Comparing Simulation Methods for Modelling the Errors of Stand Inventory Data

Arto Haara

Forest management planning requires information about the uncertainty inherent in the available data. Inventory data, including simulated errors, are infrequently utilised in forest planning studies for analysing the effects of uncertainty on planning. Usually the errors in the source material are ignored or not taken into account properly. The aim of this study was to compare different methods for generating errors into the stand-level inventory data and to study the effect of erroneous data on the calculation of specieswise and standwise inventory results. The material of the study consisted of 1842 stands located in northern Finland and 41 stands located in eastern Finland. Stand-level ocular inventory and checking inventory were carried out in all study stands by professional surveyors. In simulation experiments the methods considered for error generation were the 1nn-method, the empirical errors method and the Monte Carlo method with log-normal and multivariate log-normal error distributions. The Monte Carlo method with multivariate error distributions was found to be the most flexible simulation method. This method produced the required variation and relations between the errors of the median basal area tree characteristics. However, if the reference data are extensive the 1nn-method, and in certain conditions also the empirical errors method, offer a useful tool for producing error structures which reflect reality.

Author´s address Finnish Forest Research Institute, Joensuu Research Centre, P.O.Box 68, FIN-80101 Joensuu, Finland
E-mail arto.haara@metla.fi

1 Introduction

Reliable inventory data are essential for forest growth modelling and forest planning. In Nordic countries data are usually collected using subjective forest inventory methods (e.g. ocular inventory methods). The basic unit of a subjective forest inventory is a forest stand, which is a homogeneous forest region about 0.5–20 hectares in acreage. The criteria for the delineation of a stand is based on relevant stand characteristics, e.g. the site fertility, stand age and composition of the tree species. Forest stands are also used as management planning units.

In Finland, forest inventories are mainly carried out standwisely using Bitterlich (1984) sample plots. The basal area of the stand is assessed as an average of the representative sample plots which have been selected subjectively by the surveyor. Therefore, the sampling errors are difficult to estimate. Tree heights and diameters at breast height are not measured. Instead, trees are counted using a relascope and one basal area median diameter tree per tree species per stand is assessed by the surveyor.

The specieswise mean stand characteristics are used as dependent variables in diameter distribution models (e.g. Kilkki and Päivinen 1986, Mykkänen 1986, Kilkki et al. 1989). With these models the specieswise theoretical diameter distributions of the stand are estimated. Sample trees from this distribution are used to describe the current situation of the stand. For example, the volumes of the sample trees from the theoretical diameter distribution can be calculated by taper curve models and by summarizing these volumes the total or merchantable stand volume can be calculated.

Standwise field data include sampling, assessment and classification errors. Furthermore, derived stand characteristics of the inventory data, e.g. the stand volume per hectare, are calculated from the assessments of the stand characteristics using a variety of models. Prediction errors of statistical models are derived from four main sources: 1) the model misspecifications, 2) the random estimation errors of the model coefficients, 3) the residual variation of the models, and 4) the errors in the independent variables of the models, which can include sampling error, measurement error, grouping error and prediction error (e.g. Kangas 1999). Judgemental aspects in the predictions can also be a source of error (Alho 1990).

Because stand-level inventory is a subjective method there is considerable variation in its accuracy. In Scandinavian forests the standard error of the stand-level inventory of the basal area (G) can vary from 13 per cent to 22 per cent and the bias of the basal area can be 10 per cent (e.g. Mähönen 1984, Laasasenaho and Päivinen 1986, Nersten and Næsset 1992, Ståhl 1992, Pigg 1994). From the derived stand characteristics the standard error of the total stand volume varies from 15 to 45 per cent (e.g. Poso 1983, Laasasenaho and Päivinen 1986, Pussinen 1992, Ståhl 1992, Hyyppä et al. 2000). The variation of the errors of the basal area and the total stand volume can, for example, be due to different age and forest site distributions of the study area and the diverse experience of the surveyors (e.g. Laasasenaho and Päivinen 1986).

The errors of the basal area median tree characteristics are correlated (e.g. Ståhl 1992, Pigg 1994). The variance of the mean age is about 20 per cent and that of the mean diameter and mean height is 10 to 20 per cent (e.g. Mähönen 1984, Laasasenaho and Päivinen 1986, Ståhl 1992, Pigg 1994). However, there are only few studies on the specieswise errors of the stand-level inventory in mixed forests in Finland (e.g. Pussinen 1994).

Usually the errors of the stand-level inventory are estimated by measuring a systematic net of stand checking sample plots on each stand. Relascope sample trees from these plots are measured for composing the empirical diameter distribution of the stand. The empirical stand characteristics (e.g. basal area, stand volume per hectare) are calculated from this distribution. However, the use of systematic sampling includes some problems; the error estimates are not accurate and possible systematic variation of the stand can affect the estimate of the variance.

There are two widely used methods to estimate the uncertainty: the Monte Carlo method (e.g. Mäkelä 1988, McRoberts et al. 1994, Kangas 1996, Kangas 1999) and the variance propagation method (e.g. Gertner 1987, Kangas 1996). In both methods the total error is composed of

several sources of errors.

Monte Carlo methods involve the repeated sampling of the probability distribution for model parameters, driving variables, boundary conditions and initial conditions and use of generated sets of samples in a simulation (Rubinstein 1981). The probability distribution of the model prediction is then derived from the combination of model predictions resulting from repeated simulations based on the sampled inputs. The advantages of Monte Carlo techniques include that the precision can be assessed without an independent measurement data set and the effect of certain assumptions or models can be studied separately. One reservation with the Monte Carlo method is that simulations produce only a lower limit for the true variance because all the error sources may not be known and cannot therefore be taken into account (Kangas 1999). Monte Carlo methods also require massive computations for large areas.

Variance propagation methods, such as the Taylor series expansion (e.g. Mowrer and Frayer 1986, Gertner 1987, Mowrer 1991, Kangas 1996), require the computation of a deterministic output trajectory for the model, followed by the quantification of the effects of various small amplitude sources of input uncertainty or uncertainties about the reference trajectory (Burges and Lettenmaier 1975, Argantesi and Olivi 1976). Although the use of variance propagation methods can be difficult in complex situations because of their highly restricted demands, these methods can be more suitable than Monte Carlo methods when the simulation data are large (Gertner et al. 1995, Kangas 1996).

Forest management planning requires information about the uncertainty inherent in the available data. Inventory data, including simulated errors, are infrequently utilised in forest planning studies for analysing the effects of uncertainty on planning. Usually the errors in the source material are ignored or not taken into account properly. For example, in stand-level inventories the correlation between assessment errors is ignored. However, the quantification and modelling of errors related to the state variables of a model(s) are important. The aim of this study was to compare different ways to generate errors into the stand-level inventory data and also to study how errone-

ous data affect the calculation of specieswise and standwise inventory results. The simulation methods compared were: 1) the 1nn-method, 2) the empirical errors method and Monte Carlo methods with 3) log-normal distributions and 4) joint log-normal distributions.

2 Material and Methods

2.1 Study Material

The study material comprised two independent stand checking data sets. The first checking data (CC1) were measured during 1990–1994 in northern Finland. The checking was carried out by measuring a systematic net of relascope sample plots on each stand. Tree species, diameter at breast height and the height of each tree from the plots were measured. The average amount of the sample plots was 7.7 plots per stand. From this data set (CC1) 90 stands were selected using random sampling. These sample stands formed data set CC1b and the remaining 1752 stands formed data set CC1a.

The other stand checking data (CC2) were measured in eastern Finland in 1999. The checking was done by measuring a systematic net of circular sample plots on each stand. The radius of the plots varied from 3.99 meters (young stands) to 10 meters (mature stands). The average amount of the sample plots was 8 plots per stand.

The error distributions of the different stand characteristics and relations between measurement errors were studied using data set CC1a. Data set CC1b and the second stand checking data set (CC2) were used for testing different error generation methods. The data sets are introduced in Table 1. Both stand checking data sets included sampling errors from the systematic sampling of the checking sample plots. This caused the checking results to underestimate the accuracy of the stand-level inventory. However, for simplicity's sake and because the sampling errors for systematic sampling are somewhat problematic, this complication was ignored in this study.

An ocular stand-level inventory was carried out in both stand checking data sets by professional surveyors. In the case of checking data CC1, 24 surveyors carried out the assessment. In the case

Table 1. The mean stand characteristics of the stand checking inventory data.

	n	Vol (m³ha⁻¹)	G (m²ha⁻¹)	D (cm)	H (m)	N (Nha⁻¹)	Area (ha)
CC1a							
* Pine	1388	44.9	7.3	17.8	10.7	529	
* Spruce	632	18.8	3.7	14.1	9.8	424	
* Deciduous trees	831	21.2	4.6	10.4	8.9	1031	
* Stands	1752	60.3	10.8	15.1	9.7	1382	6.9
CC1b							
* Pine	70	57.5	8.9	18.2	11.6	536	
* Spruce	55	19.4	4.0	13.4	9.6	437	
* Deciduous trees	60	23.3	5.1	10.6	9.0	1019	
* Stands	90	69.5	12.2	16.2	9.8	1368	6.5
CC2							
* Pine	39	76.7	8.6	25.3	19.2	351	
* Spruce	40	102.8	10.1	20.3	17.2	330	
* Deciduous trees	33	21.2	2.4	21.2	18.9	115	
* Stands	41	187.9	19.6	24.7	19.8	714	3.1

n = number of observations, Vol = stand volume per hectare, G = mean basal area, D = basal area mean diameter, H = basal area mean height, N = number of stems, Area = average area of the stand

of checking data CC2, the same surveyor carried out the whole assessment. Both inventories were carried out by establishing some relascope sample plots within each forest stand. The average values from these plots were recorded as stand characteristics. In data CC1 the relative accuracy of the stand volume per hectare estimates varied considerably between surveyors (from 13.7 to 49.3 percent).

2.2 Basic Calculations

In all data sets empirical diameter distributions of the stands were derived from the stand checking data. The volumes of the trees from different diameter classes of the stand were calculated using Laasasenaho's (1982) taper curve models. The stand volumes per hectare of this true data were obtained by summarizing these tree volumes.

Data including model errors (stand-level inventory data) were created from the true data. Basal area and basal area median tree characteristics (i.e. median diameter (D), height (H) and age) were calculated from the empirical tree diameter distributions of data sets CC1a, CC1b and CC2. Then, using these true stand characteristics, theoretical diameter distributions were estimated with

Weibull diameter distribution models (Kilkki et al. 1989, Maltamo 1998). The heights of the sample trees of the theoretical distribution were calculated using Veltheim's (1991 (1987)) height models and the estimated heights were calibrated with the height of the basal area median tree of the stand. Volumes of the sample trees from the theoretical distribution were calculated using Laasasenaho's (1982) taper curve models. The errors of the data including model errors came from the errors of all these models. The sampling errors of the checking data were not included in the error calculations. The independent variables of the models were assumed to be error free. The stand volume per hectare and the mean sawn wood volume of the stands were calculated by summarizing volumes of the sample trees from the theoretical distribution.

The data including model and assessment errors of the data sets CC1b and CC2 were generated from the data including model errors by generating the errors of the stand characteristics with all simulation methods. Relative errors were used instead of absolute errors because large generated errors of the stand characteristics could have caused negative values when added to small true values. The stand volume per hectare and mean sawn wood volume were also calculated with
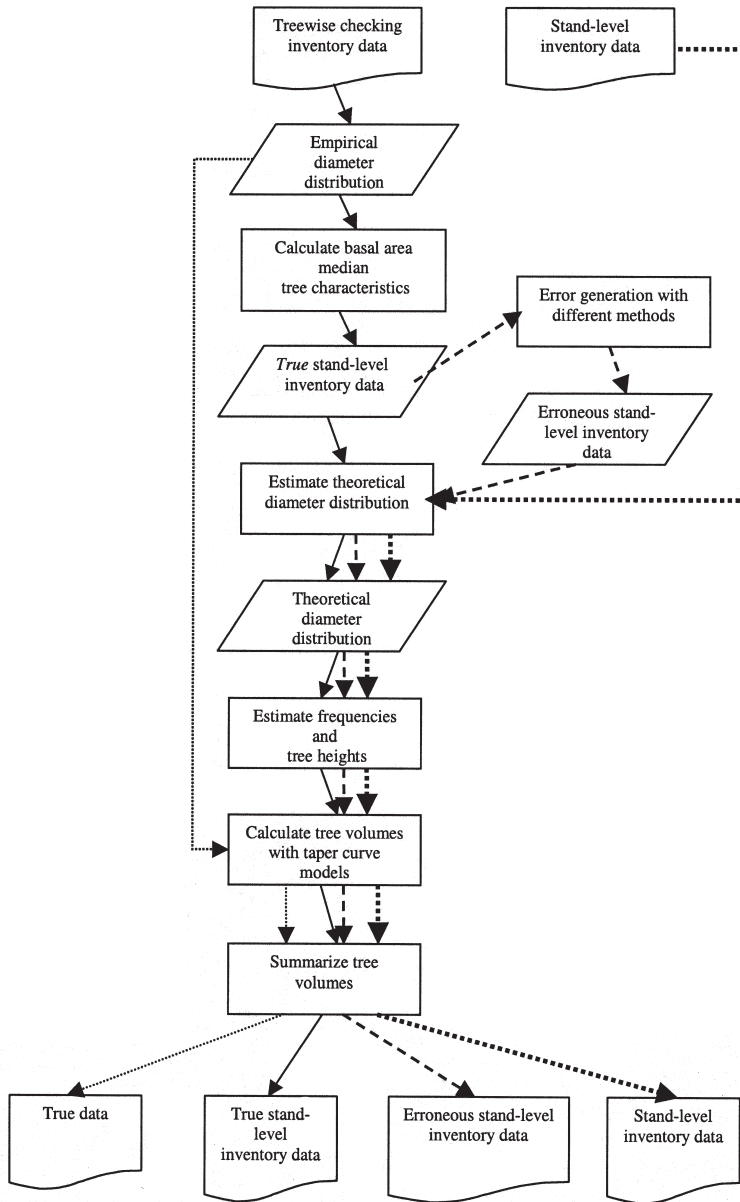
Fig. 1. Flow chart of the calculation of the inventory results of the stand with different methods.

each data including model and assessment errors. The check assessment was considered to be true data with all simulation methods. The generation of true data, data including model errors and data including model and assessment errors are illustrated in Fig. 1.

2.3 The One Nearest Neighbour Method

Nearest neighbour non-parametric regression (Härdle 1989, Altman 1992) offers a useful tool for detecting stands with similar empirical assessment errors. When the amount of nearest

neighbours is restricted to one the method can be called a one-nearest neighbour method (1nn-method). The idea behind using the 1nn-method is the supposition that the level and the structure of the assessment errors are similar when stands are similar. The target stand is a stand which is excluded from the reference data and to which errors are generated. The search for a similar reference stand was done using commonly measured stand characteristics, which were used as the distance function's variables. The standardization of these variables was performed by subtracting the mean of the variable and dividing it by the standard deviation of the variable. The standardization was used to eliminate the influence of different scales of the variables. The stand checking data set CC1a was used as reference data and data sets CC1b and CC2 as target stand data.

In the 1nn-method similarity distance functions were applied depending on the tree species. The similarity distance function for tree species i in target stand k was (Function 1)

$$d_{k_i} = |Gp_{k_i} - Gp_{j_i}| + |G_{k_i} - G_{j_i}| + |Dgm_{k_i} - Dgm_{j_i}| + neighbour_{k_i l} \quad (1)$$

where

Gp_{k_i} the proportion of the basal area per hectare of tree species i in the target stand k

Gp_{j_i} the proportion of the basal area per hectare of tree species i in the reference stand j

G_{k_i} the basal area per hectare of tree species i in the target stand k

G_{j_i} the basal area per hectare of tree species i in the reference stand j

Dgm_{k_i} the basal-area median diameter of tree species i in the target stand k

Dgm_{j_i} the basal-area median diameter of tree species i in the reference stand j

$neighbour_{k_i l}$ the neighbour parameter for target stand k in the lth simulation.

The random parameter $neighbour_{k i l}$ was used to add variation into the selection of the nearest neighbour. A random variate from the normal distribution N(0,d), where d is the standard deviation of the distance function, was added into the parameter when the stand was chosen as a nearest neighbour. Otherwise the same nearest neighbour would have been chosen each time the simulation

was carried out. The relative differences between the original assessment and check assessment of the neighbour stand characteristics of the nearest neighbour were used as the assessment errors of the target stand.

2.4 The Empirical Errors Method

The use of empirical errors as a source of erroneous inventory data removes the uncertainties caused by the assumptions which must be made during the simulation of the errors in Monte Carlo methods. The error variances of the stand characteristics correspond closely to reality if the empirical errors of the stand characteristics are added from another stand that is as similar as possible.

When the empirical errors method was used the reference data (CC1a) were first classified specieswise into four basal area median tree diameter classes (0–9.99 cm, 10–14.99 cm, 15–19.99 cm, 20+ cm). The specieswise diameter class which included the tree species of the target stand was chosen. The specieswise reference unit was chosen from the class using random sampling. Then the differences between the original assessment and the check assessment of the neighbour stand characteristics were added from the reference unit. The sampling was done 50 times for each stand.

2.5 The Monte Carlo Methods

Monte Carlo methods were also used for generating errors into the stand characteristics. The error sources considered were the errors in the independent variables of the models (diameter distribution, height and stem curve) and the standard errors of the parameter estimates. The errors of the independent variables were simulated in two ways: 1) by using log-normal error distributions (D, H, Age, N, G) and 2) by using multivariate log-normal error distributions (D, H, Age, G) and log-normal error distributions (N).

There were not enough stem number assessments for estimating correlation matrices for multivariate distributions in stand-level inventory data CC1a. Thus log-normal error distributions of the stem number were used instead of multivariate

distributions.

The stand characteristics of the stand-level inventory were highly biased in test data CC2 because the surveyor was a beginner. Thus Monte Carlo simulations with bias added were also carried out. The RMSEs (root mean square error) of the stand characteristics were divided into their two components: bias and variance (Formula 2).

$$RMSE^2 = Variance + Bias^2 \qquad (2)$$

Bias was added to generated random variates with a simple Formula 3:

Estimated value =
    True value + Random error + Bias   (3)

The biases in test data CC2 were used as the estimates of the biases. The biases were fixed for all stands. The possibility of using trend in error simulation was also utilized. The trend was added using the Formula 4:

Estimated value =
    $a + b$*True value + Random error   (4)

Parameters a and b were calculated from the data CC1a. Systematic error was generated by varying the size of the parameters a and b. If the parameter b has a value of less than 1, then low true values are overestimated and, respectively, high true values are underestimated.

2.5.1 Independent Errors from the Log-Normal Distribution

The errors were generated into the stand characteristics (D, H, Age, G and N) from the log-normal distribution. The variation of the measurement errors of the stand variables was obtained from the stand checking data. The errors of the stand characteristics were assumed to be independent. Random variates (x) were simulated for each variable from the normal distribution ($X \sim N(\mu,\sigma^2)$). Then the errors ($Y = e^x$) had the lognormal distribution with p.d.f (Flewelling and Pienaar 1981):

$$f_y(y) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma y} \exp\left[-\dfrac{(\ln y - \mu)^2}{2\sigma^2} \right], & 0 \le y \le \infty \\ 0, & \text{otherwise} \end{cases} \qquad (5)$$

Error generations were made 100 times for each tree species within the stand.

2.5.2 Errors from the Multivariate Log-Normal Distribution

First the variances of the error distributions of the stand characteristics were obtained from the testing data. The specieswise errors of the mean diameter at breast height, mean height and mean age of the stand were supposed to be dependent. Thus, the specieswise errors were generated using a joint distribution function. Furthermore the errors of the basal areas of the tree species within a stand were supposed to be dependent. Therefore a joint distribution was also used in generating the errors of the basal areas. A random n-variate normal vector $X = (X_1,\ldots,X_n)$ has a multinormal distribution $N(\mu,\Omega)$ if the p.d.f. is given (Rubinstein 1981)

$$f_x(x) = \dfrac{1}{(2\pi)^{n/2}|\Omega|^{1/2}} \exp\left[-\dfrac{1}{2}(x-\mu)^T \Omega^{-1}(x-\mu) \right] \quad (6)$$

where $\mu = (\mu_1,\ldots,\mu_i,\ldots,\mu_n)$ with $\mu_i = E(X_i)$ and Ω is the covariance (n x n) matrix (7)

$$\Omega = \begin{vmatrix} \sigma_{11}\sigma_{12}\ldots\sigma_{1n} \\ \sigma_{21}\sigma_{22}\ldots\sigma_{2n} \\ \vdots \\ \sigma_{n1}\sigma_{n2}\ldots\sigma_{nn} \end{vmatrix} \qquad (7)$$

The multivariate distribution was computed using Choleski's technique (Johnson 1987, Rubinstein 1981). Random variates were retrieved from the multivariate normal distribution and the final errors were produced from the log-normal distribution by multiplying the value of the stand characteristics with the exponent of the estimated error. The joint distribution used, i.e. multivariate log-normal distribution, was computed with standard software library IMSL.

The use of the multivariate log-normal distributions produces errors the variation of which derives from the reference data. However, it is possible to increase or decrease the variation of

the errors by multiplying covariances and variances with a constant (Lappi 1993). The errors of the mean tree values and the specieswise errors of the basal areas were supposed to be homoscedastic. These two error groups were also assumed to be mutually independent.

The error distributions of the median tree characteristics were simulated for each tree species within the stand. The errors of the basal areas of the tree species were generated concurrently using a joint distribution. The errors of the specieswise stem numbers were simulated from the log-normal distribution. Error generations were made 100 times for each tree species within the stand.

2.6 Comparison of the Methods

The test criteria used in the comparison of the quality of the estimation of the stand characteristics (G, D, H) and predicted stand characteristics (stand volume per hectare and mean sawn wood volume) were relative bias (%), standard deviation of the prediction errors (s_b) and relative RMSE (root mean square error) (Formulas 8, 9 and 10). The correlation of the errors of the stand characteristics were also studied.

$$bias_\% = 100 \frac{1}{n} \sum_{i=1}^{n} \left[\left(\overset{\wedge}{Y_i} - Y_i \right) / Y_i \right] \qquad (8)$$

$$s_b = \frac{1}{n} \sum_{i=1}^{n} \sqrt{\left(e_i - bias_\% \right)^2} \qquad (9)$$

$$RMSE = \sqrt{s_b^2 + bias_\%^2} \qquad (10)$$

where

Y_i the true stand characteristics
$\overset{\wedge}{Y_i}$ the predicted stand characteristics
e_i the relative prediction error (%) in stand i

Table 2. The correlation coefficients of some forest stand characteristics. For variable codes, see Table 1.

| | Correlation between | | |
	D-H	D-Age	H-Age
Pine	0.813**	0.546**	0.403**
Spruce	0.726**	0.394**	0.322**
Deciduous trees	0.628**	0.241**	0.240**

** Correlation is significant at the 0.01 level

3 Results

In the reference data (CC1a) the specieswise assessment errors of the variables of the basal-area diameter tree were correlated (Table 2). The correlation between the errors of the basal areas of the tree species within a stand was not so obvious (pine–spruce: –0.306; pine–deciduous trees: –0.08; spruce–deciduous trees –0.293). The dependence of these errors allowed the generation of specieswise multivariate normal distributions for basal area median tree characteristics and also for the errors of the specieswise basal areas.

The specieswise correlations of the observed errors of the basal area median tree characteristics in the stand-level inventory showed that these errors were dependent (Table 3). This dependency was also retained in estimates of variances when using multivariate distributions and when using the 1nn-method for all tree species and the empirical errors method for pine.

The estimates of the specieswise relative biases and standard deviations of the differences of the basal area were quite alike with all the simulation methods in data set CC1b (Fig. 2). In the observed errors there was a tendency to overestimate small basal areas and underestimate large basal areas in all tree species. The use of trend in both Monte Carlo methods brought estimates of the biases clearly closer to the observed biases.

In data set CC2 the estimates of the relative standard deviations of the basal area were quite alike with all simulation methods (Fig 3). The largest exception occurred when the empirical errors method was used in error simulation of the large basal area of spruce. The difference came from the composition of the reference data. The tendency of the surveyor to overestimate the

Table 3. Correlation coefficients of some specieswise errors of the stand characteristics of the basal area median tree with stand-level inventory (observed correlations) and with different error simulation methods. For variable codes, see Table 1.

| | | Correlation between | | | | | | | |
| | | D-H | | | D-Age | | | H-Age | |
CC1b	Pine	Spruce	Birch	Pine	Spruce	Birch	Pine	Spruce	Birch
Stand-level inventory	0.751**	0.897**	0.344**	0.560**	0.665**	0.278*	0.466**	0.612**	0.266*
Empirical errors method	0.638**	0.669**	0.707**	–0.209**	0.189**	0.059	–0.07	0.282**	0.245**
1nn-method	0.643**	0.737**	0.693**	0.360**	0.584**	0.304**	0.445**	0.438**	0.451**
Multivariate distrib. without trend	0.578**	0.438**	0.629**	0.110**	0.299**	0.063**	0.271**	0.299**	0.044*
Log-normal distrib. without trend	0.111**	0.060	0.196**	–0.033	0.070	0.109**	–0.055	0.046	–0.005
Multivariate distrib. with trend	0.582**	0.711**	0.651**	0.126**	0.299**	0.043*	0.273**	0.299**	0.038*
Log-normal distrib. with trend	0.104**	0.011	0.171**	–0.020	0.067	0.076*	–0.053	0.045	–0.006
CC2									
Stand-level inventory	0.646**	0.876**	0.775**	–0.150	0.626**	0.401*	0.011	0.646**	0.624*
Empirical errors method	0.605**	0.582**	0.651**	0.224**	0.382**	0.185**	0.220**	0.297**	0.206**
1nn-method	0.665**	0.859**	0.382	0.299**	0.719**	0.260*	0.500**	0.746**	0.018
Multivariate distrib. without bias	0.558**	0.733**	0.571**	0.259**	0.298**	0.236**	0.509**	0.343**	0.200**
Log-normal distrib. without bias	0.067	0.138**	0.053	0.119**	0.144**	0.191**	0.122**	0.191**	0.059
Multivariate distrib. with bias	0.581**	0.737**	0.605**	0.250**	0.291**	0.115**	0.516**	0.343**	0.130**
Log-normal distrib. with bias	0.093*	0.143**	0.074	0.118**	0.141**	0.190**	0.124**	0.079	0.058

*   Correlation is significant at the 0.05 level
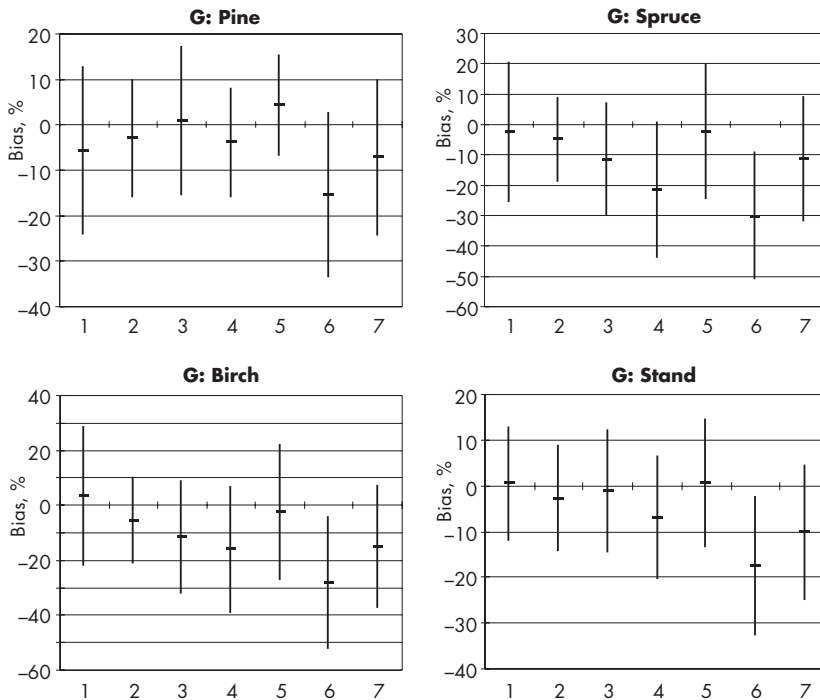** Correlation is significant at the 0.01 level

Fig. 2. Relative biases and relative standard deviations of the residuals of the basal area (G) in compartment checking data CC1b for pine, spruce, birch and stand with different simulation methods.

(1 = Compartment inventory; 2 = 1nn-method; 3 = Empirical errors method; 4 = Log-normal distributions with trend; 5 = Log-normal distributions, no trend; 6 = Multivariate log-normal distributions with trend; 7 = Multivariate log-normal distributions, no trend)
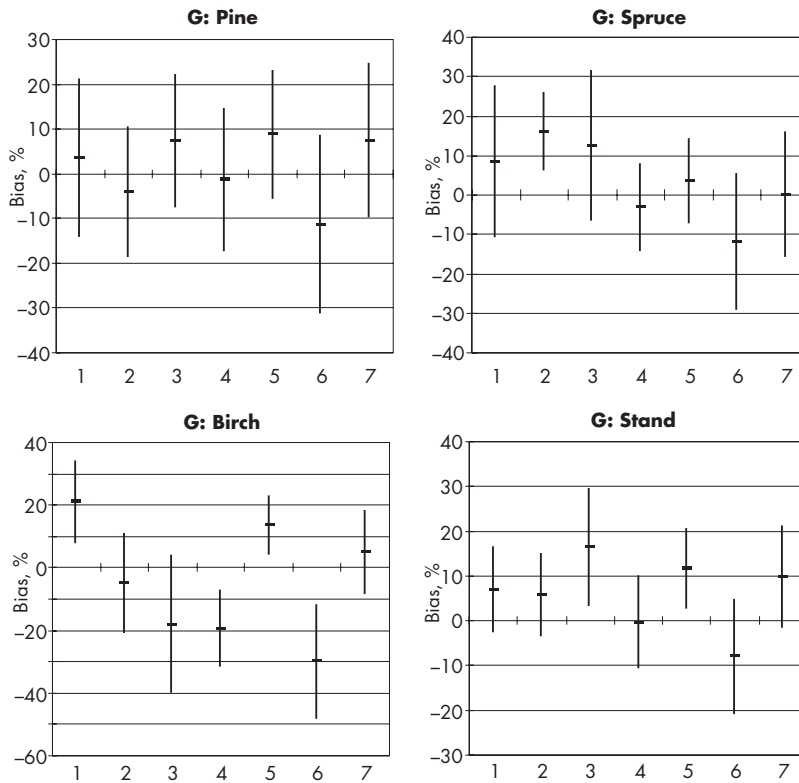
Fig. 3. Relative biases and relative standard deviations of the residuals of the basal area (G) in compartment checking data CC2 for pine, spruce, birch and stand with different simulation methods.

(1 = Compartment inventory; 2 = 1nn-method; 3 = Empirical errors method; 4 = Log-normal distributions; 5 = Log-normal distributions and correction of bias; 6 = Multivariate log-normal distributions; 7 = Multivariate log-normal distributions and correction of bias)

smaller basal areas and underestimate the larger basal areas was detected only with spruce. One reason for this is that spruce was the dominant species and pine and birch were mostly dominated in data set CC2. The total stand basal areas were clearly overestimated with small basal areas and clearly underestimated with large basal areas. These kinds of biases were also achieved with the 1nn-method, empirical errors methods and both Monte Carlo methods. Including the bias in the simulations produced data that with respect to accuracy was similar to that obtained in the standwise inventory.

The classwise observed errors of the median basal area diameter classes showed that there was considerable variation between diameter classes and tree species with both test areas CC1b and CC2. When including the bias, with both Monte Carlo methods the results were similar to the observed errors when the diameter classes were large. However, with small diameter classes simulation results varied. The standard deviation and the bias of the median diameter were small for birch in the CC1b data set. This could be noted with all Monte Carlo methods. Small diameter classes were overestimated when the tree species was dominant, e.g. pine in data set CC1b and spruce in data set CC2. Correspondingly, in small diameter classes the median diameters of the dominated tree species, e.g. spruce in data set CC1b and pine in data set CC2, were underestimated.

The observed relative standard errors of the derived stand characteristics were quite high in the test data for both test areas (Table 4). In the data set CC1b the estimates of the standard errors

Table 4. The relative errors and biases of the derived stand characteristics of the stand-level inventory in stand checking data CC1b and CC2.

	Stand volume per hectare RMSE (%)	Stand volume per hectare Bias (%)	Mean sawn wood RMSE (%)	Mean sawn wood Bias (%)
CC1b				
Data including model errors	4.01	1.50		
Stand-level inventory	31.10	–2.80	77.01	–7.51
1nn-method	38.40	–0.64	78.90	–6.81
Empirical errors method	38.06	–1.38	85.00	6.69
Multivariate distribution without trend	47.14	–11.97	113.10	–18.22
Log-normal distribution without trend	44.05	1.32	110.31	25.84
Multivariate distribution with trend	45.79	–9.16	120.10	–8.76
Log-normal distribution with trend	38.81	4.37	121.89	43.91
CC2				
Data including model errors	5.30	2.70	18.50	9.70
Stand-level inventory	34.34	19.56	82.20	59.20
1nn-method	38.33	15.46	79.69	34.22
Empirical errors method	56.50	27.61	94.98	48.70
Multivariate distribution without bias	36.86	–2.92	77.79	74.65
Log-normal distribution without bias	35.41	7.79	98.78	84.75
Multivariate distribution with bias	47.96	23.21	126.74	21.62
Log-normal distribution with bias	42.81	23.45	133.71	50.48

were 7–16 per cent larger than the observed errors with all simulation methods. When the trend in stand characteristics was used the estimates of the errors approached the observed errors. The estimates of the volume biases differed from the observed errors considerably with multivariate distributions without trend. The estimate of the error of the sawn wood differed the least with the 1nn-method. The estimates of the biases of the sawn wood were closest to the observed biases with the 1nn-method and the Monte Carlo method when using multivariate distributions.

In the CC2 data set the estimates of the error variances of the stand volume were closest to the observed variances with two Monte Carlo methods without bias (Table 4). The relative bias was closest to the bias made by the surveyor when using the 1nn-method and multivariate distribution and including bias. The use of the bias produced stand volume biases clearly closer to observed biases.

For pines and spruces, in all the simulation methods with data set CC1b the specieswise variations of the stand characteristics were quite close to the observed variations in stand-level inventory (Table 5). Only the errors of the median diameter

of the birch, which were small, produced considerable differences between the methods. The log-normal distributions methods could only note the difference between small diameter errors and remarkably larger height errors. The estimates of errors of the specieswise stand characteristics followed the variation of the observed errors in data set CC2 very well. The variations of the errors of the derived specieswise volumes were rather similar to all the simulation methods in both data sets. For data set CC1b, the biases of the specieswise stand characteristics were closest to the observed biases in the stand-level inventory with both Monte Carlo methods without trend and for data set CC2 they came closest with both Monte Carlo methods with bias correction.

4 Discussion

This study deals with the simulation of erroneous stand-level inventory data for further use in, for example, studies on the consequences of using this kind of stand data in a planning context. In simulation experiments the considered methods for error generation were the 1nn-method, the

Table 5. The specieswise errors (RMSE-%) of the basal area median tree characteristics and stand volume of the stand-level inventory and of the simulations in stand checking data CC1b and CC2. For variable codes, see Table 1.

	Pine			Spruce			Birch		
	D	H	Vol	D	H	Vol	D	H	Vol
CC1b									
Data including model errors			6.20			7.30			4.31
Stand-level inventory	29.85	19.96	43.62	36.64	28.36	78.97	3.87	21.78	86.55
1nn-method	20.93	21.45	46.09	27.78	26.44	65.69	30.35	25.64	55.72
Empirical errors method	17.82	18.17	57.70	23.68	23.38	71.43	29.48	24.78	57.17
Multivariate distrib. without trend	27.60	22.74	60.62	34.34	34.54	70.27	4.71	2.87	60.63
Log-normal distrib. without trend	29.99	19.33	44.39	37.03	27.75	77.21	5.73	22.14	82.54
Multivariate distrib. with trend	27.10	22.54	59.71	33.39	22.54	59.71	7.99	2.21	56.43
Log-normal distrib. with trend	28.59	19.03	42.51	33.40	27.36	68.47	8.54	21.67	67.90
CC2									
Data including model errors			6.50			6.31			17.49
Stand-level inventory	45.77	14.40	60.65	32.26	21.47	35.79	37.35	18.63	68.91
1nn-method	15.46	10.40	49.64	23.46	17.90	62.26	18.01	10.38	46.44
Empirical errors method	12.68	11.95	63.92	12.73	10.39	95.72	15.37	12.08	79.09
Multivariate distrib. without bias	35.55	28.21	52.81	29.47	27.26	52.80	33.03	21.58	51.39
Log-normal distrib. without bias	42.26	13.95	53.14	33.61	20.64	48.86	35.52	16.85	46.00
Multivariate distrib. with bias	48.62	34.00	75.50	32.04	28.00	62.07	50.27	27.08	72.09
Log-normal distrib. with bias	50.49	14.68	58.60	36.08	20.93	58.60	38.94	20.83	79.95

empirical errors method and the Monte Carlo method with log-normal and multivariate log-normal error distributions.

The Monte Carlo method with multivariate error distributions was found to be the most flexible simulation method. The method produced the required error variance with relations between the errors of the median basal area tree characteristics. However, if the reference data are extensive the 1nn-method and in certain conditions also the empirical errors method offer useful tools for producing error structures which can be expected in reality.

For all tree species, correlations between the errors of the basal area median diameter tree characteristics were positive in the stand-level inventory. This was expected because the median tree characteristics are measured from the same tree. Ståhl (1992) also found significant correlation between these variables. Correlations between the basal areas of the tree species were slightly negative. Thus the errors of the basal area of one tree species were opposite to the other tree species within the stand.

The 1nn-method and empirical error method produce error structures which are expected in practise; the relations between the errors of the different tree and stand characteristics can be kept reasonable because the reference data are chosen from among genuine samples. The use of both methods is very easy; they do not require any assumptions about error distributions or relations between the errors of the stand characteristics. However, when using the methods it is necessary that the reference data be as large as in this study. The error variation has to be the same in the reference data and in real conditions, otherwise the variation of the estimated errors does not correspond to reality. The number of nearest neighbours is restricted to one. However, the variation of the most similar neighbours can be added using the random parameter in distance functions.

The reference data used in this study was found to be adequate for the purposes of this study. It is not certain if the more extensive reference data would have given better simulation results with both the 1nn-method and the empirical errors methods. The specieswise variation between the two methods and the stand-level inventory originates mostly in the composition of the reference data.

The structure of the errors of the stand-level

inventory varies with the size of a stand. Data from small stands usually have better precision than data from large stands (Poso 1983, Ståhl 1992, Pigg 1994). The size of the stand also has an influence on the calculation of the true values if the sampling error is not properly noted when the amount of the checking plots is defined (e.g. Laasasenaho and Päivinen 1996). The use of the size of a stand can also be tested as a variable in distance measure in extensive reference data.

The error variances estimated with both Monte Carlo methods were quite similar to the observed variance in the stand-level inventory. However, this demanded specific information about error variances of the stand characteristics of the test areas. In complex systems like stand simulation models the use of independent errors in stand characteristics frequently produces unrealistic relationships between the stand characteristics. Thus the errors of the derived tree characteristics can be surprisingly high. The use of multivariate log-normal distributions allowed taking the relations between the errors of the tree characteristics of the basal area median diameter tree into account. If the assumptions of the joint distributions of the errors are adequate and correct then the Monte Carlo methods offer a useful and flexible tool for simulating erroneous inventory data for further studies.

In Monte Carlo methods the systematic errors of the stand characteristics can be taken into account by adding a simple systematic error term in the simulation of the erroneous inventory data. However, the magnitude of the bias has to be known before it can be taken into account. The direction of the bias of the stand and tree characteristics may also vary (e.g. Päivinen et al. 1992). The use of trend can be very worthwhile when there is a tendency to overestimate small stand characteristics and underestimate large ones (Ståhl 1992). In the 1nn-method, and to a lesser extent in the empirical errors method, the trend is included implicitly.

The precision of check assessments varied remarkably between surveyors in checking data CC1. However, the influence of the surveyor was not considered in this study. In practise the same surveyor measures the stands from the same area and information about the surveyor is mostly available. Thus, it is possible to make a personal error model for the surveyor and calibrate the multivariate distribution with the surveyor's error variance if there is suitable checking material available. It is also possible to make personal multivariate distribution error models for each surveyor and use these regionally in simulations. However, because of the lack of suitable reference data it would be advisable to categorize reference data as a beginner surveyor or an experienced surveyor. The influence of the surveyor can be noted easily with the 1nn-method and with the empirical errors method by categorizing the reference data or using surveyor or experience of the surveyors as a distance function variables.

If the error structure of the standwise inventory varies considerably regionally there is a need for regional error models. However, earlier studies (e.g. Laasasenaho and Päivinen 1986) have shown that variation between regions is similar at least for homogeneous stands. The influence of the dominance of the tree species on assessment errors also needs further study.

The magnitude of the inventory errors is needed for the calibration of the error distributions. The use of normal variances of the errors from earlier studies produces an error source which can influence assumptions of the simulation studies at least when studying forestry units.

This study dealt with subjective inventory methods. However, the results can be utilized in objective inventory methods studies in which the errors are correlated. The variation of the study data was rather high. If the composition of the study data had been different it would have affected the results at least with respect to the 1nn-method and empirical errors method. When utilizing these methods it would be useful to have reference data and target data that were as similar as possible.

Acknowledgements

References

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

Altman, N.S. 1992. The introduction to kernel and nearest neighbour nonparametric regression. American Statistician 46: 175–185.

Argantesi, F. & Olivi, L. 1976. Statistical sensitivity analysis of a simulation model for the bio-mass-nutrient dynamics in aquatic ecosystems. Proceedings, 4th Summer Computer Simulation Conference. Simulation Council, La Jolla, Ca. p. 389–393.

Bitterlich, W. 1984. The relascope idea. Commonwealth Agricultural Bureaux. Farnham Royal. 242 p.

Burges, S.J. & Lettenmaier, D.P. 1975. Probabilistic methods in stream quality management. Water Resources Bulletin 11: 115–130.

Flewelling, J.W. & Pienaar, L.V. 1981. Multiplicative regression with lognormal errors. Forest Science 27(2): 281–289.

Gertner, G. 1987. Approximating precision in simulation projections: an efficient alternative to Monte Carlo methods. Forest Science 33: 239–244.

— , Cao, X. & Zhu, H. 1995. A quality assessment of a Weibull based growth projection system. Forest Ecology and Management 71: 235–250.

Härdle, W. 1989. Applied nonparametric regression. Cambridge University, Cambridge. 333 p.

Hyyppä, J., Hyyppä, H., Inkinen, M., Engdahl, M., Linko, S. & Zhu, Y.-H. 2000. Accuracy comparison of various remote sensing data sources in the retrieval of forest stand attributes. Forest Ecology and Management 128: 109–120.

Johnson, M.E. 1987. Multivariate Statistical Simulation. Wiley, New York. 240 p.

Kangas, A. 1996. On the bias and variance in tree volume predictions due to model and measurement errors. Scandinavian Journal of Forest Research 11: 281–290.

— 1999. Methods for assessing uncertainty of growth and yield predictions. Canadian Journal of Forest Research 29(9): 1357–1364.

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.

Kleijnen, J. & van Groenendaal, W. 1992. Simulation: A Statistical perspective. John Wiley & Sons, England. 241 p.

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

— & Päivinen, R. 1986. Kuvioittaisen arvioinnin tarkastamisesta. Summary: On the checking of inventory by compartments. Folia Forestalia 664. 19 p.

Lappi, J. 1993. Metsäbiometrian menetelmiä. University of Joensuu. Silva Carelica 24. 190 p.

Mähönen, M. 1984. Kuvioittaisen arvioinnin luotettavuus. Pro gradu -työ. University of Helsinki, Department of Forest Resource Management. 56 p. (In Finnish)

Mäkelä, A. 1988. Performance analysis of a process-based stand growth model using Monte Carlo techniques. Scandinavian Journal of Forest Research 3: 315–331.

Maltamo, M. 1998. Basal area diameter distribution in estimating the quantity and structure of growing stock. University of Joensuu, Faculty of Forestry. 43 p.

McRoberts, R.E., Hahn, J.T., Hefty, G.J. & Van Cleve, J.R. 1994. Variation in forest inventory field measurements. Canadian Journal of Forest Research 24: 1766–1770.

Mowrer, H.T. 1991. Estimating components of propagated variance in growth simulation model projections. Canadian Journal of Forest Research 21: 379–386.

— & Frayer, W.E. 1986. Variance propagation in growth and yield projections. Canadian Journal of Forest Research 16: 1196–1200.

Mykkänen, R. 1986. Weibull-funktion käyttö puuston läpimittajakauman estimoinnissa. University of Joensuu, Faculty of Forestry. 80 p. (In Finnish)

Nersten & Næsset 1992. Nøyaktighet av bestandtaksering med relaskop. Accuracy of standwise relascope survey. Medd. Skogforsk. 45: 1–22. (In Norwegian with English summary)

Päivinen, R., Nousiainen, M. & Korhonen, K. 1992. Puustotunnusten mittaamisen luotettavuus. Summary: Accuracy of certain tree measurements. Folia Forestalia 787. 18 p.

Pigg, J. 1994. Keskiläpimitan ja puutavaralajijakauman sekä muiden puustotunnusten tarkkuus Metsähallituksen kuvioittaisessa arvioinnissa. Metsänarvioi-

mistieteen pro gradu -työ. University of Helsinki, Department of Forest Resource Management. 92 p. (In Finnish)

Poso, S. 1983. Kuvioittaisen arvioimismenetelmän perusteita. Summary: Basic features of forest inventory by compartments. Silva Fennica 17: 313–343.

Pussinen, A. 1992. Ilmakuvat ja Landsat TM -satelliittikuva välialueiden kuvioittaisessa arvioinnissa. MSc. Thesis, Faculty of Forestry, University of Joensuu. 48 p. (In Finnish)

Rubinstein, R.Y. 1981. Simulation and the Monte Carlo Method. John Wiley & Sons, New York. 278 p.

Ståhl, G. 1992. En studie av kvalitet i skogliga avdelnigsdata som imsamlats med subjektiva inventeringsmetoder. Sveriges Landbruksuniversitet, Institutionen för biometri och skogsindelning. Rapport 24. 128 p. (In Swedish)

Veltheim, T. 1991. Pituusmallit männylle, kuuselle ja koivulle. Metsänarvioimistieteen pro gradu -työ. Helsingin yliopisto [1987]. In: Mäkelä, H. & Salminen, H. (eds.). Metsän tilaa ja muutoksia kuvaavia puu- ja puustotunnusmalleja. Metsäntutkimuslaitoksen tiedonantoja 398. p. 32–34. (In Finnish)

Total of 34 references