# Performance of Percentile Based Diameter Distribution Prediction and Weibull Method in Independent Data Sets

Annika Kangas and Matti Maltamo

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

Diameter distribution is used in most forest management planning packages for predicting stand volume, timber volume and stand growth. The prediction of diameter distribution can be based on parametric distribution functions, distribution-free parametric prediction methods or purely non-parametric methods. In the first case, the distribution is obtained by predicting the parameters of some probability density function. In a distribution-free percentile method, the diameters at certain percentiles of the distribution are predicted with models. In non-parametric methods, the predicted distribution is a linear combination of similar measured stands. In this study, the percentile based diameter distribution is compared to the results obtained with the Weibull method in four independent data sets. In the case of Scots pine, the other methods are also compared to k-nearest neighbour method. The comparison was made with respect to the accuracy of predicted stand volume, saw timber volume and number of stems. The predicted percentile and Weibull distributions were calibrated using number of stems measured from the stand. The information of minimum and maximum diameters were also used, for re-scaling the percentile based distribution or for parameter recovery of Weibull parameters. The accuracy of the predicted stand characteristics were also compared for calibrated distributions. The most reliable results were obtained using the percentile method with the model set including number of stems as a predictor. Calibration improved the results in most cases. However, using the minimum and maximum diameters for parameter recovery proved to be inefficient.

**Keywords** diameter distribution prediction, Weibull function, nearest neighbour method, distribution-free method, calibration estimation, stand structure
**Authors' addresses** *Kangas*, Finnish Forest Research Institute, Kannus Research Station, P.O. Box 44, FIN-69101 Kannus, Finland; *Maltamo*, Finnish Forest Research Institute, Joensuu Research Station, P.O. Box 68, FIN-80101 Joensuu, Finland
**Fax** +358 6 874 3201 **E-mail** annika.kangas@metla.fi
**Received** 8 November 1999 **Accepted** 3 October 2000

# 1 Introduction

Different theoretical distributions, for example beta, Weibull and Johnson's SB functions, have been widely used to describe tree stock mathematically (e.g. Loetsch et al. 1973, Bailey and Dell 1973, Hafley and Schreuder 1977). In Finland, researchers have mainly applied the Weibull distribution (e.g. Kilkki and Päivinen 1986, Kilkki et al. 1989, Maltamo 1997). Two main methods have been used to predict a parametric distribution function, namely the parameter prediction method and the parameter recovery method (Hyink 1980, Hyink and Moser 1983). The characteristics considered in parameter recovery can be either moments or percentiles of diameter distribution (see Knoebel and Burkhart 1991). It is also possible that some of the parameters are predicted and others are solved using a parameter recovery approach (Burkhart et al. 1982, Maltamo 1998).

Distribution-free methods do not rely on any predefined functional form. For example, Borders et al. (1987) developed a percentile based diameter distribution prediction method. In this method, the diameter distribution is obtained by first predicting diameters at certain percentiles and then interpolating the distribution function between these percentiles. Maltamo et al. (2000) used the percentile based approach to predict irregular stem frequency diameter distributions of stands in a natural state. Kangas and Maltamo (2000b) presented two sets of models for each species in Finland, to predict 12 percentiles of the basal area diameter distribution.

Another possibility to describe diameter distribution is to rely purely on non-parametric methods. For example, Haara et al. (1997) and Maltamo and Kangas (1998) predicted the basal area diameter distribution using nearest neighbour based approaches. The predicted diameter distribution was a weighted average of the distributions of the sample plots that are most similar to the stand of interest. These plots were selected from a database of previously measured stands.

In most applications, the predicted basal area diameter distribution is scaled to the measured basal area. However, there is no guarantee that the other stand characteristics obtained from the predicted diameter distribution correspond to the measured stand characteristics (except for characteristics that are used for parameter recovery of theoretical distributions or measured percentiles used in the percentile method). Kangas and Maltamo (2000a) calibrated stem frequency diameter distribution with basal area using an approach adopted from sampling theory, calibration estimation, originally presented by Deville and Särndal (1992). In their study, both number of stems and basal area were known but the predicted distribution did not give correct estimates for both of these stand characteristics. With calibration, correct estimates for all known stand characteristics were obtained.

In this study, the performance of the percentile based method is compared to the parameter prediction of the Weibull distribution, and, in the case of Scots pine, to nearest neighbour based method. The models are thoroughly tested in varying conditions, with available independent data sets. The calibration estimation method is applied to basal area diameter distribution when additional information is assumed to be available.

# 2 Material and Methods

## 2.1 Test Materials

The models were tested with several independently measured data sets. These include angle-count sample plot data from southern Finland, fixed area sample plot data from mineral soils across Finland, peatland stands and different kinds of mixed stands. The purpose was to examine the accuracy of the models in different situations. These data sets also include stands (mixed stands and stands in northern Finland) which have been found to be especially problematic for current diameter distribution models (e.g. Siipilehto 1999). Independent data sets were restricted standwise according to the following rules: the basal area of the considered tree species had to be over 1 $m^2$/ha and the number of stems over 50 per hectare; the number of measured trees in the sample plots had to be at least 10; the basal area median diameter had to be over 5 cm; and the range between both minimum and median and maximum and median diameter had to be over 2 cm.

A private forest enterprise collected the first test data (see e.g. Suutarla 1985). Six to twelve relascope (angle-count) sample plots were systematically located in each stand. Diameter at breast height (dbh) was recorded in 1 cm classes from each tree included in the sample plots using the basal area factor two ($m^2$/ha). The basal area diameter distributions were formed by summing the trees in the relascope sample plots in the stand. This data set is later referred to as ANGLE.

The other test data sets include the permanent sample plots (INKA and SINKA) measured by the Finnish Forest Research Institute (FFRI), originally for growth modelling purposes (Gustavsen et al. 1988, Penttilä and Honkanen 1986). The INKA sample plots were established on mineral soils across Finland. The SINKA sample plots are mainly located on ditched peatland (later called DITCHED) in northern Finland. A small data set including unmanaged peatland stands was also considered (later called UNMANAGED). The data sets include sample plots which consist of a cluster of three circular plots within a stand. When testing the diameter distribution prediction methods, these circular plots were combined. The diameter of all trees within a plot were measured to the nearest 0.1 cm. In the case of the INKA material, the test data was also divided by geographical area. Tests were made separately in southern Finland and in northern Finland in the provinces of Oulu and Lapland. These data sets include 208 (southern Finland), 189 (Oulu) and 124 (Lapland) sample plots for Scots pine, 195 (southern Finland), 109 (Oulu) and 29 (Lapland) sample plots for Norway spruce, and 77 (southern Finland), 56 (Oulu) and 32 (Lapland) sample plots for birch species

Stand age was not measured in all stands in the SINKA data. Therefore, a simple standwise age model was constructed from those parts of the SINKA data which included stand age. The model is

$$\text{Age} = 87.888 + 4.283 * d_{gM} - 0.078 * TS \qquad (1)$$

where $d_{gM}$ is basal area median diameter and $TS$ temperature sum, d.d. (+5 °C threshold). The degree of determination of the model is 0.30 and standard error 24.87. The standard error of the model is quite large, which is to be expected when the age on peatlands is considered. However, as in forestry practise the stands with missing values cannot be dismissed, they were not dismissed in this test. In such a case, the use of imputed values is reasonable. Consequently, the estimated model was used to predict the stand age on such peatland stands, where it was not measured.

Finally, the models were tested in mixed stands. These data sets include pine-spruce sample plots (MIXED 1), pine-birch sample plots (MIXED 2) and spruce-birch sample plots (MIXED 3). The mixed pine-spruce stand data was originally collected for studying the productivity of mixed forests stands (Pukkala et al. 1994). The sample plots from mixed pine-birch and spruce-birch stands were originally from the growth and yield studies of Mielikäinen (1980, 1985). These stands are located in southern and eastern Finland. In the case of pine-birch stands three circular plots were located within a stand. In the current study these plots were combined. In the case of spruce-birch stands one circular plot, and in the case of pine-spruce stands one rectangular sample plot was measured from each stand. In this data diameter at breast height was measured in 1 mm classes from all trees.

A summary of the sample plot type of different data sets, number of observations and the mean values of the most important stand characteristics are presented in Table 1.

## 2.2 The Compared Methods

In the percentile method, the diameters at certain percentiles of the distribution function are predicted with models. These diameters characterise an empirical distribution function. In this study, the percentiles are defined with respect to basal area in a stand. By interpolating between the predicted diameters, a basal area diameter distribution function is obtained. From this distribution, the basal area and number of stems in desired diameter classes can be calculated. Interpolation is done using Späth's rational spline, in order to obtain a monotone distribution (see Kangas and Maltamo 2000b for details). The models used to predict the diameters at different percentiles are those estimated by Kangas and Maltamo (2000b).

**Table 1.** The method of data collection, number of observations (*n*), and the mean value of basal area (*G*), number of stems (*N*), basal area median diameter ($d_{gM}$) and age in each data set by species.

| Data | Type of sample plots | Var | Scots pine | Norway spruce | Birch |
|---|---|---|---|---|---|
| ANGLE | Angle count | $n$ | 379 | 280 | 55 |
| | | $G$ | 11.1 | 13.8 | 5.6 |
| | | $N$ | 675 | 615 | 346 |
| | | $d_{gM}$ | 20.3 | 20.9 | 18.9 |
| | | Age | 69.7 | 79.6 | 81.2 |
| INKA | Fixed cluster | $n$ | 521 | 333 | 165 |
| | | $G$ | 12.1 | 12.6 | 4.7 |
| | | $N$ | 1193 | 881 | 502 |
| | | $d_{gM}$ | 15.6 | 17.3 | 15.4 |
| | | Age | 67.2 | 77.0 | 81.8 |
| DITCHED | Fixed cluster | $n$ | 361 | 121 | 325 |
| | | $G$ | 8.5 | 5.6 | 8.8 |
| | | $N$ | 958 | 605 | 1404 |
| | | $d_{gM}$ | 13.3 | 12.8 | 11.1 |
| | | Age | 84.7 | 90.3 | 80.4 |
| UNMANAGED | Fixed cluster | $n$ | 60 | 23 | 28 |
| | | $G$ | 4.0 | 5.7 | 3.7 |
| | | $N$ | 824 | 668 | 671 |
| | | $d_{gM}$ | 10.4 | 11.5 | 9.6 |
| | | Age | 88.4 | 104.9 | 89.6 |
| MIXED1 | Fixed rectangular | $n$ | 43 | 43 | - |
| | | $G$ | 14.9 | 10.9 | - |
| | | $N$ | 391 | 954 | - |
| | | $d_{gM}$ | 25.9 | 17.8 | - |
| | | Age | 68.5 | 68.5 | - |
| MIXED2 | Fixed cluster | $n$ | 92 | - | 92 |
| | | $G$ | 14.6 | - | 11.6 |
| | | $N$ | 385 | - | 392 |
| | | $d_{gM}$ | 25.0 | - | 21.9 |
| | | Age | 74.1 | - | 71.6 |
| MIXED3 | Fixed circle | $n$ | - | 64 | 64 |
| | | $G$ | - | 17.6 | 10.8 |
| | | $N$ | - | 1142 | 578 |
| | | $d_{gM}$ | - | 20.0 | 20.2 |
| | | Age | - | 59.9 | 61.7 |

They estimated two model sets: in the first set the number of stems was not included as a predictor (Percentiles 1) and in the second set it was (Percentiles 2).

The percentile based basal area diameter distribution models were compared to the Weibull distribution parameter prediction models presented by Maltamo (1997). These models have been estimated separately for Scots pine and Norway spruce from the same empirical data which was also used for estimating the models for diameters at different percentiles. The parameter models of Scots pine were used for birch species. Stand characteristics measured by tree species (e.g. basal area and basal area median diameter) were used as predictors in these parameter models. In the case of ditched peatland also the Weibull-models presented by Hökkä et al. (1991) were tested for Scots pine and birch distributions. These models have been estimated for Scots pine and birch dominated stands earlier using the same peatland data (SINKA).

The nearest neighbour based approach for predicting diameter distributions presented by Maltamo and Kangas (1998) was applied to the Scots pine stands of the INKA material, since there were no distance functions available for other tree species. The K-nearest neighbour method is based on distance-weighted nearest neighbour estimation, where k most similar stands are used for predicting the diameter distribution of the target stand. Before the nearest neighbour method can be applied, 1) the form of distance function to be used to find the most similar reference stands; 2) the number of nearest neighbours to be used; and 3) the form of the weight function for weighting the reference stands must be established. In this study the results of the study by Maltamo and Kangas (1998) were applied. The used form of the distance function was based on absolute differences between stand characteristics. Basal area median diameter and median height of the stand were the stand characteristics used when searching for nearest neighbours. The coefficients of these variables were 5 for basal area median diameter and 1 for median height. The number of nearest neighbours used was 14. The weighting parameter of the chosen nearest neighbours was 4.8.

In nearest neighbour approach, INKA-stands were used as target stands, for which diameter distributions are calculated. The INKA stands were chosen as target stands in order to get a large scale alternative, which could then be compared with other diameter distribution prediction methods. Both INKA and ANGLE data were used as databases from which the reference stands were chosen. When the stands of the INKA data were used both as target and as reference stands, the target stand was excluded from the reference stands. Using the ANGLE data as a reference database can be justified by the idea that there is common inventory data which can be used in many applications. Using the INKA material as a reference database can be justified by the idea that one should search for the most similar neighbours from stands which correspond to the target stands as closely as possible.

In the application stage, the estimate of the relative basal area in each diameter class $[d_1,d_2]$ was calculated from the cumulative distribution of diameters F as $F(d_2)–F(d_1)$. For the Weibull method, the cumulative distribution is obtained analytically. For the percentile method, the value of the empirical distribution F was obtained by interpolating the percentiles as a function of the predicted diameters with Späth's rational spline interpolation (Späth 1974, Lether 1984, see Maltamo et al. 2000). In the case of the k-nearest neighbour method, the original diameter classes of the chosen reference stands were used. In each case, the relative basal areas were scaled to the measured basal area in the stand to obtain an absolute value of basal area $b_k$ in each diameter class $k$. Finally, the frequency $f_k$ in each diameter class $k$ was calculated from the class basal area by dividing it with the basal area of the mean tree in this class:

$$f_k = \frac{b_k}{g_k} \qquad (2)$$

Volumes for each diameter class were calculated with Laasasenaho's models (1982), using diameter at breast height as a predictor. Saw timber volume was defined as the volume of trees larger than or equal to 16.5 cm.

### 2.3 Calibrating the Predicted Basal Area Diameter Distribution

The predicted basal area diameter distribution was calibrated with an approach presented by Deville and Särndal (1992). Kangas and Maltamo (2000a) used this approach to calibrate the predicted frequencies of a diameter distribution. In the present study, the calibration estimator was used to modify the predicted basal area $b_k$ of each diameter class $k$. The modification was carried out so that the modified class basal areas $w_k$ are as close as possible to the predicted basal areas $b_k$, while respecting the calibration equation(s). The calibration equation for number of stems is

$$\sum_{k=1}^{K} w_k / g_k = N \qquad (3)$$

where $N$ is the number of stems per hectare, $g_k$ is the basal area of the mean tree in diameter class $k$ (m$^2$/ha) and $K$ is the number of 1-cm diameter classes.

Further, it was required that the basal area, which was used in scaling the relative basal area, remained correct also after calibration, by using a constraint

$$\sum_{k=1}^{K} w_k = G \tag{4}$$

where $G$ is the stand basal area ($m^2$/ha). In addition, the basal area below the midpoint of $d_{gM}$ diameter class was set to half of the total basal area as

$$\sum_{k=1}^{K} w_k \gamma_k = G/2 \tag{5}$$

where

$$\gamma_k = \begin{cases} 1, d_k < d_{gM} \\ 1/2, d_k = d_{gM} \\ 0, d_k < d_{gM} \end{cases} \tag{6}$$

The distance measure used was the logarithmic distance (Deville and Särndal 1992)

$$\sum_{k=1}^{K} w_k \log(w_k / b_k) - w_k + b_k \tag{7}$$

Minimising this distance measure (7) while respecting the calibration equation(s) (3, 4, and 5) is a constrained non-linear optimisation problem. The problem was solved by reformulating it using Lagrange multipliers (see Deville and Särndal 1992 for details). The resulting group of non-linear equations was then solved using IMSL subroutines.

If the minimum and maximum diameters were assumed to be known, the distribution obtained with percentile method was re-scaled to the correct interval. The minimum (maximum) was set to the observed value, and the other diameters between minimum and mean diameter (mean and maximum diameter) were scaled according to

$$\hat{d}_i^* = d_{\min} + \left((d_{gM} - d_{\min}) / (d_{gM} - \hat{d}_0)\right)\left(\hat{d}_i - \hat{d}_0\right) \tag{8}$$

where $d_{\min}$ is the observed minimum diameter, $\hat{d}_0$ is the predicted minimum diameter, $\hat{d}_i$ is the predicted diameter at $i$th percentile and $\hat{d}_i^*$ is the re-scaled diameter. In the case of Weibull distribution, the minimum and maximum diameters were utilised using parameter recovery approach. The calibration was performed for ANGLE, INKA and DITCHED datasets.

## 2.4 Performance of the Methods

The performance of the different approaches (percentile, Weibull and k-nearest neighbour; both calibrated and uncalibrated), was examined by calculating the root mean square errors and biases of stand volume estimates ($m^3$/ha) obtained with these methods. The absolute root mean square error (RMSE) was calculated as

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(V_i - \hat{V}_i\right)^2}{n}} \tag{9}$$

where $n$ is the number of sample stands, $V_i$ is the true volume of stand $i$ and $\hat{V}_i$ is the volume of stand $i$ estimated from the predicted distribution. The relative RMSE of the volume estimate was calculated by dividing the absolute RMSE by the true mean volume $\overline{V}$ of the stands. The bias of the predictions was calculated as

$$\text{bias} = \frac{\sum_{i=1}^{n}\left(V_i - \hat{V}_i\right)}{n} \tag{10}$$

In addition to stand volume, the saw timber volume and number of stems were considered. The saw timber volume was not calculated for data sets including mostly pulpwood trees (DITCHED, UNMANAGED and Oulu region of INKA birches). Finally, it was calculated what proportion of the 25th and 75th percentiles had been correctly predicted into their diameter class in INKA dataset. This was done to examine the ability of different methods to predict stand structure characteristics. The 25th and 75th percentiles were chosen because they are not directly predicted when using the percentile based method.

# 3 Results

## 3.1 Uncalibrated Test Results

Some examples of predicted distributions of different tree species and models are presented in Figs. 1–3. All examples are from the INKA data set. In the case of Scots pine (Figs. 1a–f) the fit of the percentile based method which uses number of stems as a predictor (Percentiles 2) is good and it can also reproduce stand structure quite precisely. Both the Weibull method and especially the percentile based method (Percentiles 1), neither of which uses number of stems as a predictor, produce overestimates of small trees. The situation is quite similar in the case of Norway spruce

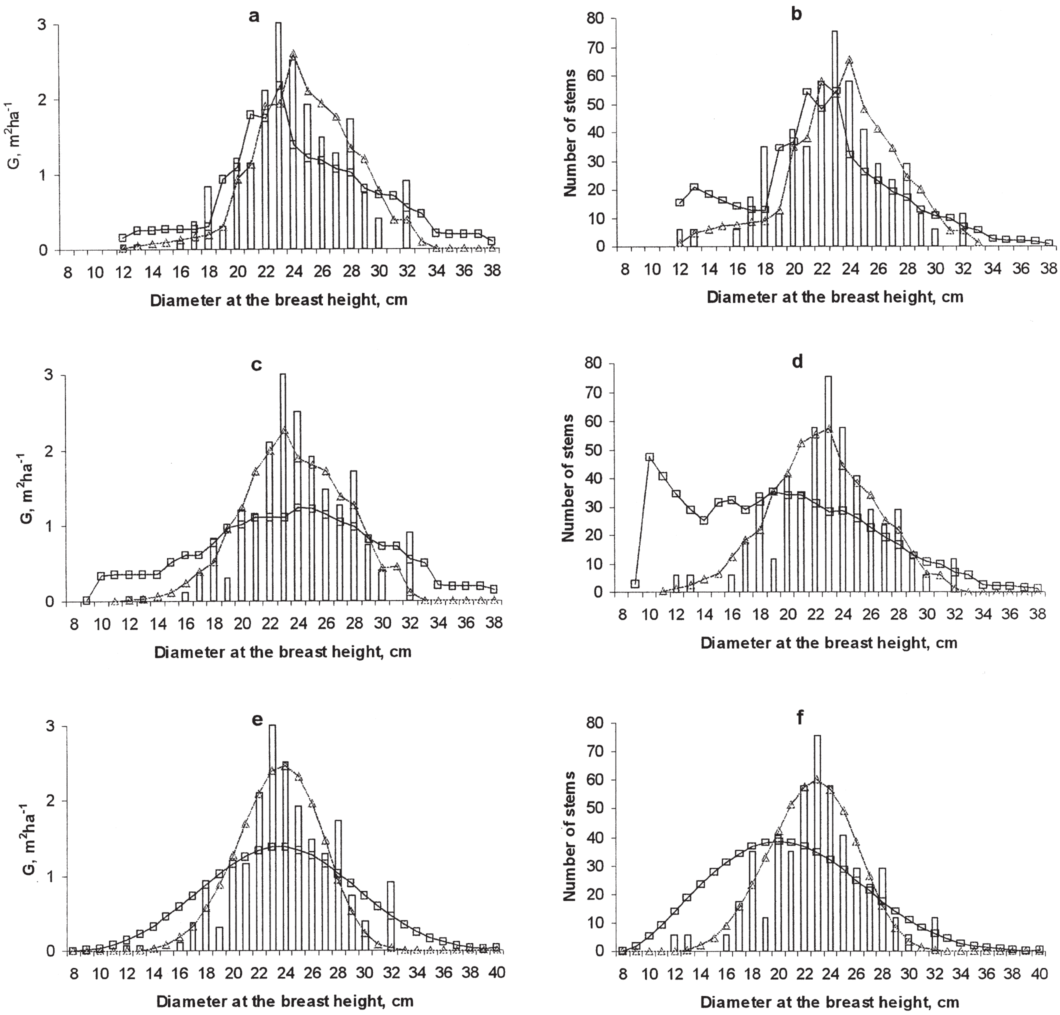


**Fig. 1.** An example of predicted and calibrated Scots pine diameter distributions. Subfigures a, c and e show basal area diameter distributions and subfigures b, d and f corresponding stem frequency diameter distributions. Subfigures a and b show uncalibrated (□) and calibrated (△) Percentiles 2 method, subfigures c and d show uncalibrated (□) and calibrated (△) Percentiles 1 method, and subfigures e and f uncalibrated (□) and calibrated (△) Weibull method. Subfigures a, c and e show the histogram of empirical basal area diameter distributions and subfigures b, d and f the histogram of empirical stem frequency diameter distributions.

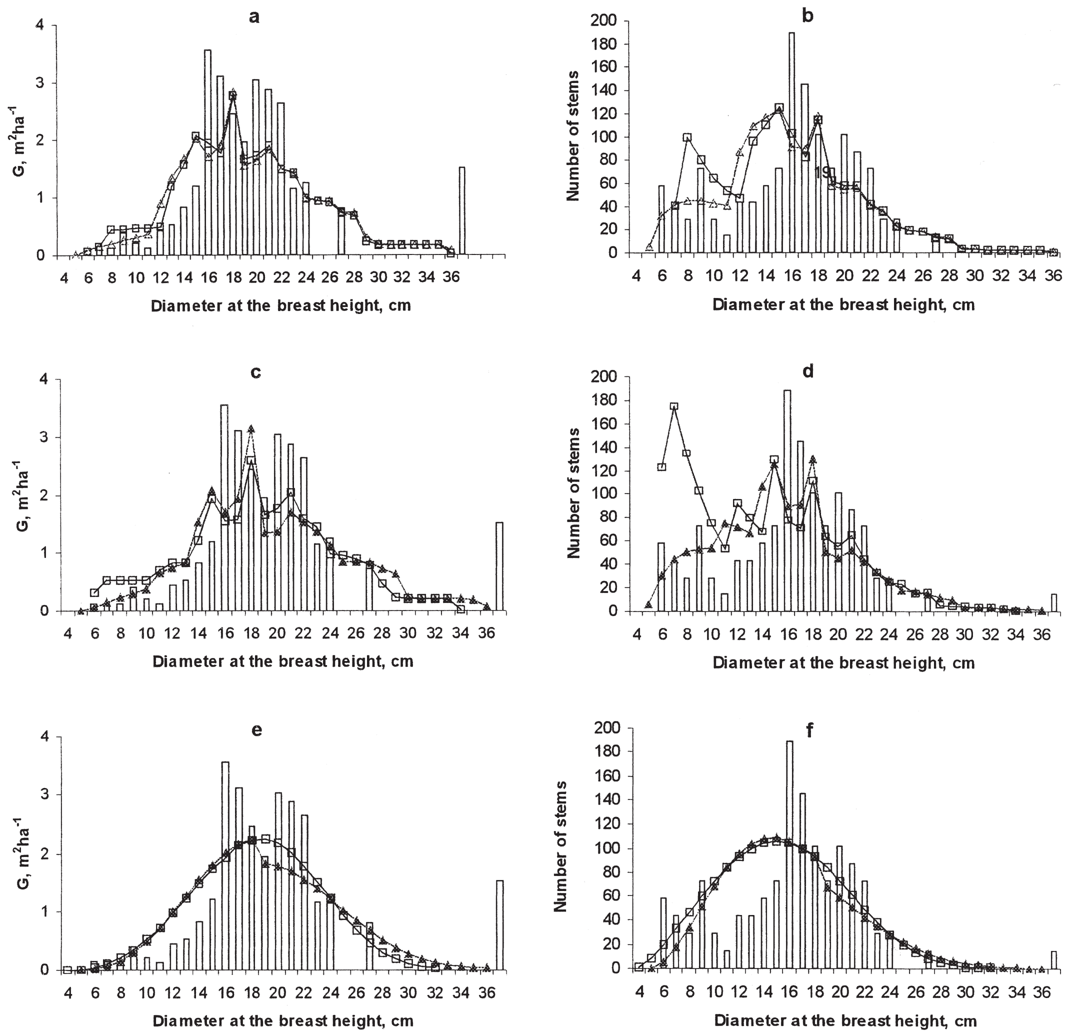


**Fig. 2.** An example of predicted and calibrated Norway spruce diameter distributions. Subfigures a, c and e show basal area diameter distributions and subfigures b, d and f corresponding stem frequency diameter distributions. Subfigures a and b show uncalibrated (□) and calibrated (△) Percentiles 2 method, subfigures c and d show uncalibrated (□) and calibrated (△) Percentiles 1 method, and subfigures e and f uncalibrated (□) and calibrated (△) Weibull method. Subfigures a, c and e show the histogram of empirical basal area diameter distributions and subfigures b, d and f the histogram of empirical stem frequency diameter distributions.

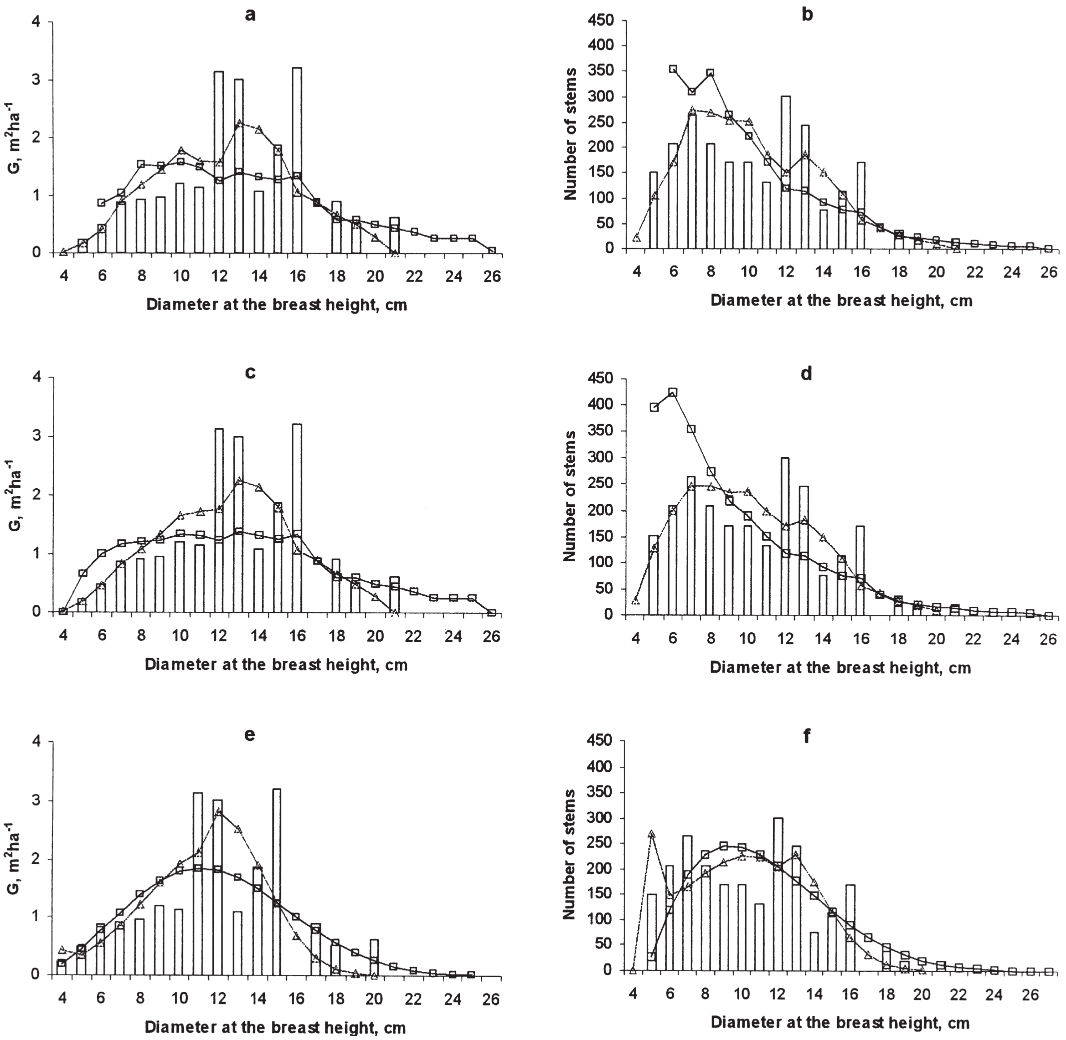


**Fig. 3.** An example of predicted and calibrated birch species diameter distributions. Subfigures a, c and e show basal area diameter distributions and subfigures b, d and f corresponding stem frequency diameter distributions. Subfigures a and b show uncalibrated (□) and calibrated (△) Percentiles 2 method, subfigures c and d show uncalibrated (□) and calibrated (△) Percentiles 1 method, and subfigures e and f uncalibrated (□) and calibrated (△) Weibull method. Subfigures a, c and e show the histogram of empirical basal area diameter distributions and subfigures b, d and f the histogram of empirical stem frequency diameter distributions.

(Figs. 2a–f). The fit of the Percentiles 2 is good. The Percentiles 1 method overestimates small trees and the Weibull method cannot describe the shape of the distribution properly. Finally, in the case of birch species (Figs. 3a–f) the fit of the Percentiles 2 is not as good as for other tree species. The situation is same for the Percentiles 1 method but the fit of the Weibull function is good.

The relative root mean square errors (RMSE) and the absolute biases of stand volume ($m^3$/ha), stand timber volume ($m^3$/ha) and number of stems (per hectare) in four independent test data sets are presented in Tables 2, 3 and 4. On average, the results for Scots pine were the most reliable, and the results for birch were the least reliable. With all species, the most reliable results were obtained with the Percentiles 2 method. When compared to the Percentiles 1 method, the Percentiles 2 method produced on average 23% smaller RMSE

**Table 2.** The results of the prediction of basal area diameter distribution of Scots pine in four independent test data sets (ANGLE, INKA, DITCHED and UNMANAGED).

| Prediction method | Material | RMSE% | Bias | RMSE% | Bias | RMSE% | Bias |
|---|---|---|---|---|---|---|---|
| | | Volume | | Saw timber | | Number of stems | |
| Percentiles 1 | ANGLE | 2.03 | 0.54 | 10.04 | 0.09 | 29.98 | –50.20 |
| Percentiles 2 | ANGLE | 1.15 | –0.08 | 8.56 | –1.47 | 5.93 | –1.21 |
| Weibull | ANGLE | 1.81 | 0.03 | 10.27 | –1.56 | 33.98 | –22.17 |
| Percentiles 1 | INKA | 2.21 | –0.04 | 14.75 | –1.66 | 26.80 | –8.43 |
| Percentiles 2 | INKA | 1.33 | –0.23 | 13.28 | –2.23 | 4.84 | 24.76 |
| Weibull | INKA | 2.38 | –0.59 | 15.49 | –2.68 | 27.06 | 62.06 |
| K-nn 1 | INKA | 2.04 | –0.10 | 13.52 | –0.45 | 23.20 | –23.20 |
| K-nn 2 | INKA | 2.73 | 0.51 | 15.09 | 1.09 | 25.39 | 0.53 |
| Percentiles 1 | DITCHED | 2.74 | 0.00 | 31.32 | –2.53 | 29.29 | –139.00 |
| Percentiles 2 | DITCHED | 2.20 | –0.51 | 29.50 | –2.96 | 4.69 | 32.49 |
| Weibull | DITCHED | 2.82 | –0.30 | 27.22 | –2.02 | 21.46 | –10.86 |
| Weibull2 | DITCHED | 2.67 | –0.09 | 25.38 | –1.30 | 17.40 | 45.01 |
| Percentiles 1 | UNMANAGED | 3.74 | –0.07 | 30.46 | –0.26 | 54.55 | –195.99 |
| Percentiles 2 | UNMANAGED | 2.40 | –0.10 | 28.60 | 0.00 | 6.29 | 35.37 |
| Weibull | UNMANAGED | 4.49 | –0.10 | 31.70 | 0.08 | 40.58 | –89.40 |

**Table 3.** The results of the prediction of basal area diameter distribution of Norway spruce four independent test data sets (ANGLE, INKA, DITCHED and UNMANAGED).

| Prediction method | Material | RMSE% | Bias | RMSE% | Bias | RMSE% | Bias |
|---|---|---|---|---|---|---|---|
| | | Volume | | Saw timber | | Number of stems | |
| Percentiles 1 | ANGLE | 3.60 | 2.42 | 10.39 | 3.53 | 41.75 | –136.92 |
| Percentiles 2 | ANGLE | 2.23 | –0.52 | 7.31 | –2.56 | 5.54 | 7.50 |
| Weibull | ANGLE | 2.83 | –0.03 | 8.89 | –2.88 | 36.51 | 26.96 |
| Percentiles 1 | INKA | 3.29 | 0.75 | 11.85 | –0.91 | 22.47 | –40.69 |
| Percentiles 2 | INKA | 2.15 | –0.34 | 9.71 | –2.51 | 6.88 | 38.14 |
| Weibull | INKA | 3.41 | –0.79 | 12.50 | –4.08 | 25.51 | 78.86 |
| Percentiles 1 | DITCHED | 4.86 | 0.10 | 26.27 | –1.58 | 31.46 | –113.19 |
| Percentiles 2 | DITCHED | 3.97 | –0.48 | 24.79 | –1.86 | 7.63 | 28.76 |
| Weibull | DITCHED | 4.87 | –0.49 | 26.16 | –2.35 | 29.50 | –24.59 |
| Percentiles 1 | UNMANAGED | 3.74 | –0.04 | 27.21 | –1.72 | 30.88 | –172.08 |
| Percentiles 2 | UNMANAGED | 3.97 | –0.64 | 27.51 | –1.68 | 5.30 | 24.16 |
| Weibull | UNMANAGED | 4.23 | –0.61 | 29.58 | –2.48 | 28.05 | –99.79 |

**Table 4.** The results of the prediction of basal area diameter distribution of birch species in four independent test data sets (ANGLE, INKA, DITCHED and UNMANAGED).

| Prediction method | Material | RMSE% | Bias | RMSE% | Bias | RMSE% | Bias |
|---|---|---|---|---|---|---|---|
| | | Volume | | Saw timber | | Number of stems | |
| Percentiles 1 | ANGLE | 4.07 | 1.13 | 19.32 | 0.38 | 59.80 | –123.49 |
| Percentiles 2 | ANGLE | 1.79 | 0.11 | 13.42 | –1.18 | 14.31 | –17.89 |
| Weibull | ANGLE | 3.08 | 0.27 | 12.14 | –1.02 | 45.63 | –53.94 |
| Percentiles 1 | INKA | 5.01 | 0.25 | 30.68 | –0.65 | 40.11 | –57.16 |
| Percentiles 2 | INKA | 2.88 | –0.23 | 23.89 | –1.32 | 11.85 | 1.44 |
| Weibull | INKA | 3.39 | –0.28 | 18.94 | –0.23 | 41.24 | –6.93 |
| Percentiles 1 | DITCHED | 4.08 | –1.03 | – | - | 30.26 | –251.18 |
| Percentiles 2 | DITCHED | 4.62 | –1.65 | - | - | 8.15 | –26.93 |
| Weibull | DITCHED | 3.17 | –0.40 | - | - | 39.69 | –192.20 |
| Weibull2 | DITCHED | 3.04 | –0.18 | - | - | 17.17 | –1.01 |
| Percentiles 1 | UNMANAGED | 4.55 | –0.50 | - | - | 35.65 | –166.26 |
| Percentiles 2 | UNMANAGED | 5.81 | –0.80 | - | - | 7.76 | –28.12 |
| Weibull | UNMANAGED | 3.18 | 0.02 | - | - | 40.46 | –0.89 |

for stand volume, 16% smaller RMSE for saw timber volume and as much as 79% smaller RMSE for number of stems.

The percentile based method not using number of stems as a predictor and the Weibull method produced nearly equally reliable results for pine and spruce. For birch, however, the results obtained with Weibull method were clearly more accurate than those obtained with Percentile 1 method. It is worth noting that the Weibull results for birch were calculated using models estimated for pine. The number of stems estimates were in many cases badly biased if number of stems was not used as a predictor.

The performance of k-nearest neighbour estimates with the Scots pines in the INKA material was close to that of the other methods not using number of stems as a predictor (Table 2). When the INKA stands were used as reference stands (K-nn 1), the accuracy of the predicted stand characteristics was slightly better than with the Percentiles 1 method and the Weibull method. When ANGLE stands were used as reference stands (K-nn 2), the accuracy of volume estimates was worse than with other methods. For saw timber volume and number of stems the results were quite similar to the Percentiles 1 and Weibull methods.

The results obtained using peatland data, SINKA, were generally poorer than the results obtained in mineral soils. This is especially true for the unmanaged peatland and for birch. However, the poorer results can partly be explained by the fact that the SINKA sample plots were measured from more northern areas than the modelling data. Weibull models based on the DITCHED material (Weibull2, Hökkä et al. 1991) produced only slightly more accurate results than general Weibull models. It is also notable that with respect to peatland the percentile based method which used number of stems as a predictor produced in some cases worse results than the one not using number of stems. This was especially true for birch species.

The effect of the geographical areas on the RMSE of the different forest variables is presented in Table 5 using the INKA data: the results were clearly poorer in the northern areas. Although the Percentiles 1 method and the Weibull method produced generally nearly equally reliable results, it seems that in extreme conditions, such as on peatland which has not been ditched (UNMANAGED) and most of northern Finland (INKA in Lapland), the results are usually more accurate with the percentile based method.

For the INKA data set, the proportion of 25th and 75th percentiles that were predicted to be in the correct diameter class are presented in Table 6. The results are presented only for INKA as

**Table 5.** The results of the prediction of basal area diameter distribution in three different geographical areas of the INKA data.

| Prediction method | Data | RMSE% Volume | | | RMSE% Saw timber | | | RMSE% Number of stems | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Pine | Spruce | Birch | Pine | Spruce | Birch | Pine | Spruce | Birch |
| Percentiles 1 | Southern | 1.71 | 3.02 | 4.67 | 12.87 | 10.55 | 22.63 | 20.88 | 20.55 | 41.03 |
| Percentiles 2 | Southern | 0.90 | 1.70 | 2.07 | 11.95 | 7.93 | 14.79 | 3.52 | 6.62 | 13.03 |
| Weibull | Southern | 1.53 | 3.05 | 2.20 | 13.34 | 10.80 | 11.33 | 22.13 | 24.66 | 42.49 |
| Percentiles 1 | Oulu | 2.15 | 3.80 | 5.15 | 14.73 | 15.08 | - | 32.38 | 26.68 | 39.46 |
| Percentiles 2 | Oulu | 1.41 | 3.40 | 4.79 | 13.42 | 15.86 | - | 4.50 | 5.60 | 7.31 |
| Weibull | Oulu | 2.29 | 4.21 | 5.56 | 14.31 | 17.20 | - | 29.91 | 25.99 | 38.66 |
| Percentiles 1 | Lappland | 3.60 | 3.53 | 4.61 | 20.05 | 12.29 | 24.52 | 27.05 | 23.54 | 29.28 |
| Percentiles 2 | Lappland | 2.18 | 2.73 | 3.65 | 16.53 | 12.03 | 24.28 | 6.36 | 9.83 | 12.02 |
| Weibull | Lappland | 4.46 | 3.97 | 5.33 | 23.81 | 17.09 | 25.59 | 31.85 | 28.26 | 33.72 |

**Table 6.** Relative proportion of diameter quartiles of number of stems and basal area which are predicted to correct diameter class. Results are presented in INKA test data set with (C) and without (U) calibration.

| Prediction method | Tree species | Number of stems 25 % | | Number of stems 75 % | | Basal area 25 % | | Basal area 75 % | |
|---|---|---|---|---|---|---|---|---|---|
| | | U | C | U | C | U | C | U | C |
| Percentiles 1 | Scots pine | 41.3 | 84.8 | 73.1 | 91.6 | 70.2 | 85.2 | 54.3 | 69.1 |
| Percentiles 2 | Scots pine | 77.7 | 85.0 | 86.6 | 92.3 | 82.3 | 86.0 | 54.9 | 68.9 |
| Weibull | Scots pine | 38.0 | 53.9 | 71.6 | 76.8 | 72.0 | 71.0 | 54.7 | 58.9 |
| Percentiles 1 | Norway spruce | 31.8 | 73.9 | 40.8 | 77.8 | 48.9 | 77.5 | 40.5 | 54.1 |
| Percentiles 2 | Norway spruce | 68.2 | 72.1 | 76.6 | 77.5 | 75.7 | 74.5 | 38.1 | 52.6 |
| Weibull | Norway spruce | 27.6 | 43.2 | 48.9 | 58.6 | 63.7 | 61.9 | 51.1 | 40.5 |
| Percentiles 1 | Birch species | 25.5 | 72.7 | 25.5 | 79.4 | 30.9 | 73.3 | 23.6 | 55.2 |
| Percentiles 2 | Birch species | 55.8 | 70.3 | 71.5 | 79.4 | 49.1 | 71.5 | 23.0 | 54.5 |
| Weibull | Birch species | 26.1 | 46.1 | 61.8 | 69.1 | 56.4 | 61.8 | 43.0 | 41.2 |

the differences between the data sets were not great. However, there were notable differences between the model sets and different percentiles. The Percentiles 2 method produced the percentiles considered most accurately. The other methods produced better results only for the 75th percentile of basal area in the case of spruce and for both basal area percentiles in the case of birch.

The 75th percentile with respect to number of stems and the 25th percentile with respect to basal area could be predicted correctly more often than the other percentiles. This can be explained by the fact that a small deviation from the true distribution in the smallest diameter classes can change the number of stems considerably, but not the basal area. On the other hand, a deviation from the true distribution in the largest diameter classes may affect the basal area considerably, but may have a negligible effect on the number of stems. One reason for the good results in predicting the 75th percentile of number of stems is also that this diameter is usually very close to the basal area median diameter, which was assumed to have been assessed and thus known.

In the mixed data sets (Table 7) the results for Scots pine were very good with all methods. However, the results for Norway spruce and birch were generally much poorer. This may be partly due to the nature of the species: Scots pine does

**Table 7.** The results of the prediction of basal area diameter distribution in data of mixed stands of Scots pine, Norway spruce and birch species (MIXED1, MIXED2 and MIXED3).

| Prediction method | Data | RMSE% Volume | | | RMSE% Saw timber | | | RMSE% Number of stems | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Pine | Spruce | Birch | Pine | Spruce | Birch | Pine | Spruce | Birch |
| Percentiles 1 | MIXED1 | 1.28 | 4.04 | - | 4.00 | 12.48 | - | 13.00 | 21.55 | - |
| Percentiles 2 | MIXED1 | 0.46 | 3.60 | - | 1.76 | 10.84 | - | 3.34 | 11.23 | - |
| Weibull | MIXED1 | 1.08 | 4.61 | - | 3.03 | 13.78 | - | 19.71 | 30.10 | - |
| Percentiles 1 | MIXED2 | 1.57 | - | 4.36 | 5.73 | - | 14.76 | 14.37 | - | 44.87 |
| Percentiles 2 | MIXED2 | 0.54 | - | 1.34 | 3.73 | - | 7.61 | 2.63 | - | 8.17 |
| Weibull | MIXED2 | 1.43 | - | 2.51 | 4.97 | - | 9.17 | 10.92 | - | 19.57 |
| Percentiles 1 | MIXED3 | - | 2.96 | 4.83 | - | 10.71 | 21.12 | - | 26.27 | 40.26 |
| Percentiles 2 | MIXED3 | - | 1.55 | 1.75 | - | 7.51 | 12.93 | - | 7.42 | 9.19 |
| Weibull | MIXED3 | - | 2.25 | 2.57 | - | 8.94 | 13.08 | - | 28.83 | 26.92 |

not form understorey in the same way as Norway spruce does (e.g. Siren 1955). It has also been found that the natural forest structure is still retained in many managed Norway spruce stands (Esseen et al. 1997). The distribution of nearly even-aged dominant trees is easier to predict than that of the suppressed trees of uneven-aged stands. The data of Norway spruce and especially pubescent birch may contain more dominated or suppressed trees than that of Scots pine. The different tree storeys were not separated in this study. Excluding the understorey trees would probably have improved the results for Norway spruce.

## 3.2 Calibration Results

Calibration was performed for the percentile based methods and for the Weibull method. K-nearest neighbour estimates or alternative Weibull models on peatland were not calibrated, since these methods were only used in one case. Calibration was carried out in all different data sets except UNMANAGED and different mixed stands. In each data set with each method, the distribution was calibrated using number of stems (Table 8). Also basal area and $d_{gM}$ were used in calibration, to ensure correct values for these variables also after calibration. Observed minimum and maximum diameters were used to re-scale the predicted distribution to the correct interval in the case of percentile based method (Table 9). If minimum and maximum diameters were assumed to be known, these were used to recover the Weibull parameters instead of using prediction models (Table 9).

Calibrating the basal area diameter distribution with number of stems proved to be problematic (Tables 8 and 9). Feasible solutions to the calibration equations were not found in all cases using logarithmic distance function. With the quadratic distance function (see Kangas and Maltamo 2000a), a feasible solution was found in all cases, but the obtained basal area diameter distribution included negative frequencies, which are not acceptable. Thus, if feasible solution was not found in a stand using the logarithmic distance function, the original predicted distribution was used for this stand.

In addition, the calibrated results were in some cases worse than without calibration, especially in the case of birch. Usually, calibrating the distribution obtained without using number of stems as a predictor (Percentiles 1) produced poorer results than could have been obtained by using the number of stems as a predictor (Percentiles 2). Calibrating a distribution obtained with Percentiles 2 method with respect to number of stems seems to be a more reasonable alternative. Number of stems as such is not a good predictor for stand volume, it needs to be related to some information concerning the size of the trees; in the Percentiles 2 models it was related to the basal area of the stands.

In average, calibrating with number of stems reduced the RMSE of stand volume by 19% in

**Table 8.** The results of the calibration of basal area diameter distribution with stem number in three independent data sets ANGLE, INKA and DITCHED for Scots pine, Norway spruce, and birch.

| Prediction method | Data | RMSE% Volume | | | RMSE% Saw timber | | | RMSE% Number of stems | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Pine | Spruce | Birch | Pine | Spruce | Birch | Pine | Spruce | Birch |
| Percentiles 1 | ANGLE | 1.19 | 2.19 | 2.58 | 8.81 | 7.07 | 16.64 | 23.48 | 19.43 | - |
| Percentiles 2 | ANGLE | 1.13 | 2.22 | 2.03 | 8.50 | 7.37 | 15.48 | - | - | - |
| Weibull | ANGLE | 1.39 | 2.18 | 1.87 | 9.58 | 7.81 | 11.10 | 23.87 | 30.29 | 6.77 |
| Percentiles 1 | INKA | 1.82 | 2.24 | 3.30 | 14.09 | 9.61 | 24.79 | 3.80 | - | 12.13 |
| Percentiles 2 | INKA | 1.34 | 2.16 | 3.01 | 13.20 | 9.60 | 24.96 | - | - | - |
| Weibull | INKA | 1.77 | 3.14 | 2.90 | 13.81 | 11.84 | 18.43 | 10.07 | 12.26 | 13.92 |
| Percentiles 1 | DITCHED | 2.19 | 3.76 | 5.46 | 30.26 | 25.81 | - | 0.43 | - | - |
| Percentiles 2 | DITCHED | 1.96 | 3.66 | 4.83 | 28.64 | 24.60 | - | - | - | - |
| Weibull | DITCHED | 1.84 | 4.16 | 2.82 | 24.66 | 25.69 | - | 5.96 | 15.16 | 5.14 |

**Table 9.** The results of the calibration of basal area diameter distribution with stem number in three independent data sets ANGLE, INKA and DITCHED for Scots pine, Norway spruce, and birch. The distributions are re-scaled according to observed minimum and maximum diameters.

| Prediction method | Data | RMSE% Volume | | | RMSE% Saw timber | | | RMSE% Number of stems | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Pine | Spruce | Birch | Pine | Spruce | Birch | Pine | Spruce | Birch |
| Percentiles 1 | ANGLE | 1.21 | 1.67 | 1.92 | 7.73 | 4.87 | 10.67 | - | - | - |
| Percentiles 2 | ANGLE | 1.12 | 1.92 | 1.93 | 7.78 | 5.46 | 10.66 | - | - | - |
| Weibull | ANGLE | 2.32 | 2.11 | 2.55 | 10.44 | 6.13 | 11.71 | 27.48 | 25.19 | 11.41 |
| Percentiles 1 | INKA | 1.15 | 2.20 | 1.99 | 9.85 | 6.65 | 10.88 | - | - | - |
| Percentiles 2 | INKA | 1.09 | 2.42 | 2.12 | 10.08 | 6.71 | 11.66 | - | - | - |
| Weibull | INKA | 2.09 | 3.73 | 3.26 | 14.93 | 13.48 | 18.38 | 18.09 | 19.14 | 19.06 |
| Percentiles 1 | DITCHED | 1.48 | 3.91 | 1.96 | 17.98 | 18.71 | - | - | - | - |
| Percentiles 2 | DITCHED | 1.40 | 3.49 | 2.02 | 18.38 | 18.02 | - | - | - | - |
| Weibull | DITCHED | 3.01 | 5.69 | 3.20 | 28.77 | 28.33 | - | 13.53 | 26.53 | 12.98 |

the case of pine, 16% in the case of spruce and 9 % in the case of birch. The RMSE of saw timber volumes did not reduce as much, in average by 6% for pine, 8% for spruce and 4% for birch. The RMSE of number of stems, on the other hand, could be reduced in average by 80%.

If the minimum diameter and maximum diameter were also assumed to be known, the results concerning the saw timber volume and birch stand volume could be greatly improved (Table 9). The RMSE of saw timber volume could be reduced in average by 18% in the case of pine, 25% in the case of spruce and 31% in the case of birch. The RMSE of stand volume could be reduced in average by 29% in the case of birch. The re-scaling of the distribution worked very well for the percentile methods. Instead, the parameter recovery of Weibull parameters using minimum and maximum diameters did not seem very efficient. Especially using the maximum diameter in parameter recovery is problematic. For Weibull method, other calibrating variables would probably be better.

Examples of the calibrated distributions are also presented in Figs. 1–3. With all tree species, the calibration worked quite well. In each case, calibration removed both large and small trees from the predicted distribution. The skewness and kurtosis of the predicted distributions were also modified in some cases. In the case of Scots pine (Figs. 1a–f), calibration produced outstanding results, especially Percentiles 1 and Weibull predictions were improved considerably. In the case of Norway spruce (Figs. 2a–f), for Weibull method, the effect of calibration is minor in this stand. In the case of birch stand (Figs. 3a–f) the Weibull prediction is modified to slightly bimodal distribution

The calibration also improved the proportion of correctly predicted percentiles, especially for the Percentiles 1 method (Table 6). In most cases, the calibrated Percentiles 1 method provided as accurate or even more accurate results than the Percentiles 2 method, which uses number of stems as a predictor. In any case, measuring the number of stems is worthwhile if an accurate estimate of the stand structure is of interest.

# 4 Discussion

The prediction of diameter distribution worked fairly well in most cases, the largest RMSEs of stand characteristics occurring in the northernmost Finland and in the peatland. However, the results were quite accurate also in this data; the biggest relative RMSEs were still less than 6% for stand volume. From the stand characteristics considered, stand volume could be predicted most accurately and number of stems least accurately. There were 5%–7% RMSE in the estimate of number of stems even if it was assumed to be known. With regard to the tree species, the Scots pine results were the most accurate, and the birch results the least accurate. The results obtained were quite similar to those obtained in the modelling data set (Kangas and Maltamo 2000b), which indicates that the models are reliable in varying conditions.

The accuracy of the percentile based method using number of stems as a predictor (Percentiles 2) was superior compared to other existing models. The accuracy of the percentile based method not using number of stems as a predictor (Percentiles 1) and the accuracy of the Weibull method were close to each other. This is natural because these models utilise the same amount of information. However, their superiority differed in different data sets. This same phenomenon has been established also in earlier studies (e.g. Maltamo 1997, Siipilehto 1999). However, it seems that percentile based methods are more flexible with respect to varying stand structure and therefore produce more accurate results in extreme conditions.

In this study number of stems per hectare was obtained either by counting all the trees in a plot (circular and rectangular sample plots) or by computing from treewise basal area (angle-count sample plots). A number of stems estimate, especially one obtained from an angle-count sampling plot, may not be precise. This may also explain the great biases and errors which were obtained in the predictions of number of stems with data based on angle-counts. In addition, if number of stems is measured in forest inventories all trees may not be considered. For example in advanced stands, the smallest trees may not be measured. Then the number of stems is smaller than that which was used in this study. All in all, it is obvious that the measurement and use of number of stems require further investigation. Especially, the effects of the measurement errors of number of stems on the prediction of the distribution and on its calibration were not considered in this study and remain to be studied.

K-nearest neighbour estimates were calculated for Scots pine stands of the INKA material. These estimates were based on distance functions, number of nearest neighbours and value of weighting parameter from the results of the study by Maltamo and Kangas (1998). If these characteristics had been optimised from material used in the current study the results would have been even more accurate. The use of number of stems in distance functions would have further improved the results. Presumably, the accuracy of non-parametric estimates would then have been close to the accuracy of the percentile based models which include number of stems as a predictor, since the amount of information seems to be more important than the prediction method. Also other non-parametric methods, such as most similar neighbour (MSN) approach (Moeur and Stage 1995), can be used in description of diameter distribution.

With respect to number of stems, the results can be compared to those of Siipilehto (1999). When number of stems was not used as a predictor the accuracy of percentile based method was worse than that of the corresponding Johnson's SB models of Siipilehto (1999). If number of stems was used as a predictor, the percentile based method produced quite similar estimates except for the birch. For stand volume and timber volume, the results cannot be directly compared, because of different usage of height models: the height estimate can have a profound effect on the accuracy of volume estimates. Siipilehto (1999) used a height model, whereas in this study height model was not used.

Calibration seems to be worthwhile especially with respect to stand structure. However, calibration with number of stems did not always improve the accuracy of the results. The diameter distribution in some stands may be so heterogeneous that more information, for example about median diameters, would be required to make the calibration worthwhile. Re-scaling the predicted distribution with extreme diameters seemed to be worthwhile, especially with respect to saw timber volume. However, using these extreme diameters for parameter recovery in the case of Weibull method did not work so well.

The calibration of predicted distributions using additional information measured from the corresponding stand seems to be a very promising method. The results of this study and those of Kangas and Maltamo (2000a) indicate that good estimates are obtained in most cases. The calibration can be carried out for any diameter distribution prediction method, including non-parametric estimates. The calibration possibilities are not restricted purely to number of stems. If stem frequency diameter distributions are used, the calibration using basal area is very effective (Kangas and Maltamo 2000a). In irregular stands, assessing other mean diameter characteristics than basal area median diameter, e.g. the arithmetic mean diameter, may help. In tropical countries, other stand mean characteristics, such as dominant diameter, may be used. Also the basal area or number of stems in certain diameter classes, e.g. for trees bigger than any predefined value, could be useful in calibration.

In the case studied, feasible solutions for calibration with number of stems were not found if the number of stems estimate obtained from the original distribution was very poor. In those cases, the original predicted distribution was used. However, if the minimum and maximum diameters were assumed known, the feasible solution was found more often. Feasible solutions were always found with a quadratic distance function, but they included negative frequencies. Consequently, different forms of distance function may be useful in future studies (see e.g. Théberge 1999).

More attention should also be paid to prediction of tree heights. In this study, the height information was not utilised, since the tree volumes were predicted using a volume model with diameter as the only predictor. However, the effect of errors in tree heights can be important, especially when the volumes for the different timber sortiments are needed. In the future, the whole procedure from data acquisition to the calculation of the stand results needs to be considered.

## Acknowledgements

The study material was provided by the Finnish Forest Research Institute, University of Joensuu, Faculty of Forestry and UPM. The work was carried out in the project "Integrating landscape ecology into forest planning", which is part of the Finnish Biodiversity Research Programme FIBRE, funded by the Finnish Academy. The authors would like to thank Prof. Timo Pukkala, Doc. Juha Lappi and Lic.For. Jouni Siipilehto for their valuable comments on the earlier drafts of the manuscript.

## References

Bailey, R.L. & Dell, T.R. 1973. Quantifying diameter distribution with the Weibull-function. Forest Science 19: 97–104.

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

Burkhart, H.E., Knoebel, B.R. & Beck, D.E. 1982. A growth and yield model for thinned stands of yellow-poplar. Paper presented at IUFRO subject group S4.01 meeting, Oct. 4–8, 1982, Vienna, Austria.

Deville, J-C. & Särndal, C-E. 1992. Calibration estimators in survey sampling. Journal of the American Statistical Association 87: 376–382.

Esseen, P.-A., Ehnström, B., Erickson, L. & Sjöberg, K. 1997. Boreal forests. Ecological Bulletin 46: 16–47.

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., 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 even-aged stands. Canadian Journal of Forest Research 7: 481–487.

Hökkä, H., Piiroinen, M.-L & Penttilä, T. 1991. The estimation of basal area-dbh distribution using the Weibull-function for drained pine- and birch dominated and mixed peatland stands in north Finland. Folia Forestalia 781. 22 p.

Hyink, D.M. 1980. Diameter distribution approaches to growth and yield modeling. In: Brown, K.M. & Clarke, F.D. (eds.). Forecasting stand dynamics. School of Forestry, Lakehead University, Thunder Bay, Ontario. p. 138–163.

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

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

— & Maltamo, M. 2000b. Percentile based basal area diameter distribution models for Scots pine, Norway spruce and birch species. Submitted manuscript.

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

— , Maltamo, M., Mykkänen, R. & Päivinen, R. 1989. Use of the Weibull function in estimating the basal-area diameter distribution. Silva Fennica 23: 311–318.

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

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

Loetsch, F., Zöhrer, F. & Haller. K.E. 1973. Forest Inventory 2. BLV. Verlagsgesellschaft, München. 469 p.

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

— 1998. Basal area diameter distribution in estimating the quantity and structure of growing stock. Joensuun yliopisto. Metsätieteellisen tiedekunnan tiedonantoja 67. 43 p.

— & Kangas, A. 1998. Methods based on k-nearest neighbor regression in estimation of basal area diameter distribution. Canadian Journal of Forest Research 28: 1107–1115.

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

Mielikäinen, K. 1980. Structure and development of mixed pine and birch stands. Communicationes Instituti Forestalis Fenniae 99. 82 p.

— 1985. Effect of an admixture of birch on the structure and development of Norway spruce stands. Communicationes Instituti Forestalis Fenniae 133. 79 p.

Moeur, M. & Stage, A.R. 1995. Most similar neighbor: an improved sampling inference procedure for natural resource planning. forest science 41: 377–359.

Penttilä, T. & Honkanen, M. 1986. Suometsän pysyvien kasvukoealojen (SINKA) maastotyöohjeet. Metsäntutkimuslaitoksen tiedonantoja 226. 98 p.

Pukkala, T., Vettenranta, J., Kolström, T. & Miina, J. 1994. Productivity of mixed stands of Pinus sylvestris and Picea abies. Scandinavian Journal of Forest Research 9: 143–153.

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

Sirén, G. 1955. The development of spruce forest on raw humus sites in northern Finland and its ecology. Acta Forestalia Fennica 62. 363 p.

Suutarla, T. 1985. Kuvioittaisen inventoinnin päivitys ja sen luotettavuus. Metsänarvioimistieteen pro gradu -tutkielma. Helsingin yliopisto. 49 p.

Théberge, A. 1999. Extensions of calibration estimators in survey sampling. Journal of the American Statistical Association 94: 635–644.

*Total of 31 references*