

Prediction of Stem Attributes by Combining Airborne Laser Scanning and Measurements from Harvesters

Johan Holmgren, Andreas Barth, Henrik Larsson and Håkan Olsson

Holmgren, J., Barth, A., Larsson, H. & Olsson, H. 2012. Prediction of stem attributes by combining airborne laser scanning and measurements from harvesters. *Silva Fennica* 46(2): 227–239.

In this study, a new method was validated for the first time that predicts stem attributes for a forest area without any manual measurements of tree stems by combining harvester measurements and Airborne Laser Scanning (ALS) data. A new algorithm for automatic segmentation of tree crowns from ALS data based on tree crown models was developed. The test site was located in boreal forest (64°06'N, 19°10'E) dominated by Norway spruce (*Picea abies*) and Scots Pine (*Pinus sylvestris*). The trees were harvested on field plots, and each harvested tree was linked to the nearest tree crown segment derived from ALS data. In this way, a reference database was created with both stem data from the harvester and ALS derived features for linked tree crowns. To estimate stem attributes for a tree crown segment in parts of the forest where trees not yet have been harvested, tree stems are imputed from the most similar crown segment in the reference database according to features extracted from ALS data. The imputation of harvester data was validated on a sub-stand-level, i.e. 2–4 aggregated 10 m radius plots, and the obtained RMSE of stem volume, mean tree height, mean stem diameter, and stem density (stems per ha) estimates were 11%, 8%, 12%, and 19%, respectively. The imputation of stem data collected by harvesters could in the future be used for bucking simulations of not yet harvested forest stands in order to predict wood assortments.

Keywords ALS data, pre-harvest inventory, tree detection

Addresses *Holmgren, Larsson & Olsson*: Swedish University of Agricultural Sciences, Forest Resource Management, Umeå, Sweden; *Barth*: The Forestry Research Institute of Sweden, Uppsala, Sweden

E-mail johan.holmgren@slu.se

Received 10 November 2011 **Revised** 16 March 2012 **Accepted** 20 March 2012

Available at <http://www.metla.fi/silvafennica/full/sf46/sf462227.pdf>

1 Introduction

There is a demand for accurate and high-resolution data from pre-harvest inventories to implement modern tools for optimized wood flow based on industrial requirements (Wilhelmsson et al. 2002, Moberg and Nordmark 2006). In Sweden, most planning in practical forestry is based on stand records or pre-harvest field inventories normally delivering stand averages, but these data are neither accurate enough nor detailed enough for optimization of round-wood deliveries. Accurate predictions of round-wood assortments require information on distributions of stem geometric attributes, e.g. diameters and height, as well as tree species and information of stem defects. Pre-harvest inventories including detailed field inventory in every stand are expensive and therefore remote sensing methods are considered. Stem diameter distributions have been estimated using Airborne Laser Scanning (ALS) with field data from inventories of sample plots as reference data (Gobakken and Næsset 2004, Maltamo et al. 2009). It is however expensive to collect field data needed for stem diameter estimations. Also, common field inventory methods only include measurements of the stem diameter at one place along the stem (DBH) whereas data for the whole stem profile is important for the forest industry.

Detailed stem data are recorded by cut-to-length (CTL) harvesters performing final felling. These data are normally only utilized to measure logging production and provide input data for controlling transportation to the mills. However, these data could potentially be used to estimate the round-wood assortments of not yet harvested stands with similar forest characteristics. This could be possible if there is a data source that can be used to link stem data from harvested areas to non-harvested trees with similar characteristics. Airborne laser scanning (ALS) can be used for tree detection and estimation of tree attributes for individual trees using Individual Tree Crown (ITC) methods (Hyypä et al. 2008). The tree crown delineation algorithms provide estimates of position and characteristics of individual tree crowns. The tree crown segments can then be used to extract variables from ALS data for individual tree crowns in order to identify tree species or

estimate variables at a tree level, for example tree height, stem diameter, and stem volume (Hyypä et al. 2001, Persson et al. 2002, Koch et al. 2006, Solberg et al. 2006). In recent studies, non-parametric estimation methods, for instance the Most Similar Neighbour (MSN) method (Moeur and Stage 1995) or the Random Forest (RF) method (Breiman 2001) have been found useful for estimation of stem attributes using several variables derived from laser reflection located inside individual tree crowns. In Finland, tree species, tree height, stem diameter, and stem volume have been estimated simultaneously using both the MSN-method and RF with explanatory variables derived with alpha-shape models and the height and intensity distribution of laser data within manually delineated tree crowns (Vauhkonen et al. 2010). In Norway, stem volume was estimated for all trees on plots and for trees of different tree species applying both the MSN-method and the RF method (Breidenbach et al. 2010). To reduce bias due to segmentation errors, the analysis was done on a tree crown segment level: zero, one, or several tree stems were linked to each tree crown segment. Then a crown segment was imputed to a target tree crown segment without stem data all trees that had been linked to the crown segment were also imputed to that target segment. This approach efficiently reduced the bias that otherwise is due to segmentation errors and occurs after aggregation of tree level predictions for an area, e.g. a raster cell or forest stand.

In this study, the development of a method for imputation of tree stems was motivated because attributes of actual tree stems are needed to be able to predict the wood assortments that can be expected from trees within forest stands. Today, detailed stem data are available for each tree from the harvester's bucking system but data are obtained too late to support decisions regarding allocations of harvester operations. If harvesters in the future are equipped with high precision positioning systems, tree positions could be measured during harvesting, making it possible to automatically link tree stem data with tree crown polygons derived from ALS data. A future operational system could continuously receive stem data sent from operating harvesters and these data could be used to continuously update an ALS based reference database enabling improved

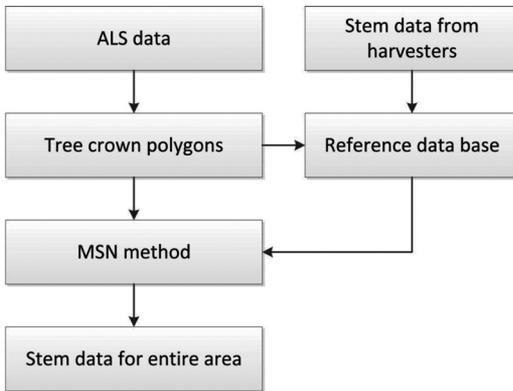


Fig. 1. Flow chart of the complete methods for prediction of stem attributes by combining ALS and harvester data.

prediction of stem attributes for not yet harvested forest stands. Thus, the proposed method uses existing measurements from the CTL-harvester and would with a high precision positioning system not require any manual field work. There are technical solutions that could be used for high precision positioning of the harvester head relative to the harvester but so far there have not been enough incentives for the machine manufacturers to add this functionality to new machines. The potential for enhanced use of geographical information, for example the concept presented in this article, could be one incentive for manufacturers to equip harvesters with improved positioning systems. In this study, a high precision positioning system was not available on-board the harvester and therefore the tree positions on plots were manually measured and linked to harvester measurements of tree stems using identification numbers. As an important part of the complete method an algorithm was validated for automatic segmentation of tree crowns based on geometric tree crown models. The segmentation method is trained with known tree positions, thus the segmentation results could be continuously improved as more ground truth are automatically collected by harvesters in the region.

The objective of this study was to demonstrate a new method for which automatically measured tree stems are imputed based on tree crown segmentation from ALS data (Fig. 1). The

method could in the future be used to predict stem attributes for not yet harvested forest. The stem data could then be imported to already existing systems for prediction of wood assortments in order to improve information for planning of harvesting operations.

2 Material and Methods

2.1 Study Area

The study was located at a test site in northern Sweden (64°06'N, 19°10'E) where 20 forest stands, planned to be harvested, were selected. The forest stands were dominated by Norway spruce (*Picea abies* L. Karst.) or Scots Pine (*Pinus sylvestris* L.). Four field plots were systematically placed within each forest stand using a grid with 50 m inter-node distance. Due to problems with data collection by the harvester described in section 2.3 two plots in one stand and one plot in another stand were excluded for the analysis of this study. The total area of all field plots within a stand is referred to here as a sub-stand with an area of 1257 m² in the most common case when all four plots within a stand were used.

2.2 Field Data

The field plot inventory was performed from June to August 2008. For all trees on a 10 m radius field plot, the stems were callipered 1.3 m above ground level (DBH) if the DBH was at least 40 mm. For all callipered trees, tree species was recorded and stem position was measured relative to the field plot centre using ultrasonic trilateration with three transponders. Also for all callipered trees on the plot, an identification number was painted on the tree stem. The plot centre position was measured using differential GPS logging data during the time when the tree positions were measured relative to the plot centre. The GPS data were then post-processed with data from a reference station, which should produce sub-meter accuracy under optimal conditions. There is a description of the field data in Table 1.

Table 1. Descriptive statistics of sub-stands from the field inventory; stem diameter (DBH), stem density (stems), basal area (BA), and proportion of species (Scots pine, Norway spruce and deciduous trees).

Sub-stand	Mean DBH (mm)	Std. DBH (mm)	Min. DBH (mm)	Max. DBH (mm)	Stems (ha ⁻¹)	BA (m ² ha ⁻¹)	Pine (%)	Spruce (%)	Dec. (%)
1	183	80	46	341	764	24	89	9	2
2	201	64	44	365	923	32	91	6	3
3	240	127	41	428	334	19	87	13	0
4	190	92	50	382	780	27	13	81	6
5	221	87	46	426	684	30	19	79	1
6	155	94	40	395	1130	29	44	48	8
7	153	89	41	373	915	22	11	80	9
8	162	76	42	370	1106	28	7	83	10
9	155	83	41	387	1098	27	64	30	6
10	178	112	42	405	740	26	75	25	0
11	194	89	49	411	971	35	2	95	3
12	207	84	47	447	963	38	0	100	0
13	166	75	44	300	796	21	0	93	7
14	171	62	45	380	1162	30	1	82	17
15	195	100	42	387	775	29	50	45	4
16	215	83	52	395	621	26	18	77	5
17	209	60	52	351	844	31	0	95	5
18	207	60	72	400	676	25	29	71	0
19	217	58	43	375	772	31	97	0	3
20	221	84	42	383	605	26	25	59	16

2.3 Harvester Data

Trees on the plots were harvested with a Ponsse Buffalo Dual harvester during winter 2009. A special module of a harvester GIS program was developed for registration of trees on the plots used in this research project. The position of the harvester and the field plots were visible on a digital map and if a tree was harvested within a field plot, the operator was requested to enter the tree identification number painted on the tree stem. The stem data were recorded according to the StanForD standard (Standard for Forest ... 2012) and the tree identification number was saved in an associated stem-file. The stem diameter was continuously measured by the harvester along the processed part of the tree stem and registered with 10 cm intervals. The program for registration of tree stems did not operate in an optimal way for the specific harvester used in this case, which produced linking errors. The linking was manually verified and because of errors three plots were excluded from further analysis: these were two plots in one stand and one plot in another stand.

2.4 Airborne Laser Scanner Data

The laser data were acquired on the 3rd and 5th of August 2008 using the TopEye MKII system S/N 425 operated on-board a helicopter from an altitude of 500 m and 250 m above ground level for main strips and cross strips, respectively. Because of overlapping strips the measurement density varied between plots but the average density was approximately 15 laser returns per m², including both first and last returns. The laser wavelength was 1064 nm, the pulse length 4 ns, and the pulse repetition frequency 50 kHz. A palmer scanner was used with a maximum scan angle of $\pm 20^\circ$ across flight direction and $\pm 14^\circ$ along flight direction.

2.5. Field Plot Matching

The spatial patterns of tree positions and tree sizes from field data and tree detection with ALS data were used as input to an earlier developed algorithm (Olofsson et al. 2008) for correction of

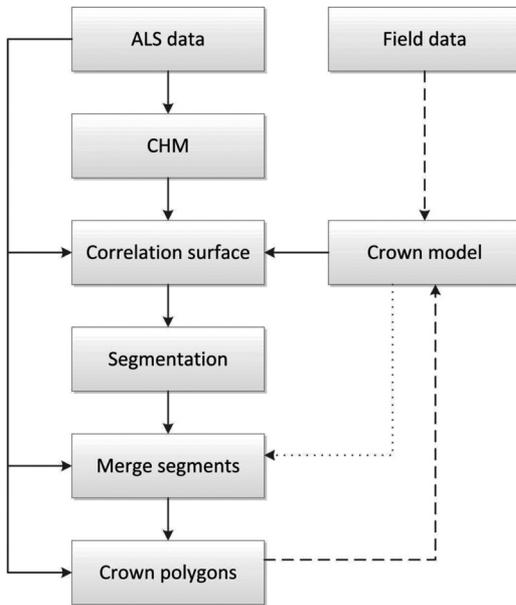


Fig. 2. Flow chart of the method used for automatic delineation of tree crowns based on ALS data.

DGPS (Differential Global Positioning System) measured field plot centre and tree positions. The algorithm used an estimated position of the field plot centre to define a search area that must contain the real field plot centre. The search area was in this study ± 20 m from the estimated field plot centre obtained from DGPS measurements. Tree detection from ALS data was collected to a list with tree position coordinates, x and y , and tree height representing the tree size. For the field plots a similar list was required with tree position coordinates, x and y , relative to the plot centre and stem diameter representing the tree size. The tree lists were used to create two single tree position images. For each image, a tree was represented as a Gaussian function where the x and y coordinates determined the position within the image, the tree size variable determined the amplitude of the Gaussian function, and the standard deviation of the Gaussian function was set to the expected tree position precision. Since large trees are often detected from above whereas small trees are often hidden the maximum surface was used, i.e. from all of the Gaussian functions that cover the same area, the highest value was chosen. The two single

tree position images were then cross correlated to find the closest match between the patterns in the two images. The field plot image was rotated with one degree steps within an interval of ± 10 degrees between each correlation run in order to compensate for possible compass errors and translated with 0.5 m steps within the search area. The position and rotation with the highest correlation coefficient was assumed to be the place where the field plot was located.

2.6 Tree Crown Segmentation

The segmentation procedure consisted of two phases: 1) a training phase where known tree locations on sample plots were used in order to predict optimal parameter settings as a function of variables that can be derived from ALS data, and 2) a prediction phase in order to produce tree crown segments for the entire area. The segmentation algorithm has four steps: create a height model, create a correlation surface, segmentation, and finally merge segments (Fig. 2). The used raster cell size was 0.25 m. In the prediction phase, the algorithm uses a ratio between crown radius and crown height which is a function of ALS derived variables. The ratio is predicted in the training phase on the field plots with known tree coordinates before the algorithm can be applied on the entire laser scanned forest. The algorithm was recently validated and compared with other algorithms for forests in Norway, Sweden, Germany, and Brazil (Vauhkonen et al. 2012).

2.6.1 Create a Height Model

A Digital Surface Model (DSM) was created by assigning each raster cell the highest z -value of laser returns located within the cell. A Digital Elevation Model (DEM) was created and nDSM was calculated as the difference between DSM and DEM. A canopy height model (CHM) was created to describe the height of tree crowns at each location. A raster cell value of the CHM was set to the nDSM value if the cell value was above a height threshold (2 m), otherwise the value was set to zero. Some raster cells with a zero value could be inside a tree crown due to the lack of

laser returns at that location or because the laser beam penetrated through gaps in the canopy. In order to set a value to these raster cells with no data, these areas need to be distinguished from areas outside of the tree crown, i.e. the crown area needs to be defined. First, a binary crown area raster was derived from the nDSM, in which a raster cell was set to the value of one if the corresponding nDSM raster cell had a value greater than a height threshold (2 m); otherwise the raster cell was left as a zero value. The crown area raster was then updated by morphological closing with a structure element of 3×3 cells. The next step was to update the CHM based on the crown area raster in order to set non-zero values of the CHM at crown locations, i.e. value of one in the crown area raster. 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, then that value was assigned; otherwise, the mean value of non-zero values was assigned.

2.6.2 Create a Correlation Surface

The cell values of a correlation raster were set to the maximum found correlation from tests with geometric tree models with the origin placed at the centre of a raster cell. For each raster cell, different geometric models, i.e. generalized ellipsoids (Pollock 1996), with the shape parameter set to two (Eq. 1) were used to calculate the height of the ellipsoid surface (h), where a and b are parameters and r is the horizontal distance from model origin. The correlation was calculated between h and z -values of laser returns within the horizontal model radius, parameter b .

$$h = a \sqrt{1 - \left(\frac{r}{b}\right)^2} \quad (1)$$

Different models were tested for a raster cell if the CHM had a height value greater than zero: parameter a was set to the value of the CHM at the model origin, and different radii were tested, $b = 0.5, 0.7, r_{max}$, where r_{max} is the maximum expected radius set as a proportion of model height (parameter a). In order to obtain higher correla-

tion values with the solid geometric crown model some laser returns were removed (Solberg et al. 2006): a laser return was removed if there was another laser return within a horizontal distance of 0.3 m with a higher z -value. This pre-processing increased the correlation because returns inside tree crowns were discarded making the correlation with a solid geometric model higher. The correlation surface was smoothed by filtering the surface three times with a 3×3 Gaussian kernel. The degree of smoothing was chosen to produce segments in the next step that are small enough to include only one of the smallest trees in the data but large enough to allow for tests of models to decide if segments should be merged or not.

2.6.3 Segmentation

A starting point (i.e. a seed) was placed at each raster cell with a non-zero CHM value, and with a positive value of the correlation surface. For each seed, the current location was updated by changing the position to the neighbour cell with the highest value of the smoothed correlation surface. This was repeated until the position could not be updated because a local maximum of the smoothed correlation surface had been reached. The seeds with the final location at the same local maximum defined a segment. A raster cell that had not been included into a segment but was enclosed by raster cells that were all from the same segment was assigned to the surrounding segment.

2.6.4 Merge Segments

The aim of the preceding steps was to create segments that were small enough to include small trees but large enough to make it possible to test models, i.e. include more than a minimum number of laser returns. A neighbour segment was defined as a segment having an edge in common with the tested segment. The next step was to merge segments with the aim of removing segments that only covered parts of a tree. First, small segments containing less than a minimum number of returns, in this study set to eight, were merged with the neighbour segments with the longest

edge in common with the small segment. This was repeated until no more segments were merged. For each of the remaining segments, models (i.e. Eq. 1) were used to decide if a segment should be merged or not to a neighbour segment. The model origin was placed at the raster cell with the highest value of the correlation surface (segment centre) within the tested segment and a test value was calculated using only laser data within that segment. The model was also placed at the centre of a neighbour segment, i.e. another segment with a common border with the segment, and a test value was calculated using only laser data within the tested segment. If the value at the tested segment centre yielded a higher test value than any model where the origin was at a neighbour segment centre, the segment was not merged; otherwise, the segment was merged with the neighbour segment for which the highest test value was calculated. This was repeated until no more segments were merged. The test value used to decide whether to merge two segments was the weighted correlation between h and z -values (weight = distance above ground level) multiplied by a penalty factor, p , calculated with Eq. 2, where r_p is the predicted and r_o is the observed radius-height ratio for a tree crown.

$$p = e^{-\frac{(r_p - r_o)^2}{2\sigma^2}} \quad (2)$$

The σ value in Eq. 2 was set to 0.1 but could be adjusted according to the expected uniformity of crown shapes in the specific forest.

2.6.5 Prediction of Radius-Height Ratio

In order to predict the best radius-height ratio (i.e. the r_p value) to be used for the segmentation, training data were used from field plots with known tree locations. Several segmentations, each with one expected radius-height ratio (0.06, 0.08, ... , 0.14), were tested and the results were evaluated by counting the number of trees within each of the segments for the segments that were located inside a field plot. For each segment laser variables were also derived. For this study five classes (percentile 20, 40 ... 100) were defined using the maximum height of the segment, but other variables, such as variables related to tree

species, could also be used. The radius-height ratio with the highest proportion of one field tree within a segment was judged as the best radius-height ratio for a class. For this study a cross validation procedure was used by excluding field data for a specific square kilometre when producing the training data used for segmentation of that square kilometre.

2.6.6 Segmentation of the Entire Area

After prediction of the radius-height ratio, segmentation can be performed for the entire laser scanned area with different predicted radius-height ratios for different classes derived from laser data, i.e. percentiles. The segments defined the area of individual tree crowns and based on the segmentation results variables were derived to be used as independent variables for the imputation of tree stems. One of these variables was the tree crown width which was calculated using the area of the crown segment and the area equation for a circle. All other variables were derived from the laser data extracted within the tree crown segment which were above a height threshold. The height distribution of the extracted laser returns was used to calculate height percentiles (20 ... 90, 95, and 100 percentiles), average height, and standard deviation of heights. Also, the mean intensity (i.e. amplitude) of the laser returns, proportion of single returns, number of returns from the crown, and laser derived crown base height were derived. The used height threshold was 1 m and 10% of the maximum height within the segment.

2.7 Imputation

Tree stems were linked to a laser derived crown polygon if they were inside the polygon. If a tree stem was outside any crown polygon the stem was linked to the nearest tree crown polygon (Fig. 3). The MSN method (Moeur and Stage 1995) was used for imputation of tree crowns using the `yalmp` package as part of the R-project (The R foundation...2012). Only segments fully contained within a field plot were used for training but all segments with predicted tree top location within a field plot were used for prediction. The sum of

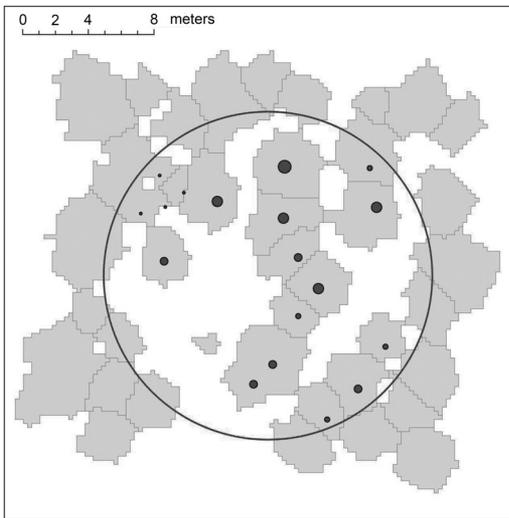


Fig. 3. Tree crown segments and harvester measured trees (small circles, size proportional to stem diameter) within a field plot (large circle).

stem volume, DBH, stem diameter at last cut and tree height were used as depended variables for the imputation. The sums were calculated with zero, one, or several tree stems that had been linked to a segment. The above described variables extracted from ALS data for each crown segment were used as the independent variables for the imputation.

2.8 Validation

Predictions of forest attributes were validated on a sub-stand level. A cross validation procedure was applied: the sub-stand that was predicted was excluded from the training data and this was repeated until all sub-stands had been predicted. The RMSE and Bias were calculated using Eq. 3 and 4, where \hat{y}_i is the predicted and y_i is the observed variable for stand i and N is number of stands.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (3)$$

$$\text{Bias} = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i) \quad (4)$$

An error index (EI) (Reynolds et al. 1988) was calculated with a fixed number of intervals ($m = 10$) and fixed sized intervals (50 mm) with Eq. 5 where p_j is the predicted and o_j is the observed number of trees in class j and n is the total number of trees:

$$EI = \frac{1}{n} \sum_{j=1}^m |p_j - o_j| \quad (5)$$

3 Results

The performance of the tree detection algorithm was validated by counting field measured trees that were linked to a segment derived from ALS data. For the segments fully contained within a plot 64%, 16%, 5%, and 2%, were linked to one, two, three, and four or more trees, respectively. Also, 13% of all segments fully contained within a plot were not linked to any harvester measured tree stems and were therefore regarded as false segments. The proportion of detected trees was calculated as the number of one-to-one links, between harvester measured trees and ALS derived crown segments, divided by the total number of harvester processed trees. If only one laser detected tree was linked to each ALS derived segment, 73% of the harvester measured trees would have been classified as detected trees. The proportion of segments that were not linked to any harvester measured tree was 11% of the number of harvester measured trees that were linked to a segment fully contained within a plot. The detection rate was highest for tall trees: a higher proportion of harvester measured trees with a large stem diameter could be linked to a tree crown segment compared with trees with small stem diameter (Fig. 4). The detected trees that were classified as false (commission errors) were usually small trees: a higher proportion of tree crown segments with a low height could not be linked to a harvester measured tree compared with tall trees (Fig. 5).

The imputed and harvester measured sums of stem volume were compared at segment level where imputed false segments contributed with zero stem volume (Fig. 6). When aggregated at sub-stand level the RMSE values were between 8 and 19% for all validated variables using cross

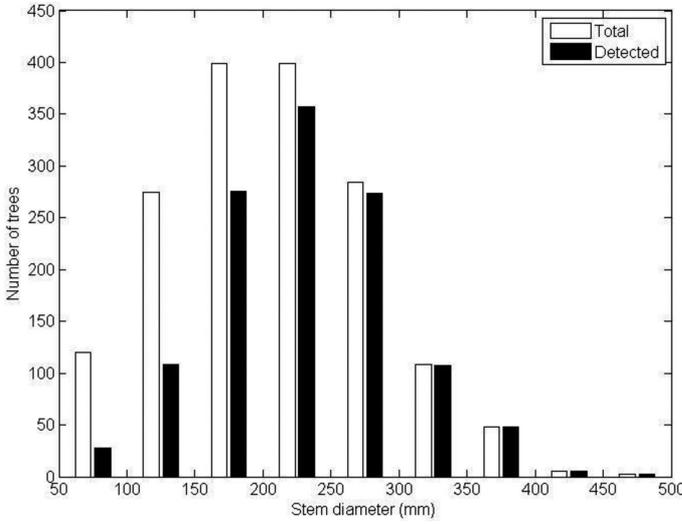


Fig. 4. The stem diameter (DBH) distribution for all trees on the plots (total) and the stem diameter for linked trees on all plots (detected). The maximum stem diameter was used if several trees were linked to the same crown segment.

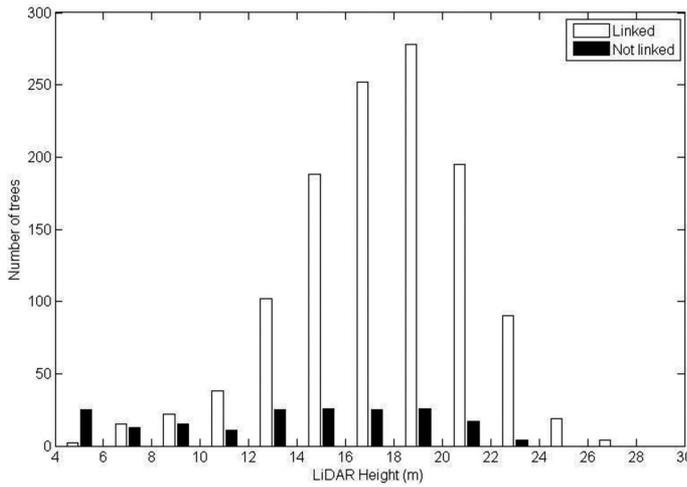


Fig. 5. The height distributions of crown segments on all plots that were linked or not linked to a field tree, respectively.

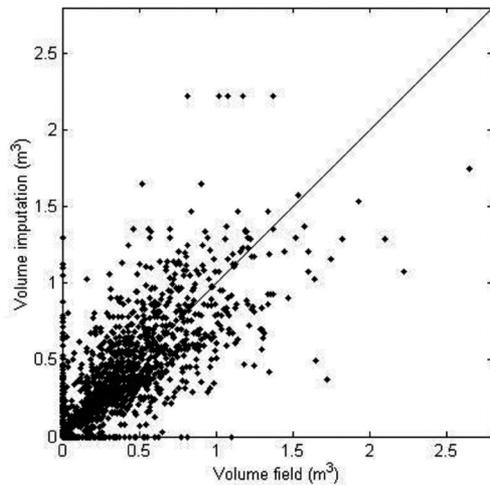


Fig. 6. Stem volume per segment (m^3) from imputation plotted against harvester measured stem volume; validation done by excluding segments from same sub-stand from training data.

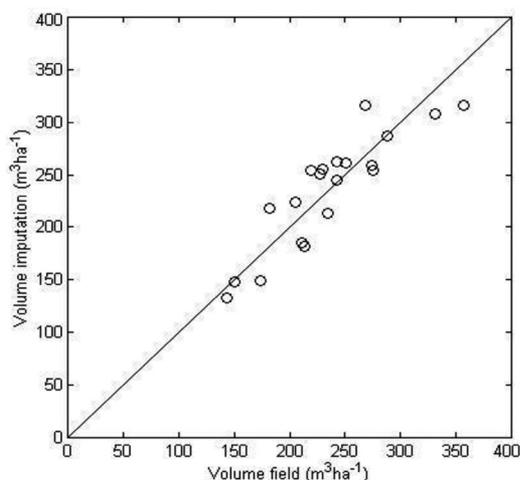


Fig. 7. Predicted stem volume at sub-stand level plotted against harvester measured stem volume. Validation was done by excluding segments from the same sub-stand from training data.

validation with all segments within the same sub-stand excluded from training data. The bias was near zero for stem volume estimates, but there was a small negative bias for mean tree height and stem diameter, and a small positive bias for stem number (Table 2). A high deviation from the one-to-one line was found for three sub-stands with high volume (Fig. 7). The accuracy of the stem diameter distribution predictions was visualized by histograms for the forest stand with highest and lowest EI (Fig. 8). For stem diameter distribution estimations the average EI value was 0.44 at sub-stand level. However, a high variation could be observed in estimation accuracy of the stem diameter distribution.

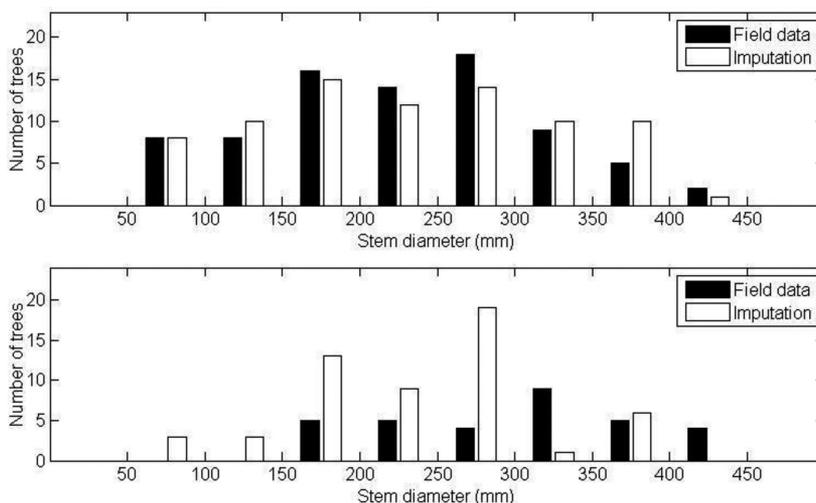


Fig. 8. Number of stems in 50 mm stem diameter (DBH) intervals from field data and imputed trees; Top: the sub-stand with lowest EI (0.2); Bottom: the sub-stand with highest EI (1.4). Validation was done by excluding segments from the same sub-stand from training data.

Table 2. Imputation results validated at sub-stand level by excluding segments from the same sub-stand from training data.

	RMSE	Rel. RMSE (%)	Bias	Rel. Bias (%)
Stem volume ($m^3 \cdot ha^{-1}$)	25.0	10.58	0.05	0.02
Mean tree height (m)	1.13	7.75	-0.42	-2.88
Mean stem diameter (mm)	24.95	11.76	-11.72	-5.52
Stem number (ha^{-1})	135	18.72	34.09	4.73

4 Discussion

The aim was to develop an automatic method where tree crown polygons delineated from ALS data are used for imputation of tree stems measured by harvesting machinery. In this study, the imputation of tree stem measurements obtained from a harvester was for the first time imputed with high resolution ALS data.

The accuracy of tree stem imputations will depend on how ALS metrics can be extracted from ALS data. One error source is segments with no linked field trees, usually referred to as commission errors. The non-linked segments could be caused by false tree crown segments, linking errors, but also the lower size limit for a tree to be harvested. The linking errors could be caused by relative positioning errors of field trees, leaning trees, or problems to automatically correct the position of field plots with the automatic field plot matching algorithm (Olofsson et al. 2008). A high proportion of non-linked segments with low heights could be the result of the lower size limit for trees to be harvested. The problem with either no or several trees linked to a segment was solved by using all crown segments for the imputation and using the sum of stem attributes, e.g. total stem volume of all trees linked to a segment, as reference variables (Breidenbach et al. 2010). The tree crowns with no linked trees were usually estimated as segments with no or low volume (i.e. usually less than 0.5 m³). The estimates were aggregated to sub-stand level which produced estimates with low bias for stem volume, mean tree height, mean stem diameter, and stem density. The imputation results could possibly be improved by deriving more variables from ALS data, for example, those related to tree species. High accuracy tree species classification results have been obtained in earlier studies using variables extracted from ALS data within crown segments (Holmgren and Persson 2004; Holmgren et al. 2008). However, only the detected trees were classified. The results obtained from imputation using the semi-ITC approach, i.e. using the sum of stem volume for zero, one, or several trees linked to ALS derived segments, are similar to results obtained by Breidenbach et al. (2010).

The RMSE values were similar to those

obtained in other studies in Scandinavian forests with validation performed at stand level but were low compared to other studies where validations were performed at plot level (Næsset et al. 2004). In this study, the imputation results were validated at sub-stand level, i.e., 2–4 aggregated field plots within a forest stand, thus making the validation unit more similar to large plots than forest stands. One advantage of the imputation of tree stems, compared with using for example regression analysis for prediction of stem attributes, is that imputed tree stems can be used for bucking simulations in order to predict wood assortments. To make this method operational, the harvester should be equipped with a high precision positioning system in order to automatically measure the position of harvested trees. A positioning system that saves the position of each tree in the associated stem-file would make the method completely automatic and a reference database with tree stems linked to ALS data could be continuously updated. For this study, the harvester's geographical information system was supplied with functionality to add manually entered identification numbers in stems-files for trees harvested within field plots. In the future, the position of trees should instead be automatically inserted by a high precision positioning system. This functionality should be supplied by the manufacturers of harvesters. It is possible that MSN methods produce low accuracy estimates if a small training dataset is used because then the probability is high that there are no similar data points in the reference database. If tree stems were geo-referenced by a positioning system on the harvester the stem data could be linked to remote sensing data automatically. It would then be possible to produce large reference databases which probably would make it possible to improve the estimation results. It is also possible that the imputation results could be improved with other imputation methods, other variables extracted from the ALS data within each tree crown segment, or transformation of these variables. The influence of the independent variables in the MSN-method is determined by the correlation with predicted variables, thus transformation could improve the results. The imputation of tree crown segments with linked tree stems produced unbiased estimates of stem volume. However, the predicted stem diameter

distribution was only a result of the trees that were linked to the imputed tree crown segments and therefore a high variation for the accuracy of stem diameter estimations could be observed.

In the future the potential of the imputation of tree stems will be evaluated using bucking simulations. Further research is needed to find out if not only tree stems with similar sizes, as was evaluated in this study, can be imputed, but also if quality parameters can be predicted, such as tree species and number of branches. Three dimensional structures obtained from high resolution ALS data could possibly provide us with the information needed.

Acknowledgements

This research was financed by the Swedish Research Council Formas, as part of the IRIS project within the ERA-Net Wood Wisdom, the Nordic Forest Research Co-operation Committee (SNS), and by the Research Programme Heureka. We would like to thank Alexandra Frank and her colleagues at the forest company Sveaskog for planning the harvesting operation and Lena Gustafsson at the Unit for Field-based Forest Research at the Swedish University of Agricultural Sciences for supporting the project. The entrepreneur Lars Anders Gustafsson harvested and entered identification numbers for trees on the field plots. The programming of the harvester's software for collection of stem files with identification numbers was done by Thorbjörn Cruse. The method for correction of the field plot positions was developed by Dr. Kenneth Olofsson. We would also like to thank Dr. Heather Reese for valuable comments.

References

- Breidenbach, J., Naesset, E., Lien, V., Gobakken, T. & Solberg, S. 2010. Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. *Remote Sensing of Environment* 114(4): 911–924. doi:10.1016/j.rse.2009.12.004.
- Breiman, L. 2001. Random forests. *Machine Learning* 45(1): 5–32. doi:10.1023/a:1010933404324.
- Gobakken, T. & 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(6): 529–542. doi:10.1080/02827580410019454.
- Holmgren, J. & Persson, Å. 2004. Identifying species of individual trees using airborne laser scanner. *Remote Sensing of Environment* 90(4): 415–423. doi:10.1016/S0034-4257(03)00140-8.
- Holmgren, J., Persson, Å. & Söderman, U. 2008. Species identification of individual trees by combining high resolution LIDAR data with multi-spectral images. *International Journal of Remote Sensing* 29(5): 1537–1552. doi:10.1080/01431160701736471.
- Hyypä, J., Kelle, O., Lehtikoinen, M. & Inkinen, M. 2001. A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. *IEEE Trans Geosci Remote Sensing* 39(5): 969–975.
- , Hyypä, H., Leckie, D., Gougeon, F., Yu, X. & Maltamo, M. 2008. Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing* 29(5): 1339–1366. doi:10.1080/01431160701736489.
- Koch, B., Heyder, U. & Weinacker, H. 2006. Detection of individual tree crowns in airborne lidar data. *Photogrammetric Engineering and Remote Sensing* 72(4): 357–363.
- Maltamo, M., Næsset, E., Bollandas, O.M., Gobakken, T. & Packalen, P. 2009. Non-parametric prediction of diameter distributions using airborne laser scanner data. *Scandinavian Journal of Forest Research* 24(6): 541–553. doi:10.1080/02827580903362497.
- Moberg, L. & Nordmark, U. 2006. Predicting lumber volume and grade recovery for Scots pine stems using tree models and sawmill conversion simulation. *Forest Products Journal* 56(4): 68–74.
- Moeur, M. & Stage, A.R. 1995. Most similar neighbor – an improved sampling inference procedure for natural-resource planning. *Forest Science* 41(2): 337–359.
- 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(6): 482–499.

- doi:10.1080/02827580410019553.
- Olofsson, K., Lindberg, E. & 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. Paper presented at the SilviLaser 2008, 8th international conference on LiDAR applications in forest assessment and inventory, Edinburgh, 17–19 September 2008.
- Persson, Å., Holmgren, J. & Söderman, U. 2002. Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering and Remote Sensing* 68(9): 925–932.
- Pollock, R. 1996 The automatic recognition of individual trees in aerial images of forests based on a synthetic tree crown image model. University of British Columbia, Vancouver, Canada.
- Reynolds, M.R., Burk, T.E. & Huang, W.C. 1988. Goodness-of-fit tests and model selection procedures for diameter distribution models. *Forest Science* 34(2): 373–399.
- Solberg, S., Naesset, E. & Bollandsas, O.M. 2006. Single tree segmentation using airborne laser scanner data in a structurally heterogeneous spruce forest. *Photogrammetric Engineering and Remote Sensing* 72(12): 1369–1378.
- Standard for forest machine data and communication. 2012. Available at: www.skogforsk.se/en/About-skogforsk/Collaboration-groups/StanForD/. [Cited 15 Mars 2012].
- The R foundation for statistical computing (R-project). 2012. Available at: www.r-project.org. [Cited 15 Mars 2012].
- Vauhkonen, J., Korpela, I., Maltamo, M. & Tokola, T. 2010 Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. *Remote Sensing of Environment* 114(6): 1263–1276. doi:DOI 10.1016/j.rse.2010.01.016.
- , Ene, L., Gupta, S., Heinzel, J., Holmgren, J., Pitkänen, J., Solberg, S., Wang, Y., Weinacker, H., Hauglin, K.M., Lien, V., Packalén, P., Gobakken, T., Koch, B., Næsset, E., Tokola, T. & Maltamo, M. 2012. Comparative testing of single-tree detection algorithms under different types of forest. *Forestry* 85(1): 27–40. doi: 10.1093/forestry/cpr051.
- Wilhelmsson, L., Arlinger, J., Spangberg, K., Lundqvist, S.O., Grahn, T., Hedenberg, O. & Olsson, L. 2002. Models for predicting wood properties in stems of *Picea abies* and *Pinus sylvestris* in Sweden. *Scandinavian Journal of Forest Research* 17(4): 330–350.

Total of 22 references