Article

Retrieval of Forest Aboveground Biomass and Stem Volume with Airborne Scanning LiDAR

Ville Kankare [1],*, Mikko Vastaranta [1], Markus Holopainen [1], Minna Räty [2], Xiaowei Yu [3], Juha Hyyppä [3], Hannu Hyyppä [4,5], Petteri Alho [6] and Risto Viitala [7]

[1] Department of Forest Sciences, University of Helsinki, PL 27, FI-00014 Helsinki, Finland;
E-Mails: mikko.vastaranta@helsinki.fi (M.V.); markus.holopainen@helsinki.fi (M.H.)

[2] Forest Resources and Climate Unit, Institute for Environment and Sustainability, Joint Research Centre, Via E. Fermi 2749, I-21027 Ispra (VA), Italy; E-Mail: minna.raty@jrc.ec.europa.eu

[3] Department of Remote Sensing and Photogrammetry, Finnish Geodetic Institute, PL 15, FI-02431 Masala, Finland; E-Mails: xiaowei.yu@fgi.fi (X.Y.); juha.hyyppa@fgi.fi (J.H.)

[4] School of Science and Technology, Aalto University, PL 14100, FI-00076 Aalto, Finland;
E-Mail: hannu.hyyppa@aalto.fi

[5] Helsinki Metropolia University of Applied Sciences, PL 4000, FI-00079 Helsinki, Finland;
E-Mail: hannu.hyyppa@metropolia.fi

[6] Department of Geography and Geology, University of Turku, FI-20014 Turku, Finland;
E-Mail: petteri.alho@utu.fi

[7] HAMK University of Applied Sciences, Saarelantie 1, FI-16970 Evo, Finland;
E-Mail: risto.viitala@hamk.fi

* Author to whom correspondence should be addressed; E-Mail: ville.kankare@helsinki.fi;
Tel.: +358-409-669-109; Fax: +358-9-191-58100.

Abstract: Airborne scanning LiDAR is a promising technique for efficient and accurate biomass mapping due to its capacity for direct measurement of the three-dimensional structure of vegetation. A combination of individual tree detection (ITD) and an area-based approach (ABA) introduced in Vastaranta *et al*. [1] to map forest aboveground biomass (AGB) and stem volume (VOL) was investigated. The main objective of this study was to test the usability and accuracy of LiDAR in biomass mapping. The nearest neighbour method was used in the ABA imputations and the accuracy of the biomass estimation was evaluated in the Finland, where single tree-level biomass models are available. The relative root-mean-squared errors (RMSEs) in plot-level AGB and VOL imputation were 24.9%

and 26.4% when field measurements were used in training the ABA. When ITD measurements were used in training, the respective accuracies ranged between 28.5%–34.9% and 29.2%–34.0%. Overall, the results show that accurate plot-level AGB estimates can be achieved with the ABA. The reduction of bias in ABA estimates in AGB and VOL was encouraging when visually corrected ITD (ITD$_{visual}$) was used in training. We conclude that it is not feasible to use ITD$_{visual}$ in wall-to-wall forest biomass inventory, but it could provide a cost-efficient application for acquiring training data for ABA in forest biomass mapping.

1. Introduction

Remote sensing (RS) methods, such as optical and microwave satellite imaging, digital aerial photography and light detection and ranging (LiDAR), are highly useful in various forest-monitoring tasks. Until recently, knowledge of forest biomass and its changes has been based on ground measurements and coarse or medium resolution satellite images. Therefore, the accuracy of biomass estimations, especially at the local level (e.g., forest stands or sample plots), is poor. Stand biomass is highly correlated with tree heights that can be determined accurately by means of LiDAR, e.g., [2]. It is expected that LiDAR applications will enhance the accuracy of forest biomass estimates at all levels from single-tree to nationwide inventory applications.

LiDAR measurements can be divided into profiling and scanning approaches. Airborne scanning LiDAR (small-footprint) has two main methods for deriving forest information: an area-based approach (ABA, [3]) and individual tree detection (ITD, [4]). LiDAR is a promising technique for efficient and accurate biomass detection, due to its capacity for direct measurement of vegetation structure and forest characteristics at different levels (from tree to stand level), e.g., [5,6]. With LiDAR's ability to directly measure forest structure, including canopy height (CH) and crown dimensions, it is increasingly being used for forest inventories at different levels. Previous studies have shown that LiDAR data can be used to estimate a variety of forest inventory attributes, including tree, plot, and stand level estimates for tree height [4,7–9], stem volume (VOL) [3,10,11], basal area (BA) [12–14] and tree species [15–19]. In relation to biomass changes, LiDAR is also a promising method for monitoring forest hazards and defoliation, due to its ability to derive vegetation structure properties [20–24].

Inventory of stand aboveground biomass (AGB) can be based on LiDAR data at single time points and multitemporal LiDAR can be used in monitoring biomass changes. Lefsky *et al.* [12] showed that LiDAR-derived features such as quadratic mean CH could explain 80% of the variance in AGB. The structure of the forest canopy and leaf area index (LAI) affect the penetration of laser pulses in tree crowns [25]. The LAI value represents the area of leaf surface per unit area of ground surface. Changes in AGB have also been estimated, using changes in LAI. The ground truth for LAI can be determined, using special measuring devices, or by estimation from the LiDAR data [21,25–27]. Popescu *et al.* [28] studied the accuracy of individual tree crown diameter at the plot-level and its effect on AGB estimation from LiDAR data. The study area was covered by coniferous, deciduous and mixed stands of varying age classes. The effect on biomass estimation was notable; the estimate of crown diameter

alone explained 78% of the variance in biomass and the R^2 values and root-mean-squared errors (RMSEs) improved by up to 0.24 and 7 t/ha, respectively.

Popescu *et al*. [29] combined small-footprint LiDAR and multispectral airborne data to estimate the plot-level VOL and AGB in deciduous and pine (*Pinus* L.) forests, using ITD in which the average VOL ranged between 123 and 163 m³/ha. The maximum R^2 values for AGB were 0.32 for deciduous trees and 0.82 for pines. The respective RMSEs were 44 t/ha and 29 t/ha. Bortolot and Wynne [30] also used ITD for AGB estimation in young forests (ages between 11 and 16 yr) in which the correlation (r) varied from 0.59 to 0.82 and RMSEs from 13.6 to 140.4 t/ha. Van Aardt *et al*. [31] estimated the VOL and AGB with LiDAR point height metrics as predictors in a per-segment estimation in deciduous, coniferous and mixed forests. The adjusted R^2 and RMSE values for deciduous AGBs were 0.58 and 37.41 t/ha. Næsset [32] used regression methods to estimate AGB for 143 sample plots in young and mature coniferous forests. The sample plot data were divided into three strata (I: young forest, II: mature forest with poor site quality and III: mature forest with favourable site quality). The regression models explained 92% of the variability in the AGB for all the forest types. Jochem *et al*. [33] used a semiempirical model that was originally developed for VOL estimation to estimate the AGB in Norway spruce (*Picea abies* L.) dominated alpine forests. The model was extended with tree canopy transparency parameters (CTPs) extracted from LiDAR. The model was calibrated, using 196 selected sample plots. The R^2 values for the fitted AGB models were 0.70 with no CTP and varied from 0.64 to 0.71 with different CTPs. The standard deviations (stds) varied from 87.4 t/ha (35.8%) to 101.9 t/ha (41.7%). Latifi *et al*. [34] tested the ABA in southwestern Germany in VOL and AGB mapping. They found that the random forest method was superior to other nearest neighbour (NN) methods and achieved relative errors of 23.3%–31.4% in plot-level VOL and 22.4%–33.2% in AGB prediction, depending on the feature sets and feature selection used.

Breidenbach *et al*. [35] introduced a method, semi-ITC (Individual Tree Crown), that combines a modified ITD approach with ABA. The study introduced a practical solution for solving the problems with tree detection and the estimation bias of ITD. The semi-ITC method predicts volume for the entire segment, which can include none, one or more trees. The use of ABA in forest parameter estimation requires large amounts of training data, which are currently collected mainly with field measurements. Since field measurements are expensive and slow to make, new applications for acquiring training data are needed. Vastaranta *et al*. [1] introduced a method that combined ITD and ABA for forest parameter retrieval. They also used visual interpretation to diminish tree detection problems in ITD. Breidenbach *et al*. [36] used a new data collection strategy employing high-density scanning LiDAR from overlapping flight strips to improve the accuracy of *k* nearest neighbour (kNN) estimates in the areas with low-density scanning LiDAR. The semi-ITC method was used with high-density LiDAR to make forest parameter predictions and the accuracy was tested, using external data with AGB as a response variable. The relative RMSEs varied from 26.7% to 41.56%, depending on the selection of reference plots used.

The objective of this study was to use a combination of two airborne scanning LiDAR-based applications, ITD [4] and the ABA [3,14], in a manner similar to that used in [1], to map AGB and VOL estimates in the same study area [1]. The ABA was trained separately with field and ITD measurements. The main goal of the study was to find the usability and accuracy of the previously introduced method in AGB and VOL mapping. The accuracy of the biomass estimation was evaluated

in Finland, where single-tree-level biomass models are available. The models used were evaluated with values of tree biomass measured in the laboratory. The study focuses on developing a new cost-efficient biomass mapping that would provide spatially explicit estimates.

2. Materials and Methods

2.1. Study Area and Field Data

The study area is located at Evo, Finland (61.19°N, 25.11°E). The area belongs to the southern Boreal Forest Zone and is comprised of mainly managed boreal forests, but Evo is also a popular recreation area. Therefore a broad mixture of forest stands, varying from natural to intensively managed, is available. Scots pine (*Pinus sylvestris* L.) and Norway spruce are the dominant tree species in the study area, contributing approximately 78.2% of the total VOL. See Vastaranta *et al.* [1] for more detailed description of the study area.

The field data comprised 509 circular plots of radius 10 m within all trees with a diameter-at-breast-height (DBH) larger than 7 cm were measured. Sampling the field plots was based on pre-stratification of existing stand inventory data to distribute plots over various site types, tree species and stand development classes. Pre-stratification was done separately in 2007 and 2009 [37]. The field reference data in 2007 were sampled, using stratified sampling [37]. The stratification of the study area was based on previous stand-level inventory data. An initial grid of points with 100 m × 100 m spacing was created, each point representing the centre of a plot, for which the stand variables used in the stratification were allocated. The input variables used in the stratification of the initial plots were volumes of pine, spruce, larch (*Larix decidua* Mill.) and deciduous species and characteristics of the site type (mineral/peatland and site fertility class). All the input variables were standardized to a mean of 0 and std of 1.

The latter set of field plots measured in 2009 were sampled on the basis of the airborne scanning LiDAR and aerial image data acquired in 2007 using an initial grid of first phase points generated equidistantly in the middle of the grid of 2007 [37]. The plots were stratified on the basis of a smaller subset of the airborne scanning LiDAR and aerial image features selected by Holopainen *et al.* [38]. The features used in the stratification included three LiDAR height features, two LiDAR texture features and two aerial image features. The procedure applied in the stratification was otherwise similar to that used in 2007.

The initial plots were clustered into 40 strata using the k-means clustering algorithm. Proportional allocation was applied in deriving the field sample, *i.e.*, the number of the field plots allocated to each stratum was based on the number of initial plots in the strata, with the exception of very small strata in which additional plots were added to the proportional share (at least two plots per stratum). Within the strata, the sample plots were allocated systematically, based on their geographical location.

The field plots were located with a GEOXM 2005 global positioning system (GPS) device (Trimble Navigation Ltd., Sunnyvale, CA, USA), and the locations were post-processed with local base station data resulting accuracy of approximately 0.6 m [39]. The field measurements were collected in 2007 and 2009 from the study plot. The SIMO (Simulation and Optimization) forest management planning calculation system [40] was used for adding tree-level growth to the trees measured in 2007.

2.2. Calculation of AGB

The tree-level AGB was calculated with models introduced by Repola [41,42] for Scots pine, Norway spruce and birch (*Betula pendula* Roth and *Betula pubenscens* Ehrh.) in Finland. The models were developed for total biomass of the tree and for following individual tree components: stemwood and bark, living and dead branches, foliage or needles, stump and roots. The model predictors were tree species, DBH and height. The AGBs for trees were predicted, using total AGB models (1–3) and field-measured DBH and height. The total AGB models are listed below.

Scots pine [42]:

$$\ln(y_{ki}) = b_0 + b_1 \frac{d_{Ski}}{(d_{Ski} + 12)} + b_2 \frac{h_{ki}}{(h_{ki} + 20)} + u_k + e_{ki} \tag{1}$$

Norway spruce [42]:

$$\ln(y_{ki}) = b_0 + b_1 \frac{d_{Ski}}{(d_{Ski} + 20)} + b_2 \ln(h_{ki}) + u_k + e_{ki} \tag{2}$$

Birch [41]:

$$\ln(y_{ki}) = b_0 + b_1 \frac{d_{Ski}}{(d_{Ski} + 12)} + b_2 \frac{h_{ki}}{(h_{ki} + 22)} + u_{5k} + e_{5ki} \tag{3}$$

where y_{ki} is the total biomass for tree i in stand k (kg), d_{Ski} is $2 + 1.25 d_{ki}$ (d_{ki} = the tree DBH for tree i in stand k, cm) and h_{ki} is the tree height for tree i in stand k (m). Factor u_k is the variance of random stand k parameter and e_{ki} is the residual error for tree i in stand k. The AGB model parameters are presented in Table 1.

Table 1. Parameters for aboveground biomass for models 1–3 by Repola [41,42].
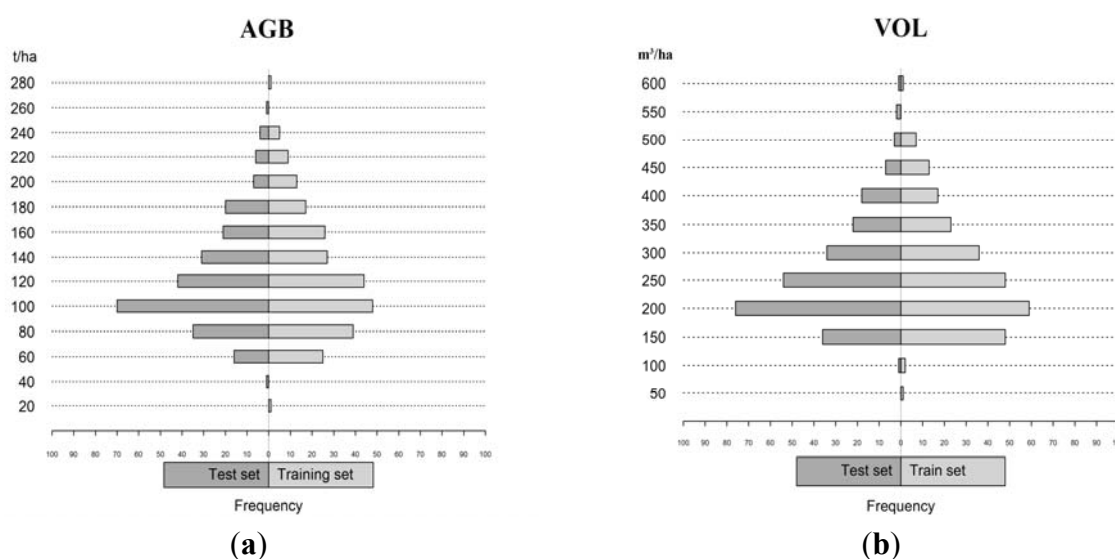
Variable	Tree Species		
	Scots Pine	**Norway Spruce**	**Birch**
b0	−3.198	−1.808	−3.654
b1	9.547	9.482	10.582
b2	3.241	0.469	3.018
var(uk)	0.009	0.006	0.00068
var(eki)	0.01	0.013	0.00727

Table 2. Plot-level statistics for field measurements. 254 circular, fixed-radius (r = 10 m). The plots were used as the test set and 255 plots as the training set for ABA.

Data set	n	Variable	Min	Max	Mean	Std
Test	254	VOL, m³/ha	43.6	533.3	183.4	89.5
Test	254	Dg, cm	11.1	62.2	24.0	8.4
Test	254	Hg, m	7.9	31.7	19.1	4.7
Test	254	AGB, t/ha	19.5	239.3	92.6	41.7
Training	255	VOL, m³/ha	46.8	586.2	189.2	99.3
Training	255	Dg, cm	10.0	49.8	23.7	7.3
Training	255	Hg, m	8.2	30.9	19.1	4.6
Training	255	AGB, t/ha	22.4	270.8	96.2	47.5

Plot-level AGB estimates were obtained by summing the tree data. The VOLs were calculated with standard Finnish models [43]. The field data were divided randomly for the training (255 plots) and test datasets (254 plots). The statistics of the forest variables of AGB, weighted mean diameter (Dg), weighted mean height (Hg) and VOL, in the test and training datasets are presented in Table 2. The frequency distributions of the AGBs and VOLs within the test and training plots are shown in Figure 1.

Figure 1. (**a**) AGB and (**b**) VOL frequency distributions for test and training sets with class interval of 20 t/ha (AGB) and 50 m^3/ha (VOL).

(**a**)　　　　　　　　　　　　　　(**b**)

2.3. Tree Analyses for Evaluating the Biomass Models Used

The data for evaluating the existing biomass models developed by Repola [41,42] were collected in summer 2010 from the study area. A total of 38 sample trees were measured: 19 Scots pines and 19 Norway spruce (Table 3). The field measurements and biomass calculations are described in further detail in Kankare *et al*. [44] and were conducted on the same principle as that described in Repola [41]. A short summary is presented below.

Table 3. Sample tree statistics for evaluation of Repola [42] models.

	Scots Pine			Norway Spruce		
	Mean	**Std**	**Range**	**Mean**	**Std**	**Range**
DBH, cm	19.6	4.0	15.6	20.9	7.0	26.6
Height, m	18.8	2.4	9	19.2	6.4	20
Age, year	49	5	16	68.6	25.8	88
Crown ratio	0.57	0.08	0.30	0.25	0.11	0.35
Liv. branch, kg	46	25	91	106	69	279
Dead branch, kg	4.2	3.3	13	4.0	4.5	16
Stem mass, kg	270	122	446	338	281	1034
Total mass, kg	321	146	541	448	351	1332

The field measurements were collected in a destructive manner. All the sample trees were felled, thinned and the trunks were cut into logs. The following information was recorded: tree height, heights of the living and dead crowns, weight of the branches (including cones and needles) and logs, diameters along the trunk and moisture content of the bark, stem wood and branches. Samples of bark, stem wood and branches were dried in an oven at 70 °C for 2–3 days. The biomass components were modelled, using ratio estimation methods. The biomass is referred to as dry weight in Table 3.

2.4. LiDAR Data

The LiDAR data was acquired in midsummer 2009, using a Leica ALS50-II SN058 system (Leica Geosystems AG, Heerbrugg, Switzerland). The data was acquired with following parameters: (1) the flying altitude was 400 m at a speed of 80 knots, (2) an open angle of 30°, (3) beam divergence of 0.15 mrad and (4) a pulse rate of 150 kHz. The pulse density within the plots was 10 hits per m^2. A digital terrain model (DTM) was computed by the data provider.
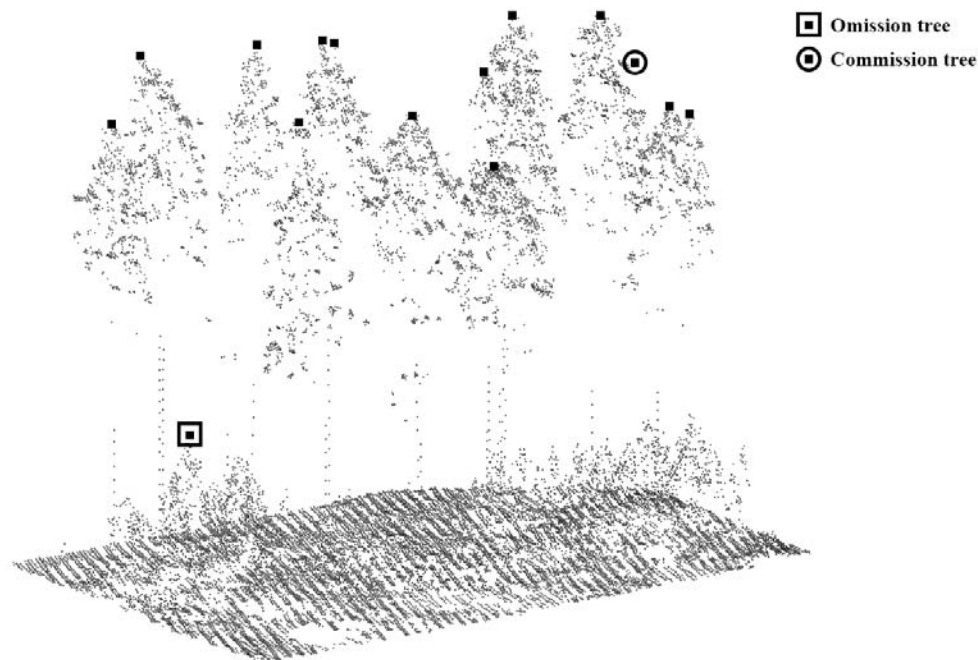
2.5. Individual Tree Detection

2.5.1. Automatic Tree Detection and Visual Interpretation

A summary of individual tree detection methods used follows and the methods used are the same as introduced elsewhere in detail [1]. A raster canopy height model (CHM) was created from normalized data (DTM subtracted from the height of the vegetation points). Single tree segmentations were performed in the CHM images, using a minimum curvature-based region detector (see [45,46]). During the segmentation processes, the tree crown shapes and locations of individual trees were determined. The CHM was smoothed with a Gaussian filter to remove small variations on the crown surface. The degree of smoothness was determined by the value of the std (Gaussian scale) and kernel size of the filter. An std of 0.7 and bandwidth of five pixels (pixel size 0.5 m) were used based on trial and error. Then the minimum curvatures were calculated (see [46]). The minimum curvature is one of the principal curvatures and, for a surface such as a CHM, a higher value of minimum curvature describes the treetop. The smoothed CHM image was linearly scaled, based on the minimum curvature computed, resulting in a smoothed, yet contrast-stretched image. Finally, the local maxima were searched in a sliding neighbourhood of 5×5 pixels. They were considered as treetops and used as markers in the following marker-controlled watershed transformation [47] for tree crown delineations. Each segment was considered to represent a singletree crown and the highest laser point height within the segment was used as an estimate for tree height.

The results of above described automatic tree detection (ITD$_{auto}$) were visually inspected in TerraScan software (TerraSolid Ltd, Helsinki, Finland). The omission and commission errors (Figure 2) were searched and reduced in a three-dimensional (3D) environment with visual interpretation. This procedure combining ITD$_{auto}$ and visual interpretation is further referred to as visual ITD (ITD$_{visual}$).

Figure 2. ITD$_{auto}$-detected trees plotted with black solid squares. Omission tree marked (shallow black square) from understorey. Commission error from plot with a single tree (shallow black circle) is easily reduced in visual interpretation.

2.5.2. Retrieval of AGB and VOL from Individual Tree Detection

The plot-level AGB and VOL, based on the ITD results, were calculated for ITD$_{auto}$ and ITD$_{visual}$, with no field measurements. A DBH was predicted for the trees, using the laser-based tree height and model developed by Kalliovirta and Tokola [48] for boreal forest stands in southern Finland. The AGB and VOL were calculated, using the laser-based tree height and predicted DBH, with the same models [41–43] used in calculation of the AGB and VOL of the field measurements (see Section 2.2). In AGB and VOL modelling, the main tree species of the plots were assumed to be known.

2.6. Area-Based Approach

The ABA used statistical LiDAR features as predictors that were calculated from the CH or vertical distribution of the laser returns [14,49] for estimating the plot-level AGB and VOL. The AGB and VOL were imputed for the test plots, using ABA with three differently acquired training sets, field measurements, ITD$_{auto}$ and ITD$_{visual}$. Methods are referred as ABA$_{field}$, ABA$_{ITDauto}$ and ABA$_{ITDvisual}$. The following features were derived individually per plot from the normalized laser data, minimum height (Hmin), maximum height (Hmax), mean height (Hmean) calculated as the arithmetic mean of the laser heights, STD of laser heights (Hstd), coefficient of variation (CV), penetration, percentiles calculated from 10% to 100% of the CH distribution at 10% intervals (h10–h100) and canopy cover percentiles expressed as proportions of the first returns below a given percentage (10–90%) of the total height (p10–p90).

The nearest neighbour (NN) imputations were applied in ABA. The selected NN imputation method was the k-most similar neighbour (k-MSN) [50]. In k-MSN, the similarity is based on canonical

correlations and the Mahalanobis distance [50]. The R yaImpute library [51] was applied in the NN imputations. The number of NNs was set at 5. LiDAR features used as predictors were selected, using the least absolute shrinkage and selection operator (LASSO) method introduced by Tibshirani [52]. LASSO regression was applied, using the glmnet-package [53] in R software for statistical computing [54]. The method uses penalizing of the absolute size of the regression coefficients and this can produce a situation in which the size of the coefficients can shrink to zero, which enables feature selection. The method recommended for selecting optimal penalty value (lambda) is cross-validation, 10-fold cross-validation was used in this study. The features were selected separately for ABA$_{field}$, ABA$_{ITDauto}$ and ABA$_{ITDvisual}$.

2.7. Statistical Analyses

The accuracies of the tree-level biomass models and imputed plot-level AGBs and VOLs were evaluated by calculating the bias and RMSE:

$$BIAS = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{n} \tag{4}$$

$$BIAS\% = 100 * \frac{BIAS}{\bar{y}} \tag{5}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{6}$$

$$RMSE\% = 100 * \frac{RMSE}{\bar{y}} \tag{7}$$

where *n* is the number of observations, y_i the value estimated from the field data for observation *i*, $\hat{y}_i$ the predicted value for observation *i* and $\bar{y}$ the mean of the variable in question.

3. Results

3.1. Validation of the Biomass Estimation Models Used

The accuracies of the biomass models used were evaluated at the tree level, using 38 trees measured in the laboratory. The estimates of AGB were calculated, using field-measured DBH, h and species. These estimates were compared with the biomasses determined in the laboratory (Figure 3). The residuals showed a systematic shift in estimates larger than 250 kg. In this case, the biases for the Scots pines were 9.1 kg (5.8%) and for the Norway spruce 35.2 kg (15.0%). The RMSEs, combining both bias and deviation, were 15.0 kg (9.5%) and 54.6 kg (23.2%), respectively. The errors in tree-level AGB estimations are shown in Figure 4.

3.2. Accuracy of the ABA

The accuracy of the ABA in plot-level AGB and VOL imputation was assessed in 254 test plots. Training data for the ABA were measured in the field and with ITD from 255 plots. Tree-level AGBs were calculated with the models reported by Repola [41,42].

Figure 3. Total biomass of tree (kg) predicted with biomass models [42] *vs.* laboratory-derived biomasses, red = Scots pine and green = Norway spruce.
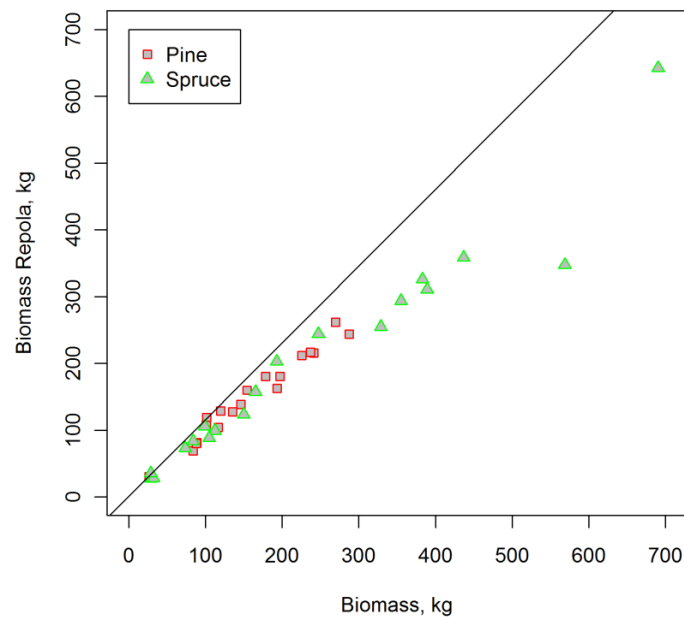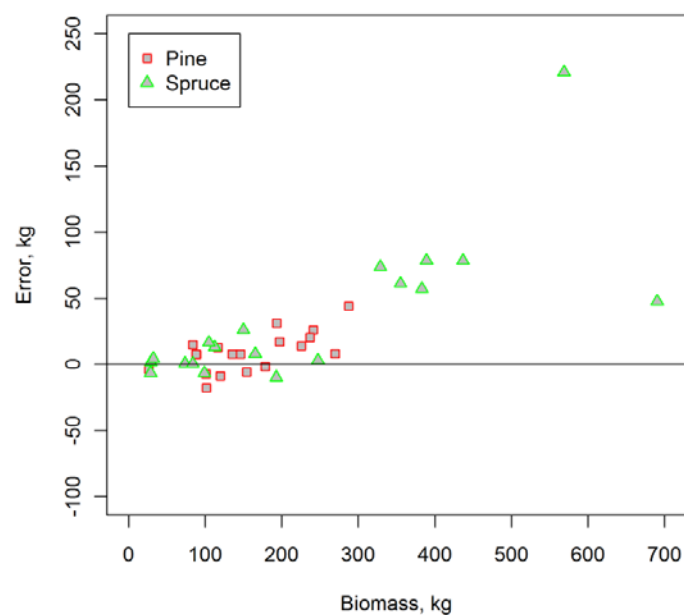

Figure 4. Residual plot for AGB at tree level. AGB prediction error in kg is presented in y-axis for Scots pine (red) and for Norway spruce (green).

The final statistical LiDAR features used in the imputations were selected, using LASSO (Table 4). The difference for selection of the best features was relatively small between the ABA methods, depending on the training data, and the number of predictors varied from 8 to 11.

The accuracies of the various ABAs are presented in Table 5 and residual plots in Figure 5. The ABA$_{ITDvisual}$ AGB imputation accuracy was similar to that of ABA$_{field}$. The RMSEs were 24.9%, 34.9% and 28.5% for ABA$_{field}$, ABA$_{ITDauto}$ and ABA$_{ITDvisual}$, respectively. The ABA$_{ITDvisual}$ resulted in even smaller bias for the AGB than did ABA$_{field}$. The adjusted R^2 values (Table 6, Figure 6) for the ABA methods in the AGB imputations varied from 0.71 to 0.72.

VOL was imputed most accurately with ABA$_{field}$ (Table 5) and an RMSE of 26.4% was achieved. The ABA$_{ITDvisual}$ imputation accuracy was better than that of ABA$_{ITDauto}$, especially the bias, which was reduced from −12.5% to −1.9%.

Table 4. Statistical LiDAR features used in k-MSN imputations.

Feature	ABA$_{field}$	ABA$_{ITDauto}$	ABA$_{ITDvisual}$
Hmax	●	●	●
Hstd			●
CV			●
h10	●	●	
h20	●		
h30	●		●
h40		●	
h50	●		●
h60	●	●	
h90			●
p10	●	●	●
p30		●	
p40	●		●
p50		●	
p70	●		●
p80		●	●
p90	●		●

Table 5. AGB and volume imputation accuracies, RMSE and bias, for the test plots (n = 254).

Method	Variable	Bias	Bias (%)	RMSE	RMSE (%)
ABA$_{field}$	AGB, t/ha	−3.2	−3.5	23.0	24.9
ABA$_{ITDauto}$	AGB, t/ha	−12.9	−13.9	32.3	34.9
ABA$_{ITDvisual}$	AGB, t/ha	−2.8	−3.1	26.4	28.5
ABA$_{field}$	VOL, m^3/ha	−4.9	−2.7	48.4	26.4
ABA$_{ITDauto}$	VOL, m^3/ha	−22.9	−12.5	62.4	34.0
ABA$_{ITDvisual}$	VOL, m^3/ha	−3.4	−1.9	53.6	29.2

**Figure 5.** Residual plots for AGB (t/ha) with the ABA.
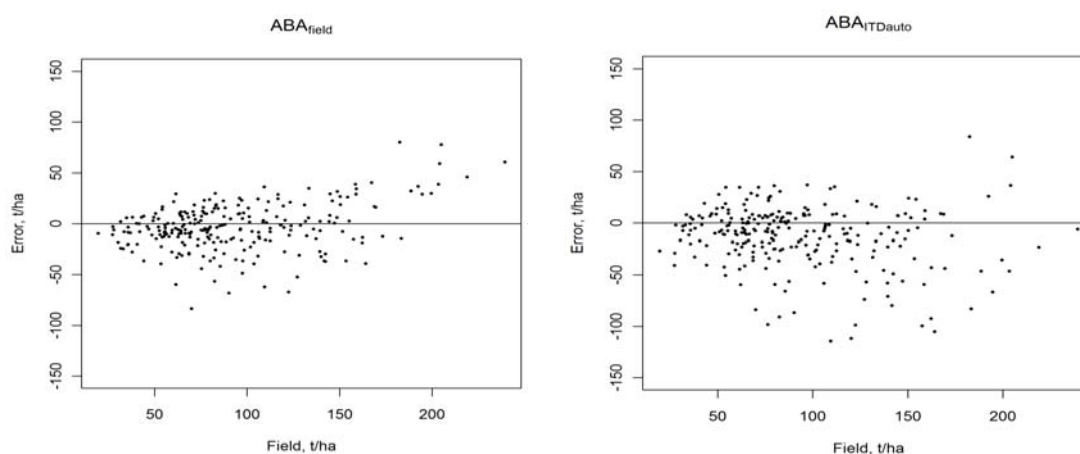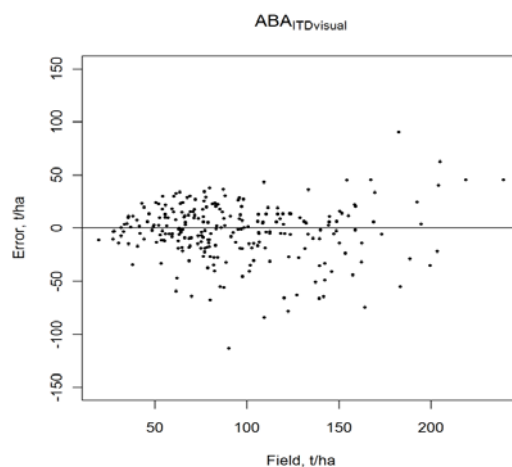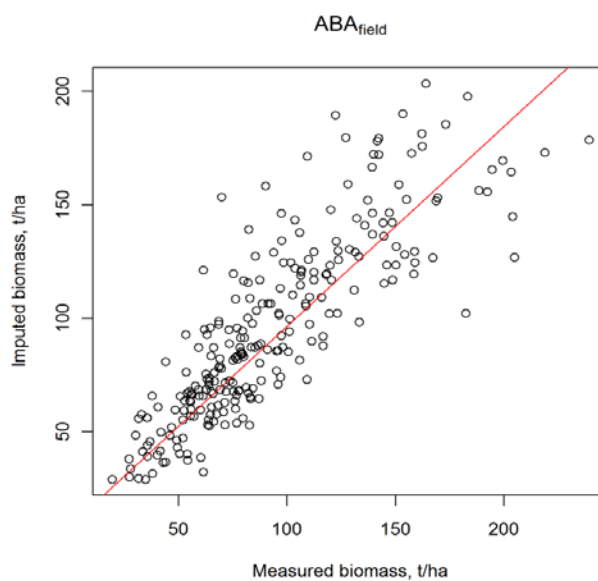
Figure 5. *Cont.*

Table 6. Adjusted R^2 for ABA methods in AGB imputations compared with test set.

Method	Tree Species	n	Adjusted R^2
ABA_{field}	Scots pine	144	0.73
ABA_{field}	Norway spruce	68	0.68
ABA_{field}	Deciduous trees	42	0.72
ABA_{field}	All	254	0.71
$ABA_{ITDauto}$	Scots pine	144	0.75
$ABA_{ITDauto}$	Norway spruce	68	0.68
$ABA_{ITDauto}$	Deciduous trees	42	0.71
$ABA_{ITDauto}$	All	254	0.71
$ABA_{ITDvisual}$	Scots pine	144	0.76
$ABA_{ITDvisual}$	Norway spruce	68	0.69
$ABA_{ITDvisual}$	Deciduous trees	42	0.71
$ABA_{ITDvisual}$	All	254	0.72

Figure 6. Field-measured biomass *vs.* ABA_{field}-imputed biomass, t/ha.

4. Discussion

The estimation of forest AGB and VOL with airborne scanning LiDAR was tested in the present study. The results showed that LiDAR is capable of retrieving AGB accurately at the plot level. Validated current tree-level biomass models proved to be useful in AGB calculations, providing accuracies similar to those of general tree VOL models. Systematic shifts were found in validation of the current tree-level biomass models for biomasses larger than 250 kg, which meant that the large biomasses were underestimated. The bias in larger trees can be the reason for the increasing deviance that is seen in Figure 6. Therefore the usability of these models should be studied in forests with varying development levels.

The AGB imputation accuracies (RMSEs) varied in this study from 24.9% (23 t/ha) to 34.9% (32.3 t/ha) at the plot level. The results were in line with previous studies, e.g., [29,30,55,56], in which the RMSEs varied from 14% to 42%. Latifi *et al.* [34] achieved plot-level accuracies of 22.2%−45.5%, using LiDAR data, and the best results were achieved with a random forest approach in comparison to other NN methodologies. The AGB imputation accuracy [34] for the k-MSN method varied from 43.3% to 45.8%, depending on the variable set used. The results for the AGB imputation with the ABA$_{ITD}$-methods were, in general, similar to those in Latifi *et al.* [34], but to make adequate comparison of the NN methods used, all external factors (e.g., site-specific or the selection of model features) should be removed.

The best features for each ABA method were selected, using LASSO regression. The number of features selected was dependent on the penalty value (lambda), which was searched using cross-validation. The values selected were quite small and therefore the number of features in the models was large. This could cause overfitting, especially if highly correlated feature groups are present. With higher lambda value, fewer features were selected for the model, but this resulted in rapid increase in mean squared error (MSE) in cross-validation. Detailed comparison of the various feature selection methods when LiDAR features are used would be an interesting research topic since the correlated features here were selected for the optimal feature set. The impact of variable selection was also investigated by calculating the estimates for ABA$_{ITDauto}$ and ABA$_{ITDvisual}$ using selected features from ABA$_{field}$. Only small increase in bias and RMSE was found, e.g., the relative RMSE increased 1.6% and 0.5% for ABA$_{ITDauto}$ and ABA$_{ITDvisual}$, respectively.

The AGBs imputed with the various ABA methods explained 71%–72% of the variation at the plot level. The R^2 values for the Scots pine, Norway spruce and deciduous plots varied between 0.73 and 0.76, 0.68 and 0.69 and 0.71 and 0.72. In comparison to previous studies, the total R^2 values were similar or somewhat better, e.g., [28,30].

The accuracy of tree VOL imputation at different levels (tree, plot or stand) is one of the most studied of forestry subjects, e.g., [4,13,14,34]. The accuracy of plot-level VOL imputation in the present study varied from 26.4% (48.4 m³/ha) to 34.0% (62.4 m³/ha), depending on the ABA training data used. Latifi *et al.*'s [34] imputation accuracies for VOL varied between 23.3% and 31.4%, depending on the feature set and selection method used. Popescu *et al.* [29] achieved VOL RMSEs of 52.84 m³/ha and 47.9 m³/ha for deciduous and pine plots, respectively. The VOL accuracy obtained in the present study with the ABA$_{ITD}$ methods was similar to that in the above-mentioned studies. The best result was from ABA$_{field}$, but the results from the ABA$_{ITD}$ methods were not far behind.

In large-scale forest biomass inventories, acquisition of extensive field references may be a major source of expense. In the present study, the training datasets for ABA were measured from LiDAR data and the imputation accuracies were compared with the accuracies imputed, using traditional field data. This type of approach could provide savings, especially in areas with difficult terrain or sparse road networks, and still maintain the spatial accuracy of the estimates. However, it is crucial to obtain unbiased estimates, which often presents a bottleneck if ITD is used in acquiring training data for the ABA. In the present study, $ABA_{ITDvisual}$ resulted in even smaller bias for AGB and VOL than ABA_{field}. The relative bias was 0.8%–1.2% smaller for $ABA_{ITDvisual}$. This was most likely caused by the growth simulation used with the sample plots measured in 2007. It should be noted that the knowledge of tree species has a large impact on ABA estimation results and it was given as an auxiliary input data for the ITD training datasets in this study. The bias estimates might therefore be too low since tree species were not estimated using the LiDAR data. However, multiple promising studies have been published on this subject [16–19] and tree species classification should be integrated to the method in operational forestry.

Tree detection has been a major error source in ITD. Under Nordic forest conditions, tree detection accuracies with ITD_{auto} methods have varied from 20% to 90% [57]. Other sources of bias in ITD methods include prediction of DBH and tree species classification [58]. In this study, the composition of tree species was relatively simple. In more diverse forests, the biomass estimation and especially biomass model development would be more challenging. We used visual interpretation to reduce bias caused by commission and omission errors of tree detection. The bias for AGB was −13.9% with $ABA_{ITDauto}$. Visual interpretation reduced the bias to −3.1%. The bias in ITD could also be reduced by using plot-level biomass calibration data measured in the field or more sophisticated DBH prediction models [24,58,59].

5. Conclusions

The airborne scanning LiDAR-based forest mapping method [1], which utilizes ITD to measure training data for ABA, was tested in plot-level AGB and VOL imputations in the same study area used in [1]. The main goal was to find the usability and accuracy of the introduced method in AGB mapping in Finland where biomass models are available. The results show that spatially accurate AGB estimates can be achieved, using the ABA. Biomass inventory can be even more accurate and effective if ITD measurements are used in training. The use of ITD as training data for the ABA provides a new application for forest biomass inventories in areas that are difficult to reach. Accurate local biomass models were used in this study, but in most of the areas these types of models are not available. Therefore, the use of tree-level LiDAR features should be studied in AGB estimation and model creation.

Acknowledgments

Conflict of Interest

The authors declare no conflict of interest.

References

1. Vastaranta, M.; Kankare, V.; Holopainen, M.; Yu, X.; Hyyppä, J.; Hyyppä, H. Combination of individual tree detection and area-based methodology in imputation of forest variables using airborne laser data. *ISPRS J. Photogram.* **2012**, *67*, 73–79.
2. Kellndorfer, J.M.; Walker, W.S.; LaPoint, E.; Kirsch, K.; Bishop, J.; Fiske, G. Statistical fusion of LiDAR, InSAR, and optical remote sensing data for forest stand height characterization: A regional-scale method based on LVIS, SRTM, Landsat ETM+, and ancillary data sets. *J. Geophys. Res.* **2010**, *115*, doi: 10.1029/2009JG000997.
3. Næsset, E. Determination of mean tree height of forest stands using airborne laser scanner data. *ISPRS J. Photogram.* **1997**, *52*, 49–56.
4. Hyyppä, J.; Inkinen, M. Detecting and estimating attributes for single trees using laser scanner. *Photogramm. J. Fin.* **1999**, *16*, 27–42.
5. Koch, B. Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. *ISPRS J. Photogram.* **2010**, *65*, 581–590.
6. Holopainen, M.; Hyyppä, J.; Vaario, L.-M.; Yrjälä, K. Implications of Technological Development to Forestry. In *Forest and Society—Responding to Global Drivers of Change*; Mery, G., Katila, P., Galloway, G., Alfaro, R.I., Kanninen, M., Lobovikov, M., Varjo, J., Eds.; IUFRO-World Series: Vienna, Austria, 2010; pp. 157–182.
7. Falkowski, M.; Smith, A.; Hudak, A.; Gessler, P.; Vierling, L.; Crookston, N. Automated estimation of individual conifer tree height and crown diameter via two-dimensional spatial wavelet analysis of LiDAR data. *Can. J. Remote Sens.* **2006**, *32*, 153–161.
8. Magnussen, S.; Eggermont, P.; LaRiccia, V.N. Recovering tree heights from airborne laser scanner data. *For. Sci.* **1999**, *45*, 407–422.
9. Maltamo, M.; Mustonen, K.; Hyyppä, J.; Pitkänen, J.; Yu, X. The accuracy of estimating individual tree variables with airborne laser scanning in a boreal nature reserve. *Can. J. For. Res.* **2004**, *34*, 1791–1801.
10. Hyyppä, J.; Kelle, O.; Lehikoinen, M.; Inkinen, M. A segmentation-based method to retrieve stem volume estimates from 3-dimensional tree height models produced by laser scanner. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 969–975.
11. Wallerman, J.; Holmgren, J. Remote sensing based imputation of sample plot data for forest management planning utilizing single-tree forecast models. *Remote Sens. Environ.* **2007**, *110*, 501–508.
12. Lefsky, M.A.; Harding, D.; Cohen, W.B.; Parker, G.; Shugart, H.H. Surface lidar remote sensing of basal area and biomass in deciduous forests of Eastern Maryland, USA. *Remote Sens. Environ.* **1999**, *67*, 83–98.
13. Means, J.E.; Acker, S.A.; Fitt, B.J.; Renslow, M.; Emerson, L.; Hendrix, C.J. Predicting forest stand characteristics with airborne scanning lidar. *Photogramm. Eng. Remote Sensing* **2000**, *66*, 1367–1371.

14. Næsset, E. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sen. Environ.* **2002**, *80*, 88–99.

15. Van Aardt, J.A.N.; Wynne, R.H.; Scrivani, J.A. Lidar-based mapping of forest volume and biomass by taxonomic group using structurally homogenous segments. *Photogramm. Eng. Remote Sensing* **2008**, *74*, 1033–1044.

16. Brandtberg, T. Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar. *ISPRS J. Photogram.* **2009**, *61*, 325–340.

17. Holmgren, J.; Persson, Å. Identifying species of individual trees using airborne laser scanner. *Remote Sens. Environ.* **2004**, *90*, 415–550.

18. Korpela, I.; Ørka, H.O.; Maltamo, M.; Tokola, T.; Hyyppä, J. Tree species classification using airborne LiDAR—Effects of stand and tree parameters, downsizing of training set, intensity normalization, and sensor type. *Silva Fenn.* **2010**, *44*, 319–339.

19. Vauhkonen, J.; Korpela, I.; Maltamo, M.; Tokola, T. Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. *Remote Sens. Environ* **2010**, *114*, 1263–1276.

20. Solberg, S.; Næsset, E.; Hanssen, K.H.; Cristiansen, E. Mapping defoliation during a severe insect attack on Scots pine using airborne laser scanning. *Remote Sens. Environ.* **2006**, *102*, 384–376.

21. Solberg, S.; Næsset, E. Monitoring Forest Health by Remote Sensing. In Proceedings of Symposium on Forests in a Changing Environment—Results of 20 years ICP Forests Monitoring, Göttingen, Germany, 25–28 October 2006; Schriftenreihe der Forstlichen Fakultät Göttingen und der Nordwestdeutschen Forstlichen Versuchsanstalt: Göttingen, Germany, 2007; pp. 99–104.

22. Kantola, T.; Vastaranta, M.; Yu, X.; Lyytikäinen-Saarenmaa, P.; Holopainen, M.; Talvitie, M.; Kaasalainen, S.; Solberg, S.; Hyyppä, J. Classification of defoliated trees using tree-level airborne laser scanning data combined with aerial images. *Remote Sens.* **2010**, *2*, 2665–2679.

23. Solberg, S. Mapping gap fraction, LAI and defoliation using various ALS penetration variables. *Int. J. Remote Sens.* **2010**, *32*, 1227–1244.

24. Vastaranta, M.; Holopainen, M.; Yu, X.; Haapanen, R.; Melkas, T.; Hyyppä, J.; Hyyppä, H. Individual tree detection and area-based approach in retrieval of forest inventory characteristics from low-pulse airborne laser scanning data. *Photogram. J. Fin.* **2011**, *22*, 1–13.

25. Solberg, S.; Brunner, A.; Hanssen, K.H.; Lange, H.; Næsset, E.; Rautiainen, M.; Stenberg, P. Mapping LAI in a Norway spruce forest using airborne laser scanning. *Remote Sens. Environ.* **2009**, *113*, 2317–2327.

26. Solberg, S. Mapping gap fraction, LAI and defoliation using various ALS penetration variables. *Int. J. Remote Sens.* **2008**, *31*, 1227–1244.

27. Korhonen, L.; Korpela, I.; Heiskanen, J.; Maltamo, M. Airborne discrete-return LIDAR data in the estimation of vertical canopy cover, angular canopy closure and leaf area index. *Remote Sens. Environ.* **2011**, *115*, 1065–1080.

28. Popescu, S.C.; Wynne, R.H.; Nelson, R.F. Measuring individual tree crown diameter with lidar and assessing its influence on estimation forest volume and biomass. *Can. J. Remote Sens.* **2003**, *29*, 564–577.

29. Popescu, S.C.; Wynne, R.H.; Scrivani, J.A. Fusion of small-footprint lidar and multispectral data to estimate plot-level volume and biomass in deciduous and pine forests in Virginia, USA. *For. Sci.* **2004**, *50*, 551–565.

30. Bortolot, Z.; Wynne, R.H. Estimating forest biomass using small footprint LiDAR data: An individual tree-based approach that incorporates training data. *ISPRS J. Photogramm.* **2005**, *59*, 342–360.

31. Van Aardt, J.A.N.; Wynne, R.H.; Oderwald, R.G. Forest volume and biomass estimation using small-footprint lidar distributional parameters on a per-segment basis. *For. Sci.* **2006**, *52*, 636–649.

32. Næsset, E. Estimation of above- and below-ground biomass in boreal forest ecosystems. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2004**, *36*, 145–148.

33. Jochem, A.; Hollaus, M.; Rutzinger, M.; Höfle, B. Estimation of aboveground biomass in Alpine forests: A semi-empirical approach considering canopy transparency derived from airborne LiDAR data. *Sensors* **2011**, *11*, 278–295.

34. Latifi, H.; Nothdurft, A.; Koch, B. Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: Application of multiple optical/LiDAR-derived predictors. *Forestry* **2010**, *83*, 395–407.

35. Breidenbach, J.; Næsset, E.; Lien, V.; Gobakken, T.; Solberg, S. 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 Sens. Environ.* **2010**, *114*, 911–924.

36. Breidenbach, J.; Næsset, E.; Gobakken, T. Improving k-nearest neighbor predictions in forest inventories by combining high and low density airborne laser scanning data. *Remote Sens. Environ.* **2012**, *117*, 358–365.

37. Tuominen, S. Personal Communication, 2013.

38. Holopainen, M.; Haapanen, R.; Tuominen, S; Viitala, R. Performance of Airborne Laser Scanning- and Aerial Photograph-Based Statistical and Textural Features in Forest Variable Estimation. In Proceedings of SilviLaser 2008: 8th International Conference on LiDAR Applications in Forest Assessment and Inventory, Edinburgh, UK, 17–19 September 2008; pp. 105–112.

39. Vastaranta, M.; Holopainen, M.; Haapanen, R.; Yu, X.; Melkas, T.; Hyyppä, J.; Hyyppä, H. Comparison between an Area-based and Individual Tree Detection Method for Low-pulse Density Als-based Forest Inventory. In Proceedings of Laser Scanning, Paris, France, 1–2 September 2009; pp. 147–151.

40. Rasinmäki, J.; Kalliovirta, J.; Mäkinen, A. An adaptable simulation framework for multiscale forest resource data. *Comput. Electron. Agr.* **2009**, *66*, 76–84.

41. Repola, J. Biomass equations for birch in Finland. *Silva Fenn.* **2008**, *42*, 605–624.

42. Repola, J. Biomass equations for Scots pine and Norway spruce in Finland. *Silva Fenn.* **2009**, *43*, 625–647.

43. Laasasenaho, J. Taper curve and volume functions for pine, spruce and birch. *Commun. Inst. For. Fenn.* **1982**, *108*, 74.

44. Kankare, V.; Holopainen, M.; Vastaranta, M.; Puttonen, E.; Yu, X.; Hyyppä, J.; Vaaja, M.; Hyyppä, H.; Alho, P. Individual tree biomass estimation using terrestrial laser scanning. *ISPRS J. Photogram.* **2013**, *75*, 64–75.

45. Chandra, S.; Sivaswamy, J. An Analyses of Curvature based Ridge and Valley Detection. In Proceedings of International Conference on Acoustics, Speech and Signal Processing, Toulouse, France, 14–19 May 2006; pp. 737–740.

46. Yu, X.; Hyyppä, J.; Vastaranta, M.; Holopainen, M.; Viitala, R. Predicting individual tree attributes from airborne laser point clouds based on random forests technique. *ISPRS J. Photogram.* **2011**, *66*, 28–37.

47. Mathworks, 2010. Image Processing Toolbox. Available online: http://www.mathworks.se/help/images/examples/marker-controlled-watershed-segmentation.html (accessed on 2 June 2012).

48. Kalliovirta, J.; Tokola, T. Functions for estimating stem diameter and tree age using tree height, crown width and existing stand database information. *Silva Fenn.* **2005**, *39*, 227–248.

49. Magnussen, S.; Boudewyn, P. Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. *Can. J. For. Res.* **1998**, *28*, 1016–1031.

50. Moeur, M.; Stage, A.R. Most similar neighbor: An improved sampling inference procedure for natural resource planning. *For. Sci.* **1995**, *41*, 337–359.

51. Crookston, N.; Finley, A. yaImpute: An R Package for k-NN Imputation. *J. Stat. Softw.* **2008**, *23*, 1–16.

52. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. Royal Stat. Soc. B Stat. Meth.* **1996**, *58*, 267–288.

53. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **2010**, *33*, 1–22.

54. R Development Core Team. R: A Language and Environment for Statistical Computing. In *R Foundation for Statistical Computing*; 2011. Available online: http://www.R-project.org/ (accessed on 1 September 2011).

55. Nelson, R.; Krabill, W.; Tonelli, J. Estimating forest biomass and volume using airborne laser data. *Remote Sens. Environ.* **1988**, *24*, 247–267.

56. Lim, K.S.; Treitz, P.M. Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. *Scand. J. For. Res.* **2004**, *19*, 558–570.

57. Kaartinen, H.; Hyyppä, J. *EuroSDR/ISPRS Project, Commission II "Tree Extraction"*; Final Report; EuroSDR Official Publication No. 53; European Spatial Data Research: Dublin, Ireland. 2008.

58. Vastaranta, M.; Holopainen, M.; Yu, X.; Hyyppä, J.; Mäkinen, A.; Rasinmäki, J.; Melkas, T.; Kaartinen, H.; Hyyppä, H. Effects of ALS individual tree detection error sources on forest management planning calculations. *Remote Sens.* **2011**, *3*, 1614–1626.

59. Yu, X.; Hyyppä, J.; Vastaranta, M.; Holopainen, M. Predicting individual tree attributes from airborne laser point clouds based on random forest technique. *ISPRS J. Photogram.* **2011**, *66*, 28.