

Analysing the Agreement Between an Airborne Laser Scanning Based Forest Inventory and a Control Inventory – a Case Study in the State Owned Forests in Finland

Tarja Wallenius, Risto Laamanen, Jussi Peuhkurinen, Lauri Mehtätalo and Annika Kangas

Wallenius, T., Laamanen, R., Peuhkurinen, J., Mehtätalo, L. & Kangas, A. 2012. Analysing the agreement between an Airborne Laser Scanning based forest inventory and a control inventory – a case study in the state owned forests in Finland. *Silva Fennica* 46(1): 111–129.

Airborne laser scanning based forest inventories have recently shown to produce accurate results. However, the accuracy varies according to the test area and used methodology and therefore, an unambiguous and practical quality assessment will be needed as a part of each inventory project. In this study, the accuracy of an ALS inventory was evaluated with a field sampling based control inventory. The agreement between the ALS inventory and the control inventory was analysed with four methods: 1) root mean square error (RMSE) and bias, 2) scatter plots with 95% confidence intervals, 3) Bland-Altman plots and 4) tolerance limits within Bland-Altman plots. Each method has its own special features which have to be taken into account when the agreement is analysed. The pre-defined requirements of the ALS inventory were achieved. A simplified control inventory approach with a slightly narrower focus is proposed to be used in the future. The Bland-Altman plots with the tolerance limits are proposed to be used in quality assessments of operational ALS inventories. Further studies to improve the efficiency of quality assessment are needed.

Keywords quality assessment, forest inventory, airborne laser scanning

Addresses *Wallenius* and *Laamanen*, Metsähallitus, P.O. Box 94, FI-01301 Vantaa, Finland; *Peuhkurinen*, Oy Arbonaut Ltd, Helsinki, Finland; *Mehtätalo*, University of Eastern Finland, School of Forest Sciences, Joensuu, Finland; *Kangas*, University of Helsinki, Department of Forest Sciences, Helsinki, Finland **E-mail** tarja.wallenius@metla.fi

Received 2 May 2011 **Revised** 21 October 2011 **Accepted** 28 December 2011

Available at <http://www.metla.fi/silvafennica/full/sf46/sf461111.pdf>

1 Introduction

In a traditional compartmentwise forest inventory, a surveyor visits every compartment and measures 2–8 angle count plots at subjectively selected locations in each compartment (Koivuniemi and Korhonen 2006, Kangas and Maltamo 2006). From each plot, the mean tree is visually defined and its diameter and height are measured. Recently, large inventory projects have been replaced with a continuous updating of the compartmentwise data as a part of operational harvest and silvicultural planning. The accuracy of the compartmentwise inventory data is analysed by comparing the assessments to control field measurements (e.g. Haara and Korhonen 2004, Koivuniemi 2003, Pigg 1994, Laasasenaho and Päivinen 1986).

Airborne laser scanning (ALS) based forest inventories have recently been shown to produce accurate results with reasonable costs (e.g. Næsset et al. 2004, Suvanto et al. 2005, Packalén and Maltamo 2006, Uuttera et al. 2006, Packalén and Maltamo 2008). In these studies an area-based approach has been used. In an area-based ALS inventory, a set of field sample plots are measured (reference plots). Using the field sample and the features extracted from the ALS data, the stand characteristics are first predicted for primary prediction units (plots or grid cells). A grid consists of square cells with a fixed length of a side that varies between different studies. Several techniques have been used for predicting: for example regression techniques (e.g. Næsset 2004), k-most similar neighbour techniques (e.g. Packalén and Maltamo 2006) or Sparse Bayesian estimation (Junttila et al. 2008). Then, using the estimates of the primary prediction units, the stand characteristics can be generalised to larger area units like compartments.

In the Finnish state forest enterprise Metsähallitus, compartmentwise inventory has been the prevailing data acquisition method. The quality of the data has been analysed in the late 1990's using a control inventory. Compartmentwise method is, however, considered laborious and expensive. Therefore, Metsähallitus aims at improving the efficiency of operational planning with the help of ALS-based inventories. The first ALS-based forest inventory of a large area was carried out

in 2008 in the Kuhmo district in eastern Finland. The inventory covered an area of approximately 50 000 hectares. The objective of the inventory was to produce forest resource data first for developing the operational harvest planning process and the planning tools, and for actual operational harvest planning as well. Metsähallitus outsourced the inventory to a private commercial service provider. A contract of a complete ALS-based inventory was made with a company providing ALS inventory services. According to the contract, the inventory included laser scanning, aerial photography, reference plot sampling and plot measurements, delineation of microsegments, modelling, prediction of stand estimates for a grid with a cell size of 16 m × 16 m and finally generalisation of estimates for microsegments. Microsegments with a mean size from 0.6 ha to 0.8 ha here refer to a result of a segmentation algorithm which concentrates on delineating different timber types apart, ignoring management (Leppänen et al. 2008). Metsähallitus was mainly interested in the results for microsegments which were considered to fit best with the operational level planning process. In the contract, the demand for the quality of the estimates was defined so that acceptable root mean square error (RMSE) at plot and compartment levels should not exceed those in the study of Packalén (2006).

Using results of area based ALS research reports for setting targets for quality of operational inventories is to some extent problematic, as the test arrangements especially concerning stand level analysis are often different between studies. For example, in some studies a separate test and modelling data have been used (Packalén and Maltamo 2006), whereas in others the same set has been used both for modelling and testing (Packalén and Maltamo 2007). In some cases, the stand-level results are calculated as a mean of the plots within a stand (e.g. Packalén and Maltamo 2007), in some studies they have been derived from a grid (e.g. Næsset 2004) and in some studies they have been predicted directly for the stands using regression models designed at plot level (Suvanto et al. 2005, Uuttera et al. 2006).

The results on the accuracy have varied a lot between studies and have even been contradictory within studies. For instance, while k-nn proved to be better with respect to the species-specific results, fuzzy classification was better for predict-

ing total volumes (Packalén and Maltamo 2006). Different studies have given different stand level error figures for total volume, varying from 10% (Packalén and Maltamo 2007) to 19% (Holmgren 2004). In some studies where a separate validation data has been used, a bias has been detected (Uutera et al. 2006, Holmgren 2004). The effect of the different study areas, the modelling methods and validation methods on the error estimates remains largely unknown.

Very few evaluations of operational ALS inventories have been reported. Næsset (2004) carried out an evaluation of the first Nordic operational standwise ALS-based inventory (49 000 hectares) in Nordre-Land in south-east Norway. In this evaluation, test plots measured from two locations – the other local (39 plots) and the other 80 km from the inventory area (15 plots) – were used. For total volume, the results were unbiased with both test datasets. The standard deviation of total volume was 6.5% with the local test data and 13.4% with the more distant test data. Næsset (2007) evaluated an operational inventory (6500 hectares) in the Hole area in south-east Norway. Results were reported for two strata: for mature stands with high or medium site the standard deviation was 10.6% and for mature stands with poor site productivity it was 14.0%. No serious bias was detected.

Raaterova (2009) reported of a control measurement study of a pilot ALS inventory carried out jointly by Metsähallitus and the Forestry Centre in Lapland. A subjective selection of 50 stands was carried out aiming at an even distribution of young stands, near mature stands and mature stands in the test material. A grid of 5–10 sample plots was laid in each stand with same centre points as the ALS grid cells. Three different ALS predictions were produced by three different consultants for the inventory area. Results were predicted first for a grid and then generalised for stands. At stand level, RMSEs between 15.3%–16.6% were found for total volume.

In addition to the problems related to comparing RMSE values derived in different test or evaluation arrangements, there are some other problems in using them. Firstly, there is no research on validation of area based ALS inventories in which the results would have been derived to micro-segments, which were the most important level

of inventory results in the Metsähallitus Kuhmo inventory. Secondly, the interpretation of RMSE values is difficult in practical forestry. In setting quality targets for ALS inventories or in evaluating them, unambiguous and easily understandable methods should be used. Kangas and Lappi (2011) have suggested methods such as Bland-Altman plots and tolerance limits for analysing the agreement between two inventories.

The objective of this study is to evaluate the accuracy of the Kuhmo ALS-based forest inventory at microsegment level with a field sampling based control inventory. Different methods are applied for analysing the agreement between the ALS inventory and the control inventory. Another objective was to propose approaches to a simple and practical quality assessment method based on the experiences of this study.

2 Material and Methods

2.1 Study Area

The Kuhmo study area covers 50 403 hectares of commercially managed forest in eastern Finland (Fig. 1). The inventory area is mainly dominated by Scots pine (*Pinus sylvestris*). The share of pine is approximately 73% of total volume. Norway spruce (*Picea abies*) and deciduous trees mostly occur as non-dominant species.

2.2 The ALS-Based Inventory

The LiDAR data were collected using a Leica ALS50-II scanning system operated from Cessna 401 B aircraft. The flying altitude was 2000 meters above ground level, flying speed 120 knots, scan rate 52.1 Hz, pulse rate 58 900 Hz and the field of view $\pm 15^\circ$. These settings resulted a nominal pulse density of ~ 1 pulse per square meter. Due to almost constant cloudy weather, the scanning mission had to be canceled several times. The whole area was scanned with two separate mobilizations to the project area. The first mobilization took place between 23th July and 28th July. During this period 5 scanning flights were done. The second mobilization was between 19th September and

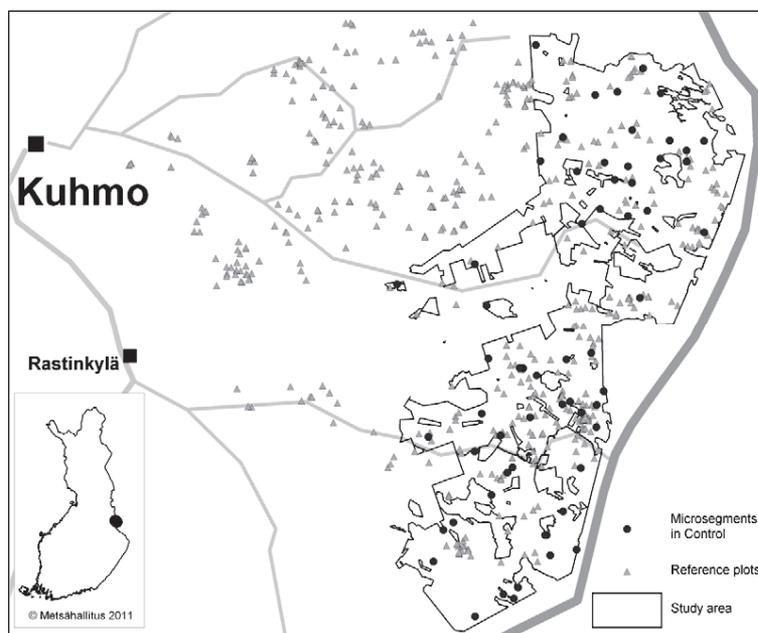


Fig. 1. Location of study area, microsegments in control and reference plots.

22nd September. This period included two flight missions to complete the project. A data quality check was done right after the scanning flight to detect possible data gaps, cloud cover or sensor failures.

The field reference data for modelling was collected between October in 2008 and January in 2009. Totally, 471 reference plots with a fixed radius of 9 meters were measured. Approximately half of the reference plots were located neighboring the study area, because of a collective purchase (see Fig. 1). The plots were allocated to the inventory area by using weighted random sampling. The weights were defined as 85th laser height percentile calculated for 16 meter raster. The weighting was used to allocate more plots to stands with a large average mean tree size. A simple random sample would have resulted only few plots in mature stands, since the development class distribution was left skewed. Every tree inside a plot with a DBH >5 cm was tallied and measured for DBH and species. Besides tally trees, one height sample tree from every species per plot was sampled and measured for height. The measured plot data was used for

calculating the mean height (HgM) weighted by the basal area, the dominant height (Hdom), the mean diameter (DgM) weighted by the basal area, the number of stems (N), the basal area (G) and the volume (V) both as totals for all species and separately for pine, spruce and deciduous trees.

Aerial photography failed because of adverse weather conditions. Therefore, the aerial photographs per annum 2004 were used to extract information about the species proportions. The species proportions were estimated by classifying the image pixels as broadleaved, coniferous or non-vegetation pixels using threshold of image band differences. The field reference data was used in training. Since the aerial images were of inconsistent quality, the threshold values were adjusted image by image.

The area was automatically segmented into microsegments according to laser height, laser density and share of broadleaved trees using the limited region growing algorithm (Leppänen et al. 2008). Laser height refers to the 85th height percentile of the LiDAR pulses, laser density to the percentage of LiDAR pulses returning from vegetation and share of broadleaved trees was the

one estimated from the aerial images. The laser height and density were at first calculated for a 8 meter raster, which were then interpolated to 4 meter rasters. The share of broadleaved species was estimated to 4 meter raster. The aim of the segmentation was to delineate the inventory area into homogenous areas according to timber related variables. The goal was to separate areas from each other if there was more than 2 m²/ha change in G or more than 2 meters change in HgM, or if the main tree species changed. The minimum area of a segment was set to 0.05 hectares. Segments smaller than that were joined to adjacent segments. The final segmentation resulted in a total of 56 477 polygons with a mean size of 0.64 hectares.

Forest parameters (HgM, Hdom, DgM, N, G, V) were estimated both as totals and for pine, spruce and deciduous trees with the Sparse Bayesian method (Junttila et al. 2008). The independent variables were extracted for field reference plots and for 16-meter grid cells, which was the initial estimation area unit. The independent variables were laser height percentiles from 10, 20.... ...,100 percentiles and the proportion of vegetation returns from all returns calculated from first pulse height data, the proportions of pulses with value lower than a threshold value from all pulses calculated from last pulse height data and first and last pulse intensity data and the species proportions estimated from the aerial images. Besides these, 30 laser and aerial image extracted independent variables, the linearizing power transformations of the independent laser variables were used in estimation. The Sparse Bayesian method automatically selects the variables, so that a typical number of independent variables in a model is less than 10. The models were validated and tested for bias at plot-level by using cross-validation. The Sparse Bayesian method estimates models for each dependent variable independently. To obtain logical results, the initial estimates are adjusted so that the variance structure of the dependent variables is preserved. This is done by optimizing the initial results allowing to them corrections, the magnitude of which is proportional to model errors. For example, species-specific volume estimates have to sum up to the total volume estimate. The optimizing algorithm adjusts the total and species-specific

volumes so that the volume rule is met by allowing larger correction to those volumes, typically volumes of minor species, which have larger model errors. The microstand-level estimates were aggregated from a grid cell-level estimates by using area-weighted mean. The cells were cut using the microstand borders. The cells which were smaller than half of the area of a complete cell were merged with adjacent cells before calculation of the independent features.

The quality report was provided in the plot level. The leave-one-out cross validation method was employed to calculate RMSE and bias values.

2.3 Field Control

The target group in this study was defined to be microsegments that have a relevant size for planning purposes. Therefore, the smallest microsegments (those under 0.2 ha, comprising approximately 20% of the total number and 3% of the total area) were left out of sampling. In addition, wastelands, roads, water areas, clear cutting areas and young stands with height less than 4.5 meters were ignored. The size of the target group in the sampling was 27 750 microsegments with a mean size of 0.74 ha.

ALS inventory results were used to stratify the target group into six strata (Table 1). The strata were formed according to two criteria: 1) estimated need of treatments in the near future and 2) estimated tree species mixture. This classification was used because the main use of the data is the timing of the next treatment. In addition, the accuracy of ALS-estimated stand parameter estimates in mixed stands was considered important to be assessed. The need for thinning and regeneration was based on recommendations of the Forestry Development Centre Tapio (2006). Single tree species stands were considered to be those where major species covers over 80% of the stand volume. The sample size was defined to be 60 microsegments due to limited resources. The allocation was resolved so that first a fixed minimum of eight sampled microsegments was set for each stratum. Secondly, some additional microsegments were sampled in mixed species strata, where the need for treatment was found in the ALS inventory. The strata and the number of

Table 1. Number of microsegments in stratum (N_h), proportion of stratum ($W_h = N_h/N$), number of sampled units in stratum (n_h) and proportion of sample in stratum ($w_h = n_h/N_h$).

Stratum	N_h	% (W_h)	n_h	% (w_h)
a) Single tree species, thinning	295	1.1	8	2.7
b) Mixed, thinning	1801	6.5	12	0.7
c) Single tree species, regeneration	666	2.4	8	1.2
d) Mixed, regeneration	1838	6.6	12	0.7
e) Single tree species, no management	14045	50.6	10	0.1
f) Mixed, no management	9105	32.8	10	0.1
Total	N = 27750	100	n = 60	0.2

microsegments to be sampled from each stratum are shown in Table 1.

A systematic network of fixed sized circle field plots with a 9 meter radius was laid on the sampled microsegments. Field plots with the centre point inside the microsegment were measured. The mirage method (Schmid 1969, see Gregoire and Valentine 2008) was used on plots which crossed the edge of microsegment. The distance between field plots varied according to the area of microsegment. The following distances were used: 20 m (area <0.5 ha), 30 m (0.5 ha ≤ area <1.2 ha), 40 m (1.2 ha ≤ area <2.0 ha) and 50 m (area ≥ 2.0 ha). Totally, 483 field plots in 60 microsegments were selected and measured. Thus, on the average, 8 plots per microsegment were measured. The number of plots in a microsegment varied between 5 and 14. The measurements were carried out between 13th May and 26th June 2009. In total, 121 working days equaling to 795 hours in the field were recorded. The time reported excludes traveling time by car.

Tree species and diameter (cm) at breast height were measured for each tree with a diameter over 5 cm. Height (m) was measured for every 7th tree by tree species. A linear mixed model with Näslund's (1936) equation was used to estimate the height for other tally trees. Stem volumes for pine, spruce and birch trees were calculated with species-specific regression models based on diameter and height (Laasasenaho 1982). The volume for other deciduous trees was estimated with the model for birch.

Table 2. Sample ranges and means of different stand characteristics in the control and ALS-based stand parameter estimates.

	Control data		ALS-based stand parameter estimates	
	Range	Mean	Range	Mean
Pine				
V, m ³ ha ⁻¹	9.6–218.6	65.9	14.5–182.1	70.6
G, m ² ha ⁻¹	1.6–23.1	9.8	3.0–19.6	10.5
N, ha ⁻¹	46–1395	580	157–1160	625
d, cm	10.7–37.2	18.0	9.9–33.1	18.0
h, m	7.6–21.3	12.6	8.5–20.4	12.7
Spruce				
V, m ³ ha ⁻¹	0–136.1	21.4	0–118.7	18.5
G, m ² ha ⁻¹	0–19.7	3.5	0–16.5	2.9
N, ha ⁻¹	0–1549	274	0–900	207
d, cm	5.5–37.3	15.5	9.7–22.6	16.2
h, m	3.6–21.4	11.4	8.3–14.8	11.8
Deciduous				
V, m ³ ha ⁻¹	0.1–92.0	17.0	0–69.7	14.2
G, m ² ha ⁻¹	0.0–15.7	3.1	0–9.8	2.2
N, ha ⁻¹	5–1749	396	0–883	204
d, cm	6.1–35.1	12.7	3.1–19.8	12.2
h, m	4.7–20.5	10.8	6.0–15.5	11.5
Total				
V, m ³ ha ⁻¹	23.7–285.8	104.3	39.6–254.4	103.3
G, m ² ha ⁻¹	5.3–35.8	16.4	7.0–32.3	15.6
N, ha ⁻¹	400–2594	1250	497–1990	1035
d, cm	9.1–32.6	16.4	8.5–27.5	17.0
h, m	7.4–20.9	12.1	7.9–17.7	12.5
Hdom, m	9.0–22.7	14.0	11.0–20.0	14.0

2.4 Methods in Agreement Analyses

Firstly, the compliance to the required study of Packalén (2006) was checked by calculating RMSE

$$RMSE = \sqrt{\sum_{h=1}^L \left(W_h \frac{\sum_{i=1}^{n_h} (\bar{Y}_{hi} - \hat{Y}_{hi})^2}{n_h} \right)} \quad (1)$$

where \bar{Y}_{hi} is the mean value of attribute y in microsegment i of stratum h in the control inventory, \hat{Y}_{hi} is the corresponding ALS estimate, n_h is the number of sampled microsegments in stratum h , L is the number of strata and W_h is the proportion of stratum h in the population. However, with Eq. 1 the RMSE values will be overestimated because the control measurements include a sampling error. Assuming that this error is independent of the errors of ALS-based stand parameter estimates, the error variance of the ALS-based stand parameter estimates can be unbiasedly estimated by subtracting the estimated error variance of the control measurements from the MSE (Kangas 2006, Haara and Korhonen 2004, Laasasenaho and Päivinen 1986). Using the estimator of mean for stratified sampling for the microsegment-specific variances of means, the estimated error variance of the control measurements is

$$\begin{aligned} & \sum_{h=1}^L \left(W_h \frac{1}{n_h} \sum_{i=1}^{n_h} \text{var}(\bar{Y}_{hi}) \right) \\ &= \sum_{h=1}^L \left(W_h \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{n_{hi}} \frac{(Y_{hij} - \bar{Y}_{hi})^2}{n_{hi}(n_{hi} - 1)} \right) \end{aligned} \quad (2)$$

where \bar{Y}_{hi} is a mean value of observed attribute y in microsegment i in stratum h , Y_{hij} is an observed attribute y in sample plot j in microsegment i in stratum h and n_{hi} is a number of plots in microsegment i in stratum h . The RMSE without sampling error

$$RMSE_{(-se)} = \sqrt{\sum_{h=1}^L \left(W_h \frac{\sum_{i=1}^{n_h} (\bar{Y}_{hi} - \hat{Y}_{hi})^2}{n_h} \right) - \sum_{h=1}^L \left(W_h \frac{1}{n_h} \sum_{i=1}^{n_h} \text{var}(\bar{Y}_{hi}) \right)} \quad (3)$$

was also calculated. The bias

$$bias = \sum_{h=1}^L \left(W_h \frac{\sum_{i=1}^{n_h} (\bar{Y}_{hi} - \hat{Y}_{hi})}{n_h} \right) \quad (4)$$

was calculated and compared with the bias reported in the quality report of provider. Significance of the bias was tested as well using Student's t -test. Furthermore, RMSE and bias relative to the mean value were calculated and compared to study of Packalén (2006) and to the quality report of the provider.

Secondly, the level of agreement was analysed by creating confidence intervals as Maltamo et al. (2009)

$$\bar{Y}_i \pm t \times SE(\bar{Y}_i) \quad (5)$$

for field measured control attribute y in microsegment i . In Eq. 5, the standard error of observed attribute y in microsegment i is

$$SE(\bar{Y}_i) = \sqrt{\frac{\sum_{j=1}^n (Y_{ji} - \bar{Y}_i)^2}{n_i(n_i - 1)}} \quad (6)$$

where Y_{ij} is a measured value of attribute y in sample plot j in microsegment i and n_i is a number of sample plots in microsegment i . In this study, t -value for 95% confidence was used. The confidence intervals of the observed attributes were plotted against ALS-based stand parameter estimates. Then it was checked how often the confidence intervals crossed the identity line. If the confidence interval covers the identity line, the control inventory does not show an error of statistically significant magnitude for that microsegment.

Thirdly, as suggested by Kangas and Lappi (2011), the Bland-Altman plots were produced to assess the agreement. In the Bland-Altman plots, the difference between the attribute value in the control and the ALS-based estimation method is plotted against the average of control value and ALS estimated value. The mean of the differences is shown with the continuous horizontal line in the plots showing the bias in the sample. The trend in the differences describes the difference between the variances of the control and ALS-based method (e.g. Krümmenauer et al. 2006). The standard deviations were calculated to create the confidence lines for the mean of the differences. If the distribution of the differences is normal, 95% of observations should be between the lines of mean difference $\pm 2SD$.

Finally, the tolerance limits were defined with respect to the mean of control and ALS estimates

and they were combined with the Bland-Altman plots. In this study, the tolerance limit was set to 20% based on an analysis of information needs in operational harvest planning (Laamanen and Kangas 2011). The proportion of plots falling between the defined tolerance limits will directly confirm if the agreement is acceptable.

3 Results

3.1 RMSE and Bias

In Table 3 both the absolute and the relative RMSE values are presented with and without subtracting the sampling error of the control measurements. The same table also includes the RMSE values of the

Table 3. Estimated RMSE values over the inventory area with and without the sampling error at microsegment level in four first columns. RMSE values in the quality report of provider at plot level and in the study of Packalén and Maltamo (2007) at stand level in last four columns.

	RMSE	RMSE%	RMSE -se _i	RMSE% -se _i	RMSE provider	RMSE% provider	RMSE P&M 2007	RMSE % P&M 2007
Pine								
V, m ³ ha ⁻¹	21.8	33.1	17.6	26.6	38.0	49	27.7	28.1
G, m ² ha ⁻¹	3.0	31.1	2.4	24.5	5.1	46	3.3	27.1
N, ha ⁻¹	217	37.4	176	30.4	299	51	232	40.8
d, cm	2.9	15.9	2.4	13.6	2.7	14	3.4	16.9
h, m	1.3	10.6	1.2	9.5	1.3	9	1.4	8.5
Spruce								
V, m ³ ha ⁻¹	13.6	63.3	10.1	47.3	26.1	104	27.0	32.6
G, m ² ha ⁻¹	2.4	68.1	2.0	57.1	3.7	94	3.1	31.3
N, ha ⁻¹	226	82.6	211	77.1	203	78	240	38.1
d, cm	6.2	40.0	6.1	39.0	4.1	24	2.6	20.2
h, m	4.3	37.5	4.2	36.8	2.1	16	2.0	17.6
Deciduous								
V, m ³ ha ⁻¹	11.8	69.4	10.2	59.8	17.0	148	13.7	62.3
G, m ² ha ⁻¹	2.3	74.4	2.1	67.5	2.4	134	1.6	52.5
N, ha ⁻¹	379	95.7	359	90.7	198	99	151	47.6
d, cm	5.0	39.3	4.8	37.7	4.9	34	2.9	25.3
h, m	3.8	35.0	3.7	34.4	2.4	19	2.3	18.4
Total								
V, m ³ ha ⁻¹	15.1	14.5	6.8	6.5	28.2	25	21.1	10.4
G, m ² ha ⁻¹	2.6	15.7	1.7	10.3	3.9	23	2.1	8.6
N, ha ⁻¹	420	33.6	390	31.2	293	28	241	15.9
d, cm	1.6	9.6	1.3	7.7	2.3	12	-	-
h, m	1.1	9.6	1.1	9.1	1.1	8	-	-
Hdom, m	1.3	8.9	-	-	1.1	7	-	-

quality report of the provider and the RMSE values in the study of Packalén and Maltamo (2007). The RMSE values in this study are mostly lower than in the quality report of the provider. The RMSE values in the quality report were calculated for the plot level. However, the RMSE values for diameter and height are almost at the same level or even higher in this study than in the quality report of the provider. The relative RMSE values of most attributes are higher in this study than in the reference study. The reason for this is that the means of stand characteristics in this study are lower than in the reference study. For example, the mean stand volume (see Table 2) in this study is only half of the mean stand volume in the reference study. The absolute RMSE values both with or without the sampling error reductions are mostly lower in this study than in the reference study. Exceptions, marked with bold in Table 3, can be found in the stem numbers of total and deciduous trees, as well as diameter and height of spruce and deciduous trees. In addition, the absolute RMSE for the basal area of deciduous trees is higher in this study than in the reference study.

Both the absolute bias and the relative bias are presented in Table 4. Some slight underestimates (e.g. total volume and basal area) as well as slight overestimates (e.g. mean height and mean diameter) occur. However, the biases are not statistically significant in any of the cases.

3.2 Confidence Intervals

The (95%) confidence intervals were calculated for each attribute in each microsegment. Fig. 2 shows the confidence intervals of total volume, basal area, mean diameter and mean height. The confidence interval of control value for stand volume is not crossing the identity line in 3.3% of the segments. The percentage is 5%, 25% and 11.7% for basal area, mean height, and mean diameter, respectively. The high proportion in mean height and mean diameter would suggest that there is some bias in the estimates, although it may be caused by high RMSE values as well. However, previously calculated RMSE and bias values for the whole inventory show that despite of divergences the errors were not significant, i.e. the results are marginally unbiased. The difference can partly be due to a

Table 4. Estimated bias over the inventory area. Bias values in the quality report of provider in last two columns.

	Bias	Bias %	Bias Provider	Bias % Provider
Pine				
V, m ³ ha ⁻¹	-4.7	-7.1	-3.2	-4
G, m ² ha ⁻¹	-0.6	-5.8	-0.6	-6
N, ha ⁻¹	-34.1	-5.9	-13	-2
d, cm	0.4	2.3	0.1	0
h, m	-0.2	-1.3	0.1	1
Spruce				
V, m ³ ha ⁻¹	4.1	18.9	2.2	9
G, m ² ha ⁻¹	0.9	24.2	0.5	12
N, ha ⁻¹	73.2	26.7	29	11
d, cm	1.5	9.4	-1.5	-8
h, m	1.3	11.6	-0.6	-5
Deciduous				
V, m ³ ha ⁻¹	2.4	14.1	-0.1	-1
G, m ² ha ⁻¹	0.8	27.6	-0.0	-1
N, ha ⁻¹	201.7	50.1	6	3
d, cm	2.2	17.3	-0.8	-6
h, m	1.1	9.8	-0.1	-1
Total				
V, m ³ ha ⁻¹	1.8	1.7	-1.1	-1
G, m ² ha ⁻¹	1.1	6.9	-0.2	-1
N, ha ⁻¹	240.7	19.3	22	2
d, cm	-0.4	-2.3	-0.4	-2
h, m	-0.5	-3.9	-0.1	-1
Hdom, m	0.1	0.5	-0.1	-1

trend in bias, which means that the results are not unbiased for certain subpopulations. Partly it may be due to the fact that the confidence intervals of total volume, basal area, mean height and mean diameter were narrow. The situation was the same with height and diameter of pine (Figs. A3 and A4 in Appendix 1). It must be kept in mind that RMSE and bias are analysed for the whole inventory area. Instead, the analysis with confidence intervals focuses on the sampled microsegments only. Therefore, the interpretation of the results may differ between these two analyses.

3.3 Bland-Altman Plots

The agreement analysed with the Bland-Altman plots can be seen in Fig. 3 and Appendix 2. The

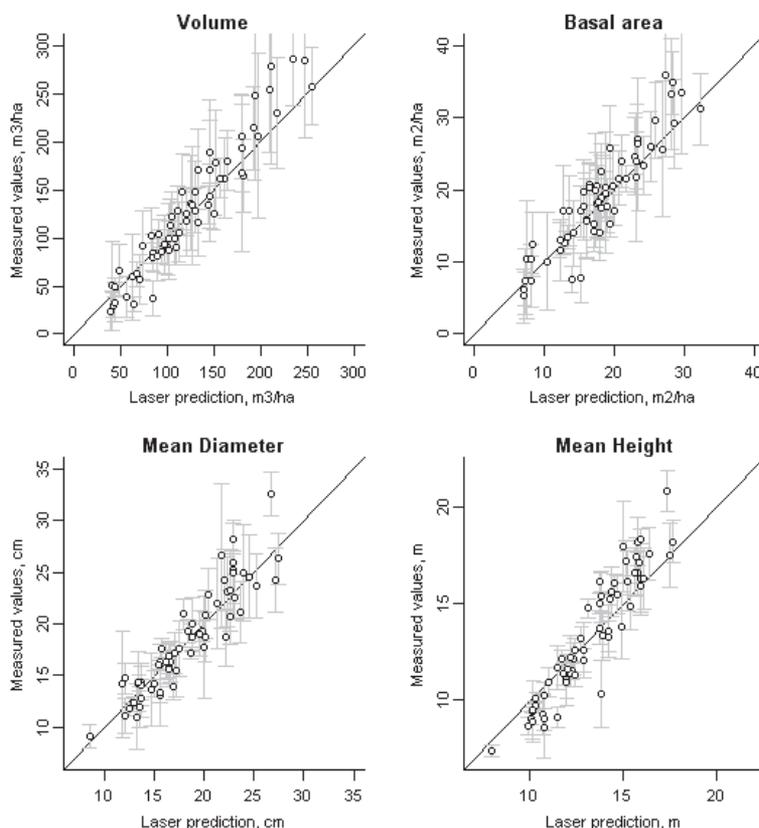


Fig. 2. The confidence intervals of total volume (m³/ha), basal area (m²/ha), mean diameter (cm) and mean height (m). The vertical lines indicate the confidence intervals in microsegments. The confidence interval lines should be intersected the identity line in 95% of estimations.

bias is described by the Bland-Altman plots with a mean of difference line (control – laser). Since the bias here embodies the bias in the sample the results may differ or even be contradictory (e.g. the volume in Fig. A5 and basal area of pine in Fig. A6 in Appendix 2) with the bias calculated with Eq. 4 for the whole inventory. The trend line describes the relationship of (between-microsegment) variances between ALS-based inventory and control measurements. The upward trend line demonstrates that the variance of the control measurements was higher than the variance of ALS-based method, as seen in most plots in Appendix 2. The range of attributes in control measurements is wider in most cases (Table 2). To some extent, the upward trend line shows also trend in bias: it means that

large values of control inventory tend to be underestimated in the ALS inventory and small values of the control inventory tend to be overestimated in the ALS inventory. The opposite trend can be seen in three plots (diameter of spruce, height of spruce and height of deciduous). In the plot of the diameter of deciduous trees, the trend line is horizontal (Appendix 2).

Although results seem to be good in general, those observations outside the $\pm 2SD$ lines are in strata which are important to operational forestry, namely thinning and regeneration (see Fig. 3). A precisely estimated basal area provides a better basis for the timing of thinning and an accurate diameter estimation respectively a better basis for the timing of regeneration.

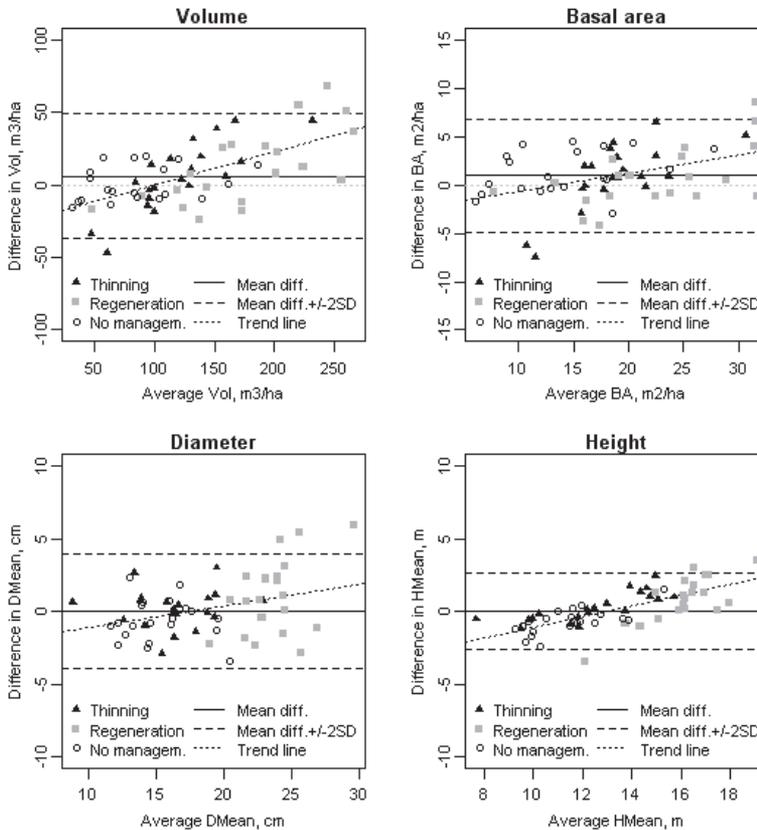


Fig. 3. The Bland-Altman plot for total volume (m³/ha), basal area (m²/ha), mean diameter (cm) and mean height (m). Difference is control – laser. Thinning, regeneration and no management based on laser inventory.

When the tolerance limits are added to the Bland-Altman plots, microsegments achieving the set accuracy limits can be seen immediately (Fig. 4). The 20% tolerance limits were set according to the findings by Laamanen and Kangas (2011), where the majority of interviewed team leaders accepted a 20% error for total volume estimates in harvest planning. The tolerance limits could be absolute values as well. Inspection of the Bland-Altman plots for mean diameter and mean height in Fig. 4 reveals that 95% of observations are between 20% tolerance limits. Furthermore, it can be seen that 72% of volume estimates and 77% of basal area estimates are between the 20% tolerance limits.

4 Discussion

The relative RMSE values of most attributes were higher in this study than in the study of Packalén and Maltamo (2007), however the absolute RMSE values were generally lower. This satisfies the requirements set in the contract for the service provider as the same error level as in the reference study was achieved in the ALS inventory. The inventory results showed a slight underestimate for total volume (1.8 m³/ha) and for basal area (1.1 m²/ha). These biases are nevertheless non-significant and do not cause a serious conflict with the requirements of the contract.

Bias and RMSE are the error terms that are generally used when agreement of two inventories

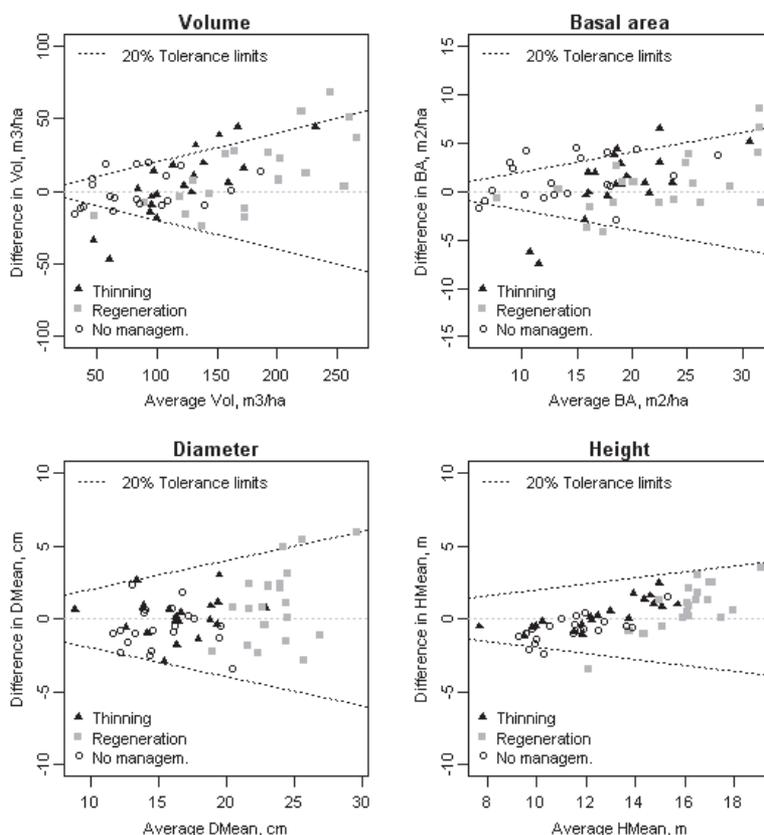


Fig. 4. The Bland-Altman plots with 20% tolerance limits for total volume (m^3/ha), basal area (m^2/ha), mean diameter (cm) and mean height (m). Thinning, regeneration and no management based on laser inventory.

is analysed. Bias is simple and easy to understand as it should always be as close to zero as possible. The RMSE value has been widely used for assessment of inventory methods and it is a well known statistical method. However, interpreting the RMSE values by operational forestry staff may be difficult due the inadequate statistical skills, and therefore, other measures are useful. Utilization and connection of RMSE values in decision making in operational planning should be studied. A problem in analysing agreement is to find relevant references for comparison. Scientific studies have varying research arrangements, they are located in different geophysical conditions and often do not have separate test data for validation.

Even though the RMSE analysis indicates that the ALS inventory fulfills the requirements, the

agreement looks different when analyzed with the confidence intervals. In this study, the 95% confidence intervals were used for observations in control measurements. When the confidence interval crosses the identity line, the ALS estimation corresponds with the control measurement with 95% confidence. As seen in results, 25% of confidence intervals for mean height and 11.7% of confidence intervals for mean diameter do not cross the identity line. The microsegments were delineated according to laser height, laser density and share of broadleaved trees. Therefore, the standard error of mean height was relatively small, which leads to narrow confidence limits and although most estimates were close to the identity line the confidence intervals do not cross the line. In practice some error in ALS estimates

is tolerated and results should rather be checked against tolerance limits which could perhaps be combined with identity lines. However, based on the confidence intervals there seems to be a trend in bias so that the bias is larger with the extreme values of height and diameter, and this may be problematic in applications.

In many cases (see Appendix 1) the problem is that the confidence intervals may be too wide to be useful for analyzing the agreement. The width of the intervals is affected by both the homogeneity of microsegments related to attributes in concern and the number of sample plots in each microsegment. Therefore, an acceptable sampling error at the microsegment level should be defined before the control inventory. Then, it should be possible already in the control field work to calculate the actual need of sample plots in each microsegment to reach the pre-defined error level. However, to reach the acceptable accuracy at tree species level may lead to an unrealistic number of sample plots.

With the Bland-Altman plots, accuracy can be assessed with tolerated differences between methods. The assessment of accuracy is considered to be objective, because neither of the methods is assumed to be the absolute truth (Kangas and Lappi 2011). With the Bland-Altman plots the interpretation is relatively easy after users get used to them. The acceptable levels of standard deviation and the width of the \pm SD lines should be set according to the purpose of use. One problem is that the mean difference represents only the bias in the sample, not in the whole inventory area as the sample proportions varies between strata. In addition, it is difficult to distinguish the bias and the variance in one observation. Combining the tolerance limits with the Bland-Altman plots makes it easier to interpret the results. The tolerance limits and proportion of observations within the tolerance limits could be defined on order of the ALS-based inventory. Thereafter, two questions still remain: how should the observations outside the limits be regarded and how far from tolerance limits should they be allowed to be. The limits and acceptable errors must be linked to the forest management decisions that are made with the inventory information. Errors in an inventory data – especially in estimates of basal area, mean diameter and site index – lead to wrong decisions in forest planning and diminish the quality of planning (Vanhatalo 2010).

In this study, it was deemed necessary to have enough microsegments sampled that are important to operational forest management. Therefore, microsegments were first stratified to six non-overlapping strata according to the need for forest management in near future. Tree species mixture was also used in stratification. Total sample size was strictly limited due to available resources for field measurements. The sampling was allocated to each stratum according to pre-defined minimum of sample size and by giving more weight to the strata with a need for management. The optimal allocation could not be utilised in advance due to lack of prior information of variance. Pre-sampling to resolve the prior information was not used due to limited resources. The optimal allocation should be a topic of further studies. The proportional allocation was rejected because it would have led to inadequate sample size in small strata. Furthermore, equal allocation was not used because the size of the strata varied remarkably. If the population could be stratified to homogeneous strata, stratified random sampling with optimal allocation would probably be the most efficient sampling method. However, when several characteristics are estimated it is not easy to define homogeneous strata for those all.

A quality assessment carried out in this study is time-consuming and expensive. An average of 50 minutes was used per each field plot including the time spent for navigation to the plot. Total field costs in the control inventory were 0.7 € per hectare for the whole study area (50403 hectares). In this study some measurements were carried out which are unnecessary in further field control measurements. In one third of the plots, tree distance from the plot centre point was measured. This data will be used in studies of adequate and convenient size of sample plots. Furthermore, the height of every 7th tree by tree species in every field plot was measured. On the average, this meant about 24 height measurements in each microsegment. The actual need of height sample trees should be analyzed in a separate study. In addition, a more convenient and efficient sampling design should be studied. Cluster sampling of microsegments might cut travel costs. However, with cluster sampling the total sample size may have to be increased to achieve the same precision.

The efficiency of the control inventory can perhaps be improved by measuring only tree diameters. Thereafter, quality assessment of ALS-based forest inventory could be based on analysing agreement of basal area and diameter only. Field measurements would be much easier and faster without height measurements. Height measurements in field may have significant measuring errors (see e.g. Päivinen et al 1992). Even though height is an important parameter in making thinning decisions, the risk of errors and the time consumed for measurements may mean that the cost of controlling height outweighs the benefit.

In the future, quality assessment should be a simple and easy part of an inventory process. A control inventory in the field, like in this study, is expensive, time-consuming and the results are obtained too late in respect of the acceptance of the ALS inventory results. Therefore, other ways to control the quality are needed. A quick and light control inventory could be one solution and it would have to have a much narrower focus than used in this study. Then, Bland-Altman plots with tolerance limits would be the main method for analysing the agreement in operational forestry.

One way to get control data is to analyse the clear-cut stands by comparing the harvester volume measurements and laser inventory predictions made for the stand. This method, even though it is very simple and inexpensive, is very slow and can produce data for mature stands only.

Perhaps the main emphasis should be put on controlling the ALS-based forest inventory process itself. The inventories for large areas may cause challenges. The models may deteriorate because of probable greater variety in large areas. More reference plots may be needed, although resources may be limited due to costs. The effect of sampling method may evolve. Remote sensing may be challenging because several mobilizations are required for large areas and probability of cloudy weather may increase. Detecting those challenges may be worth the effort. A control approach of the process itself should be designed together with the contractor, so that elements of the inventory process that are important from the quality point of view would be analysed in certain steps during the inventory.

References

- Gregoire, T.G. & Valentine H.T. 2008. Sampling strategies for natural resources and the environment. 474 p.
- Haara, A. & Korhonen, K.T. 2004. Kuvioittaisen arvioinnin luotettavuus. *Metsätieteen aikakauskirja* 4/2004: 489–508. (In Finnish).
- Holmgren, J. 2004. Prediction of tree height, basal area and stem volume in forest stands using airborne laser scanning. *Scandinavian Journal of Forest Research* 19: 543–553.
- Junttila, V., Maltamo, M. & Kauranne, T. 2008. Sparse Bayesian estimation of forest stand characteristics from airborne laser scanning. *Forest Science* 54 p. 543–552.
- Kangas, A. & Lappi, J. 2011. On comparing agreement of forest variable estimates obtained with different methods. 50 p. (Unpublished manuscript).
- & Maltamo, M. (eds.) 2006. *Forest inventory: methodology and applications*. 362 p.
- Koivuniemi, J. 2003. *Metsiköihin ja paikannettuihin koealoihin perustuvan kuvioittaisen arvioinnin tarkkuus. The accuracy of the compartmentwise forest inventory based on stands and located sample plots. Helsingin yliopiston metsävarojen käytön laitoksen julkaisuja 36. Dissertation*. 160 p.
- & Korhonen, K.T. 2006. Inventory by compartments. In: *Forest inventory – methodology and applications. Managing forest ecosystems, Vol. 10*. Springer.
- Krummenauer, F., Genevrière, I., & Nixdorff, U. 2000. The biometric comparison of cardiac imaging methods. *Computer Methods and Programs in Biomedicine* 62:21–34.
- Laamanen, R. & Kangas, A. 2011. Large-scale forest owner's information needs in operational planning of timber harvesting – some practical views in Metsähallitus, Finnish state-owned enterprise. *Silva Fennica* 45(4): 711–727.
- Laasasenaho, J. 1982. Taper curve and volume functions for pine, spruce and birch. *Communicationes Instituti Forestalis Fenniae* 108.
- & Päivinen, R. 1986. Kuvioittaisen arvioinnin tarkistamisesta. *Folia Forestalia* 664. p. 19 (In Finnish).
- Leppänen, V. J., Tokola, T., Maltamo, M., Mehtätalo, L., Pusa, T. & Mustonen, J. 2008. Automatic delineation of forest stands from LIDAR data. Conference presentation in GEOBIA 2008 – Pixels,

- Objects, Intelligence: GEOgraphic Object-Based Image Analysis for the 21st Century. p. 6.
- Maltamo, M., Packalén, P., Suvanto, A., Korhonen, K.T., Mehtätalo, L. & Hyvönen, P. 2009. Combining ALS and NFI training data for forest management planning: a case study in kuortane, Western Finland. *European Journal of Forest Research* 128: 305–317.
- Metsätalouden kehittämiskeskus Tapio. 2006. Hyvän metsänhoidon suosituksset. 100 p.
- Næsset, E. 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remot Sensing of Environment* 80: 88–99.
- Næsset, E. 2004. Accuracy of forest inventory using airborne laser scanning: evaluating the first Nordic full-scale operational project. *Scandinavian Journal of Forest Research* 19: 554–557.
- 2007. Airborne laser scanning as a method in operational forest inventory: Status of accuracy assessments accomplished in Scandinavia. *Scandinavian Journal of Forest Research* 22: 433–442
- Næsset, E., Gobakken, T., Holmgren, J., Hyypä, H., Hyypä, J., Maltamo, M., Nilsson, M., Olsson, H., Persson, Å & Söderman, U. 2004. Laser scanning of forest resources: the Nordic experience. *Scandinavian Journal of Forest Research* 19: 482 - 499.
- Näslund, M. 1936. Skogsförsöksanstaltens gallringsförsök I tallskog, Meddelanden från Statens skogsförsöksanstalt 28(1).
- Packalén, P. 2006. Using airborne laser scanning data and digital aerial photographs to estimate forest growing stock. Licenciate thesis. University of Joensuu.
- & Maltamo, M. 2006. Predicting the plot volume by tree species using airborne laser scanning and aerial photographs. *Forest Science* 52: 611–622.
- & Maltamo, M. 2007. The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. *Remote Sensing of Environment* 109:328–341. doi:10.1016/j.rse.2007.01.005.
- & Maltamo, M. 2008. Estimation of species-specific diameter distributions using airborne laser scanning and aerial photographs. *Canadian Journal of Forest Research* 38: 1750–1760.
- Pigg, J. 1994. Keskiläpimitan ja puutavaralajijakauman sekä muiden puustotunnusten tarkkuus Metsähallituksen kuvioittaisessa arvioinnissa. Metsänarvioimistieteen pro gradu -työ. Helsingin yliopisto, metsävarojen käytön laitos. 92 p. (In Finnish).
- Päivinen, R., Nousiainen, M. & Korhonen, K.T. 1992. Puutunnusten mittaamisen luotettavuus. Summary: Accuracy of certain tree measurements. *Folia Forestalia* 787. 18 p.
- Raaterova, J. 2009. Laserkeilausmenetelmän luotettavuus metsävarojen inventoinnissa – tutkimustuloksia Pisa-Louevaaran alueelta. Rovaniemen ammattikorkeakoulun opinnäytetyö. 46 p.
- Suvanto, A., Maltamo, M., Packalén, P. & Kangas, J. 2005. Kuviokohtaisten puustotunnusten ennustaminen laserkeilauksella. *Metsätieteen aikakauskirja* 4/2005: 413–428.
- Utterä, J., Anttila, P., Suvanto, A. & Maltamo, M. 2006. Yksityismetsien metsävaratiedon keruuseen soveltuvilla kaukokartoitusmenetelmillä estimoitujen puustotunnusten luotettavuus. *Metsätieteen aikakauskirja* 4/2006: 507–519.
- Vanhatalo, K.O. 2010. Metsävaratiedon virheiden vaikutus hakkuiden toimenpidesuunnitteluun. Metsänarvioimistieteen pro gradu -työ. Helsingin yliopisto, metsätieteiden laitos. 73 p. (In Finnish).

Total of 30 references

Appendix 1 (Confidence intervals)

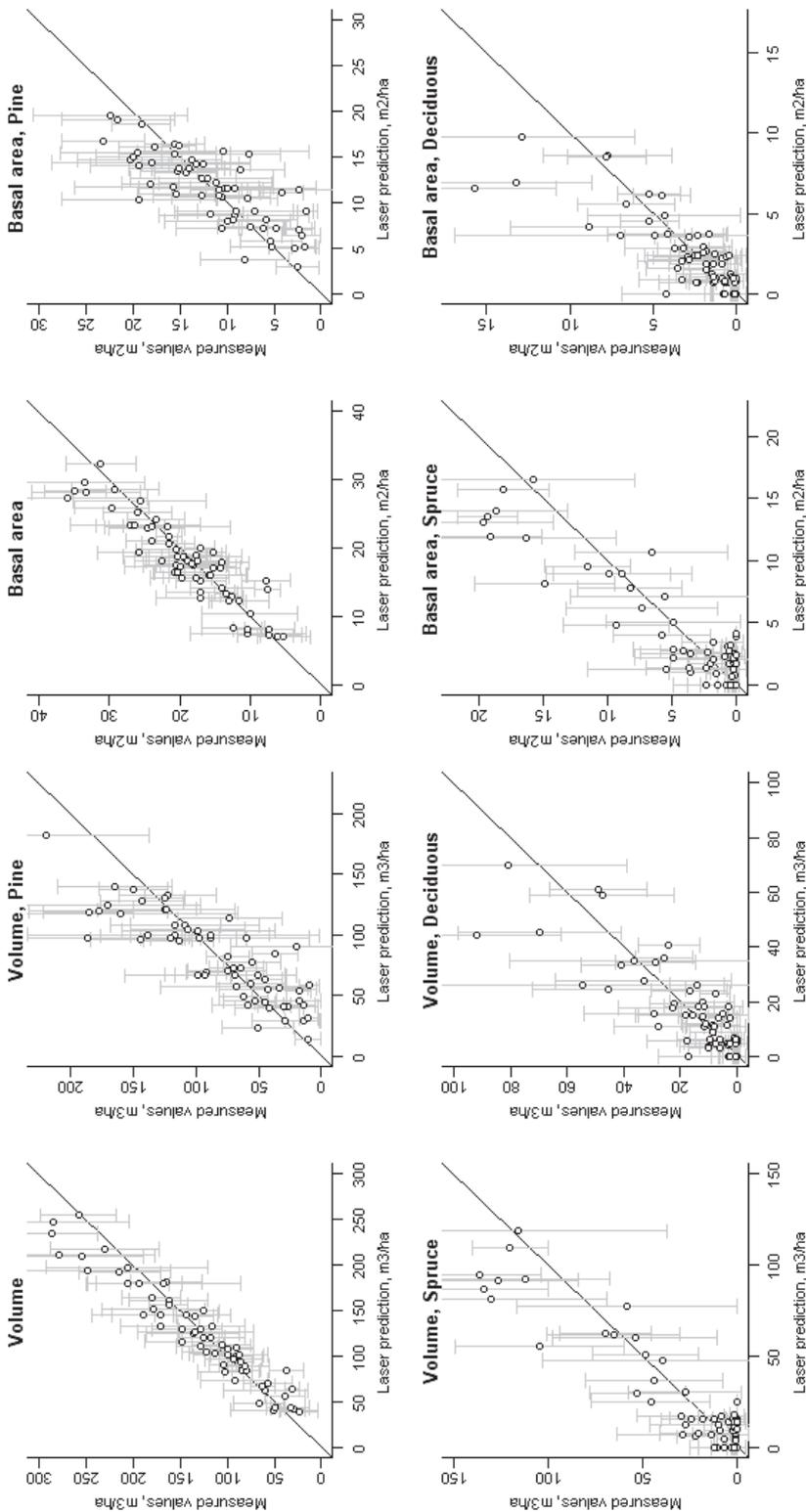


Fig. A1. Confidence intervals of Volume, m³/ha (both for total and for different tree species).

Fig. A2. Confidence intervals of Basal area, m²/ha (both for total and for different tree species).

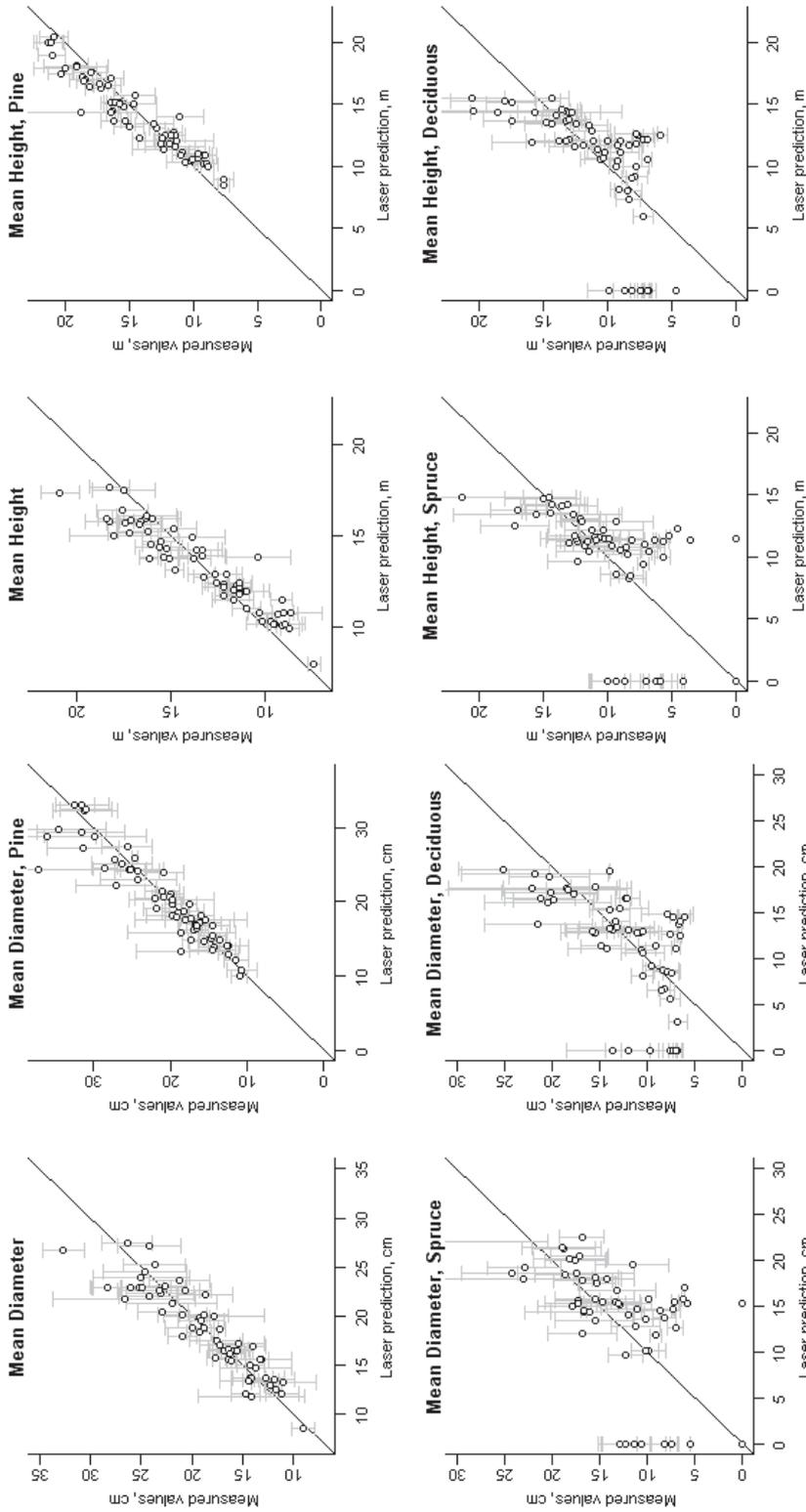


Fig. A3. Confidence intervals of mean Diameter, cm (both for total and for different tree species). **Fig. A4.** Confidence intervals of mean Height, m (both for total and for different tree species).

Appendix 2 (Bland-Altman plots)

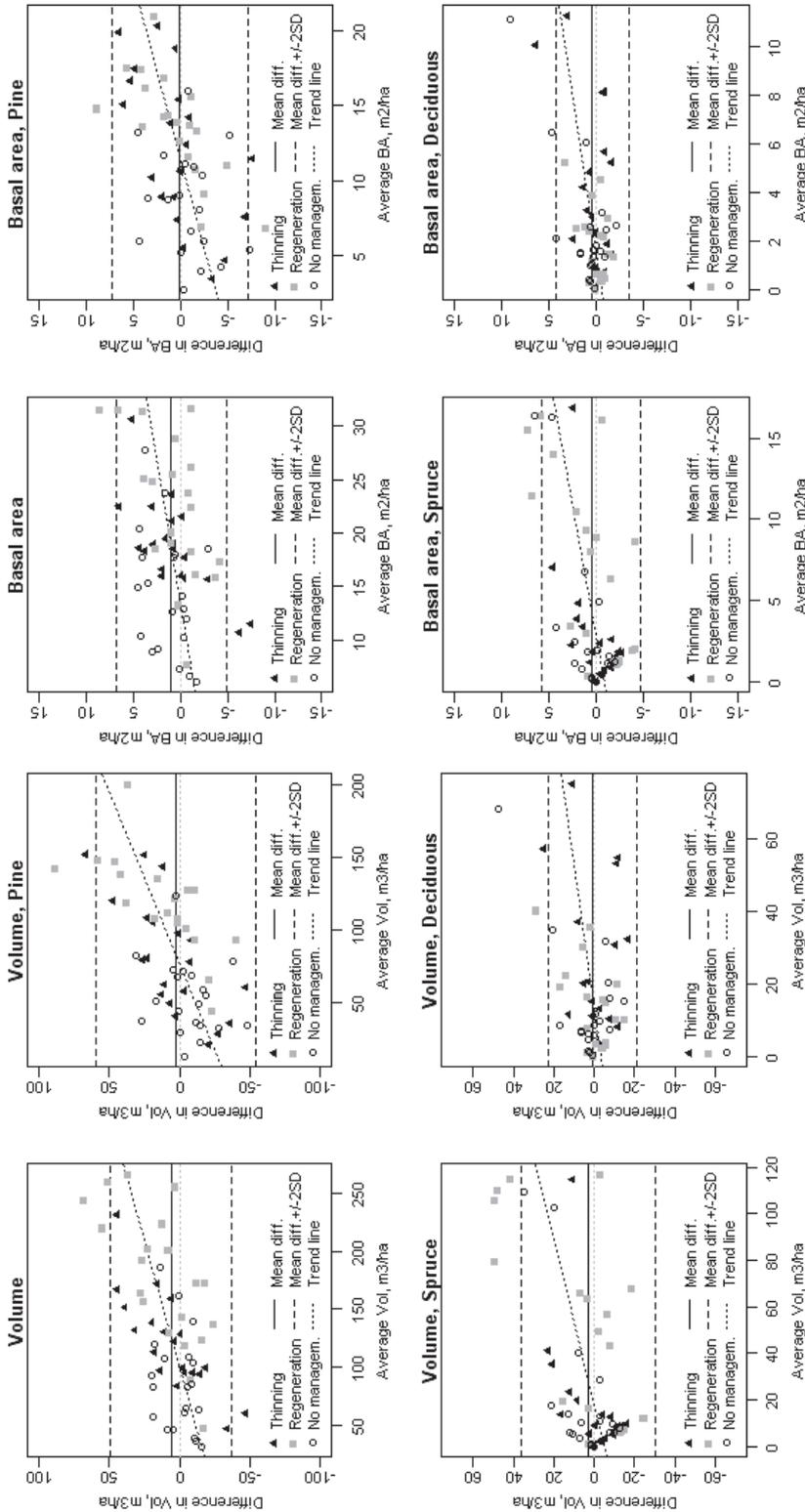


Fig. A5. The Bland-Altman plot for Volume, m³/ha. Difference is control – laser. Thinning, regeneration and no management based on laser inventory.

Fig. A6. The Bland-Altman plot for Basal Area, m²/ha. Difference is control – laser. Thinning, regeneration and no management based on laser inventory.

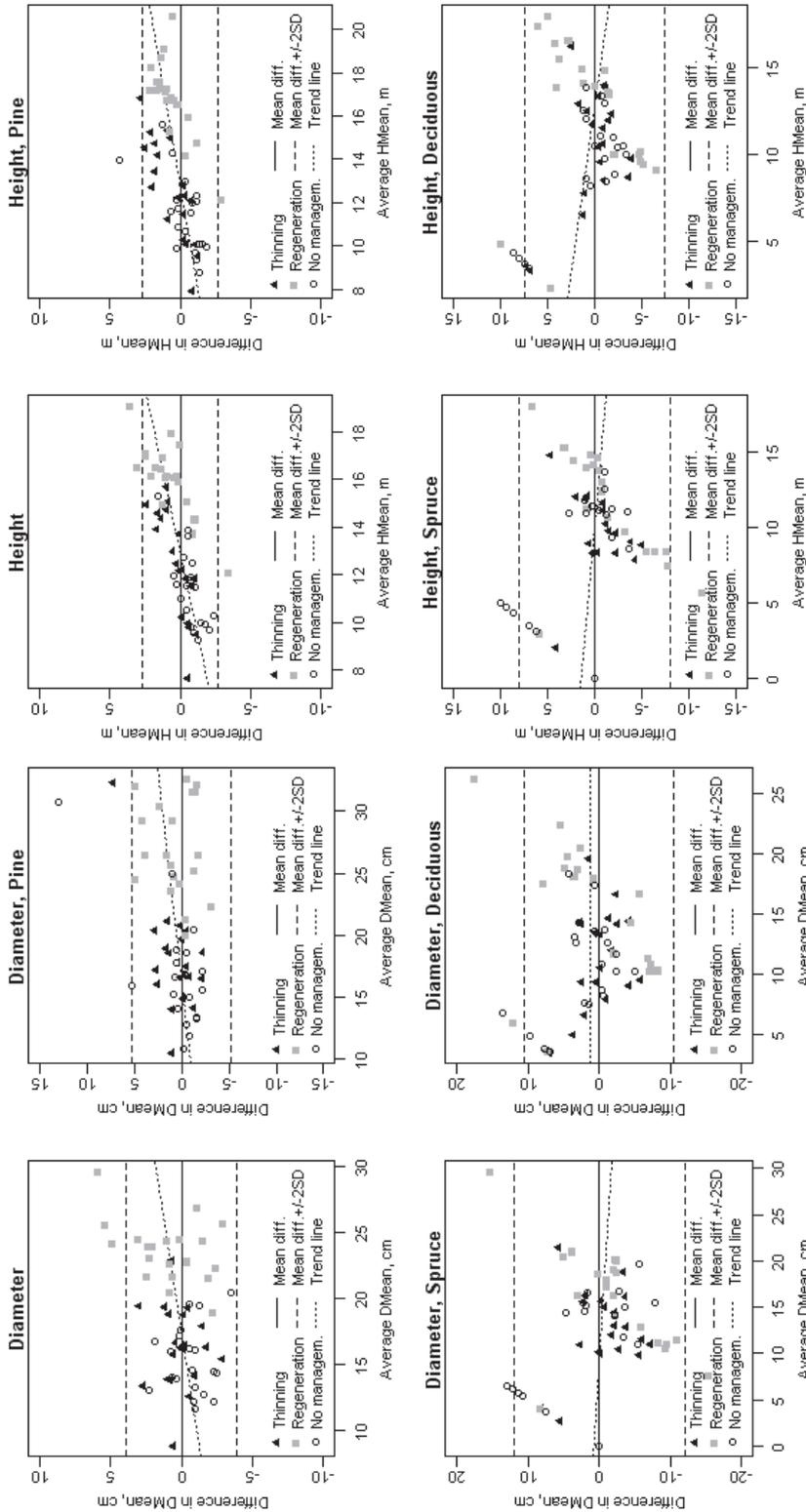


Fig. A7. The Bland-Altman plot for Diameter, cm. Difference is control – laser. Thinning, regeneration and no management based on laser inventory.

Fig. A8. The Bland-Altman plot for Height, m. Difference is control – laser. Thinning, regeneration and no management based on laser inventory.