Methods and applications for improving parameter prediction models for stand structures in Finland

Jouni Siipilehto

Department of Forest Sciences
Faculty of Agriculture and Forestry
University of Helsinki

Academic dissertation

To be presented, with the permission of the Faculty of Agriculture and Forestry of the University of Helsinki, for public critisism in Lecture Hall B5, Building of Forest Sciences, on September 30th, 2011 at 12 o'clock.

ABSTRACT

This thesis report attempts to improve the models for predicting forest stand structure for practical use, e.g. forest management planning (FMP) purposes in Finland. Comparisons were made between Weibull and Johnson's SB distribution and alternative regression estimation methods. Data used for preliminary studies was local but the final models were based on representative data. Models were validated mainly in terms of bias and RMSE in the main stand characteristics (e.g. volume) using independent data.

The bivariate SBB distribution model was used to mimic realistic variations in tree dimensions by including within-diameter-class height variation. Using the traditional method, diameter distribution with the expected height resulted in reduced height variation, whereas the alternative bivariate method utilized the error-term of the height model. The lack of models for FMP was covered to some extent by the models for peatland and juvenile stands. The validation of these models showed that the more sophisticated regression estimation methods provided slightly improved accuracy.

A flexible prediction and application for stand structure consisted of seemingly unrelated regression models for eight stand characteristics, the parameters of three optional distributions and Näslund's height curve. The cross-model covariance structure was used for linear prediction application, in which the expected values of the models were calibrated with the known stand characteristics. This provided a framework to validate the optional distributions and the optional set of stand characteristics. Height distribution is recommended for the earliest state of stands because of its continuous feature. From the mean height of about 4 m, Weibull dbh-frequency distribution is recommended in young stands if the input variables consist of arithmetic stand characteristics. In advanced stands, basal area-dbh distribution models are recommended. Näslund's height curve proved useful. Some efficient transformations of stand characteristics are introduced, e.g. the shape index, which combined the basal area, the stem number and the median diameter. Shape index enabled SB model for peatland stands to detect large variation in stand densities. This model also demonstrated reasonable behaviour for stands in mineral soils.

ACKNOWLEDGEMENTS

LIST OF ORIGINAL ARTICLES

This thesis consists of an introductory review followed by five research articles, which are listed below and referred to in the text by the Roman numerals I-V. Articles I-IV are reproduced with the kind permission from the publishers, while study V is the author version of the submitted manuscript.

I   Siipilehto, J. 1999. Improving the accuracy of predicted basal-area diameter distribution in advanced stands by determining stem number. Silva Fennica 33(4): 281–301.
http://www.metla.fi/silvafennica/full/sf33/sf334281.pdf

II   Siipilehto, J. 2000. A comparison of two parameter prediction methods for stand structure in Finland. Silva Fennica 34(4): 331–349.
http://www.metla.fi/silvafennica/full/sf34/sf344331.pdf

III Siipilehto, J., Sarkkola, S., & Mehtätalo, L. 2007. Comparing regression estimation techniques when predicting diameter distributions of Scots pine on drained peatlands. Silva Fennica 41(2): 333–349.
http://www.metla.fi/silvafennica/full/sf41/sf412333.pdf

IV   Siipilehto, J. 2008. Modelling stand structure in young Scots pine dominated stands. Forest Ecology and Management 257: 223–232.
doi:10.1016/j.foreco.2008.09.001

V   Siipilehto, J. A compact family of models for flexible prediction of stand structure: The BLUP application for Scots pine-dominated stands in Finland. Submitted.

AUTHOR'S CONTRIBUTION

I was responsible for most of the analysis and most of the writing in Paper III. The material on the main differences between regression estimation techniques was written by Mehtätalo, while Sarkkola described the characteristics of peatlands. The final text of the manuscript was jointly prepared by all authors.

CONTENTS

1 INTRODUCTION

1.1 Finnish forests in brief

Finland, 'the land of the thousand lakes', could as well be called 'the land of millions of forests'. Indeed, the land area of Finland comes to 30 million hectares, of which 87% is classified as forestry land. Of the total forestry land, the proportion owned by about 900,000 non-industrial private forestowners amounts to 52% and that owned by the state amounts to 35%, while companies own 8%. The total standing volume amounts to about 2,206 million m^3, for an average of 107 m^3ha^{-1}. However, in Southern Finland, the mean volume (132 m^3ha^{-1}) is almost twice as great as in northern Finland (76 m^3ha^{-1}). Nature conservation and wilderness areas (3.0 million hectares) are found mainly on state land in the northern part of Finland (Finnish ... 2009). Stands growing on drained peatlands are very important natural resources in Finland: about 52% of the country's 10 million hectares of peatland has been drained for forestry purposes in order to increase wood production (Hökkä et al. 2002).

About half of the growing stock consists of Scots pine (*Pinus sylvestris* L.), while the proportion of Norway spruce (*Picea abies* (L.) Karst.) is 30% and that of broadleaf species, mostly birch (*Betula* spp.), 20% (Finnish ... 2009). Scots pine is the most common tree species both on drained and on pristine peatlands. It covers 77% of the total drained forested peatland area in Finland. At present, young and advanced seedling stands amounted to 21% of the total forest area of Finland. The proportion of artificial regeneration has been slightly increasing and is now 44% in the whole country and 55% in southern Finland, according to the latest national forest inventory (NFI) results. Pine has been dominant as an artificially regenerated species, but the proportion of spruce has been increasing gradually and surpassed that of pine in 2005 (ibid.).

1.2 Stand structure

1.2.1 Definition of stand structure

The structure of the forests in general can be described in terms of the mean and sum characteristics of interest, as done above. A tree stand is defined as a relatively uniform group of forest trees that can be differentiated from the surrounding stands by its structure, tree species composition, and site. In practical forestry, a tree stand is closely related to a stand 'compartment', which is a management unit. Many definitions, from tree to landscape level, have been used for describing stand structure in research and practical forestry (see Sarkkola 2006, p 15–16). Throughout this study, the concept of stand structure refers to tree-species-specific size distributions of living trees in a stand. More specific description of the stand structure may take into account spatial distribution and species diversity in addition to the variation in tree dimensions (e.g., Kuuluvainen et al. 1996, Stoyan and Penttinen 2000, Pommerening 2002). The size distributions can describe distribution of stem number, stand basal area, stand volume, etc. by diameter classes (e.g., Bailey and Dell 1973, Kilkki and Päivinen 1986, Loetsch et al. 1973), by height classes (e.g., Westfall et al. 2004, Maltamo et al. 2004), or in terms of both – by diameter and height classes as a joint bivariate distribution (e.g., Schreuder and Hafley 1977, Zucchini et al. 2001, Li et al. 2002). Diameter distributions describing stem number or basal area by diameter classes are most commonly used. The tree diameter is typically defined as diameter-at-breast-height (dbh).

The structure of a forest stand in terms of its diameter distribution is of great importance. In practical forestry, the diameter distribution is useful for determining the stand's stage of development (e.g., Cajanus 1914); in combination with a height model for estimating, for example, the stand's total volume (e.g., Li et al. 2002, Mabvurira et al. 2002); and for assessing the quantity (e.g., Päivinen 1980, Mack and Burk 2004) and sometimes also quality (Kärki et al. 2000) of timber assortments. In addition, it enables prediction and simulation of the future yields and the target stand states for management objectives, such as cutting regimes (Hyink and Moser 1983, McTague and Bailey 1987, Bowling et al. 1989, Franklin et al. 2002, Newton et al. 2005). Furthermore, the effect of genetic improvement, management activities (e.g., vegetation control), or disturbances (e.g., moose browsing) on stand structure can be described through changes in diameter or height distribution (see Knowe et al. 1992, Siipilehto and Heikkilä 2004, Smith 2007, Weng et al. 2010). The diameter distribution of living trees is also a relevant basis for characterising stand diversity (e.g., Buongiorno et al. 1994, Uuttera et al. 1997, Staudhammer and LeMay 2001, Pommerening 2002).

1.2.2 Factors affecting stand structure

Stand structure is the product of natural processes and many influencing factors such as geographical location, site, species-specific dynamics and management. The location of a stand has usually been taken into account in the models for stand development via a temperature sum factor (see Hynynen et al. 2002). Shade-tolerant Norway spruce dominates on fertile sites, whereas less fertile sites are dominated by the shade-intolerant Scots pine. The decreasing diameter distribution is characteristic of the natural late-successional spruce stands (Linder et al. 1997, Kuuluvainen et al. 1998b). Along with succession, the shape of the diameter distribution can also become more symmetrical because of the mortality of the smallest trees (Laiho et al. 1994, Linder 1998). Pine-dominated old-growth forests typically consist of different age and species cohorts, which form a patchy and multi-layered canopy structure (Kuuluvainen 2002, Lähde et al. 1994ab, Rouvinen and Kuuluvainen 2005). Thus, diameter distributions in natural or semi-natural old-growth forests are often bimodal or multimodal (e.g., Kuuluvainen et al. 1998a, Linder et al. 1997). If the different tree species are examined separately, the distributions seemed unimodal in most cases examined (Siipilehto 2001b, Rouvinen and Kuuluvainen 2005; see also Merganič and Sterba 2006 concerning virgin forests in Slovakia).

The great majority of the forests in Finland are commercially managed. The diameter distribution of a managed stand is typically less wide and more symmetrical than that of a natural, unmanaged stand in Finland (see Siipilehto 2001b, Siipilehto and Siitonen 2004, Rouvinen and Kuuluvainen 2005). In recent decades, the human impact on forests has changed strongly. For example, tar burning, slash-and-burn cultivation, and woodland grazing of cattle all had their effects until the early 1900s (Kaila 1931, Lehtonen 1998), and some of their effects on stand structures could still be seen in recent studies (Siitonen et al. 2000, Uotila et al. 2002). At first, the commercial assortments included only sawlogs, which resulted in the setting of cutting diameter limits. The rapid increase in activity in the pulp and paper industry in the '60s (Finnish ... 2007) opened the markets to pulpwood and thus enabled forest management practice to turn toward even-aged forestry (Kuusela 1990). In the '60s and '70s, levels of, for example, annual reforestation and drainage reached their maximum areas because of the government's strong stake in funding of forestry projects (see Koskimaa 1985, MERA ... 1969). In the early 1950s, the area artificially regenerated annually (mostly seeded)

amounted to only about one fourth of that from the late 1960s to modern years. Thus, planted stands are still relatively young and concurrently extremely rare in mature forests.

On pristine peatland, the stands are often sparse and have a heterogeneous age, size, and spatial structure (Gustavsen and Päivänen 1986, Norokorpi et al. 1997). The size distribution shifted slowly toward a bell-shaped diameter distribution with respect to increasing dominant age (Sarkkola 2006, p 39). The stocking level remains so low that self-thinning plays only a minor role in tree mortality. Drainage is clearly one of the most important silvicultural methods that have affected stand structures in Finland. After drainage, the structural inequality in the size distribution of a peatland stand may increase on account of the improved regeneration and growing conditions for the trees (Hökkä and Laine 1988, Sarkkola et al. 2003, Sarkkola et al. 2005). During post-drainage succession, increasing inter-tree competition results in decreasing stem number even if no cuttings are carried out. Thus, the size distribution changes because of the mortality of the smallest trees.

In conclusion, unmanaged natural forests are few in number in Finland. In managed forests, the management history is totally different in existing old forests from that of the younger forests. In addition, according to Maltamo et al. (1997), minor differences in forest stand structures can be found between forest-owner groups, most probably due to differences in forest management.

1.3 Modelling stand structure

1.3.1 Stand characteristics

The first stage of describing a stand is to assess its site and stand characteristics. Stand characteristics can be modelled, for example, as a function of stand age, location, and site factors by tree species. This is the base for more detailed description of stand structures. In order to avoid laborious measurement in the practical forest management planning (FMP) field work, the description of a stand is commonly simplified to visually assessed mean and sum characteristics. FMP as applied on private estates in Finland is in the process of changing (see Koivuniemi 2003, Holopainen and Hyyppä 2009, Tikkanen et al. 2010). The same tendency can be seen in FMP carried out for state-owned or forest-company-owned stands and to some extent in the NFI. Therefore, until the late '80s, stand variables such as mean age, mean diameter and mean height, total stem number, and basal area were considered adequate to characterize the entire growing stock. Tree species were characterized by their proportion of the stand basal area. Today, stand characteristics are assessed by tree species, and they are described separately for each storey (see PATI-maastotyöohje 2004, Solmun ... 1997, Valtakunnan ... 2009). Determining the stem number or basal area is typically optional in FMP field work. In practice, stem number is assessed in young stands by means of fixed-area sample plots while basal area is assessed in advanced stands using relascope (angle-gauge) sample plots (i.e., probability proportional to tree basal area). The guidelines for FMP field work are very much the same for state-owned forests (Laamanen et al. 1997) and also for landscape-related ecological management planning (Karvonen 2000). However, sometimes stem number is required additionally to basal area and basal-area-weighted variables in FMP for the forest-company-owned stands (Kuvioittainen ... 1998).

Today, utilisation of satellite images and laser scanning data in Finnish FMP is under intensive study (e.g., Peuhkurinen et al. 2007, Närhi et al., 2008, Tomppo et al. 2008). These methods seems to have increased the accuracy in the number of stems (Suvanto et al. 2005, Uuttera et al. 2006, Packalén and Maltamo 2007, Vohland et al. 2007) when compared with

field work (Kangas et al. 2004). Wider utilisation of the laser scan data is leading to a new kind 'precision forestry' as described by Holopainen and Hyyppä (2009). Laser scanning has already been used operationally for some years now for large-area forest inventory in Norway (e.g. Næsset 2007, Næsset et al. 2004) and was begun in the whole of Finland in 2010 as an inventory system for the private forests (Tapio ... 2009).

Alternative choices related to the stand characteristics assessed may cause problems through the use of unequal FMP inventory data as input variables in simulators. Accordingly, a need for modelling individual stand characteristics or relationships between stand characteristics arises from the changes and alternatives in FMP or NFI practices (e.g. Nuutinen 1986, Eid 2001, Nissinen 2002).

1.3.2 Size distributions

Finnish simulators such as MELA (see DemoMELA, Siitonen et al. 1996), MOTTI (see MOTTI software, Hynynen et al. 2005), and MONSU (see MONSU, Pukkala 2004) are based on tree-level data. The SIMO simulator incorporates both stand-level and tree-level simulation options, but, in any case, distribution models are needed for calculation of assortment volumes (Kalliovirta 2006, Tokola et al. 2006, Holopainen et al. 2010). Consequently, the next step in modelling stand structure is to convert stand-level information into tree-level information through size distribution modelling. This means selecting the *distribution function*, selecting the scale of *weighting*, and selecting the distribution *modelling approach*.

Many studies have carried out probability density function (pdf) comparisons empirically in order to find the most appropriate pdf (e.g., Hafley and Schreuder 1977, Kamziah et al. 2000, Zhang et al. 2003, Palahi et al. 2007). Another way of comparing the flexibility of alternative distributions is more theoretical, by means of possible kurtosis-skewness ranges (e.g., Hafley and Schreuder 1977, Wang and Rennolls 2005). Skewness, or asymmetry, is defined as a departure from symmetry about the mean where negative values indicate a distribution with a long tail to the left (i.e., negatively skewed, or left-skewed) and positive values a long tail to the right (i.e., positively skewed or rightskewed). Kurtosis is a relative measure of the flatness or peakedness of a distribution; the larger the value, the more peaked the distribution, and vice versa: the lower the value, the flatter the distribution. In this kind of theoretical description of flexibility, the normal, exponential, and uniform distributions are all represented by a point in skewness-kurtosis space, a verification that they all have but one shape. The gamma, lognormal, and Weibull distributions are represented by the lines demonstrating their capability to assume a variety of shapes. The gamma and lognormal distributions are limited to shapes that have positive skewness, whereas the Weibull has the ability to describe both positive and negative skewness. The beta and Johnson's SB distributions are flexible in covering a region in the skewness-kurtosis space. The logit logistic distribution (Tadikamalla and Johnson 1982) has recently been presented for forestry applications, and it seems to be the most flexible parametric distribution in view of the possible skewness-kurtosis variation (Wang and Rennolls 2005).

In practical applications, dbh distributions are presented either *unweighted* with respect to tree frequency (i.e., dbh-frequency distribution) or *weighted with respect to tree basal area* (i.e., basal area-dbh distribution) (see Gove and Patil 1998). Weighting affects the shape of the distribution. For example, if we assume that the dbh-frequency distribution is symmetrical, the basal area-dbh distribution is skewed to the left. This skewness is more pronounced if the volume-dbh distribution is presented (Loetsch et al. 1973, p 44). Consequently, weighting may have some effect on the goodness of fit and on the predictability of the selected distribution –

especially in the case of decreasing dbh-frequency distributions, weighting has increased the predictability (see Hökkä et al. 1991, Gove 2003a).

The great majority of the alternative distribution models in Finland are based on basal area-dbh distribution, which is partly a result of relascope-sampled data and partly because of its ability to emphasise the large and the most valuable trees (Päivinen 1980). Elsewhere, basal area-dbh distribution models are rarely used (see Gove and Patil 1998), even though they were introduced as early as 1967 by McGee and Della-Bianca and 1971 by Lenhart and Clutter. Frequency distributions have traditionally been used in Scandinavia (e.g., Mønnes 1982, Tham 1988, and Holte 1993), but they are few in number and also comparatively recent in Finland: Sarkkola et al. (2003, 2005) presented the Weibull model and Maltamo et al. (2000) a Weibull- and percentile-based prediction model, and, more recently, Maltamo et al. (2007) compared dbh-frequency distribution with a basal area-dbh distribution model using Weibull. Basal-area-weighted models can be considered quite unpractical for young stands, because of the unweighted stand variables assessed. However, models specifically for young stands are almost totally absent in Finland and consist only of models for planted spruce stands by Valkonen (1997).

There are two main *approaches* for predicting the parametric diameter distribution of a stand by using mean and sum stand characteristics only. In the *parameter prediction method*, estimated regression models using stand characteristics as explaining variables are applied for prediction of the pdf of the target stand (e.g., Rennolls and Rollinson 1985, Robinson 2004). The alternative approach is the *parameter recovery* method, in which the relationships between stand variables (moments or percentiles) and distribution parameters are solved from the system of equations (e.g., Bailey and Dell 1973, Burk and Newberry 1984, Lindsay et al. 1996).

Most of the distribution models in Finland are based on straightforward parameter prediction. Such models include the beta distribution (e.g., Päivinen 1980, Siipilehto 1988, Maltamo et al. 1995) and the Weibull distribution (e.g., Kilkki and Päivinen 1986, Mykkänen 1986, Kilkki et al. 1989, Maltamo et al. 1995, Maltamo 1997). There are few parameter recovery models in Finland. *Percentile-based recovery* models have incorporated the effect of moose browsing (Siipilehto and Heikkilä 2005) or retained trees and stand edges on the height distribution for sapling stands (Valkonen et al. 2002, Siipilehto 2006a, Ruuska et al. 2008). Apart from older studies by Cajanus (1914) and Ilvessalo (1920), *moment-based recovery models* are not found in Finland.

It needs to be mentioned that some applicable methods do not involve parametric distribution functions. Such methods are percentile-based distribution (e.g., Borders et al. 1987, Kangas and Maltamo 2000b), k-nearest neighbour (k-NN), or k-most-similar neighbour (k-MSN) (e.g., Mouer and Stage 1995, Haara et al. 1997, Maltamo and Kangas 1998). Recently, the k-NN method has been studied actively in relation to remote sensing techniques (e.g., Peuhkurinen et al. 2008, Holopainen et al. 2009, Järnstedt 2010). Kernel smoothing has been used too, but it is not suitable for prediction purposes (e.g., Droessler and Burk 1987, Uuttera et al. 1996, Maltamo et al. 1997, Koivuniemi 2003).

1.3.3 Bivariate distribution of tree diameters and heights

Going one step further in the modelling of tree-level information means incorporating the *within-dbh-class height variation* into the model. The more sophisticated the tree-specific growth and survival models are (e.g., in terms of competition indices), the more detailed and reasonable the predicted stand structure should be (Biging and Doppertin 1992, Zhang

et al. 1997). The social status of a tree depends not only on its relative diameter but also on its relative height in the stand. These features are reflected in a tree's further development by means of tree growth and mortality. One practical motivation is that knowledge of the between- and within-diameter-class height variations increases the possibilities for imitating different types of thinning (Hafley and Buford 1985) whereas the typical motivation is simply the ability to provide a more realistic picture of the stand structure (e.g., Tewari and Gadow 1999). Stand structure as a joint distribution of tree diameters and heights can be described by means of bivariate pdf. Johnson's SBB distribution has been used for this purpose in a number of studies (e.g., Hafley and Schreuder 1977, Hafley and Buford 1985, Siipilehto 1996, Tewari and Gadow 1999). No other bivariate generalisation of the alternative univariate parametric distributions has been able to provide such reasonable marginal distributions, joint bivariate distribution, and diameter–height relationship in closed form (Schreuder and Hafley 1977, Wang and Rennolls 2007). However, using alternative copulas (i.e., methods that couple bivariate distribution function with their one-dimensional marginal distributions and dependence structure), Wang and Rennolls (2007) presented satisfactory bivariate extension with the logit logistic, beta, and SB distribution as marginal distributions, while Li et al. (2002) presented that for gamma distribution. Zucchini et al. (2001) presented a bivariate model based on the mixture of two bivariate normal distributions. Thus, the height–dbh relationship was described by two straight lines, with different slopes. The early Finnish application by Kilkki and Siitonen (1975) presented a bivariate model based on the beta function as diameter distribution and conditional height distributions together with Näslund's height curve describing expected height. The bivariate model may have practical application such as predicting the missing heights with random variation for tally trees or in general for generation of model-based data, as in the study by Kilkki and Siitonen (ibid.).

1.4 Objectives

The common objective of these studies is to develop and improve parameter prediction models for predicting size distributions of Finnish forest stands for practical use, such as forest management planning purposes. Naturally, models should include all of the main tree species, but in this thesis, some models are introduced for Scots pine only. More detailed objectives are more or less methodological or practical.

1. Comparing and ranking of the alternative regression estimation methods (Papers **III** and **IV**)

2. Specifying efficient transformations in order to linearize the dependence and to find close correlation between modelled parameters and explaining stand characteristics (Papers **I**, **III**, and **IV**)

3. Developing alternative methods to mimic realistic variation in tree dimensions by including within-dbh-class height variation (Paper **II**)

4. Introducing a family of simultaneous models for stand characteristics and parameters of the optional dbh distributions and a height curve including a cross-model covariance structure for linear prediction application (Paper **V**)

5. Covering the lack of models to some extent by developing models for peatland (Paper **III**) and juvenile stands (Paper **IV**)

6. Developing simple models for the relationship between tree diameter and height for forest management planning purposes (Papers **I**, **II**, **IV**, and **V**)

7. Ensuring the applicability of the alternative models, requiring alternative input data (Paper **V**)

8. Comparing the additional stand characteristics in terms of their ability to improve the models' performance (Paper **V**)

9. Developing a framework for flexible validation of the optional input *variables* and optional *models* in order to allow giving recommendations for their practical use (summary)

When I searched for additional stand characteristics in order to improve the accuracy of the predicted distributions, I paid special attention to the additional stem number (in advanced stands) as well as to the dominant tree characteristics. This means that the responses to these variables were checked with the optional distribution models by means of some examples in this summary. In addition, the effects of the additional knowledge on these stand characteristics were checked in terms of bias and RMSE in the main stand characteristics through application of the validation data. Also, the optional models for height–diameter relationships were validated with an example and via validation data. I presume that this kind of validation in this summary offers me the keys to be able to recommend a certain model for certain conditions.

2 MATERIAL

2.1 Stand plots

When one focuses on tree size distribution, the number of observations within a stand is essential for reliable estimation of the distribution function. Shiver (1988) stated that, regardless of the estimation method, a sample size of approximately 50 trees is required for reproducing marginal distributions in classes with less than 10% error. Considerable reduction in variance, bias, and RMSE has been found in the Weibull parameters when sample size changed from 30 to 50, and the further reduction thereafter had a decreasing rate (ibid.). In study of the bivariate structure of tree diameters and heights, quite likely twice as many observations are needed. For example, sample sizes of greater than 100 trees have been used by Schreuder and Hafley (1977) and Wang and Rennols (2007). Modelling the development of stand characteristics does not necessarily require as large samples in terms of measured trees; e.g., Koivuniemi (2003) recommends 30 trees, but, more obviously, a representative sample is needed. Quite likely, the most suitable existing data for simultaneous modelling of stand characteristics and diameter distributions in mineral soils were TINKA and INKA data sets, for young and advanced stands, respectively (see Gustavsen et al. 1988).

Each TINKA and INKA sample consisted of a cluster of three permanent, objectively located circular plots (300–2,500 m^2) within a stand. The whole cluster represented a stand, in order to yield enough observations for modelling the distributions and stand characteristics in Papers **IV** and **V**. The re-measurements were carried out twice, five and 10 years after establishment. Typically, subsamples of INKA and TINKA data were used in this thesis. The test data applied in Papers **I** and **II** were selected from the second measurement round of INKA, either having H_{dom} greater than or equal to 10 m or restricted to Scots pine-dominated stands, as for Paper **III**. In the selection of the modelling data for Papers **IV** and **V**, the main criterion was domination by Scots pine, but each measurement round was included (see Figure 1 and Table 1). In addition, in Paper **V**, a mean diameter greater than 1.5 cm was required for avoiding the anomalies in stand characteristics resulting from the low proportion of trees above breast height in stands in the youngest state.

The data from pine–birch and spruce–birch mixtures (Mielikäinen 1980, 1985) were based on temporal sample plots, originally established for study of the effect of birch admixture on the growth and yield of the stands. The representativeness of the data was questionable, because of the location of the stands in the south-eastern part of Finland (see Figure 1). In addition, the sample plots in conifer–birch mixtures were placed subjectively within the stands to represent a conifer-dominated and a birch-dominated plot in addition to a plot with a birch admixture of about 50%. The whole cluster represented a stand, in order to yield enough observations for fitting the univariate and bivariate distributions in Papers I and **II**, respectively. Combining the plots had the disadvantage of diminishing the variation in the proportion of the birch admixture (30–65%). The main stand variables are presented in Table 1 in Paper **I** and Paper **II**. These data sets have been described in detail by Mielikäinen (1980, 1985).

The drained peatland stand data set (see Paper **III**) was based on permanent sample plots, originally established to study the effect of drainage and forest management. The measurement period varied, but the re-measurements were typically carried out with either five- or 10-year intervals. Some of the study plots were followed for about 80 years since drainage. The advantages of the data were in the large variation in the successional stage in terms of years since drainage (see Table 1), resulting in a wide range in stand characteristics as well as in the shape of the diameter distributions. The data covered all Scots pine-dominated peatland

site types (see Table 1), and the 14 study areas covered roughly the whole of Finland (see Figure 1). The sample plots used in this study have been subjectively selected from the larger available data set. The chosen Scots pine-dominated stands on drained peatlands had long developmental series, varying stand density and site fertility. For details of the selection, see the paper by Sarkkola et al. (2005). The main stand variables are presented in Table 1 in Paper **III**.

Quite typically, data have been randomly divided into two groups, one for modelling and the other for testing (e.g., Cao 2004). In Paper **V**, 25% of the data were randomly selected for model validation. I preferred using totally independent data sets for modelling and testing if such data were available. Consequently, the modelling work utilizes all data. The capability of the model for prediction is then critically tested with data that are not generated in the same way as the modelling data (see Papers **I–IV**). This procedure will more likely reveal the critical cases when models are used for predicting new stands. Test data may include extreme treatments such as a continuum from heavy thinnings to unthinned stands (Paper **III**) or new kinds of cleaning practices (Paper **IV**). If the model is intended for practical use, their critical testing is a great benefit. An excellent example of this critical testing is the percentile-based distribution models by Kangas and Maltamo (2000b), which were tested with a large number of different, independent data sets (Kangas and Maltamo 2000a).

An additional test data set (HARKO) for validation of models in Paper **III** consisted of 52 Scots pine-dominated, permanent sample plots on drained peatlands in 14 distinct peatland areas in central and northern Finland (see Figure 1). The HARKO data covered the same peatland site types as the modelling data (see Table 1). However, these stands were mainly found in different geographical and climatic regions than the modelling peatland data (see Figure 1). In contrast to the modelling data, the HARKO data included a relatively broad range of thinning intensities. In addition to the unthinned controls, relatively heavy thinnings were included. Thinning removal was about 80% of the total basal area, at its greatest. Thus, said data provided important additional information about the validity of the models. For more details, see Table 2 in Paper **III**.

Two additional test data sets were included, to validate the models for juvenile Scots pine stands in Paper **IV**. Both data sets were local and represented mineral soil sites of *Myrtillus* type (MT) (see Table 1). The first test data set ('Establishment') was originally used for studying the establishment of pine stands immediately after planting on harrowed soil. The second test data set ('Cleaning'), in Paper **IV**, was originally used for studying the effect of cleaning treatments: 1) no cleaning, 2) point cleaning of broadleafs within a one-metre radius of crop-tree pines, 3) total cleaning of all broadleafs, and 4) topping of competing broadleafs (cutting the stem at a height of 1.3 m). For more details, see Table 2 in Paper **IV**.

2.2 Measurements

In this thesis, the interesting random variables were the tree dimensions, dbh, and height – and indirectly also tree volume. In the mixed conifer–birch data sets in Papers **I** and **II**, all of the roughly 120 trees per stand plot were measured for tree dbh and height. In the INKA and TINKA data sets, the total number of trees tallied was about 100–120 per stand (i.e., in the cluster of three sample plots). In the TINKA data set, originally established in sapling stands, all crop trees were measured for tree species, dbh (if $h > 1.3$ m) and height. However, in the INKA data set, a smaller radius (one third of the total area) was used for selection of sample trees, which were measured for tree height (and other more detailed measurements). Because the dbh distribution model and height–dbh relationship was modelled simultaneously with

stand characteristics for Paper **V**, the need for a sufficient sample was fulfilled by combining the cluster of three plots to represent a stand. In addition, for effective utilisation of the data, the whole tally tree data set was used and the missing heights were predicted from the fitted Näslund's height curve, including the random variation in the validation data set (see Paper **V**). For more details on the TINKA and INKA data sets, see Table 1 of Papers **IV** and **V**, respectively, and the work of Gustavsen et al. (1988). In Paper **III**, the only random variable studied was tree dbh, which was measured from all the trees.
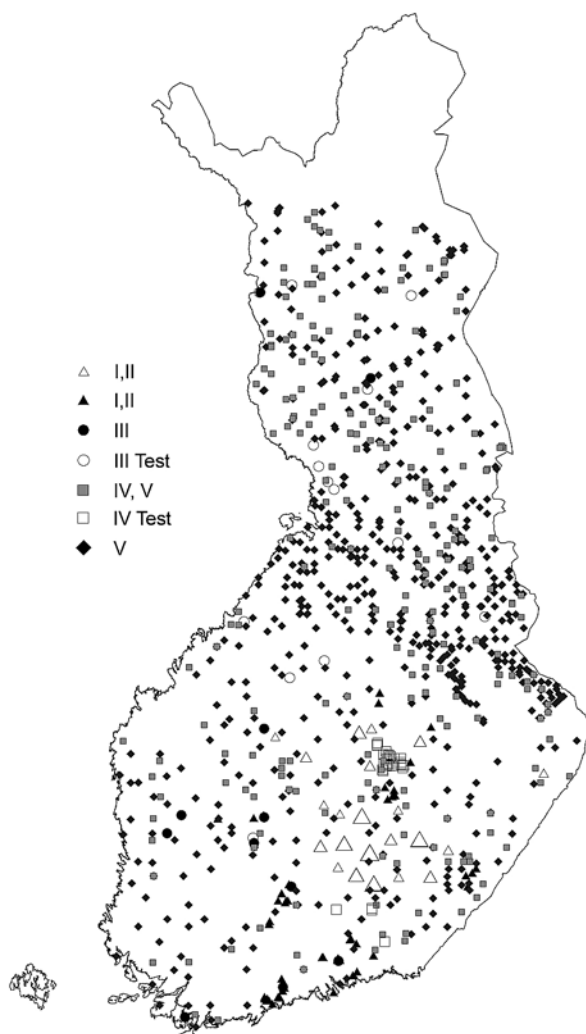
Figure 1. Location of the modelling and test data sets: pine–birch (△) and spruce–birch (▲) mixtures in Papers **I** and **II**, pine-dominated drained peatland areas for modelling (●) and HARKO data for testing (○) in Paper **III**, pine-dominated TINKA data for modelling (■) and local data sets for testing (□) in Paper **IV**, and pine-dominated INKA data for modelling and testing in Paper **V** (◆) in addition to TINKA (■) data.

Table 1. The number of observations (stand × measurement occasions) in age classes and the distribution of site types for the stands in the different data sources.

Age	Mixed pine–birch	Mixed spruce–birch	Peatland[a]	TINKA	INKA	HARKO	Cleaning	Estab.
<20			143	450	22	21	27	81
20–39	16	9	180	188	189	259		
40–59	30	36	210	6	206	138		
60–79	43	10	90	1	199	113		
80–99	2	5	36		172			
100–119					105			
>120					62			
Site								
OMaT		2			1			
OMT	10	43			8			
MT	58	15	11	33	128	32	5	27
VT	23		33	112	284	20		
CT			21[b]	33	55	4		
CIT					4			

[a] Years since drainage.

[b] The site type was actually DsT (i.e. dwarf-shrub site type).

3 METHODS

3.1 Distribution functions

3.1.1 Weibull distribution

The Weibull distribution has been widely used to describe and predict diameter distributions. Its advantages include simplicity of mathematical derivation, the small number of parameters to be estimated, the closed-form cumulative function, and its flexibility in description of different shapes of unimodal distributions (Bailey and Dell 1973). The three-parameter Weibull probability density function (*f*) is as follows:

$$f(x) = c/b[(x-a)/b]^{c-1} \exp\left\{-[(x-a)/b]^c\right\}$$

(1)

where *x* is the random variable, the observed diameter or height in a stand plot, and *a*, *b*, and *c* are the location, scale, and shape parameters of the Weibull function, respectively. In the two-parameter Weibull function, *a* is fixed at 0 (see Papers **I, IV**, and **V**). Note that a negative exponential distribution results when shape parameter *c* is given the value of 1. The Weibull distribution is skewed to the right when $c < 3.6$, symmetrical with a value of 3.6, and left-skewed when $c > 3.6$. The cumulative Weibull function has a closed-form expression, as:

$$F(x) = 1 - \exp\left\{-[(x-a)/b]^c\right\}$$

(2)

The Weibull function has some convenient features. The cumulative function helps, for example, when one is sampling trees from the Weibull distribution. Probabilities can be produced easily without the need for numerical integration. According to size-biased theory, Gove and Patil (1998) showed that weighting the initial two-parameter Weibull frequency distribution with tree basal area leads to a standard gamma distribution with parameter $k = (1+2/c)$ of gamma function $\Gamma(k)$. Correspondingly, in the case of returning the initial basal area-dbh distribution to represent frequency distribution yields a gamma distribution with parameter $k = (1-2/c)$.

3.1.2 Johnson's SB distribution

Johnson's SB distribution (Equation 3) is based on a transformation (Equation 4) to standard normality (Johnson 1949b) as follows:

$$f(d) = \frac{\delta}{\sqrt{2\pi}} \frac{\lambda}{(d-\xi)(\xi+\lambda-d)} \exp(-0.5z^2)$$

(3)

and

$$z = \gamma + \delta \ln\left[\frac{d-\xi}{\lambda+\xi-d}\right]$$

(4)

where γ and δ are the shape parameters, ξ and λ are the location and range parameters, and *d* is the tree diameter observed in a stand plot. In the three-parameter SB function, location ξ is fixed at 0. Quite extreme shapes of SB distribution exist with a low value of shape parameter δ (ibid.). Value $\delta = 1/\sqrt{2}$ results in a flat, almost uniform distribution. Parameter δ values close to 1 represent almost decreasing dbh distributions, if γ is simultaneously close to 1. One

interesting property is that the SB distributions have a bimodal shape if δ is given a value less than or equal to 0.5. This property was used in exclusion of bimodal distribution from the modelling data (see Papers **I–III**). For the work described throughout this thesis, we applied a three-parameter SB function in which the location (minimum) was set to 0 (see also Kamziah et al. 1999).

3.1.3 Bivariate SBB distribution

The bivariate Johnson's SBB function (see Equation 5) is based on the bivariate normal distribution (Johnson 1949a). In the SBB, both marginal distributions follow a univariate SB distribution. The original variables, diameters, and heights were transformed into standard normal variates by means of equations 6 and 7.

$$P(z_d, z_h) = (2\pi\sqrt{1-\rho^2})^{-1}\exp[-0.5(1-\rho^2)^{-1}(z_d^2 - 2\rho z_d z_h + z_h^2)] \tag{5}$$

and

$$z_d = \gamma_d + \delta_d \ln\big[(d - \xi_d)/(\lambda_d + \xi_d - d)\big] \tag{6}$$

$$z_h = \gamma_h + \delta_h \ln\big[(h - \xi_h)/(\lambda_h + \xi_h - h)\big] \tag{7}$$

where subscripts d and h denote distribution of diameters and heights, respectively, and ρ is the correlation coefficient between z_d and z_h. The SBB distribution was applied as a basal-area-weighted distribution (see Paper **II**).

3.2 The relationship between tree diameter and height

One of the properties of interest is the regression relationship between the tree diameter and height obtained from the bivariate SBB distribution. The usual mean regression is complicated, but the median regression takes a simpler form (Johnson 1949a, Schreuder and Hafley 1977, Paper **II**: Equation 7). The regression curve can have various forms, depending on the relationship between the two parameters, φ and θ (Johnson 1949a, Paper **II**: Figure 1). The typical sigmoid form of the height curve is obtained if both parameters, φ and θ, are greater than 1. To avoid unreasonable height curves, Schreuder and Hafley (1977) recommended constraining φ to be greater than or equal to 1 in fitting of the distribution. Both unconstrained and constrained solutions for SBB parameters were studied in Paper **II**. The conditional height distribution for a given dbh also follows SB distribution (ibid.), but the shape of the conditional height distribution changes with the changing diameter (see Siipilehto 1996).

The Näslund's (1936) height curve (Equation 8) was fitted in the linearized form (Equation 9) in Papers **I, II**, and **V**.

$$h = \frac{d^\alpha}{(\beta_0 + \beta_1 d)^\alpha} + 1.3 + \varepsilon_h \tag{8}$$

$$\frac{d}{(h-1.3)^{\alpha^{-1}}} = \beta_0 + \beta_1 d + \varepsilon_z \tag{9}$$

where β_0 and β_1 α are the parameters of the model, power $\alpha = 2$ for pine and birch, $\alpha = 3$ for spruce; and ε_z is the random error of the linearized model. In Paper **II**, we were interested in the height distribution conditional to known dbh. One approach was based on Equation 9 and its residual variation, which was assumed to be homogenous and normally distributed (see Näslund 1936, p. 52). The derivation of the formula for the random variation in the initial scale of tree diameters and heights is detailed in Paper **II**: equations 10–12.

In juvenile stands (addressed in Paper **IV**), one goal was to construct a rather simple and flexible model for the tree dbh from the known tree height and stand characteristics. Different candidate formulations were examined, based on, for example, relative tree size as presented by Nishizono et al. (2005). We selected the multiplicative model as a basis (see Fahlvik and Nyström 2004). The model was fitted in linearized form with logarithmic transformation as:

$$\ln(d) = b_0 + \sum \beta_i \times X_i + \varepsilon \tag{10}$$

where b_0 is a constant, β_i is the ith coefficient for the ith independent variable X_i, and ε is the random component. The error term was divided into stand-level, measurement-occasion-level, and tree-level random components when the mixed-effect model was estimated according to the MIXED procedure and REML estimation in SAS (see Paper **IV**).

3.3 Approaches to modelling

3.3.1 Parameter recovery

The parameter recovery method (PRM) is briefly discussed here because some of its features are commonly utilized also in parameter prediction methods. In PRM, the relationships between stand variables and distribution parameters are derived in a closed form and the parameters estimated for the target stand are solved for on the basis of the resulting system of equations. PRM is possible only for as many parameters as there are known distribution-related stand variables. Furthermore, only stand variables that are mathematically related to the diameter distribution can be used.

For the two-parameter Weibull function, two percentiles with a known value of the random variable and two unknown parameters can be solved for with the system of equations in closed form for parameters b and c (e.g., Bailey and Dell 1973). Dubey (1967) showed that the most efficient and asymptotically normal percentile estimators are the 24th and 93rd when both of the parameters, b and c of the Weibull function, are unknown. Gobakken and Næsset (2004, 2005) and Siipilehto (2006a) applied these in their *percentile-based recovery* model. Sometimes the 50th percentile has been used, because median could be considered a known variable (e.g., Siipilehto and Heikkilä 2005 for the Weibull; see Newberry and Burk 1985, Knoebel and Burkhart 1991 for the SB distribution).

Moment-based parameter recovery is commonly based on the first-order arithmetic mean and the second-order quadratic mean diameter (D_q), the latter having direct relation to stem number (N) and basal area (G) as shown in Equation 11:

$$D_q = \sqrt{(\sum d^2)/n} = \sqrt{G/kN} \tag{11}$$

where $k = [\pi/(2 \cdot 100)^2]$ (see, for example, Gove and Patil, 1998). In the case of the three-parameter Weibull function, the location parameter (a – that is, the minimum) is typically

predetermined. The systems of two equations for the parameters of the dbh distribution can be written as follows (e.g., Cao 2004):

$$b=(D–a)/\Gamma_1 \tag{12}$$

$$D_q^2+a^2–2aD–b^2\Gamma_2=0 \tag{13}$$

where $\Gamma_i = \Gamma(1+i/c)$, $\Gamma(\cdot)$ is the complete gamma function and a is a predetermined location parameter. Naturally, in the case of the two-parameter model, the terms including parameter a are eliminated.

If all three parameters of the Weibull function are recovered, the model incorporates mean, variance, and skewness of dbh distribution (see Burk and Newberry 1984, Lindsay et al. 1996). In the prediction application, the selected percentiles (e.g., 50th and 93rd) or moments such as D and D_q (or N and G instead of D_q) may be known, but in the case of a pure three-parameter recovery method, variance and skewness have to be predicted because they do not belong to standard stand characteristics. Moment-based recovery seems useful for diameter distribution, but in the case of height distribution, the second-order height characteristics have no direct relation to standard stand characteristics, such as what Equation 11 shows for diameters.

3.3.2 Linear regression estimation for parameter prediction

In the parameter prediction method (PPM), *a priori* estimated regression models are applied for prediction of the pdf of the target stand. To model size distributions, distributions are typically first fitted to data, and the estimates obtained, treated as true values for the stand, are then modelled against stand variables. All of the distributions in this study were estimated with the method of maximum likelihood (ML). The basal-area-based distributions in line with the Weibull, SB, and SBB functions were estimated using the ML method represented by Bailey and Dell (1973), Johnson (1949a), and Schreuder and Hafley (1977) but, instead of frequency distribution, with basal area weighting applied (Paper **I**: Equations 3 and 8, Paper **II**: Equation 3). The dbh-frequency Weibull distributions were estimated using PROC NLIN in SAS, with modification of the code that Cao (2004) provided in the appendix.

If the errors are independent and have equal variances, the efficient estimator of coefficients is the Ordinary Least Squares (OLS) estimator. Allowing correlation between observations, the efficient estimator is the Generalized Least Squares (GLS) estimator. A multivariate model is a set of single regression models that are estimated from the same data. Such a situation is common when several parameters have to be modelled according to the PPM approach (e.g., Robinson 2004). Zellner (1962) showed that an OLS estimator would be efficient if the residuals of the separate models were not correlated and the residual variances of the models were equal. Furthermore, OLS is efficient also if the design matrices are the same across models even though residuals are correlated. However, in other cases, the seemingly unrelated regression (SUR) approach can be utilized such that the models are first estimated with separate OLS fits and then re-estimated by GLS.

OLS assumptions can be violated as a result of the hierarchy of the data, meaning that each observation belongs to a class of observations, several observations are available from a single class, and the modelling data represent only a random sample of classes of the population. A recommended approach for these cases is mixed-effects modelling (Laird and Ware 1982) wherein the error variance is divided into between-class and within-class

components (McCulloch and Searle 2000, Diggle et al. 2002). On account of correlations between observations achieved by hierarchy, the fixed parameters should be estimated by means of GLS instead of OLS (Gregoire et al. 1995, McCulloch and Searle 2000). In repeated measurement, using longitudinal data with several responses, an efficient estimation method is able to take into account both the hierarchy (autocorrelation) of the data and the correlations between models. A multivariate mixed-effects model (or hierarchical multiresponse model), combining the mixed-effects modelling and SUR approaches, is the most appropriate (Goldstein 1995). A model of this kind can be treated as a special case of hierarchical (mixed) model, where an additional level of hierarchy is added to the model for longitudinal data (Snijders and Bosker 1999) and the implied assumptions about between-models covariances are parameterized in the covariance matrix of the observations.

In a generalized linear modelling (GLM) approach, the fitting of the Weibull distribution and estimation of the prediction models for the parameters is done in one stage from the treewise data (Cao 2004). In this GLM, instead of minimising the sum of squares of the error with respect to parameter b and c, the goal is to maximize the total log-likelihood of the Weibull function (ibid.; see also Paper **IV**: equations 5–7). In the hybrid method further developed with respect to the GLM approach, parameter b in the likelihood function is replaced with the moment estimator (see Equation 12) while c is substituted for with the prediction equation (see Paper **IV**). Thus, parameter c is estimated conditionally to the moment-based recovered parameter b, the goal therefore being to maximize the total log-likelihood of the Weibull function conditional to an equal mean from the sample and from the predicted Weibull distribution.

3.3.3 Predictors of the parameters

When it comes to regression modelling, one pays special attention to finding the most appropriate formulation of the prediction function. This means that many kinds of transformations may be used in order to find unbiased behaviour across all the variation in the predictor variables. When the common FMP (SOLMU) data are available, the variation in the shape of the dbh distribution within one particular site, stand age, fixed basal area, and median diameter can be projected only by the variation in the median height. Median height can be included as an explaining variable (Päivinen 1980) or expressed as a form (slenderness) of the median tree ($F = h_{gM}/d_{gM}$). The behaviour of the SB model with respect to the slenderness of the median tree was shown for pine and spruce (see Paper **I**: Figure 3).

The transformation named 'shape index' (Equation 14) was introduced in Paper **I** and utilized in Papers **II** and **III**.

$$Shape\ index = \frac{G}{g_M N} \tag{14}$$

where $g_M = (\pi/4)(d_{gM}/100)^2$. The idea was to compare observed basal area with the 'calculated basal area' (i.e., $g_M N$) because I presumed that the ratio between them has to be connected with the shape of the dbh distribution. The shape index was calculated by tree species. Note that the shape index can be determined also as a squared proportion of the quadratic mean (D_q) and basal-area-median diameter ($D_q^{\ 2}/d_{gM}^{\ 2}$). Typically, this proportion is less than 1, which means that d_{gM} is greater than D_q (see Paper **III**: Figure 1). The behaviour of the shape index was studied with varying dbh-frequency and corresponding basal area-dbh distributions used (Paper **I**: Figure 1).

Also, the derived transformation used for predicting parameters of the height distribution is based on the ratio of two different mean characteristics. Much of the variation in the Weibull parameter c could be characterized in a linearized form by means of transformation $1/\ln(H_{dom}/H)$ (Paper **IV**: Figure 1). Again, the above transformation was not just a 'trial and error' finding; by contrast, I derived it from the percentile estimator by Dubey (1967) (see Paper **IV**).

3.3.4 Useful explicit solutions

As Cao (2004) and Fonseca et al. (2009) noted, the approach applied does not need to be pure; it can be a combination of several methods. Typically this means that a moment (mean) or percentile (median) is utilized to solve for a parameter such that the predicted distribution produces it correctly. In Finland, numerous Weibull applications characterising basal area-dbh distributions have applied the known basal-area-median diameter (d_{gM}) this way (e.g., Kilkki et al. 1989, Maltamo et al. 1995, Maltamo 1997). If two out of three parameters were predicted, the third parameter was solved using one of the following equations:

$$a = d_{gM} - b(-\ln(0.5))^{(1/c)} \tag{15}$$

$$c = \ln[-\ln(0.5)]/[\ln(d_{gM} - a)/b] \tag{16}$$

$$b = (d_{gM} - a)/(-\ln(0.5))^{1/c} \tag{17}$$

Similarly, d_{gM} can be set for the median of the predicted basal area-dbh SB distribution. As the values of parameter ξ and median d_{gM} are known and the values of δ and λ are predicted, the parameter γ is solved for by means of the formula below:

$$\gamma = \delta\ln(\lambda + \xi - d_{gM}) - \delta\ln(d_{gM} - \xi) \tag{18}$$

Paper **IV** specified an option for a two-parameter Weibull function, where parameter c was predicted and b was recovered from a moment, mean height (H). Thus, Equation 12 was used for scale parameter b by substitution of mean diameter D with mean height H. In addition, Equation 17 was used in Paper **I** and Equation 18 was used in Papers **I–III** in order to achieve compatibility in the median dbh.

3.3.5 An alternative modelling approach using BLUP

Unlike in the regression modelling, we don't have to fix beforehand which variables are used in application of the linear prediction theory (see Lappi 1993). Linear prediction is based on random variables. The error terms of statistical models are random variables. The cross-model error variance–covariance matrix is valuable when one is calibrating the expected value (μ_1) by means of linear prediction theory. In the notation of Lappi (1991, 1993), the best linear unbiased predictor (BLUP) for variable x_1 is:

$$\hat{x}_1 = \mu_1 + \sigma_{12}\textstyle\sum_{22}^{-1}(x_2 - \mu_2) \tag{19}$$

where x_1 is a scalar of dependent unknown variable and x_2 is a vector of known stand variables, σ_{12} is a row vector including the covariances between unknown dependent and known

variables, and Σ_{22} is the variance–covariance matrix between known variables. The variance of the prediction error after calibration of dependent variable x_1 is:

$$var(\hat{x}_1 - x_1) = \sigma_{11} - \sigma_{12} \Sigma_{22}^{-1} \sigma'_{12} \qquad (20)$$

where σ_{11} is a scalar of the initial variance of the residual error of the dependent variable and σ_{12}' is a transpose of the row vector σ_{12}. In the case of logarithmic transformation in the dependent variable, the bias correction term $s_\varepsilon^2/2$, and for inverse transformation ($1/c$) the bias correction term s_ε^2/x_1^2, had to be added to the intercept (e.g., Lappi 1993). The variance (20) was recalculated whenever the calibrating variable for prediction (19) was changing.

The best linear unbiased predictor approach has much in common with linear regression estimation. However, key differences appear. The most important difference is that each of the variables in the set of BLUP models can be predicted/calibrated with any combination of remaining variables in the set of models as far as they are known and, thus, the residual between known and expected value can be calculated (ibid., p. 77). This feature was utilized in Paper **V** and in the summary in validation of the models presented in this thesis.

In summary, six alternative regression estimation methods were applied in prediction of stand and distribution characteristics. The regression approaches were as follows:

1) Linear model estimated by ordinary least squares (**I–III**)

2) Linear mixed-effects model with random intercept (**III, IV**)

3) Multivariate linear model estimated according to the seemingly unrelated regression approach (**III, V**)

4) Multivariate mixed model estimated as a mixed-effects model with the additional level of hierarchy to allow simultaneous estimation (**III, IV**)

5) Generalized linear model estimated by maximising the total log-likelihood (**IV**)

6) A hybrid method, wherein the generalized linear model was estimated by maximising the total log-likelihood conditional to compatibility in the mean achieved by moment-based recovery (**IV**)

3.4 Model evaluation

The estimates generated from model application for the stem number, 'volume' (Σd^3), and 'value' (Σd^4) were compared with the values derived from the original dbh measurements (see Paper **III**). The advantage of using Σd^3 and Σd^4 in dbh distribution validation is that they do not require height information, while they still provide reasonable estimates of the accuracy in the volume and value of the growing stock, respectively (see Kilkki and Päivinen 1986, Maltamo et al. 1995). Thus, the outcomes are based completely on observed or predicted dbh distributions. However, height–diameter relationships were also modelled and examined in Papers **I**, **II**, **IV**, and **V**. In those Papers and in the summary, stand volume and timber assortments were calculated according to models for individual tree volume and taper curve as a function of known tree dbh and height (Laasasenaho 1982). The accuracies of the models constructed were validated in terms of bias (21), RMSE (22), and precision (23) in the generated stand characteristics (e.g., stem number, basal area, dominant diameter and height, volume of total growing stock, timber assortments, and waste wood fraction). Relative bias

and RMSE were calculated by dividing RMSE by the mean value of the observed Y_i, and they were expressed as percentages. The precision describes error variation excluding the effect of bias (see Papers **I–III**). A Kolmogorov-Smirnov (KS) goodness-of-fit test at alpha 0.1 level was used for validating size distribution models (see Papers **II–IV**). The KS test is based on the greatest absolute difference – i.e., the 'supremum' – between observed and predicted distributions. The error index by Reynolds et al. (1998) was also applied for ranking the models with respect to the errors in the predicted frequencies in height classes in addition to the KS test (Paper **IV**: Equation 12).

$$bias = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i) \tag{21}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n-1}} \tag{22}$$

$$s_b = \frac{1}{n} \sum_{i=1}^{n} \sqrt{(e_i - bias)^2} \tag{23}$$

where Y_i is the observed and $\hat{Y}_i$ is the predicted stand characteristic, and e_i is the relative prediction error (%) in stand i.

In the summary, the alternative height models were validated in terms of bias and RMSE in H and H_{dom}. In this validation, tree diameters were sampled from dbh-frequency distribution for young stands and from basal area-dbh distribution for advanced stands using the BLUP models (Paper **V**) for the Weibull parameters. Predictions were based on the common FMP stand characteristics. The Näslund's curve according to the BLUP model in Paper **V** was compared with Näslund's model presented in Paper **I**. The former applied N, D, and H in young stands and G, d_{gM}, and h_{gM} in advanced stands for the calibration of Näslund's parameters, respectively. The latter, instead, applied N, D, and H for calibrating d_{gM} and h_{gM} in young stands in order to predict the Näslund's curve from them, whereas d_{gM} and h_{gM} were known and used as such in advanced stands. In addition, height models by Veltheim (1987) and Mehtätalo (2005) were examined simultaneously, but it needs to be mentioned that these two models did not have any stand-level height characteristics as an explaining variable. Furthermore, the correct form of the height curve means unbiased behaviour with respect to relative tree size. Therefore, the biases in tree heights were presented with respect to relative diameter (d/D) by means of alternative models for diameter–height relationship.

For concluding validation, the accuracy levels of the most promising models were compared from the randomly selected test material of Paper **V**. Here I focused on only a few combinations of input variables, regarded as the most relevant ones. Accordingly, the FMP (SOLMU) data set was the basis and the arithmetic variables (N, D, and H) were used for young stands ($H \leq 9$ m; 249 observations) and basal-area-based variables (G, d_{gM}, and h_{gM}) for advanced stands ($H > 9$ m; 280 observations). The effect of the additional knowledge of N, D_{dom}, and H_{dom} on the generated stand characteristics was tested. When calculating the volumes for the model validation, the diameter distribution model was combined with Näslund's height model from Paper **V**. The hybrid height distribution model (GLM+M) in combination with the dbh prediction model from Paper **IV** was checked as an option for young stands. However, distribution models in Paper **I** were excluded from this evaluation because of the unrepresentative data. Furthermore, the SB distribution models in Paper **I** have been tested widely with alternative models in Maltamo et al. (2002).

4 RESULTS

4.1 BLUP models for stand characteristics

In total, eight stand characteristics (G, N, D, d_{gM}, D_{dom}, H, h_{gM}, and H_{dom}) were simultaneously modelled with the SUR method (Paper **V**). Stand age and temperature sum were the driving variables of the models. Site types by Cajander (1925) were used as dummy variables to estimate the differences from the average sub-xeric *Vaccinium* site type (VT) for Scots pine. Site fertility had considerable effect, not only on mean characteristics but also on basal area. The effect of stoniness and paludification was significant and evidently stronger in the arithmetic means than in basalareaweighted means or dominant tree characteristics (see Paper **V**: Table 3). The expectation values for the dimensions were logical when compared with each other as a function of stand age (i.e., $D < d_{gM} < D_{dom}$ and $H < h_{gM} < H_{dom}$).

Multiplicative models for stand variables were linearized by logarithmic transformation. Multiplicative structure and the logarithmic linearization can be justified by the known dependence among the quadratic mean diameter (D_q), stem number, and basal area (see Gove and Patil 1998, Paper **V**). From Equation 11, it follows that basal area and stem number can be written in logarithmic form as $\ln(G) = \ln(k) + 1 \ln(N) + 2 \ln(D_q)$ and $\ln(N) = \ln(1/k) + 1 \ln(G) - 2 \ln(D_q)$, where $k = [\pi/(2 \cdot 100)^2]$. Replacing D_q with other mean characteristics (D, d_{gM}, or D_{dom}) in the linear prediction for $\ln(G)$ or $\ln(N)$ is assumed to follow the above structures but with different coefficients (estimated effects).

A typical application is predicting values for unknown stand characteristics, which are required in, for example, the models implemented in stand simulators. With respect to Finnish size distribution models (e.g., Päivinen 1980, Mykkänen 1986, Maltamo 1997, Kangas and Maltamo 2000b) and the common FMP data, this means the need for G, d_{gM}, and h_{gM} for young stands when only N, D, and H are known. As an example, when one is calibrating G with N and D, the RMSE decreased from 43.3% to 15.9% and, further, with H, to 15.4% according to estimated covariance structure (Paper **V**: Table 5, 7). Similarly, the RMSE of the expected d_{gM} (20.0%) and h_{gM} (19.9%) were reduced to 10.5% and 8.4%, respectively. Thus, mean height reduced the RMSE of basal area only marginally but that of h_{gM} by 3.6 percentage points (see Paper **V**: Table 7).

4.2 Models providing relationship between diameter and height

4.2.1 Näslund's height curve prediction

A linearized Näslund's (1935) height curve (see Equation 9) was selected as the basis for modelling the relationship between dbh and height in advanced stands. The dimensions of the basal area median tree explained 61%, 75%, and 64% of the variation in β_1 for pine, spruce, and birch, respectively (see Paper **I**). After prediction of β_1, β_0 was solved for from Equation 24 such that the curve passes through a known/predicted point, (d_{gM}, h_{gM}).

$$\beta_0 = d_{gM}/[(h_{gM}-1.3)^{\alpha^{-1}}] - \beta_1 d_{gM} \tag{24}$$

(See the notation for equations 8 and 9). The height curve for pine typically bends more than that of spruce (see Paper **II**: Figure 2). The height–dbh relationships presented were regarded as auxiliary models, but they are, of course, significant for prediction of tree volume. Later, it was found that the assumed linear relationship between the parameters of the

Näslund's curve and the median tree characteristics did not hold within the greater variation in the stand developmental stage (see Paper **V**). Instead, the relationships between β_1 and d_{gM}, h_{gM}, D_{dom}, and H_{dom} were shown as curvilinear, but they could be linearized by taking logarithms of both sides (see Figure 2). The basic models including stand age, origin, and site factors explained 36% and 78% of the variation in $\ln\beta_0$ and $\ln\beta_1$, respectively (see Paper **V**: Table 4). The corresponding RMSEs were 0.25 and 0.11. When the expectations of $\ln\beta_0$ and $\ln\beta_1$ were calibrated with common FMP data (e.g., G, d_{gM}, and h_{gM}), the RMSEs were reduced to 0.22 and 0.08, and, further, with additional knowledge of D_{dom} and H_{dom} to 0.19 and 0.07, respectively (see Paper **V**: Figure 3). The corresponding degrees of determination were 52% and 87% in the former and 74% and 90% in the latter case for calibrated $\ln\beta_0$ and $\ln\beta_1$, respectively. Thus, parameters β_0 and β_1 could be efficiently calibrated with the BLUP models.
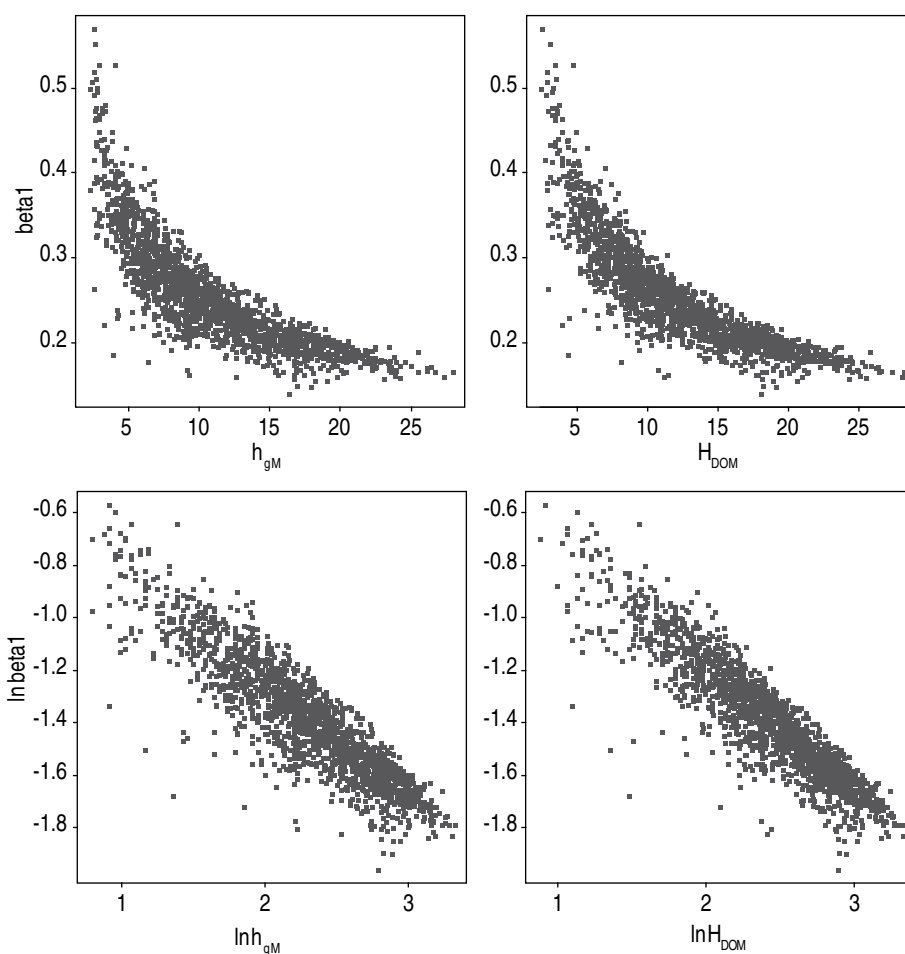
Figure 2. The relationship between parameter β_1 of the Näslund's height curve and the median (h_{gM}) and dominant height (H_{dom}) on the initial and logarithmic scale.

4.2.2 Comparison of models for tree height and diameter

The relationships between diameter and height formulated in Papers **I**, **IV**, and **V** were compared with each other. This example represents a 28-year-old Scots pine stand on a VT site (plot 1010 in the INKA data set) having a dominant height of 10 m. The standard error of the fitted Näslund's curve was utilized for predicting random variation in tree height for the tallied trees. These tree heights are shown as open symbols in contrast to measured sample trees with filled symbols, in Figure 3. When height curve was predicted using the models in Paper **I** for Näslund's parameters, the relationship between dbh and height was too nearly linear and obviously biased in this stand. Instead, a BLUP model for the Näslund's curve, calibrated with three predictors (N, D, and H) or especially with additional D_{dom} and H_{dom}, showed a good and an excellent fit to the treewise data, respectively (Figure 3). The pairs of stand characteristics, such as D and H or d_{gM} and h_{gM}, are useful for the slope, whereas dominant tree characteristics are useful for the bending of the Näslund's curve.

The almost linear dbh curve seems to show good fit (see Figure 3). Because the residual errors of dbh or height models are calculated against the opposite axis, the shape of the dbh–height curve and of the height–dbh curve differ more or less from each other in any case, especially if they provide unbiased estimates. The evident difference is that the dbh curve is less concave against the dbh-axis than is the height curve. This principal difference is important in selection of the approach for generating the trees of the initial stand for a simulator. If we have, for example, a threshold height, where the height distribution approach is replaced with a dbh distribution approach, the threshold should be selected such that the generated trees are as similar as possible around this threshold.
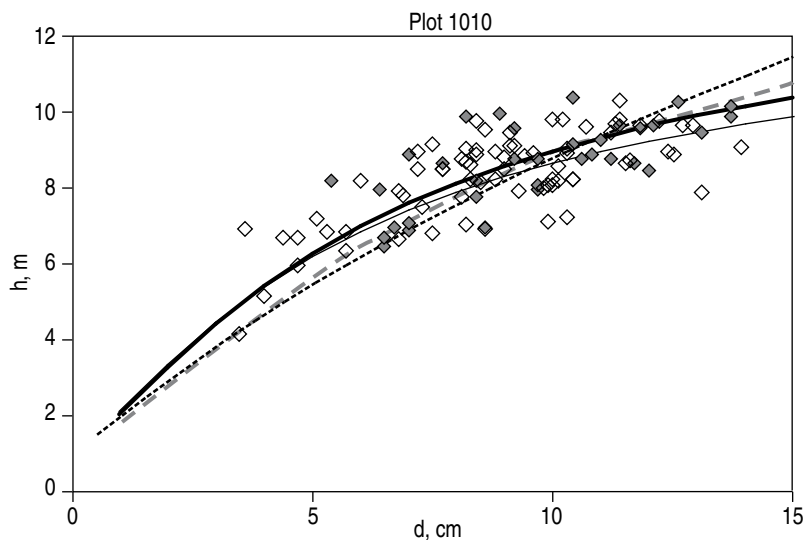
Figure 3. The relationship between tree dbh and height according to models in Papers **I** (– –), **IV** (····), and **V** (— for three and — for four calibrating variables). Note the difference that the residual error in the dbh–height model (**IV**) is calculated against height-axis while errors in height models (**I, V**) are calculated against dbh-axis. Measured sample trees are indicated by a filled symbol ◆ and tallied trees with expected height ± random error by '◊'.

Bias and RMSE for the height characteristics (see Table 2) are calculated for the original predictions and after forcing of the height curve to pass through the point (d_{gM}, h_{gM}) (i.e., Equation 24 was used for the BLUP model for the Näslund's curve and the calibrating coefficient was calculated for Veltheim's model). The BLUP model for the Näslund's curve showed superior performance in both young and advanced stands with the randomised test data set. Its use resulted in an RMSE of 3.2–14.4% in height characteristics. Models by Veltheim (1987) and by Mehtätalo (2005) provided much larger RMSEs, of 14–27% and 12–19%, respectively, because of their strongly averaging nature (see Table 2). The commonly applied solution forcing the Veltheim's height curve through d_{gM}, h_{gM} (e.g., Packalén and Maltamo 2008) reduced the bias and RMSE of the model, achieving accuracy quite comparable with that of the Näslund's curve. Nevertheless, the BLUP model still provided the best performance (see Table 2). The model by Mehtätalo (2005) could be efficiently calibrated with treewise data, but the option of using sample trees was ignored here.

Given the bias in tree height with respect to d/D, the BLUP model for the Näslund's curve that utilized additional dominant tree characteristics performed slightly better than that with the common FMP variables in the validation test data set (see Figure 4). The model for the Näslund's curve for Socts pine described in Paper **I** performed generally quite well for advanced stands. That model was unbiased around the mean diameter because the curve passed through the point (d_{gM}, h_{gM}). However, the height of the suppressed trees ($d/D < 0.4$) was overestimated by about 50 cm (15%). In young stands, the BLUP model for the Näslund's curve produced unbiased height across the variation from suppressed to dominating trees (see Figure 4). The model in Paper **I** was applied outside the range of its modelling data in young stands, resulting in a more severe underestimate in tree height the smaller the relative dbh was (Figure 4). Each model provided practically unbiased tree heights for the thickest trees ($d/D > 1.3$) in young and in advanced stands.

Table 2. Bias and, in brackets, RMSE, as percentage of the mean and dominant height generated with the Weibull distribution with Näslund's model in Paper **V** (BLUP), models by Veltheim (1987) and by Mehtätalo (2005); the effect of the calibration through the point (d_{gM}, h_{gM}) is given for models in Paper **V** and Paper **I** and in Veltheim (1987)

		Expectations			Through point (d_{gM}, h_{gM})		
		BLUP	Veltheim	Mehtätalo	BLUP	Paper I	Veltheim
Young	H	-0.61	13.66	-6.74	-0.59	1.27	-1.30
		(3.2)	(22.2)	(16.4)	(2.3)	(2.4)	(3.8)
	H_{dom}	4.25	16.14	-4.23	4.21	1.01	1.48
		(14.4)	(27.0)	(19.4)	(14.1)	(14.3)	(14.3)
Advanced	H	-2.06	9.38	-0.27	-2.05	-2.91	-1.11
		(7.6)	(17.0)	(16.6)	(7.5)	(8.2)	(7.8)
	H_{dom}	-0.08	8.24	3.02	-0.05	0.77	-2.10
		(3.4)	(14.0)	(12.1)	(3.2)	(3.6)	(4.0)
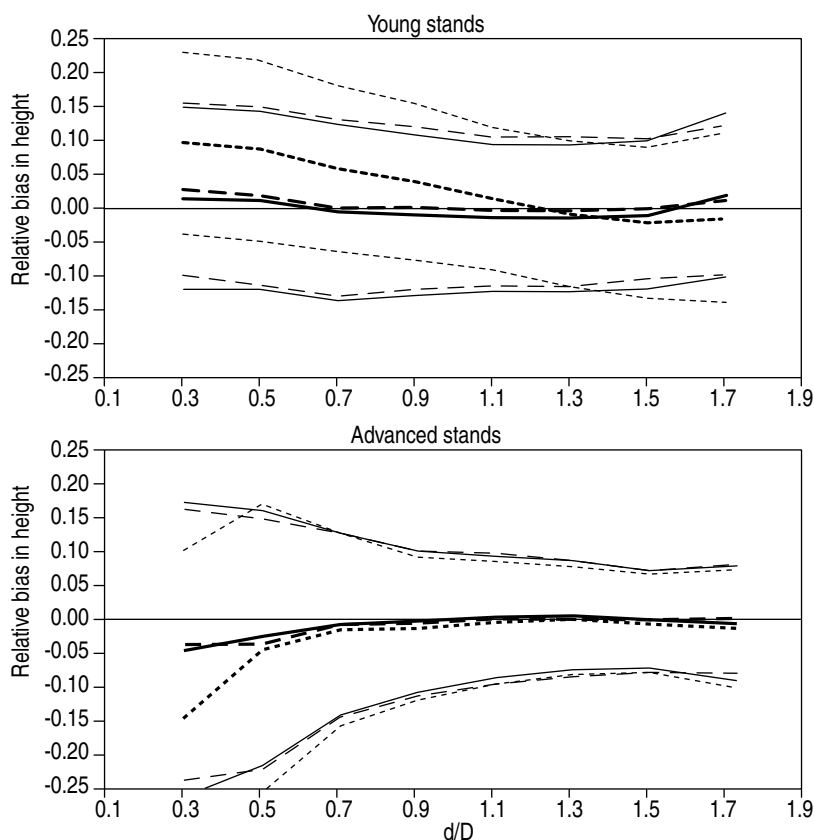
Figure 4. Relative bias (± standard deviation) in tree height with the following models applied for the Näslund's height curve: BLUP model (Paper **V**) with FMP input variables (—), BLUP model with additional D_{dom} and H_{dom} (– –), and OLS model (Paper **I**) with FMP variables (- - -).

4.2.3 Prediction of within-dbh-class height variation

A method using the predicted error term of the linearized Näslund's curve provided a reasonable basis for handling individual tree diameters and heights (Paper **II**). In this approach, the height distribution, conditional to known dbh, remains symmetrical over the variation in dbh. An alternative approach, the bivariate SBB distribution, enabled reasonable description of tree dimensions as a joint distribution of diameters and heights (see Paper **II**). The shape of the conditional height distribution changed from positively skewed to symmetrical and finally negatively skewed as relative dbh increased (see Siipilehto 1996).

The height variation seemed to be greater for pine than for spruce, particularly within the smallest dbh classes (see Paper **II**: Figure 2). It is obvious that the traditional method using dbh distribution and expected height diminishes the variation in tree heights, especially for shade-intolerant Scots pine. This was shown in KS goodness-of-fit tests for Scots pine stands. When including random variation in tree height, marginal distribution for tree heights rejected only 1–4% of the cases, whereas without error-structure, rejections amounted to 11

and 21% for the modelling and test data sets, respectively (see Paper **II**: Table 8). However, this phenomenon was not as evident with shade-tolerant Norway spruce, because of the originally wider dbh distributions and a less concave relationship between tree dbh and height. Consequently, differences between the KS tests using a univariate and a bivariate approach were only marginal for Norway spruce (Paper **II**: Table 5).

Visual assessments of the randomly generated tree dimensions from the above-mentioned bivariate systems did not prove any great differences between the methods applied. Generated tree dimensions for Scots pine were regarded as outliers in 0.28% of the cases with Näslund's model and 0.15% with the constrained SBB model (Paper **II**: figures 5 and 7). The unconstrained SBB model generated a considerably greater proportion of outliers (0.5%) quite probably because the unrealistic height–diameter relationships were accepted (see Paper **II**: Figure 6). Also marginal were the differences in the total volume and timber assortments (Paper **II**: tables 6 and 7 for spruce and 9 and 10 for pine).

An important practical application using the height–diameter relationship is to generalize tree height for tally trees (see Lappi et al. 2006). Depending on the application, it may be wise to include random variation in the expected value in order to produce a realistic picture of the tree dimensions and their variations. If so, we will have greater variation in tree heights, which, in turn, results in greater variation in tree volume and, thus, in total and assortment volumes in comparison with the variation using the expectation values. Using Näslund's height model in combination with the error-term seems convenient because the linearized form of the Näslund's curve is easy to fit and the error can be assumed to be homogeneous and normally distributed (Näslund 1936, p. 52, Paper **II**). This method was applied in Paper **V** in order not to overstate the accuracy in height and volume characteristics, used as validation criteria in the model comparisons.

4.3 Prediction of the size distributions

4.3.1 Model formulation

The form of the median tree (h_{gM}/d_{gM}) was applicable in explaining some of the variation in the shape of SB distribution (see Paper **I**). This was probably because slenderness reflected changes in management history (Hynynen and Arola 1999, Niemistö 1994). From the model formulation point of view, it is noteworthy that the dimensions and the form of the median tree do not necessarily change much as a result of thinning, unlike basal area and stem number. This means that the shape index (see Equation 14) can follow immediate changes in the growing stock of a stand. Thus, considerable improvement was found in the models for the maximum (λ) and shape (δ) of the SB pdf when one assumes knowledge of stem number and applies it in the form of the shape index and its transformations. This was especially the case for the shape parameter, since the degree of determination increased from 5–28% to 38–50% (Paper **I**: tables 4 and 6). However, a disadvantage of the model formulation was also found. The models constructed in Paper **I** did not respond reasonably to increased stand density outside the range of the modelling data (i.e., when the shape index fell to below 0.55). The shape index (denoted as ψ) varied from 0.17 to 0.99 in the drained peatland stands, and it proved to be an extremely useful predictor. Indeed, its linearized transformation $\ln(\psi+1.6)^8$ alone explained 58% and 74% of the variation in $\ln(\lambda)$ and $\ln(\delta)$, while the degree of determinations of the final OLS models was 62% and 82%, respectively (see Paper **III**: Figures 1 and 2; see also Appendix 1). The degree of determination is a highly data-dependent characteristic, but, by contrast, the RMSE of the model is more comparable with the other RMSEs between the

optional models and between different modelling data sets. Thus, the RMSEs of the OLS model for $\ln(\delta)$ for Scots pine was 0.30 without the shape index and 0.25 with the shape index (see Paper **I**) and 0.16 in the OLS model using the transformations $\ln(\psi+1.6)^8$, $\ln(\psi+1)^3$ and $\ln(\psi+1)^4$ (see Paper **III**). These improvements were mainly due to model formulation.

It is well known that the Weibull distribution is theoretically not as flexible as the SB distribution with respect to variation in the shape of the distribution (e.g., Hafley and Schreuder 1977, Paper **I**). In practice, this feature may have major or minor relevance, depending on the ability to predict the variation of the ML estimated parameters from available stand characteristics. The degree of determination of the Weibull parameter c was rather low – namely, 39, 22, and 18% with the common FMP input data (G, d_{gM}, and h_{gM}) for pine, spruce, and birch, respectively (see Paper **I**). Including the shape index improved the models for spruce ($r^2 = 51\%$) and birch (24%) but not that for pine.

The Weibull distribution was not able to perform efficient prediction with the available FMP inventory data or even assuming the additional knowledge of stem number. Therefore, I proposed trying additional dominant tree characteristics. Dominant height showed its potential in relation to the Weibull distributions for tree heights in juvenile Scots pine stands (see Paper **IV**: Figure 1). Indeed, the transformation $1/\ln(H_{dom}/H)$ could on its own explain about 90% of the variation in the shape parameter c. When it comes to BLUP estimation for the dbh distribution parameters in Paper **V**, some aspects are worthy of mention. If we assume that the ratio of two stand characteristics is important for prediction/calibration of the model, the multiplicative model structure takes this automatically into account. Assume that $Y_i = f[\ln(x_1/x_2)]$, it follows that $Y_i = f[\ln(x_1)-\ln(x_2)]$. Because c of the height distribution was closely correlated with $1/\ln(H_{dom}/H)$, it follows that $1/c$ has to be closely correlated with $\ln(H_{dom}/H)$. Therefore, in calibration of the model for $1/c$ of the dbh-frequency distribution with D and D_{dom}, the ratio can be written in the form $\ln(D_{dom})-\ln(D)$. In the above case, the coefficient is calculated for each previous term from the variance–covariance matrix of the corresponding residuals by means of Equation 19. The same assumptions applied also for models for δ and c of the SB and Weibull basal area-dbh distributions, respectively. The inverse of δ and inverse of c of the basal area-dbh distributions were also closely related to the ratios of two mean stand characteristics, but the logarithmic transformations $\ln(\delta)$ and $\ln(c)$ performed slightly better in general (see Paper **V**). One general advantage of the logarithmic transformation is that it ensured the requirement of the positive value for c and δ.

4.3.2 Regression estimation techniques

The regression model estimation techniques including OLS, mixed, SUR, and the multivariate mixed model denoted as MSUR were validated in Paper **III**. Although each of these models provided quite excellent estimates for stem number, 'volume' as $\sum dbh^3$ and 'value' as $\sum dbh^4$ of the stock, it was evident that the more advanced techniques provided enhanced model performance. Thus, the mixed models with a random stand component were superior to OLS even though the mixed-effect models were validated as fixed models. SUR took into account the correlation between the residuals of the regression models ($r = 0.74$), while MSUR accounted for both the random stand effect and the crossmodel error correlation. Both of the last mentioned were superior to the OLS and mixed models. Surprisingly, the ranking of the model through generated stand characteristics (i.e., model application) seemed opposite the ranking of the error terms of the estimated models (i.e., model statistics). Indeed, s_e of 0.262 with OLS was smaller than $s_e + s_{stand}$ of 0.265 in the mixed model, which was smaller than $s_e + s_{stand}$ of 0.270 in MSUR for parameter λ (see Paper **III**: Appendix 1). In any case, the

significance of the estimated parameters is more reliable in the models with mixed effects because the dependencies among repeated measurements are better accounted for (Lappi 1993, p. 68).

Each model for height distribution was statistically sophisticated (see Paper **IV**). Parameter prediction models, using PPM, were estimated according to the mixed-effect SUR approach, whereas the generalized linear model fitted the distributions and estimated the prediction model in a single step (Cao 2004). Note that the latter model, because of its complicated structure, did not include mixed effects. Both approaches were further combined with the moment estimator (see Equation 12), denoted as PPM+M and GLM+M. Again, each model provided excellent results and the differences between estimated models were quite marginal. The similar performance was simply due to exceptionally close correlation between the shape parameter c and the ratio between H_{dom} and H. However, the overall best performance was provided by the hybrid GLM+M (see Paper **IV**: Table 5). Consequently, we can say that the advanced estimation technique resulted in improved model performance.

As before, the accuracy of the applications could not be seen directly in the estimated models – namely, in the approximate standard errors (ASEs) of the estimated parameters (see Paper **IV**: Table 3). This was simply because in the traditional PPM, the Weibull distribution parameters first were estimated for each stand and in the second step they were used as true values (without error) in PPM estimation. Instead, in the GLM, the ASEs were calculated from the tree-level model fit and showed about three times higher values than the ASEs in the PPM approach. In any case, the inclusion of the moment estimator in GLM+M reduced the ASEs of the remaining estimated parameters in comparison with the original GLM (Paper **IV**: Table 3).

4.3.3 Behaviour of the distribution models

The optional diameter distribution prediction models were compared, first with respect to varying stem numbers in the assumed 25-year-old pine stand with fixed d_{gM} = 10 cm and G = 10 m²ha⁻¹. Distributions are predicted by means of stem number N = 3,100 ha⁻¹ for high density, N = 2,500 ha⁻¹ for moderately high density, N = 1,900 ha⁻¹ for moderate density, and N = 1,300 ha⁻¹ for low density (shape indices between 0.41 and 0.98) of a forest stand. The most obvious changes in the shape of the predicted distribution could be seen with the SB MSUR model (see Figure 5A). Simultaneously, the resultant errors in stem numbers were only 1, 3, -5, and -4%, from the highest to the lowest stand density. Distributions according to the SB_G BLUP model (see Figure 5B) did not vary enough with respect to stem number. The corresponding errors in N were 19, 16, 7, and -11%. Even more inadequate response to stem number variation was found with the Weibull models. Distributions from the W_G BLUP model (see Figure 5C) resulted in errors in N of 34, 23, 6, and -28%. When the W_N BLUP model (see Figure 5D) was scaled to a known basal area of 10 m²ha⁻¹, the resulting errors in N were 27, 17, 4, and -16%. Thus, response to the variation in stem number was better achieved in W_N as compared with the W_G model, but both SB models were superior to the Weibull models.

The high-density pine stands under the SB ($G+N$) model showed bimodal diameter distributions (see Paper **I**: Figure 6), while distributions for dense stands resembled almost decreasing distributions with the SB MSUR model (see Figure 5). Additionally, even much higher densities could be included in the SB MSUR model successfully, resulting in an inverse J-shaped distribution. For example, N of 6,000 ha⁻¹ (shape index 0.21) still resulted in less than 9% error in the stem number generated. In conclusion, efforts to improve the ability of a model
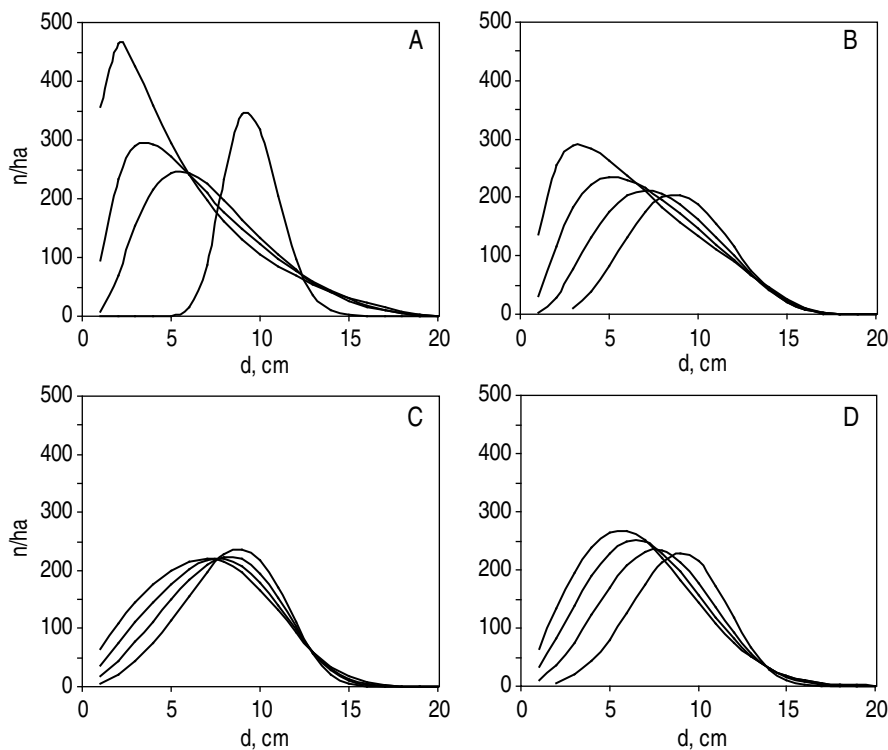
Figure 5. Variation in the shape of the predicted distribution with respect to variation in stem number (N of 3,100, 2,500, 1,900 and 1,300 ha^{-1}) when d_{gM} was 10 cm and basal area 10 m^{2}ha^{-1}. The models applied are: (A) SB MSUR (Paper **III**), (B) SB$_G$, (C) W$_G$, and (D) W$_N$ BLUP (Paper **V**).

to respond to variation in stem number succeeded rather well with the SB pdf, especially by means of the SB MSUR model presented in Paper **III**.

In the next comparison, the behaviour of the distribution model was checked with variation in the dominant diameter assumed while the mean diameter, D, was fixed to 8 cm for a 25-year-old pine stand. D_{dom} of 12, 16, 20, and 24 cm represented D_{dom}/D ratios of 1.5 to 3. The dominant diameter that was calculated from the predicted distribution was generally quite close to that of the given input variable. The relative differences from the smallest to the highest value of D_{dom} were as follows: SB MSUR in Figure 6A: -10.9, -3.7, 4.8, and 12.0%; SB$_G$ BLUP in Figure 6B: -2.8, -2.3, -1.7, and -0.7%; W$_G$ BLUP in Figure 6C: 0.2, -0.7, -1.2, and -1.2%; and, finally, W$_N$ BLUP in Figure 6D: -4.4, -2.9, 3.6, and -4.1%. Thus, the BLUP models detected 99–103%, whereas the SB MSUR model detected 65% (D_{dom} 13.3–21.1 cm), of the given 12–24-centimetre variation. According to the accuracy calculations, the most accurate response to the dominant diameter was obtained with the W$_G$ and SB$_G$ BLUP models. Nevertheless, the SB MSUR model detected the variation in D_{dom} considerably well through calibration of input variables N, G, and d_{gM} with known D and D_{dom}.
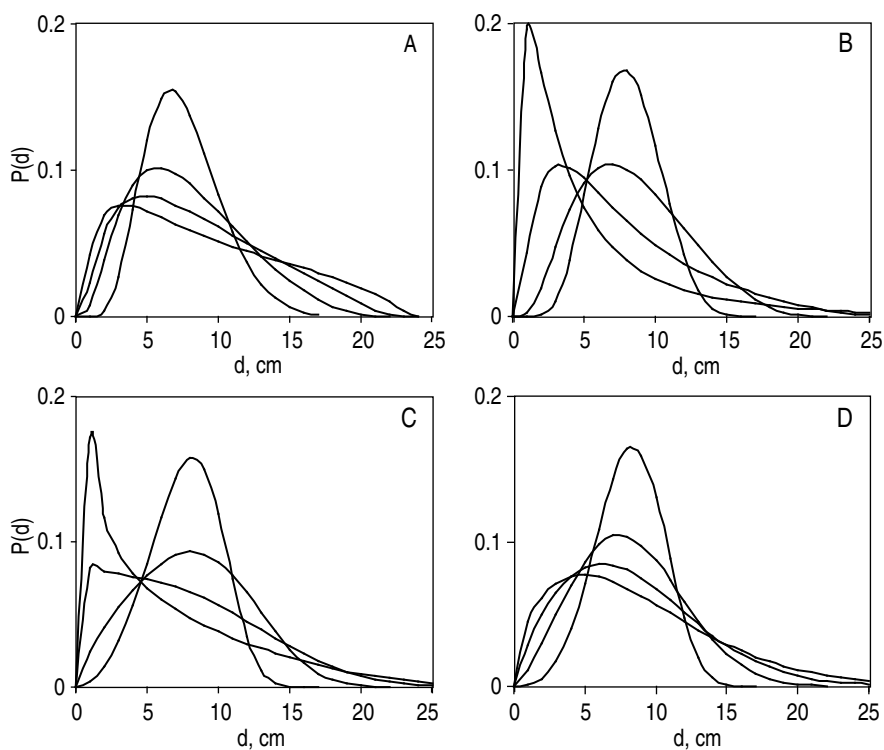
Figure 6. Variation in the shape of predicted distribution with respect to variation in dominant diameter (D_{dom} of 12, 16, 20, and 24 cm) when D was 8 cm and age 25 years. The models applied are: (A) SB MSUR (Paper **III**), (B) SB_{G}, (C) W_{G}, and (D) W_{N} BLUP (Paper **V**).

4.4 Final model validation

4.4.1 Model performance for advanced stands

Each distribution model performed similarly, underestimating sawlog volume and simultaneously overestimating the pulpwood volume. The SB_{G} BLUP model with additional N resulted in the least bias in total and commercial wood volumes, whereas SB MSUR provided smaller biases than the W_{N} BLUP model did (see Table 3). In general, differences between biases among basal area-dbh distribution models were quite small. An exception was the considerably greater bias in waste wood volume with the SB MSUR model (see Table 3). In terms of bias, the most beneficial additional stand characteristic was stem number.

The differences between the distribution models in the RMSE of total volume were hardly visible. The highest RMSE in total volume was generated with W_{N} BLUP using the additional knowledge of dominant dbh and height (2.4%), while the lowest RMSE was provided by W_{G} BLUP and SB MSUR with the additional knowledge of stem number (1.7%). In general, the dbh-frequency W_{N} BLUP model provided the worst validated volume characteristics (see Figure 7). If either the common FMP data or additional dominant tree characteristics were available, the W_{G} BLUP model provided the most accurate total and timber assortment

Table 3: Bias (%) in stand characteristics for advanced stands (H ≥ 9 m) according to W_N, W_G, and SB_G models in Paper **V** and SB distribution models in Paper **III** (MSUR). The smallest biases are highlighted in bold with respect to the combination of input variables.

Model	Predictors	Volume	Logs	Pulp	Waste	N	D_{dom}	H_{dom}
W_N	G, d_{gM}, h_{gM}	1.15	9.95	-9.87	**1.75**	**-1.51**	1.65	0.40
	H_{dom}	1.12	10.04	-10.06	**2.52**	**-1.16**	1.79	0.52
	D_{dom}	1.02	9.57	-9.81	2.93	-0.86	1.75	0.37
	N	0.90	7.28	-7.37	**-0.59**	-1.89	0.76	**0.18**
	D_{dom}, H_{dom}	0.96	9.51	-9.87	2.82	-0.86	1.75	0.38
W_G	G, d_{gM}, h_{gM}	**0.17**	**1.56**	-2.22	4.99	2.39	0.10	**-0.16**
	H_{dom}	**0.19**	**1.82**	-2.56	5.47	2.41	0.29	**-0.02**
	D_{dom}	**0.20**	1.98	-2.58	2.55	**0.54**	0.23	**-0.16**
	N	0.19	1.63	-2.15	3.39	1.62	**-0.16**	**-0.18**
	D_{dom}, H_{dom}	**0.14**	1.93	-2.62	2.27	**0.49**	0.22	**-0.15**
SB_G	G, d_{gM}, h_{gM}	0.18	1.60	-2.16	3.61	2.32	-2.09	-0.80
	H_{dom}	0.28	2.60	-3.22	2.58	1.53	-1.51	-0.54
	D_{dom}	0.57	4.82	-5.13	**-1.28**	-1.13	-0.38	-0.31
	N	**0.15**	**0.57**	**-0.37**	-1.20	**-0.58**	-2.61	-0.92
	D_{dom}, H_{dom}	0.52	4.80	-5.21	**-1.63**	-1.21	-0.39	-0.29
SB MSUR	G, d_{gM}, h_{gM}	0.53	1.92	**-0.98**	-6.28	-3.14	**-0.05**	-0.22
	H_{dom}	0.53	2.00	**-1.10**	-6.16	-2.99	**-0.04**	-0.10
	D_{dom}	0.46	**1.84**	**-1.09**	-5.70	-2.86	**0.02**	-0.26
	N	0.34	1.48	-1.10	-2.80	-1.55	-1.07	-0.48
	D_{dom}, H_{dom}	0.40	**1.76**	**-1.09**	-5.94	-2.89	**0.02**	-0.25

volumes. The additional knowledge of stem number benefited the SB MSUR model in particular (see figures 7 and 8), but, simultaneously, the differences between the basal area-dbh distribution models become very slight, except those in generated waste wood volume (see Figure 7) and stem number (see Figure 8). Indeed, the SB MSUR model resulted in an RMSE of 14% in waste wood while that with W_G BLUP was 25%. Besides, stem number knowledge reduced RMSE in generated N to only 5% using the SB MSUR model, while the W_N, W_G, and SB_G BLUP models resulted in an RMSE in N of 6, 14, and 9%, respectively. When the dominant tree characteristics were included, the improvement in the D_{dom} and H_{dom} generated was the most obvious with the SB_G BLUP model. Indeed, the RMSE of 5.2% in D_{dom} and 3.5% in H_{dom} for the common FMP data fell to only 1.6% and 2.3%, respectively (see Figure 8). Unfortunately, this obvious improvement was not reflected similarly in sum characteristics, stem number, or volume. In contrast, each model provided a slightly more inaccurate total volume (RMSE of 2.1–2.4%) in comparison with the accuracy found with FMP data (RMSE of 1.7–2.2%). Further, the W_G BLUP model provided the least accurate N (RMSE of 17%). According to Figure 7, stem number seemed a more appropriate additional stand variable than dominant tree characteristics for use for calibration/prediction of dbh distribution for advanced stands. Note that the results from test data were in line with the given examples of distribution model behaviour (see figures 5 and 6).

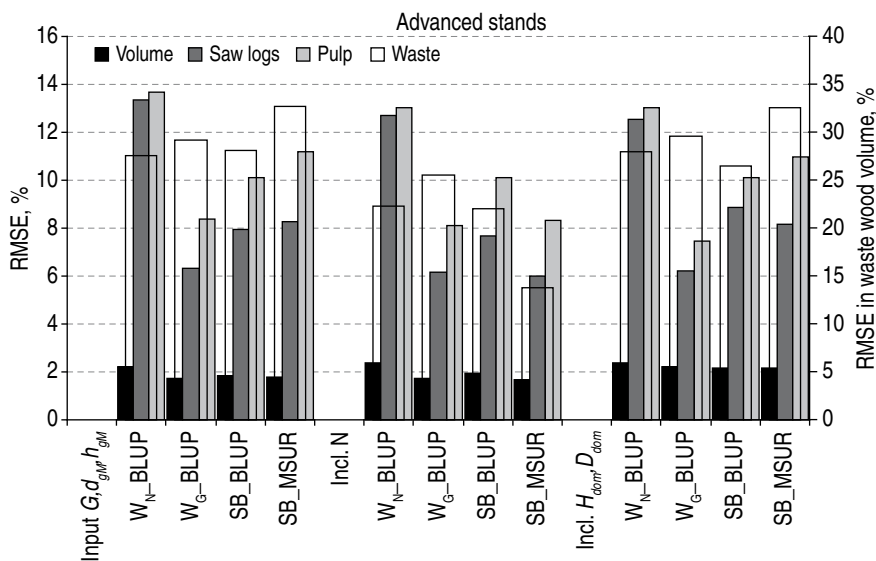**Figure 7.** RMSE (%) in assortments according to Weibull and SB models with known G, d_{gM}, and h_{gM} and including additional knowledge of N or H_{dom} and D_{dom} for advanced stands.

Figure 8. RMSE (%) in D_{dom}, H_{dom}, and N according to Weibull and SB models with known G, d_{gM}, and h_{gM} and including additional knowledge of N or H_{dom} and D_{dom} for advanced stands.
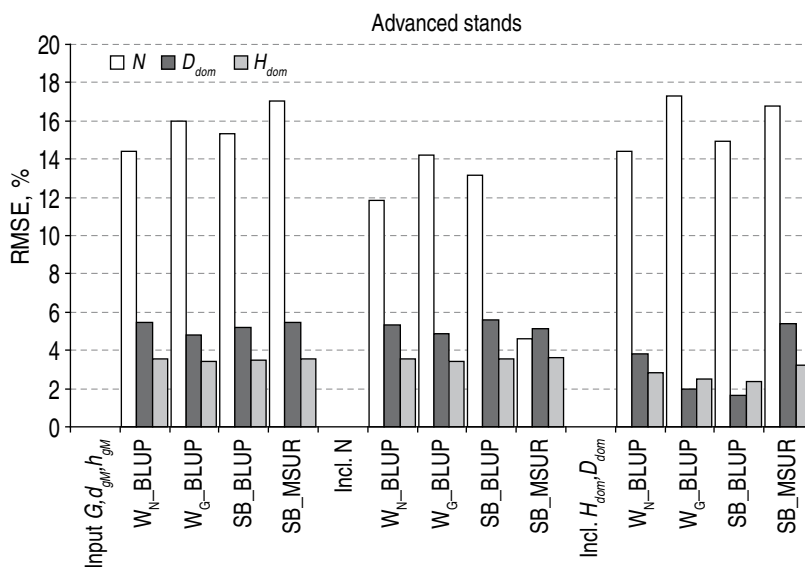
4.4.2 Model performance with young stands

Table 4. Bias in stand characteristics according to alternative Weibull and SB distribution models for young stands – models included the Weibull and SB distributions: W_N, W_G, and SBG models in Paper **V**, Weibull height distribution in Paper **IV**, and SB MSUR in Paper **III** (the smallest biases are highlighted in **bold** with respect to the combination of input variables)

Model	Predictors	Volume	Pulp	Waste	G	D_{dom}	H_{dom}
W_N	N, D, H	4.62	5.70	**0.30**	1.40	3.51	3.64
	H_{dom}	**1.22**	**1.58**	-0.25	**-0.17**	1.30	1.70
	D_{dom}	**1.72**	**2.34**	-0.84	**-0.08**	1.29	2.63
	D_{dom}, H^{dom}	**1.27**	**1.74**	-0.68	**-0.27**	1.09	1.98
W_G	N, D, H	3.61	3.04	6.12	1.55	1.17	2.64
	H_{dom}	4.22	3.81	6.01	3.84	**0.49**	1.40
	D_{dom}	8.10	8.66	5.76	6.56	1.26	2.64
	D_{dom}, H_{dom}	6.56	6.80	5.53	5.65	0.92	1.94
SB_G	N, D, H	**1.33**	**0.88**	3.27	**-0.54**	**-0.51**	1.99
	H_{dom}	3.59	3.61	3.48	2.92	-0.83	0.96
	D_{dom}	7.60	8.71	3.11	5.75	**0.60**	2.40
	D_{dom}, H_{dom}	6.18	6.93	3.13	4.93	**0.00**	1.62
W height	N, D, H	9.11	12.66	-3.73	5.04	10.06	**-0.56**
	H_{dom}	1.49	2.94	-4.19	1.72	5.83	-2.51
	D_{dom}	5.78	8.61	-4.71	3.57	7.26	-1.74
SB MSUR	N, D, H	8.59	9.14	6.32	6.73	-1.23	1.65
	H_{dom}	5.42	5.57	4.76	5.60	-1.70	**0.46**
	D_{dom}	5.66	5.91	4.59	5.33	-1.55	**1.32**
	D_{dom}, H_{dom}	5.35	5.51	4.68	5.39	-1.62	**0.70**

In young stands, the frequency-based Weibull models (height distribution of Paper **IV** and dbh distribution of Paper **V**) performed better in general than did basal area-dbh distribution models (see figures 9 and 10). The difference between these models was the most obvious in the accuracy of waste wood (see Figure 9) and basal area (see Figure 10). In terms of bias, the W_N BLUP model with H_{dom} performed the best, whereas in terms of RMSE, the combinations of the W_N BLUP model with H_{dom} and D_{dom} as additional variables to FMP data showed the best performance. Indeed, in the former case, the W_N BLUP model resulted in bias of less than 2% in volume characteristics (see Table 4 and also Paper **V**: Table 8), and in the latter, the RMSE percentage was 11–15% for total, pulpwood, and waste wood volumes (see Figure 9). Additional D_{dom} slightly improved the W_N BLUP and SB MSUR model behaviour in terms of bias (see Table 4) but, at the same time, achieved a twofold to threefold increase in the biases with W_G BLUP and SB_G BLUP models used for young stands (see Table 4). Clear underestimations in total and pulpwood volumes could be explained by the underestimated basal area and H_{dom} whenever D_{dom} was used as a calibrating variable for the W_G and SB_G
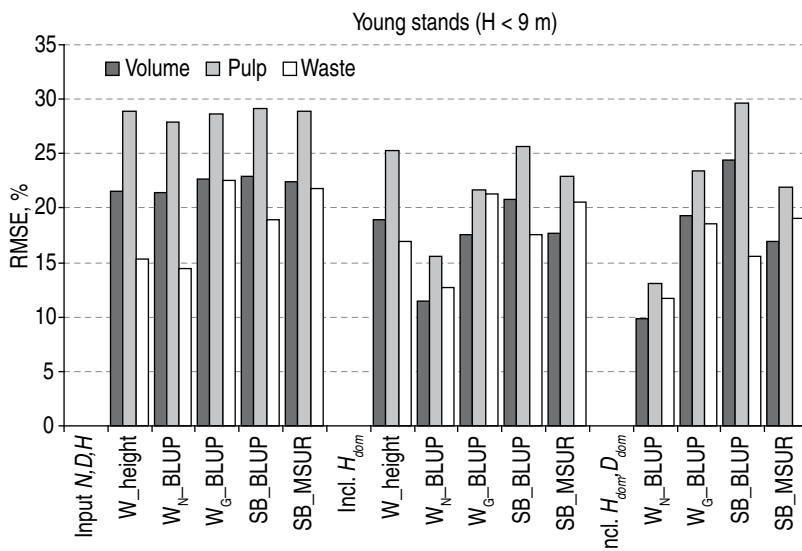
Figure 9. RMSE (%) in volume characteristics according to the Weibull and SB models with *N, D*, and *H* and including additional knowledge of H_{dom} and D_{dom} for young stands.
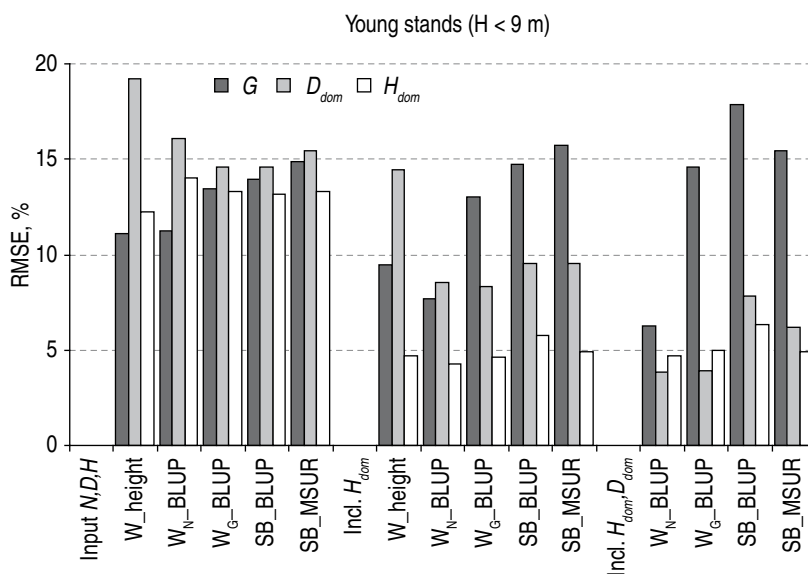
Figure 10. RMSE (%) in *G, D_{dom}*, and H_{dom} according to Weibull and SB models with *N, D*, and *H* and including additional knowledge of H_{dom} and D_{dom} for young stands.

BLUP models. In addition, the resulting RMSE in total and in pulpwood volume was about twice that of the W_N BLUP model provided for young stands (see Figure 9). Nevertheless, the most accurate D_{dom} (with an RMSE of 3.7%) was yielded by the W_G BLUP model including D_{dom} and H_{dom} as input variables.

Naturally, the height distribution model benefited from the knowledge of H_{dom}, which was an explaining variable in the model. Nevertheless, it could not produce as accurate volume characteristics as did the W_N BLUP model with the same input variables. The most likely reason was the less accurate dbh distribution, which was generated from height distribution and the model for tree dbh. Thus, the bias and RMSE in basal area and D_{dom} according to the height distribution modelling approach were considerably higher than the corresponding numbers from the W_N BLUP model (see Table 4 and also Figure 10). These models seemed to result in an almost equal RMSE of 4.3–4.7% for H_{dom} (see Figure 10).

5 DISCUSSION AND CONCLUSIONS

Many studies using Weibull and SB distribution have focused on the comparisons between distributions and their fitting methods (e.g., Hafley and Schreuder 1977, Holte 1993, Zhou and McTague 1996, Kamziah et al. 1999, 2000, Scolforo et al. 2003, Zhang et al. 2003, Wang and Rennolls 2005, Palahi et al. 2007). The SB distribution and maximum likelihood estimation method were regarded as superior in most of the comparisons. Many fitting methods needed such information about the dbh distribution that they cannot be considered applicable from a Finnish point of view without some auxiliary prediction models. Among the necessary characteristics were the minimum and maximum dbh, standard deviation of dbh, and percentile d95. Minimum and maximum belong to 'extreme order statistics', which means that they are dependent on the sample size. Nevertheless, this phenomenon has not been taken into account in the construction of models for them (see Rennols and Rollinson 1985). In general, moments are easier to predict from stand characteristics than are the parameters of the distribution function (e.g., Knoebel and Burkhart 1991). However, Burk and Newberry (1984) cautioned that moment-based recovery is very sensitive to small changes in moment predictions. Overall, if one particular method is regarded as superior, it should provide more reliable predictions for a new stand than the methods compared with it.

In this thesis, the Weibull, Johnson's SB, and SBB distributions were estimated by means of the maximum likelihood method, which is generally regarded as the best because it utilizes all information measured for the distribution (e.g., Bailey and Dell 1973, Johnson 1949b, Shiver 1988). One parameter was excluded through setting of the minimum dbh to zero. This can be justified from a number of studies (e.g., Hafley and Schreuder 1977, Maltamo et al. 1995, Kamziah et al. 1999, Parresol 2003, Gove 2003b). In general, instead of on the accuracy of the predicted parameters, the focus was placed on the applications in terms of accuracy in the stand characteristics generated (stem number, basal area, dominant tree characteristics, total volume, assortment volume, etc.).

One of the most important objectives for this thesis involved the attempts to provide more variation and simultaneously increase the accuracy in the shape of predicted dbh distributions in comparison with previous models in Finland. Indeed, some previous models could provide only one distribution from the known median diameter and basal area (e.g., Mykkänen 1986, Kilkki et al. 1989, Maltamo's 1997 model for pine). Slight variation could be included through age variation in the models by Kilkki and Päivinen (1986), Maltamo et al. (1995), and Maltamo (1997) for spruce or for variation in mean height by Päivinen (1980). The percentile-based model of Kangas and Maltamo (2000b), including basal area and stem number, could provide a wide range of shapes of dbh distributions, including bimodal or multimodal distribution as a possible outcome. In previous Nordic studies, SB distribution parameters were predicted by means of stem number with mean height (Mønness 1982, Holte 1993) or with additional basal area (Tham 1988).

The species admixtures were not included in the models for species-specific distributions, because of statistical insignificance (Papers **I** and **II**; see also Tham 1988), but were used earlier as dummy variables by Päivinen (1980) and as a species proportion by Maltamo et al. (1995) in distribution models for the whole stand. Furthermore, site factors are only rarely included in Finnish parameter prediction models (Päivinen 1980, Sarkkola et al. 2005) but have been commonly used for predicting size distribution elsewhere, especially in terms of site index (e.g., Mønness 1982, Bowling et al. 1989, Holte 1993, Liu et al. 2004, Westfall et al. 2004).

It was shown that the additional knowledge of the stem number enabled variation in the shapes of predicted distributions and had potential to improve the accuracy of the dbh distribution considerably (see Papers **I** and **III**). Stem number was included in the models by means of the transformation introduced as the 'shape index'. Later, it has been applied as an optional measure of stand diversity, by Rouvinen and Kuuluvainen (2005). For Paper **I**, the modelling data were too limited for obtaining successful formulation of the relationship between shape index and distribution parameters; therefore, some anomalies remained when the models were applied outside the range of the modelling data. By contrast, the data in Paper **III** included extremely wide variation in stand characteristics, such as mean diameter, stand basal area, and stem number. Thus, said data provided an excellent basis for studying the relationship between SB distribution parameters and stand characteristics. According to independent data sets, the models formulated from the drained peatland data (Paper **III**) proved to be superior to the previous Weibull model by Mykkänen (1986) and at least comparable with percentile-based model from Kangas and Maltamo (2000b). This was also the case when these models were applied for pine stands on mineral soil (Paper **III**: Table 4). Furthermore, the SUR and MSUR models addressed in Paper **III** did not show any trend in the bias of stem number with respect to thinning intensity from unthinned to heavily thinned stands. As a whole, models by Kangas and Maltamo (2000b), Mabvurira et al. (2002), and Mehtätalo et al. (2011) have shown the improvement effect of additional stem number on the accuracy of the predicted dbh distributions.

A Weibull function was used to describe height distribution for juvenile Scots pine stands in Paper **IV**. In this case, an efficient transformation was formulated as a ratio between dominant and mean height. It was derived from the percentile estimator for the shape parameter c. Quite a similar transformation was earlier utilized by Sarkkola et al. (2003, 2005). Instead, Nord-Larsen and Cao (2006) used the ratio between D_q and H_{dom}. A model for $1/c$ of the Weibull dbh-frequency distribution was found to benefit greatly from the knowledge of two different dimensions – e.g., D_{dom} and D. Models for $1/c$ of the Weibull pdf (Paper **V**) have been presented also by Robinson (2004) and Mabvurira et al. (2002).

In general, the basal area-dbh distribution models proved superior to the dbh-frequency distribution models for advanced stands, in particular because of the reliability of volume estimates, while for young stands the situation was the opposite (see Paper **V**). The ranking of the basal-area-based models was dependent on the content of known stand variables used for prediction/calibration. In any event, differences, especially in total volume, were usually quite marginal. The most obvious differences were found in the accuracy of the waste wood volume and stem number, both indicating the accuracy versus inaccuracy in the left-tail of the distribution and, thus, in the potential energy wood fraction.

A slightly controversial feature appears in relation to two optional approaches for young stands. Firstly, models can be constructed for continuous height distribution, with dbh modelled from tree height for trees above breast height (Paper **IV**). Secondly, dbh distribution can be modelled for trees above breast height, with height modelled from the known dbh (e.g., Paper **V**). The trees generated from these two optional methods would be quite different, not least because of the differences between the models for individual tree dbh and height. Applying dbh distribution models as early as possible could be wise if one wishes to minimise the differences in generated trees between these two approaches. The motivation is that for the early stage of a stand, the relationship between diameter and height is almost linear but as succession progresses and especially after canopy closure, the relationship becomes more and more curvilinear. The reasonable stage of stand development for switching from a height distribution approach to dbh distribution approaches could be at the mean height of 4 m.

Indeed, according to Paper **IV**, practically all trees have reached breast height, and the dbh–height relationship was still almost linear, when H was 4 m (see the paper's Figure 5).

To my knowledge thus far, an SB MSUR model could be recommended for practical use if the stem number is known in addition to basal area in advanced stands. This model was capable of capturing the wide range in stand densities while yielding only slight biases in stem number. Simultaneously, SB_G BLUP captured about 60%, W_N BLUP 40%, and W_G BLUP only 20% of the range in stem number in the example given (see Figure 6). The most relevant optional models in this situation are the species-specific percentile-based models by Kangas and Maltamo (2000b). In the case of broadleaved species, the SB ($G+N$) models for birch (Paper **I**) are an option. Indeed, both of the above-mentioned models have proved to perform well among the alternative models in use in Finland (Maltamo et al. 2002). Without stem number knowledge, the W_G BLUP model provided a relevant option for the SB MSUR model in advanced stands. When the dominant tree characteristics were included as explaining variables, the W_G BLUP model turned out to be superior to the SB MSUR model, in particular because of improved accuracy in timber assortments (see Table 3 and also Figure 8). However, its obvious weakness was the inaccuracy in the waste wood fraction and stem number, no matter the set of input variables (see figures 6, 8, and 9). The Weibull model for the height distribution, especially the hybrid model, is recommended for practical use for stands in their youngest stage. Even though the model was estimated for Scots pine, it is recommended for all species in seedling and sapling stands, on account of the lack of relevant models for juvenile stands. Furthermore, the Weibull frequency distribution (W_N BLUP) is a recommended option for sapling stands above a mean height of 4 m, not least because its volume estimates are potentially superior to those yielded by the height distribution approach (see Table 4 and also figures 10 and 11).

Finally, if there is a need to mimic forest stand structure that include within-dbh-class height variation, use of Näslund's model in combination with the error structure introduced in Paper **II** and applied in Paper **V** is recommended. Siipilehto (2001a) gave an example from a mature Scots pine stand where the variation in expected tree height was extremely low, from 23 to 26 m, and inclusion of the predicted within-dbh-class height variation increased the variation in tree height from 19 to 28 m, close to the observed variation (see Figure 4 in Siipilehto 2001a). Simultaneously with inclusion of random variation in tree height, the Kolmogorov-Smirnov tests showed improved goodness of fit to the observed height distributions.

The bivariate SBB model is also suitable, especially if a concave height–diameter relationship is a required constraint. It is absolutely necessary that the minimum and maximum endpoints of the marginal distributions be beyond the observed data (e.g., Schreuder and Hafley 1977, Wang and Rennols 2007). If the endpoints are fixed to the minimum and maximum observations, the unreasonable height–diameter curve from SBB is not only a possible but a definite outcome (see Li et al. 2002). Furthermore, fixing the endpoints according to data disables the tails of bell-shaped distribution, resulting in unnecessary lack of fit.

Many studies have dealt with the height–diameter relationship in a stand (e.g., Curtis 1967, Newton and Amponsah 2007, Leduc and Goltz 2009). Among the sigmoid curves applied have been ChapmanRichards, Gomperts, Korf, Richards, Schumacher, Sloboda, and Weibull functions. However, no function has been found to be superior (Mehtätalo 2004). Sometimes the model is based on relative height as a function of relative dbh (Kiviste et al. 2003, Nishizono et al. 2005), forcing the curve to pass through the point (D,H) or close to the point (D_{dom},H_{dom}), as the model by Eerikäinen (2003) does. Korf functions (e.g., Parresol 1992) have been commonly used in Finland (Lappi 1997, Eerikäinen 2003, Hökka and Ojansuu 2004, Mehtätalo 2004, 2005). A model $h = \exp(\beta X)$, in which X includes only tree diameters

(Kilkki 1983) or additional stand and site factors (Veltheim 1987, Eerikäinen 2009), has been linearized using the logarithmic transformation. Models by Veltheim (1987) and Mehtätalo (2005) for tree height did not include any height-based stand characteristics (i.e., H, h_{gM}, H_g, or H_{dom}), resulting in a strong averaging character. Indeed, Newton and Amponsah (2007) noticed the evident improvement in their alternative height models when dominant height was included. As a conclusion we can say that simple height models utilising the set of stand characteristics of common FMP are lacking in Finland.

The Näslund's curve is typically used in order to generalize height for tally trees by fitting the curve to sample tree data (e.g., Heinonen 1994, Elfving and Kiviste 1997) but is seldom modelled for prediction purposes in Finland (Siipilehto 1996, Kangas and Maltamo 2002, Maltamo et al. 2007), the exception being this thesis. The simple height curve models described in Paper **I** have already been utilized in many connections by, for example, Kangas and Maltamo (2003), Pukkala and Miina (2005), and Maltamo et al. (2006). However, a word of caution must be given. I noticed that the assumed linear relationship between the parameter β_1 and the stand characteristics on the original scale did not hold for the data with a wide range of development stages for stands. Instead, the logarithmic transformation of both sides linearized the relationship (see Figure 2). Recently, Korpela and Tokola (2006) found a slight bias when using the height model of Paper **I**. When models by Veltheim (1987) and Mehtätalo (2005) were compared with Näslund's model as specified in this thesis, the resulting RMSEs in H and H_{dom} were manifold. The *a posteriori* calibration of the alternative height models resulted in quite comparable accuracy, but, still, the height model in Paper **V** proved slightly better, whereas Veltheim's model and the model in Paper **I** provided almost identical accuracy. Also, Packalén and Maltamo (2008) noticed similarity in terms of volume estimates when using the height model in Paper **I** or the calibrated height model by Veltheim (1987). Note the principal difference that the model for Näslund's parameters in Paper **V** is able to utilize various *stand characteristics* flexibly whereas the recent mixed-effect models from Mehtätalo (2005) and Eerikäinen (2009) can be localised efficiently by the *sample tree data*, if available. Regardless, Mehtätalo et al. (2007) noticed the incompatibility between the Korf-function-based height model (Mehtätalo 2005), calibrated with one sample tree, and the fitted Näslund's curve as a 'ground truth height'.

The models formulated utilized some stand characteristics that are not always collected in forest management planning field work. Stem number is typically optional for basal area in the forest inventory field work for private forest estates. However, still-current instructions for forest company field work include both as required characteristics (Kuvioittainen ... 1998). A similar situation obtains with dominant height. In Paper **V**, the main stand characteristics were modelled simultaneously. In this case, the models were presented for Scots pine, but similar models have been formulated for Norway spruce (Siipilehto 2006b) and birch species (unpublished) for the MOTTI simulator. In the absence of the characteristics necessary for the size distribution model used or the height–dbh relationship, the predictions are provided by these models; either the expected value with the minimum required input variables (tree species, stand age, location as a temperature sum, and site characteristics) or the information available on the stand characteristics is used for calibration of the issued unknown parameters. Linear prediction theory provided a convenient approach because the models could be formulated without being fixed to a particular existing inventory system. Indeed, in the case where all eight stand characteristics can be selected freely for calibration of unknown parameters, we have 255 different models for them described by the basic model and the error variance–covariance matrix.

It is well known that the stand variables collected in the field include considerable measurement error (e.g., Heikkinen 2002, Ojansuu et al. 2002, Haara and Korhonen 2004), the most inaccurate variables being stem number (RMSE of 50–80%) and basal area (20–40%). Recently, considerable improvement in the accuracy of stand characteristics has been found by means of laser scanning data – according to Suvanto et al. (2005) and Næsset (2002), RMSE in stem number and basal area was 18–35% and 8–21%, respectively. If one wants to emphasise differences in the accuracy between stand variables, the average measurement errors could be added to the diagonal variances (Paper **V**: Table 5), resulting in a diminishing calibration effect (see Equation 19). When Siipilehto (2006b) checked this option by using the corresponding models for spruce, the resultant, considerably less efficient calibration achieved sometimes irrelevant combinations of the predicted stand variables. The expectation value is, of course, sensitive to the error in the driving variable, stand age. However, the calibration was able to correct the predictions efficiently despite the errors that were generated in stand age (ibid.).

In this thesis, I focused mostly on Scots pine, which is the most common dominating species in Finland. Some of the models presented for pine may provide a relevant basis for prediction of the distribution for other species, too. I regard height distributions for pine-dominated young stands (see Paper **IV**) and SB distributions for drained Scots pine-dominated peatland stands (see Paper **III**) as such promising models. In addition, the latter could be validated for natural and semi-natural forests in Finland because it performed well with unthinned peatland stands. The family of the BLUP models (see Paper **V**) for Scots pine should be extended to cover Norway spruce and birch species. This would be especially useful because of the lack of diameter distribution models for young stands up to an H_{dom} of about 10 m and also for improved Näslund's height curves. Further comparisons between recently published models for stand characteristics, especially that for H_{dom} in juvenile pine stands by Huuskonen and Miina (2006) and the BLUP models (see Paper **V**), should be made. The effect of the accuracy of the predicted distributions on growth and yield prediction remains to be studied. Also, parameter recovery methods should be compared with parameter prediction models, in particular, when required percentile or moment based stand characteristics are known. Laser scanning is becoming current practice for large-scale operations; therefore, new kinds of approaches are needed for describing stand structure (e.g., models for stand characteristics and distributions based on laser pulses, k-NN methods, etc.). NFI field data collection has recently moved toward more detailed description of stands, so the models presented may become more useful with respect to studies and calculations based on NFI field data.

The results described in this thesis showed that:

- Selection of distribution function was important but not a crucial issue. A theoretically more flexible SB function was not as superior as could be expected in comparison with the Weibull function. This was simply because the parameters of the SB function were not as closely correlated with stand characteristics as were those of the Weibull function. Consequently, the variation in the estimated parameters of the Weibull distribution could be more accurately predicted/calibrated in some situations than that for SB distribution. The two-parameter Weibull distribution seemed an appropriate choice even though it was not compared with the three-parameter one in this thesis (see Maltamo et al. 1995). As a frequency distribution it can be recommended for dbh and height distribution in young stands. Choosing the best basal area-dbh distribution for advanced stands was a matter of the set of known input variables. With the common FMP data, as well as with

the additional information apart from stem number, Weibull and SB distributions were comparable. The additional knowledge of the stem number benefited the calibration and modelling of the SB distribution more than it did the Weibull distribution.

- Model formulation played the most important role in the two studies conducted, when one was constructing the models for the parameters of SB for basal area-dbh distribution in drained peatland stands and in modelling of Weibull for height distribution in sapling stands. In both cases, particular attention was paid to finding effective transformations in order to linearize the dependence and homogenise the variance between the modelled parameters and the combinations of stand characteristics as explaining variables. The final formulations can be considered successful.

- The choice of statistical method played a minor role in the ability of a model to predict stand structure (see also Liu et al. 2009). Two studies focused on the alternative methods for fitting the model. In both cases, each model provided reasonable accuracy but the model was improved to some extent by adjustment of the statistical method. On the one hand, the mixed-effect seemingly unrelated regression method for SB distribution and, on the other hand, the generalized linear model including the moment estimator for the Weibull distribution can be regarded as the most sophisticated methods applied for the present thesis, and they also showed superior performance. More obvious differences between statistical approaches have been found in the case where the parameter prediction model could not explain much of the variation in the dependent variables (e.g., Cao 2004).

- Inputting of data by means of the known stand characteristics played a slightly controversial role. The common FMP data were quite effective and evidently the minimum requirement for reliability in volume characteristics. Inclusion of additional input variables could improve the model's accuracy, but, unfortunately, improvement was clearly evident in the accuracy of the included variable only. Thus, including stem number as a calibrating variable for advanced stands improved the accuracy of the stem number generated but not, for example, that of the dominant tree characteristics, and, likewise, inclusion of dominant tree variables did not improve the accuracy of the stem number generated. Nevertheless, including the stem number (an important sum characteristic) resulted in improved accuracy also in the volume characteristics. On the other hand, the additional dominant height proved advantageous whereas the dominant diameter turned out to be detrimental in prediction of the structure of young stands by means of the BLUP application for the basal area-dbh distribution models. As a curiosity, including arithmetic mean dbh and basal-area-median dbh simultaneously showed an interesting ability to reduce bias and RMSE in most of the validated characteristics (see Paper **V**: Tables 8–11).

- The Näslund's height curve provided an excellent basis for characterising height–diameter relationship. Successful linearization required taking the logarithm of both sides, explained parameters and explaining stand characteristics. The Näslund's curve could be accurately predicted/calibrated with the common forest management planning input data (with D,H or d_{gM},h_{gM} giving the slope), and, in particular, with additional knowledge of the dominant tree characteristics, the bending of the curve was captured by the model. Thus, choosing Näslund's function provided simplicity in the fitting of the curve to sample tree data, in model formulation as well as in model application.

REFERENCES

Bailey, R.L. & Dell, T.R. 1973. Quantifying diameter distributions with the Weibull function. Forest Science 19(2): 97–104.

Biging, G.S. & Doppertin, M. 1992. A comparison of distance-dependent competition measures for height and basal area growth of individual conifer trees. Forest Science 38(3): 695–720.

Borders, B.E., Stouter, R.A., Bailey, R.L., & Ware, K.D. 1987. Percentile-based distributions characterize forest stand tables. Forest Science 33(2): 570–576.

Bowling, E., Burkhart, H.E., Burk, T.E., & Beck, D. 1989. A stand-level multispecies model for Appalachian hardwoods. Canadian Journal of Forest Research 19: 405–412.

Buongiorno, J., Dahir, S., Lu, H.C., & Lin, C.R. 1994. Tree size diversity and economic returns in uneven-aged forest stands. Forest Science 40: 83–103.

Burk, T.J. & Newberry, J.D. 1984. A simple algorithm for moment-based recovery of Weibull distribution parameters. Forest Science 30(2): 329–332.

Cajanus, W. 1914. Über die Entwicklung gleichaltriger Waldbestände: Eine Statistische Studie. Acta Forestalia Fennica 3. 142 p.

Cao, Q.V. 2004. Predicting parameters of a Weibull function for modelling diameter distribution. Forest Science 50(5): 682–685.

Curtis, R.O. 1967. Height–diameter and height–diameter–age equations for second-growth Douglas-fir. Forest Science 13(4): 265–357.

DemoMELA. Available from http://www.metla.fi/metinfo/mela/index-en.htm (visited on 16 Sep. 2008).

Droessler, T.D. & Burk, T.E. 1989. A test of nonparametric smoothing of diameter distributions. Scandinavian Journal of Forest Research 4(3): 407–415.

Eerikäinen, K. 2003. Predicting the height–diameter pattern of planted *Pinus kesiya* stands in Zambia and Zimbabwe. Forest Ecology and Management 175: 355–366.

Eerikäinen, K. 2009. A multivariate linear mixed-effects model for the generalization of sample tree heights and crown ratios in the Finnish National Forest Inventory. Forest Science 55(6): 480–493.

Eid, T. 2001. Models for prediction of basal area mean diameter and number of trees for forest stands in south-eastern Norway. Scandinavian Journal of Forest Research 16(5): 467–479.

Elfving, B. & Kiviste, A. 1997. Constructing of site index equations for *Pinus sylvestris* L. Using permanent plot data in Sweden. Forest Ecology and Management 98: 125–134.

Fonseca, T.F., Marques, C.P., & Parresol, B.R. Describing maritime pine diameter distribution with Johnson's SB distribution using a new all-parameter recovery approach. Forest Science 55(4): 367–373.

Franklin, J.F., Spies, T.A., Pelt, R.V., Carey, A.B., Thornburgh, D.A., Berg, D.R., Lindenmyer, D.B., Harmon, M.E., Keeton, W.S., Shaw, D.C., Bible; K., & Chen, J. 2002. Disturbances and structural development of natural forest ecosystems with silvicultural implications, using Douglas-fir forest as an example. Forest Ecology and Management 155: 399–423.

Gobakken, T. & Næsset, E. 2004. Estimation of diameter and basal area distributions in coniferous forest by means of airborne laser scanner data. Scandinavian Journal of Forest Research 19: 529–542.

Gobakken, T. & Næsset, E. 2005. Weibull and percentile models for lidar-based estimation of basal area distribution. Scandinavian Journal of Forest Research 20: 490–502.

Gove, J.H. 2003a. Estimation and applications of size-biased distributions in forestry. In: A. Amaro, D. Reed, & P. Soares (eds), Modelling Forest Systems, pp. 201–212.

Gove, J.H. 2003b. Moment and maximum likelihood estimators for Weibull distributions under length- and area-biased sampling. Environmental and Ecological Statistics 10: 455–467.

Gove, J.H. & Patil, G.P. 1998. Modelling the basal area–size distribution of forest stands: A compatible approach. Forest Science 44(2): 285–297.

Gustavsen, H.G. & Päivänen, J. 1986. Luonnontilaisten soiden puustot kasvullisesalla metsämaalla 1950-luvun alussa [Summary: Tree stands on virgin forested peatlands in the early 1950's in Finland]. Folia Forestalia 673: 1–27.

Gustavsen, H.G., Roiko-Jokela, P., & Varmola, M. 1988. Kivennäismaiden talousmetsien pysyvät (INKA ja TINKA) kokeet. Suunnitelmat, mittausmenetelmät ja aineistojen rakenteet. Metsäntutkimuslaitoksen tiedonantoja 292, 212 p.

Haara, A. & Korhonen, K.T. 2004. Kuvioittaisen arvioinnin luotettavuus. Metsätieteen aikakauskirja 4: 489–508.

Haara, A., Maltamo, M., & Tokola, T. 1997. The k-nearest-neighbour method for estimating basal–area diameter distribution. Scandinavian Journal of Forest Research 12: 200–208.

Hafley, W.L. & Schreuder, H.T. 1977. Statistical distributions for fitting diameter and height data in evenaged stands. Canadian Journal of Forest Research 7: 481–487.

Hafley, W.L. & Buford, M.A. 1985. A bivariate model for growth and yield prediction. Forest Science 31(1):237–247.

Heinonen, J. 1994. Koealojen puu- ja puustotunnusten laskentaohjelma KPL. Käyttöohje. Metsäntutkimuslaitoksen tiedonantoja (Finnish Forest Research Institute), Research Papers 504, 80 p.

Hökkä, H. & Laine, J. 1988. Suopuustojen rakenteen kehitys ojituksen jälkeen. [Summary: Post-drainage development of strctural characteristics in peatland forest stands]. Silva Fennica 22: 45–65.

Hökkä, H., Piiroinen, M.-L., & Penttilä, T. 1991. Läpimittajakauman ennustaminen Weibulljakaumalla Pohjois-Suomen mänty- ja koivuvaltaisissa ojitusaluemetsiköissä. Folia Forestalia 781, 22 p.

Hökkä, H., Kaunisto, S., Korhonen, K.T., Päivänen, J., Reinikainen, A. & Tomppo, E. 2002. Suomen suometsät 1951–94. Metsätieteen aikakauskirja 2B/2002: 201–357.

Hökkä, H. & Ojansuu, R. 2004. Height development of Scots pine on peatlands: Describing change in site productivity with a site index model. Canadian Journal of Forest Research 34: 1081–1092.

Holopainen, M. & Hyyppä, J. 2009. Kohden lasermittauksiin perustuvaa täsmämetsätaloutta. Metsätieteen aikakauskirja 4/2009.

Holopainen, M., Tuominen, S., Karjalainen, M., Hyyppä, J., Hyyppä, H., Vastaranta, M., Hujala, T. & Tokola, T. 2009. Korkearesoluutioisten E-SAR-tutkakuvien tarkkuus puustotunnusten koealatason estimoinnissa. Metsätieteen aikakauskirja 4/2009: 309–323

Holopainen, M., Mäkinen, A., Rasinmäki, J., Hyytiäinen, K., Bayazidi, S., & Pietilä, I. 2010. Comparison of various sources of uncertainty in stand-level net present value estimates. Forest Policy and Economics 12: 377–386.

Holte, A. 1993. Diameter distribution functions for even-aged (*Picea abies*) stands. Meddelelser fra Skogforsk 46(1): 1–47.

Huuskonen, S. & Miina, J. 2007. Stand-level growth models for young Scots pine stands in Finland. Forest Ecology and Management 241: 49–61.

Hyink, D.M. & Moser, J.W. 1983. A generalized framework for projecting forest yield and stand structure using diameter distributions. Forest Science 29: 85–95.

Hynynen, J. & Arola, M. 1999. Ensiharvennusajankohdan vaikutus hoidetun männikön kehitykseen ja harvennuksen kannattavuuteen. Metsätieteen aikakauskirja 1/1999:5–23.

Hynynen, J., Ojansuu, R., Hökkä, H., Siipilehto, J., Salminen, H. & Haapala, P. 2002. Models for predicting stand development in MELA system. Finnish Forest Institute, Research Papers 835, Vantaan Research Center. 116 s.

Hynynen, J., Ahtikoski, A., Siitonen, J., Sievänen, R., & Liski, J. 2005. Applying the MOTTI simulator to analyse the effect of alternative management schedules on timber and non-timber production. Forest Ecology and Management 207: 5−18.

Ilvessalo, Y. 1920. Tutkimuksia metsätyyppien taksatoorisesta merkityksestä nojautuen etupäässä kotimaiseen kasvutaulujen kehitystyöhön. Acta Forestalia Fennica 15(3): 1–157.

Järnstedt, J. 2010. Erittäin korkean resoluution ilmakuvista tuotettu pintamalli puustotunnusten estimoinnissa. Master's thesis. University of Helsinki. Department of Forest Sciences, 52 p.

Johnson, N.L. 1949a. Bivariate distributions based on simple translation systems. Biometrika 36: 297–304.

Johnson, N.L. 1949b. Systems of frequency curves generated by methods of translation. Biometrika 36: 149–176.

Kaila, E.E. 1931. Tervanpolton leviäminen Suomessa 1700-luvun puolimaissa. [Summary: Tar-burning in Finland in the middle of the 18th century]. Silva Fennica 21: 1–38.

Kalliovirta. J. 2006. SIMO simulointi. SIMO-hankkeen loppuseminaari 2.11.2007. University of Helsinki. Published online at http://www.simoproject.org/download/ seminars/2007/1102/SIMO_loppuseminaari.pdf.

Kamziah, A.K., Ahmad, M.I., & Lapongan, J. 1999. Nonlinear regression approach to estimating Johnson SB parameters for diameter data. Canadian Journal of Forest Research 29: 310–314.

Kamziah, A.K., Ahmad, M.I. & Ahmad Zuhaidi, Y. 2000. Modelling diameter distribution in even-aged and uneven-aged stands. Journal of Tropical Forest Science 12(4): 669–681.

Kangas, A. & Maltamo, M. 2000a. Calibrating predicted diameter distribution with additional information. Forest Science 46(3): 390–396.

Kangas, A. & Maltamo, M. 2000b. Percentile based basal area diameter distribution models for Scots pine, Norway spruce and birch species. Silva Fennica 34(4): 371–380.

Kangas, A. & Maltamo, M. 2002. Anticipating the variance of predicted stand volume and timber assortments with respect to stand characteristics and field measurements. Silva Fennica 36(4): 799–811.

Kangas, A. & Maltamo, M. 2003. Calibrating predicted diameter distribution with additional information in growth and yield predictions. Canadian Journal of Forest Research 33(3): 430–434.

Kangas, J., Heikkinen, E., & Maltamo, M. 2004. Accuracy of partially visually assessed stand characteristics: A case study of Finnish forest inventory by compartments. Canadian Journal of Forest Research 34(4): 916–930.

Kärki, T., Maltamo, M., & Eerikäinen, K. 2000. Diameter distribution, stem volume and stem quality model for grey alder (*Alnus incana*) in eastern Finland. New Forests 20: 65–86.

Karvonen, L. 2000. Guidelines for landscape ecological planning. Forestry publications of Metsähallitus 36, 46 p.

Kilkki, P. 1983. Sample trees in timber volume estimation. Acta Forestalia Fennica 1982, 35 p.

Kilkki, P. & Siitonen, M. 1975. Metsikön puuston simulointimenetelmä ja simuloituun aineistoon perustuvien puustotunnusmallien laskenta. [Summary: Simulation of artificial stands and derivation of growing stock models from this material]. Acta Forestalia Fennica 145. 33 p.

Kilkki, P. & Päivinen, R. 1986. Weibull-function in the estimation of basal area DBH-distribution. Silva Fennica 20(2): 149–156.

Kilkki, P., Maltamo, M., Mykkänen, R., & Päivinen, R. 1989. Use of the Weibull function in estimating the basal area DBH-distribution. Silva Fennica 23(4): 311–318.

Kiviste, A., Nilson, A., Hordo, M., & Merenäkk, M. 2003. Diameter distribution models and height–diameter equations for Estonian forests. In: A. Amaro, D. Reed, & P. Soares (eds), Modelling Forest Systems. CAB International, ISBN 0–85199–693–0.

Knoebel, B.R. & Burkhart, H.E. 1991. A bivariate distribution approach to modelling forest diameter distributions at two points in time. Biometrics 47: 241–253.

Knowe, S.A., Harrington, T.B., & Shula, R.G., 1992. Incorporating the effects of interspecific competition and vegetation management treatments in diameter distribution models for Douglas-fir saplings. Canadian Journal of Forest Research 22: 1255–1262.

Koivuniemi, J. 2003. Metsiköihin ja paikannettuihin koealoihin perustuvan kuvioittaisen arvioinnin tarkkuus [Summary: The accuracy of the compartmentwise forest inventory based on stands and located sample plots]. University of Helsinki, Department of Forest Resource Management Publications 36: 1–160.

Korpela, I. & Tokola, T. 2006. Potential of aerial image-based monoscopic and multiview single-tree forest inventory: A simulation approach. Forest Science 52(2): 136–147.

Koskimaa, E. 1985. MERA-kausi vakiinnutti rahoitusta. Metsälehti, p. 16.

Kuuluvainen, T. 2002. Natural variability of forests as a reference for restoring and managing biological diversity in boreal Fennoscandia. Silva Fennica 36(1): 97–125.

Kuuluvainen, T., Järvinen, E., Hokkanen, T.J., Rouvinen, S., & Heikkinen, K. 1998a. Structural heterogeneity and spatial autocorrelation in natural mature *Pinus sylvestris* dominated forest. Ecography 21(2): 159–174.

Kuuluvainen, T., Syrjänen, K., & Kalliola, R. 1998b. Structure of a pristine *Picea abies* forest in northeastern Europe. Journal of Vegetation Science 9: 563–574.

Kuusela, K. 1990. The dynamics of boreal coniferous forests. SITRA 112. Helsinki. ISBN 951563274–9.

Kuvioittainen arviointi ja tietojen päivitys 1998. UPM-Kymmene Metsä, 31 p.

Laamanen, R., Leskinen, J., & Hallman, E. 1997. PATI-maastotyöohje. Helsinki: Metsähallitus, 58 p.

Lähde, E., Laiho, O., Norokorpi, Y., & Saksa, T. 1994a. Structure and yield of all-sized Scots pine-dominated stands. Annales des Sciences Forestiéres 51: 111–120.

Laiho, O., Lähde, E., Norokorpi, Y., & Saksa, T. 1994b. Varttuneiden metsiköiden rakenne 1950–luvun alussa [Summary: Stand structure of advanced forests in the early 1950s in Finland]. Metsäntutkimuslaitoksen tiedontoja 495: 90–128.

Laird, N.M. & Ware, J.H. 1982. Random-effects models for longitudinal data. Biometrics 38: 963–974.

Lappi, J. 1993. Metsäbiometrian menetelmiä. Silva Carelica 24. University of Joensuu, 182 p.

Lappi, J. 1997. A longitudinal analysis of height/diameter curves. Forest Science 43(4): 555–570.

Lappi, J., Mehtätalo, L., & Korhonen, K.T. 2006. Generalizing sample tree information. In: A. Kangas & M. Maltamo (eds), Forest Inventory: Methodology and Applications. Managing Forest Ecosystems 10, pp. 85–105.

Leduc, D. & Goelz, J. 2009. A height–diameter curve for longleaf pine plantations in the Gulf Coastal Plain. Southern Journal of Applied Forestry 33(4): 164–170.

Lehtonen, H. 1998. Fire history recorded on pine trunks and stumps: Influence of land use and fires on forest structure in Northern Karelia. Scandinavian Journal of Forest Research 13(4): 462–468.

Li, F., Zhang, L., & Davis, C.J. 2002. Modeling the joint distribution of tree diameters and heights by bivariate generalized beta distribution. Forest Science 48(1): 47–58.

Linder, P. 1998. Structural changes in two virgin boreal forest stands in central Sweden over 72 years. Scandinavian Journal of Forest Research 13(4): 451–461.

Linder, P., Elfving, B. & Zackrisson, O. 1997. Stand structure and successional trends in virgin boreal forest reserves in Sweden. Forest Ecology and Management 98: 17−33.

Lindsay, S.R., Wood, G.R. & Woollons, R.C. 1996. Stand table modelling through the Weibull distribution and usage of skewness information. Forest Ecology and Management 81: 19−23.

Liu, C., Zhang, S.Y., Lei, Y., Newton, P.F. & Zhang, L. 2004. Evaluation of three methods for predicting diameter distributions of black spruce (*Picea mariana*) plantations in central Canada. Canadian Journal of Forest Research 34: 2424–2432.

Liu, C., Bealieu, J., Pregent, G. & Zhang, S.Y. 2009. Applications and comparison of six methods for predicting parameters of the Weibull function in unthinned *Picea glauca* plantations. Scandinavian Journal of Forest Research 24:67–75.

Loetsch, F., Zöhrer, F., & Haller, K. 1973. Forest inventory, Vol. II. BLV Verlagsgesellschaft München. ISBN 3–405–10812–8.

Mabvurira, D., Maltamo, M., & Kangas, A. 2002. Predicting and calibrating diameter distributions of *Eucalyptus grandis* (Hill) Maiden plantations in Zimbabwe. New Forests 23: 207–223.

Mack, T.J. & Burk, T. E. 2004. Equations for predicting merchantable yield and diameter distribution for Lake States red pine. Northern Journal of Applied Forestry 21(2): 107–110.

Maltamo, M. 1997. Comparing basal area diameter distributions estimated by tree species and for the entire stock in a mixed stand. Silva Fennica 31(1): 53–65.

Maltamo, M. & Kangas, A. 1998. Methods based on k-nearest neighbor regression in the prediction of basal area diameter distribution. Canadian Journal of Forest Research 28(8): 1107–1115.

Maltamo, M. & Uuttera, J. 1998. Angle-count sampling in the description of forest structure. Forest & Landscape Research. Paino, Finland.

Maltamo, M., Puumalainen, J., & Päivinen, R. 1995. Comparison of beta and Weibull functions for modelling basal area diameter distribution in stands of *Pinus sylvestris* and *Picea abies*. Scandinavian Journal of Forest Research 10: 284–295.

Maltamo, M., Uuttera, J., & Kuusela, K. 1997. Differences in forest stand structures between forest ownership groups in Central Finland. Journal of Environmental Management 51: 145–167.

Maltamo, M., Kangas, A., Uuttera, J., Torniainen, T., & Saramäki, J. 2000. Comparison of percentile based prediction methods and the Weibull distribution in describing the diameter distribution of heterogeneous Scots pine stands. Forest Ecology and Management 133: 263–274.

Maltamo, M., Haara, A., Hirvelä, H., Kangas, A., Lempinen, R., Malinen, J., Nalli, A., Nuutinen, T., & Siipilehto, J. 2002. Läpimittajakaumamalleihin perustuvat vaihtoehdot kuvauspuiden muodostamiseen puuston keskitunnustietojen avulla. Metsätieteen aikakauskirja 3/2002: 407–423.

Maltamo, M., Eerikäinen, K., Pitkänen J., Hyyppä, J., and Vehmas, M. 2004. Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. Remote Sensing of Environment 90: 319–330.

Maltamo, M., Suvanto, A., & Packalén, P. 2007. Comparison of basal area and stem frequency diameter distribution modelling using airborne laser scanner data and calibration estimation. Forest Ecology and Management 247: 26–34.

McTague, J.P. & Bailey, R.L. 1987. Compatible basal area and diameter distribution models for thinned loblolly pine plantations in Santa Catarina, Brazil. Forest Science 33(1): 43–51.

Mehtätalo, L. 2004. A longitudinal height-diameter model for Norway spruce in Finland. Canadian Journal of Forest Research 34: 131–140.

Mehtätalo, L. 2005. Height-Diameter (H-D) models for Scots pine and birch in Finland. Silva Fennica 39(1): 55–66.

Mehtätalo, L., Comas, C., Pukkala, T., & Palahi, M. 2011. Combining a predicted diameter distribution with an estimate based on a small sample of diameters. Canadian Journal of Forest Research 41: 750–762.

Merganič, J. & Sterba, H. 2006. Characterisation of diameter distribution using the Weibull function: Method of moments. European Journal of Forest Research 125: 427–439.

Mielikäinen, K. 1980. Mänty-koivusekametsikön rakenne ja kehitys. [Summary: Structure and development of mixed pine and birch stands]. Communicationes Instituti Forestalis Fenniae 99(3):1–82.

Mielikäinen, K. 1985. Koivusekoituksen vaikutus kuusikon rakenteeseen ja kehitykseen. [Summary: Effect of an admixture of birch on the structure and development of Norway spruce stands]. Communicationes Instituti Forestalis Fenniae 133:1–73.

Mønness, E.N. 1982. Diameter distributions and height curves in even-aged stands of *Pinus sylvestris* L. Meddelelser fra Norsk Institutt for Skogforskning 36(15): 1–43.

MONSU. Forest planning program. Available via http://www.monsu.net/ (visited on 16 Apr. 2010).

MOTTI software. Available from http://www.metla.fi/metinfo/motti/index-en.htm (visited on 16 Sep. 2008).

Mykkänen, R. 1986. Weibull-funktion käyttö puuston läpimittajakauman estimoinnissa. Master's thesis. University of Joensuu, 80 p.

Näslund, M. 1936. Skogsförsöksanstaltens gallringsförsök i tallskog. Meddelanden från Statens Skogsförsöksanstalt 28(1).

Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sensing of Environment 80: 88–99.

Næsset 2007. Airborne laser scanning as a method in operational forest inventory: Status of accuracy assessments accomplished in Scandinavia. Scandinavian Journal of Forest Research 22(5): 433–442.

Næsset, E., Gobakken, T., Holmgren, J., Hyyppä, H., Hyyppä, J., Maltamo, M., Nilsson, M., Olsson, H., Persson, A., & Söderman, U. 2004. Laser scanning of forest resources: The Nordic experience. Scandinavian Journal of Forest Research 19(6): 482–499.

Newton, P.F. & Amponsah, I.G. 2007. Comparative evaluation of five height–diameter models developed for black spruce and jack pine stand-types in terms of goodness-of-fit, lack-of-fit and predictive ability. Forest Ecology and Management 247(1–3): 149–166.

Newton, P.F., Lei, Y., & Zhang, S.Y. 2005. Stand-level diameter distribution yield model for black spruce plantation. Forest Ecology and Management 209: 181–192.

Niemistö, P. 1994. Männikön ensiharvennus ala-, ylä- tai laatuharvennusta käyttäen. Folia Forestalia - Metsätieteen aikakauskirja 1994(1): 19–32.

Nishizono, T., Iehara, T., Kuboyama, H., & Fukuda, M. 2005. A forest biomass yield table based on an empirical model. Journal of Forest Research 10: 211–220.

Nissinen, J. 2002. Improving compatibility between prediction of basal area diameter distributions and assesments of young stands. Master's thesis. University of Joensuu. 48 p.

Nuutinen, T. 1986. Puustotunnusten väliset relaatiot. Master's thesis. University of Joensuu.

Ojansuu, R., Halinen, M., & Härkönen, K. 2002. Metsätalouden suunnittelujärjestelmän virhelähteet männyn ensiharvennuskypsyyden määrittelyssä. Metsätieteen aikakauskirja 3/2002: 441–457.

Packalén, P. & Maltamo, M. 2007. The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. Remote Sensing of Environment 109(3): 328–341.

Packalén, P. & Maltamo, M. 2008. Estimation of species-specific diameter distributions using airborne laser scanning and aerial photographs. Canadian Journal of Forest Research 38: 1750–1760.

Päivinen, R. 1980. Puiden läpimittajakauman estimointi ja siihen perustuva puustotunnusten laskentamenetelmä. [Summary: On the estimation of stem diameter distribution and stand charactaristics]. Folia Forestalia 442, 28 p.

Parresol, B.R. 1992. Baldcypress height–diameter equations and their prediction confidence intervals. Canadian Journal of Forest Research 22: 1429–1434.

Parresol, B.R. 2003. Recovering parameters of Johnson's SB distribution. USDA Forest Service. Southern Research Station. Research Paper SRS-31, 9 p.

Peuhkurinen, J., Maltamo, M., & Malinen, J. 2007. Preharvest measurement of marked stands using airborne laser scanning. Forest Science 53(6): 653–661.

Peuhkurinen, J., Maltamo, M., & Malinen J. 2008. Estimating species-specific diameter distributions and saw log recoveries of boreal forest from airborne laser scanning data and aerial photographs: A distribution based approach. Silva Fennica 42(4): 625–641.

Pommerening, A. 2002. Approaches to quantifying forest structures. Forestry 75(3): 305–324.

Pukkala, T. 2004. Dealing with ecological objectives in the Monsu planning system. Silva Lusitana 12 Especial: 1–15.

Pukkala, T. & Miina, J. 2005. Optimizing the management of a heterogeneous stand. Silva Fennica 39(4): 525–538.

Rennols, K., Geary, D.N., & Rollinson, T.J.D. 1985. Characterising diameter distributions by the use of the Weibull distribution. Forestry 58(1): 57–66.

Reynolds, M.R., Burk, T.E., & Huang, W.C. 1988. Goodness-of-fit tests and model selection procedures for diameter distribution models. Forest Science 34: 373–399.

Robinson, A. 2004. Preserving correlation while modelling diameter distributions. Canadian Journal of Forest Research 34: 221–232.

Rouvinen, S. & Kuuluvainen, T. 2005. Tree diameter distributions in natural and managed old *Pinus sylvestris*-dominated forests. Forest Ecology and Management 208(1–3): 45–61.

Sarkkola, S. 2006. Stand structural dynamics on pristine and managed boreal peatlands. Dissertationes Forestales 29, 78 p.

Sarkkola, S., Alenius, V., Hökkä, H., Laiho, R., Penttilä, T., & Päivänen, J., 2003. Changes in structural inequality in Norway spruce stands after water-level drawdown. Canadian Journal of Forest Research 33: 222–231.

Sarkkola, S., Hökkä, H., Laiho, R., Päivänen, J. & Penttilä, T. 2005. Stand structural dynamics on drained peatlands dominated by Scots pine. Forest ecology and Management 206: 135–152.

Schreuder, H.T. & Hafley, W.L. 1977. A useful bivariate distribution for describing stand structure of tree heights and diameters. Biometrics 33: 471–478.

Scolforo, J.R.S., Tabai, F.C.V., de Madeco, R.L.G., Acerbi, F.W.J., & de Assis, A.L. 2003. Sb distribution's accuracy to represent the diameter distribution of *Pinus taeda*, through five fitting methods. Forest Ecology and Management 175: 489–496.

Shiver, B.D. 1988. Sample sizes and estimation methods for the Weibull distribution for unthinned slash pine plantation diameter distributions. Forest Science 34(3): 809–814.

Siipilehto, J. 1988. Metsätalouskuvioiden läpimittajakaumien ennustaminen beta-funktiolla. Master's thesis. University of Helsinki. 54 p.

Siipilehto, J. 1996. Metsikön läpimitta- ja pituusjakauman kuvaaminen kaksiulotteisen todennäköisyysfunktion avulla. Licentiate's thesis. University of Helsinki.

Siipilehto, J. 2001a. Improving the accuracy of predicted diameter and height distributions. In. V. LeMay and P. Marshall (eds.). Forest modelling for ecosystem management, forest certification, and sustainable management. Proceedings of the conference held in Vancouver, BC, Canada August 12 to 17, 2001, p 385–391

Siipilehto, J. 2001b. Puuston läpimittajakaumien erot luononmetsien ja varttuneiden talous-metsien välillä. Julkaisussa: Siitonen, J. (toim.). Monimuotoinen metsä. Metsäluonnon monimuotoisuuden tutkimusohjelman loppuraportti. Metsäntutkimuslaitoksen tiedon-antoja 812: 11–23.

Siipilehto, J. 2006a. Height distributions of Scots pine sapling stands affected by retained tree and edge stand competition. Silva Fennica 40(3): 473–486.

Siipilehto, J. 2006b. Linear prediction application for modelling the relationships between a large number of stand characteristics of Norway spruce stands. Silva Fennica 40(3): 517–530.

Siipilehto, J. & Heikkilä, R. 2005. The effect of moose browsing on the height structure of Scots pine saplings in a mixed stand. Forest Ecology and Management 205: 117–126.

Siipilehto, J. & Siitonen, J. 2004. Degree of previous cutting in explaining the differences in diameter distributions between mature managed and natural forests. Silva Fennica 38(4): 425–435.

Siitonen, J., Martikainen, P., Punttila, P., & Rauh, J. 2000. Coarse woody debris and stand characteristics in mature managed and old-growth boreal mesic forests in southern Finland. Forest Ecology and Management 128: 211–225.

Siitonen, M., Härkönen, K., Hirvelä, H., Jämsä, J., Kilpeläinen, H., Salminen, O., & Teuri, M. 1996. MELA Handbook, 1996 Edition. The Finnish Forest Research Institute, Research Paper 622, 452 p.

Smith, B.C. 2007. Modeling the diameter distribution in juvenile loblolly pine (*Pinus taeda* L.) from diverse genetic provenances under deficient and optimum nutrition regimes. North Carolina State University, Faculty of Forestry. Master's thesis available on the Internet: http://www.lib.ncsu.edu/theses/available/etd-11092007-113109/unrestricted/etd.pdf.

Solmun maastotyöopas. 1997. Metsätalouden kehittämiskeskus Tapio. Hakapaino Oy, 82 p.

Staudhammer, C.L. & LeMay, M. 2001. Introduction and evaluation of possible indices of stand structural diversity. Canadian Journal of Forest Research 31: 1105–1115.

Stoyan, D. & Penttinen, A. 2000. Recent applications of point process method in forestry statistics. Statistical Science 15: 61–78.

Suvanto, A., Maltamo, M., Packalén, P., & Kangas, J. 2005. Kuviokohtaisten puustotunnusten ennustaminen laserkeilauksella. Metsätieteen aikakauskirja 4/2005: 413–428.

Tadikamalla, P.P. & Johnson, N.L. 1982. Systems of frequency curves generated by transformations of logistic variables. Biometrika 69: 461–465.

Tapio. 2009. Laserinventoinnin kehitys etenee. Tapion kehittämispalvelut. Newsletter available on the Internet: http://www.tapio.fi/uutiskirje_uutinen?pid=15107755.

Tewari, V.P. & Gadow, K.V. 1999. Modelling the relationship between tree diameters and heights using SBB distribution. Forest Ecology and Management 119: 171–176.

Tham, Å. 1988. Structure of mixed *Picea abies* (L.) Karst. and *Betula pendula* Roth and *Betula pubescens* Ehrh. stands in South and Middle Sweden. Scandinavian Journal of Forest Research 3: 355–369.

Tikkanen, J., Hokajärvi, R., & Hujala, T. 2010. Development phases of forest planning on non-industrial private lands in Finland: Perspective of planners' work. Small-scale Forestry 9: 331–347.

Tokola, T., Kangas; A., Kalliovirta, J., Mäkinen, A., & Rasinmäki, J. 2006. SIMO - SIMulointi ja Optimointi uuteen metsäsuunnitteluun. Metsätieteen aikakauskirja 1: 60–65.

Tomppo, E., Olsson, H., Ståhl, G., Nilsson, M., Hagner, O. & Katila, M. 2008. Combining national forest inventory field plots and remote sensing data for forest databases. Remote Sensing of Environment 112: 1982–1999.

Uuttera, J., Maltamo. M., & Kuusela, K. 1996. Impact of forest management on the state of forests in relation to natural forest succession: Comparative study, North Karelia, Finland vs. Republic of Karelia. Forest Ecology and Management 83(1–2): 71–85.

Uuttera, J., Maltamo. M., & Hotanen, J.P. 1997.The structure of forest stands in virgin and managed peatlands: a comparison between Finnish and Russian Karelia. Forest Ecology and management 96(1–2): 125–138.

Uuttera, J., Anttila, P., Suvanto, A., & Maltamo. M. 2006. Yksityismetsien metsävaratiedon keruuseen soveltuvilla kaukokartoitusmenetelmillä estimoitujen puustotunnusten luotettavuus. Metsätieteen aikakauskirja 4/2006: 507–519.

Valkonen, S. 1997. Viljelykuusikon alkukehityksen malli. Metsätieteen aikakauskirja 3: 321–347.

Valkonen, S., Ruuska, J. & Siipilehto, J. 2002. Effect of retained trees on the development of young Scots pine stands in Southern Finland. Forest Ecology and Management 166(1–3): 227–243.

Valtakunnan metsien 11. inventointi (VMI11). Maastotyön ohjeet 2009. Available from http://www.metla.fi/ohjelma/vmi/vmi11-maasto-ohje09-2p.pdf (visited on 4 Oct. 2010).

Veltheim, T. 1987. Pituusmallit männylle, kuuselle ja koivulle. University of Helsinki, 60 p.

Wang, M. & Rennolls, K. 2005. Tree diameter distribution modelling: Introducing the logit-logistic distribution. Canadian Journal of Forest Research 35: 1305–1313.

Wang, M. & Rennolls, K. 2007. Bivariate distribution modelling with tree diameter and height data. Forest Science 53(1): 16–24.

Weng, Y., Park, Y., Simpson, D., Tosh, K., & Fullarton, M. 2010. Tree improvement effects on tree size distributions for *Picea glauca* and *Picea mariana* in New Brunswick, Canada. Scandinavian Journal of Forest Research 25: 10–20.

Westfall, J.A., Burkhart, H.E., & Allen, H.L., 2004. Young stand growth modeling for intensively-managed loblolly pine plantation in southern U.S. Forest Science 50(6): 823–835.

Zellner, A. 1962. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. Journal of the American Statistical Association 57: 348–368.

Zhang, L., Packard, K.C., & Liu, C. 2003. A comparison of estimation methods for fitting Weibull and Johnson's Sb distribution to mixed spruce-fir stands in northeastern North America. Canadian Journal of Forest Research 33: 1340–1347.

Zhang, S., Amateis, R.L., & Burkhart, H.E. 1997.Constraining individual tree diameter increment and survival models for loblolly pine plantations. Forest Science 43(3): 414–423.

Zhou, B.L. & McTague, J.P. 1996. Comparison and evaluation of five methods of estimation of the Johnson system parameters. Canadian Journal of Forest Research 26: 928–935.

Zucchini, W., Schmidt, M., & von Gadow, K. 2001. A model for the diameter–height distribution in an uneven-aged beech forest and a method to assess the fit of such models. Silva Fennica 35(2): 169–183