Local Prediction of Stand Structure Using Linear Prediction Theory in Scots Pine-Dominated Stands in Finland

Jouni Siipilehto

This study produced a family of models for eight standard stand characteristics, frequency and basal area-based diameter distributions, and a height curve for stands in Finland dominated by Scots pine (*Pinus sylvestris* L.). The data consisted of 752 National Forest Inventory-based sample plots, measured three times between 1976 and 2001. Of the data, 75% were randomly selected for modelling and 25% left out for model evaluation. Base prediction models were constructed as functions of stand age, location and site providing strongly average expectations. These expectations were then calibrated with the known stand variables using linear prediction theory when estimating the best linear unbiased predictor (BLUP). Three stand variables, typically assessed in Finnish forest management planning fieldwork, were quite effective for calibrating the expectation for the unknown variable. In the case of optional distributions, it was essential to choose the weighting of the diameter distribution model such that the available input variables and the model applied were based on the same scale (e.g. arithmetic stand variables for frequency distribution). Additional input variables generally improved the accuracy of the validated characteristics, but the improvements in the predicted distributions were most noteworthy when the arithmetic mean and basal area-weighted median were simultaneously included in the BLUP estimation. The BLUP method provided a flexible approach for characterising relationships among stand variables, alternative size distributions and the height–diameter curve. Models are intended for practical use in the MOTTI simulator.

1 Introduction

In Finland, the description of a stand is commonly simplified to visually assessed mean and sum characteristics in forest management planning (FMP) fieldwork. Stand characteristics are assessed by tree species (see Solmun… 1997). The number of stems is typically assessed in young stands, and the basal area in advanced stands. However, the stem number is sometimes required additionally for basal area and basal area-weighted variables (Kuvioittainen… 1998). Alternative choices related to the stand characteristics assessed may cause problems in the use of FMP inventory data as input variables in simulation systems. 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 while the SIMO (Kalliovirta 2006) simulator incorporates both stand-level and tree-level simulation options. In any case, diameter distributions are needed for the calculation of assortment volumes (Kalliovirta 2006). Thus, tree size distribution models are needed for converting stand-level information into tree-level information.

Most of the diameter distribution models in Finland are based on basal area and basal area-weighted medians (e.g. Päivinen 1980, Mykkänen 1986, Maltamo 1997, Kilkki et al. 1989). More recently developed Johnson's SB and percentile-based distribution models attempt to enhance the accuracy in predicted distributions by using the stem number as an additional independent variable (Siipilehto 1999, Kangas and Maltamo 2000a, Siipilehto et al. 2007). Stem frequency-based Weibull models have been presented for drained peatlands by Sarkkola et al. (2003, 2005). Models for young stands based on Weibull height distribution (Valkonen 1997, Siipilehto 2009) and the models for tree height in Finland (Veltheim 1987, Siipilehto 1999, Mehtätalo 2004, 2005, Eerikäinen 2009) require different sets of input variables. Therefore, we may still have an incomplete set of input variables for some present models and/or unsuitable models for the data in use.

Thus, the need for modelling individual stand characteristics or relationships between stand characteristics arises from the difference between known and needed input variables for applied models, but also because of the changes and alternatives in FMP or National Forest Inventory practices (e.g. Nissinen 2002, Nuutinen 1986, Eid 2001). At present, the utilisation of satellite images and laser scanning data may create pressure for changes – for example, by including the number of stems in the variables assessed, because of the increased accuracy (Packalén and Maltamo 2007, Maltamo et al. 2007, Vohland et al. 2007) in comparison with traditional FMP fieldwork (e.g. Haara and Korhonen 2004, Kangas et al. 2004).

To overcome the problem of missing input variables, Siipilehto (2006) presented a linear prediction application for eight stand characteristics of Norway spruce stands. Basic models for spruce were function of stand age, location, site and origin. The expectations for unknown stand characteristics were then calibrated with the known stand variables using linear prediction theory (see Lappi 1993, 2006). For example, the original RMSE of 34% for basal area was reduced to 18% when the mean diameter and stem number were used for calibration (Siipilehto 2006). This work is a direct follow-up and extension to that study. Extension consisted of simultaneous prediction models for three optional diameter distributions and a height curve in order to provide all the necessary tools for stand structure in the same package.

The purpose of this study is to develop a family of models providing tools to predict the structure of Scots pine stands in Finland. Simultaneously to stand characteristics, models for unknown parameters of the diameter distributions and height–diameter relationship are estimated. The models are intended to provide 1) availability of prediction with a limited number of known characteristics as a basic model; 2) calibration of the prediction with an arbitrary set of known stand variables based on linear prediction theory; 3) a theoretically consistent framework for comparing alternative size distribution models; and 4) a broad range of applicability ranging from sapling stands to mature stands. Tree diameter is defined as diameter-at-breast-height (dbh) and its measurement is assumed as being error-free. In this study, we focused on the extension to Siipilehto (2006), i.e. a prediction of stand structure in terms of dbh distribution and the height-dbh curve.

The distributions studied are Weibull for a dbh-frequency distribution (W_N) and basal area-dbh distribution (W_G) and Johnson's SB distribution for basal area-dbh distribution (SB_G). Height was modelled with a function by Näslund (1936). Models are intended for practical use in a MOTTI simulator.

2   Material and Methods

2.1   Modelling and Test Data

The data was obtained from permanent sample plots, established in seedling stands between 1984–1989 (TINKA: $H_{dom} < 5$ m when established) and more advanced stands (INKA: $H_{dom} \geq 5$ m) between 1976–1983. The stands are based on a subsample of the Seventh National Forest Inventory. Each standwise sample plot consisted of a cluster of three circular plots within a stand. The radius was selected such that the total number of trees tallied was about 100–120 per stand. A smaller radius has been applied within each circular plot to select one-third of the sampled area for height and other more detailed measurements (Gustavsen et al. 1988). For the purpose of the present study, the cluster of three larger-radius sample plots were combined for reliable stand characteristics and especially for reliable distributions (see Shiver, 1988) representing the whole stand. Expected heights were given from Näslund's height curve for trees with missing tree height in INKA data set, whereas in sapling stand TINKA data, all crop trees have been measured for height and dbh. In the INKA data set, the smallest trees (dbh < 5 cm) were not measured if they were considered unsuitable for growing (e.g. having inadequate growing space).

The number of Scots pine-dominated INKA and TINKA stands was 562 and 190 respectively. In this study, data was restricted to a mean diameter (D) greater than 1.5 cm. This was done in order to avoid the anomalies in stand characteristics resulting from the low proportion of trees above breast height in the youngest state of stands. The remeasurements were carried out twice between 1981–2001, typically five and ten years after establishment. Because of the repeated

Table 1. Stand variables and estimated parameters [a] for the height curve and dbh distribution for Scots pine stands. P denotes the proportion of Scots pine out of the stand total basal area.

Model data (n = 1391)	Mean	Std	Min	Max
DDY (°C)	1022	162.6	674	1361
T (years)	59.9	34.5	10	188
G (m²ha⁻¹)	12.6	6.7	0.4	37.6
P	0.89	0.14	0.29	1.00
N (ha⁻¹)	1259	750	88	4626
D (cm)	11.7	6.0	1.7	34.4
H (m)	9.6	5.0	1.6	26.8
d_{gM} (cm)	14.8	6.3	3.0	37.0
h_{gM} (m)	11.1	5.1	2.5	28.1
D_{dom} (cm)	19.4	7.0	4.2	41.4
H_{dom} (m)	12.4	5.1	2.5	28.6
β_0	1.225	0.367	0.430	2.595
β_1	0.257	0.061	0.141	0.574
B	13.0	6.2	1.9	36.9
C	3.2	1.3	1.1	10.5
b_G	16.1	6.6	3.4	38.8
c_G	4.3	1.3	1.7	10.7
λ	28.4	11.9	5.0	81.1
δ	1.6	0.5	0.5	4.0
γ	–0.09	0.84	–2.35	2.31

[a] β_0, β_1, parameters of the Näslund's height curve; b, c, parameters of the Weibull function for frequency distribution; b_G, c_G, those for Weibull basal area-dbh distribution; λ, δ, γ, parameters of Johnson's SB function for basal area-dbh distribution.

measurements, the total number of observations in the study was 1858. Data was randomly divided into a modelling data set (75%) for estimation of the models and a validation data set (25%) for evaluating the estimated models. The final number of observations was 1392 for modelling (see Table 1) and 467 for validation (see Table 2). The modelling data covered dominant heights of 2.5 to 28.6 m (as seen in Table 1).

In current forest management planning field-work, the typical set of variables characterising stand structure (i.e. dbh and height distributions) is either 1) the number of trees per hectare, the mean diameter, and the mean height for young stands; or 2) the basal area per hectare, the basal area median dbh, and the corresponding height for advanced stands (see Solmun… 1997, Kuvioit-tainen… 1998). Hence, the randomly selected test data was divided into young and advanced stands due to the assumption of different combi-

Table 2. The mean and range of the validated stand characteristics for young (H<9 m) and advanced (H≥9 m) stands from validation data.

	G	N	D_{dom}	H_{dom}	Vtot	Logs	Pulp	Waste
Young stands	n=248							
Mean	8.9	1590	14.7	8.7	40.1	-	32.4	7.6
Min	0.4	251	4.4	2.8	0.9	-	0.0	0.9
Max	24.1	4552	30.7	17.1	117.3	-	101.0	29.1
Advanced stands	n=219							
Mean	17.4	872	25.3	17.0	134.4	78.7	51.7	4.0
Min	2.4	122	15.2	11.0	14.8	0.0	6.3	0.4
Max	34.4	2988	36.4	26.8	341.6	249.8	128.0	18.3

nations of the known stand characteristics. In this study, the threshold was set to a mean height of 9 m, resulting in the subsamples of almost same size (see Table 2). Test data covered dominant heights of 2.8 to 26.8 m. Note that calculated stand characteristics are accurate compared with the visually assessed characteristics of FMP due to assumed error-free tree-level measurements. In order to avoid overstating the accuracy in height and volume characteristics in model validation, random error was added to the expected height for trees with missing height measurement in the test data set.

2.2 Overview of the Modelling Approach

This study consists of different modelling phases before the final prediction for stand structure. The basic idea was to include all the necessary models for stand structure in simultaneous estimations. This means including all the candidate stand characteristics for potential calibration in addition to the unknown parameters of the selected dbh distributions and height-dbh functions. A brief introduction is given here in order to summarise the methods and modelling steps:

1) Dbh distribution and height-dbh models are required for the tree-level information of stand structure. Selected functions and fitting methods for the parameters are given in section 2.3.

2) The basic prediction models for the obtained parameters are constructed simultaneously with the models for stand characteristics in section 2.4. The estimated parameters of the basic models are given in the results, Tables 3 and 4. These basic models are intended to provide logical but strongly averaging age-dependent predictions for different sites and locations of stands.

3) In practice, the expectations and error variances for unknown parameters and unknown stand characteristics are calibrated with the set of known stand characteristics in order to get the final, improved prediction for the stand structure. This method is called linear prediction and it is presented in section 2.5. The idea is that the predictors (calibrating variables) are not fixed beforehand; instead, arbitrary set of included stand characteristics can be used for calibration. Linear prediction application utilises a cross-model error variance-covariance matrix for calibration (i.e. estimating the best linear unbiased predictor). The estimated matrix (along with the correlation coefficients) is given in two parts in the results, in Tables 5 and 6.

4) Given examples and the final model evaluation (Tables 7–10) looks for the efficient stand characteristics for calibrating the unknown parameters of the included dbh distributions and height curve.

2.3 Modelling of Tree Dimensions

The diameter distribution was characterised by Weibull and Johnson's SB distribution functions. The two-parameter Weibull function for tree diameter d has the following probability density function (PDF) (e.g. Bailey and Dell 1973):

$$f(d) = c/b(d/b)^{c-1} \exp\left\{-(d/b)^c\right\} \tag{1}$$

where $b > 0$ is the scale and $c > 0$ the shape parameter. 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, where the parameter $k = (1 + 2/c)$ of the gamma function $\Gamma(k)$. Similarly, if the fitted basal area-dbh Weibull distribution is returned into the dbh-frequency distribution, it leads to a gamma distribution with parameter $k = (1 - 2/c)$. These features were convenient when the samples were taken either from frequency distribution representing young stands or from basal area distributions representing advanced stands in the model evaluation.

Johnson (1949) presented a family of parametric distributions, which are based on transforming the distribution of the original variable into a standard normally distributed variable by applying different transformation functions. One of them, SB distribution, has been used for size distribution in forestry (e.g. Hafley and Schreuder 1977, Mønness 1982). It has the following PDF (Eq. 2) which applies the bounded transformation function shown in Eq. 3.

$$f\left(d\right) = \frac{\delta}{\sqrt{2\pi}} \frac{\lambda}{\left(d - \xi\right)\left(\xi + \lambda - d\right)} \exp\left(-0,5z^2\right) \qquad (2)$$

$$z = \gamma + \delta \ln\left[\frac{d - \xi}{\lambda + \xi - d}\right] \qquad (3)$$

where γ and δ are shape parameters, ξ and λ are location and scale parameters, and d is the diameter observed in a stand plot.

The height–diameter relationship was characterised through the use of the height curve (Eq. 4) by Näslund (1936) fitted in the linearised form (Eq. 5).

$$h = \frac{d^2}{\left(\beta_0 + \beta_1 d\right)^2} + 1.3 + \varepsilon_h \qquad (4)$$

$$\frac{d}{\sqrt{\left(h - 1.3\right)}} = \beta_0 + \beta_1 d + \varepsilon_z \qquad (5)$$

where β_0 and β_1 are the estimated parameters. The residual e_z from Eq. 5 was assumed to have

homogenous and normally distributed standard error (see Näslund 1936, p. 53).

Distribution functions were fitted to each stand separately through the maximum likelihood (ML) method (see Bailey and Dell 1973, Schreuder and Hafley 1977). The Weibull dbh-frequency distribution was fitted with the NLIN procedure in the SAS software application (see the appendix of Cao 2004) and both of the basal area-dbh distributions through the use of the author's Fortran programs as in the work of Siipilehto (1999). A three-parameter form of the SB distribution was applied by fixing the minimum $\xi = 0$. In order to avoid huge maximum endpoints (outliers in the modelling point of view), the parameter λ was restricted to be two times the maximum observed dbh at most. Näslund's height curve was fitted to each stand separately by tree species in SAS by means of the NLIN procedure, convenient for saving the estimated stand-specific parameters and the standard errors to the output data. The error term was used to generate random variation around expected tree heights for those tallied trees missing a height in the test data set. For more details on the error structure see Siipilehto (2000). The estimated parameters (see Table 1) were regarded as true values for stands, when the models for these parameters were formulated against stand age and site factors.

2.4 Construction of the Basic Models

The structure of the basic models was assumed to be multiplicative as with previous models by Nuutinen (1986), Eid (2001) and Siipilehto (2006) for stand characteristics and models by e.g. Kilkki et al. (1989), Mykkänen (1986) and Siipilehto et al. (2007) for distribution parameters. The models were fitted by means of multiple linear regression after logarithmic transformations in order to linearise the models and homogenise the residuals. The following 17 dependent variables were fitted simultaneously: 1) total basal area (G, m^{2}ha^{-1}), 2) total stem number (N, ha^{-1}), 3) arithmetic mean diameter (D, cm), 4) basal area median diameter (d_{gM}, cm), 5) dominant diameter (D_{dom}, cm), 6) mean height (H, m), 7) basal area median height (h_{gM}, m), 8) dominant height (H_{dom}, m), 9) parameter β_0 and 10) parameter β_1 of Näslund's

height curve, 11) parameters b and 12) c of the Weibull dbh-frequency distribution, 13) parameters b_G and 14) c_G of the Weibull basal area-dbh distribution, 15) parameters λ 16) and δ of the SB basal area-dbh distribution, and 17) parameter γ from the same. The basic models represented the average development of each dependent variable over the stand total (biological) age. The common structure of the candidate standwise model was as follows:

$$\ln(Y+m) = a_0 + a_1 (T+l)^k + a_2 \ln(DDY) \\ + a_3\ Origin + a_i\ S_i + a_j\ Thinn_j + \varepsilon \quad (6)$$

where $m=1$ in the model for D, d_{gM}, and D_{dom}, and for parameter b of the Weibull function and λ for the SB function, and otherwise $m=0$; $T=$ total age (in years); the candidate power k is $-1,-0.9,\ldots,-0.5$; constant l is either 0 or 10; $DDY=$ degree days (i.e. the average annual sum of the temperature on days with a mean temperatures above 5 °C); $Origin$ is a dummy variable (with a value of either 0 or 1) for artificial regeneration methods; S consists of dummy variables associated with a certain site (i) defined as forest types by Cajander (1925), and supplementary site characteristics such as stoniness and paludification; $Thinn$ consists of dummy variables associated with a thinning treatment (j) being either intermediate thinning or heavy thinning preparing for natural regeneration; a_0 to a_j, are estimated parameters of the model; and ε is the random error.

Dominant trees represented the hundred thickest trees per hectare. The models for the sum characteristics were formulated for the whole stand as a potential of the site (total stem number and total basal area). The proportion P ($0 < P \leq 1$) of Scots pine was given adding $\ln(P_{pine})$ to the model with a fixed (restricted) parameter of 1.0. Thus, if the stand was a pure pine stand, the effect was zero. The inverse of the parameter c was found to be the best transformation for the Weibull frequency distribution. The inverse transformation can be motivated by the known percentile estimator (Dubey 1967) and previous studies by Mabvurira et al. (2002) and Robinson (2004). However, a model for $\ln(c_G)$ was preferred to the inverse of c_G for the basal area-dbh distribution because of its slightly superior behaviour in advanced stands.

The models were multivariate; i.e. several models were fitted simultaneously (systems of equations) to the same data set. As a result of the correlation between the errors of different models, fitting the individual models separately would not be an efficient approach to estimation. Instead, the error structure of the ordinary least square (OLS) fit was utilised in the re-estimation of parameters of correlated response variables using the seemingly unrelated regression (SUR) method (Zellner 1962). Estimation was performed with the SYSLIN procedure in SAS (see SAS/ETS User's Guide 1993), giving the estimated parameters of the first step OLS models, the re-estimated parameters of SUR models and the cross-model error variance–covariance matrix (and its transpose and inverse matrix) as an output.

2.5 Linear Prediction Application and Model Validation

The error terms of statistical models are random variables. In this study, the error term consisted of the residual term only. The cross-model error variance-covariance matrix is valuable when calibrating the expected value (μ_1) with known variables x_2 using linear prediction theory (e.g. Lappi 1991, 1993). The best linear unbiased predictor (BLUP) for variable x_1 is

$$\hat{x}_1 = \mu_1 + \sigma_{12} \Sigma_{22}^{-1} (x_2 - \mu_2) \quad (7)$$

where x_1 is a scalar (dependent unknown variable) and x_2 is a vector (known stand variables), σ_{12} is a row vector (covariances between unknown dependent and known variables), and Σ_{22} is the variance-covariance matrix of x_2 (between known variables). The variance of the prediction error after calibrating the dependent variable x_1 is:

$$\mathrm{var}(\hat{x}_1 - x_1) = \sigma_{11} - \sigma_{12} \Sigma_{22}^{-1} \sigma_{12}' \quad (8)$$

where σ_{11} is a scalar (initial variance of the residual error of the dependent variable), and σ_{12}' is a transpose of the row vector σ_{12}. When assuming the residual errors of the logarithmic models to be multinormally distributed, half of the error variance ($s_\varepsilon^2 / 2$) had to be added to the intercept in order to avoid bias when transforming the loga-

rithmic model back to the original scale. With the inverse transformation $(1/c)$, the bias correction term is $s_\varepsilon^2 / \hat{x}_1^2$ (see Lappi 1993, p. 91–93). Thus, in the applied method, the variance had to be recalculated whenever the vector of known stand variables (x_2) for prediction was changing. An example of the BLUP estimation for Näslund's parameters is given in the appendix.

In development and validation of the structure of the candidate models, the RMSE for the fitted model was examined, as was the residual variation of each model with respect to its expected value, temperature sum and stand age. In addition, models were checked visually to ensure logical behaviour of the models when one is using a wide range of stand ages. The effect of known stand characteristics on BLUP estimation was examined with respect to RMSE – i.e. the square root of the variance shown in Eq. 8 for the final models. Theoretically, any combination of the stand variables presented could be used for linear prediction. However, the linear prediction application was briefly checked, predicting basal area-weighted characteristics (d_{gM}, h_{gM}, and G) with arithmetic stand characteristics (D, H, and N), and vice versa as in Siipilehto (2006). As for the predictability of the size distributions and height curve, additional stand characteristics were tested as calibrating variables in order to study their ability to improve reliability in the unknown parameters. Some of the combinations are more realistic than others, because the dominant height is sometimes included in the characteristics assessed for FMP, and stem number is required additionally for basal area and basal area-weighted variables (Kuvioittainen… 1998).

The final validation for the family of models generating individual trees (i.e. the dbh distribution and height curve) was performed by means of bias and RMSE for sum and dominant tree characteristics as well as for total, saw timber, pulpwood and waste wood volume. Stem total and assortment volumes were calculated by means of volume and taper curve functions by Laasasenaho (1982). Waste wood is potential energy wood and therefore an interesting characteristic at present (e.g. Tahvanainen et al. 2007, Maltamo et al. 2007). In addition, accuracy in terms of waste wood indicates the accuracy of the left tail of the predicted distribution. On the other hand,

accuracy in dominant tree characteristics was examined in order to yield an idea of the reliability of the height curve and the right tail of the distribution. These results were calculated from systematic samples of 40 trees taken from the predicted distribution.

For effective utilisation of the data, young sapling stands were also included in the data. This means that in the case of a mean height of, let us say, less than 4 m, a considerable proportion of trees can be less than 1.3 m high. Thus, an auxiliary logistic model (9) was estimated in order to describe the probability of a tree being above breast height:

$$P(h > 1.3) = \exp(a_0 + a_1 H) / [1 + \exp(a_0 + a_1 H)] \qquad (9)$$

where $a_0 = -3.530$ and $a_1 = 2.3241$. The model was fitted in SAS with the NLMIXED procedure, which fits non-linear mixed models by maximising an approximation to the likelihood integrated over random (stand) effects. The estimated proportion was used to correct the total number of stems to apply to trees above breast height when, for example, basal area and volume (volume of trees $h < 1.3$ m was ignored) were calculated. Without this correction, they could be severely overestimated in sapling stands. Indeed, according to the estimated model (9), 23%, 76%, 97% and 99.7% of Scots pine stems were higher than 1.3 m when the mean height H was 1, 2, 3 and 4 m respectively. In advanced stands, distributions were scaled according to the basal area instead of stem number. The effect of each stand variable on the vector of unknown dependent variables was focused by including them one by one as a known variable in the family of BLUP models. Some examples are given to illustrate the model application.

3 Results

3.1 Estimated Models

The SUR method proved to be advantageous, by improving the significance of some of the independent variables from the original OLS fit, Using mainly the same variables and transforma-

Table 3. Models for the stand characteristics. Estimates were highly significant (P<0.001) unless shown by **(P<0.01) or *(P<0.05).

Model	$\ln(G)$	$\ln(N)$	$\ln(D+1)$	$\ln(d_{gM}+1)$	$\ln(D_{dom}+1)$	$\ln(H)$	$\ln(hgM)$	$\ln(H_{dom})$
Intercept	−9.202	6.736	−1.595	−0.712	−0.130	−4.322	−3.308	−2.666
$\ln(T)$		−0.455						
T^{-1}	−36.443							
$T^{-0.5}$						−10.113	−9.174	−8.450
$T^{-0.6}$			−10.511	−9.288	−8.343			
$\ln(DDY)$	1.806	0.325	0.733	0.623	0.565	1.143	1.003	0.914
OMT	0.453	−0.100	0.338	0.294	0.262	0.309	0.272	0.245
MT	0.206		0.145	0.125	0.113	0.164	0.140	0.125
CT	−0.290	0.045	−0.177	−0.164	−0.147	−0.183	−0.169	−0.155
ClT	−0.317	0.659	−0.485	−0.462	−0.416	−0.624	−0.597	−0.559
Planted	0.376	−0.058*	0.203	0.176	0.145	0.141	0.113	0.097
Sown	0.182		0.083	0.068	0.059	0.066	0.051	0.041
Stoniness	−0.031		−0.016*	−0.014*	−0.014*	−0.037	−0.031	−0.028
Paludified	−0.252	0.106*	−0.154	−0.145	−0.128	−0.178	−0.168	−0.155
Thinn	−0.033	−0.059*				0.019	0.020	0.020
Thinn_R	−0.189	−0.428	0.124	0.119	0.101	0.063	0.069	0.064
R^2	0.661	0.531	0.757	0.767	0.781	0.811	0.836	0.844
RMSE	0.433	0.445	0.234	0.200	0.173	0.240	0.199	0.178

Table 4. Models for the parameters of the Näslund's height curve, Weibull and SB distribution. Estimates were highly significant (P<0.001) unless shown by **(P<0.01) or *(P<0.05).

Model	$\ln(\beta_0)$	$\ln(\beta_1)$	$\ln(b+1)$	$1/c$	$\ln(b_G+1)$	$\ln(c_G)$	$\ln(\lambda+1)$	$\ln(\delta)$	γ
Intercept	4.917	−0.174	−1.522	1.386	−0.331	−2.352	1.866	−0.801	6.836
$\ln(T)$	0.157								
T^{-1}				3.120				−2.717	
$T^{-0.5}$		3.867							4.356
$T^{-0.6}$			−9.780		−9.095	−3.156	−8.065		
$(T+10)^{-1}$	16.185								
$\ln(DDY)$	−0.818	−0.258	0.726	−0.158	0.577	0.586	0.324	0.185	−1.084
OMT	−0.086	−0.084	0.325	−0.064	0.286	0.177	0.286	0.156	−0.164
MT	−0.086	−0.036	0.138	−0.022	0.121	0.065	0.088		−0.164
CT		0.061	−0.162		−0.161	−0.051	−0.129		0.115
ClT		0.229	−0.467		−0.460	−0.051	−0.443		0.115
Planted	−0.032*		0.202	−0.047	0.168	0.127	0.095		−0.333
Sown	−0.032*		0.083	−0.013**	0.063	0.062	0.034*		−0.119**
Stoniness	0.061		−0.013*		−0.014*		−0.020*		
Paludified	0.040*	0.063	−0.153		−0.137	−0.082	−0.090		0.162
Thinn		−0.007*		−0.007*				0.023	0.018*
Thinn_R		−0.007*	0.065		0.126		0.142	0.023	−0.241**
R^2	0.360	0.775	0.740	0.323	0.769	0.350	0.605	0.126	0.091
RMSE	0.253	0.106	0.230	0.104	0.195	0.233	0.263	0.300	0.799

Table 5. Cross-model error covariances above the diagonal, variances on the diagonal (in **bold**) and correlation coefficients below the diagonal (*italics*) for the modelled stand characteristics.

	G	N	D	d_{gM}	H	h_{gM}	D_{dom}	H_{dom}
G	**0.1878**	0.0895	0.0498	0.0389	0.0667	0.0532	0.0476	0.0479
N	*0.4639*	**0.1980**	–0.0513	–0.0420	–0.0256	–0.0199	–0.0165	–0.0112
D	*0.4918*	*–0.4936*	**0.0546**	0.0397	0.0495	0.0361	0.0309	0.0289
d_{gM}	*0.4476*	*–0.4712*	*0.8485*	**0.0402**	0.0348	0.0334	0.0316	0.0268
H	*0.6419*	*–0.2393*	*0.8831*	*0.7228*	**0.0576**	0.0433	0.0296	0.0364
h_{gM}	*0.6156*	*–0.2248*	*0.7754*	*0.8365*	*0.8234*	**0.0397**	0.0284	0.0339
D_{dom}	*0.6352*	*–0.2147*	*0.7658*	*0.9106*	*0.6544*	*0.8251*	**0.0299**	0.0248
H_{dom}	*0.6230*	*–0.1421*	*0.6955*	*0.7523*	*0.7806*	*0.9569*	*0.8087*	**0.0315**

tion in the models aided in finding the logical behaviour of the models. For example, the curves for diameters or for the corresponding heights with SUR models did not cross each other (i.e. $D<d_{gM}<D_{dom}$ and $H<h_{gM}<H_{dom}$), even though some of them crossed in checking the first step OLS models. Using the transformation $\ln(Y+1)$ for diameters and the various powers in the transformation of age for dbh characteristics ($T^{-0.6}$) and height characteristics ($T^{-0.5}$) helped to achieve a reasonably small dbh value for the short trees (i.e. logical behaviour just above breast height). Degrees of determination were quite high for dbh and height characteristics (76–84%) but also considerable for basal area and stem number, at 66 and 53% (Table 3).

The most typical site for Scots pine was *Vaccinium*-type (VT) sub-xeric heath (Cajander 1925), representing the reference level of the models. The effects of site fertility on stand variables were characterised using dummy variables. For example, on *Oxalis-Myrtillus*-type grove-like sites (OMT sites), the heights (H_{dom}, h_{gM}, and H) were about 28–40% greater than with VT, while on *Calluna*-type xeric heath (CT) these heights were about 15% and for *Cladonia*-type (ClT) about 40% lower (Table 3). The effect of site fertility on the stem number and basal area was obvious and could be explained by the differences in growth potential with respect to site classes. On the other hand, stoniness and paludification reduced stand mean variables by around 3% and 14%, respectively. Artificial regeneration showed a significant effect on stand characteristics and the effect of planting was greater than

that of sowing. Intermediate thinning (Thinn) was significant for sum and height characteristics, while heavier thinning in preparation for natural regeneration (Thinn$_R$) was significant for each stand characteristic (see Table 3).

Site factors, stand origin and past thinnings had an effect on the height curve as well as on dbh distributions. Note that scale parameter b of the Weibull functions and the maximum λ of the SB function had the same model structure as dbh characteristics because of the close correlation between them (see Table 4). The logical dependencies of $D<b$, $d_{gM}<b_G$, and $D_{dom}<\lambda$ were fulfilled. Degrees of determination were considerable for the parameters modelled (32–78%), except for the shape parameters of the SB distribution (9–13%). Note that the adjacent dummy variables have been combined where they have equal estimated parameters in Table 4 (e.g. OMT and MT in the model for $\ln(\beta_0)$).

Covariances and correlations of the cross-model residual errors of the stand characteristics are given in Table 5. It was not surprising that correlation between many stand characteristics was very strong. A correlation coefficient above 0.7 was found in 14 out of 28 cases (see Table 5). The highest correlation, 0.96, was between h_{gM} and H_{dom}. Covariances and correlations between the residuals of the modelled parameters of size distributions and height curve and stand characteristics are shown in Table 6 and Fig. 1. Analysis revealed relatively high absolute values of correlation coefficients between the height curve parameters and some stand characteristics (e.g. $r=-0.4$ between β_0 and G and H, and
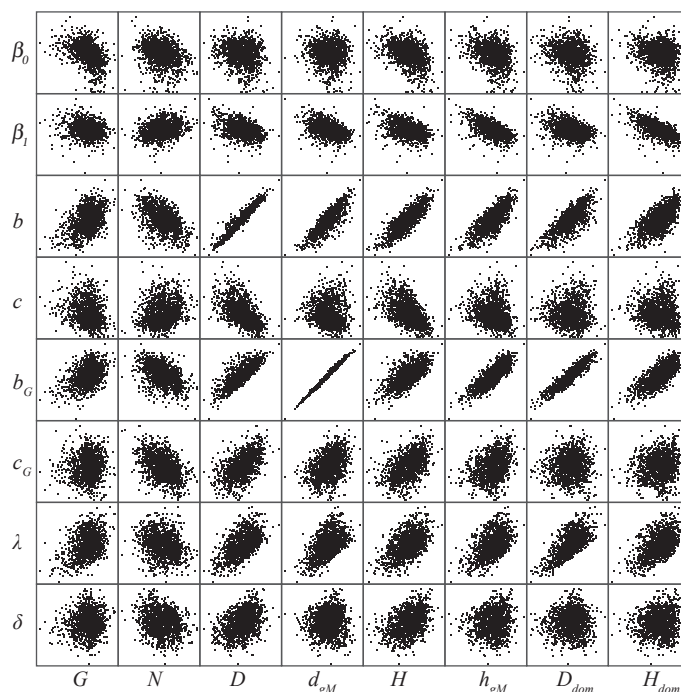
Fig. 1. Correlations of the cross-model residual errors between the modelled stand characteristics and the parameters. Stand characteristics from left to right were G, N, D, d_{gM}, H, h_{gM}, D_{dom} and H_{dom}, and the parameters from top to bottom were β_0 and β_1 of the Näslund's curve, parameters b and c of the W_N, b_G and c_G of W_G and λ and δ of the SB_G distribution.

$r=-0.6$ between β_1 and h_{gM} and H_{dom}). Because parameter b represents the 63rd percentile of the Weibull distribution, b of W_N was highly correlated with mean D ($r=0.97$) and b_G of W_G with basal area median dbh, d_{gM} ($r=0.99$). There was no such close correlation between the residuals of the shape parameter c and individual stand characteristics. The highest correlation was found between c and D ($r=-0.53$). According to correlation scatter plots in Fig. 1, the assumption of the linear correlation of BLUP held quite nicely.

3.2 The Effect of BLUP Estimation on the RMSE

BLUP is a flexible approach – any of the known stand variables can be used for predicting (cali-

brating the expected value of the basic model using Eq. 7) the unknown dependent variable (see Lappi 1993, 2006). Similar to Siipilehto (2006), stand age was the only stand variable required in the basic models. The prediction efficiency of the individual stand variables in terms of RMSE (i.e. the square root of Eq. 8) has been given for stand characteristics for Norway spruce (see Table 4 in Siipilehto 2006). Models for Scots pine behaved in astonishingly similar ways. For example, calibrating d_{gM} with N, D and H resulted in RMSE of 10.8 and 10.5% for spruce and pine, respectively. However, calibration with a sum and a mean dbh characteristic decreased the original RMSEs of 43.2 and 44.5% for G and N to only 13.6% and 20.3% respectively. Thus, the calibrated sum characteristics for pine proved more accurate than those for spruce (18.0 and 24.3%) even though the

Table 6. Cross-model error covariances and correlation coefficients (*italics*) between modelled stand characteristics and parameters. Correlations |r|>0.5 are highlighted in **bold**.

		G	N	D	d_{gM}	H	h_{gM}	D_{dom}	H_{dom}
β_0	cov	–0.0464	–0.0306	–0.0092	–0.0055	–0.0241	–0.0164	–0.0081	–0.0112
	corr	*–0.4235*	*–0.2713*	*–0.1552*	*–0.1080*	*–0.3968*	*–0.3260*	*–0.1840*	*–0.2484*
β_1	cov	–0.0077	0.0133	–0.0102	–0.0092	–0.0121	–0.0125	–0.0070	–0.0121
	corr	*–0.1688*	*0.2835*	*–0.4139*	*–0.4341*	*–0.4783*	***–0.5948***	*–0.3818*	***–0.6467***
b	cov	0.0450	–0.0523	0.0517	0.0398	0.0468	0.0360	0.0307	0.0288
	corr	*0.4525*	***–0.5126***	***0.9657***	***0.8660***	***0.8513***	***0.7870***	***0.7753***	***0.7060***
c	cov	–0.0081	0.0107	–0.0128	–0.0016	–0.0128	–0.0037	0.0002	–0.0024
	corr	*–0.1810*	*0.2327*	***–0.5279***	*–0.0773*	***–0.5141***	*–0.1812*	*0.0134*	*–0.1318*
b_G	cov	0.0382	–0.0388	0.0374	0.0389	0.0328	0.0323	0.0315	0.0263
	corr	*0.4526*	*–0.4470*	***0.8209***	***0.9951***	***0.7007***	***0.8313***	***0.9355***	***0.7592***
c_G	cov	0.0096	–0.0436	0.0308	0.0190	0.0260	0.0162	0.0040	0.0087
	corr	*0.0953*	*–0.4211*	***0.5677***	*0.4067*	*0.4663*	*0.3504*	*0.1001*	*0.2105*
λ	*cov*	0.0388	–0.0301	0.0346	0.0282	0.0325	0.0268	0.0307	0.0251
	corr	*0.3408*	*–0.2575*	***0.5635***	***0.5347***	***0.5144***	***0.5117***	***0.6762***	***0.5385***
δ	cov	0.0086	–0.0299	0.0254	0.0003	0.0258	0.0069	0.0006	0.0064
	corr	*0.0662*	*–0.2244*	*0.3627*	*0.0044*	*0.3591*	*0.1157*	*0.0114*	*0.1202*
γ	cov	–0.0047	0.0517	–0.0264	–0.0448	–0.0152	–0.0267	–0.0080	–0.0105
	corr	–0.0136	0.1453	–0.1416	–0.2797	–0.0791	–0.1677	–0.0582	–0.0738

original RMSEs for pine were higher. Nevertheless, due to overall similarity, the other detailed results between stand characteristics are not worth giving here, but they can be easily calculated using the Eq. 8 and the covariance matrix in Table 5.

The effect of a stand characteristic in improving the precision of the size distribution and height curve parameters varied clearly between the characteristics utilised. It was obvious that knowing both one dbh and one height characteristic (D and H or d_{gM} and h_{gM}) was relatively efficient for calibrating the parameters of the height curve in terms of reduction in RMSE (Fig. 2: A, C). In addition, known D_{dom} and H_{dom} increased the accuracy in β_0 and β_1 but the additional sum characteristics, either N or G, had no effect. Calibration of the parameters of the Weibull and SB distributions was relatively efficient with D or d_{gM}. Adding mean height and sum characteristics (H and N or h_{gM} and G) affected the RMSE for size distribution parameters only marginally. However, adding dominant tree characteristics clearly reduced the RMSE for each parameter. As for the sum characteristics, additional knowledge of the basal area did not improve the accuracy

(Fig. 2: A, B), but the additional stem number was relatively efficient in reducing the RMSE for each size distribution parameter (Fig. 2: C, D). The most efficient calibration was achieved with the additional mean diameter. Thus, including both D and d_{gM} in the BLUP models reduced the RMSE for each parameter as shown in the bar furthest to the right in Fig. 2.

Fig. 2 proves that the dbh-frequency distribution could be more efficiently calibrated with stand variables that are directly associated with frequency distribution (i.e. D, H, and N). Stand variables that are typically assessed by means of relascope sample plots (i.e. d_{gM}, h_{gM}, and G) were surprisingly inefficient for calibrating the shape parameter of the Weibull distribution. Knowledge of the additional stem number seemed quite essential for the shape parameter of W_N and W_G but simultaneously also enhanced b of W_N (see Fig. 2: C, D). Mean and dominant diameter and stem number were significant additional variables for calibration of the models for the SB_G parameters, especially when we were dealing with the stand characteristics on the arithmetic scale (Fig. 2: B). On the other hand, when starting with the basal area-based characteristics (Fig. 2: D),
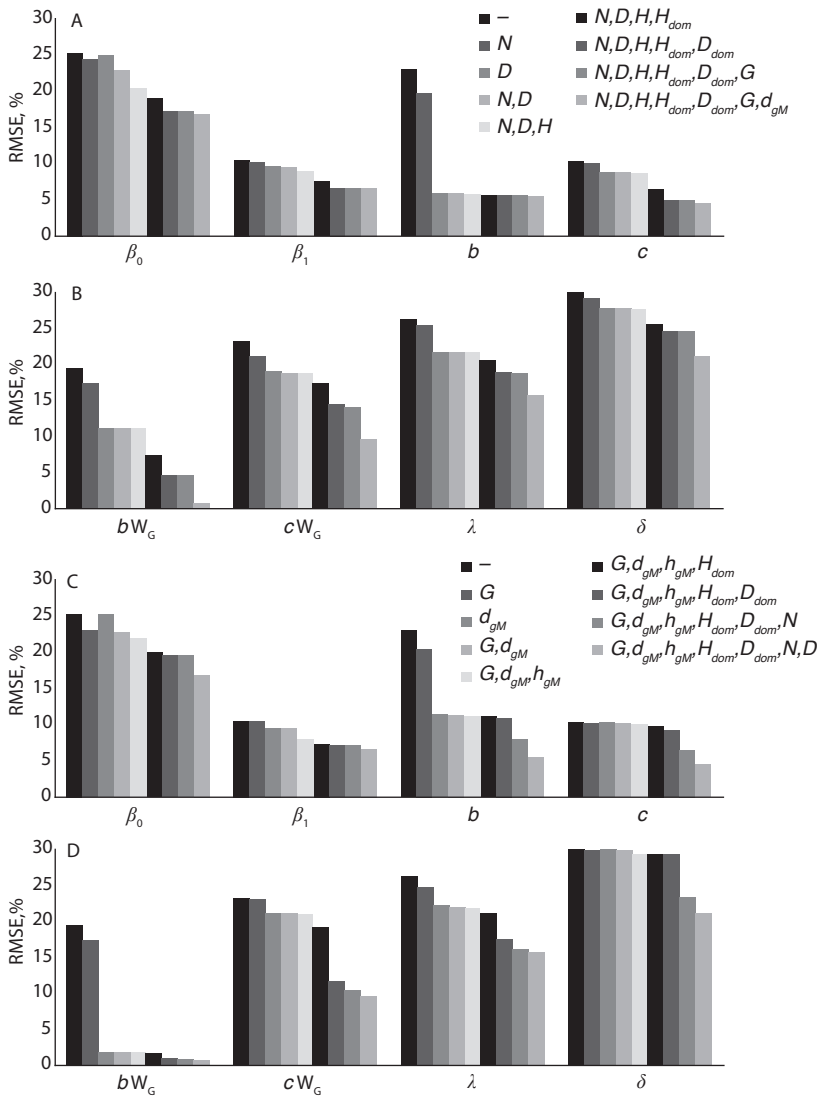
Fig. 2. The effect of BLUP estimation on the RMSE for the modelled parameters. Calibration was based on stand characteristics which were either mainly arithmetic (A, B) or basal area-weighted (C, D). The parameters are β_0 and β_1 of the Näslund's curve, b and c of the Weibull frequency distribution or that of basal area-dbh distribution (W_G) and λ and δ of the SB distribution. Variables used in calibration are shown in the legend.

the prediction for λ benefited each of the stand characteristics included but the RMSE for δ was reduced only marginally until the additional stem number or additional mean dbh was included. It is worth noting that the effect of stand charac-teristics on the estimate for γ of the SB distribution was quite marginal (figure not shown). The RMSE of 80% for γ was reduced somewhat only when additional D_{dom} (78–64%), N (61%), D or d_{gM} (60%) was included in the BLUP model.

3.3 Model Behaviour with Respect to Stand Density

The following example illustrates the effect of stand density on the stand characteristics and dbh distribution. Let three given hypothetical stands represent a 25-year-old Scots pine stand sown on a sub-xeric (VT) site in central Finland (annual temperature sum of 1200 °C), and assume that the density of these Scots pine stands is 2000 ha^{-1}, 4000 ha^{-1}, and 8000 ha^{-1}, respectively. Simultaneously with increased density, the shapes of the distributions shifted from more or less symmetrical bell shapes to being more and more skewed to the right as shown in Fig. 3. In addition, dbh distributions moved to the left, resulting in mean diameters of 8.0, 6.5, and 5.3 cm, respectively. The above mean diameters from the predicted Weibull distributions followed the corresponding density-calibrated BLUP estimates for D – namely, 7.8, 6.3, and 5.1 cm. Also, the calculated basal areas (11.7, 15.8, and 21.3 m^{2}ha^{-1}) followed quite a similar pattern to that of the density-calibrated BLUP estimates for G (11.2, 15.3, and 20.9 m^{2}ha^{-1}). Similar comparisons were also made for 25-year-old naturally regenerated stands and 20-year-old planted stands. According to these comparisons, the BLUP estimation for D and G differed only slightly from the respective characteristic calculated from the calibrated Weibull distribution.

3.4 Calibration of the Dbh Distribution to a Real Stand

The following example shows the calibration effect of the known stand characteristics on the predicted dbh distribution against one existing stands. Size distributions of the basic model represent the average for a particular site, stand age and origin. Thus, because of natural variation and varying management practices, the observed structure and the expected structure can be considerably different. The example, plot no. 1017, represents a 21-year-old planted stand on an MT site in southern Finland (DDY: 1245 °C). Prediction for this stand was performed with the W_N model. The observed N, D, and H were 950 ha^{-1}, 11.0 cm, and 7.1 m, respectively. The respective
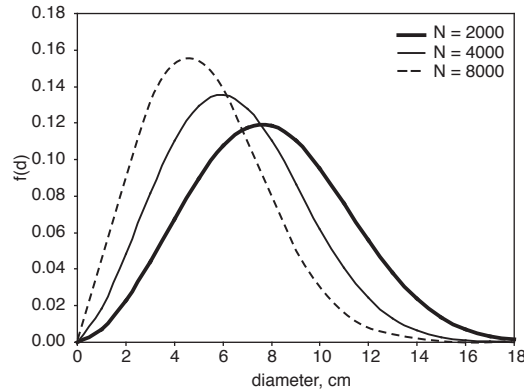

Fig. 3. The effect of stand density ($N = 2000$–8000 ha^{-1}) on the dbh distribution of a 20-year-old Scots pine stand.

expected values of the basic models were as follows: 1224 ha^{-1}, 9.1 cm, and 7.2 m. The additional stand characteristics checked for calibration were G, d_{gM}, D_{dom} and H_{dom}, which had values of 9.6 m^{2}ha^{-1}, 12.4 cm, 15.7 cm and 7.7 m. The respective expected values were as follows: 8.6 m^{2}ha^{-1}, 11.8 cm, 15.9 cm and 9.6 m.

The expected W_N distribution for the basic models ($W(0)$) was to the left of the observed distribution, because of the lower expected D (see Fig. 4). In the second step, prediction was conducted with N, D, and H as common known variables for young stands. The resulting dbh distribution was situated comfortably over the observed distribution, but the tail toward the thickest trees was too long, resulting in about a 5 m^{3}ha^{-1} overestimation in total volume. The additional dominant tree characteristics increased the peakedness of the distribution, which clearly improved the fit. In this case, the change was resulted in mainly by H_{dom} because the observed D_{dom} was close to its expectation. Calibrating with additional G slightly decreased but with d_{gM} significantly increased the fit of the Weibull distribution as compared with the three common calibration variables (see Fig. 4). The two best-fitting distributions were as follows: 1. calibrated with dominant tree characteristics, which resulted in an error in total volume of 1.0 m^{3}ha^{-1} and 2. calibrated with d_{gM} , which resulted in an error in total volume of –1.8 m^{3}ha^{-1}.
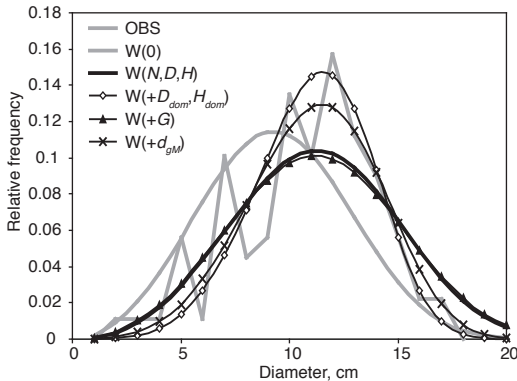
Fig. 4. Example of the BLUP estimation for the Weibull frequency distribution for a 21-year-old planted Scots pine stand on MT site in southern Finland. W(0) represents expected Weibull distribution of the basic model. W(N,D,H) was calibrated with N, D and H and W(+D_{dom},H_{dom}) with additional dominant tree characteristics, etc.

If the same stand's prediction was done with G, d_{gM}, and h_{gM} (common for advanced stands) the resulted distribution differed only slightly from the expected one. Thus, calibrating the Weibull frequency distribution with the above variables did not improve the fit as efficiently as the arithmetic characteristics did and the dominant tree characteristics were also less efficient. Instead, calibration with the additional N or D resulted in a good fit.

3.5 Final Validation of the Family of Models

The linear prediction application provided a consistent basis for comparison of alternative size distribution models. The family of models was evaluated through use of the combinations of stand characteristics for the linear prediction proceeding from the common FMP variables. In young stands, the bias in total volume showed 1.3–4.6% underestimation with the three independent variables (see Table 7). In general, when additional predictors were included in BLUP models, the W_N model yielded the least biased volume estimates. Inclusion of G-reduced biases most efficiently with basal area-dbh dis-

tribution models – i.e. SB_G and W_G – whereas H_{dom} was efficient with the dbh-frequency W_N model (see Table 7). Note that whenever D_{dom} was included, it resulted in clear underestimation in volume characteristics with the SB_G and W_G models for young stands. It was noted that the greatest underestimates were found when the difference between D and D_{dom} was great ($D_{dom} > 2 \times D$).

When the W_N model was used, the effect of each additional variable on the RMSE for volume and assortments was clear, most obviously when d_{gM} was included, and smallest when G was included (see Table 8). However, the calibration of the basal area-based distribution models reduced the RMSE only slightly, most when H_{dom} or d_{gM} was included. All in all, d_{gM} reduced the RMSE value for total and pulpwood volume by 3–12% and 6–16 percentage points, respectively. The combination for the best two additional variables was evidently d_{gM} and h_{gM} for W_N and SB_G models. Apart from those, H_{dom} and D_{dom} obviously enhanced the accuracy in the validated characteristics with the W_N model while G and H_{dom} did the same with the W_G and SB_G models.

In terms of bias and RMSE in the validated characteristics, W_N was by far the best model for young stands. It was superior in terms of bias and RMSE in volume characteristics and basal area. However, dominant height was least biased with the SB_G model. On average, the W_G model generated the most accurate dominant diameter, providing the least bias for five cases and the smallest RMSE in six out of 11 cases (see Tables 7 and 8).

In advanced stands, total volumes were practically unbiased and the differences in RMSEs were rather small between the basal area-dbh distribution models whereas W_N resulted in about 1% underestimation and systematically higher RMSEs than the basal area-dbh distribution models (see Tables 9 and 10). The bias and RMSE in commercial assortments were generally low, with standard three stand variables with the W_G and SB_G models (see Tables 9 and 10), but they were still reduced with the addition of D or N, whereas H_{dom} reduced RMSE only (Table 10). Also, when the W_N model was used, a considerably high bias of about 10% in timber assortments was reduced to 6–7% with additional D or N (Table 9). RMSE

Table 7. The bias (%) in stand characteristics using the alternative distribution models for young stands ($H<9$ m). The first row represents SOLMU variables (N, D, H) and below that are the additional stand variables used for predictions. The best validating criterions are highlighted in **bold**.

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
	G	2.43	2.84	**0.69**	**0.15**	1.93	2.81
	d_{gM}	**2.06**	2.86	−1.20	**0.23**	1.50	2.75
	G, H_{dom}	**1.06**	**1.33**	−0.05	−0.37	0.95	1.77
	G, D_{dom}	**2.02**	**2.74**	−0.94	**0.10**	1.42	2.74
	D_{dom}, H_{dom}	**1.27**	**1.74**	−0.68	−0.27	1.09	1.98
	d_{gM}, h_{gM}	**1.59**	2.26	−1.18	**0.13**	1.42	2.16
	G, D_{dom}, H_{dom}	**1.56**	**2.13**	−0.79	−0.09	1.21	2.08
	All	**2.08**	**2.84**	−1.01	**0.24**	1.39	2.40
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
	G	**0.39**	**−0.98**	6.63	−0.47	**−0.16**	1.92
	d_{gM}	2.29	**1.61**	5.25	2.21	**0.57**	2.35
	G, H_{dom}	2.90	2.11	6.36	2.57	**0.01**	1.43
	G, D_{dom}	7.10	7.44	5.66	5.60	1.41	2.72
	D_{dom}, H_{dom}	6.56	6.80	5.53	5.65	0.92	1.94
	d_{gM}, h_{gM}	2.04	**1.30**	5.28	2.20	**0.57**	1.83
	G, D_{dom}, H_{dom}	5.56	5.59	5.47	4.69	1.02	2.02
	All	7.49	7.06	9.35	6.93	1.86	2.73
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**
	G	−0.43	−1.46	4.19	−1.32	−1.74	**1.32**
	d_{gM}	1.97	1.78	2.75	1.24	−1.88	**1.32**
	G, H_{dom}	3.04	2.85	3.87	2.37	−1.22	**1.04**
	G, D_{dom}	7.46	8.52	3.17	5.57	**0.85**	**2.44**
	D_{dom}, H_{dom}	6.18	6.93	3.13	4.93	**0.00**	**1.62**
	d_{gM}, h_{gM}	1.50	**1.23**	2.65	1.17	−1.81	**0.81**
	G, D_{dom}, H_{dom}	6.03	6.75	3.10	4.74	**0.26**	**1.74**
	All	5.66	6.48	2.30	4.41	**0.65**	**2.06**

Table 8. The RMSE (%) for stand characteristics using the alternative distribution models for young stands. The first row represents SOLMU variables (N, D, H) and below that are the additional stand variables used for predictions. The best validating criterions are highlighted in **bold**.

Model	Predictors	Volume	Pulp	Waste	G	D_{dom}	H_{dom}
W_N	N, D, H	**21.54**	**28.11**	**14.43**	**11.28**	16.07	14.07
	H_{dom}	**11.39**	**15.41**	**12.61**	**7.62**	8.43	**4.35**
	D_{dom}	**11.38**	**15.32**	**11.82**	**6.45**	**3.94**	**8.91**
	G	18.62	24.04	13.45	9.76	13.90	12.65
	d_{gM}	**9.52**	12.43	10.30	**5.45**	7.83	9.05
	G, H_{dom}	**11.53**	**15.44**	**12.39**	**7.44**	8.03	**4.35**
	G, D_{dom}	**11.38**	**15.39**	**11.88**	**6.53**	**4.05**	**8.94**
	D_{dom}, H_{dom}	**9.91**	**13.50**	**11.75**	**6.34**	**3.75**	**4.80**
	d_{gM}, h_{gM}	8.70	11.30	**10.24**	5.45	7.70	6.66
	G, D_{dom}, H_{dom}	**9.93**	**13.58**	**11.82**	**6.42**	**3.89**	**4.84**
	All	**8.43**	**10.99**	**11.12**	**5.28**	**5.82**	**6.41**
W_G	N, D, H	22.65	28.66	22.46	13.43	**14.60**	13.38
	H_{dom}	17.87	22.07	21.25	13.32	**8.32**	4.84
	D_{dom}	22.67	27.52	18.94	15.78	4.59	9.20
	G	21.97	27.25	21.62	14.39	**12.83**	**12.12**
	d_{gM}	19.19	22.39	21.06	15.52	**7.31**	**8.63**
	G, H_{dom}	17.33	21.12	21.14	12.64	**7.89**	4.82
	G, D_{dom}	23.83	28.91	19.01	16.35	5.40	9.34
	D_{dom}, H_{dom}	19.19	23.30	18.53	14.56	3.93	5.21
	d_{gM}, h_{gM}	19.06	22.23	20.96	15.51	**7.31**	**6.49**
	G, D_{dom}, H_{dom}	20.47	24.84	18.75	15.03	4.17	5.28
	All	35.50	40.34	32.63	29.40	9.12	7.59
SB_G	N, D, H	23.07	29.26	18.85	13.97	14.62	**13.21**
	H_{dom}	21.64	26.61	17.59	15.34	9.61	6.01
	D_{dom}	27.45	33.56	15.71	19.12	9.06	10.35
	G	23.77	29.51	17.55	15.61	13.50	12.18
	d_{gM}	17.49	20.98	13.05	12.21	8.36	8.65
	G, H_{dom}	21.23	25.91	17.25	15.07	9.53	6.02
	G, D_{dom}	27.58	33.74	15.78	19.07	9.47	10.41
	D_{dom}, H_{dom}	24.35	29.62	15.50	17.78	7.77	6.54
	d_{gM}, h_{gM}	16.37	19.57	12.93	11.92	8.35	6.57
	G, D_{dom}, H_{dom}	24.89	30.35	15.53	17.92	7.90	6.55
	All	23.82	28.97	13.71	17.60	8.41	7.36

Table 9. The bias (%) in stand characteristics using the alternative distribution models for advanced stands (H≥9 m). The first row represents SOLMU variables (G, d_{gM}, h_{gM}) and below that are the additional stand variables used for predictions. The best validating criterions are highlighted in **bold**.

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	0.66	5.57	–6.03	1.16	**–0.74**	0.95	**0.09**
	N, H_{dom}	0.88	7.41	–7.59	–0.17	–1.69	0.91	0.31
	N, D_{dom}	0.83	7.31	–7.60	–0.01	–1.63	0.98	0.18
	D_{dom}, H_{dom}	0.96	9.51	–9.87	2.82	–0.86	1.75	0.38
	D, H	0.42	5.38	–6.32	**0.78**	**–0.81**	0.94	**–0.16**
	N,D_{dom},H_{dom}	0.78	7.26	–7.65	**–0.02**	–1.58	0.98	0.19
	All	0.39	5.42	–6.42	**1.03**	–0.75	0.95	**–0.07**
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	0.14	1.42	–1.95	3.14	1.46	**0.23**	–0.20
	N, H_{dom}	**0.20**	**1.87**	**–2.50**	3.76	1.64	**0.04**	–0.04
	N, D_{dom}	**0.20**	**1.99**	**–2.51**	1.71	**0.03**	**0.07**	–0.17
	D_{dom}, H_{dom}	0.14	**1.93**	**–2.62**	2.27	0.49	**0.22**	–0.15
	D, H	**–0.05**	**1.23**	–2.20	2.89	1.46	**0.23**	–0.42
	N,D_{dom},H_{dom}	**0.15**	**1.94**	**–2.55**	1.48	**0.00**	**0.06**	–0.17
	All	**0.00**	**1.61**	**–2.46**	1.34	**–0.08**	**0.31**	–0.32
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	**0.11**	**0.62**	**–0.59**	**–0.79**	–1.10	–2.32	–0.95
	N, H_{dom}	0.25	1.59	–1.54	–2.04	**–1.44**	–1.96	–0.64
	N, D_{dom}	0.62	3.82	–3.48	–4.82	–4.20	–0.48	–0.32
	D_{dom}, H_{dom}	0.52	4.80	–5.21	**–1.63**	–1.21	–0.39	–0.29
	D, H	–0.12	0.44	**–0.84**	–1.53	–1.30	–2.31	–1.19
	N,D_{dom},H_{dom}	0.63	3.95	–3.60	–5.18	–4.24	–0.54	–0.33
	All	0.33	3.47	–3.63	–5.80	–5.01	–0.69	–0.59

in waste wood volume was about 30%, indicating inaccuracy in the left-hand tail of the distribution. Including D improved the accuracy in the waste wood fraction greatly, especially with the W_N and SB_G distribution models. The W_G model provided practically unbiased dominant tree characteristics, while the W_N and SB_G models produced slight biases (<3%) in D_{dom}. Biases in stem numbers were typically less than 2%. Surprisingly, D was superior to N in reducing RMSE in the stem number (see Table 10).

When the models were compared with each other in terms of bias, the W_G model seemed slightly superior in the validated stand characteristics. However, SB_G also performed well in volume estimates and W_N yielded the least biased waste wood volume. The smallest RMSEs for stem number and waste wood was provided by the W_N model. The two basal area-dbh distribution models produced nearly equal RMSEs in dominant tree characteristics and were superior to the dbh-frequency distribution model.

Table 10. The RMSE (%) for stand characteristics using the alternative distribution models for advanced stands. The first row represented SOLMU variables (G, d_{gM}, h_{gM}) and the variables below represented the additional information used for predictions. The best validating criterions are highlighted in **bold**.

Model	Predictors	Volume	Logs	Pulp	Waste	N	D_{dom}	H_{dom}
W_N	G, d_{gM}, h_{gM}	2.20	13.37	13.66	**27.58**	**14.43**	5.47	3.55
	H_{dom}	2.24	13.17	13.52	28.09	**14.28**	5.10	2.63
	D_{dom}	2.23	12.64	12.91	27.84	**14.38**	3.84	3.41
	N	2.39	12.69	13.02	22.36	**11.84**	5.31	3.57
	D	2.06	9.92	10.50	**11.59**	**5.35**	4.88	3.55
	N, H_{dom}	2.49	12.76	13.11	22.97	**11.85**	4.90	2.67
	N, D_{dom}	2.42	12.39	12.55	**22.31**	**11.89**	3.67	3.45
	D_{dom}, H_{dom}	2.38	12.56	13.00	28.03	**14.38**	3.84	2.80
	D, H	2.44	9.86	10.53	**10.55**	**5.27**	4.80	3.42
	N,D_{dom},H_{dom}	2.54	12.33	12.71	**22.61**	**11.88**	3.67	2.81
	All	2.65	9.46	10.11	**10.54**	**5.24**	3.48	2.85
W_G	G, d_{gM}, h_{gM}	**1.70**	**6.32**	**8.39**	29.22	16.00	**4.78**	**3.44**
	H_{dom}	**1.81**	**6.03**	**7.92**	29.80	15.69	**4.19**	2.34
	D_{dom}	**1.95**	**6.17**	**7.37**	29.42	17.25	2.01	3.18
	N	**1.74**	**6.14**	**8.10**	25.47	14.19	**4.89**	**3.42**
	D	**1.66**	**5.68**	**7.39**	21.92	12.20	**4.43**	**3.40**
	N, H_{dom}	**1.88**	**5.91**	**7.71**	26.05	13.83	**4.28**	2.37
	N, D_{dom}	**1.96**	**5.93**	**7.04**	27.50	15.73	**2.18**	3.19
	D_{dom}, H_{dom}	2.20	**6.21**	**7.46**	29.59	17.31	2.00	2.50
	D, H	**2.02**	**5.25**	**7.60**	21.98	12.19	**4.42**	3.21
	N,D_{dom},H_{dom}	**2.21**	**5.97**	**7.11**	27.72	15.79	2.17	2.49
	All	2.43	**5.40**	**6.95**	26.03	14.96	1.96	2.69
SB_G	G, d_{gM}, h_{gM}	1.83	7.97	10.08	28.12	15.36	5.22	3.51
	H_{dom}	1.96	7.81	9.55	**28.00**	15.15	4.34	**2.26**
	D_{dom}	1.99	8.94	9.92	**26.24**	14.72	**1.66**	**3.10**
	N	1.96	7.67	10.12	**21.98**	13.17	5.60	3.55
	D	1.74	6.85	8.85	13.26	10.85	5.19	3.53
	N, H_{dom}	2.11	7.23	9.32	**21.81**	13.30	4.56	**2.33**
	N, D_{dom}	2.27	8.16	9.28	23.07	18.28	2.15	**3.15**
	D_{dom}, H_{dom}	**2.18**	8.85	10.09	**26.54**	14.91	**1.67**	**2.35**
	D, H	1.86	6.27	8.70	12.13	10.60	5.13	**3.28**
	N,D_{dom},H_{dom}	3.14	8.65	9.40	23.52	18.22	**2.07**	**2.42**
	All	**2.24**	7.06	8.75	16.98	21.31	**1.74**	**2.42**

4 Discussion and Conclusions

One of the main advantages of the applied approach is the flexible prediction with varying sets of known stand variables. The essential predictors were the age of the stand, site factors, and the average temperature sum corresponding to the location of the stand. The temperature sum for a stand can be calculated according to the location in terms of latitude, longitude and altitude (Ojansuu and Henttonen 1983), but tabulated munici-

pal average temperature sums are also available in MOTTI. The formulation of the prediction equations for eight stand characteristics was only slightly modified from the corresponding models for Norway spruce that were presented by Siipilehto (2006). It is worthy of mention that the models for Norway spruce were not sensitive to errors in stand age, if correct stand variables were available for linear prediction. Indeed, an assumed 25% bias in stand age resulted in the bias of only 2–5% in the five stand characteristics

out of the three calibrating stand variables (see Siipilehto 2006).

The expectations of the basic models provided reasonable development for stand characteristics by site and stand origin. In fact, dominant heights at a total age of 100 years for different site types (Cajander 1925) and stand origins seemed comparable with the existing Finnish H_{100} site index curves, for naturally regenerated stands by Gustavsen (1980) and for artificially regenerated stands by Vuokila and Väliaho (1980). However, the above-mentioned site index curves showed slower early ($T < 40$ years) development in comparison with the basic model for H_{dom} in this study. Similarly, the initial dominant height development for young artificially regenerated Scots pine stands as described by Varmola (1993) had a more pronounced sigmoid form and produced smaller H_{dom} values for the first ten years, but the later development was comparable with our results.

In a MOTTI simulator, we have user-defined initial stand as an option. It is convenient for a non-experienced user to include the possibility of a simulator to generate expected age, species and site-dependent values for standard stand characteristics rather than force the user to guess reasonable values for them. However, a typical MOTTI application is to simulate young stands with various user-defined initial stand densities. The brief application for sapling stands showed that the effect of stand density on dbh distribution was reasonable and consistent with the density-calibrated BLUP estimates for diameter-based stand characteristics. In addition, the effect of stand density on the stand characteristics was generally in line with the models of Varmola (1993) and Huuskonen and Hynynen (2006). In contrast, stand density had a much smaller effect on mean and dominant diameter or mean height in the models by Huuskonen and Miina (2007).

The two-parameter Weibull function was applied for both dbh-frequency distribution (W_N) and the basal area-dbh distribution (W_G) models motivated by the simplicity in weighting (Gove and Patil 1998) but also because Maltamo et al. (1995) reported that the two-parameter Weibull provided systematically better volume estimates than the three-parameter Weibull did. In line with this study's findings, Gobakken and Næsset (2004)

and Palahi et al. (2007) noticed that determining two-parameter Weibull distributions for advanced stands as a basal area-dbh distribution instead of a dbh-frequency distribution results in more accurate stand characteristics. On the other hand, Maltamo et al. (2007) did not notice much difference between these two approaches. Furthermore, Gobakken and Næsset (2005) were surprised that the two-parameter Weibull could produce results comparable with the flexible percentile-based distributions. It is well known that the SB function is more flexible than the Weibull function (e.g. Hafley and Schreuder 1977). Nevertheless, the superior flexibility had a minor impact because the shape parameters of the SB distribution were not calibrated efficiently with the stand characteristics that were included.

Typical Finnish combinations of the variables assessed proved to be relatively efficient, but, concurrently, at least three variables were required for reliable stand volume estimates. For example, RMSE in the total volume for advanced stands was 39% with the basic model for W_N distributions and it was reduced to 10% with two calibrating variables and further to 2% in calibration with the common three variables. In line with the findings of Siipilehto (1999), Kangas and Maltamo (2000a, 2000b), and Mabvurira et al. (2002), additional stem number knowledge in combination with G, d_{gM}, and h_{gM} produced variability in the shapes of predicted size distributions, which generally improved the fit. Knowledge of the stem number reduced the RMSE for N to 12–14% from the predicted distributions when the basal area was simultaneously correct. In comparison, distribution models by Kangas and Maltamo (2000b) and Siipilehto et al. (2007) including N as an additional predictor yielded a 5% RMSE in N with the independent INKA test data set.

Maltamo et al. (2000) have shown the ability of medians of frequency (d_M) and basal area-dbh distributions (d_{gM}) to predict the shape of the Weibull frequency distribution and also multimodality in the case of predicted percentile-based distribution for Scots pine-dominated heterogeneous stands. When D and d_{gM} were combined with other common FMP variables in this study, the accuracy in volume characteristics was considerably enhanced, especially with the W_N model. In advanced stands, RMSEs of about 30% in waste

wood volume indicated inaccurate left tails of the predicted distributions. Additional D improved the accuracy of waste wood volume most efficiently (Table 10). Furthermore, D had an even greater ability to improve the accuracy in the stem number generated than the inclusion of N itself did.

Dominant tree characteristics showed a rather confusing calibration effect. On one hand, they showed an ability to reduce RMSEs for the modelled parameters of the distribution functions and height curve (see Fig. 2). Applying them for calibration resulted in reduced RMSEs for each validated stand characteristic in the test data when W_N was used for prediction involving young stands (see Table 9). Furthermore, derived from the model by Siipilehto (2009), $1/c$ of the dbh-frequency Weibull distribution could be assumed to be closely correlated with $\ln(D_{dom}/D)$, which, in turn, can be written as $\ln(D_{dom}) - \ln(D)$. However, D_{dom} as a calibrating variable resulted in clearly biased volume estimates when basal area-dbh distribution models were applied for young stands (see Table 8). In advanced stands, despite their ability to improve accuracy for dominant tree characteristics of the predicted distributions, their effect on the accuracy of stand volume was not as evident. The reason was probably related to the fact that dominant tree characteristics were too closely correlated with basal area medians so could not provide as useful additional information. The ratio between dominant and mean characteristics has been used previously for predicting the Weibull distribution by Siipilehto (2009), while Sarkkola et al. (2005) used a similar formulation but utilised the maximum diameter instead of the dominant diameter.

In a study by Siipilehto (1999), parameter β_1 of Näslund's function was constructed to be linearly dependent on d_{gM} and h_{gM}, and β_0 was solved for such that the curve crossed the point (d_{gM}, h_{gM}). This height model has been used in numerous studies since (e.g. Kangas and Maltamo 2003) and recently, Korpela and Tokola (2006) found a slight bias when using it. Note that the more representative data of the present study clearly showed that the dependence between these variables was non-linear in the original scale, and linearisation between the parameters of Näslund's curve and stand characteristics required logarithmic trans-

formation on both sides (see fig. 2 in Siipilehto 2011). The BLUP estimated height curves in this study passed close to the point of means (D, H) or medians (d_{gM}, h_{gM}), while the knowledge of D_{dom} and H_{dom} was useful in characterising the bending of the curve (see Eerikäinen 2003). In contrast, stand density did not show any effect on the accuracy of the height curve parameters (see Fig. 2C).

Naturally, the models presented for stand characteristics can be used independently of the size distribution models given, in combination with any existing size distribution model. In that case, we may need the prediction for d_{gM}, h_{gM}, and G for young stands, or N for advanced stands, in order to be able to use the prediction model presented, for example, by Siipilehto et al. (2007) or Kangas and Maltamo (2000a). The height distribution model describes the complete distribution, whereas that of the dbh distribution describes the right tail of the truncated distribution until all the trees reach breast height. When a mean height of 4 m was reached, practically all of the crop trees had reached breast height. If dbh distribution is used for a stand in its early state, instead of the total number of stems, the number of stems above breast height should be used when one is calculating stand basal area. However, it is recommended to apply the height distribution model by Siipilehto (2009) up to a mean height of about 4 m because of its continuous feature.

The approach in this study shows one option for flexible description of stand structure. It is flexible in the sense that size distributions and height–dbh relationships can be predicted in 255 different ways, if the eight stand characteristics that were modelled in this study can be combined freely as explaining (calibrating) variables. The flexibility in combining the stand characteristics may become of great interest whenever new techniques (e.g. laser scanning) find new combinations of variables to be more efficient or accurate to assess as compared with the stand characteristics traditionally collected in the field.

According to this study, it would appear essential to choose the size distribution model such that the available data and the model applied are based on the same scale. Thus, if the basis lies in the common FMP stand characteristics, frequency distribution is superior to basal area

distribution for young stands and, vice versa, basal area distribution is superior to frequency distribution in advanced stands. In general, considerable improvements in the accuracy of the predicted distribution is possible when arithmetic mean and basal area-median diameters are simultaneously included in the prediction model. It is worth noting, however, that some combinations can also result in severely biased estimates of volume, as was shown when the W_G and SB_G models were applied for young stands with D_{dom} used in addition to the common variables N, D, and H. This was probably due to violation of the assumed linear dependence of the cross-model errors. Such violation was not noticed with respect to the W_N model for young stands or any model for advanced stands. Thus, this was indirect evidence that the assumed linearity between cross-model errors held in general.

The presented BLUP models are intended to replace the previous generation of models (e.g. Siipilehto 2006) in the MOTTI simulator. One methodological drawback was the fact that we did not manage to include the effect of repeated measurements in the model, most probably due to the large number of simultaneous models. However, according to Liu et al. (2004) the influence can be considered small. In addition, Siipilehto et al. (2007) showed the comparable results with SUR and mixed-effect SUR models when predicting SB distribution, whereas they were both superior to OLS and separately fitted mixed models.

The accuracy of the present model in terms of total stand volume and timber assortments should be carefully compared with the other models available in order to find the most reliable combinations of models. The linear prediction application presented provides a theoretically consistent basis for comparing alternative size distribution models with alternative combinations of known stand characteristics. Also, the option of utilising methods for compatibility in the known moment or percentile in order to improve the fit of the presented dbh distribution is worth exploring (see Merganič and Sterba 2006, Gobakken and Næsset 2004). Obviously, the optimal distribution model and the best set of variables changes with the stand's stage of development and with respect to the specific stand characteristic of interest (Kangas and Maltamo 2003).

Acknowledgements

References

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

Cajander, A.K. 1925. The theory of forest types. Acta Forestalia Fennica 29. 108 p.

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

DemoMELA. Available at: http://mela2.metla.fi/mela/demomela/index.htm. [Cited 18 Nov 2010].

Dubey, S.D. 1967. Some percentile estimators for Weibull parameters. Technometrics 9: 119–129.

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

— 2009. A multivariate linear mixed-effects model for 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.

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.

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

Gustavsen, H.G. 1980. Talousmetsien kasvupaikkaluokittelu valtapituuden avulla. Summary: Site index curves for conifer stands in Finland. Folia Forestalia 454. 31 p.

—, Roiko-Jokela, P. & Varmola, M. 1988. Kivennäis-maiden talousmetsien pysyvät (INKA ja TINKA) kokeet. Suunnitelmat, mittausmenetelmät ja aineis-tojen rakenteet. Metsäntutkimuslaitoksen tiedonan-toja 292. 212 pp. (in Finnish).

Haara, A. & Korhonen, K.T. 2004. Kuvioittaisen arvi-oinnin luotettavuus. Metsätieteen aikakauskirja 4/2004: 489–508. (In Finnish).

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

Huuskonen, S. & Hynynen, J. 2006. Timing and inten-sity of precommercial thinning and their effect on the first commercial thinning in Scots pine stands. Silva Fennica 40(4): 645–662.

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

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.

Johnson, N.L. 1949. Systems of frequency curves gen-erated by methods of translation. Biometrika 36: 149–176.

Kalliovirta, J. 2006. SIMO simulointi. SIMO-hankkeen loppuseminaari 2.11.2007. University of Helsinki. Available at: http://www.simo-project.org/down-load/seminars/2007/1102/SIMO_loppuseminaari.pdf.

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

— & Maltamo, M. 2000b. Performance of percentile based diameter distribution prediction and Weibull method in independent data sets. Silva Fennica 34(4): 381–398.

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

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

Kilkki, P., Maltamo, M., Mykkänen, R. & Päivinen,

R. 1989. Use of the Weibull function in estimat-ing the basal area dbh-distribution. Silva Fennica 23(4): 311–318.

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.

Kuvioittainen arviointi ja tietojen päivitys 1998. UPM-Kymmene Metsä. 31 p. (In Finnish).

Laasasenaho, J. 1982. Taper curve and volume func-tions for pine, spruce and birch. Communicationes Instituti Forestalis Fenniae 108. 74 p.

Lappi, J. 1991. Calibration of height and volume equa-tions with random parameters. Forest Science 37(3): 781–801.

— 1993. Metsäbiometrian menetelmiä. Silva Carelia 24. University of Joensuu. 182 p. (In Finnish).

— 2006. A multivariate, nonparametric stem-curve prediction method. Canadian Journal of Forest Research 36: 1017–1027.

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

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.

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

—, Puumalainen, J. & Päivinen, R. 1995. Compari-son 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.

—, Kangas, A., Uuttera, J., Torniainen, T. & Sara-mäki, J. 2000. Comparison of percentile based prediction method and the Weibull distribution in describing the diameter distribution of heterogene-ous Scots pine stands. Forest Ecology and Manage-ment 133: 263–274.

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

Mehtätalo, L. 2004. A longitudinal height–diameter

model for Norway spruce in Finland. Canadian Journal of Forest Research 34: 131–140.

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

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

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.

MOTTI software. Available at: http://www.metla.fi/metinfo/motti/index-en.htm. [Cited 18 Nov 2010].

Mykkänen, R. 1986. Weibull-funktion käyttö puuston läpimittajakauman estimoinnissa. Metsätalouden syventävien opintojen tutkielma. University of Joensuu. 80 p. (In Finnish).

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

Nissinen, J. 2002. Improving compatibility between prediction of basal area diameter distributions and assessments of young stands. Metsäsuunnittelun ja -ekonomian pro gradu. University of Joensuu. 48 p.

Nuutinen, T. 1986. Puustotunnusten väliset relaatiot. Metsätalouden suunnittelun syventävien opintojen tutkielma. University of Joensuu. (In Finnish).

Ojansuu, R. & Henttonen, H. 1983. Kuukauden keskilämpötilan, lämpösumman ja sademäärän paikallisten arvojen johtaminen ilmatieteen laitoksen mittaustiedoista. Summary: Estimation of local values of monthly mean temperature, effective temperature sum, and precipitation sum from the measurements made by the Finnish meteorological office. Silva Fennica 17(2): 143–160.

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.

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

Palahi, M., Pukkala, T. Blasco, E. & Trasobares, A.

2007. Comparison of Johnson's SB, Weibull and truncated Weibull functions for modelling the diameter distribution of forest stands in Catalonia (north-east of Spain). European Journal of Forest Research 126: 563–571.

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

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

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.

—, 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.

SAS/ETS User's Guide. Version 6, second edition. 1993. SAS Institute, Inc. Cary, NC, USA. 1022 p.

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

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. 1999. Improving the accuracy of predicted basal-area diameter distribution in advanced stands by determining stem number. Silva Fennica 33(4): 281–301.

— 2000. A comparison of two parameter prediction methods for stand structure in Finland. Silva Fennica 34(4): 331–349.

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

— 2009. Modelling stand structure in young Scots pine-dominated stands. Forest Ecology and Management 257: 223–232.

— 2011. Methods and applications for improving parameter prediction models for stand structures in Finland. Dissertationes Forestales 124.

—, Sarkkola, S. & Mehtätalo, L. 2007. Comparing regression estimation techniques when predicting diameter distributions of Scots pine on drained peatlands. Silva Fennica 41(1): 333–349.

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.

Solmun maastotyöopas. 1997. Metsätalouden kehittämiskeskus Tapio. 82 p.

Tahvanainen, O., Kaartinen, K., Pukkala, T. & Maltamo, M. 2007. Comparison of approaches to integrate energy wood estimation into the Finnish compartment inventory system. Silva Fennica 41(1): 123–135.

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

Valkonen, S. 1997. Viljelykuusikon alkukehityksen malli. Metsätieteen aikakauskirja 3: 321–347. (In Finnish).

Varmola, M. 1993. Viljelymänniköiden alkukehitystä kuvaava metsikkömalli. Summary: A stand model for early development of Scots pine cultures. Folia Forestalia 813. 43 p.

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

Vohland, M., Stoffels, J., Hau, C. & Schüler, G. 2007. Remote sensing techniques for forest parameter assessment: Multispectral classification and linear spectral mixture analysis. Silva Fennica 41(3): 441–456

Vuokila, Y. & Väliaho, H. 1980. Viljeltyjen havumetsiköiden kasvatusmallit. Summary: Growth and yield models for conifer cultures in Finland. Communicationes Instituti Forestalis Fenniae 99(2) 271 p.

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

Total of 73 references

Appendix

BLUP vs. OLS

The best linear unbiased predictor for the variable x_1 was given in Eq. 7 and the variance of the calibrated dependent variable in Eq. 8. As an example of the BLUP application we estimated the coefficient for the residuals of $\ln(G)$, $\ln(d_{gM}+1)$, and $\ln(h_{gM})$ in order to calibrate the predictions for Näslund's height curve parameters. Residuals in the equation are shown as rG, $rdgM$, and $rhgM$, and the expected value for the basic model was given as $E(y|T, DD, S)$. The variances and covariances needed were chosen from Tables 5 and 6. The initial error variances for β_0 and β_1 were 0.0615 and 0.0114, respectively (see Table 4).

$$\ln\left(\hat{\beta}_0\right) = E\left(y|T,DD,S\right) + \begin{bmatrix} -.0465 & -.0055 & -.0164 \end{bmatrix} \begin{bmatrix} .1878 & .0389 & .0532 \\ .0389 & .0402 & .0334 \\ .0532 & .0334 & .0397 \end{bmatrix}^{-1} \begin{bmatrix} rG \\ rdgM \\ rhgM \end{bmatrix}$$

Calculation of the above matrices resulted in the following coefficients:

$\ln(\beta_0) = E(y|T, DD, S) - 0.179\ rG + 0.606\ rdgM - 0.684\ rhgM$
Similarly, the calibrated model for $\ln(\beta_1)$ was as follows:

$\ln(\beta_1) = E(y|T, DD, S) + 0.085\ rG + 0.153\ rdgM - 0.558\ rhgM$

The error variance of the BLUP model for $\ln(\beta_0)$ is given below, with the inverse of the above variance–covariance matrix being calculated:

$$\text{var}\left(\hat{\beta}_0 - \beta_0\right) = 0.0641 - \begin{bmatrix} -.0465 & -.0055 & -.0164 \end{bmatrix} \begin{bmatrix} 8.812 & 4.339 & -15.459 \\ 4.339 & 84.779 & -77.140 \\ -15.459 & -77.140 & 110.804 \end{bmatrix} \begin{bmatrix} -.0465 \\ -.0055 \\ -.0164 \end{bmatrix}$$

Calibration yielded an error variance of 0.0478. The corresponding error variance for β_1 was 0.0062 for the calibrated model.

    The corresponding OLS model for the parameters was estimated by including $\ln(G)$, $\ln(d_{gM}+1)$, and $\ln(h_{gM})$ in the basic model for β_0 and β_1. The degrees of determination of the OLS models were 52% and 87%, and the error variances were just above those of the BLUP model, at 0.0483 and 0.0064 for β_0 and β_1, respectively. The OLS estimates were quite similar to the BLUP estimates, as shown below.

$\ln(\beta_0) = E(y|T, DD, S) - 0.136\ \ln(G) + 0.640\ \ln(d_{gM}+1) - 0.776\ \ln(h_{gM})$

$\ln(\beta_1) = E(y|T, DD, S) + 0.067\ \ln(G) + 0.191\ \ln(d_{gM}+1) - 0.553\ \ln(h_{gM})$