Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Epsilon Open Archive http://epsilon.slu.se

Estimation of tree lists from airborne laser scanning by combining single-tree and area-based methods

Eva Lindberg*†, Johan Holmgren†, Kenneth Olofsson†, Jörgen Wallerman † and Håkan Olsson †
† Swedish University of Agricultural Sciences, Department of Forest Resource Management, SE-90183 Umeå, Sweden
*Corresponding author. Email address: Eva.Lindberg@srh.slu.se

Abstract

Individual tree crown segmentation from airborne laser scanning (ALS) data often fails to detect all trees depending on the forest structure. This paper presents a new method to produce tree lists consistent with unbiased estimates at area level. First, a tree list with height and diameter at breast height (DBH) was estimated from individual tree crown segmentation. Second, estimates at plot level were used to create a target distribution by using a k-nearest neighbour (k-NN) approach. The number of trees per field plot was rescaled with the estimated stem volume for the field plot. Finally, the initial tree list was calibrated using the estimated target distribution. The calibration improved the estimates of the distributions of tree height (error index (EI) from 109 to 96) and DBH (EI from 99 to 93) in the tree list. Thus, the new method could be used to estimate tree lists that are consistent with unbiased estimates from regression models at field plot level.

1 Introduction

Estimates of the size and position of every tree in a forest has become a realistic option with the commercial development of high-frequency airborne laser scanners and high-resolution digital aerial cameras. Many modern systems for forest management planning will require forest information at the individual tree level (Söderberg and Ledermann 2003, Backéus *et al*. 2005, Kärkkäinen *et al*. 2008) or, at the very least, information about the diameter at breast height (DBH) distribution (Maltamo *et al*. 2007). For the purpose of forest resource planning, it is also essential that the estimates are unbiased.

Data from airborne laser scanning (ALS) can give information about the height and density of the tree canopy. Recent developments in ALS technology have increased the pulse repetition frequency, making high spatial coverage possible without decreased ground sampling rate. There are already commercial projects in Sweden where high ground sampling rate (≥10 emitted pulses m-2) ALS data have been retrieved for large areas (≥ 100 000 ha) to produce estimates of forest variables for the forest industry. The high ground sampling rate ALS data can be used to measure tree canopy height and shape. By creating a canopy height model (CHM) from the ALS data and using an algorithm that identifies maxima in the canopy heights and divides the tree canopy into tree crown segments based on the height, it is possible to segment individual tree crowns (Brandtberg 1999, Hyyppä *et al*. 2001, Persson *et al*. 2002, Holmgren and Wallerman 2006, Solberg *et al*. 2006). Field data from the laser scanned

area are needed to create statistical models to estimate variables relevant to forestry such as tree height, DBH and stem volume. A great advantage of single-tree methods is that the tree species can be estimated from canopy shape (Holmgren and Persson 2004). In coniferous-dominated boreal forests, individual tree detection methods could provide lists of estimated tree stems with associated variables such as position, tree height, DBH and tree species, at least for most of the dominant and subdominant trees. However, tree crown segmentation often fails to detect trees below the dominant tree layer or trees standing close together, and one tree crown segment may contain more than one tree (Persson *et al*. 2002). Therefore, the output from the tree crown segmentation will result in an underestimation of the stem density and stem volume. Further development of these methods is therefore needed to produce unbiased estimates from individual tree detection methods for field plots, forest stands or any other area unit.

ALS is used operationally in Scandinavian forestry (Næsset and Bjerknes 2001, Næsset 2007). Typically, the height distribution of laser returns at georeferenced field plots is used to create regression models to estimate forest variables (Næsset 2004). A non-parametric k most similar neighbour (k-MSN) method has also been applied to a combination of ALS data and aerial photographs (Packalén and Maltamo 2007). Once the regression models for plot-level estimates have been established with coefficients unique for the acquisition, estimates of forest variables can be aggregated to stand level (Næsset 2002). These methods are usually referred to as area-based methods

because the estimates are mean values of forest variables in an arbitrary area. Area-based methods have become operational in coniferous-dominated boreal forests because they produce estimates of mean tree height, mean DBH, basal area and stem volume with errors of the same size as from accurate field inventories (i.e. using several field plots within each forest stand) (Næsset and Bjerknes 2001, Næsset *et al.* 2005, Maltamo *et al.* 2006). In addition, lower ground sampling rate data could be used than for individual tree detection methods, typically about one emitted pulse per square metre, which makes area-based methods cost-efficient.

Area-based methods using only low ground sampling rate ALS data will, however, not provide information about tree species or tree lists. Thus, individual tree detection methods are an attractive option for the future. To address the underestimation of individual tree detection methods, Maltamo *et al.* (2004) combined a theoretical distribution function with the tree height distribution of the tree crown segments from ALS data and estimated a tree height distribution. The estimated distribution was used as a complement to predict the number of stems with smaller DBH. The tree crown segmentation resulted in an underestimation of stem density (root mean square error (RMSE)=74%, bias=61%) and stem volume (RMSE=25%, bias=24%), but estimation of tree height distribution reduced the error (stem density RMSE=49%, bias=6%; stem volume RMSE=16%, bias=8%) for field plots with approximately 100 trees each (plot size 625-1600 m2). Flewelling (2008) estimated the number of trees associated with each detected tree crown by using probability models, where features derived from the crown area were used as independent variables. The models produced unbiased estimates at stand level even though the proportion of matched tree crown segments was only 48% of the total number of matched trees and 74% of total matched basal area. However, a separate model was needed to handle unseen trees, that is trees that had not been matched to any of the detected trees. To further improve the estimation of individual trees, results from an area-based method could be used to deduce whether the individual tree crown segmentation has identified all trees in an area.

When deriving stand properties from ALS data and field plots, the remotely sensed data and field data need to be accurately co-registered to reduce errors (Gobakken and Næsset 2008), preferably with submetre accuracy. However, field data with such accurate positions are not always available. There are a number of techniques to automatically correct poorly co-registered datasets: Dorigo *et al.* (2010) found 67 % of the true field plot positions by searching for the best match between an ALS-derived tree crown height model and field plot tree positions and heights; Korpela *et al.* (2007) suggested a method that combines photogrammetric observations of tree tops and field triangulation; and Flewelling (2008) used a computer-assisted system and Voronoi tessellations to associate trees with individual tree crown segments. A method to cross-correlate tree position images for the co-registration of field and remotely sensed data has also been proposed by Olofsson *et al.* (2008). This technique requires input data from a remote sensing individual tree detection method (e.g. Erikson and Olofsson 2005).

The aim of this study was to develop and validate a method to produce a list of individual trees, with each tree having estimated variables such as tree height, DBH and stem volume. The most important requirement is that the estimated tree list should produce unbiased estimates if estimated tree variables, such as stem volume, are aggregated over an area, for example a forest stand. The developed method combines analysis at the individual tree level and the field plot level. The idea is to use the information about most individual trees that can be measured directly in the high ground sampling rate ALS data and calibrate these results with estimates from area-based methods at plot level with high accuracy and low bias.

2 Materials

2.1 Study area

The study area is 1989 ha and located in the north of Sweden (latitude 64 ° 25' N, longitude 14 ° 50' E) with elevation ranging from 325 to 658 m asl, which means that the site is located close to the local limit for productive forest. The dominating tree species are Norway spruce (*Picea Abies*), birch (*Betula spp.*) and Scots pine (*Pinus Sylvestris*).

2.2 ALS data

The ALS data acquisition was performed on 7 and 8 August 2007 using a Leica ALS50-II system carried by a helicopter. The flying altitude was 600 m above ground level and the scan angle ±16°, resulting in a scan width of 375 m and a scan density of about 10 emitted pulses m-2. The first and last returns for each pulse were recorded and used in the further analysis. Laser returns were classified as ground or non-ground using a progressive triangular irregular network (TIN) densification method (Axelsson 1999, 2000) in TerraScan software (Soininen 2004), and the ground returns were used to derive a digital elevation model (DEM). Laser canopy height (LCH) was derived by subtracting the relevant DEM value from each laser return.

2.3 Field data

The area was divided into five strata using an existing forest map with associated stand register, and a total of 179 field plots were allocated (table 1).

Different field plot radius and minimum DBH were selected for cost-efficiency because of the large number of stems in young forest compared to old forest.

Table 1: Summary of field plot data.

Stratum	Selection criterion: Age (years)	Selection criterion: Species composition	Number of field plots	Field plot radius (m)	Minimum DBH (mm)
1	25-74	>=60% pine	23	6	40
2	25-74	>=60% spruce	29	6	40
3	25-74	mixed forest	33	6	40
4	≥75	>=60% spruce	60	8	60
5	≥75	>=60% pine or mixed forest	34	8	60

Stratum	Species composition (%)			Stem volume (m^3 ha^{-1})			Stem density (ha^{-1})		
	Pine	Spruce	Other	Mean	5 percentile	95 percentile	Mean	5 percentile	95 percentile
1	61	25	14	40	28	59	1484	539	2847
2	0	65	35	49	13	122	1524	654	2493
3	31	40	29	43	6	132	1299	601	2440
4	9	74	17	119	41	218	895	540	1450
5	36	56	8	140	51	261	895	413	1577

The positions of the field plots were measured using a Global Navigation Satellite System (GNSS). The Forest Management Planning Package (Jonsson *et al.* 1993) was used to measure the trees in the field plots. Within the field plots, all trees with a DBH greater than the minimum DBH were measured using a calliper, and the tree species were recorded. The height was also measured for a subsample of the trees selected randomly with probability proportional to size (PPS) sampling proportional to basal area. The positions of the trees were registered relative to the centre of each field plot by measuring azimuth and distance with a compass and ultrasonic device, respectively. No position was registered for trees with a large inclination, that is with a horizontal distance greater than 3 m between the tree top and the position at breast height (3% of the total number of trees). The tree positions were clustered and the canopy had larger gaps than most managed forests in Scandinavia. The mean stem density was 1147 stems ha -1 and the mean stem volume was 96 m3 ha-1.

3 Methods

The method developed in this study is outlined in figure 1. The first step is estimation of individual trees. This includes automatic segmentation of individual tree crowns from ALS data and field plot matching to co-register the coordinates. The number of trees per segment is estimated and their height and DBH are estimated from regression models. The second step is estimation at the field plot level, which includes estimation of stem volume from regression, and estimation of percentiles of the tree height and DBH distributions by using a k-nearest neighbour (k-NN) approach. The last step is calibration of the tree list from the first step using a target distribution of tree height and DBH estimated at the field plot level from the second step. This includes removal of excess trees and addition of missing trees.
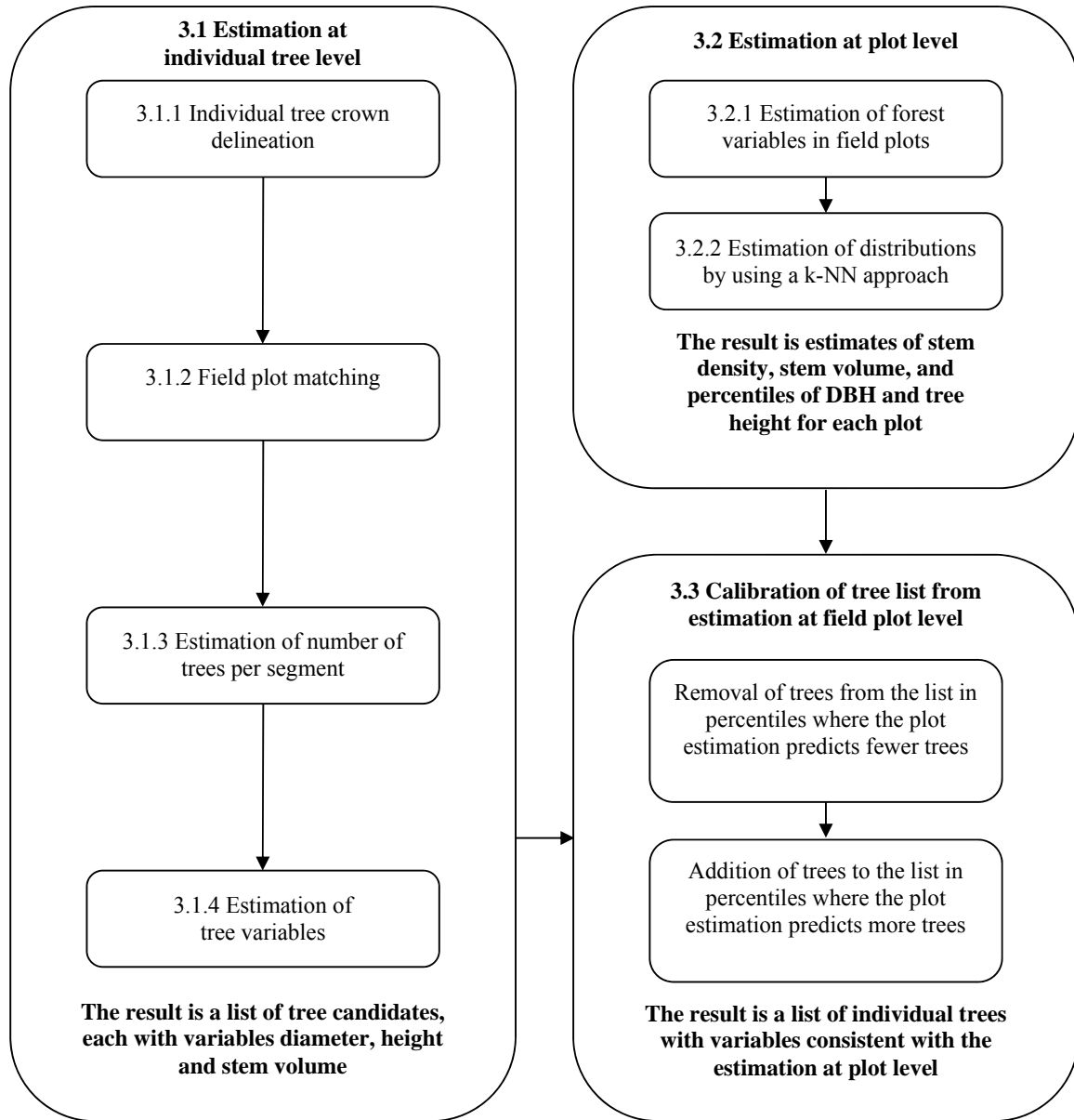
3.1 Estimation at individual tree level

> **3.1.1 Individual tree crown delineation**

> **3.1.2 Field plot matching**

> **3.1.3 Estimation of number of trees per segment**

> **3.1.4 Estimation of tree variables**

The result is a list of tree candidates, each with variables diameter, height and stem volume

3.2 Estimation at plot level

> **3.2.1 Estimation of forest variables in field plots**

> **3.2.2 Estimation of distributions by using a k-NN approach**

The result is estimates of stem density, stem volume, and percentiles of DBH and tree height for each plot

3.3 Calibration of tree list from estimation at field plot level

> Removal of trees from the list in percentiles where the plot estimation predicts fewer trees

> Addition of trees to the list in percentiles where the plot estimation predicts more trees

The result is a list of individual trees with variables consistent with the estimation at plot level

Figure 1. An outline of the method used in this study.

3.1 Estimation at individual tree level

3.1.1 Individual tree crown segmentation.

Tree crowns were segmented automatically based on geometric tree crown models (Holmgren and Wallerman 2006). The segmentation process included several steps with different raster, all with a cell size of 0.25 m × 0.25 m. A CHM was created from LCH data by setting the raster cell value to the maximum LCH value within each raster cell if this value was above a 2 m height threshold, otherwise the cell was given a zero value. The height threshold was used to avoid the influence of rocks and low vegetation. Raster cells containing no laser returns were set to zero. However, some of these raster cells could be in a tree crown. Such zero-values within crowns were regarded as noise and were therefore filled using a morphological closing operation. This involved several steps. First, a binary canopy raster was derived from the CHM, in which a raster cell was set to one if the corresponding CHM raster cell had a value greater than zero, or was left as a zero-value otherwise. This binary canopy raster should define areas containing crowns. To remove zero-values of canopy raster cells located within crowns, closing was done with a structure element of 3 × 3 cells. The next step was to update the CHM based on the binary canopy raster in order to set non-zero values of the CHM at crown locations. Zero-values of the CHM were updated if the raster cell was at a tree crown location according to the binary canopy raster (value = one). Neighbour cells with values greater than zero within the smallest window needed to cover at least one value greater than zero were used to update the CHM raster cell. If only one non-zero value was found within the window, that value was assigned, otherwise the mean was used.

A correlation raster was created based on both the CHM raster and LCH data. The correlation between geometric models of tree crowns and LCH values was calculated at each raster cell location of the correlation raster using generalized ellipsoids of revolution (GER). Pollock (1996) introduced the use of GER-based models of tree crowns for automatic detection of individual trees in aerial images. At each raster cell location, the cell value of the correlation raster was set to the maximum correlation from tests with several GER models. A correlation value was calculated by placing the centre of the GER at the cell location, setting the height (i.e. the vertical radius of the ellipsoid) to the value of the CHM at the cell location, and testing a horizontal radius. The tested horizontal radii of the GER were 0.5, 0.7,…, 4.0 m. To reduce the amount of false trees, a correlation value was only calculated for a GER model with a radius to height ratio ≥ 0.1 because, based on field experience, a smaller ratio would probably model an individual branch and not a whole tree. In addition, to avoid overfitting, correlation values were only calculated for GER models with at least 25 laser returns within the radius.

The correlation raster was smoothed and used for segmentation. A starting point, a seed, was placed at each raster cell with a CHM value above a 2 m height threshold and with a positive correlation value. For each seed, the current location was updated by changing the position to the position of the neighbour cell (eight neighbours) with the highest value. This was repeated until the position could not be updated because a local maximum was reached. The seeds with the final location at the same local maxima defined a tree crown segment. A raster cell that had not been included in a segment but was enclosed by raster cells from the same segment was assigned to the surrounding segment.

The result of the segmentation was tree crown segments, each including an individual tree or a group of trees. The maximum LCH value (H) within the segment was used to estimate the tree height. The area of a segment (A) was derived by counting the number of raster cells (0.25 m × 0.25 m) of a segment and the width (W) of a segment was then derived assuming that a tree crown was circular.

3.1.2 Field plot matching.

The field plot-measured tree positions had a precision of < 1 m but lower accuracy. This was because tree positions relative to the plot centre were measured with approximately decimetre accuracy but GNSS measurements of the plot centre could have a bias of several metres depending on the satellite configuration and poor receiving conditions of the satellite signals due to tree crowns. Therefore, the two coordinate systems, that is the local tree positions from the field inventory and the global coordinate system of the ALS data, needed to be co-registered before the field plot trees could be linked to the trees detected from ALS. To achieve this, the method by Olofsson *et al*. (2008) was used. In the search process, the estimated field plot centre was displaced up to ±20 m in the north-south direction and up to ±20 m in the east-west direction, giving a search area estimated from the expected GNSS error to be large enough to contain the real field plot centre. To compensate for possible compass errors, the field plot image was rotated 2° between each calculation (with minimum/maximum field plot rotation end values of ±16°). The standard deviation of the tree displacement precision was set to 0.75 m based on the precision of the field plot-measured tree positions. The size of the raster cell used in the field plot matching was set to 0.25 m × 0.25 m based on the stem density.

3.1.3 Estimation of number of trees per segment.

Individual tree crown segmentation often produces segments that, in reality, contain several trees, which causes underestimation of the stem density (Persson *et al*. 2002). However, the shape and size of each tree crown segment could contain information about the actual number of trees, where, for instance, large and elongated segments probably contain several trees. To obtain an estimate of the correct stem density, a training dataset was used where the number of field-measured trees for each tree crown segment was known.

Segments close to a field plot boundary usually cover ground outside the field plot and it is not known how many trees they contain. The models for individual trees were therefore based only on segments where the centre was located inside a field plot and at least 2 m from the boundary, from now on referred to as *reference segments* (figure 2). The estimation of variables at the individual tree level did, however, include all segments where the centre was located inside a field plot. We excluded trees with a large inclination from the field data when estimating the models at the individual tree level, but included them when estimating the models at the field plot level and validating the estimated results.

Figure 2. Example of polygons from segmentation of ALS data and field-measured trees shown as circles with radius proportional to DBH. The outer large circle represents the field plot and the inner large circle represents the area where the centres of the reference segments are located.

Table 2: Variables used for analysis at the tree crown segment level. Variables 1 and 2 were derived from the laser data and variables 3-11 were derived from the field data.

	Variable	Description
1	W	$2\sqrt{A/\pi}$, A is the area of the segment
2	W/W_s	W_s is the mean of the segment widths within the same field plot
3	N	Number of field-measured trees within a segment
4	D_{max}	Maximum field-measured DBH found within a segment
5	D_{rest}	Sum of field-measured DBH for the rest of the trees within a segment
6	H_{max}	Maximum field-measured tree height found within a segment
7	H_{rest}	Mean of field-measured tree heights for the rest of the trees within a segment
8	B_{max}	Maximum field-measured basal area found within a segment
9	B_{rest}	Sum of field-measured basal areas for the rest of the trees within a segment
10	V_{max}	Maximum field-measured stem volume found within a segment
11	V_{rest}	Sum of field-measured stem volumes for the rest of the trees within a segment

As a first step, mean values for the number of trees per segment were estimated for reference segments in the training dataset based on the variables W_j and $W_j/W_{s,i}$, for each segment j, where $W_{s,i}$ is the mean of segment widths within the same field plot i (table 2). W_j and $W_j/W_{s,i}$, were selected because they showed the highest

correlation with the number of trees per segment of all available measures. The variables were divided into eight intervals of equal size. The number of intervals was selected based on the data to have a sufficient number of trees in each interval. Each reference segment was placed in the interval determined by the value of W_j and $W_j/W_{s,i}$ for that segment, resulting in 64 intervals. For each interval, the mean number of field-measured trees was calculated. As a second step, the number of trees within segments was predicted for all segments. To estimate the number of trees inside a segment, it was first placed in an interval determined by W_j and $W_j/W_{s,i}$ for that segment. The number of trees m_j inside the segment was estimated as the mean number of field-measured trees for segments in the same interval of W_j and $W_j/W_{s,i}$ in the training dataset.

3.1.4 Estimation of tree variables.

Regression models for the field-measured variables 4-11 in table 2 were estimated using the reference segments. The regression was done separately for segments containing different numbers of field-measured trees. For each segment, the variables 4-11 in table 2 were estimated as the weighted mean value of the regression result for segments containing different numbers of field-measured trees. The result was an estimate of variables for the largest tree in each segment (i.e. 4, 6, 8, and 10 in table 2) and an estimate of variables for the rest of the trees in the segment (i.e. 5, 7, 9, and 11 in table 2). The latter estimate $\hat{A}_j$ was divided by the estimated number of trees minus one to obtain an estimate for the average tree:

$$\hat{A}_j' = \frac{\hat{A}_j}{\hat{m}_j - 1} \qquad (1)$$

If the result for the DBH was below the minimum value for the field-measured DBH, $\hat{m}_j$ was reduced iteratively by one until the resulting DBH was above the minimum value. If the DBH was below the minimum value for the field-measured DBH even for one tree, the estimate was discarded. The estimates for the largest tree and the rest of the trees in all segments formed a list of tree candidates.

3.2 Estimation at field plot level

3.2.1 Estimation of forest variables in field plots.
The mean of LCH, the standard deviation of LCH, percentiles of LCH, and the vegetation ratio were derived based on the LCH distribution by using vegetation returns at the field plot level. The vegetation ratio is defined as the number of vegetation returns divided by the total number of laser returns. To exclude laser returns from below the tree canopy (e.g. shrubs and rocks), vegetation returns were defined as

laser returns with a vertical distance to the DEM > 1 m and > 10% of the maximum height within the field plot (e.g. Næsset 2002).

A regression model was used to estimate the stem volume per hectare (V_r) for each field plot i. To investigate the contribution from the available variables, different combinations were tested with best subset regression, and the following regression model was selected:

$$ln(V_{r,i}) = \alpha_0 + \alpha_1 ln(p_{30,i}) + \alpha_2 ln(p_{95,i}) + \alpha_3 ln(R_{v,i}) + \epsilon_i$$
(2)

where p_{30} is the LCH 30th percentile, p_{95} is the LCH 95th percentile and R_v is the vegetation ratio, defined as the number of vegetation returns divided by the total number of laser returns in a field plot. All variables had a p-value ≤ 0.06 in the model and the variance inflation factor (VIF) was 1.207-4.183.

The stem density (number of stems per hectare, d_r) was estimated for each field plot i from a regression model:

$$d_{r,i} = \beta_0 + \beta_1 p_{90,i} + \beta_2 R_{v,i} + \epsilon_i \qquad (3)$$

where z_{stdh} is the standard deviation of LCH, p_{90} is the LCH 90th percentile and R_v is the vegetation ratio. All variables had a p-value ≤ 0.001 in the model and the variance inflation factor (VIF) was below 1.083.

The distributions of tree heights, DBH and basal area have been estimated previously from ALS data. The Weibull distribution has been commonly used to describe tree size-related distributions such as DBH and basal area (Gobakken and Næsset 2004, Maltamo *et al*. 2007). Some results (Breidenbach *et al*. 2008) indicate that distribution methods may fail in strata with large gaps in the canopy. Distributions of DBH have also been estimated using percentiles with non-parametric methods (Gobakken and Næsset 2005, Maltamo *et al*. 2006, Bollandsås and Næsset 2007). A comparison of a Weibull DBH distribution and DBH percentiles showed no significant difference between the two validated methods (Gobakken and Næsset 2005). Maltamo *et al*. (2006) compared estimated DBH percentiles and a Weibull DBH distribution. Of these two, DBH percentiles resulted in the best result for estimation of stem volume. Studies have also shown that the parametric distribution methods are not suitable for estimation of multilayered forest (Kangas and Maltamo 2000). For the present study, we used tree height and DBH percentiles to estimate the target distributions because theoretical distributions were not suitable for use with the small field plots.

For each field plot, the DBH percentiles and tree height percentiles were estimated from seemingly unrelated regression (SUR) models (Zellner1962), (table 3), where z_{avg} is the mean of LCH, z_{stdh} is the

standard deviation of LCH, and the other symbols are as defined earlier.

3.2.2 Estimation of distributions by using a k-NN approach.

To estimate the final tree height and DBH distributions, we used a k-NN approach. The estimates of the stem volume, stem density and tree height and DBH percentiles (equations (2) and (3) and table 3)

Table 3. Seemingly unrelated regression models for percentiles of tree height and DBH.

	SUR model, separate for tree height and DBH percentiles
Percentiles of tree height	$H_{25,i} = \alpha_{0H25} + \alpha_{1H25}p_{10,i} + \alpha_{2H25}p_{70,i} + \alpha_{3H25}z_{avg,i} + \alpha_{4H25}z_{stdh,i} + \alpha_{5H25}R_{v,i}$
	$H_{50,i} = \alpha_{0H50} + \alpha_{1H50}p_{70,i} + \alpha_{2H50}z_{avg,i} + \alpha_{3H50}z_{stdh,i} + \alpha_{4H50}R_{v,i}$
	$H_{75,i} = \alpha_{0H75} + \alpha_{1H75}p_{95,i} + \alpha_{2H75}z_{avg,i} + \alpha_{3H75}z_{stdh,i} + \alpha_{4H75}R_{v,i}$
	$H_{90,i} = \alpha_{0H90} + \alpha_{1H90}p_{95,i} + \alpha_{2H90}z_{avg,i} + \alpha_{3H90}z_{stdh,i} + \alpha_{4H90}R_{v,i}$
	$H_{100,i} = \alpha_{0H100} + \alpha_{1H100}p_{95,i} + \alpha_{2H100}z_{avg,i} + \alpha_{3H100}z_{stdh,i} + \alpha_{4H100}R_{v,i}$
Percentiles of DBH	$D_{25,i} = \alpha_{0D25} + \alpha_{1D25}p_{10,i} + \alpha_{2D25}p_{70,i} + \alpha_{3D25}z_{avg,i} + \alpha_{4D25}z_{stdh,i} + \alpha_{5D25}R_{v,i}$
	$D_{50,i} = \alpha_{0D50} + \alpha_{1D50}p_{70,i} + \alpha_{2D50}z_{avg,i} + \alpha_{3D50}z_{stdh,i} + \alpha_{4D50}R_{v,i}$
	$D_{75,i} = \alpha_{0D75} + \alpha_{1D75}p_{70,i} + \alpha_{2D75}z_{avg,i} + \alpha_{3D75}z_{stdh,i} + \alpha_{4D75}R_{v,i}$
	$D_{90,i} = \alpha_{0D90} + \alpha_{1D90}p_{95,i} + \alpha_{2D90}z_{avg,i} + \alpha_{3D90}z_{stdh,i} + \alpha_{4D90}R_{v,i}$
	$D_{100,i} = \alpha_{0D100} + \alpha_{1D100}p_{95,i} + \alpha_{2D100}z_{avg,i} + \alpha_{3D100}z_{stdh,i} + \alpha_{4D100}R_{v,i}$

were used to identify field plots with similar tree height and DBH distributions with the k-NN approach. The 10 field plots with the smallest sums of squared differences between the values were selected as the nearest-neighbour field plots.

Tree height percentiles and DBH percentiles were derived from the field-measured trees in the nearest-neighbour field plots. The field-measured trees at the nearest-neighbour plots were placed in a distribution matrix where each row corresponded to a tree height percentile class and each column to a DBH percentile class (table 4).

The number of trees n_{fij} in each matrix cell was multiplied by a scaling factor S:

$$S = \frac{\hat{V}_r}{V_{k-MSN}} \quad (4)$$

where $\hat{V}_r$ is the stem volume per hectare estimated from regression and V_{k-MSN} is the stem volume per hectare from nearest-neighbour field plots. The result was an estimated target distribution matrix (table 4), where each element corresponded to a tree height percentile and DBH percentile, and the value n_{tij}

corresponded to the number of trees in the field plot in each combination of percentiles. The aim was to make the sum of the volume of the trees in the target distribution matrix consistent with $\hat{V}_r$ at the field plot level.

3.3 Calibration of tree list from estimation at field plot level

The aim was to predict tree lists with variables that produce unbiased estimates if aggregated over an area. Therefore, the list of individual tree candidates from the segmentation was calibrated using estimates of stem volume and estimates of tree height and DBH distributions at the field plot level. First, we calculated the distribution of the tree candidates from the segmentation by summing the number of tree candidates n_{stij} in each percentile class given by the target distribution (table 4). Tree candidates with a height or DBH greater than the 100th percentile were excluded from the list. For each percentile class, the difference between the target distribution (i.e. the distribution estimated at field plot level by using the k-NN approach) and the candidate distribution was calculated. If the number of tree candidates was too large according to the target distribution, that number of trees was excluded from the list.

Table 4. The distribution matrices used in the analysis.

Field distribution	Target distribution	Tree candidate distribution

$$
\begin{bmatrix}
 & D_{10} & D_{20} & \dots & D_{100} \\
H_{10} & n_{f1,1} & n_{f1,2} & \dots & n_{f1,10} \\
H_{20} & n_{f2,1} & \dots & \dots & \dots \\
\dots & \dots & \dots & \dots & \dots \\
H_{100} & n_{f10,1} & \dots & \dots & n_{f10,10}
\end{bmatrix}
\begin{bmatrix}
 & D_{10} & D_{20} & \dots & D_{100} \\
H_{10} & n_{t1,1} & n_{t1,2} & \dots & n_{t1,10} \\
H_{20} & n_{t2,1} & \dots & \dots & \dots \\
\dots & \dots & \dots & \dots & \dots \\
H_{100} & n_{t10,1} & \dots & \dots & n_{t10,10}
\end{bmatrix}
\begin{bmatrix}
 & D_{10} & D_{20} & \dots & D_{100} \\
H_{10} & n_{st1,1} & n_{st1,2} & \dots & n_{st1,10} \\
H_{20} & n_{st2,1} & \dots & \dots & \dots \\
\dots & \dots & \dots & \dots & \dots \\
H_{100} & n_{st10,1} & \dots & \dots & n_{st10,10}
\end{bmatrix}
$$

If the number of tree candidates was too small, that number of trees was added to the list by selecting trees with correct height and DBH for the specific percentile class at random from the list of field-measured trees in the nearest-neighbour field plots. The aim was a list of trees with size distribution and stem volume consistent with the estimates at the field plot level. The results were aggregated to the field plot level, and the procedure was repeated 50 times to study the average accuracy of the estimation.

3.4 Validation

All steps in the method were validated by using leave-one-out cross-validation. One field plot at a time was excluded from the dataset, and the rest of the dataset was used as reference data to estimate forest variables for the excluded field plot. This was done both for estimates at both the individual tree level and the field plot level.

The RMSE and bias for stem volume were calculated as follows:

$$
RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{V}_i - V_i)^2}{n}} \tag{5}
$$

$$
bias = \frac{\sum_{i=1}^{n} (\hat{V}_i - V_i)}{n} \tag{6}
$$

where V_i is the observed and $\hat{V}_i$ is the predicted stem volume of field plot i and n is the number of field plots. Corresponding expressions were used to calculate the RMSE and bias for stem density.

The error index (EI) for tree height, DBH and basal area in each field plot was also calculated. The EI measures the proportion of mismatches between two histograms based on given class boundaries and is, according to Reynolds *et al*. (1988), defined as:

$$
EI = \frac{1}{N_T} \sum_{j=1}^{m} |T_j - I_j| \tag{7}
$$

where I_j is the number of estimated trees to histogram class j, T_j is the number of actual trees in class j, m is the number of histogram classes, and N_T is the total number of actual trees.

In this study, the tree height, DBH and basal area were divided into ten intervals each.

4 Results

4.1 Estimation of number of trees per segment

Table 5 shows the result for the estimation of the number of trees per segment. Most of the segments contained one or two trees. The accuracy was high for segments that contained one or two trees (0.71 and 0.65, respectively) and much lower for segments that contained three trees or more. The estimation had an overall accuracy of 0.60.

4.2 Estimation of stem volume and stem density

Estimation at the individual tree level only (table 6, methods 1a and 1b) resulted in a large RMSE and a large underestimation for stem density (52%, -35% and 45%, -18%, respectively). Estimation of the number of trees per segment with mean values (table 6, method 1b) resulted in a smaller RMSE and bias both for stem volume (34%, -3%) and stem density (45%, -18%) than estimation with one tree per tree crown segment (36%,-14% and 52%, -35%, respectively) (table 6, method 1a).

Table 5. Result for the classification of the number of trees in each segment. The columns correspond to the correct number of trees per segment according to field data and the rows correspond to the estimated number of trees per segment.

Estimated number of trees	Correct number of trees											Correct fraction
	0	1	2	3	4	5	6	7	8	9	10	
0	0	3	0	1	0	0	0	0	0	0	0	0.00
1	0	472	52	13	3	2	0	0	0	0	0	0.87
2	0	166	130	65	28	8	4	1	0	0	0	0.32
3	0	21	5	4	0	4	4	0	2	0	1	0.10
4	0	6	12	9	8	0	1	0	2	0	1	0.21
Correct fraction	0.00	0.71	0.65	0.04	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.60

Table 6. RMSE and bias for stem volume and stem density estimates, cross-validation results aggregated over field plots, relative to the mean values at the field plot level using the methods:

1a, Estimation at the individual tree level; 1b, Estimation at the individual tree level including estimation of the number of trees per segments; 2, Estimation at the field plot level; 3a, Calibration of tree candidate list from results at the field plot level; 3b, Calibration of tree candidate list including estimation of the number of trees per segment from results at the field plot level.

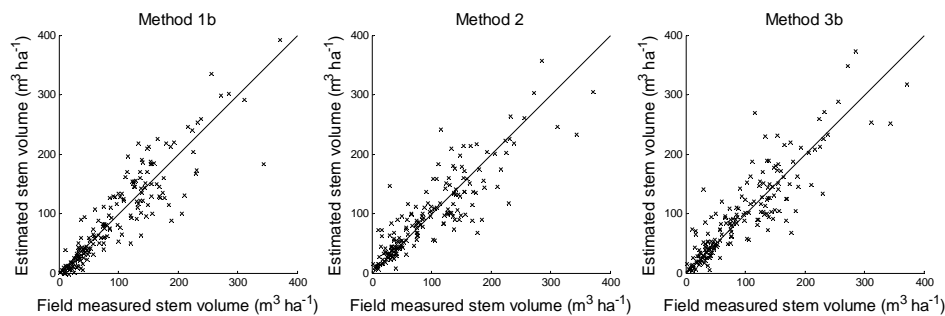| Method	Stem volume				Stem density			
	RMSE		Bias		RMSE		Bias	
	(m^3ha^{-1})	(%)	(m^3ha^{-1})	(%)	(ha^{-1})	(%)	(ha^{-1})	(%)
1a. Individual tree level	35	36	-14	-14	595	52	-403	-35
1b. Individual tree level + estimation of the number of trees per segment	33	34	-2	-3	515	45	-208	-18
2. Field plot level	35	36	0	0	421	37	-33	-3
3a. Calibration of 1a with 2	36	37	4	4	421	37	-34	-3
3b. Calibration of 1b with 2	36	37	2	2	421	37	-36	-3

Figure 3. Field-measured stem volume versus stem volume from three of the methods: 1b. Estimation at the individual tree level including estimation of the number of trees per segment, 2. Estimation at the field plot level, 3b. Calibration of tree candidate list including estimation of the number of trees per segment from results at the field plot level

.

Estimation at the field plot level (table 6, method 2) resulted in very little or no bias for stem volume (0%) and stem density (-3%). However, these are not estimates for individual trees. Calibration of the tree candidate list from the estimation at the field plot level (table 6, method 3a and 3b) reduced the RMSE and bias for stem density (37%,-3% and 37%, -3%, respectively). For stem volume, the results were

similar for all methods except for estimation at the individual tree level with one tree per tree crown segment which was an underestimation (-14%). Figure 3 shows the field-measured stem volume and the volume estimated from different methods. The estimates are distributed on both sides of the one-to-one line. The estimates are more accurate for small stem volume than for large stem volume.

4.3 Estimation of tree-size distributions

The EI was used to validate if the combined method improved the estimation of the tree-size distribution compared to estimates at the individual tree level. The EI, which measures the proportion of mismatches between two histograms, decreased after the calibration of the tree candidate list with the estimation at the field plot level. This was observed for the distributions of tree heights (109 to 96), DBH (99 to 93) and basal area (92 to 89), although the difference was most obvious for the distributions of tree heights and DBH (table 7). The EI was considerably lower for estimation of distributions at the field plot level (82, 76 and 69, respectively).

Table 7. Error index for the distributions of tree heights, DBH, and basal area at the field plot level using the methods: 1a, Estimation at the individual tree level; 1b, Estimation at the individual tree level including estimation of the number of trees per segment;

2, Estimation at the field plot level; 3a, Calibration of tree candidate list from results at the field plot level; 3b, Calibration of tree candidate list including estimation of the number of trees per segment from results at the field plot level.

Method	Error index		
	Tree height	DBH	Basal area
1a. Individual tree level	98	97	90
1b. Individual tree level + estimation of the number of trees per segment	109	99	92
2. Field plot level	82	76	69
3a. Calibration of 1 a with 2	96	92	89
3b. Calibration of 1b with 2	96	93	89

5 Discussion

The use of individual tree-based methods based on ALS data is limited without including methods that also model the non-detected trees. The non-detected trees cause biased estimates if the individual tree estimates are simply aggregated to field plots or forest stands. This study has introduced a method to produce a list of trees (not just the trees that can be detected by finding local maxima in ALS data) using a combination of analysis at individual tree level and plot level. The method used individual tree detection followed by estimation of the number of trees per segment. In the next step, target distributions of tree heights and DBH at plot level were estimated by using a k-NN approach and rescaling with the plot-wise volume estimates. These target distributions were used to compensate for missing trees in the original individual tree detection.

As expected, the individual tree detection underestimated the stem density (Persson *et al.* 2002, Maltamo *et al.* 2004). For this study area, clustered trees made the tree detection task more difficult compared to the more evenly distributed trees in most managed forests in Scandinavia. However, the estimation of the number of trees per segment using variables derived from ALS data reduced the negative bias for stem density estimates from -35% to -18%. For stem volume, the negative bias of -14% was replaced by a weak positive bias by using this method. One reason why the bias was not completely eliminated may be that individual tree detection underestimates the number of trees because a segment cannot be defined if there are too few laser returns within a tree crown. This usually occurs for small tree crowns.

The area-based method resulted in an RMSE for stem volume at the plot level comparable to previous studies in spruce-dominated forest although the relative RMSE was larger due to a smaller mean stem volume (Maltamo *et al.* 2006, Næsset 2002, Næsset 2004). One error source may be that the field plot size was small. Trees standing close to a field plot boundary may have a large part of their branches on the other side. The proportion of deciduous forest was high compared to most boreal forests in Scandinavia, approximately 30% in strata 2 and 3, which may also degrade the accuracy (Næsset *et al.* 2004). For other studies, stratification of data and the use of separate regression models for different strata have improved the accuracy (Næsset *et al.* 2004). The strata definitions are based on, for instance, tree species composition and the strata are mapped by the interpretation of aerial photographs. For this study area, it would have been difficult to use

separate regression models based on tree species because many field plots contained a mixture of tree species. However, the aim of this study was to compare the different methods and their results relative to each other using the same dataset.

We used a new technique to first estimate individual trees in whole field plots by using a k-NN approach and then multiply the number of stems with a scaling factor derived from stem volume estimates. This technique efficiently eliminated underestimations of stem volume and stem density. In several other studies, k-MSN approaches have been used to estimate forest variables with ALS data as input. One advantage of these techniques is that the covariance structure is reasonably maintained. For instance, Packalen and Maltamo (2007) predicted species-specific stand variables by using the k-MSN method without producing unrealistic combinations of estimates. For the present study, the k-NN approach was used to estimate distributions at field plot level because better estimates were achieved with this compared to regression estimates of percentiles and stem density. The elimination of the bias for stem density estimates after estimation of tree lists by using a k-NN approach and rescaling with the stem volume indicates that the applied method is useful to achieve accurate tree lists. Thus, there is no bias for stem volume and stem density estimates and the covariance structure is realistic.

The estimation of tree height and DBH percentiles at the field plot level resulted in a much lower EI than did estimation at the individual tree level. After calibration of the tree list from the individual tree detection, the EI and RMSE increased as compared to the estimates at field plot level that defined the target distribution. One reason for this is that there are few trees in each field plot, and a small error in the estimated tree height or DBH will change the histogram. The target distribution includes more trees, and small errors will not change the histogram. Previous studies have reported EI for DBH distributions of 30%-43% in large field plots (3121 to 4219 m2) (Gobakken and Næsset 2004) and 55%-78% in smaller field plots (300-1000 m2) (Gobakken and Næsset 2005, Bollandsås et al. 2007).

6. Conclusions

In this study, the bias for the estimates of stem volume and stem density from individual tree detection was reduced by using estimation of the number of trees per segment. The best estimates of distributions of tree height and DBH were obtained by using a k-NN approach with estimated percentiles at field plot level as reference variables. The stem volume of the trees in the calibrated tree list aggregated to field plot level had

an estimation error (RMSE and bias) that was similar to the direct regression estimates of stem volume.

The estimates of stem volume at field plot level were based on variables derived from the LCH. One problem with analysis at the field plot level is that the variables are derived from a population of trees and the accuracy of the estimates of these variables depends on the size of the field plots. For small field plots, tree crown segmentation is more flexible. To include variables derived from the tree crown segments, such as mean values, might improve the regression models. A mixture of tree species within field plots is also a motivation for tree crown segmentation because tree species information may be derived from the tree crown segments (Brandtberg et al. 2003, Holmgren and Persson 2004, Holmgren et al. 2008). The results for the calibrated tree list rely on the estimates from the area-based method, and better estimates would probably improve the end-result. To validate the method, it should be applied to another study area with managed boreal forest. To further improve the results, other input variables for the k-NN approach could be used. The selection method in this study was based on the Euclidean distance between field plot level estimates of tree height and DBH percentiles. A better option may be to use optimized weights for the percentiles of LCH. The EI was lower for the tree lists calibrated with results from the area-based method. Individual tree detection works best for larger trees and the area-based method probably adds most information for smaller trees. It may be possible to improve this by deriving larger trees from individual tree detection and calibrating the distribution according to results from the area-based method for smaller trees. This is in line with the method used by Maltamo et al. (2004).

7. Acknowledgements

References

AXELSSON, P., 1999, Processing of laser scanner data – algorithms and applications. ISPRS Journal of Photogrammetry and Remote Sensing, 54, pp. 138–147.

AXELSSON, P., 2000, DEM generation from laser scanner data using adaptive TIN models. In International Archives of Photogrammetry and Remote Sensing, 33, pp. 110–117.

BACKÉUS, S., WIKSTRÖM, P., and LÄMÅS, T., 2005, A model for regional analysis of carbon sequestration and

timber production. Forest Ecology and Management, 216, pp. 28–40.

BOLLANDSÅS, O. M. and NÆSSET, E., 2007, Estimating percentile-based diameter distributions in uneven-sized Norway spruce stands using airborne laser scanner data. Scandinavian Journal of Forest Research, 22, pp. 33–47.

BRANDTBERG, T., 1999, An algorithm for delineation of individual tree crowns in high spatial resolution aerial images using curved edge segments at multiple scales. In International Forum: Automated Interpretation of High Spatial Resolution Digital Imagery for Forestry, pp. 41–54.

BRANDTBERG, T., WARNER, T.A., LANDENBERGER, R.E., and MCGRAW, J.B., 2003, Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in North America. Remote Sensing of Environment, 85, pp. 290–303.

BREIDENBACH, J., GLÄSER, C., and SCHMIDT, M., 2008, Estimation of diameter distributions by means of airborne laser scanner data. Canadian Journal of Forest Research, 38, pp. 1611–1620.

DORIGO, W., HOLLAUS, M., SCHADAUER, K., and WAGNER, W., 2008, A new automated approach for co-registration of national forest inventory and airborne laser scanning data. 31, pp. 1133-1153.

ERIKSON, M. and OLOFSSON, K., 2005, Comparison of three individual tree crown detection methods. Machine Vision and Applications, 16, pp. 258–265.

FLEWELLING, J. W., 2008, Probability models for individually segmented tree crown images in a sampling context. In Proceedings of SilviLaser 2008: 8th international conference on LiDAR applications in forest assessment and inventory. Hill R.A., Rosette, J. and Suárez, J. (Eds). September 2008. Edinburgh, UK, pp. 284–293.

GOBAKKEN, T. and NÆSSET, E., 2004, Estimation of Diameter and Basal Area Distributions in Coniferous Forest by Means of Airborne Laser Scanner Data. Scandinavian Journal of Forest Research, 19, pp. 529–542.

GOBAKKEN, T. and NÆSSET, E., 2005, Weibull and percentile models for lidar-based estimation of basal area distributions. Scandinavian Journal of Forest Research, 20, pp. 490–502.

GOBAKKEN, T. and NÆSSET, E., 2008, Assessing effects of sample plot positioning errors on biophysical stand properties derived from airborne laser scanner data. In Proceedings of SilviLaser 2008: 8th international conference on LiDAR applications in forest assessment and inventory. Hill R.A., Rosette, J. and Suárez, J. (Eds). September 2008. Edinburgh, UK, pp. 14–21.

GONZALEZ, R. C. and WINTZ, P., 1987, Digital Image Processing. Second edition. Addison-Wesley Publishing Company. ISBN 0-201-11026-1. Correlation coefficient: equation (8.4-4), pp. 426.

HOLMGREN, J. and PERSSON, Å., 2004, Identifying species of individual trees using airborne laser scanner. Remote Sensing of Environment, 90, pp. 415–423.

HOLMGREN, J. and WALLERMAN, J., 2006, Estimation of tree size distribution by combining vertical and horizontal distribution of LIDAR measurements with extraction of individual trees. In T. Koukal and W. Schneider (Eds.). In Workshop on 3D Remote Sensing in Forestry, 14-15 February 2006, University of Natural Resources and Applied Life Science, Vienna, Austria, pp. 168–173.

HOLMGREN, J., PERSSON, Å., and SÖDERMAN, U., 2008, Species identification of individual trees by combining high resolution LIDAR data with multi-spectral images. International Journal of Remote Sensing, 29, pp. 1537–1552.

HYYPPÄ, J., KELLE, O., LEHIKOINEN, M., and INKINEN, M., 2001, A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. IEEE Transactions on Geoscience and Remote Sensing, 39, pp. 969–975.

JONSSON, B., JACOBSSON, J., and KALLUR, H., 1993, The forest management planning package. Theory and application. Studia Forestalia Suecica 189, Swedish University of Agricultural Sciences, Umeå, Sweden, 56 p. ISBN 91 -576-4698-8

KANGAS, A. and MALTAMO, M., 2000, Performance of percentile based diameter distribution prediction and Weibull method in independent data sets. Silva Fennica, 34, pp. 381–398.

KORPELA, I., TUOMOLA, T., and VÄLIMÄKI, E., 2007, Mapping forest plots: an efficient method combining photogrammetry and field triangulation. Silva Fennica, 41, pp. 457–469.

KÄRKKÄINEN, L., MATALA, J., HÄRKÖNEN, K., KELLOMÄKI, S., and NUUTINEN, T., 2008, Potential recovery of industrial wood and energy wood raw material in different cutting and climate scenarios for Finland. Biomass and Bioenergy, 32, pp. 934–943.

MALTAMO, M., EERIKÄINEN, K., PITKÄNEN, J., HYYPPÄ, J., and WEHMAS, M., 2004, Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. Remote Sensing of Environment, 90, pp. 319–330.

MALTAMO, M., EERIKÄINEN, K., PACKALÉN, P., and HYYPPÄ, J., 2006, Estimation of stem volume using laser scanning-based canopy height metrics. Forestry, 79, pp. 217–229.

MALTAMO, M., SUVANTO, A., and PACKALÉN, P., 2007, Comparison of basal area and stem frequency diameter distribution modeling using airborne laser scanner data and calibration estimation. Forest Ecology and Management, 247, pp. 26–34.

NÆSSET, E. and BJERKNES, K., 2001, Estimating tree heights and number of stems in young forest stands using airborne laser scanner data. Remote Sensing of Environment, 78, pp. 328–340.

NÆSSET, E., 2002, Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sensing of Environment, 80, pp. 88–99.

NÆSSET, E., GOBAKKEN, T., HOLMGREN, J., HYYPPÄ, H., HYYPPÄ, J., MALTAMO, M., NILSSON, M., OLSSON, H., PERSSON Å., and SÖDERMAN U., 2004, Laser Scanning of Forest Resources: The Nordic Experience. Scandinavian Journal of Forest Research, 19, pp. 482–499.

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, pp. 554–557.

NÆSSET, E., 2007, Airborne laser scanning as a method in operational forest inventory: Status of accuracy assessments accomplished in Scandinavia. Scandinavian Journal of Forest Research, 22, pp. 433–442

NÆSSET, E., BOLLANDSÅS, O. M., and GOBAKKEN, T., 2005, Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data. Remote Sensing of Environment, 94, pp. 541–553.

OLOFSSON, K., LINDBERG, E., and HOLMGREN, J., 2008, A method for linking field-surveyed and aerial-detected single trees using cross correlation of position images and the optimization of weighted tree list graphs. In Proceedings of SilviLaser 2008: 8th international conference on LiDAR applications in forest assessment

and inventory. Hill R.A., Rosette, J. and Suárez, J. (Eds). September 2008. Edinburgh, UK, pp. 95–104.

PACKALÉN, P. and 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, pp. 328–341.

PERSSON, Å., HOLMGREN, J., and SÖDERMAN, U., 2002, Detecting and measuring individual trees using an airborne laser scanner. Photogrammetric Engineering and Remote Sensing, 68, pp. 925–932.

POLLOCK, R. J., 1996, The automatic recognition of individual trees in aerial images of forests based on a synthetic tree crown image model. PhD thesis, University of British Columbia, Canada. ISBN 0-612-14815-7.

REYNOLDS, M. R., BURK, T. E., and HUANG, W. C., 1988, Goodness-of-fit tests and model selection procedures for diameter distribution models. Forest Science, 34, pp. 373–399.

SOININEN, A., 2004, TerraScan for MicroStation, User's Guide. Terrasolid Ltd., Jyvaskyla, Finland. 132 p.

SOLBERG, S., NÆSSET, E., and BOLLANDSÅS, O. M., 2006, Single tree segmentation using airborne laser scanner data in a structurally heterogeneous spruce forest. Photogrammetric Engineering and Remote Sensing, 72, pp. 1369–1378.

SÖDERBERGH, I. and LEDERMANN, T., 2003, Algorithms for simulating thinning and harvesting in five European individual-tree growth simulators: a review. Computers and Electronics in Agriculture, 39, pp. 115–140.

ZELLNER, A., 1962, An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. Journal of the American Statistical Association, 57, pp. 348–368.