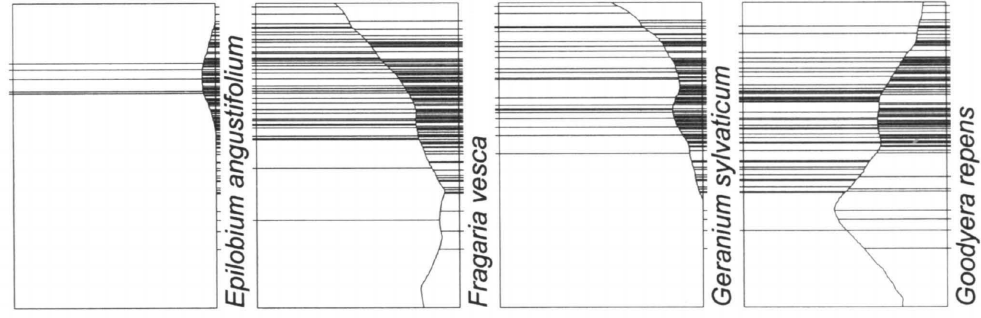
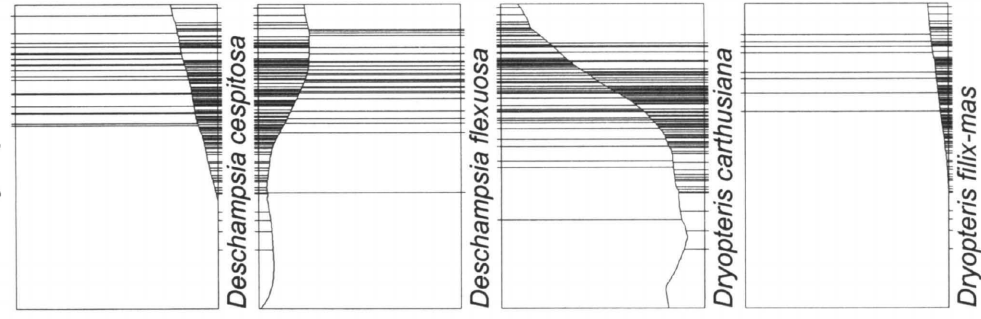
Scots pine forests



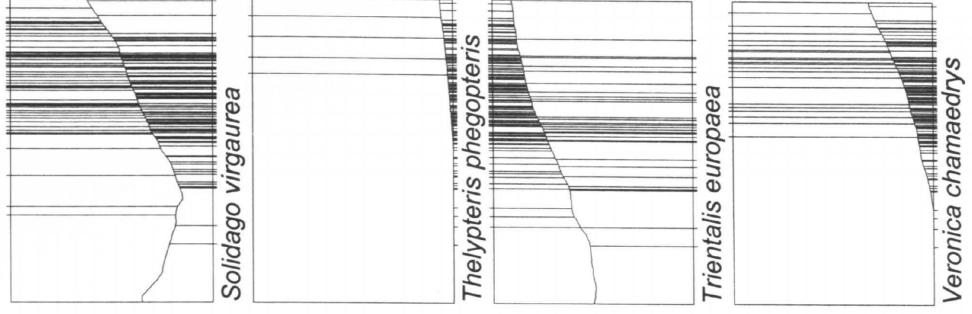
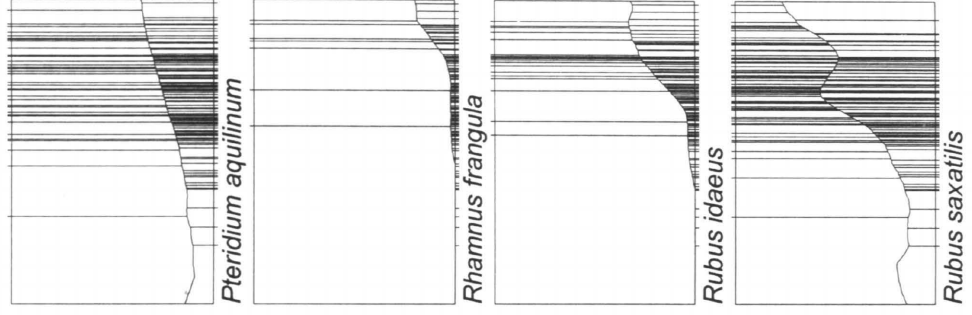
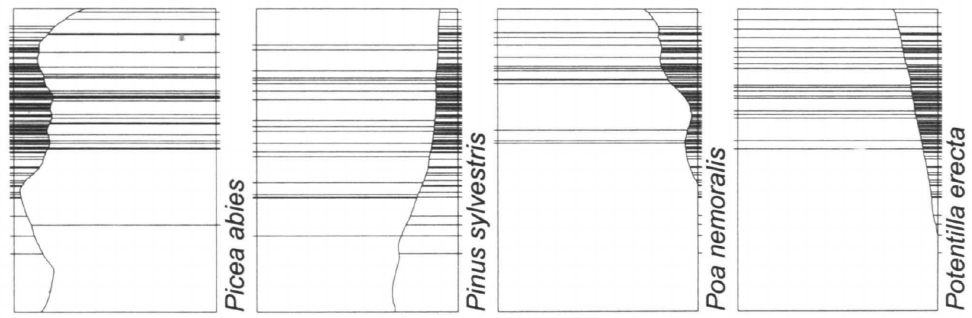
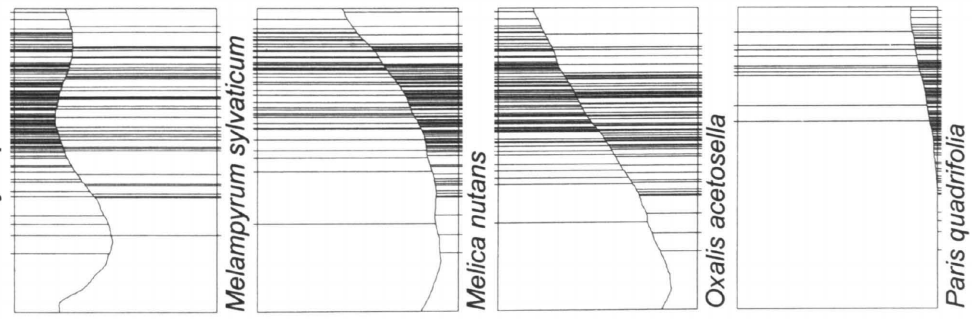
Norway spruce forests

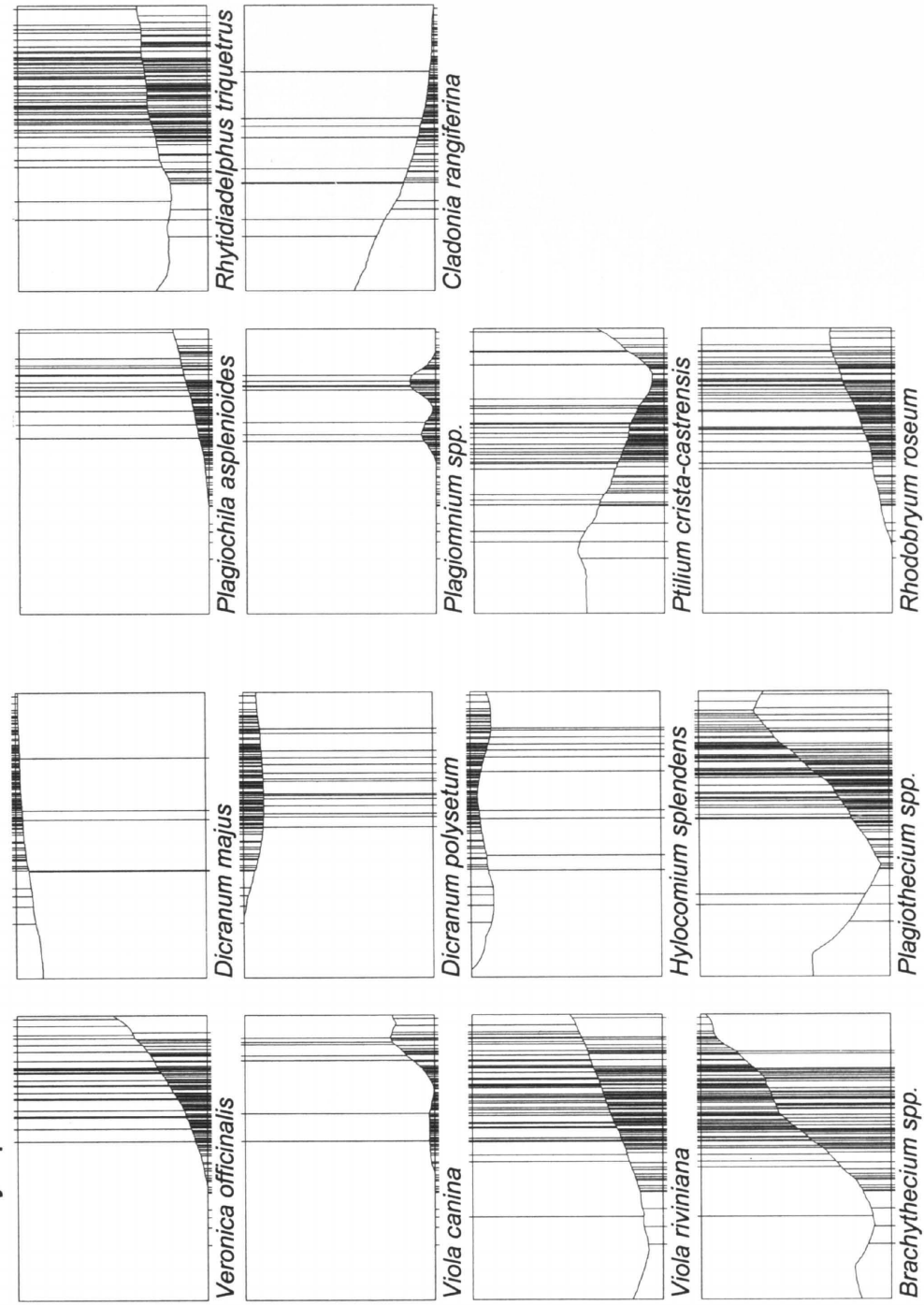


Norway spruce forests



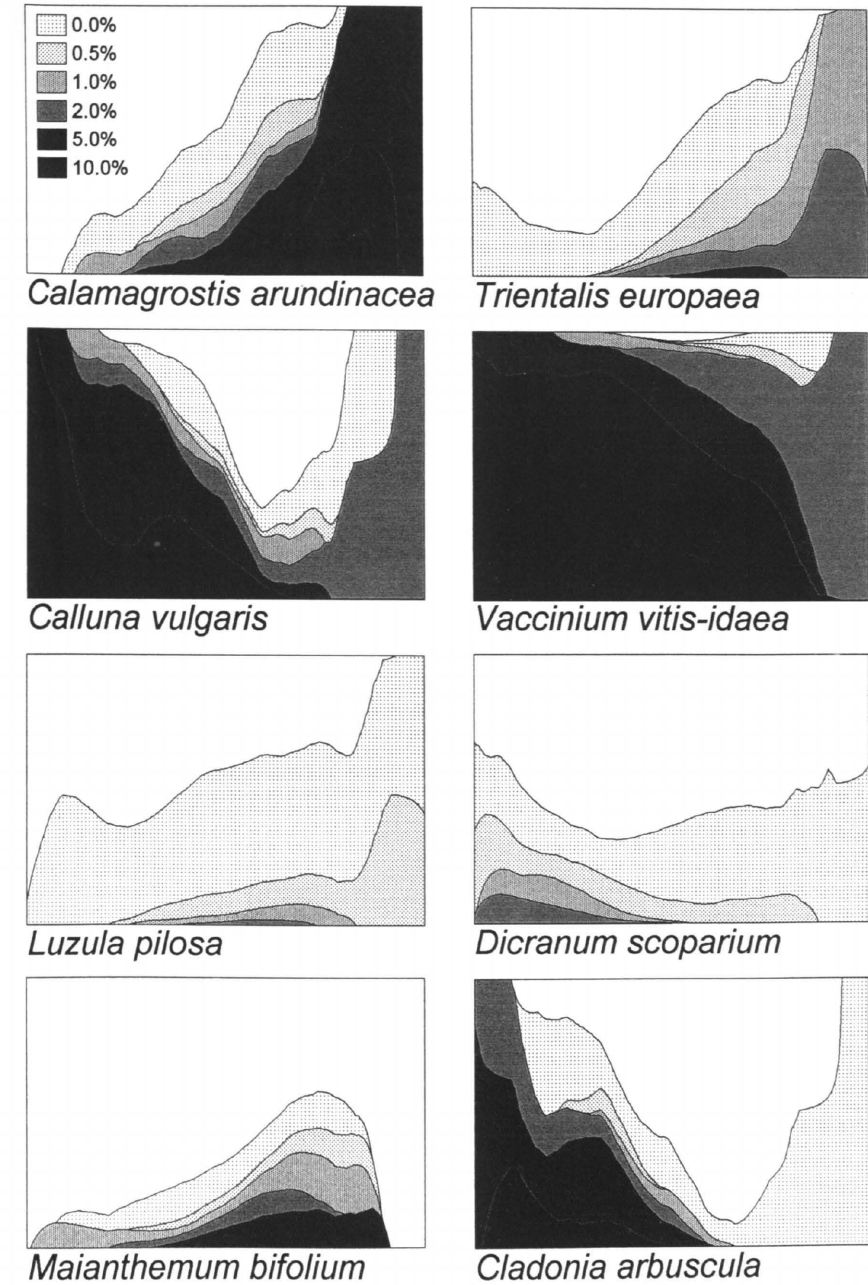
Norway spruce forests



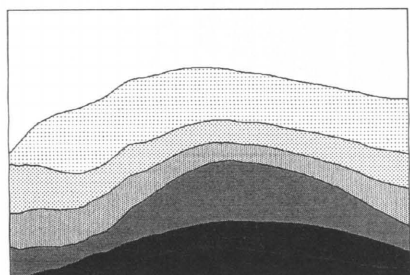


Appendix 3. Presence-absence response curves of abundance data (pseudospecies) for a selection of taxa in pine- and spruce-dominated forests, estimated with an adaptive kernel method. For each taxon, six limit abundances (0 %, 0.5 %, 1 %, 2 %, 5 % and 10 %) were used for estimating the response curves. The range of the H_{100} values on the x-axis is 10–35 meters. The y-axis shows the probability of occurrence (range from zero to one).

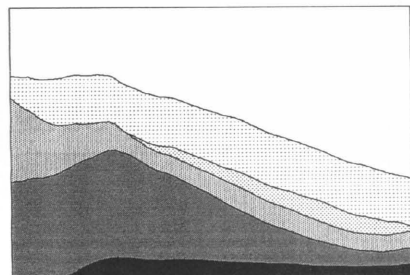
Scots pine forests



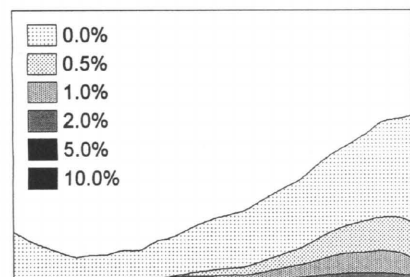
Norway spruce forests



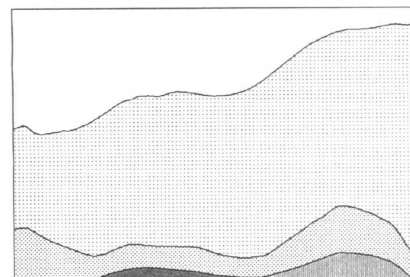
Calamagrostis arundinacea



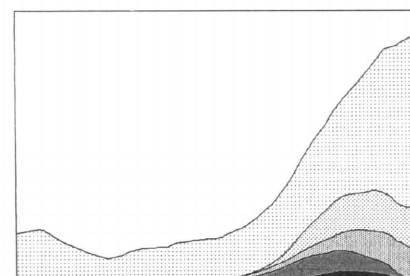
Linnaea borealis



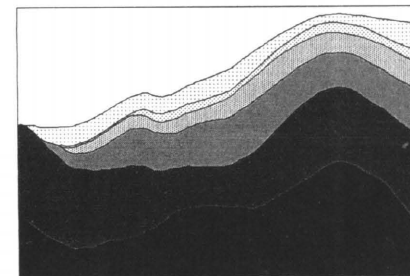
Carex digitata



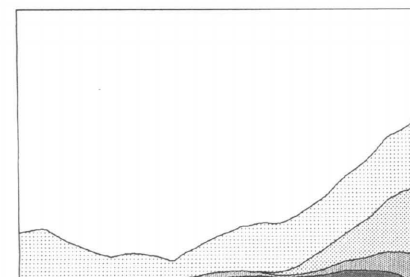
Luzula pilosa



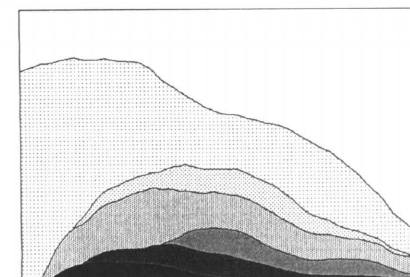
Dryopteris carthusiana



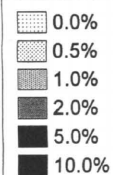
Maianthemum bifolium



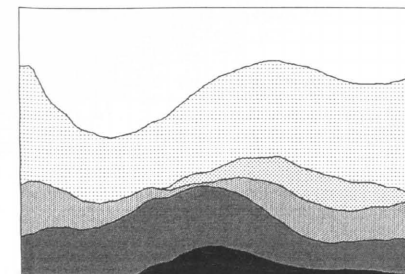
Fragaria vesca



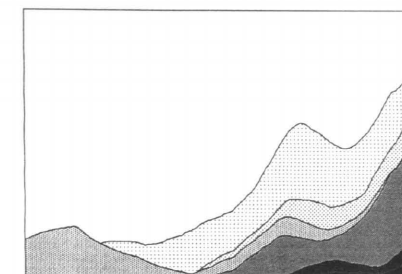
Melampyrum pratense



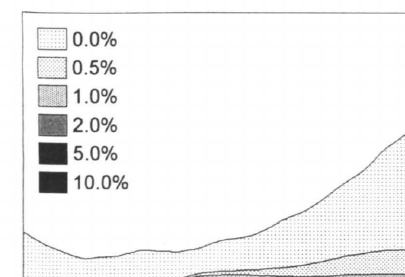
Norway spruce forests



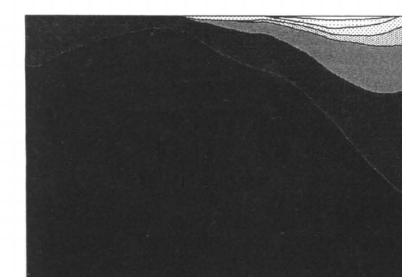
Melampyrum sylvaticum



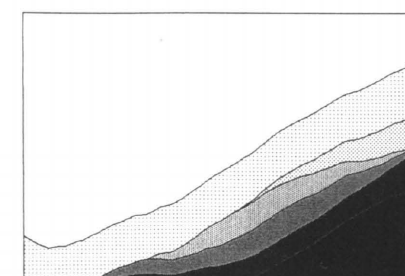
Rubus saxatilis



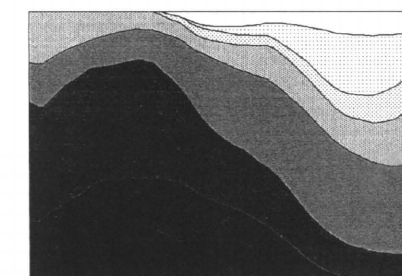
Melica nutans



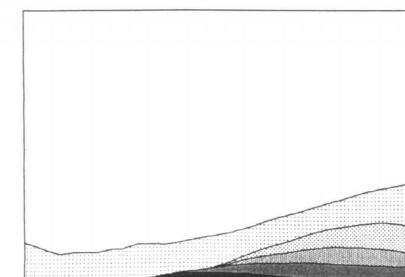
Vaccinium myrtillus



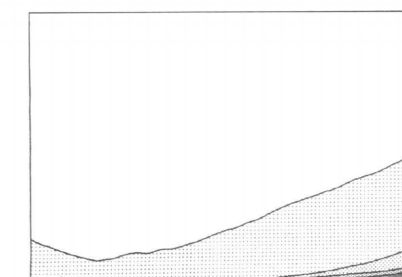
Oxalis acetosella



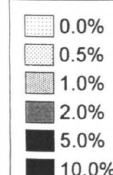
Vaccinium vitis-idaea



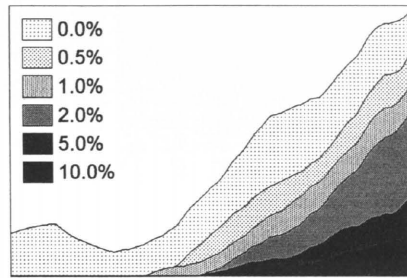
Pteridium aquilinum



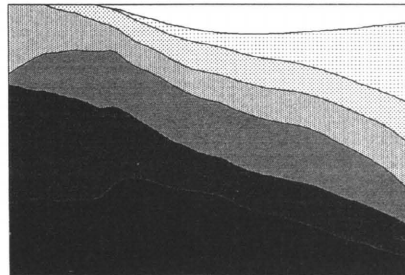
Viola riviniana



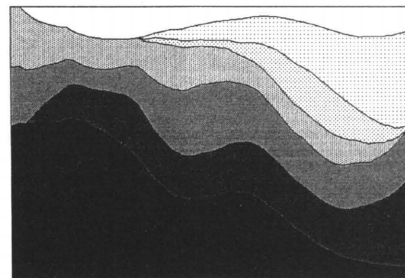
Norway spruce forests



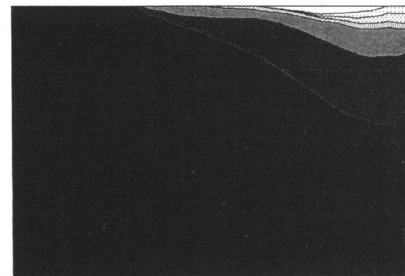
Brachythecium spp.



Dicranum polysetum



Hylocomium splendens



Pleurozium schreberi

Submission of Manuscripts Manuscripts should be submitted in triplicate to Acta Forestalia Fennica, Unioninkatu 40 A, FIN-00170 Helsinki, Finland. Detailed instructions to authors are printed in Acta Forestalia Fennica 244. Offprints of Instructions are available on request.

Publication Schedule Acta Forestalia Fennica is published intermittently about five numbers in a year.

Subscriptions and Exchange Subscriptions and orders for back issues should be addressed to Academic Bookstore, Subscription Services, P.O. Box 23, FIN-00371 Helsinki, Finland, Phone +358 0 121 4430, Fax +358 0 121 4450. Subscription price for 1995 is 70 FIM per issue. Exchange inquiries should be addressed to the Finnish Society of Forest Science, Unioninkatu 40 B, FIN-00170 Helsinki, Finland, Phone +358 0 658 707, Fax +358 0 191 7619.

Statement of Publishers Acta Forestalia Fennica has been published since 1913 by the Finnish Society of Forest Science. In 1989 Acta Forestalia Fennica was merged with Communicationes Instituti Forestalis Fenniae, started in 1919 by the Finnish Forest Research Institute. In the merger, the Society and Forest Research Institute became co-publishers of Acta Forestalia Fennica. The Finnish Society of Forest Science is a nonprofit organization founded in 1909 to promote forest research. The Finnish Forest Research Institute, founded in 1917, is a research organization financed by the Ministry of Agriculture and Forestry.

Abstracting Articles in Acta Forestalia Fennica are abstracted and indexed in Agrindex, Biological Abstracts, Current Advances in Ecological Sciences, Current Advances in Plant Sciences, Ecological Abstracts, Forest Products Journal, Forestry Abstracts, International Bibliography of Periodical Literature, Life Sciences Collection.

ACTA FORESTALIA FENNICA

The Finnish Society of Forest Science
The Finnish Forest Research Institute

- 243 **Rauno Väisänen** and **Kari Heliövaara**: Assessment of insect occurrence in boreal forests based on satellite imagery and field measurements.
- 244 **Eero Kubin** and **Lauri Kemppainen**: Effect of soil preparation of boreal spruce forest on air and soil temperature conditions in forest regeneration areas.
Instructions to Authors.
- 245 **Reijo Mykkänen**: Aspiration-based utility functions in a planning model for timber flow management.
- 246 **Tapani Lahti**: Understorey Vegetation as an Indicator of Forest Site Potential in Southern Finland.

ISBN 951-40-1435-9
ISSN 0001-5636

