STUDIA FORESTALIA SUECICA

Nr 138 · 1977

Site index estimation by means of site properties Scots pine and Norway spruce in Sweden

Skattning av höjdboniteten med ståndortsfaktorer Tall och gran i Sverige

BJÖRN HÄGGLUND
Department of Forest Yield Research

JAN-ERIK LUNDMARK
Department of Forest Ecology and Forest Soils

SKOGSHÖGSKOLAN
SWEDISH COLLEGE OF FORESTRY
STOCKHOLM

Abstract

ODC 114.521—542

The aim of this study is to establish functional relationships between site index (dominant height at a total age of 100 years) and site properties, which can be used for predicting site index in forestry. The functions are derived from a mathematical growth model, assuming that the effects of different growth factors interact in a multiplicative way. Data selected from the National Forest Survey and comprising more than 3 000 plots are used to estimate the parameters of the model by means of regression analysis. In total, 10 functions are constructed—6 for Scots pine and 4 for Norway spruce. These functions are checked by an examination of residuals and by a comparison with data not used for constructing the functions. The checks gave fairly satisfactory results, and the functions seem suitable for their purpose. However, site indices predicted have an accuracy considerably lower than when using site index curves in stands suitable for that method. Hence, the use of the functions should be restricted to situations where site index curves cannot be applied.

Ms. received 7th February, 1977

LiberFörlag/Allmänna Förlaget

ISBN 91-38-03497-2, ISSN 0039-3150

Berlingska Boktryckeriet, Lund 1977

Contents

1 Introduction

This report deals with the problem of establishing functional relationships between site index (dominant height at a total age of 100 years) and site properties. Site index refers to the two most important species in Swedish forestry, Scots pine and Norway spruce.

The report is a summary of a longer report in Swedish, Hägglund & Lundmark, 1977. This summary mainly deals with methodical questions, while the quantitative results of the study are not fully reported here.

2 The problem

The most common method of estimating site index nowadays is by the use of site index curves, showing the relationship between height and age. Strictly, this means that the stand itself is used as a biological test of site productivity. Such a test is valid only if the condition of the stand is such that the site index estimated is not unduly influenced by stand history. Normally this means that the stand must be undamaged, of one species, even-aged and so on. Furthermore, site index curves do not work very well in young stands where their precision is low and of course they do not work at all on areas without forest. These restrictions mean that site index curves can only be used in 30—40 % of the total Swedish forest area (Hägglund, 1974b). Therefore, there is an urgent need for methods of unbiased site index estimation without the use of stand properties. One such method is to establish functional relationships between site index and site properties. This is the aim of our study. In order to facilitate the practical use of the results of the study, the site properties we use are restricted to those which can easily be recorded in the field.

A critical discussion of problems connected with forecasting forest yield from observations of site characteristics has been made by Tamm et. al. 1967.

3 Method

A number of sample plots from the National Forest Survey are selected as data for the study. The stands on these plots are in such condition that site index could be estimated in a reliable way with site index curves. The site characteristics of the plots are described by measurements of geographical location, ground vegetation and soil properties. The dependence of site index on site properties is expressed in mathematical functions which are based on a multiplicative growth model. The parameters of the model are estimated by means of regression analysis. This operation results in ten functions, eight concerning mineral soils and two concerning peat lands. Six functions describe site index of Scots pine in Sweden, four site index of Norway spruce. The results are checked by studies of the residuals from the functions and by comparisons with data not used to construct the functions.

4 Definitions

A large number of stand and (especially) site properties are used in this study. In order to make the text accessible for the reader we have to define the most important of these. All definitions refer to a circular sample plot with a radius of 10 m.

4.1 **Site index**

Site index is defined as the dominant height at a total age of 100 years, h_{100}. The dominant height is the arithmetic mean height of the 100 largest (by diameter) trees per hectare. In the data used, dominant height is calculated as the arithmetic mean height of the two largest trees on a circular plot with the radius 10 m. Investigations (Fries, 1974) have shown that this way of calculating dominant height gives a fairly unbiased estimate of the arithmetic mean height of the ten largest trees on a 0.1 hectare plot.

Site index is estimated using site index curves of Hägglund (1972, 1973, 1974a). These curves show the relation between dominant height and age at breast height, for different levels of site index.

4.2 **Geographical location**

The geographical location of the plots is described using latitude (LAT), altitude (ALT) and some climatic regions. These regions are defined according to Ångström, 1958. Of special interest in this investigation are

- region M2. A region of maritime climate, situated along the eastern coast of southern Sweden
- region K3. A region of continental climate, situated in the central part of southern Sweden.

Besides these climatic regions, the distance of the plot from the coast is used to describe geographical location in northern Sweden (DC).

A map of Sweden showing the areas under study is enclosed (figure 1).

4.3 **Vegetation**

The description of vegetation on the plots follows the principles outlined by Malmström, 1947. This means that the different vegetation layers are described separately. Here we use descriptions of the field and bottom layers. In connection with this vegetation description, there is a differentiation made between mineral soils and peat lands according to the following definitions.

Mineral soils:	On the whole of the plot or on some part of it there are mineral soils or rocks within 30 cm from the ground surface.
Peat lands:	On the whole of the plot there is peat to a depth of at least 30 cm.

The description of the field layer is the same for mineral soils and peat lands, while different descriptions of the bottom layer are used for the two types of soil. In addition, on mineral soils, we combine the field and bottom layers to forest types, while we on peat lands keep these layers apart. In the following, mainly mineral soils are considered.

Field layer

The description of the field layer comprises four main types, ranked from fertile to poor, namely herb types, grounds without field layer, grass types and dwarf-shrub types. On wet soils also Equisetum and Carex types are used.

At the more detailed differentiation of herb types, the occurrence of indicator plants and their cumulative covering of the plot are used. We differentiate between nemoral herbs (Nemo), tall herbs (Th) and low herbs (Lh). For the two latter types, any considerable covering of dwarf-shrubs is also noted. In this case we get subtypes as ThD.sh+/My, tall herbs with Vaccinium myrtillus.

In dense stands there might be no field layer and these types we call "grounds without field layer", N fl. If more than 25 % of the plot is covered with grasses, it is defined as a grass type (Gr).

The dwarf-shrub types are defined by dominance, and we differentiate between Vaccinium myrtillus type (My), Vaccinium vitis-idea type (Vm), Empetrum type (Em), Calluna type (Cm) and Ledum palustre + Vaccinium uliginosum type (Lu).

Each plot is assigned one unique field layer type according to a special routine, the main principle of which is that more fertile types precede poorer ones.

Bottom layer on mineral soils

The occurrence of mesic-soil mosses (Hylocomium, Pleurozium and others), swamp mosses (Sphagnum and others) and lichens are the most important characteristics of the bottom layer on mineral soils. From the relations between the covering of the three groups mentioned, types are defined mainly to describe the moisture but also to some extent the fertility of the site. In this investigation we use the following four types

no bottom layer: The covering of the bottom layer is less than 1 % of the plot area.

moss type: In the existing bottom layer, mesic-soil mosses and swamp mosses dominate. The lichens cover less than 25 % of the existing bottom layer.

lichen-rich type: In the existing bottom layer mesic-soil mosses and swamp mosses dominate. The lichens cover between 25 and 50 % of the existing bottom layer.

lichen type: Lichens cover more than 50 % of the existing bottom layer.

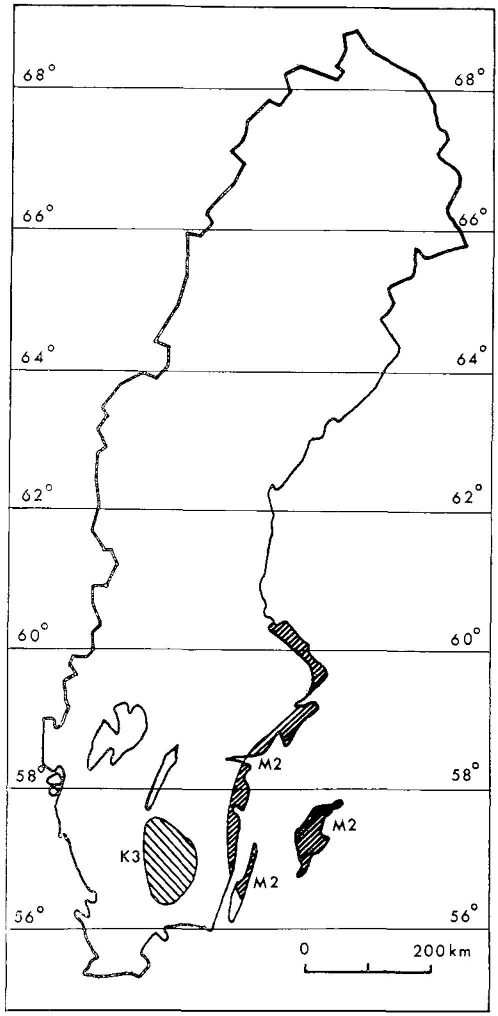


Figure 1. The location of regions M2 (with maritime climate) and K3 (with continental climate).

Forest type

On mineral soils, the registrated field and bottom layers of each plot are combined to define a forest type according to the principles of Tamm & Holmen, 1961. This forest type is purely botanic and, in our case, is founded on ground vegetation only.

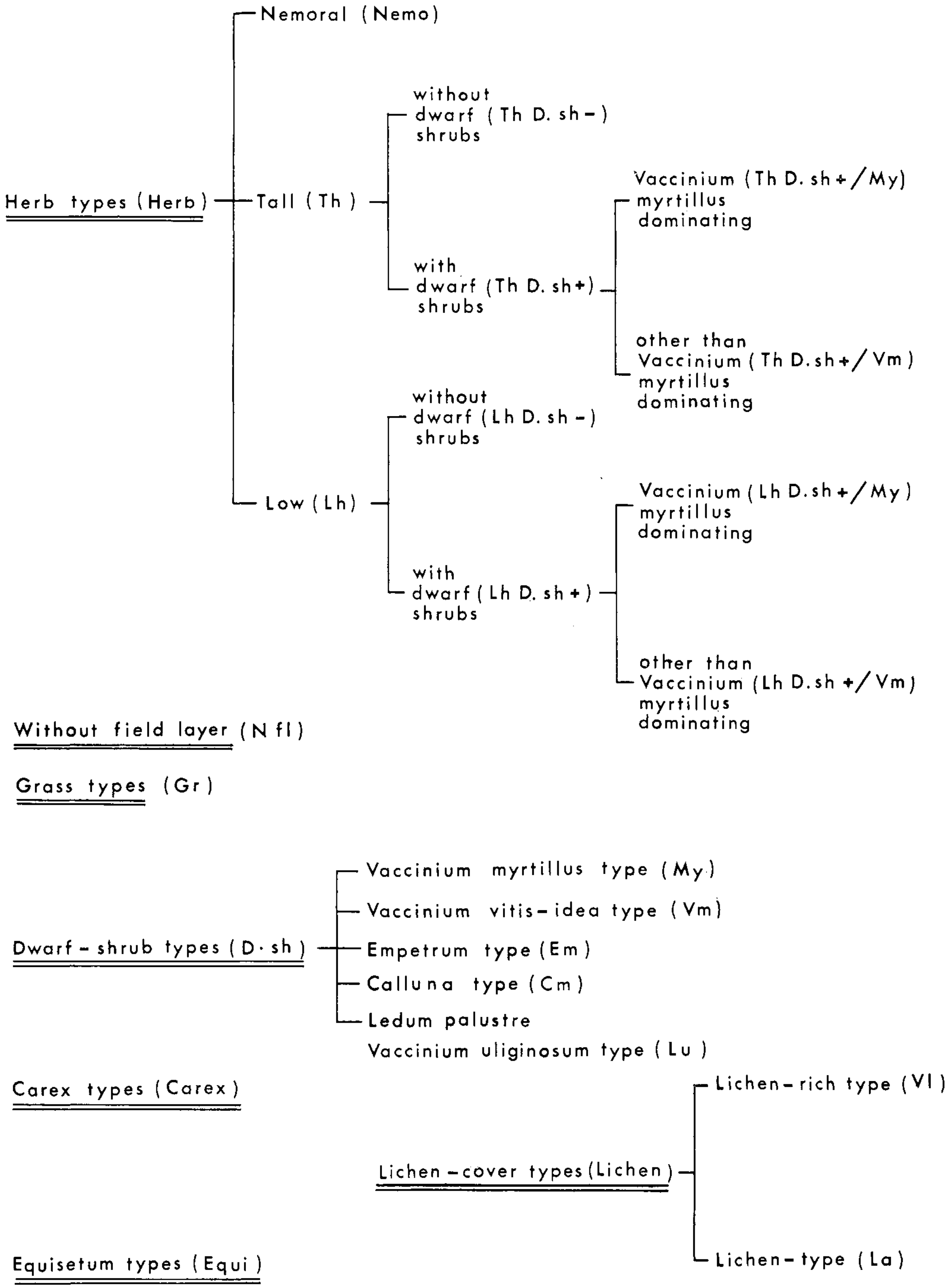


Figure 2. The differentiation of forest types. Main types are underlined (=).

The formation of forest types is hierarchic. First, the bottom layer is considered. Lichen-rich (Vl) and lichen types (La) form forest types with the same names. Notice that no attention is paid to the field layer at this time. When lichen-rich and lichen types are separated from the data, the rest of the plots are classified according to the field layer only. The resulting forest types have the same names as the corresponding field layer types.

The formation of forest types is preceded by a test of biological relevance. In this test we sort out some very rare and inhomogeneous types, for instance the combination of herbs and lichens.

The scheme of forest types used is illustrated in figure 2.

4.4 **Soil**

The following soil properties are the most important of those used in the investigation.

Soil depth

Four classes are distinguished, according to average soil depth within sample plot
deep: more than 70 cm
fairly shallow: 20—70 cm
very shallow: less than 20 cm
displacements: great differences within sample plot

Mechanical composition, textural index

The mechanical composition of the soil is determined by the use of a particle size scale and the rolling test. The composition is recorded as the average of five soil pits. By means of results of empirical studies (Lundmark, 1974a) mechanical composition is converted to a textural index, the numerical value of which is used in further analysis.

Textural index	Main particle size
>4.0	stone
3.5	gravel
2.7	sand
2.2	*sand*-fine sand
1.8	*sand-fine sand*
1.4	sand-*fine sand*
1.2	silt
1.0	clay

Thickness of humus layer

Recorded in centimeters up to 30 cm, after that in 5-cm classes up to 99 cm.

Soil moisture

The position of the watertable is the main criterion for delimiting classes of soil moisture. Certain topographical configurations and the occurrence of moist depressions in the surface are auxiliary to this assessment. Five classes are distinguished.

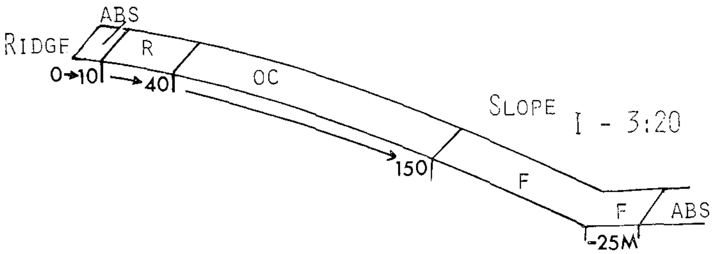
Very dry (VD):	Deep glacifluvial deposits and thick tills with the watertable at great depth.
Dry (D):	Areas with the watertable deeper than 2 m.
Mesic (ME):	Watertable at a depth of 1—2 m.
Moist (MO):	Watertable at a depth of less than 1 m but not visible.
Slightly waterlogged (SW):	The watertable is visible in depressions or seems to be close to the surface of the ground.

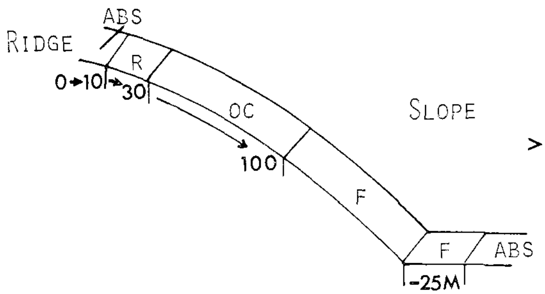
Surface/subsurface water flow

The occurrence of surface/subsurface water flow is assessed from the inclination of the surface and the length of the slope above the sample plot, according to the figure below.

The symbols in the figure are for the classes of surface/subsurface water flow used. These are:

ABS (Absent):	Sites without surface/subsurface water flow	OC (Occasional):	Sites where surface/subsurface water flow occurs during shorter periods
R (Rare):	Sites where surface/subsurface water flow occurs rarely	F (Frequent):	Sites where surface/subsurface water flow occurs during longer periods





5 Model

5.1 Principal features

As stated earlier, the aim of this study is to construct functions describing the relationship between site index (h_{100}) and site properties. These functions are founded on the model described below. The main features of the model originate from Jonsson, 1969. Lundmark, 1974a, has used a variant of the same model.

The basic assumption in the model is that the effects of different growth factors work together in a multiplicative way (Baule, 1917). With this assumption, dominant height growth might be written as

$$h'_t = f_1(t) \cdot f_2(t) \cdot \ldots\ldots\ldots\ldots\ldots f_n(t) \quad \ldots\ldots \quad (1)$$

h'_t: dominant height growth at a stand age of t years

$f_i(t)$: the effect of growth factor no i at age t years

n: number of growth factors. Jonsson, 1969, distinguishes between

- exterior, non-climatic factors, for example soil properties
- exterior climatic factors
- interior factors, i e those processes within a tree which determine the size of the "answer" in terms of growth which is obtained at a given level of exterior factors.

In our case we want to describe h_{100}. Thus, (1) is integrated over age from 0 to 100 years.

$${}_0\!\int^{100} h'_t\,dt = h_{100} = {}_0\!\int^{100} f_1(t) \cdot f_2(t) \cdot \ldots\ldots\ldots f_n(t)dt \quad \ldots\ldots\ldots \quad (2)$$

Assume that $f_i(t)$, $i = 1, 2, \ldots\ldots, n$ is continuous and does not change sign on the interval 0 to 100 years. We use the mean value theorem for integrals on (2) and get

$$h_{100} = f_1(t_1) \cdot f_2(t_2) \cdot \ldots\ldots\ldots f_{n-1}(t_{n-1}) \cdot {}_0\!\int^{100} f_n(t)dt \quad \ldots\ldots\ldots \quad (3)$$

The ages $t_1, t_2, \ldots\ldots, t_{n-1}$ are situated on the interval 0—100 years and are chosen so that the right sides of (2) and (3) are equal. We cannot assume that these ages are the same for all trees or plots. One way to make (3) a useful development of (2), is to prescribe that the effects $f_1, f_2, \ldots\ldots, f_{n-1}$ must be independent of t. This means that the growth factors 1, 2, ..., n—1 must have the same values during the whole rotation. Site factors often fulfil this restriction, i e they are often reasonably time-independent. Further, the functions f must have the same shape during the whole rotation. The expression $f_n(t)$ stands for effects of interior and time-dependent exterior growth factors. As the later factors are not of interest in this investigation, we simply express their total effect during 100 years as the product of a mean level for all stands and a stochastic component per stand, $c \cdot (1 + \varepsilon)$. We summarize:

- the ages $t_1, t_2, \ldots\ldots, t_{n-1}$ are not of interest as the factors 1, 2, ..., n—1 are assumed to be independent of t. From now on we write $f_i(t_i)$ as f_i
- the integral of $f_n(t)$ is replaced by the product of a constant and a stochastic component, $c \cdot (1 + \varepsilon)$.

The model shall be used for many plots. An index of plot no, $j = 1, 2, \ldots\ldots, m$ is therefore introduced. We write:

$$h_{100}(j) = c \cdot f_1(j) \cdot f_2(j) \cdot \ldots\ldots\ldots f_{n-1}(j) \cdot (1 + \varepsilon(j)) \quad \ldots\ldots\ldots \quad (4)$$

Observe that the index (j) now stands for plot no. and not, as earlier, for age. Taking the logarithms of both sides in (4) we get

$$\ln h_{100}(j) = \ln c + \ln f_1(j) + \ldots + \ln f_{n-1}(j) + \ln(1 + \varepsilon(j)) \quad \ldots \quad (5)$$

If we restrict ourselves very much in formulating the effects f_i, it might be possible to fit some expression like (5) to data with regression analysis. However, a further development gives us a better chance of formulating the f_i:s in a proper way. Therefore we transfer the constant ln c to a new constant c_o, which comprises the mean effects of all factors involved in (5). Thus, the constant c_o expresses the mean level of logarithmic h_{100} for all stands. The effects f_i are transferred to expressions $(1 + g_i)$, where g_i describes deviations from the mean level mentioned. If $(1 + g_i)$ is reasonably close to 1, we might approximately write

$$\ln(1 + g_i) \approx g_i \quad \ldots \quad (6)$$

We assume that g_i is of such a magnitude for all i:s that (6) can be used.

At our existing level of knowledge, many of the (n—1) exterior growth factors are unknown and are only represented as mean levels in c_o. Hence, these factors cause an increase in the magnitude of the stochastic component. We therefore introduce a new stochastic component ε_o, which includes both $\ln(1 + \varepsilon)$ and effects of factors not taken into account. For simplicity, also measurement errors of $\ln(h_{100})$ are included in ε_o. The component ε_o is assumed to have the mean value 0, and to be normally distributed with constant variance. It must be pointed out that this schematic way of handling the stochastic component is a weak part of the modelling.

If the number of known growth factors included in the model is k we get

$$\ln h_{100}(j) = c_o + g_1(j) + \ldots + g_k(j) + \varepsilon_o(j) \quad \ldots \quad (7)$$

The model (7) is the analytical form used in this investigation. However, a difficult problem remains, namely to design the relations between the factors i and their effects g_i. These are here called "dose-response relations" and are dealt with in the next chapter.

5.2 **Dose-response relations**

The design of the dose-response relations is mainly founded on earlier studies by Lundmark (1966, 1967a, 1967b, 1974a, 1974b). In general, Lundmark found it difficult to create dose-response relations which are appropriate for all sites. Often the effects of specific site factors differ from site to site—there exist complicated interactions. In order to decrease the influence of this complication, the data set is divided in groups before analysis and the model is then fitted separately to each group. These "processing groups" are defined by tree species (dominant species on the plot), soil moisture and main forest type. On mineral soils eight groups are formed, on peat lands two, according to table 1 below.

Table 1. The processing groups

	Soil moisture class	Forest type	Scots pine	Norway Spruce
	very dry, dry (VD + D)	all	X	
Mineral soils	mesic (ME)	herb types, grass types, without field layer	X	X
		dwarf-shrub types	X	X
		lichen types	X	
	moist, somewhat waterlogged (MO, SW)	all	X	X
Peat lands			X	X

The description of various dose-response relations follows below. It is not complete as some less important relations are excluded here. The symbols used for different variables are explained in section 4.

5.2.1 *Forest type*

The forming of processing groups is, as mentioned above, founded on among other things main forest type. However, within these wide types, finer types indicating different nutrient status etc can be distinguished. These finer types are included in the model as "dummy" variables. Each forest type has been assigned a dummy variable which has the value 1 for plots which are classified as the type in question, otherwise 0. In the model, forest type is described as

$$c_1 \cdot \text{TYPE}$$

where c_1 is a vector of coefficients to be estimated at regression analysis. TYPE is a vector of dummy variables. The content of TYPE will of course differ between processing groups. On peat lands, TYPE contains field and bottom layers.

5.2.2 *Latitude*

Latitude is used together with altitude and some other variables to describe the geographical location of a plot and thus essential features of the temperature and light climate on the plot. Many authors have tried to compile the effects of geographical location in some type of "index of climate". Such an index might for example describe the length of the vegetation period (Langlet, 1936) or some type of heat sum (Hägglund, 1972). However, these attempts imply a schematizing and thus a loss of information. For this investigation, we found it most convenient to work with the different variables which describe geographical location separately.

Returning to latitude, Lundmark (1974a) found it necessary to separate between the effects of latitude in southern and northern Sweden. In the model, this is done with the following expression

$$c_{21} \cdot (\text{LAT}-60-\text{Abs}(\text{LAT}-60)) + c_{22} \cdot (\text{LAT}-60+\text{Abs}(\text{LAT}-60))$$

The symbols c_{21} and c_{22} stand (as all c:s) for coefficients. LAT is latitude, and Abs(x) means the absolute value of x. The coefficient c_{21} will express the effect of latitude south of 60°N, c_{22} north of 60°N. When latitude is exactly 60°N, the total expression is 0.

5.2.3 *Surface/subsurface water flow*

The positive effect of mobile soil water on forest growth has been demonstrated by Troedsson, 1965 and others. This is due to the supply of nutrients and oxygen to the soil, which follows the mobile water. According to section 4.4, the degree of surface/subsurface water flow is described in four classes. These are included in the model as three dummy variables.

$$c_{31}(\text{ABS}+\text{R}) + c_{32}\,\text{OC} + c_{33}\,\text{F}$$

5.2.4 *Altitude and distance to coast*

Lundmark, 1974a has shown that the effect of altitude on site index often varies with the intensity of the surface/subsurface water flow. Therefore, the effect of altitude in the model is differentiated for areas which are not, rarely or during shorter periods influenced by mobile ground water (ABS+R+OC) and for areas which are more regularly influenced (F). Further, Lundmark has noticed that the site index maximum does not always occur at sea level. Especially on dry and mesic sites, site index culminates at an altitude of 75—200 m. This was interpreted as an effect of deficiency in precipitation in areas close to the coast. To avoid this complication, we have for northern Sweden introduced a dummy variable DC in the model. This variable is 1 for areas more than 50 km from the coast, otherwise 0. The fairly complicated interaction between altitude, surface/subsurface water flow and distance to coast is in the model expressed as:

$$(\text{ABS}+\text{R}+\text{OC}) \cdot (c_{41} \cdot \text{DC} + c_{42} \cdot \text{DC} \cdot \text{ALT} + c_{43} \cdot \text{DC} \cdot \text{ALT}^2) + \text{F} \cdot (c_{44} \cdot \text{DC} + c_{45} \cdot \text{DC} \cdot \text{ALT} + c_{46} \cdot \text{DC} \cdot \text{ALT}^2)$$

The variables ABS, R, OC, F and DC are dummies.

$$
\begin{aligned}
\ln h_{100} = {} & C_0 + C_1 \cdot TYPE + \left[C_{21} \cdot (LAT-60+Abs(LAT-60)) + C_{22} \cdot (LAT-60-Abs(LAT-60)) \right] + \\
& \left[C_{31} \cdot (ABS+R) + C_{32} \cdot (OC) + C_{33} \cdot (F) \right] + \\
& \left[(ABS+R+OC) \cdot (C_{41} \cdot DC + C_{42} \cdot DC \cdot ALT + C_{43} \cdot DC \cdot ALT^2 + C_{61} \cdot TEX + C_{62} \cdot TEX^2) + \right. \\
& \left. (F) \cdot (C_{44} \cdot DC + C_{45} \cdot DC \cdot ALT + C_{46} \cdot DC \cdot ALT^2 + C_{63} \cdot TEX + C_{64} \cdot TEX^2) \right] \\
& + C_5 \cdot LOC + C_{71} \cdot (SOIL\ DEPTH) + 3_0
\end{aligned}
$$

Figure 3. A model describing the principal relationship between site index h_{100} and site properties.

5.2.5 *Climatic regions*

As mentioned in section 4.2, some regions of maritime or continental climate can be distinguished within Sweden. These are introduced in the model as a vector of dummy variables,

$c_5 \cdot LOC$

5.2.6 *Textural index*

The textural index is a measure of the mechanical composition of the soil and thus strongly correlated with water-retaining capacity and transmission. Lundmark (1974a) has shown that the textural index should be introduced in the model in such a way, that it is possible to differentiate the effect for different intensity of surface/subsurface water flow. The relationship between logarithmic site index and textural index might be written as a second-degree polynomial. Hence, we introduce the expression

$(ABS + R + OC) \cdot (c_{61} \cdot TEX + c_{62} \cdot TEX^2) +$
$F \cdot (c_{63} \cdot TEX + c_{64} \cdot TEX^2)$

5.2.7 *Depth of soil*

According to Lundmark, 1974a, depth of soil has a strong influence on site index. We therefore introduce depth of soil as a dummy variable, SOIL DEPTH, which is 1 for deep soils, 0 for fairly shallow. Very shallow soils and displacements are excluded from the investigation. We write

$c_{71} \cdot (SOIL\ DEPTH)$

5.2.8 *Other variables*

Besides the variables mentioned some others are introduced in the model, for example inclination, direction of slope and a humidity index. To keep the report short, we will not describe the dose-response relations developed for these less important variables here. Neither do we include these variables in the summarized description of the model which follows below.

5.3 Summarized description of the model

The modelling of the relation between site index and site properties which is reported here is based on an assumption that the effects of different growth factors work together in a multiplicative way. From this assumption a principal model is derived. As it is the effects (and not the factors) which are interacting in a multiplicative way, the relations between different growth factors and their effects ("dose-response" relations) must be described. The designs of different such relations are reported in sections 5.2.1. to 5.2.8. The different relations are combined according to the principal model (see figure 3). The model comprises a large number of variables and coefficients. However, the length of the function decreases when it is applied to specific processing groups, as all variables are not applicable within every group. As we see it, a weak part of the model is the description of the effects of the geological material in the soil. These effects are only expressed through the forest type. Especially on soils where the geological material changes along a vertical gradient, the model might be too simple. However, we have agreed that it must be possible to easily record all site properties used in the field. This restriction makes it very difficult to include the composition of the geological material, especially at greater depths of soil, in the investigation.

5.4 An alternative approach

The model presented is founded on a rough conception about the biological processes involved in the development of the dominant height of a forest stand. The model comprises many variables, some of which are probably strongly correlated. This will make the interpretation of the results difficult. To overcome those difficulties caused by correlation among independent vabiables, researchers, for example in USA and Canada, have worked with a combination of principal components analysis (PCA) and regression analysis. By means of PCA, linear combinations of site factors are created. These combinations remove a major part of the interior variance of the data set and are between themselves uncorrelated. They are used as independent variables in regression analysis. However, we think this approach has drawbacks, the most important one being that it is difficult to use a priori existing biological knowledge when designing the regression functions. The interpretation of the content of the linear combinations must be done after the analysis. Attempts to use the method (Graney, 1974, Page, 1976) has shown that it does not generally give a lower residual sum of squares than does a regression analysis based on original site data.

Even if the combination of PCA and regression analysis might have some advantages in those cases when it is possible to interpret the linear combinations in a reasonable way, we cannot find the method suitable for our investigation.

6 Data

6.1 The National Forest Survey

The main source of data for this study is the National Forest Survey (NFS). NFS is an objective sample from the Swedish forests, which is taken every year. The observation unit is a temporary circular plot with a radius of 10 m. The plots are situated in clusters, designed as squares with a side length of 1 000—1 600 m. Along each side of the square, 4—7 plots are placed. The sample is every year taken from the whole of Sweden, but the sampling intensity is higher in the southern part of the country than in the northern. The total number of plots laid out on forest land is 10 000—15 000 per year.

On the plots, a large number of stand and site properties are recorded, amongst them those mentioned in chapter 4. The soil data is recorded by specialists in soil mapping. However, soil characters are not regularly measured on every plot. During the NFS of the years 1973—1974 soil mapping was performed only in southern Sweden. In the NFS of 1975 all plots were soil-mapped. It is these soil-mapped plots from the years 1973—1975 that are used in this study. The total number of such plots is before selection about 15 000.

In 1975 two special field teams complemented the work of the ordinary NFS. One team worked on areas of high altitude, the other on peat lands. On these sites, the NFS provides very few plots where site index can be estimated in a reliable way. The special teams were instructed to concentrate their data collection to stands where site index could be safely estimated with site index curves. The rules for selecting these stands are dealt with in the next section.

Figure 4 shows a plot in the NFS with some of the more important variables used here marked.

6.2 Selection of sample plots

The selection of sample plots is an extremely important operation in an investigation of the type reported here. The aim of the selection is to provide a data set, comprising stands in such a condition that site index can be safely estimated with site index curves.

This means that stand history must not have strongly influenced dominant height growth. Of course stand history is primarily unknown in a survey material like the NFS. Therefore, we have to use stand data from the time of measurement as indicators of stand history. By combining several characters of the stand, we can select plots where the probability of undesired stand history becomes reasonably low. Thus, we have introduced the following restrictions on stand data

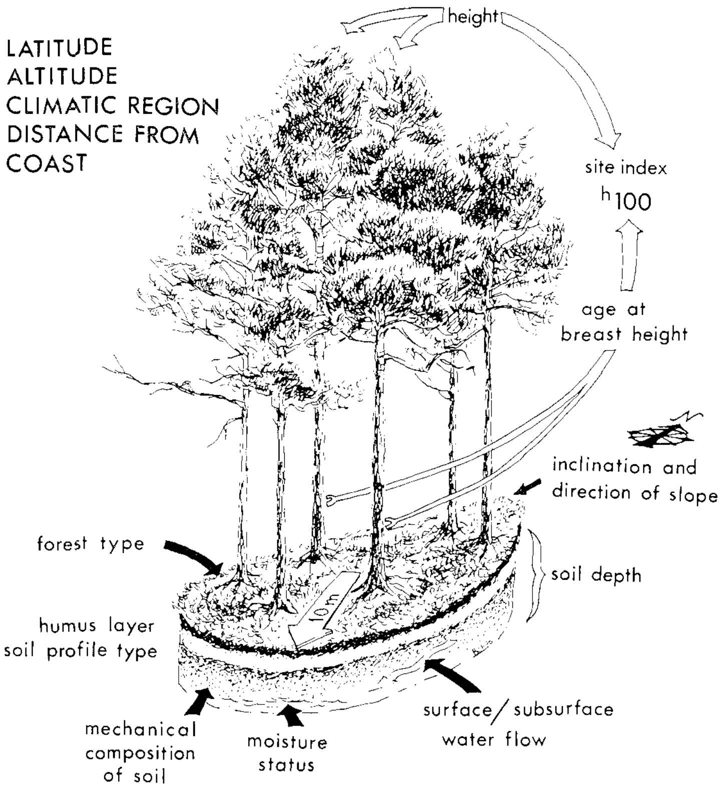
— the average age at breast height of the two "dominant height trees" on a plot shall be 20—150 years. If the age is lower than 20 years, the precision of the site index curves is too low (Hägglund, 1975). If the age is higher than 150 years, the probability of undesired stand history is too high. Furthermore, the site index curves are not applicable at ages over 150 years.

— the stand shall be classified as even-aged. As a test of this classification, we exclude plots where the difference between the ages of the dominant height trees exceeds 20 years.

— the density of the stand shall be at least 40 % of that in a fullystocked stand (In Swedish inventory work there are special conventions for estimating relative density).

— the stand shall not be classified as a residual (creamed) stand.

Figure 4. The National Forest Survey plot with some variables important in this investigation.



— the dominant height trees must not have any of the following damages: breakage, dead top, pitchy wood, damage by insect attacks.
— the dominant height must not be classified as "misleading".
— there must not have been any drainage of the site during the life of the existing stand.

Besides these restrictions concerning stand history, data has also been selected with respect to

— species composition. At least 70 % of the basal area on the plot must be either Scots pine or Norway spruce. Both the dominant height trees must be of the dominating species.
— recording errors. Ages, dominant heights, site indices and site properties are tested for obvious errors. Besides the normal control routines of the NFS, we also, for example, test the site index estimate by comparing it to an independent site classification according to Jonson, 1914.
— unlikely combinations of site properties. Some combinations of site properties are so unlikely or so rare that we have excluded them from the data set. For example, we exclude dry and very dry sites with swamp mosses.

When the restrictions were applied on the NFS data, 80 % of the data set was discarded. The most important restriction was the one concerning species composition, which resulted in excluding 45 % of the original data set. Many of these excluded plots did not however satisfy all the other restrictions.

It is hard to judge the consequences of the selections for the representativity of the remaining data set. Of course, the sorting out of uneven-aged stands for example,

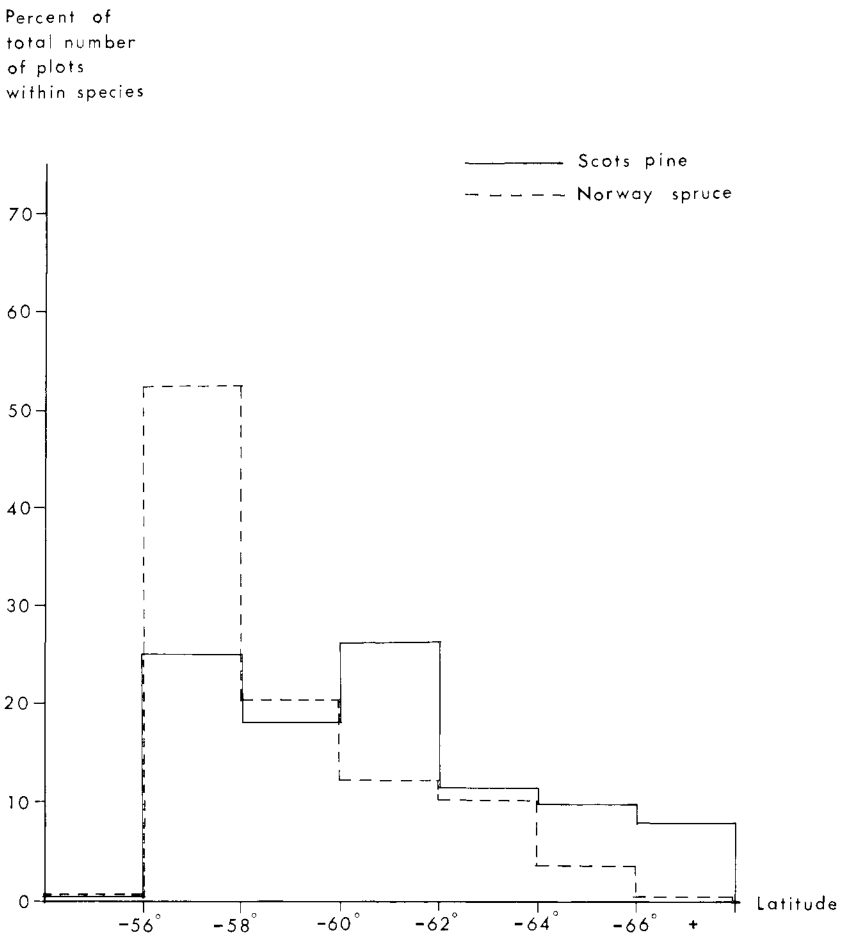


Figure 5. Distribution of data on latitude.

leads to an underrepresentation of sites where forest regeneration is difficult—on such sites stands are often uneven-aged. However, as long as the sites in question are represented in the data set and their properties are included in the model, the selection of uneven-aged stands does not introduce any bias in the calculations. Problems arise if we completely sort out some type of site or if there exist site properties which at the same time

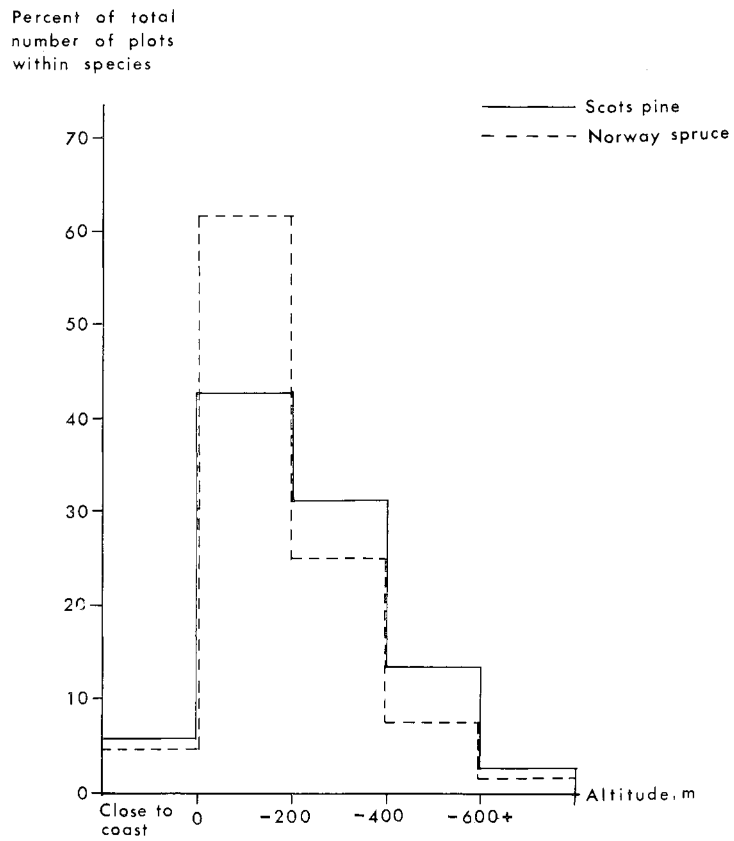
— influence site index
— are correlated with variables used for data selection
— are not represented in the model

If this is the situation, the average level of site index for a given combination of known site factors will be affected by the selection procedure. It is not possible to calculate the risk of errors of this type as the factors causing the possible error are unknown.

6.3 Description of data

The data set selected comprises 3 117 plots on mineral soils and 268 plots on peat lands. The low number of plots on peat creates special problems which are not reported here. In the rest of this summarized report we only consider mineral soils. On these, the number of Scots pine plots is 1 624 and

Figure 6. Distribution of data on altitude.



Norway spruce plots 1 493. The distributions of data on latitude and altitude are shown in figures 5 and 6. As can be seen, most of the areas of interest in this investigation are fairly well represented in the data.

Table 2 below shows the distribution of data on processing groups, the average site index and standard deviation per group. Most of the processing groups are well represented in the data set. However, the amount of data from Scots pine stands on moist and slightly waterlogged soils is not very large.

The values of average site index confirm that there are big differences between different processing groups.

6.4 **Accuracy of data**

The data used is subject to various errors. Some preliminary investigations have been made concerning the errors in the data for site index, soil properties and vegetation.

The errors in estimating site index are due to

— prediction errors. These errors are caused by influences making the actual dominant height development differ from the site index curves (Hägglund, 1975)
— height measurement errors (Eriksson, 1970)
— age measurement errors. The ages of the dominant height trees are in our data set measured by counting the growth rings in the field. On a fraction of the data, the growth rings have also been counted in a special machine, which gives very accurate estimates of age (Eklund, 1949). A comparison between the two ways of measuring age shows

Table 2. Number of plots, average site index (h_{100}) and standard deviation in percent of h_{100} for different processing groups.

Species	Soil moisture class	Forest type	No of plots	Average h_{100}, m	Standard deviation about average h_{100}, %
Scots pine	VD + D	all	225	18.0	26
	ME	Herb, Gr, N fl	251	22.5	17
		D.sh	679	19.5	20
		Lichens	302	16.1	18
	MO, SW	all	167	17.9	26
Norway spruce	ME	Herb, Gr, N fl	754	28.6	20
		D.sh	341	23.7	28
	MO, SW	all	398	24.0	30

Note: The different symbols are explained in sections 4.3 and 4.4.

that field-counting results in small, unsystematic errors. This result shall not be generalized beyond our well selected data set of coniferous trees.

None of the three errors of site index seems to be systematic. The magnitude of the errors is estimated as respectively 5.6 (prediction), 1.5 (height measurement) and 2.9 (age measurement) percent of h_{100}. The estimated errors are in terms of standard deviation per plot. The prediction errors are the most important. Adding the three errors assuming they are independent, we get a total error of h_{100} of 6.5 %. As we can make a reliable estimation of prediction errors for Scots pine only, the figure of total error is principally valid only for that species. However, we will use the estimated error for Norway spruce too.

The accuracy of soil data has been checked by means of a small-scale control mapping of 95 plots. On average, soil moisture and surface/subsurface water flow were one class mistaken on every fourth plot. It seems especially difficult to differentiate between sites where surface/subsurface water flow occurs rarely (R) and sites where the mobile water occurs during shorter periods (OC). Further, mesic and moist soils were often confused. Depth of soil was, on average, mistaken one class on every tenth plot. This error seems small, but it must be remembered that deep soils are by far the most common. It follows that even a small error might be of importance when considering the frequencies of, for example, shallow soils.

The description of vegetation was checked by means of the NFS control routine. Every year, a fraction of the inventoried plots is checked by special control teams. In 1975, more than 1 000 plots were checked with respect to description of vegetation. The check was made separately for field and bottom layers. The main result of the check is that there are often considerable differences between the results of ordinary teams and control teams. Of course, we cannot say for sure which (if any) of the teams is making the correct classification, but it seems reasonable to assume that big differences between the teams indicate big errors in classification.

Regarding the field layer, the fraction of plots which has been classified by the ordinary teams in the same way as by the control teams, varies between 79 % (grasses) and 42 % (Vaccinium vitis-idea). The corresponding frequencies for the bottom layer are from 85 % (no bottom layer) to 24 % (a combination of mesic-soil mosses and swamp mosses).

Summarizing, we must conclude that most of the data used is subject to considerable error. As for site index, h_{100}, this does not matter very much. The measures of precision obtained by regression analysis will indicate a somewhat too low precision, but this can be corrected. The errors of h_{100} will, as they seem to be random, not introduce any bias in the analysis. However, the errors of soil and vegetation data might cause bias according to the following reasoning.

We assume that the data set used to construct the functions is such, that the independent variables have the same distribution as in the data used at the application of the functions. As long as errors in independent variables are systematic and the same errors as in our data occur when collecting data for the application of the functions, in average no bias of predicted site index will occur. In this case the errors merely have the effect of "scaling" the observations. If the systematic errors in data are different when constructing and applying the functions, predicted site index will be biased. If the errors of independent variables are purely random, the coefficients of the regression functions will be absolutely too low (Madansky, 1959). It is difficult to judge the consequences of this "thinning out" of the regression functions, but it is likely that predicted site indices at extreme conditions will be biased towards more central parts of the original data set.

7 Regression analysis

The model (section 5) has been fitted to data (section 6) with regression analysis for each processing group separately. The main principle at this analysis was minimizing the residual sum of squares under the restriction that the relationships obtained should be realistic. This restriction needs some explanation.

The model comprises a large number of variables, many of them dummies. There is an evident risk of "over-fitting" the model, (see Gardner, 1972) and hence creating partial relationships which are more or less nonsense. In order to avoid this as far as possible, we have critically examined coefficients and partial relationships and compared them to the results of Lundmark's earlier studies and to common ecological knowledge. Comparisons between processing groups which logically should react the same way have also been valuable. Of course, correlations between independent variables sometimes make these examinations difficult. However, the described way of

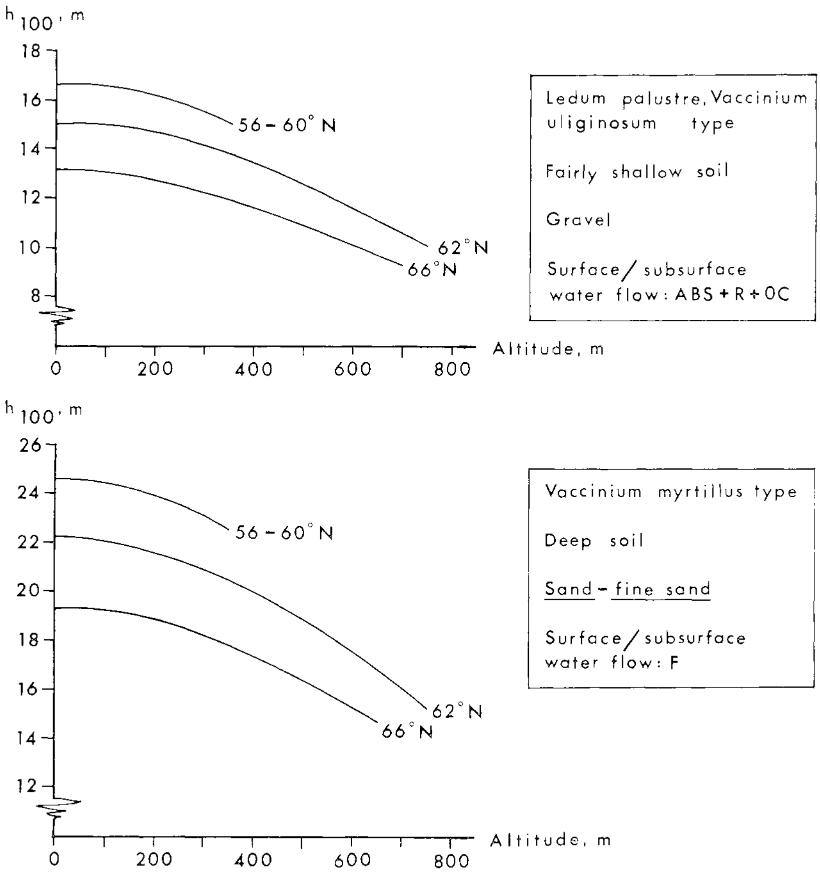
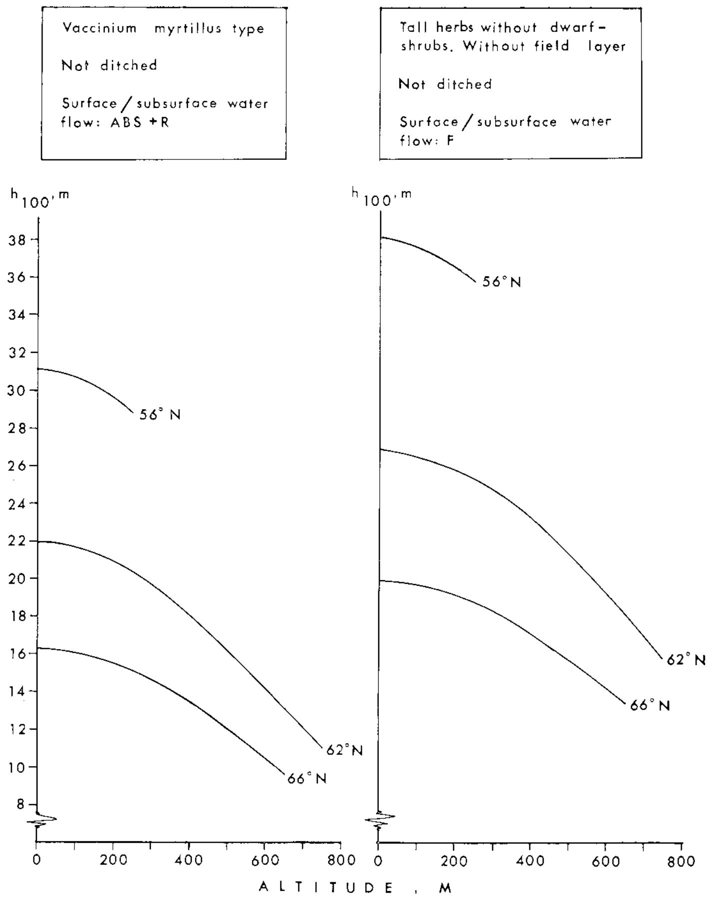


Figure 7. Scots pine on mesic soils, dwarf-shrub types. Some partial relationships.

Figure 8. Norway spruce on moist and slightly waterlogged soils. Some partial relationships.



working gave the result that none of our final functions is the one that gives the absolutely lowest sum of squares. Contributing to this is also that we tried to faciliate the practical use of the functions by neglecting variables of minor importance.

One statistical prerequisite for regression analysis is that the variance of the residuals should be constant. A preliminar analysis showed that this was not the case here. Therefore, we weighted the observations in such a way that high h_{100} got higher weights than low h_{100}. After weighting, the variance is reasonably constant.

It must be stressed that we work with a fairly small and rather inaccurate data set compared to the width of the problem under study. We cannot expect that all the relationships existing in the "real forest" are expressed by our functions. Many relations are hidden by what we call random variations. By means of the regression analysis, we get a rough and over-simplified description of some main features of an extremely complicated ecological system.

7.1 Some examples of the results

The results of the regression analysis are exemplified with the functions for two processing groups, Scots pine on mesic soils, dwarf-shrub types and Norway spruce on moist and slightly waterlogged soils, all forest types. The full quantitative results of the investigation are not reported here.

Table 3 shows the two functions obtained and in figures 7 and 8 some partial relationships are illustrated. As can be seen,

Table 3. Two examples of the results of regression analysis. Functions for Scots pine on mesic soils, dwarf-shrub types and for Norway spruce on moist and slightly waterlogged soils. Dependent variable: ln (h_{100}).

Independent variable	Scots pine		Norway spruce	
	Coefficient	Standard error, % of coefficient	Coefficient	Standard error, % of coefficient
Constant*	5.30943	—	5.59884	—
LAT—60 + Abs(LAT—60)	—0.01716	8	—0.03722	11
LAT—60 — Abs(LAT—60)	—0.00390	64	—0.02499	14
DC · (ABS + R + OC) · ALT^2/10 000	—0.00705	11	—0.01206	19
DC · F · ALT^2/10 000	—0.00651	16	—0.00937	20
OC			0.04766	42
F			0.05939	46
Deep soil	0.11580	15		
(Textural index)^2	—0.01243	18		
Ditched			0.02383	96
Tall herb types without dwarf-shrubs, without field layer			0.08075	30
Low herb types without dwarf-shrubs			0.05342	41
Herbs with dwarf-shrubs, grasses			0	—
Vaccinium myrtillus type	0.09429	22	—0.05889	33
Vaccinium vitis-idea type	0.06167	38		
Empetrum and Calluna types	0	—		
Ledum palustre, Vaccinium uliginosum type	—0.07775	76		
No. of observations	679		397	
Multiple correlation coefficient	0.62		0.81	
Standard deviation about the function	0.146		0.136	

* Corrected for logarithmic bias

some of the variables in the model in section 5 are excluded from the final functions. This is because the analysis indicated that they were of minor importance.

Because of the existence of correlations between independent variables, the interpretation of partial relationships must be done with care. In spite of this, we think it is possible to draw the following conclusions from the total result of the regression analysis.

— the effects of latitude, altitude and mobile soil water on site index are greater for Norway spruce than for Scots pine
— the more moist the soil, the stronger is the positive influence of mobile soil water
— the negative effect of altitude is greater if there is no regular flow of mobile soil water than if such a flow exists
— within a soil moisture class, the effect of mobile water is greater the less fertile the forest type is
— the climatic regions have, in some cases, strong influences on site index. The continental region K3 in the central part of southern Sweden has a positive effect on site index for Scots pine on very dry and dry soils. The maritime region M2 along the eastern coast of southern

Sweden has a negative effect on site index for both Scots pine and Norway spruce on mesic soils when the forest type is herbs, grasses or grounds without field layer
— deep soils have a positive effect on site index for Scots pine on very dry, dry and mesic soils
— textural index influences site index in most Scots pine processing groups. The coarser the soil, the lower is site index
— the influence of forest type is mainly as expected. This means that herb types, grass types and grounds without field layer are better than dwarf-shrub types which are better than lichen types.

Some comments on the forest types follow below. Herb types without dwarf-shrubs mostly give higher site index than herb types with dwarf-shrubs. When dwarf-shrubs occur within the herb types, either all dwarf-shrubs have the same effect or Vaccinium myrtillus is the best. There is also a tendency that tall herbs are better than low herbs. Grass types give approximately the same site index as the less fertile herb types. Grounds without field layer are for Scots pine mostly inferior to the herb types, while for Norway spruce they are equivalent to the best herb types.

Among the dwarf-shrub types, the Vaccinium myrtillus type is generally the best, usually followed by the Vaccinium vitis-idea type. Lack of data makes it difficult to rank the less fertile dwarf-shrub types.

Equisetum and Carex types are generally rather infertile, but we have not data enough to be sure that this is the case for Norway spruce.

7.2 The examination of residuals

The residuals from the functions have been examined mainly for systematic trends over different site properties. Among these properties are some which are included as variables in the functions and some which are not. The general conclusion from these studies is that there are no important deviations between the functions and the data set used to produce them. No systematic trends occur when the residuals are examined over variables included in the functions. However, for one variable left outside the functions—the thickness of humus layer—some trends can be seen. The main reason for not including this variable in the functions is that it is not uniformly recorded within the whole of the data set. There also in some cases might exist a positive correlation between thickness of humus layer and stand age which make this variable less suitable as a description of site conditions.

So far only residuals in relation to site properties have been discussed. However, we have also examined the residuals over a stand property, namely the average age at breast height of the two dominant height trees. Here, a very strong trend appears, which is such that the functions underestimate site index at low ages and overestimate it at high ages. The trend is illustrated for two processing groups in figure 9. The residuals in the figure are standardized, i e they are divided by the standard deviations around the functions.

One possible reason for this trend is that the site index curves used to estimate site index are extremely biased. However, these curves were thoroughly checked for systematic errors (Hägglund, 1972, 1973, 1974a) and no bias was detected that was of such a magnitude as could cause the trends in figure 9. Thus, it is not plausible that the reason for the trend is bias of the site index curves.

Another possible explanation is that nutrients are stored in the old, often low-stocked stands. After clear-cutting and regeneration these accumulated nutrients could be used by the new stands, which in this way would get a rapid start and indicate high site indices. But this explanation does not seem very likely as the trend appears over the whole age scale and for all forest types. If this explanation was correct the trends should mainly occur at low ages and with forest types indicating a good nutrient status.

In our opinion, the trend over age is mainly an effect of the fact that the NFS

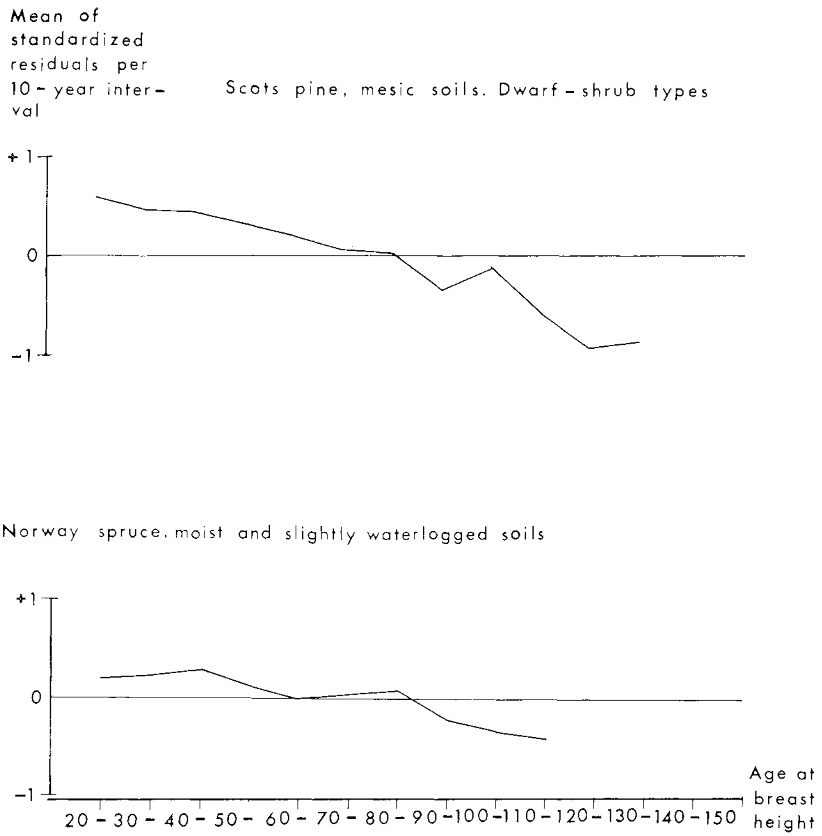


Figure 9. Standardized residuals over age at breast height for two processing groups.

data have a skewed distribution on the age/site index plane. This is a common phenomenon in most inventory data and is mainly a result of the silvicultural system. Stands on good sites are generally cut at lower ages than stands on poorer sites and thus average site index in inventory data decreases with increasing stand age. In the functions produced here, a considerable part of the variation of site index is left unexplained. This unexplained variation is mainly due to the influences of factors not included in the functions. An explanation to the trends of the residuals is that these factors mainly work the same way as the site index. This means that factors generating high site indices are more common in young stands and vice versa. Let us take an example. A typical factor of the type discussed is the genetical constitution of the stand. Stands with genetically good trees will indicate higher site indices, will grow faster and will be cut earlier than stands with bad genetical constitution. Hence the average age of the genetically good stands will be lower than the age of the bad stands. As the functions do not include any factor describing genetical constitution, they will underestimate site index in the genetically good stands and vice versa. This will cause a trend of the type showed in figure 9. The same reasoning can be applied to many other factors. However, we readily admit that this explanation is not empirically proved and thus has the character of an unconfirmed hypothesis. Here we accept this hypothesis and will therefore not introduce any corrections for stand age in the functions.

7.3 A check with new data

The examination of the residuals showed that the functions did not seem seriously biased when compared to the data used to produce them. However, this type of check does not say anything about the representativity of the selected NFS data set. To check for representativity, we have to use new data, not already used to produce the functions. One such set of new data available for us is the data from the "Great Yield Investigation" (GYI). This investigation (Näslund, 1971) was performed during the years 1940—1965 and produced 2 075 temporary plots, each with an area of 0.05 to 0.20 ha. Data from these plots has, for example, been used to construct the site index curves used in this study and to relate site index to site properties in some earlier investigations by Lundmark (1966, 1967a, 1967b).

The GYI data was considered suitable for checking the functions produced here. For every GYI plot used, we calculated site index with site index curves and with the functions. The differences between these two estimates of h_{100}, Δh_{100}, were examined.

From the total GYI data set, 513 plots in Scots pine and 296 plots in Norway spruce were selected for the check. The criterions at the selection concerned mainly species composition and age (the stands on the plots should be within the application range of the site index curves). Thus, the selection is not as strict as for the NFS data (section 6). However, the GYI plots were subjectively laid out in such a way that the need of strict selection is not as apparent as for the NFS data. Much of the selection work was done in the field. For example, the rules for the field work stipulate that the stands within the plots should be homogeneous and undamaged and that the plots should be placed on homogeneous sites. In most cases the selected GYI plots are in such condition that site index curves could be safely used to estimate site index. A few uneven-aged stands of Norway spruce however, are included in the control data.

Table 4 shows calculated average Δh_{100} for different processing groups.

Table 4. Check with data from the GYI. Δh_{100} is the difference between h_{100} estimated with site index curves and h_{100} estimated with the functions constructed in this investigation. Positive Δh_{100} indicate that the latter functions underestimate h_{100} and vice versa.

Species	Soil moisture	Forest type	No. of GYI plots	Average Δh_{100}, m	Standard deviation about average Δh_{100}, m
Scots pine	VD + D	all	128	—1.08***	2.35
	ME	Herb. Gr, N fl	23	0.99***	1.27
		D.sh	205	—0.39**	1.94
		Lichen	66	0.83***	1.94
	MO, SW	all	91	2.17***	2.04
Scots pine	Mean		513	0.25*	2.31
Norway spruce	ME	Herb. Gr, N fl	26	—2.58***	3.33
		D.sh	52	—1.41***	2.31
	MO, SW	all	218	0.85***	3.14
Norway spruce	Mean		296	0.15°	3.25
TOTAL			809	0.21*	2.69

Note: The different symbols are explained in sections 4.3 and 4.4.

For all processing groups, Δh_{100} differs significantly from 0 on the 1 % or 0.1 % level. However, average Δh_{100} change sign between groups. For two groups, average Δh_{100} are absolutely remarkably greater than for the other. For one of these two groups—Norway spruce on mesic soils, herb types, grass types and grounds without field layer—the function is founded on a large amount of NFS data. The GYI data set is small for that group and might include some uneven-aged plots. In this case it is more likely that the GYI data is biased than that the function is. The second group where a great average Δh_{100} occurs is Scots pine on moist and somewhat waterlogged soils. The function for this group is founded on relatively few NFS plots. The standard deviation around the predicted h_{100} is great. Here, it is not unlikely that the indicated systematic error is a real one.

As can be seen, Δh_{100} increases with increasing soil moisture. The functions overestimate site index on dry soils and underestimate on moist and slightly waterlogged soils. The same tendency was found by Lundmark, 1974a in a similar check. According to Lundmark, the explanation might be that small local variations of the surface/subsurface water flow are more important the more moist the soil is. Thus, local fertility variations increase with increasing soil moisture. The way of laying out plots in the GYI means that small spots of low fertility are systematically avoided. These spots are represented in the NFS data and might decrease site index predicted from site properties.

When the processing groups are pooled to larger units, the average Δh_{100} decrease. Thus, average Δh_{100} for Scots pine is 0.25 m, for Norway spruce 0.15 m and for the whole GYI data set 0.21 m. The functions constructed from NFS data seem, on average, to predict a correct level of site index. Therefore, we find the outcome of the check fairly satisfactory. We do not preclude the possibility of some of the differences noticed being of practical importance. However, as the differences are not unreasonably large, and the GYI data has some weaknesses, we do not believe that we could make the functions better by correcting them with respect to the outcome of the check.

7.4 The accuracy of the functions

The accuracy of the prediction of h_{100} from site properties has been studied mainly by means of the residual sum of squares remaining after the forming of processing groups and regression analysis. By means of grouping and regression, the sum of squares around ln (h_{100}) is reduced 49 % for Scots pine and 68 % for Norway spruce. Hence, the "multiple correlation coefficient" on species level is 0.70 and 0.82 respectively. The pooled standard deviation about predicted ln (h_{100}) is, for Scots pine, 0.1663 and for Norway spruce 0.1489.

The main sources of error which determine the magnitude of the residual sum of squares are

- there exist more factors influencing site index than those included in the functions
- those factors included in the functions might be introduced in an unsuitable analytical form. The examination of residuals indicate that this is not an important source of error
- the independent variables are estimated with errors
- the dependent variable, h_{100}, is estimated with error

The first three sources of error remain when the functions are applied. The random error originating from estimating h_{100} will, however, disappear. In section 6.4. this error was estimated to 6.5 % of h_{100}. The estimated standard deviations in ln (h_{100}) predicted with site properties were thus reduced by 0.065. The resulting standard deviations are, for Scots pine, 0.1531, for Norway spruce 0.1340. Transforming the standard deviations in terms of h_{100}, we get, for Scots pine 3.0 m and for Norway spruce 3.6 m. The corresponding figures can also be calculated from the GYI data. In this case the standard deviations (standard deviation of Δh_{100} around 0) were 2.1 m for Scots pine and 3.0 m for Norway spruce after reduction for errors originating from estimating h_{100} with site index curves. The differences in calculated standard deviations between the NFS and the GYI data are probably due to the GYI data set not containing those more extreme site conditions which increase the standard deviation in the NFS data. For practical conditions, we think the NFS data gives a realistic measure of the accuracy of estimating site index from site properties.

8 Discussion

The functions presented here are tools for predicting site index and are not primarily intended for causal interpretation. This restriction is important as it strongly influences the choice of independent variables. Even with a predictive approach, it is of course very important to select variables from a basis of existing ecological knowledge and to model the relationships in such a way that different factors interact in a realistic way. We think it is dangerous to use some very simple "trial-and error" additive model and then just let the variables step into the equations following some computer program for stepwise regression analysis.

The approach used here, to relate site index to site properties, has often been criticized in literature (see for example Daubenmire, 1976, who summarizes many viewpoints on this topic). The criticism mainly concerns the estimation of site index with site index curves. Many authors have doubted that one single set of site index curves for a species is capable of correctly describing height development for all combinations of site factors. In the case of the polymorphic sets of site index curves used here, they have been tested for different shape according to site properties (Hägglund, 1972, 1973, 1974a). Those factors considered were latitude, altitude, forest type, moisture of soil and surface/subsurface water flow. Only factor combinations within Sweden were investigated. Analysis was performed either by grouping the data for the curves according to site conditions or by introducing site properties as variables in the functional relationships between height and age which form the curves. The only site factor which was found to be of any importance was latitude for Norway spruce. The site index curves for this species therefore are somewhat different for southern and for northern Sweden. Further there exist different sets of curves for different latitudes in northern Sweden. As for Scots pine, this species seems to act in a very homogeneous way within Sweden. Our conclusion is that the site index curves are so thoroughly tested that the risk of bias for certain combinations of site factors is low. This conclusion is of course only valid within Sweden.

Another point of criticism to the approach used here is that site index is strongly influenced by stand history and thus is not a reliable measure of site productivity. We think this objection is important and we gave it fairly strong emphasis in section 6.2. We roughly solved this problem by selecting the data carefully, using the properties of the existing stand as indicators of stand history.

Looking at the independent variables used here, it is evident that our simple way of recording, for example, soil moisture and surface/subsurface water flow does not give absolutely "true" measures of the variables in question. However, the variables recorded are strongly related to site index and they are recorded in the same way when the functions are applied for predicting h_{100}. Thus, our way of recording these variables does not cause any serious problems—it is more a question of semantics than of statistics. The assessment of forest type is more questionable. As stated before, the forest type used here is a combination of ground and field layers, defined from purely botanical aspects. The composition of ground and field layers changes according to the characteristics of the existing stand. For example, on clear-cut areas, in young stands and in low-stocked older stands there will be higher frequencies of grasses than in the wellstocked mature stand. Some other changes might also occur, due to among other things the species composition of the stand. However, we think that the high frequencies of grasses which might occur occasionally is the most severe problem. Especially for dwarf-shrub types and some less fertile herb types, a confusion with grass types is often likely. Trying to quantify the problem, we found it most important on moist and slightly waterlogged soils where a confusion between the types mentioned in extreme cases might cause an error of predicted h_{100} of the magnitude 4 m. More research on vegetation dynamics is needed to get reliable assessments of forest type in all situations.

Compared to other similar studies, we have used a fairly large amount of data—more than 3 000 plots to construct the functions and almost 1 000 plots to check them. Still we think our data set is in some cases too small. When working with the analysis, it is evident that functions for processing groups which are well represented in the data set show fewer peculiarities and in many ways react more in accordance with basic ecological knowledge than functions for groups which are not so well represented. Of course, this is not only depending on the amount of data. A careful stratification of plots, measurements of high quality and so on might well be a more efficient way of increasing the reliability of the site index predictions than simply collecting more data. In our case, the data originates from a national inventory, the primary aim of which is quite different from getting a data set suitable for site index/site properties studies.

The accuracy of the prediction of h_{100} with site properties is considerably lower then when using site index curves (Hägglund, 1975) or the intercept method (Hägglund, 1976). However, the use of the latter methods are linked to such restrictions on the condition of the existing stand, that they are not valid for large areas. Our conclusion is that a method to estimate site index from site properties is a necessary part of a site evaluation system, but the use of the method shall be restricted to areas where methods using the properties of the existing stand can not be applied.

9 Acknowledgements

The data used in this investigation originates from the National Forest Survey, which is performed by the Department of Forest Survey at the Royal College of Forestry, Sweden. The work was conducted by professors Nils-Erik Nilsson and Gustaf von Segebaden. The soil-mapping of the plots was performed by the Department of Forest Ecology and Forest Soils under the management of professor Tryggve Troedsson. A major part of the mapping was financed by the Swedish Pulp and Paper Association.

The manuscript to this report has been read by professors Jöran Fries, Bertil Matérn and Carl Olof Tamm, all of the Royal College of Forestry. These readings resulted in many valuable comments and proposals, which have been incorporated in the final report. Mr Gordon Evans has scrutinized and corrected the English text. We wish to express our gratitude to those who helped ut to perform this investigation.

Stockholm and Uppsala, February 1977

Björn Hägglund *Jan-Erik Lundmark*

10 Sammanfattning

Höjdboniteten h_{100} är den övre höjd ett bestånd uppnår vid 100 års total ålder. Den används som ett samlat mått på ståndortens bördighet. Höjdboniteten skattas i dag vanligen med hjälp av höjdutvecklingskurvor, dvs. med samband mellan ett bestånds övre höjd och dess ålder i brösthöjd. Dessa kurvor bygger emellertid på en lång rad förutsättningar rörande beståndets tillstånd, och är strängt taget inte tillämpbara på mer än 30—40 % av Sveriges skogsmarksareal. Det har därför bedömts vara angeläget att utveckla en kompletterande metod för skattning av höjdboniteten, som är oberoende av det växande beståndets egenskaper. Detta kan åstadkommas genom att man använder egenskaper hos ståndorten som bonitetsindikatorer.

Syftet med den här genomförda undersökningen är att åstadkomma funktioner som uttrycker sambandet mellan höjdboniteten h_{100} och lätt mätbara ståndortsegenskaper. Funktionerna, vilka avser h_{100} för tall och gran, skall vara användbara för bonitering i det praktiska skogsbruket.

Materialet till undersökningen utgörs av observationer från riksskogstaxeringens markkarterade provytor. Endast sådana provytor utvaldes där det växande beståndets tillstånd tillät att höjdboniteten skattades med höjdutvecklingskurvor. Urvalet blev tämligen hårt, och i stort bortsorterades 80 % av det ursprungliga materialet från riksskogstaxeringen. Det kvarvarande, för undersökningen användbara, materialet består av 3 117 provytor på fastmark och 268 provytor på torvmark. Huvuddelen av undersökningen ägnades fastmarkerna, medan torvmarkerna endast blivit föremål för en orienterande studie.

Bland de utvalda provytorna på fastmark ligger 1 624 i tallbestånd och 1 493 i granbestånd. Ytorna är tämligen väl fördelade på breddgrad, höjd över havet och h_{100}. Före den egentliga analysen av materialet indelades det i åtta bearbetningsenheter, definierade av trädslag, markfuktighet och skogstypsgrupp. Bearbetningsenheterna analyserades var för sig. Motivet för detta förfarande var att olika ståndortsegenskaper kan förväntas inverka olika på höjdboniteten inom olika bearbetningsenheter. De största enheterna är tall på frisk mark, ristyper (679 ytor) och gran på frisk mark, örttyper, grästyper och marker utan fältskikt (754 ytor). Den minsta enheten är tall på frisk-fuktig och något vattensjuk mark (167 ytor).

En studie av tillförlitligheten hos insamlade data visade bl a att de variabler som beskriver markens egenskaper och markvegetationens sammansättning är behäftade med en icke försumbar osäkerhet.

En modell för sambandet mellan höjdboniteten h_{100} och olika ståndortsegenskaper byggdes upp. Denna är i princip en integrerad tillväxtmodell och förutsätter att effekterna av olika tillväxtfaktorer samverkar multiplikativt. Bland modellens egenskaper kan nämnas att breddgradens effekt på h_{100} kan bli olika i södra och norra Sverige och att effekten av höjden över havet differentieras med hänsyn till förekomsten av rörligt markvatten (över-genomsilning).

Modellen anpassades med regressionsanalys till data för var och en av de åtta bearbetningsenheterna med fastmarksprovytor. De viktigaste slutsatserna från dessa analyser var följande:

— breddgradens inverkan på h_{100} är betydligt starkare för gran än för tall

— effekten av höjden över havet är starkare för gran än för tall. Vidare blir effekten av denna variabel större om rörligt markvatten saknas än om sådant förekommer

- ju fuktigare marken är, desto kraftigare påverkas h_{100} av rörligt markvatten
- inom varje markfuktighetsklass blir effekten av rörligt markvatten större ju sämre skogstypen är
- det lokalkontinentala klimatet inom område K3 i centrala Småland har en stark positiv effekt på h_{100} för tall på mycket torr och torr mark
- det lokalmaritima klimatet inom område M2 i södra Sverige (ostkusten, Öland, Gotland) har stark negativ effekt för både tall och gran på frisk mark när skogstypen är örttyp, grästyp eller mark utan fältskikt
- jorddjupet har påtaglig inverkan för tall på mycket torr, torr och frisk mark där h_{100} är ca 10 % lägre när jorddjupet är tämligen grunt jämfört med när det är mäktigt
- texturindex (mått på jordartens kornstorlek) påverkar h_{100} för tall inom vissa bearbetningsenheter på så sätt att h_{100} sjunker med stigande kornstorlek
- skogstyperna ger i stora drag förväntade utslag i höjdboniteten, dvs. de kan rangordnas efter sjunkande h_{100} på följande sätt: Örttyp utan ris, örttyp med ris, blåbärstyp, lingontyp, övriga ristyper, lavrik typ och lavtyp. Marker utan fältskikt är för tall jämförbara med de svagaste örttyperna, för gran med de bästa. Grästyperna ger ungefär samma effekt som de svagaste örttyperna. Bland örttyperna finns en tendens att högörttyper ger högre h_{100} än lågörttyper. Starr- och fräkentyperna är sannolikt mycket svaga, men åtminstone för gran saknas data för att belägga detta.

Funktionerna för fastmarker har kontrollerats dels med residualstudier mot det använda markkarterade riksskogstaxeringsmaterialet, dels mot data från den s k stora produktionsundersökningen. Residualstudierna indikerade inga avvikelser som krävde åtgärder i form av korrektioner e d. En systematisk gång i residualerna kunde dock noteras över humuslagrets tjocklek. Funktionerna underskattar h_{100} något när humuslagret är tunt och överskattar när det är tjockt. Av flera skäl har vi dock avstått från att föra in humuslagrets tjocklek i funktionerna. Vidare finns en kraftig systematisk gång i residualerna när de läggs upp över beståndets ålder. I unga bestånd underskattar funktionerna h_{100}, i gamla överskattar de. Detta är sannolikt en följd av att taxeringsmaterialet är skevt fördelat på bonitet/ålder-planet, men detta saknar betydelse för funktionernas tillämpbarhet.

Kontrollen mot stora produktionsundersökningens material omfattade 809 provytor, vilka ej använts för att framställa funktionerna. För varje yta beräknades h_{100} dels med höjdutvecklingskurvor och dels med de här framtagna funktionerna. Skillnaden mellan dessa två skattningar av h_{100} studerades. I stora drag gav denna viktiga kontroll tillfredsställande resultat. Visserligen uppträder signifikanta avvikelser i samtliga bearbetningsenheter, men dessa jämnar ut sig när grupperna slås samman. Sålunda gäller i genomsnitt att funktionerna underskattar h_{100} för tall med 0.25 m och för gran med 0.15 m. Dessa skillnader är så små att de knappast har någon praktisk betydelse. Skillnader av betydande storleksordning eller med systematisk gång över vissa variabler förekommer dock speciellt för frisk-fuktiga och något vattensjuka marker. Även om detta delvis kan förklaras som en följd av grundläggande skillnader mellan materialen kan det inte uteslutas att funktionerna är behäftade med systematiska fel för dessa markfuktighetsklasser. Även för gran på frisk mark med örttyper, grästyper och "utan fältskikt" noterades stora skillnader, men dessa beror sannolikt på brister i kontrollmaterialet.

Funktionernas noggrannhet är betydligt lägre än den noggrannhet som erhålls när h_{100} skattas med höjdutvecklingskurvor. I medeltal är standardavvikelsen kring skattat h_{100} ca 3—3.5 m.

En fullständig, svenskspråkig redogörelse för den här sammanfattade undersökningen har publicerats som rapport nr 28 från institutionen för växtekologi och marklära vid skogshögskolan.

11 References

Baule, B. 1917. Zu Mitscherlichs Gesetz der physiologischen Beziehungen — Landw Jahrbuch, no 51, page 363—385

Daubenmire, R. 1976. The use of vegetation in assessing the productivity of forest lands — The Botanical Review, vol 42, no 2, page 115—143

Eklund, B. 1949. Skogsforskningsinstitutets årsringsmätningsmaskiner. Deras tillkomst, konstruktion och användning — Medd. från statens skogsforskningsinstitut, vol 38, no 5, page 1—77

Eriksson, H. 1970. On measuring errors in tree height determination with different altimeters — Inst för skogsproduktion, skogshögskolan. Rapporter och uppsatser no 19

Fries, J. 1974. Practical site classification — Sveriges skogsvårdsförbunds tidskrift, no 5—6

Gardner, M. J. 1972. On using an estimated regression line in a second sample — Biometrika, no 59, page 263—274

Graney, D. L. 1974. Site index predictions for shortleaf pine and oak in the mountains of Arkansas and Oklahoma. A comparison of principal components analysis and multiple regression techniques — Virginia polytechnic institute and state university

Hägglund, B. 1972. Site index curves for Norway spruce in northern Sweden — Inst för skogsproduktion, skogshögskolan. Rapporter och uppsatser no 21

— 1973. Site index curves for Norway spruce in southern Sweden — Inst för skogsproduktion, skogshögskolan. Rapporter och uppsatser no 24

— 1974 a. Site index curves for Scots pine in Sweden — Inst för skogsproduktion skogshögskolan. Rapporter och uppsatser no 31

— 1974 b. Bonitering. Föredrag vid konferensen "Framtidsskogen" — Inst för skogsproduktion, skogshögskolan. Rapporter och uppsatser no 33, page 204—213

— 1975. Estimating the accuracy of site index curves by means of simulation — Studia Forestalia Suecica, no 129

— 1976. Estimating site index in young stands of Scots pine and Norway spruce in Sweden — Inst för skogsproduktion, skogshögskolan. Rapporter och uppsatser no 39

Hägglund, B. & Lundmark, J.-E. 1977. Skattning av höjdboniteten med ståndortsfaktorer. Tall och gran i Sverige — Inst för växtekologi och marklära, skogshögskolan. Rapporter och uppsatser no 28

Jonson, T. 1914. Om bonitering av skogsmark — Svenska Skogsvårdsföreningens tidskrift, no 12, page 369—392

Jonsson, B. 1969. Studies of variations in the widths of annual rings in Scots pine and Norway spruce due to weather conditions in Sweden — Inst för skogsproduktion, skogshögskolan. Rapporter och uppsatser no 16

Langlet, O. 1936. Studier över tallens fysiologiska variabilitet och dess samband med klimatet — Medd. från statens skogsförsöksanstalt, vol 29, no 16

Lundmark, J.-E. 1966. Olika ståndortsfaktorers inverkan på skogsmarkens bonitet — Sveriges Skogsvårdsförbunds Tidskrift, no 6

— 1967 a. Höjdbonitetens beroende av några ståndortsfaktorer — Inst för växtekologi och marklära, skogshögskolan. Stencil

— 1967 b. Bestämning av höjdboniteten med hjälp av ståndortsfaktorer. Symposium angående bonitering den 30/11 1967 vid skogshögskolan. Stencil

— 1974 a. Use of site properties for assessing site index in stands of Scots pine and Norway spruce — Inst för växtekologi och marklära, skogshögskolan. Rapporter och uppsatser no 16

— 1974 b. Bonitetsklassificering med hjälp av ståndortsdata. Föredrag vid konferensen "Framtidsskogen" — Inst för skogsproduktion, skogshögskolan. Rapporter och uppsatser no 33, page 243—253

Madansky, A. 1959. The fitting of straight lines when both variables are subject to error — J Am Stat Ass, 54

Malmström, C. 1947. Appendix to SOU 1947: 36. Redogörelse avgiven av 1937 års riksskogstaxeringsnämnd — Stockholm

Näslund, M. 1971. New material for forest yield research — Studia Forestalia Suecica, no 89

Page, G. 1976. Quantitative evaluation of site potential for spruce and fir in New Foundland — Forest Science, vol 22, no 2, page 131—143

Tamm, C.-O., & Holmen, H. 1961. Klassifikationsschema för skogstyper i barrskog och blandskog — Inst för växtekologi och marklära, skogshögskolan. Stencil.

Tamm, C.-O., Troedsson, T., Lundmark, J.-E., & Persson, O. 1967. Forecasting forest yield from observations of site characteristics. A critical discussion — XIV IUFRO-Congress, Referate 2, München

Troedsson, T. 1965. The hydrology of forest soils — Kungl Skogs- och Lantbruksakademiens Tidskrift, vol 104, page 142—154

Ångström, A. 1958. Sveriges klimat — Generalstabens litografiska anstalt, Stockholm.