Anticipating the Variance of Predicted Stand Volume and Timber Assortments with Respect to Stand Characteristics and Field Measurements

Annika Kangas and Matti Maltamo

Several models and/or several variable combinations could be used to predict the diameter distribution of a stand. Typically, a fixed model and a fixed variable combination is used in all conditions. The calibration procedure, however, makes it possible to choose the measurement combination from among many possibilities, although the model used is fixed. In this study, the usefulness of utilizing additional stand characteristics for calibrating the predicted diameter distribution is examined. Nine measurement strategies were tested in predicting the total stand volume, sawlog volume and pulpwood volume. The observed errors of these variables under each strategy were modeled as a function of basal area, basal area median diameter and number of stems. The models were estimated in three steps. First, an Ordinary Least Squares (OLS) model was fitted to the observed errors. Then, a variance function was estimated using the OLS residuals. Finally, a weighted Seemingly Unrelated Regression (SUR) analysis was used to model the observed errors, using the estimated variance functions as weights. The estimated models can be used to anticipate the precision and accuracy of predicted volume characteristics for each stand with different variable combinations and, consequently, to choose the best measurement combination in different stands.

Authors´ addresses *Kangas*, University of Helsinki, Dept. of Forest Resources Management, P.O. Box 27, 00014 University of Helsinki, Finland; *Maltamo*, University of Joensuu, Faculty of Forestry, P.O. Box 111, 80101 Joensuu, Finland
Email annika.kangas@helsinki.fi, matti.maltamo@forest.joensuu.fi

1 Introduction

In compartmentwise inventory in Finland, the variables of primary interest, namely current stand volume and timber assortments, are predicted in two steps. First, the basal area diameter distribution in a stand is predicted based on assessed variables. Second, the stand volume and timber assortments are predicted using treewise height and volume models or taper curve functions, using diameters sampled from the predicted diameter distribution. The future development of the stand is predicted in a similar fashion, based on the same information.

Usually, the diameter distribution is predicted using one fixed model and a few fixed basic stand characteristics, which are assessed in the field. Yet, there are many different measurement combinations that could be used in predicting the distribution, as well as several models or modeling approaches. It can be presumed that in different conditions different measurement combinations or modeling approaches would be optimal. Nevertheless, estimating prediction models for all possible combinations and including them into a forest information system would be tedious. This problem can be overcome by using calibration estimation (see Deville and Särndal 1992, Kangas and Maltamo 2000a,c). In calibration estimation, the distribution may be predicted with the aid of basic measurements using a fixed model, and calibrated afterwards if additional information is available. This makes it possible, in principle, to use an unlimited number of different measurement combinations in predicting the diameter distribution.

Because volumes and timber assortments are predicted using predicted diameter distributions, their accuracy is difficult to assess analytically. The existing estimates of uncertainty are usually empirical estimates, which are averages for different conditions (e.g. Kangas and Maltamo 2000c). Such studies provide little information on how the different approaches perform in different conditions. By calculating the uncertainty estimates separately for different conditions, e.g. for different geographical areas, more information on the performance of the approach used can be obtained (Kangas and Maltamo 2000c). An anticipated estimate of the prediction error for each stand with different modeling approaches or dependent variables would be useful in order to choose the best model or best variable combination for each case.

The concept of anticipated variance originates from survey methodology. It is the expected variance of a sampling scheme, based on the sampling design and the assumed properties of the forest area (e.g. Mandallaz and Ye 1999). Thus, it can be used for choosing an optimal sampling scheme. For anticipating the precision and accuracy of predicted stand characteristics, a purely model-based approach is required (see e.g. Cassel et al. 1977, Gregoire 1998). Kangas (1999) used a mixed model approach to anticipate the accuracy of predicted stand volume at different points of time (for related studies from population forecasting see Alho 1990, Alho and Spencer 1997). The considered stand volume predictions were obtained with a complex simulation system including, among others, treewise growth models and a standwise mortality model. The model used for anticipating the accuracy of future predictions was estimated from the observed errors of the stand volume.

The aim of this study is to produce models for anticipating the precision and accuracy of stand volume, sawlog volume and pulpwood volume estimates in each stand, based on the characteristics of that stand and the information assumed to be available from the stand. The models can also be used to choose the variables to measure or assess in each stand, in order to produce as accurate an estimate of stand volume or timber assortments as possible.

In this study, only Scots pine observations were used, but similar approach could be used for total stand volume or volumes of other tree species as well. The basic stand measurements used are stand age, basal area and basal area median diameter. Eight other measurements, such as number of stems and maximum diameter, are available for additional information. All the measurements are assumed to be error-free. The observed errors in stand volume, sawtimber volume and pulpwood volume estimates, assuming nine different measurement combinations, are modeled using a weighted Seemingly Unrelated Regression (SUR) model.

Table 1. Range, mean and standard deviation (SD) of main characteristics of the study material. A denotes age, d_{gM} basal area median diameter, h_{gM} height of basal area median tree, d_{nM} frequency median diameter, d_m arithmetic mean diameter, d_{min} minimum diameter, d_{max} maximum diameter, N number of stems, $N_{d>6}$ stem number per ha with dbh > 6 cm, G basal area, $G_{d\leq6}$ basal area of trees with dbh ≤ 6 cm, $G_{d>16}$ basal area of trees with dbh > 16 cm, V volume, V_s sawlog volume and V_p pulpwood volume.

Variable	Min	Max	Mean	SD
A, years	15	183	67.3	31.9
d_{gM}, cm	5	35	15.6	6.3
h_{gM}, cm	3.3	27.5	11.8	5.2
d_{nM}, cm	1.3	31.8	11.8	6.1
d_m, cm	2.8	32.2	12.1	5.7
d_{min}, cm	0.2	21.7	4.1	3.1
d_{max}, cm	8.4	48.8	23.6	8.2
N, ha^{-1}	87	4237	1190	789
$N_{d>6}$, ha^{-1}	79	2315	721	433
G, m^{2}ha^{-1}	1.1	32.7	12.2	6.5
$G_{d\leq6}$, m^{2}ha^{-1}	0	5.4	0.8	0.9
$G_{d>16}$, m^{2}ha^{-1}	0	27.3	6.0	6.5
V, m^{3}ha^{-1}	3.6	310.3	80.3	61.7
V_s, cm	0	247.8	38.2	49.3
V_p , cm	0.7	144.0	36.3	27.1

2 Material

The data set includes the permanent sample plots (INKA) measured by the Finnish Forest Research Institute (FFRI), originally installed for growth modeling purposes (Gustavsen et al. 1988). The INKA sample plots were established on mineral soils across Finland. The data includes clusters of three circular plots within a stand. When testing the diameter distribution prediction methods, these circular plots were combined. Altogether 100–120 trees were measured in each stand. Of these trees about 30 were measured as sample trees. The diameter of all trees within a plot were measured to the nearest 0.1 cm. Correspondingly, tree height was measured from sample trees to the nearest 0.1 meters.

Tree height model of Näslund (1936) was first constructed separately for each stand using sample tree measurements. The height of each tally tree was then predicted with these standwise

models. A random component was added to the predicted heights from a normal distribution using the estimated standard deviation of each height model. Total, sawlog and pulpwood volumes were calculated for each tree using taper curve functions presented by Laasasenaho (1982). Finally, stand characteristics were calculated as the averages and sums of tallied trees (Table 1).

Sample tree measurements from all stands were also used for constructing a random coefficient version of Näslund height model. This model was used in calculation of the stand characteristics with different assumed measurement strategies.

3 Methods

3.1 Height Model

The available height models of Scots pine were such that the model is adjusted exactly to the observed height of the assessed mean tree (i.e. a tree with dbh = d_{gM}). However, a more efficient calibration approach is to use a random parameter model and linear prediction theory. Therefore, a random coefficient version of Näslund's model for tree k in plot j and stand i

$$y_{ijk} = b_0 + b_1 dbh_{ijk} + b_{0i} + b_{1i} dbh_{ijk} + u_{ij} + \varepsilon_{ijk} \quad (1)$$
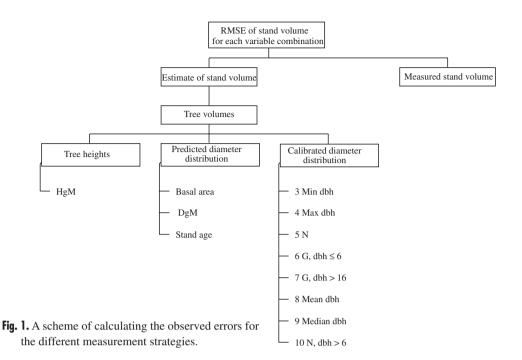
was estimated, where

$$y_{ijk} = \frac{dbh_{ijk}}{\sqrt{h_{ijk} - 1.3}} \quad (2)$$

b_0 is the (fixed) intercept term, b_1 the (fixed) coefficient of dbh, b_{0i} the random stand effect, b_{1i} is the random standwise coefficient of dbh, u_{ij} is the random plot effect and ε_{ijk} the random tree effect (residual error) (e.g. Goldstein 1995, Lappi 1991). In this model, therefore, both the intercept term and the coefficient of dbh vary from stand to stand.

Generally, the random coefficient model can be presented as

$$\mathbf{y} = \mathbf{Xa} + \mathbf{Zb} + \boldsymbol{\varepsilon} \quad (3)$$

where **y** is a vector of n observations, **X** is a n × p matrix of independent fixed variables, **Z** is a

Fig. 1. A scheme of calculating the observed errors for the different measurement strategies.

n × q design matrix, **a** is a p-vector of unknown population parameters, **b** is a q-vector of unknown random parameters and **ε** is the residual vector (e.g. Penner et al. 1995). If var(**b**) = **D** and var(**ε**) = **R**, the variance-covariance matrix of **y** can be presented as

$$\text{var}(\mathbf{y}) = \mathbf{ZDZ'} + \mathbf{R} \qquad (4)$$

Then, the estimates of unknown random parameters in each stand can be calculated as (see Lappi 1991)

$$\hat{\mathbf{b}} = \mathbf{DZ'}(\mathbf{ZDZ'} + \mathbf{R})^{-1}(\mathbf{y} - \mathbf{Xa}) = \\ \left[\mathbf{Z'R^{-1}Z} + \mathbf{D^{-1}}\right]^{-1}\mathbf{Z'R^{-1}}(\mathbf{y} - \mathbf{Xa}) \qquad (5)$$

where (**y** – **Xa**) is the vector of differences between observed and predicted values of **y**. In this study, the height for each diameter class in each stand was predicted using model (1), which was calibrated according to d_{gM}/h_{gM} relationship using equation (5).

3.2 Predicting the Forest Characteristics

In the studied case, nine different measurement strategies were tested. One of them was the basic strategy, in which stand age, basal area median diameter, height of the median tree and basal area are assumed to be known. In each of the other strategies one additional variable was assumed to be known (Fig 1). The additional variables are minimum diameter (Strategy 3), maximum diameter (4), number of stems (5), basal area of trees smaller than or equal to 6 cm (6), basal area of trees over 16 cm (7), median diameter (8), arithmetic mean diameter (9), and number of stems with dbh larger than 6 cm (10). The diameter distribution was predicted in two steps. First, the basal area diameter distribution was predicted with the percentile method. Second, the predicted distribution could be calibrated using one additional variable. Finally, each stand had the stand volume, sawlog volume and pulpwood volume estimated with nine different measurement combinations.

In the percentile method, the diameters at predefined percentiles of the distribution function are predicted with models (Borders et al. 1987, Mal-

Table 2. The calibration equations for strategies 5–10, where d_k denotes the diameter of diameter class k, w_k the modified class basal area of class k, g_k basal area of the mean tree in class k, and K the number of diameter classes. For other definitions see Table 1.

Auxiliary variable	Calibration equation		Strategy
G	$\sum\limits_{k=1}^{K} w_k = G$		all
d_{gM}	$\sum\limits_{k=1}^{K} w_k \gamma_k = G\!\big/\!2$	where $\gamma_k = \begin{cases} 1\ , d_k < d_{gM} \\ 1/2,\, d_k = d_{gM} \\ 0,\, d_k > d_{gM} \end{cases}$	all
N	$\sum\limits_{k=1}^{K} w_k\big/g_k = N$		5
$G_{d\leq6}$	$\sum\limits_{k=1}^{K} w_k \gamma_k = G_{dbh\leq6}$	where $\gamma_k = \begin{cases} 1,\, d_k \leq 6\ cm \\ 0,\, d_k > 6\ cm \end{cases}$	6
$G_{d>16}$	$\sum\limits_{k=1}^{K} w_k \gamma_k = G_{dbh>16}$	where $\gamma_k = \begin{cases} 1,\, d_k > 16\ cm \\ 0,\, d_k \leq 16\ cm \end{cases}$	7
d_{nM}	$\sum\limits_{k=1}^{K} w_k\big/g_k\, \gamma_k = N/2$	where $\gamma_k = \begin{cases} 1,\, d_k < d_{med} \\ 1/2,\, d_k = d_{med} \\ 0,\, d_k > d_{med} \end{cases}$	8
d_m	$\sum\limits_{k=1}^{K} w_k\big/g_k\, d_k = N\bar{d}$		9
$N_{d>6}$	$\sum\limits_{k=1}^{K} w_k\big/g_k\, \gamma_k = N_{dbh>6}$	where $\gamma_k = \begin{cases} 1,\, d_k > 6\ cm \\ 0,\, d_k \leq 6\ cm \end{cases}$	10

tamo et al. 2000, Kangas and Maltamo 2000b). The models used for predicting the percentiles were those presented by Kangas and Maltamo (2000b). They estimated two model sets: in the first set the number of stems was not included as a predictor and in the second set it was. The other regressors were the stand basal area, stand age and the basal area median diameter. In this study, the second set was utilized only when the stem number was assumed to be known, in all other strategies the first model set was used. By interpolating between the predicted diameters, a cumulative basal area diameter distribution function is obtained. Interpolation was carried out using Späth's rational spline, in order to obtain a monotone distribution (Maltamo et al. 2000).

The predicted basal area diameter distribution was calibrated with an approach presented by Deville and Särndal (1992). Kangas and Mal-

tamo (2000a,c) used this approach to calibrate the predicted class frequencies / basal areas of a diameter distribution. In the present study, the calibration estimator was used to modify the predicted basal area ba_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 ba_k, while respecting the calibration equation(s). The calibration equations for strategies 5–10 are presented in Table 2.

The calibration equation for mean diameter, however, is truly a calibration equation for the sum of diameters. The calibration is successful with respect to the arithmetic mean diameter only if the number of stems is also correct, which it is not in the studied case except occasionally. However, it was not possible to form a real calibration equation purely for the arithmetic mean, since the number of stems, which is subject to calibration,

is a fixed coefficient in the calibration equation (see Table 2). Similar problem occurs, for example, with dominant diameter.

If the minimum or maximum diameters were assumed to be known (strategies 3 and 4), the distribution obtained with percentile method was re-scaled to the correct interval (Kangas and Maltamo 2000c). 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

$$\hat{d}_i* = d_{\min} + ((d_{gM} - d_{\min})/(d_{gM} - \hat{d}_0)) \, (\hat{d}_i - \hat{d}_0) \quad (6)$$

or

$$\hat{d}_i* = d_{gM} + ((d_{\max} - d_{gM})/(\hat{d}_{12} - d_{gM})) \, (\hat{d}_i - d_{gM}) \quad (7)$$

where d_{min} is the observed minimum diameter, d_{max} the observed maximum diameter, $\hat{d}_0$ is the predicted minimum diameter, $\hat{d}_{12}$ the predicted maximum diameter, $\hat{d}_i$ is the predicted diameter at ith percentile and $\hat{d}_i *$ is the re-scaled diameter.

Finally, it was also required that the basal area, which was used in scaling the relative basal area, and the basal area median diameter also remained correct after calibration. The distance measure used was the square root distance of Deville and Särndal (1992). Minimizing this distance measure while respecting the calibration equation(s) is a constrained non-linear optimization problem (see Deville and Särndal 1992 for details). The resulting group of non-linear equations was solved using IMSL subroutines.

3.3 Modeling the Prediction Errors

The observed errors of each stand characteristic of interest, namely the total stand volume, sawlog volume and pulpwood volume were modeled for each of the nine strategies as a function of forest basal area, basal area mean diameter, number of stems and a dummy variable indicating whether the basal area median diameter of a stand was over 16 cm and their transformations and interactions. Thus, the errors were assumed to depend on both forest characteristics and the measurement combination used to calculate the results. The model

consisted of a fixed part, which represents the bias in the predictions, and a random part, which represents the variance of the predictions.

The model for each of the three stand characteristics and nine strategies was first fitted using an Ordinary Least Squares (OLS) regression model. The residuals of these OLS models were used to model the variance of the residuals. First, the mean of the squared residuals was calculated for predefined basal area classes. Then, a non-linear model was fitted to these class variances (see Lappi 1997). The variances of the volume characteristics were assumed to be proportional to some (unknown) power a of the stand basal area G

$$\text{var}(e) = \sigma^2 G^a + \varepsilon \quad (8)$$

meaning that for $a > 0$ the variance of the predictions increases with increasing basal area of the stand, the faster the larger the power parameter. For $a < 0$, the variance decreases with increasing basal area. The obtained estimates of the parameter a were then used to calculate weights $1/G^a$ for the observations. Finally, estimates of the fixed coefficients and σ^2 were obtained using weighted SUR model (Zellner 1962). The weights were needed because the residuals were highly heteroscedastic, and the SUR approach was used because the errors of the timber assortments and volume in a stand are correlated.

In the studied case, each strategy included one additional measurement. In this case, $9 \times 3 = 27$ models were needed. In principle, it would also have been possible to consider strategies with several additional variables. However, including two additional measurements would have increased the number of strategies considerably. Therefore, only the simple strategies were considered.

4 Results

The height model was estimated with MLWin program (Rasbash et al. 2000). The estimated parameters are presented in Table 3. The error models for the three stand variables considered were estimated with SAS REG (the OLS models), MODEL (the variance functions) and SYSLIN
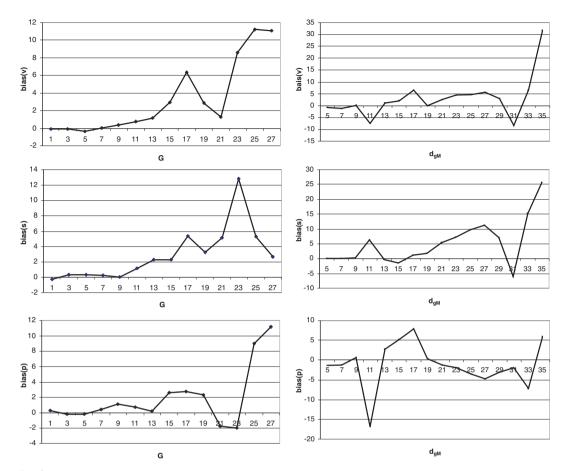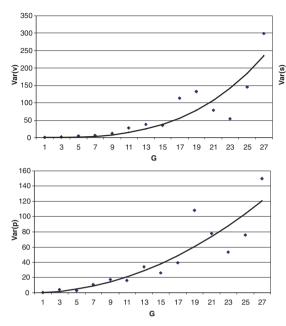
Fig. 2. The bias of stand (a), sawlog (b) and pulpwood (c) volume with respect to basal area and basal area median diameter in the basic strategy.

(the final models) procedures. The parameters of these models are presented in Tables 4–6.

The basic strategy produced estimates of total volume, in which the bias increased with increasing basal area (Fig 2.). The sawlog volume estimates were also biased; the bias increased both with respect to basal area and basal area median diameter. Pulpwood volume was most severely biased in the stands with midsize d_{gM}. These phenomena can also be detected from the coefficients of the error models (Tables 4–6). The variances of all the characteristics considered were heteroscedastic, the total stand volume the most (Fig. 3). The estimated power parameters a, which describe how the variance increases with increasing basal area varied from 1.97 for pulpwood to 3.11 for total volume (Tables 4–6).

Table 3. The random coefficient Näslund height model, where b_0 is the fixed intercept, b_1 is the fixed coefficient of diameter, b_{0i} is random stand effect, b_{1i} is random coefficient of diameter in a stand, u is random plot effect and ε is random tree effect.

Parameter	Estimated value
b_0	1.163
b_1	0.259
$Var(b_{0i})$	0.106
$Var(b_{1i})$	0.003
$Cov(b_{0i}, b_{1i})$	0.002
$Var(u_{ij})$	0.035
$Var(\varepsilon_{ijk})$	0.036

805

Fig. 3. The class variances and the estimated variance functions of the basic strategy for stand (a), sawlog (b) and pulpwood (c) volume.

When additional variables were included, both the bias and the variance component could in many cases be reduced. For example for total volume, when stem number was assumed to be known, the bias increasing with basal area could be clearly reduced. When the median diameter was assumed to be known, the bias was negligible. In most cases, however, the bias could not be entirely removed by including an additional measurement. In principle, the error models like the ones estimated in this study could be used to remove the remaining bias. In the studied case this was not intended, however, because it would be difficult to distribute the bias correction correctly over the diameter classes. Therefore, the models were only used to show the approach leading to minimum bias in certain conditions.

In the case of pulpwood volume, the use of additional variables could reduce the power parameter markedly. Then, the variance of pulpwood volume obtained from a calibrated distribution does not increase with increasing basal area as rapidly as that of the basic strategy. For example, by using the maximum diameter for calibration, the parameter *a* reduced from 1.97 to 1.46. In some cases, however, the power parameter increased when additional information was used. This was the case for sawlog volume, where
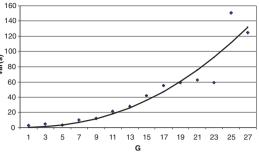
the maximum diameter increased the power parameter from 2.22 even to 3.55. Consequently, using maximum diameter can markedly improve the RMSE of pulpwood volume estimates and at the same time markedly worsen that of the sawlog volume estimates.

Figs. 4–6 show examples of the different strategies with varying basal area or basal area median diameter. The best measurement combinations depended both on the characteristics considered and on the stand conditions. For total stand volume, the best measurement combination was to measure median diameter as additional variable, and the deviation to the basic strategy increased with increasing basal area. However, basal area of trees under 7 cm, minimum and maximum diameters only worsened the results in all cases. This can be directly seen also from coefficients in Table 4.

For sawlog volume, all the additional variables increased the value of the power parameter, and consequently, the basic strategy was in all cases the best with respect to variance. The worst strategy with respect to RMSE was to measure maximum diameter as additional variable, and the best strategy was to measure minimum diameter (Fig. 6.). Measuring minimum diameter from stands with large basal area and small basal area median

Table 4. The coefficients of the error models of stand volume for the nine strategies. N denotes stem number, G basal area, d_{gM} stand basal area, σ^2 residual variance and a the power of G the variance is proportional to. For the numbers of strategies, see Fig. 4.

Strategy	intercept	N	G^2	d_{gM}	d_{gM}^2	σ^2	a
1	0.145441	–0.000534	0.013104	–	–	0.01459	3.116526
3	0.353486	–0.000872	0.013790	–	–	0.01410	3.132727
4	0.012627	–	0.007325	–	–	0.00886	3.355337
5	–0.030237	–	0.004559	–	–	0.01175	3.204359
6	1.916219	–0.000611	0.007781	–0.227484	0.005375	0.00328	3.890342
7	–0.053646	–	–	–	–	0.01708	3.049362
8	–0.077787	–	–	–	–	0.02059	2.969422
9	–0.068760	–	0.005116	–	–	0.01407	3.120086
10	–0.074209	–	–	–	–	0.01179	3.225479

Table 5. The coefficients of the error models of sawlog volume for the nine strategies. D2 is defined as

$$d2 = \begin{cases} d_{gM}, d_{gM} > 16 \\ 0, \text{otherwise} \end{cases}, \text{ other definitions see Table 4 and Fig. 4.}$$

S	int	N	G	G^2	ln(G)	d_{gM}	d_{gM}^2	d_{gM}/G	d2	d2/G	σ^2	a
1	2.7742	–0.00056	0.1285	0.0059	–	–0.4191	0.01068	–	0.0898	–	0.01376	2.2219
3	–0.0362	–	–	–	–	–	–0.00205	–	0.3115	–0.9170	0.01320	2.2965
4	–0.3690	–0.00023	–	0.0077	0.5786	–	0.00291	–	–	–0.2270	0.00604	3.5505
5	0.2839	–	0.3109	–0.0108	–0.8199	–	0.00246	–0.1309	–	–	0.00906	2.7429
6	1.3161	–	–	–	–	–0.2069	0.00395	–	0.2110	–0.6269	0.00315	3.3734
7	–0.1186	–	–	–	–	–	–	–	–	0.1673	0.01315	2.6808
8	0.9751	–	–	–	–	–0.1959	0.00743	–	–	–	0.01579	2.6317
9	–0.1182	–	–	–	–	–	–0.00094	–	0.2112	–0.6430	0.01085	2.7545
10	0.8366	–	–	–	–	–0.1740	0.00632	–	–	–	0.00897	3.2511

Table 6. The coefficients of the error models of sawlog volume for the nine strategies. For definitions see Tables 4 and 5 and Fig. 4.

S	int	N	G	G^2	ln(G)	d_{gM}	d_{gM}^2	d_{gM}/G	d2	d2/G	σ^2	a
1	–3.6323	–	–	0.0064	–	0.5759	–0.01709	–	–0.0617	–	0.01526	1.9699
3	0.5615	–0.00125	0.1854	0.0133	–	–	–	–	–0.3209	1.0587	0.01372	2.0283
4	1.2518	–0.00062	–	–	0.8372	–0.2753	–	0.2038	–	0.3344	0.00818	1.4593
5	0.7568	–0.00088	–	0.0086	–	–	–	–	–0.1268	0.3191	0.01035	1.6668
6	0.6630	–0.00089	–	0.0085	–	–	–	–	–0.2200	0.7029	0.00343	1.9448
7	0.3339	–0.00111	–	–	1.5989	–0.2180	–	0.2594	–	–	0.01405	1.8108
8	0.4600	–0.00072	–	0.0076	–	–	–	–	–0.1336	0.4037	0.01611	1.7607
9	0.5430	–0.00061	–	0.0112	–	–	–	–	–0.2681	0.9079	0.0143	1.5595
10	0.3841	–0.00061	–	0.0053	–	–	–	–	–0.1000	0.2723	0.00923	1.9603

diameter improved the RMSE. However, even if the variances increased when additional variables were used, the biases markedly decreased (Fig 7b).

The largest improvements in terms of variance and RMSE were obtained for pulpwood volume. In the case of pulpwood, measuring the minimum diameter proved to be the worst strategy and measuring the maximum diameter the best strategy (Fig. 6). This is just the opposite to the sawlog volume case. Especially with large basal area, all strategies except measuring minimum
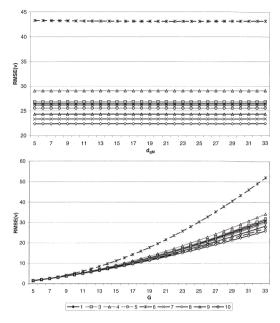
Fig. 4. The anticipated RMSE of predicted total stand volume using different strategies, with respect to basal area median diameter (A) and basal area (B), the other variables were assumed fixed. The strategies were: basic (1), minimum diameter (3), maximum diameter (4), number of stems (5), basal area of trees smaller than or equal to 6 cm (6), basal area of trees over 16 cm (7), median diameter (8), arithmetic mean diameter (9), number of stems with dbh larger than 6 cm (10).

Fig. 5. The anticipated RMSE of predicted sawlog volume using different strategies, with respect to basal area median diameter (A) and basal area (B), the other variables were assumed fixed. For the number of strategies see Fig. 4.

diameter were better than the basic strategy. However, with respect to bias the situation was not so clear (Fig. 7a). When the bias was largest, namely with the values of d_{gM} between 13–16 cm, nearly all the strategies reduced the bias markedly. However, with very small or large values of d_{gM}, most of the strategies were worse than the basic strategy.

5 Discussion

In this paper, a model-based method was presented for anticipating the precision and accuracy of interesting variables predicted with a system of models. In the studied case, the stand volume and
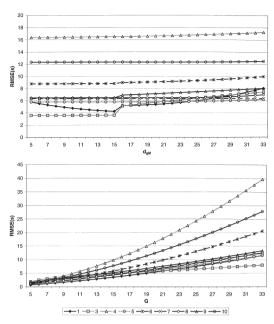
timber assortments were calculated using height and taper curve models as well as diameter distribution models and calibration, so that analytical variances of the results would be difficult to obtain. The anticipated variance of a variable is its expected variance under a chosen measurement combination and a given stand condition. It is calculated using a model estimated from the observed errors of the variable of interest. Even if the conditions included in the error model remain the same, the actual errors of the variable may vary markedly from stand to stand.

The estimated models show that estimated sawlog and pulpwood volumes are especially biased with respect to basal area median diameter of the stand. For sawlog volume, the largest biases are observed in stands with biggest d_{gM}, for pulpwood volume, the largest biases are observed in stands with d_{gM} around 15 cm. This phenomenon is most probably due to the nature of timber assortments: the volume of both pulpwood and sawlog change in steps as the mean diameter in
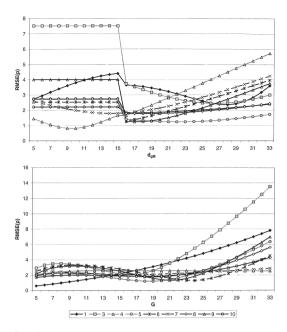
Fig. 6. The anticipated RMSE of predicted pulpwood volume using different strategies, with respect to basal area median diameter (A) and basal area (B), the other variables were assumed fixed. For the number of strategies see Fig. 4.

Fig. 7. The anticipated bias of predicted pulpwood (A) and sawlog (B) volume using different strategies, with respect to basal area median diameter. The other variables were assumed fixed. For the number of strategies see Fig. 4.

the stand increases. With calibration, these biases could be somewhat reduced. With all the stand characteristics, the variance is heteroskedastic, increasing with increasing basal area.

According to the results, the best measurement strategy varies between the characteristics of interest. For each of the three variables considered, a different strategy was best. In such a case, the accuracy of the different characteristics has to be given weights describing their relative importance, in order to choose the best strategy for the whole stand. These weights could also vary with respect to conditions. For example, in old stands, the accuracy of sawlog volume estimates could be given a large weight, and in young stands total volume may be given the largest weight.

The best strategy also varied according to stand conditions. In the case of total volume, measuring the median diameter was uniformly the best strategy, but in the case of timber assortments the best strategies varied. What was problematic, was that the best strategy for estimating pulpwood
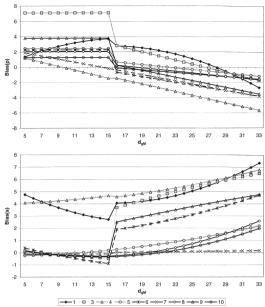
volume was often the worst for estimating sawlog volume and vice versa.

The cases of sawlog and pulpwood volume are also problematic, since the additional variables may clearly increase the variance but reduce the bias, or increase the bias but reduce the variance. In the case of one stand, the variance component may dominate the RMSE value. However, when several stands are measured according to the same strategy, the bias component may be more important. Then, the number of stands under similar conditions also may have an effect on the best measurement strategy.

In the studied case, the strategies compared included different kinds of information. A similar approach, however, could be used to compare different modeling approaches. For example, in the case of diameter distributions, Weibull function based distributions may work well in certain stands, percentile based distributions in other stands. In many studies it has been observed that the best approach varies in different data sets

(e.g. Kangas and Maltamo 2000), but so far it is unclear under which conditions the differences between these methods are negligible and under which they should be noted.

There are many interesting avenues for future research in this topic. Using a Bayesian framework for the error model would enable statements, for example, about the probability of observing an error larger than a certain value or a value within a certain interval. Accounting for the possible measurement errors in the variables assessed in the field is another important issue. For example, even if the use of the median diameter is the best strategy with respect to total volume, when measurement errors are not included, another strategy might be better when the errors are accounted for.

As a conclusion, this study demonstrates that no measurement combination is uniformly best for predicting the stand volume and timber assortments. The best measurement combination depends on the variable of interest as well as the conditions in the stand considered. To choose the best measurement combination for any one stand, weights for the the accuracy of different variables need to be applied.

Acknowledgements

References

Alho, J.M. 1990. Stochastic methods in population forecasting. International Journal of Forecasting 6: 521–530.

— & Spencer, B. 1997. The practical specification of the expected error of population forecasts. Journal of Official Statistics 13: 203–225.

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.

Cassel, C-M. Särndal, C-E. & Wretman, J.H. 1977. Foundations of inference in survey sampling. John Wiley & Sons, New York. 192 p.

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

Goldstein, H. 1995. Multilevel statistical models. 2nd. edition. Kendall's library of statistics 3. Arnold, London. 178 p.

Gregoire, T.G. 1998. Design-based and model-based inference in survey sampling: appreciating the difference. Review. Canadian Journal of Forest Research 28: 1429–1447.

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. (in Finnish)

Kangas, A. 1999. Methods for assessing the uncertainty of growth and yield predictions. Canadian Journal of Forest Research 292: 1357–1364.

— & 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. Silva Fennica 34: 371–380.

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

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

Lappi, J. 1991. Calibration of height and volume equations with random parameters. Forest Science 37: 781–801.

— 1997. A longitudinal analysis of height/diameter curves. Forest Science 43: 555–570.

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

Mandallaz, D. & Ye, R. 1999. Forest inventory with optimal two-phase, two-stage sampling schemes based on the anticipated variance. Canadian Journal

of Forest Research 29: 1691–1708.

Näslund, M. 1936. Skogsförsöksanstaltens gallrings-
försök i tallskog. Meddelanden från Statens Skogs-
försöksanstalt 29. 169 p.

Penner, M., Penttilä, T. & Hökkä, H. 1995. A method
for using random parameters in analyzing perma-
nent sample plots. Silva Fennica 29: 297–309.

Rasbash, J., Browne, W., Goldstein, H., Yang, M.,
Plewis, I., Healy, M., Woodhouse, G., Draper, D.,
Langford, I. & Lewis, T. 2000. A User's Guide to
MLwiN. University of London. 280 p.

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

Total of 21 references