Remote Sens. **2012**, *4*, 1004-1023; doi:10.3390/rs4041004

OPEN ACCESS

Remote Sensing

ISSN 2072-4292

www.mdpi.com/journal/remotesensing

Article

Comparison of Methods for Estimation of Stem Volume, Stem Number and Basal Area from Airborne Laser Scanning Data in a Hemi-Boreal Forest

Eva Lindberg [1,*] and Markus Hollaus [2]

[1] Department of Forest Resource Management, Swedish University of Agricultural Sciences, SE-90183 Umeå, Sweden

[2] Institute of Photogrammetry and Remote Sensing, Vienna University of Technology (TU Wien), Gußhausstraße 27-29, A-1040 Vienna, Austria; E-Mail: mh@ipf.tuwien.ac.at

* Author to whom correspondence should be addressed; E-Mail: eva.lindberg@slu.se; Tel.: +46-90-786-8536; Fax: +46-90-778-116.

Received: 6 February 2012; in revised form: 3 April 2012 / Accepted: 6 April 2012 / Published: 13 April 2012

Abstract: This study compares methods to estimate stem volume, stem number and basal area from Airborne Laser Scanning (ALS) data for 68 field plots in a hemi-boreal, spruce dominated forest (Lat. 58°N, Long. 13°E). The stem volume was estimated with five different regression models: one model based on height and density metrics from the ALS data derived from the whole field plot, two models based on similar combinations derived from 0.5 m raster cells, and two models based on canopy volumes from the ALS data. The best result was achieved with a model based on height and density metrics derived from 0.5 m raster cells (Root Mean Square Error or RMSE 37.3%) and the worst with a model based on height and density metrics derived from the whole field plot (RMSE 41.9%). The stem number and the basal area were estimated with: (i) area-based regression models using height and density metrics from the ALS data; and (ii) single tree-based information derived from local maxima in a normalized digital surface model (nDSM) mean filtered with different conditions. The estimates from the regression model were more accurate (RMSE 52.7% for stem number and 21.5% for basal area) than those derived from the nDSM (RMSE 63.4%–91.9% and 57.0%–175.5%, respectively). The accuracy of the estimates from the nDSM varied depending on the filter size and the conditions of the applied filter. This suggests that conditional filtering is useful but sensitive to the conditions.

Keywords: forest parameter; LiDAR; regression models; single tree parameter

1. Introduction

During the last decade, airborne laser scanning (ALS) data have been established as a standard data source for high precision topographic data acquisition and have also been used for estimation of forest variables [1]. For forestry applications, the most commonly used method is to derive measures from the ALS data in raster cells approximately the size of a field plot, 100–200 m^2, and use the measures as independent variables in regression models to estimate forest variables such as mean tree height and stem volume [2–4].

The measures derived from the ALS data may be height percentiles and measures of the density of the vegetation as the fraction of ALS reflections from vegetation relative to the total amount of ALS reflections [5]. In that case the regression models are based on an assumption that the stem volume is proportional to one or several height measures (e.g., percentiles of the height above the ground) multiplied by a density measure (e.g., the fraction of ALS data above a threshold height above the ground). Usually a log-log regression model is used. The regression model is often selected with best subset regression that selects the set of independent variables that result in the highest correlation with the dependent variable or with stepwise regression where independent variables are included or excluded in the regression model depending on their significance. The estimation of the regression model parameters is based on reference data from one study area [6]. Another approach for stem volume or biomass estimation is to use a model based on the structure of the forest by calculating the canopy volumes for different height layers and using those measures as independent variables in a linear regression model [7]. As defined in [7], the canopy volume is the entire volume between the canopy and the terrain surface. Furthermore, the different canopy height layers account for height-dependent differences in canopy structure. The forest canopy can either be described by the first echoes directly or by a rasterized digital surface model (DSM) calculated from the first echoes. As the input for the canopy volume estimation, the canopy height normalization with respect to the digital terrain model (DTM) of the first echoes and the DSM, respectively, is required. For the calibration of the linear regression model, reference data is needed, for example, from a forest inventory. Depending on its sampling design (e.g., angle count sampling, fully callipered sample plot area, stand-based), the spatial unit used to extract the ALS-based measures can vary.

If the ALS data are dense enough, individual tree crowns may be identified from the data [8–12]. The identification is usually done by deriving a normalized digital surface model (nDSM) from the ALS data and defining local maxima in the nDSM as treetops. The nDSM is calculated by subtracting the DTM from the DSM and commonly has a pixel size of 0.5 m to 1.0 m. As a second step, segmentation of the nDSM around the local maxima can be done to derive more information about the tree crowns. Commonly used raster-based segmentation methods are, for example, the watershed segmentation [13], the multi resolution segmentation [14] or an edge-based segmentation [15]. A common problem with identification of individual trees is that there is an underestimation in the result, especially for smaller trees below the dominant tree layer [16].

The analyses based on nDSMs are faster and more robust than those based directly on ALS returns. However, the nDSMs still provide information about local variations in the forest that are related to individual trees. As demonstrated in [17], the canopy volume regression model can also be applied to the rasterized ALS data. Derivation of measures from ALS data in smaller raster cells (e.g., 0.5 m to 1.0 m) could also be a way to compensate for the varying density of the ALS data [18]. The density may vary, for example, due to the pattern of the laser scanner or overlapping strips. If measures are derived from the ALS from a 100–200 m^2 raster cell, these measures are largely influenced by the parts of the raster cell with the highest pulse density.

The purpose of this study is to compare methods to estimate stem volume, stem number and basal area. The first comparison is between measures derived from ALS data in 0.5 m raster cells and variables derived from larger raster cells corresponding to the size of the field plots for estimation of forest variables. The second comparison is between the canopy volume model [7] and a model based on height percentiles and density of ALS data [3] for estimation of stem volume in hemi-boreal forest. The third comparison is between area-based regression models and individual tree-based models for estimation of stem number and basal area.

2. Material

2.1. Study Area

The study area is located in the southwest of Sweden (Lat. 58°N, Long. 13°E). The most common tree species are Norway spruce (*Picea abies*) (38.5% of basal area), Scots pine (*Pinus sylvestris*) (28.0% of basal area), birch (*Betula pendula* and *Betula pubescens*) (18.0% of basal area), oak (*Quercus robur*) (6.0% of basal area), and other broadleaved trees (9.5% of basal area).

2.2. Field Data

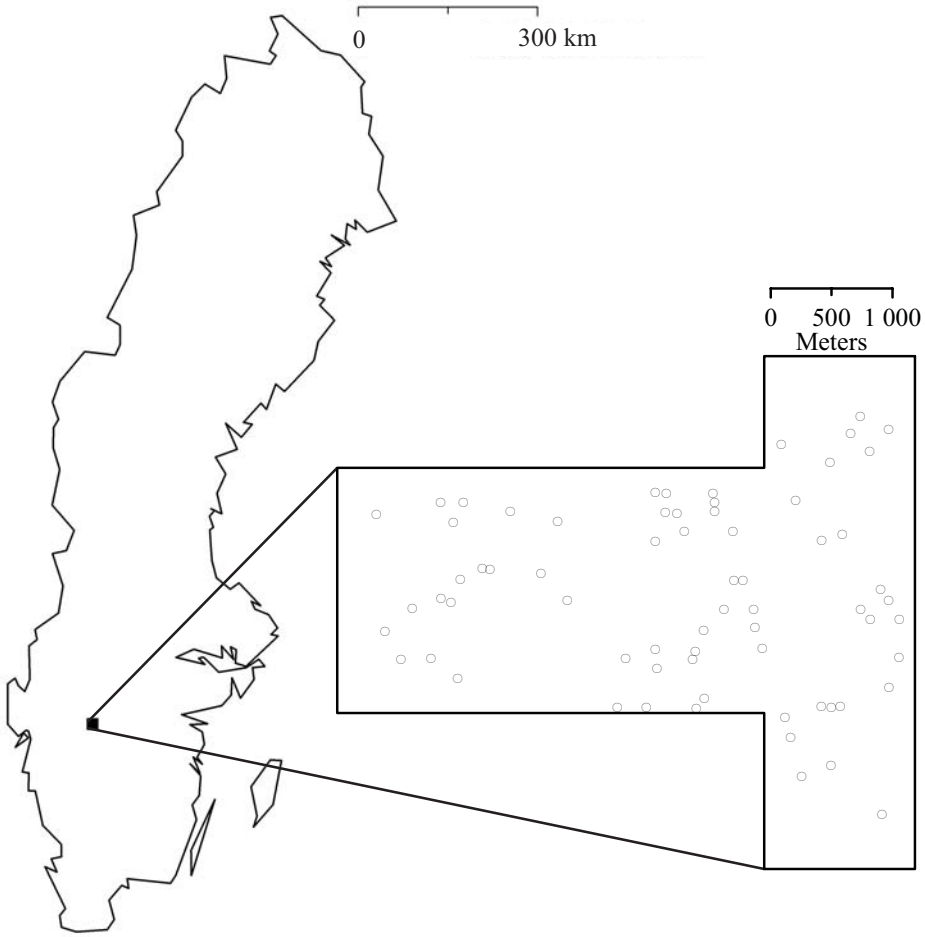
In total sixty-eight circular field plots with 12 m radius were allocated during July and August 2009 (Figure 1). The positions of the center of the field plots were measured using a DGPS with a few dm accuracy after post-processing. Within the field plots, the diameter at breast height (DBH) of all trees with DBH $\geq$ 40 mm were measured using a caliper and the tree species were recorded. For a sub-sample of trees, the heights were also measured using a hypsometer. The sub-sample was randomly selected with inclusion probability proportional to the basal area of the trees.

The stem number N in each field plot was calculated as the number of trees divided by the area A of the field plot. The stem volume of each tree in the sub-sample, where the height was measured, was calculated with specific functions for pine, spruce [19] and oak [20]. For other species, the function for birch was used [19]. To estimate the stem volume of all trees, species specific log-linear regression models were created for pine, spruce, oak, and other species based on the subsample of trees where the height was measured in all field plots simultaneously (Equation (1)).

$$ln(V_j) = \alpha_0 + \alpha_1 DBH_j + \alpha_2 ln(DBH_j) + \epsilon_j \quad (1)$$

The root mean square error (RMSE) at tree level of the regression models was 137 dm^3 (19.1%) for pine, 102 dm^3 (15.9%) for spruce, 389 dm^3 (7.7%) for oak, and 90 dm^3 (25.9%) for other species. The stem volume of all trees was estimated with the respective regression models.

Figure 1. Map of Sweden showing the location of the study area (**left**) and the study area (Lat. 58°N, Long. 13°E) with the extent of the airborne laser scanning (ALS) data and the location of the field plots (**right**).



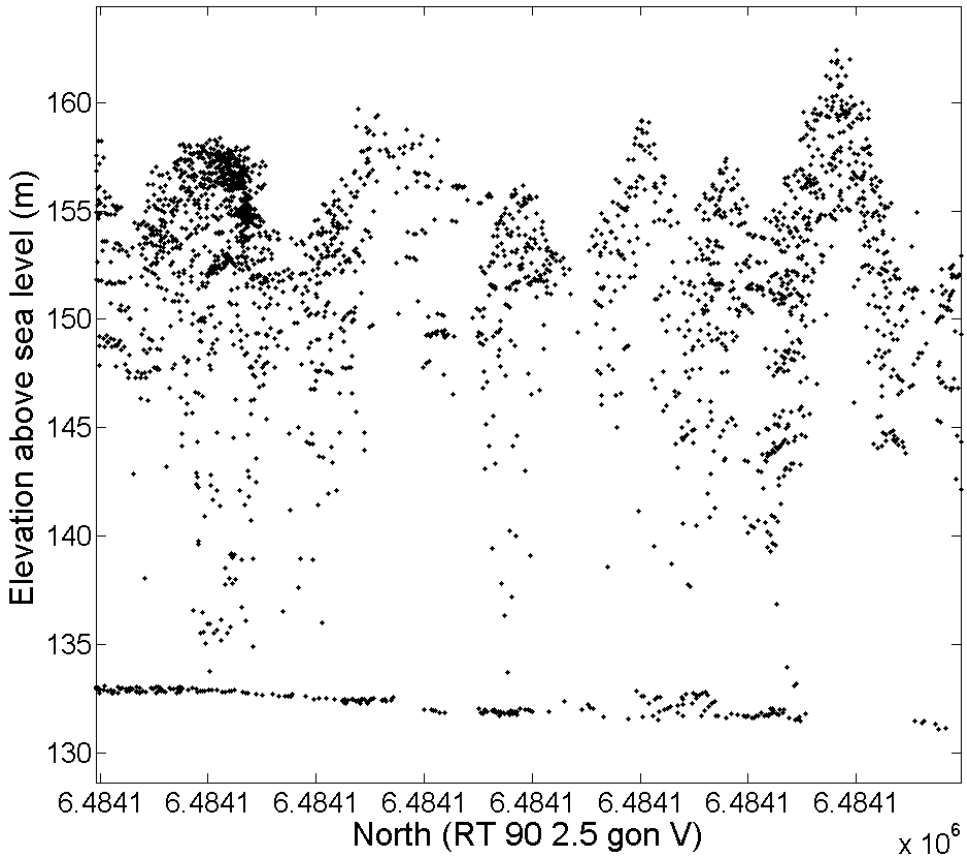
The stem volume V in each field plot was calculated as the sum of the stem volume of all trees in the field plot divided by the area A of the field plot. The basal area BA was calculated in each field plot using Equation (2):

$$BA = \frac{1}{A}\sum_{j=1}^{m} BA_j = \frac{1}{A}\sum_{j=1}^{m} DBH_j^2\pi/4 \quad (2)$$

2.3. ALS Data

The ALS data were acquired on 4 September 2008 using a TopEye MKII ALS system with a wavelength of 1064 nm carried by a helicopter. The flying altitude was 250 m above ground and the average emitted pulse density was 7 m^{-2}. The first and last returns were saved for each laser pulse and the average return density was 11 m^{-2} (Figure 2).

Figure 2. Side view of ALS data in a 10 m wide and 40 m long north-south transect in one field plot.



3. Methods

3.1. Derivation of DTM from ALS Data

ALS returns were classified as ground or non-ground using the progressive Triangular Irregular Network (TIN) densification method [21] in the TerraScan software [22]. A DTM was derived as the mean value of the ground returns in 0.5 m raster cells. TIN interpolation was used for raster cells with no data.

3.2. Statistical ALS Measures

The *z*-values of the ALS returns were normalized with respect to the DTM (Equation (3)).

$$z_{normal} = z - z_{DTM} \quad (3)$$

The following measures were derived from the ALS returns in each circular field plot with 12 m radius.

- The 10th, 20th, …, 100th percentiles of the normalized *z*-values from the ALS returns ≥ 2 m above the DTM in each field plot: p_{10}, p_{20}, …, p_{100}.
- The total number of ALS returns: N_{tot}.
- The number of ALS returns in intervals I_1, I_2, I_3, and I_4: N_1, N_2, N_3, and N_4.

- The number of ALS returns ≥ 2 m and < 34 m above the DTM: N_{veg}.
- The total number of first ALS returns: $N_{f,tot}$.
- The number of first ALS returns in intervals I_1, I_2, I_3, and I_4: $N_{f,1}$, $N_{f,2}$, $N_{f,3}$, and $N_{f,4}$.

where I_1 was $2 \leq z < 10$, I_2 was $10 \leq z < 18$, I_3 was $18 \leq z < 26$, and I_4 was $26 \leq z < 34$ m above the DTM.

The vegetation ratio in each field plot was calculated as $r_{veg} = N_{veg}/N_{tot}$. The fractions of ALS returns in intervals I_j were calculated as $r_{hj}=N_j/N_{tot}$.

The following measures were derived from the ALS returns in raster cells of 0.5 m.

- The mean normalized *z*-value of all ALS returns ≥ 2 m and < 34 m above the DTM: z_{mean}. The mean of this value in all raster cells inside each field plot: h_{mean}.
- The mean normalized *z*-value of all first ALS returns in intervals I_1, I_2, I_3, and I_4: $z_{f,mn,1}$, $z_{f,mn,2}$, $z_{f,mn,3}$, and $z_{f,mn,4}$. The mean of this value in all raster cells inside each field plot: $h_{f,mn,1}$, $h_{f,mn,2}$, $h_{f,mn,3}$, and $h_{f,mn,4}$.
- The maximum normalized z-value of all first ALS returns < 34 m above the DTM (*i.e.*, a first return nDSM): $z_{f,max}$.
- The maximum normalized *z*-value of all ALS returns < 34 m above the DTM (*i.e.*, an nDSM): z_{max}. The 99th percentile of this value in all raster cells inside each field plot: h_{99}.

To calculate the canopy volume for each interval I_j, the relative proportion (between 0 and 1) of first return DSM raster cells, whose heights fell within the interval, was used: $N_{f,j,raster}/N_{f,tot,raster}$. The maximum height of 34 m was chosen based on the maximum tree height in the field data and on the observation that ALS returns ≥ 34 m above the DTM were all erroneous returns high above the tree tops, found in a few field plots. Raster cells without ALS returns were excluded when calculating mean values and percentiles.

3.3. Canopy Volume Estimation

The canopy volume was calculated for four different height classes j = 1, 2, 3 and 4 using Equation (4) [7].

$$\mathrm{V}_{hj} = h_{f,mn,j} \times a_j \tag{4}$$

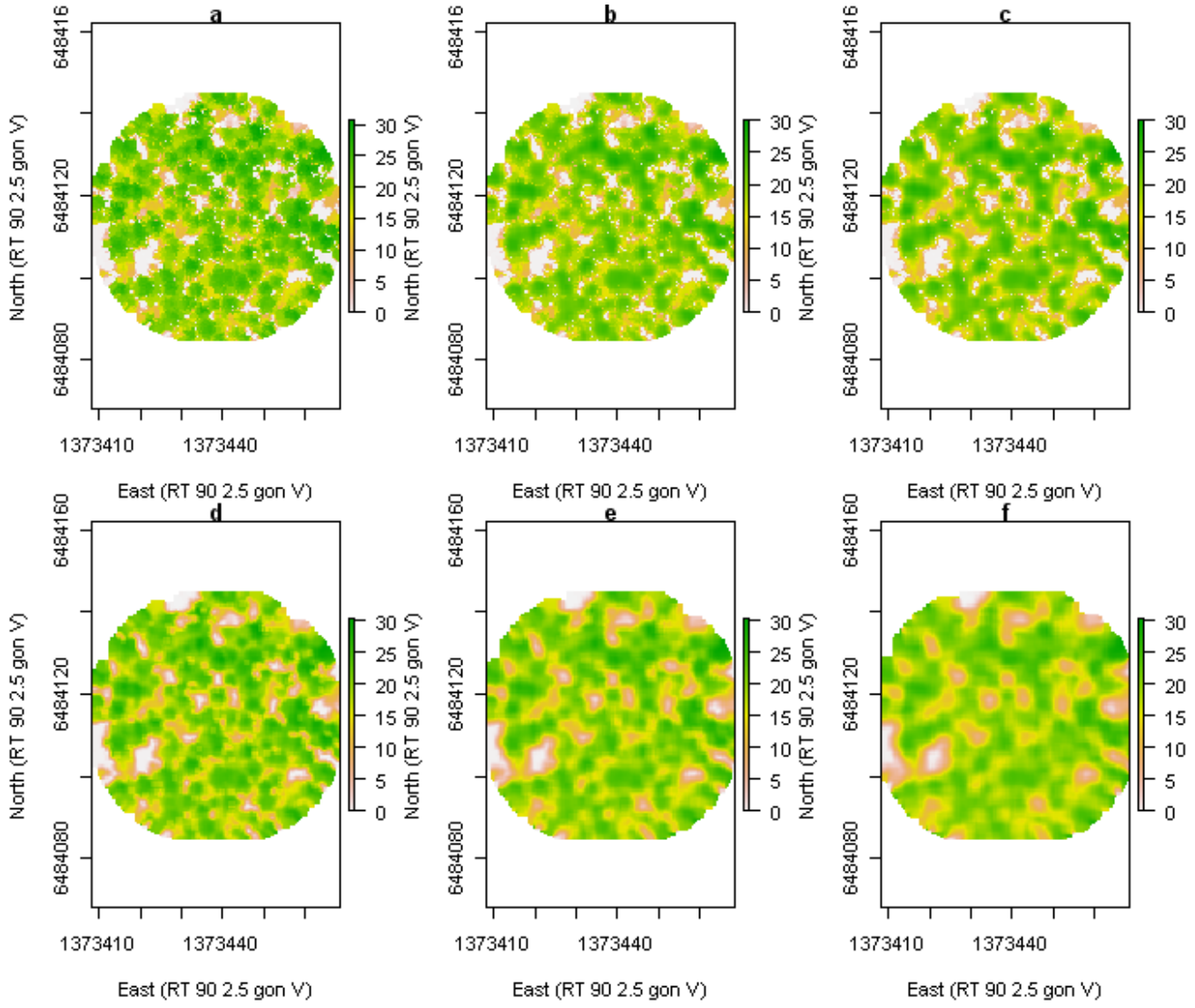
where $a_j=A \times N_{f,j}/N_{f,tot}$ and A is the total area of each field plot. For the calculation of the canopy volume, it is assumed that A is represented by the total number of first echoes $N_{f,tot}$. The canopy volume was also calculated for rasterized ALS data with $a_{j,raster}=A \times N_{f,j,raster}/N_{f,tot,raster}$.

3.4. Local Maxima Detection

Local maxima detection was used to find individual tree tops in the raster of z_{max}. Raster cells without ALS data were iteratively filled with the mean value of the eight surrounding raster cells. Before the local maxima detection was done, different filtering approaches were applied to the raster of z_{max} to remove small variations in the surface model. Three different approaches were tested: in the first case, an m × m mean filter was applied to all raster cells, in the second case, the filter was applied only if $h_{99} \geq h_{lim}$, and in the third case, the filter was applied only for local $z_{max} \geq h_{lim}$, otherwise z_{max} was used

without mean filtering. This was done for h_{lim} = 15 and 20 m and for filter sizes m × m = 3 × 3, 5 × 5, and 7 × 7 (Figure 3). For the local maxima detection, a 3 × 3 max filter was applied to the original and the filtered raster, respectively. Local maxima were defined where the raster values were equal in the raster before and after max filtering. Those raster cells represent the local maxima in the 3 × 3 windows. If several adjacent raster cells fulfilled the criterion, only the midmost raster cell was used as a local maximum. For each detected local maximum, the height of the corresponding raster cell was extracted: h_{loc}.

Figure 3. (**a**) Raster of z_{max} in one field plot, (**b**) 5 × 5 mean filter applied for h_{lim} = 20 m, (**c**) 5 × 5 mean filter applied for h_{lim} = 15 m, (**d**) 3 × 3 mean filter applied to all raster cells, (**e**) 5 × 5 mean filter applied to all raster cells, and (**f**) 7 × 7 mean filter applied to all raster cells.



3.5. Stem Volume Estimation

The stem volume was estimated with five different regression models. The independent variables of the regression models were derived from the ALS returns in each circular field plot with 12 m radius in

two cases (Equations (5) and (8)) and from the ALS returns in raster cells of 0.5 m in the other cases (Equations (6), (7) and (9)).

$$ln(V_i) = \alpha_0 + \alpha_1 ln(p_{60,i}) + \alpha_2 ln(p_{90,i}) + \alpha_3 ln(r_{veg,i}) + \epsilon_i \quad (5)$$

$$ln(V_i) = \alpha_0 + \alpha_1 ln(r_{h3,i}) + \alpha_2 ln(r_{veg,i}) + \alpha_3 ln(h_{mean,i}) + \epsilon_i \quad (6)$$

$$ln(V_i) = \alpha_0 + \alpha_1 ln(r_{h3,i}) + \alpha_2 ln(r_{veg,i}) + \alpha_3 ln(h_{99,i}) + \epsilon_i \quad (7)$$

$$V_i = \alpha_0 + \alpha_1 V_{h1,i} + \alpha_2 V_{h2,i} + \alpha_3 V_{h3,i} + \alpha_4 V_{h4,i} + \epsilon_i \quad (8)$$

$$V_{i,raster} = \alpha_0 + \alpha_1 V_{h1,i,raster} + \alpha_2 V_{h2,i,raster} + \alpha_3 V_{h3,i,raster} + \alpha_4 V_{h4,i,raster} + \epsilon_i \quad (9)$$

The models in Equations (5–7) were selected with best subset regression. For the models in Equations (5–7), the stem volume was calculated as the exponential function of the estimated values. This introduces a bias (e.g., [23,24]). Due to this, the estimates were corrected for logarithmic bias by multiplying the result with the mean value of the stem volumes from the dataset on which the regression models were based, divided by the mean value of the stem volume estimates using the dataset on which the regression models were based [25].

3.6. Stem Number and Basal Area Estimation

The stem number and the basal area were estimated with two methods: (i) an area-based approach and (ii) an individual tree-based approach.

In the area-based approach, the stem number (Equations (10) and (11)) and the basal area (Equations (12) and (13)) were estimated with different regression models. The independent variables of the regression models were derived from the ALS returns in each circular field plot with 12 m radius in two cases (Equations (10) and (12)) and from the ALS returns in raster cells of 0.5 m in two cases (Equations (11) and (13)).

$$N_i = \alpha_0 + \alpha_1 p_{30,i} + \alpha_2 p_{50,i} + \alpha_3 p_{80,i} + \alpha_4 r_{veg,i} + \epsilon_i \quad (10)$$

$$N_i = \alpha_0 + \alpha_1 r_{h3,i} + \alpha_2 r_{veg,i} + \alpha_3 h_{99,i} + \epsilon_i \quad (11)$$

$$ln(BA_i) = \alpha_0 + \alpha_1 ln(p_{10,i}) + \alpha_2 ln(p_{60,i}) + \alpha_3 ln(r_{veg,i}) + \epsilon_i \quad (12)$$

$$ln(BA_i) = \alpha_0 + \alpha_1 ln(r_{h3,i}) + \alpha_2 ln(r_{veg,i}) + \alpha_3 ln(h_{99,i}) + \epsilon_i \quad (13)$$

The models were selected with best subset regression. The models in Equations (12) and (13) were corrected for logarithmic bias [25].

In the individual tree-based approach, values of DBH were calculated using a relationship between DBH and tree height based on a regression model for the subsample of trees where the heights were measured (Equation (14)):

$$h_j = \beta_0 + \beta_1 ln(DBH_j) + \epsilon_j \quad (14)$$

where DBH_j is the DBH of tree j and h_j is the height of tree j and assuming that the heights of the local maxima h_{loc} were the tree heights. The regression model was based on all tree species since the tree species was not determined from the ALS data. The stem number was derived as the number of local maxima in a field plot divided by the area, and the basal area was calculated from the estimated DBH values.

3.7. Validation

The accuracy of the estimates from ALS data was validated using leave-one-out cross-validation for one field plot at a time: one field plot was excluded; the parameters of the models were estimated based on the remaining field plots and then applied to the excluded field plot to estimate forest variables. The accuracy was validated with the field-measured values using the RMSE and the bias (Equations (15) and (16)).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2}{n}} \quad (15)$$

$$bias = \frac{\sum_{i=1}^{n}(\hat{Y}_i - Y_i)}{n} \quad (16)$$

where Y_i is the stem volume, stem number or basal area in plot i, and n is the number of field plots. Furthermore, scatter plots were generated. The validation was done by both including all field plots as well as excluding field plots with > 80% basal area from oak, which were five field plots out of 68.

4. Results

The RMSE of the estimated stem volume was largest for the regression model in Equation (5) and smallest for the regression model in Equation (7) (Table 1). The bias was less than 2% for all regression models. For larger field-measured values, the deviation between estimated and field-measured values was larger and a few outliers were observed (Figure 4).

Table 1. RMSE and bias for stem volume from regression models in Equations (5–9), using all field plots.

	RMSE		**Bias**	
	$m^3 \cdot ha^{-1}$	% of mean	$m^3 \cdot ha^{-1}$	% of mean
Regression model in Equation (5)	75.1	41.9%	0.3	0.2%
Regression model in Equation (6)	68.6	38.2%	−0.3	−0.2%
Regression model in Equation (7)	66.9	37.3%	−0.4	−0.2%
Regression model in Equation (8)	71.5	39.8%	0.7	0.4%
Regression model in Equation (9)	68.2	38.0%	0.2	0.1%

Figure 4. Estimates of stem volume from regression models in (**a**) Equation (5), (**b**) Equation (7), (**c**) Equation (8), and (**d**) Equation (9), using all field plots.

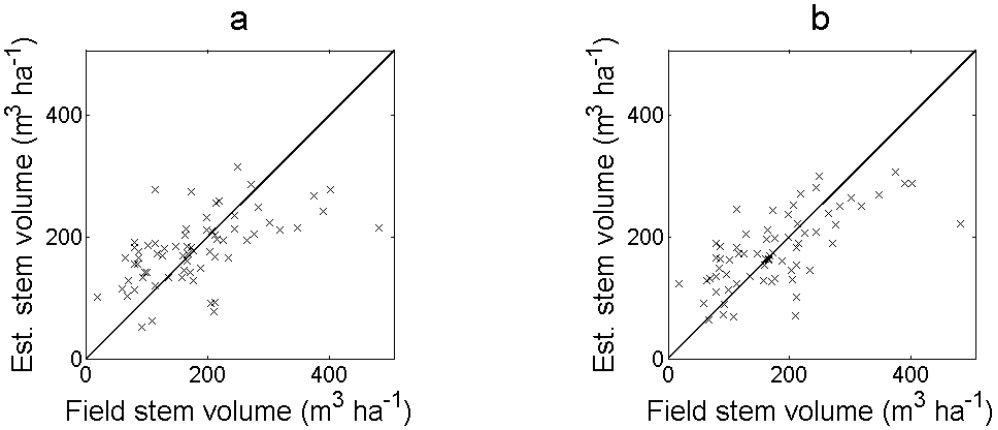
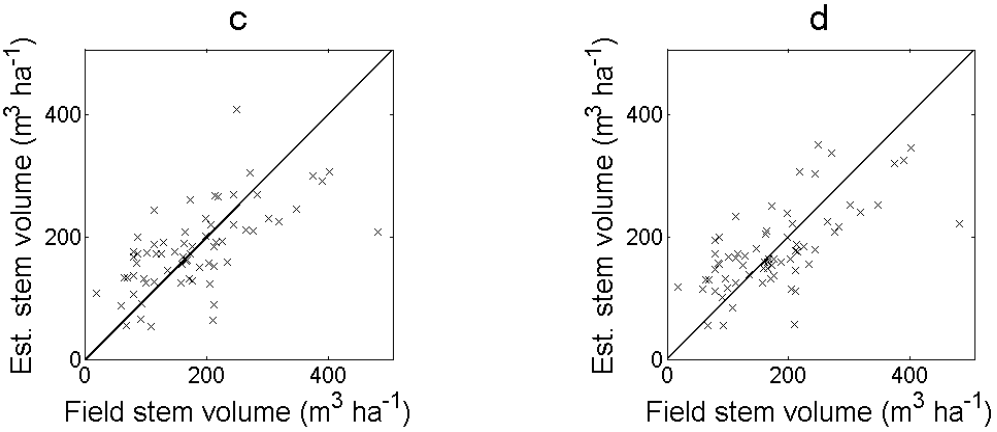


Figure 4. *Cont.*



The RMSE of the estimated stem volume was smaller when excluding field plots with > 80% basal area from oak (Table 2). The regression model in Equation (7) had the smallest RMSE also in this case. The bias was still less than 2%. The estimated values showed fewer obvious outliers (Figure 5).

Table 2. RMSE and bias for stem volume from regression models in Equations (5–9), excluding field plots with > 80% basal area from oak.

	RMSE		**Bias**	
	$m^3 \cdot ha^{-1}$	% of mean	$m^3 \cdot ha^{-1}$	% of mean
Regression model in Equation (5)	65.0	36.8%	0.2	0.1%
Regression model in Equation (6)	57.1	32.3%	−0.1	0.0%
Regression model in Equation (7)	55.7	31.5%	−0.3	−0.2%
Regression model in Equation (8)	60.1	34.0%	1.0	0.5%
Regression model in Equation (9)	58.0	32.8%	0.3	0.2%

Figure 5. Estimates of stem volume from regression models in (**a**) Equation (5), (**b**) Equation (7), (**c**) Equation (8), and (**d**) Equation (9), excluding field plots with > 80% basal area from oak.

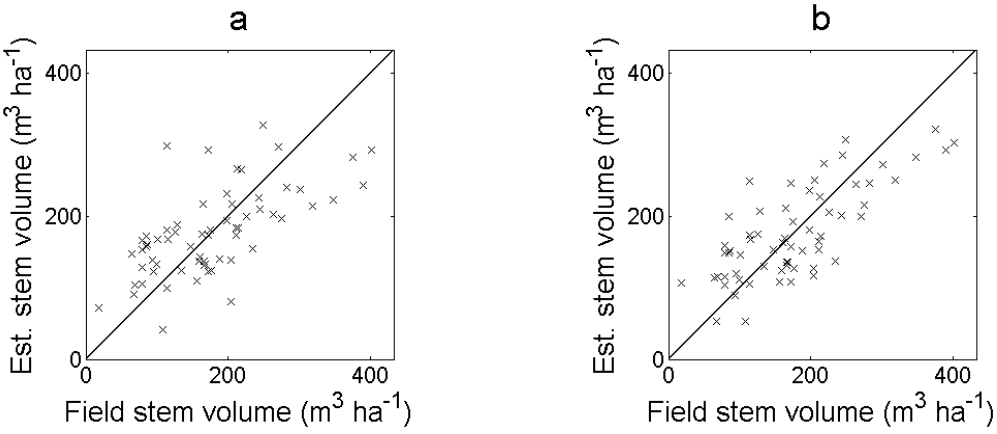
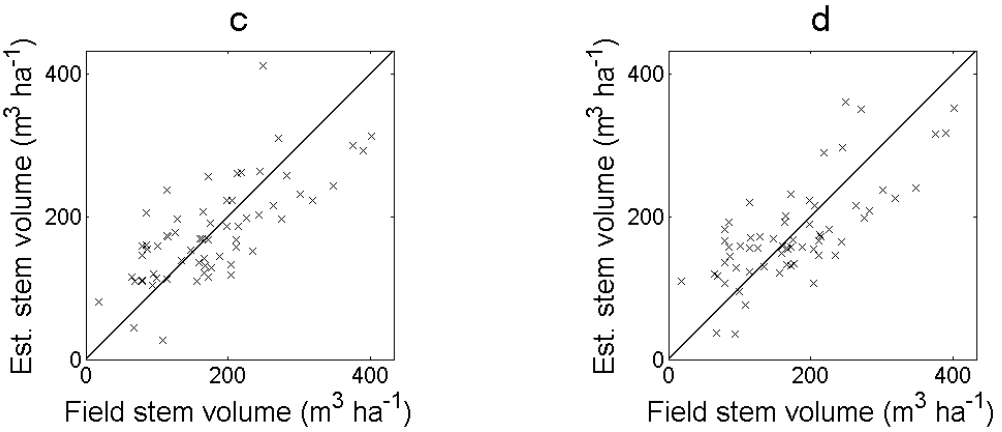


Figure 5. *Cont.*



The RMSE of the estimated stem number was smallest for the regression model in Equation (11) and second smallest for the regression model in Equation (10). The bias was close to zero in both cases (Table 3). When tree tops were identified from local maxima in the nDSM, the RMSE of the estimated stem number was in general larger for larger filter sizes and smaller for conditions when z_{max} was filtered more often (*i.e.*, always filtered or lower h_{lim}). The bias was in general lower the more z_{max} was filtered (*i.e.*, larger filter sizes or filtered more often). All cases showed outliers for large field-measured values and low estimated values (Figure 6).

Table 3. RMSE and bias for stem number from regression models in Equations (10) and (11) and derived from local maxima, using all field plots.

	RMSE		**Bias**	
	ha^{-1}	% of mean	ha^{-1}	% of mean
Regression model in Equation (10)	410.8	55.8%	−13.0	−1.8%
Regression model in Equation (11)	387.4	52.7%	−4.3	−0.6%
Mean filtered 3 × 3	507.8	69.0%	−252.9	−34.4%
Mean filtered 3 × 3 if $h_{99} \geq 15$ m	506.3	68.8%	−241.2	−32.8%
Mean filtered 3 × 3 for local $z_{max} \geq 15$ m	466.4	63.4%	65.7	8.9%
Mean filtered 3 × 3 if $h_{99} \geq 20$ m	593.0	80.6%	−3.6	−0.5%
Mean filtered 3 × 3 for local $z_{max} \geq 20$ m	675.7	91.9%	460.0	62.5%
Mean filtered 5 × 5	584.4	79.4%	−372.2	−50.6%
Mean filtered 5 × 5 if $h_{99} \geq 15$ m	582.3	79.2%	−359.5	−48.9%
Mean filtered 5 × 5 for local $z_{max} \geq 15$ m	476.2	64.7%	−17.6	−2.4%
Mean filtered 5 × 5 if $h_{99} \geq 20$ m	634.1	86.2%	−92.0	−12.5%
Mean filtered 5 × 5 for local $z_{max} \geq 20$ m	662.4	90.0%	426.2	57.9%
Mean filtered 7 × 7	633.2	86.1%	−432.0	−58.7%
Mean filtered 7 × 7 if $h_{99} \geq 15$ m	631.5	85.8%	−419.7	−57.0%
Mean filtered 7 × 7 for local $z_{max} \geq 15$ m	496.7	67.5%	−71.5	−9.7%
Mean filtered 7 × 7 if $h_{99} \geq 20$ m	653.2	88.8%	−127.8	−17.4%
Mean filtered 7 × 7 for local $z_{max} \geq 20$ m	651.0	88.5%	388.8	52.9%

Figure 6. Estimates of stem number from (**a**) regression model in Equation (10), (**b**) derived from local maxima mean filtered 3 × 3 if $h_{99} \geq 15$ m, (**c**) mean filtered 3 × 3 for local $z_{max} \geq 15$ m, (**d**) from regression model in Equation (11), (**e**) mean filtered 3 × 3 if $h_{99} \geq 20$ m, and (**f**) mean filtered 3 × 3 for local $z_{max} \geq 20$ m, using all field plots.

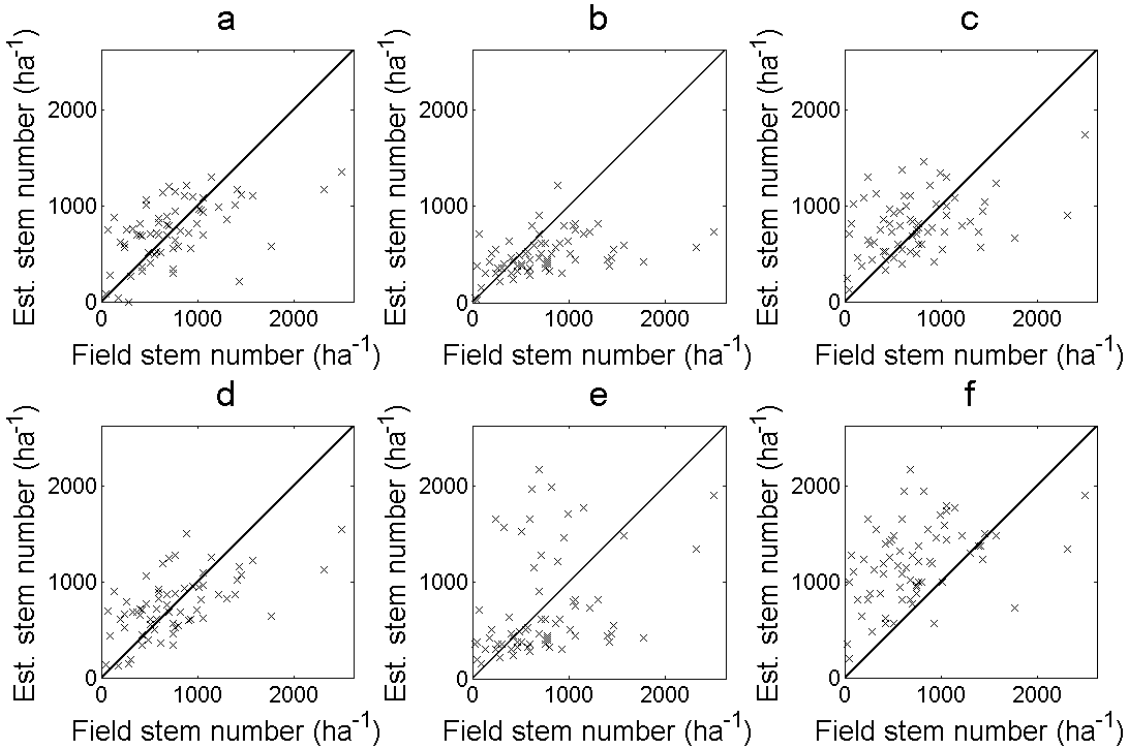
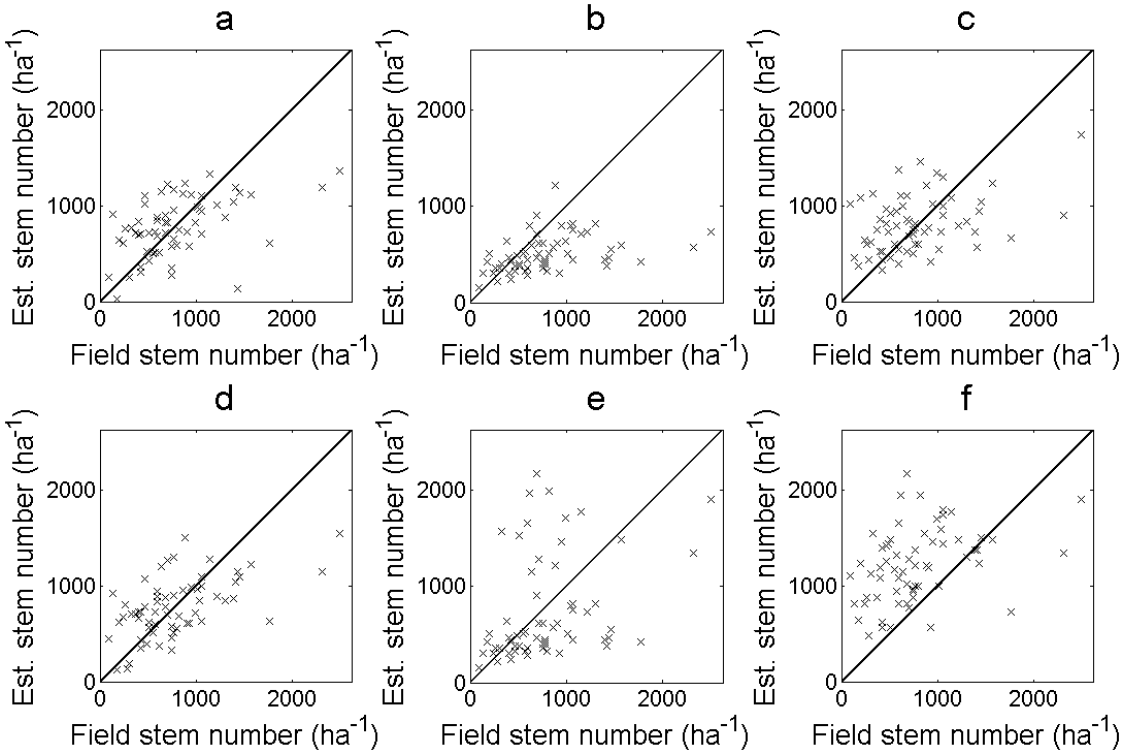


Table 4. RMSE and bias for stem number from regression models in Equations (10) and (11) and derived from local maxima, excluding field plots with > 80% basal area from oak.

	RMSE		**Bias**	
	ha^{-1}	% of mean	ha^{-1}	% of mean
Regression model in Equation (10)	420.2	53.4%	−14.4	−1.8%
Regression model in Equation (11)	393.9	50.0%	−4.5	−0.6%
Mean filtered 3 × 3	518.2	65.8%	−293.3	−37.3%
Mean filtered 3 × 3 if $h_{99} \geq 15$ m	516.6	65.6%	−280.7	−35.7%
Mean filtered 3 × 3 for local $z_{max} \geq 15$ m	447.3	56.8%	26.7	3.4%
Mean filtered 3 × 3 if $h_{99} \geq 20$ m	580.8	73.8%	−49.5	−6.3%
Mean filtered 3 × 3 for local $z_{max} \geq 20$ m	649.0	82.4%	431.9	54.9%
Mean filtered 5 × 5	604.6	76.8%	−411.6	−52.3%
Mean filtered 5 × 5 if $h_{99} \geq 15$ m	602.4	76.5%	−397.9	−50.5%
Mean filtered 5 × 5 for local $z_{max} \geq 15$ m	464.4	59.0%	−58.6	−7.4%
Mean filtered 5 × 5 if $h_{99} \geq 20$ m	630.1	80.0%	−138.6	−17.6%
Mean filtered 5 × 5 for local $z_{max} \geq 20$ m	635.4	80.7%	396.1	50.3%
Mean filtered 7 × 7	656.3	83.4%	−468.8	−59.5%
Mean filtered 7 × 7 if $h_{99} \geq 15$ m	654.5	83.1%	−455.4	−57.8%
Mean filtered 7 × 7 for local $z_{max} \geq 15$ m	491.5	62.4%	−113.3	−14.4%
Mean filtered 7 × 7 if $h_{99} \geq 20$ m	652.0	82.8%	−173.7	−22.1%
Mean filtered 7 × 7 for local $z_{max} \geq 20$ m	622.7	79.1%	356.1	45.2%

The RMSE of the estimated stem number was slightly larger for the regression model when excluding field plots with > 80% basal area from oak (Table 4 and Figure 7). When tree tops were identified from local maxima in the nDSM, the RMSE was larger when z_{max} was always mean filtered or mean filtered if $h_{99} \geq 15$ m, and smaller for the other cases. The bias changed in a negative direction for all cases. The relative order of the RMSE and bias was the same as when all field plots were included.

Figure 7. Estimates of stem number from (**a**) regression model in Equation (10), (**b**) derived from local maxima mean filtered 3 × 3 if $h_{99} \geq 15$ m, (**c**) mean filtered 3 × 3 for local $z_{max} \geq 15$ m, (**d**) from regression model in Equation (11), (**e**) mean filtered 3 × 3 if $h_{99} \geq 20$ m, and (**f**) mean filtered 3 × 3 for local $z_{max} \geq 20$ m, excluding field plots with > 80% basal area from oak.

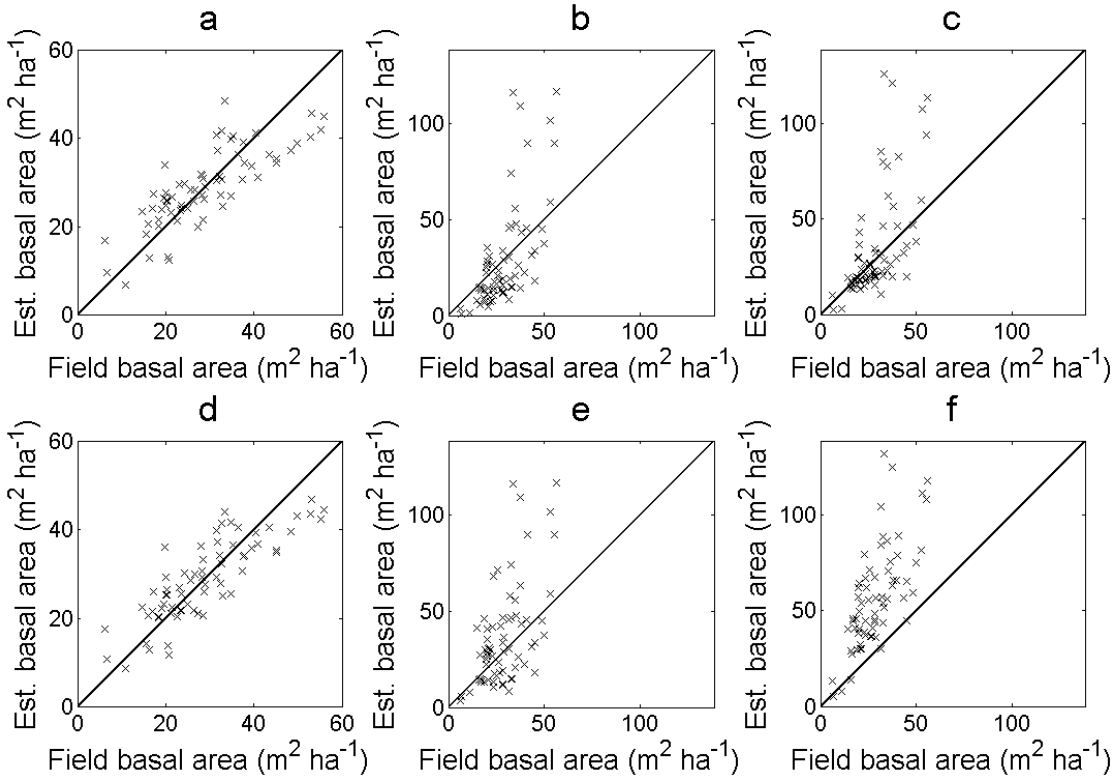


The RMSE and bias of the estimated basal area was smallest for the regression model in Equation (13) and Equation (12) (Table 5). When the basal area was calculated from the DBH derived from the local maxima, the RMSE of the estimated basal area was in general smaller for larger filter sizes and for conditions when z_{max} was filtered more often (*i.e.*, always filtered or lower h_{lim}). The bias was in general lower the more z_{max} was filtered (*i.e.*, larger filter sizes or filtered more often). The estimated values deviated more from the field-measured values for the basal area calculated from the DBH derived from the local maxima than for the regression model (Figure 8). In the first case, the basal area was overestimated for larger field-measured values.

Table 5. RMSE and bias for basal area from regression models in Equations (12) and (13) and diameter at breast height (DBH) derived from local maxima, using all field plots.

	RMSE		**Bias**	
	m^2 ha^{-1}	% of mean	m^2 ha^{-1}	% of mean
Regression model in Equation (12)	6.7	23.2%	0.0	0.0%
Regression model in Equation (13)	6.2	21.5%	0.0	−0.1%
DBH local maxima, mean filtered 3 × 3	37.6	130.1%	14.5	50.3%
DBH local maxima, mean filtered 3 × 3 if $h_{99} \geq 15$ m	37.6	130.0%	14.7	50.7%
DBH local maxima, mean filtered 3 × 3 for local $z_{max} \geq 15$ m	40.7	140.7%	19.9	68.7%
DBH local maxima, mean filtered 3 × 3 if $h_{99} \geq 20$ m	39.1	135.0%	21.0	72.8%
DBH local maxima, mean filtered 3 × 3 for local $z_{max} \geq 20$ m	50.8	175.5%	38.7	133.9%
DBH local maxima, mean filtered 5 × 5	21.6	74.8%	0.4	1.3%
DBH local maxima, mean filtered 5 × 5 if $h_{99} \geq 15$ m	21.6	74.6%	0.5	1.8%
DBH local maxima, mean filtered 5 × 5 for local $z_{max} \geq 15$ m	23.8	82.1%	6.4	22.0%
DBH local maxima, mean filtered 5 × 5 if $h_{99} \geq 20$ m	23.5	81.3%	8.1	27.9%
DBH local maxima, mean filtered 5 × 5 for local $z_{max} \geq 20$ m	34.4	119.1%	27.9	96.5%
DBH local maxima, mean filtered 7 × 7	16.6	57.2%	−7.8	−27.1%
DBH local maxima, mean filtered 7 × 7 if $h_{99} \geq 15$ m	16.5	57.0%	−7.7	−26.5%
DBH local maxima, mean filtered 7 × 7 for local $z_{max} \geq 15$ m	17.4	60.0%	−1.7	−5.8%
DBH local maxima, mean filtered 7 × 7 if $h_{99} \geq 20$ m	18.3	63.3%	0.7	2.4%
DBH local maxima, mean filtered 7 × 7 for local $z_{max} \geq 20$ m	26.2	90.7%	20.7	71.5%

Figure 8. Estimates of basal area (**a**) from regression model in Equation (12), (**b**) DBH derived from local maxima mean filtered 5 × 5 if $h_{99} \geq 15$ m, (**c**) mean filtered 5 × 5 for local $z_{max} \geq 15$ m, (**d**) from regression model in Equation (13), (**e**) mean filtered 5 × 5 if $h_{99} \geq 20$ m, and (**f**) mean filtered 5 × 5 for local $z_{max} \geq 20$ m, using all field plots.



The RMSE of the estimated basal area was smaller for the regression model in Equations (12) and (13) when excluding field plots with > 80% basal area from oak (Table 6). When the basal area was calculated from the DBH derived from the local maxima, the RMSE was larger and the bias was higher. The relative order of the RMSE and bias was the same as when all field plots were included. The estimated values showed a similar pattern as when all field plots were included (Figure 9).

Table 6. RMSE and bias for basal area from regression model in Equations (12) and (13) and DBH derived from local maxima, excluding field plots with > 80% basal area from oak.

	RMSE		**Bias**	
	m^2 ha^{-1}	% of mean	m^2 ha^{-1}	% of mean
Regression model in Equation (12)	6.5	22.1%	0.0	0.1%
Regression model in Equation (13)	6.1	20.8%	0.0	−0.1%
DBH local maxima, mean filtered 3 × 3	39.0	132.0%	16.1	54.4%
DBH local maxima, mean filtered 3 × 3 if $h_{99} \geq 15$ m	39.0	131.9%	16.2	54.8%
DBH local maxima, mean filtered 3 × 3 for local $z_{max} \geq 15$ m	42.2	142.9%	21.6	73.1%
DBH local maxima, mean filtered 3 × 3 if $h_{99} \geq 20$ m	40.6	137.2%	22.6	76.3%
DBH local maxima, mean filtered 3 × 3 for local $z_{max} \geq 20$ m	52.6	177.8%	41.1	138.9%
DBH local maxima, mean filtered 5 × 5	22.3	75.3%	1.2	4.0%
DBH local maxima, mean filtered 5 × 5 if $h_{99} \geq 15$ m	22.2	75.2%	1.3	4.5%
DBH local maxima, mean filtered 5 × 5 for local $z_{max} \geq 15$ m	24.6	83.2%	7.4	24.9%
DBH local maxima, mean filtered 5 × 5 if $h_{99} \geq 20$ m	24.3	82.3%	8.9	30.1%
DBH local maxima, mean filtered 5 × 5 for local $z_{max} \geq 20$ m	35.6	120.5%	29.5	99.9%
DBH local maxima, mean filtered 7 × 7	16.8	56.7%	−7.4	−25.1%
DBH local maxima, mean filtered 7 × 7 if $h_{99} \geq 15$ m	16.7	56.4%	−7.3	−24.6%
DBH local maxima, mean filtered 7 × 7 for local $z_{max} \geq 15$ m	17.8	60.4%	−1.2	−4.0%
DBH local maxima, mean filtered 7 × 7 if $h_{99} \geq 20$ m	18.8	63.6%	1.1	3.7%
DBH local maxima, mean filtered 7 × 7 for local $z_{max} \geq 20$ m	27.0	91.5%	21.7	73.5%

Figure 9. Estimates of basal area (**a**) from regression model in Equation (12), (**b**) DBH derived from local maxima mean filtered 5 × 5 if $h_{99} \geq 15$ m, (**c**) mean filtered 5 × 5 for local $z_{max} \geq 15$ m, (**d**) from regression model in Equation (13), (**e**) mean filtered 5 × 5 if $h_{99} \geq 20$ m, and (**f**) mean filtered 5 × 5 for local $z_{max} \geq 20$ m, excluding field plots with > 80% basal area from oak.

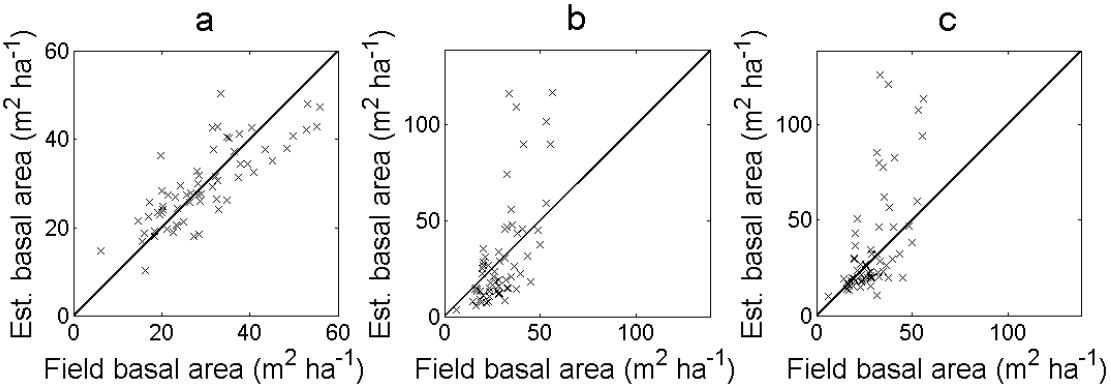
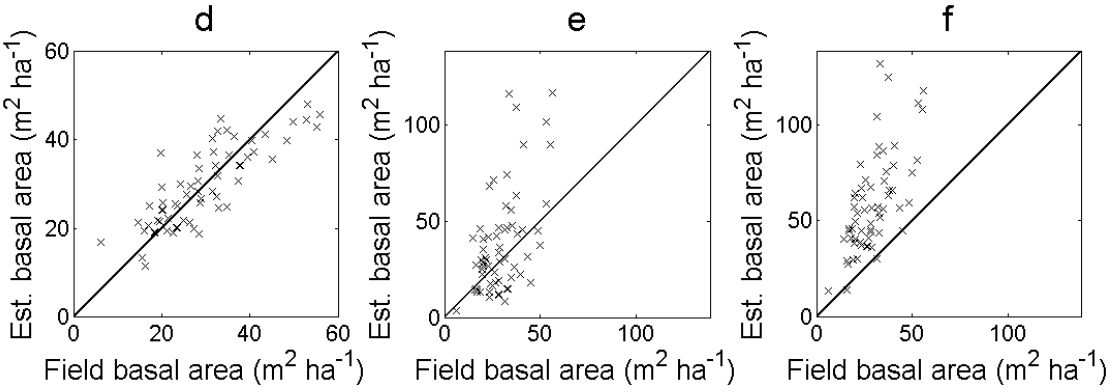


Figure 9. *Cont.*



5. Discussion

The most accurate estimates of stem volume, stem number and basal area were achieved with regression models that used rasterized (0.5 m raster cells) ALS data as input instead of 3D point cloud data directly. This suggests that the raster cells can compensate for the varying density of the ALS data and the variability of the forest properties within the field plots. For the two stem volume models that used input measures calculated at plot level from the normalized 3D point cloud directly, the canopy volume regression model was more accurate than the log-log regression model including the vegetation ratio and measures of the height of the ALS returns. However, the most accurate estimate was achieved with a log-log regression model including the vegetation ratio and a measure of the maximum height of the ALS returns derived from 0.5 m raster cells. Apart from the canopy volume models, the final models were selected with best subset regression, which means that the selection of independent variables was based on the reference data. Since the parameters of the regression model are also estimated based on the reference data, the model can be fitted very well to the reference data. However, it requires that the local reference dataset is large enough to base the models on. The canopy volume model is stable in the sense that the independent variables are not selected based on the local reference data, which might have advantages for estimation of stem volume for large areas. The stem volume used as ground truth was estimated with regression models with a comparatively high RMSE, which was around 20% for most of the trees. This makes the validation more uncertain. Excluding field plots with > 80% basal area from oak resulted in fewer outliers since most of the outliers were oak dominated field plots. Previous studies have reported larger errors for estimation of stem volume and basal area in mixed forest than in coniferous dominated forest [2,17,26] since field plots with different properties are included in the same model. The stem volume of oak is generally higher than that of most other tree species having the same tree height. In this study, only five out of 68 field plots were oak dominated. This means that the models where all field plots were included were mainly based on forest with a smaller fraction of oak and resulted in large errors when they were applied to oak dominated forest.

The estimation of the basal area showed fewer outliers for large field-measured values than the estimation of stem volume. The reason may be that the outliers for stem volume were mostly oak dominated field plots and that the relationship between DBH and tree height is more similar for oak and other tree species than the relationship between stem volume and tree height. The RMSE of the

regression estimates decreased slightly when excluding oak dominated field plots in the same way as for the estimation of stem volume. However, the RMSE and the bias of the basal area derived from local maxima increased when excluding oak dominated field plots. This may be because the regression model used for calculating DBH from the heights of the local maxima underestimated DBH for tall trees and the oaks were taller than the average tree. This negative contribution to the bias disappeared when excluding the oak dominated field plots and the result was a higher bias.

The bias of the estimated stem number changed in the negative direction when excluding oak dominated field plots. The reason was that the excluded field plots were outliers with an overestimated stem number. This may be due to the canopy of oak having more small variations than other species, which gives rise to several local maxima within the same tree crown. A few other field plots were outliers with an underestimated stem number for large field-measured values. This is expected when identifying trees from local maxima since trees below the dominant tree layer are not visible in the nDSM.

The estimates of stem number and basal area were more accurate for the regression models then when identifying tree tops from local maxima in the nDSM. The advantage of the latter is that a list of DBH estimates is produced at the same time. Distributions of DBH have previously been estimated from height and density measures from ALS data and theoretical diameter distribution models [27] and as percentiles of DBH [28]. The advantage of using local maxima in the nDSM is that they also describe the horizontal distribution of the ALS data that percentiles and density do not.

When tree tops were identified from local maxima in the nDSM, the RMSE of the estimated stem number was smallest when the nDSM was mean filtered for local $z_{max} \geq 15$ m and largest when the nDSM was mean filtered for local $z_{max} \geq 20$ m. The bias was large and positive when the nDSM was mean filtered for local $z_{max} \geq 20$ m, and closer to zero when the nDSM was mean filtered for local $z_{max} \geq 15$ m. The large positive bias in the first case was probably caused by small variations that gave rise to local maxima then identified as tree tops since the nDSM was filtered less often than for the other cases. The nDSM was filtered more often with the condition $h_{99} \geq 15$ m than local $z_{max} \geq 15$ m and the result was a lower bias in the first case. The same effect was visible for $h_{99} \geq 20$ m and local $z_{max} \geq 20$ m. The accuracy was similar when the nDSM was always mean filtered and when the nDSM was mean filtered if $h_{99} \geq 15$ m. This is probably because h_{99} was rarely below 15 m, so in the second case the nDSM was almost always filtered.

For the estimated basal area derived from local maxima, the RMSE was smallest when the nDSM was always mean filtered or mean filtered if $h_{99} \geq 15$ m, and largest when the nDSM was mean filtered for local $z_{max} \geq 20$ m. The bias was lowest in the first case and highest in the second case. An adaptive filtering may improve the identification of local maxima corresponding to tree tops but the method may also be very sensitive to parameter settings. Additionally, the filter sizes are limited to odd multiples of the size of the raster cells. This means that the conditions for setting different parameters must be chosen carefully. In future work, definition of the conditions that can be applied to different forest types will be needed.

The RMSE of the estimated stem number was smallest when the nDSM was mean filtered with a 3 × 3 filter and the bias was highest. The RMSE of the estimated stem number was larger when a 5 × 5 and 7 × 7 filter was used and the bias was lower. This suggests that the larger filter sizes removed small variations in the nDSM that would otherwise have given rise to local maxima. However, the

RMSE of the estimated basal area was smallest when a 7 × 7 filter was used and the bias was lowest. The bias was large and positive for the filter size 3 × 3 (*i.e.*, the basal area was overestimated) and the basal area was overestimated for larger field-measured values for all filter sizes. Since the basal area was calculated from the DBH and the DBH was derived from the heights of the local maxima in the nDSM, this suggests that the heights of the local maxima overestimated the tree heights. The reason may be that not all local maxima corresponded to tree tops. Tree tops below the dominant tree layer do not give rise to local maxima in the nDSM. This means that the stem number will be underestimated if the nDSM is filtered so that only tree tops give rise to local maxima. If the number of local maxima is equal to the stem number, some of the local maxima do not correspond to tree tops and the heights of the local maxima overestimate the heights of the trees below the dominant tree layer. This may explain why the basal area was overestimated with a 3 × 3 filter size even though the estimate of the stem number was most accurate.

6. Conclusions

This study has compared estimation of forest variables from regression models based on measures derived from ALS data in small (0.5 m) raster cells and based on variables derived from the 3D point cloud. The RMSE of the results achieved from regression models based on 0.5 m raster cells were approximately 2–5% lower than those achieved from the 3D point cloud, which suggests that the smaller raster cells can compensate for the varying density of the ALS data. Once the ALS data have been rasterized, the raster cells may be aggregated to any area unit suitable for the application, which means that the approach is easy to integrate in operational work flows and may have advantages in terms of computational issues. Only one raster cell size was used in the study. The size of the raster cells could be optimized or the varying density of the ALS data could be compensated for in other ways, for example, by weighting the ALS returns depending on local density. This study has also compared a canopy volume model for estimation of stem volume in hemi-boreal forest with a model based on height percentiles and density of ALS data selected with best subset regression. The most accurate estimate was achieved with a log-log regression model including the vegetation ratio and a measure of the maximum height of the ALS returns, although the RMSE was only 1% lower. Hence, both model types may be used for estimation of stem volume. However, the selection of model type can be based on many considerations, for example, if the local reference dataset is large enough for best subset regression. Finally, the study has compared area-based regression models and individual tree-based models with different filtering conditions for estimation of stem number and basal area. The most accurate results were achieved from the regression models (7–11% lower RMSE for stem number; 39–40% lower RMSE for basal area compared to the best filtering conditions). However, an advantage of the individual tree-based models is that a list of DBH is estimated at the same time. When individual trees are derived from local maxima in an nDSM, the filter sizes and the conditions for filtering the nDSM must be carefully selected. Criteria for selection of filter sizes and conditions still remain to be defined for different forest types.

Acknowledgments

The field inventory and the acquisition of ALS data were financed by the Hildur and Sven Wingquist foundation. Parts of this work were done within the project LASER-WOOD (822030), funded by the Austrian Klima- und Energiefonds in the framework of the program "NEUE ENERGIEN 2020". We would like to thank Heather Reese and Neil Cory who have checked the language in the manuscript.

References

1. Hudak, A.T.; Evans, J.S.; Smith, A.M.S. Lidar utility for natural resource managers. *Remote Sens.* **2009**, *1*, 934–951.
2. Holmgren, J. Prediction of tree height, basal area and stem volume in forest stands using airborne laser scanning. *Scand. J. For. Res.* **2004**, *19*, 543–553.
3. Næsset, E. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sens. Environ.* **2002**, *80*, 88–99.
4. Nelson, R.; Swift, R.; Krabill, W. Using airborne lasers to estimate forest canopy and stand characteristics. *J. For.* **1988**, *86*, 31–38.
5. Evans, J.S.; Hudak, A.T.; Faux, R.; Smith, A.M.S. Discrete return lidar in natural resources: Recommendations for project planning, data processing, and deliverables. *Remote Sens.* **2009**, *1*, 776–794.
6. Næsset, E. Airborne laser scanning as a method in operational forest inventory: Status of accuracy assessments accomplished in scandinavia. *Scand. J. For. Res.* **2007**, *22*, 433–442.
7. Hollaus, M.; Wagner, W.; Schadauer, K.; Maier, B.; Gabler, K. Growing stock estimation for alpine forests in austria: A robust lidar-based approach. *Can. J. For. Res.* **2009**, *39*, 1387–1400.
8. Hyyppä, J.; Inkinen, M. Detecting and estimating attributes for single trees using laser scanner. *Photogramm. J. Fin.* **1999**, *16*, 27–42.
9. Brandtberg, T.; Walter, F. Automated delineation of individual tree crowns in high spatial resolution aerial images by multiple-scale analysis. *Mach. Vis. Appl.* **1998**, *11*, 64–73.
10. Holmgren, J.; Persson, Å. Identifying species of individual trees using airborne laser scanner. *Remote Sens. Environ.* **2004**, *90*, 415–423.
11. Edson, C.; Wing, M.G. Airborne light detection and ranging (lidar) for individual tree stem location, height, and biomass measurements. *Remote Sens.* **2011**, *3*, 2494–2528.
12. Yu, X.; Hyyppä, J.; Holopainen, M.; Vastaranta, M. Comparison of area-based and individual tree-based methods for predicting plot-level forest attributes. *Remote Sens.* **2010**, *2*, 1481–1495.
13. Persson, Å.; Holmgren, J.; Söderman, U. Detecting and measuring individual trees using an airborne laser scanner. *Photogramm. Eng. Remote Sensing* **2002**, *68*, 925–932.
14. Baatz, M.; Schäpe, A. Multiresolution Segmentation: An Optimization Approach for High Quality Multi-scale Image Segmentation. In *Angewandte Geographische Informationsverarbeitung XII, Beiträge zum AGIT-Symposium Salzburg 2000*; Herbert Wichmann Verlag: Heidelberg, Germany, 2000.

15. Höfle, B.; Hollaus, M.; Lehner, H.; Pfeifer, N.; Wagner, W. Area-based Parameterization of Forest Structure Using Full-waveform Airborne Laser Scanning Data. In *Proceedings of the 8th International Conference on LiDAR Applications in Forest Assessment and Inventory (SilviLaser 2008)*, Edinburgh, UK, 17–19 September 2008; pp. 227–236.
16. Solberg, S.; Næsset, E.; Bollandsås, O.M. Single tree segmentation using airborne laser scanner data in a structurally heterogeneous spruce forest. *Photogramm. Eng. Remote Sensing* **2006**, *72*, 1369–1378.
17. Hollaus, M.; Dorigo, W.; Wagner, W.; Schadauer, K.; Höfle, B.; Maier, B. Operational wide-area stem volume estimation based on airborne laser scanning and national forest inventory data. *Int. J. Remote Sens.* **2009**, *30*, 5159–5175.
18. Nyström, M.; Holmgren, J.; Olsson, H. Change Detection of Mountain Vegetation Using Multi-temporal Als Point Clouds. In *Proceedings of the SilviLaser 2011*, Hobart, Australia, 16–20 October, 2011.
19. Brandel, G. *Volymfunktioner för Enskilda Träd: Tall, Gran Och Björk [Volume Functions for Individual Trees: Scots Pine (Pinus Sylvestris), Norway Spruce (Picea Abies) and Birch (Betula Pendula & Betula Pubescens)]*; Department of Forest Yield Research, Swedish University of Agricultural Sciences: Garpenberg, Sweden, 1990; p. 183.
20. Hagberg, E.; Matérn, B. *Tabeller för Kubering av ek Och Bok* (in Swedish); Skogshögskolan, Inst. f. skoglig matematisk statistik: Stockholm, Sweden, 1975; p. 118.
21. Axelsson, P. DEM Generation from Laser Scanner Data Using Adaptive Tin Models. In *Proceedings of XIXth ISPRS Congress 2000*, Amsterdam, The Netherlands, 16–23 July 2000; In *IAPRS*; 2000; Volume 33, pp.110–117.
22. Soininen, A. *Terra Scan for Microstation, User's Guide*; Terrasolid Ltd.: Jyvaskyla, Finland, 2004; Volume 132.
23. Baskerville, G.L. Use of logarithmic regression in the estimation of plant biomass. *Can. J. For. Res.* **1972**, *2*, 49–53.
24. Beauchamp, J.J.; Olson, J.S. Corrections for bias in regression estimates after logarithmic transformation. *Ecology (Washington DC)* **1973**, *54*, 1402–1407.
25. Holm, S. *Transformationer av en Eller Flera Beroende Variabler i Regressionsanalys (in Swedish)*; Swedish University of Agricultural Sciences: Stockholm, Sweden, 1977; p. 21.
26. Næsset, E. Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. *Scand. J. For. Res.* **2004**, *19*, 164–179.
27. Maltamo, M.; Suvanto, A.; Packalén, P. Comparison of basal area and stem frequency diameter distribution modelling using airborne laser scanner data and calibration estimation. *For. Ecol. Manage.* **2007**, *247*, 26–34.
28. Gobakken, T.; Næsset, E. Estimation of diameter and basal area distributions in coniferous forest by means of airborne laser scanner data. *Scand. J. For. Res.* **2004**, *19*, 529–542.

© 2012 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).