



This is an author produced version of a paper published in
Remote Sensing of Environment.

This paper has been peer-reviewed and is proof-corrected, but does not
include the journal pagination.

Citation for the published paper:

Mattias Nyström, Johan Holmgren, Håkan Olsson. (2012) Prediction of tree
biomass in the forest–tundra ecotone using airborne laser scanning. *Remote
Sensing of Environment*. Volume: 123, Number: August, pp 271-279.
<http://dx.doi.org/10.1016/j.rse.2012.03.008>.

Access to the published version may require journal subscription.

Published with permission from: Elsevier.

Standard set statement from the publisher:

NOTICE: this is the author's version of a work that was accepted for publication in
Remote Sensing of Environment. Changes resulting from the publishing process, such as
peer review, editing, corrections, structural formatting, and other quality control
mechanisms may not be reflected in this document. Changes may have been made to this
work since it was submitted for publication. A definitive version was subsequently
published in Remote Sensing of Environment, 123, August, 10.1016/j.rse.2012.03.008

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

Prediction of tree biomass in the forest-tundra ecotone using airborne laser scanning

Mattias Nyström*, Johan Holmgren, and Håkan Olsson

Section of Forest Remote Sensing, Department of Forest Resource Management,
Swedish University of Agricultural Sciences, Umeå, Sweden

Corresponding author

Name: Mattias Nyström
E-mail: mattias.nystrom@slu.se
Address: Swedish University of Agricultural Sciences
Department of Forest Resource Management
SE-90183 Umeå, Sweden
Telephone: +46 90 786 83 16
Cell phone: +46 72 201 80 96
Fax: +46 90 77 81 16

Abstract

The effect of ongoing climate change on sub-arctic and alpine forests has led to increased interest in monitoring potential changes in the forest-tundra ecotone. In addition to climate change, insect damage, browsing pressure by herbivores such as moose and reindeer, as well as anthropogenic impacts will contribute to changes in the forest-tundra ecotone. These changes are difficult to monitor with manual methods because of the complex mosaic pattern of the ecotone. In this study, the possibility to predict maximum tree height, above ground tree biomass and canopy cover with airborne laser scanning (ALS) was therefore tested at a forest-tundra ecotone site near Abisko in northern Sweden (Lat. N 68°20', Long. E 19°01', 420-700 m a.s.l.). The forest in the area is dominated by mountain birch (*Betula pubescens* ssp. *czerepanovii*), which has highly irregular stem and canopy forms. Predictions from two different laser data acquisitions were compared. The first laser data set had 6.1 points m⁻² and was obtained in 2008 with a TopEye MKII scanner carried by a helicopter flown at 500 m a.g.l. The second laser data set had 1.4 points m⁻² and was obtained in 2010 with an Optech ALTM Gemini scanner carried by a fixed-wing aircraft flown at 1740 m a.g.l. Linear regression models were developed for the predictions using data from 73 sample plots with ten meter radius surveyed in 2009 and 2010. The relative RMSEs obtained for the TopEye and Optech data after leave-one-out cross-validation were, respectively, 8.8% and 9.5% for maximum tree height; 18.7% and 21.2% for above ground tree biomass; and, 16.8% and 18.7% for vertical canopy cover on plot level. The results were clearly improved by introducing a new procedure to compensate for unevenly distributed laser points. In conclusion, ALS has strong potential as a data source to map mountain birch biomass in the forest-tundra ecotone, even when using sparse point density ALS data.

1 Introduction

The sub-arctic and alpine tree line in northern Scandinavia is most often made up of birch forest dominated by mountain birch (*Betula pubescens* ssp. *czerepanovii*). These forests are characterized by low, often leaning and crooked trees with umbrella-like canopies and slow growth (Fig. 1). The trees have limited commercial value and systematic monitoring of them has not been done in the past (Tømmervik et al. 2009).

There is a growing interest in monitoring this area, especially the tree line, which is expected to change with a warmer climate. Kullman (1998, 2010a, b) predicts that the tree line will advance to higher elevations. However, some studies also indicate that the tree line may retreat (e.g., Crawford et al. 2003). A large change in the tree line would affect animals and plants, as well as existing land uses like reindeer herding and tourism. In addition to climate, the tree line in this area is also influenced by other factors (Van Bogaert 2010), such as browsing by moose and reindeer, human activities, insect attacks, and possibly bog expansion (Crawford et al. 2003). Of particular significance is the autumnal moth (*Epirrita autumnata*) which defoliates large areas of sub-arctic forests on cyclic intervals of about a decade (Babst et al. 2010; Tenow et al. 2007). These factors, and in particular the local climate, will contribute to an ecotone between birch forest and alpine heath that has a very complex mosaic pattern (Payette et al. 2001). In addition to the tree line changes, an increase in the biomass of the forest-tundra ecotone can be expected (Tømmervik et al. 2009; Hedenås et al. 2011). Such ecosystem changes over large areas in the north would also influence global carbon budgets (Sjögersten and Wookey 2009; Tømmervik et al. 2009).

Several definitions of tree line exist (Van Bogaert 2010) which make comparisons between different studies problematic. Furthermore, the scattered patches of trees make it difficult to use ground based inventories to define a line separating areas with trees from alpine heath. A more convenient approach is to use complete predictions of areas based on remote sensing data. Mapping the tree line using manual interpretation of aerial photos is impractical and dependent on the interpreter (Heiskanen et al. 2008). Medium resolution optical satellite data (e.g., SPOT, Landsat) could be used for large area overviews but will not provide a detailed assessment of the structure, biomass and height of the forest-tundra ecotone (Heiskanen 2006a, b; Hill et al. 2007; Ranson et al. 2004; Weiss and Walsh 2009; Zhang et al. 2009). Laser scanning provides the possibility to efficiently retrieve 3D data of the forest, including height and canopy cover, which are the two parameters used in the FAO definition of forest (FAO 2004). So far only a few other laser scanning studies have been done for the forest-tundra ecotone. Ørka et al. (2012) successfully integrated strip samples of LiDAR with Landsat imagery to delineate the subalpine zone. Næsset (2009), Næsset and Nelson (2007), as well as Rees (2007), showed that scattered trees on the alpine heath above the forest could be detected by laser scanning. Earlier studies have shown that laser scanning works well for predicting forest features in managed boreal forests in Scandinavia when combined with field measurements on sample plots (Næsset et al. 2004).

Until now, laser data have only been available for small areas (often densely scanned), but recently several countries, including Sweden, are being laser scanned with sparse point density, mainly for the purpose of constructing new national digital elevation models. It is therefore of interest to compare biomass predictions from a sparse national laser scanning (about 1 point m⁻²) to denser alternatives (>5 points m⁻²).

This study investigated the prediction accuracy of above ground tree biomass, maximum tree height, and canopy cover of mountain birch forest in the forest-tundra ecotone when using models developed from field surveyed sample plot data and corresponding laser data. Results from two

different laser data acquisitions with different point densities, flying altitudes and scanner types were compared on sample plot level. The first scanning was performed from a helicopter with 6.1 points m⁻² and the second from fixed-wing aircraft with 1.4 points m⁻². For both laser data acquisitions, the developed models for observed above ground tree biomass were applied to 10 × 10 m grid cells over the whole area.

2 Materials and methods

2.1 Study area

The study area is about 1 km² and located 6 km southeast of Abisko in northern Sweden, centered on Lat. N 68°20', Long. E 19°01' (Fig. 2), with elevations from 420 to 700 m a.s.l. The area is characterized by a mosaic pattern of forest and alpine heath vegetation. The vegetation is predominantly mountain birch (*Betula pubescens* ssp. *czerepanovii*), but Juniper (*Juniperus communis*), Rowan (*Sorbus aucuparia*), and Willow (*Salix* spp.) taller than 2 m are also present. The birches in the study area are of the multi-stem type with several stems often sharing the same root system (polymorphism).

2.2 Laser data acquisitions

Laser data were acquired under leaf-on conditions at two occasions with different laser scanners (Table 1). The first scanning was done on August 1, 2008 with a TopEye MkII (denoted TopEye) mounted on a helicopter, and the second scanning was done on August 20, 2010 with an Optech ALTM Gemini (denoted Optech), mounted within a fixed-wing aircraft. In both cases, scanning was performed with the flying direction orthogonal to the main slope.

2.3 Field data

The field inventory was carried out over four weeks in August 2009 (88 sample plots) and two weeks in June 2010 (16 sample plots), occurring between the two laser data acquisitions. A total of 104 sample plots were placed systematically in a grid with 100 m spacing, covering a 1.3 × 1.3 km area in the forest-tundra ecotone, and each plot center was permanently marked. The grid was aligned orthogonal to the main slope direction with a randomly selected starting point. The aim was to represent the whole transition zone between closed forest and heath, thus 31 of the sample plots had no trees taller than 2 m. Table 2 gives a summary of the 73 forested sample plots used in this study. No trees were found to be recently broken in 2009, but in 2010 four trees were found broken on the sample plots that were field inventoried in 2009. These trees were therefore removed from the field data used with the Optech laser data.

Navigation to the sample plots was made with a handheld GPS. In order to avoid subjective selection of sample plot centers, GPS navigation was terminated when 10 m remained to the sample plot center and a measuring tape was used to navigate the last 10 m. The center position of the sample plot was then measured with sub-dm accuracy using a Real Time Kinematic (RTK) GPS (Trimble R7) with the base station placed 6 km away at the Abisko Scientific Research Station.

The sample plots had a radius of 10 m and the distance to the plot border was measured with an ultrasonic positioning system (UPS; Lämås 2010). The UPS could be used either with one or three ultrasonic senders. One sender was sufficient to measure distance. To measure a coordinate inside

the sample plot, three senders were placed in a triangle, 2 m apart, on a tripod in the center of the sample plot. The receiver's position could then be calculated in two dimensions.

In the field, all distances were measured parallel to the ground because this was most convenient given the hilly terrain. To achieve the correct horizontal area, a correction factor was calculated and applied for each sample plot when processing the field data. For each sample plot a plane was fitted to ground-classified laser points from the TopEye data using the least squares method. The area of a sample plot was then calculated as the horizontal projection of the fitted plane. This projected area was on average 3% and up to 12% smaller than the area of a ten meter radius plot.

When obtaining coordinates for the tree stems, trees with a common root system or overlapping crowns were grouped. For each group, position was measured with UPS and maximum tree height was measured with a measuring pole. In this study, trees are defined as having a minimum height of 2 m. DBH was measured for trees by callipering at 1.3 m from the ground along the stem. In groups with at least one tree having DBH greater than 30 mm, trees with DBH less than 30 mm were counted rather than callipered. If all trees in a group were under 30 mm DBH, the tallest tree in the group was callipered.

A measuring tape was used to measure the width and length of an ellipse describing the projected area of the canopy for the whole tree group. The width of the ellipse was always measured perpendicular to the sample plot's center. Canopy area (CA) was calculated as the ellipse area for each tree group. All tree groups were measured on the first 22 sample plots, but due to time limits, measuring of CA was reduced to every fifth tree group for the remaining sample plots. To predict CA for tree groups that were not measured, k Nearest Neighbor imputation (kNN) was used through the package `yaImpute` (Crookston and Finley 2008) in the statistical software R (R Development Core Team, 2010). The `RandomForest` package in R (Liaw and Wiener 2002) was used as a method to find neighbors using the following predictors: tree cross-sectional area, height, number of trees in the group, and number of trees less than 30 mm DBH in the group. In total, predictors from 488 tree groups were used. The number of nearest neighbors (k) was set to three, because it resulted in the smallest leave-one-out cross-validated RMSE when varying k from one to ten. A distance weight was used to calculate imputed values. Two extreme outliers and four trees with DBH equal to zero were removed from the data set. The total vertical canopy cover (VCC) for a sample plot was calculated as the sum of the area of all tree groups' ellipses (sum of CA) divided by the horizontal sample plot area.

Models to predict above ground tree biomass (AGB , denoted biomass) from measures of the tree stems exist from a previous study in the area (Dahlberg et al. 2004). Two of these biomass models were used: one model using DBH and tree height as explanatory variables, and the other model using only DBH as an explanatory variable. The choice of model was dependent on whether tree height had been measured in the field or not. The trees counted in each tree group taller than 2 m but less than 30 mm DBH were each assigned a biomass of 1.21 kg. The assigned biomass was calculated as the mean biomass of 251 trees fulfilling the conditions taller than 2 m and DBH less than 30 mm. Models from Dahlberg et al. (2004) were used to calculate biomass for these trees using DBH and height measurements.

2.4 Laser data processing

The TopEye laser data had a large overlap between the flight lines and therefore several sample plots were covered with laser data from two flight lines. To reduce problems with large scanning

angles (Holmgren et al. 2003), laser data for each sample plot were selected only from the flight line with the lowest mean scan angle. The Optech laser data were collected from a higher altitude and therefore a wider area was covered in each flight line. This resulted in only three flight lines needed to cover all sample plots in comparison to the TopEye scanning where seven flight lines were needed. In the sparse Optech laser data, removal of overlapping data was not undertaken because some sample plots (which had sparse cover and short trees) in the overlapping region had no hits at all in the vegetation if only one flight line was used. The point densities in Table 1 are based only on laser data used in the development of the models.

TerraScan (Soininen 2010) from Terrasolid was used to classify the ALS point cloud into ground and non-ground points (Axelsson 1999, 2000). The TerraScan software iteratively builds a triangulated surface model by selecting points from the ALS point cloud if they satisfy user specified settings. The settings used were (1) maximum terrain angle, i.e., steepest allowed slope in ground terrain, set to 88 degrees, (2) iteration angle, i.e., maximum angle between the point, its projection on the triangulated plane and closest triangle vertex, set to 6°, and (3) iteration distance, i.e., maximum distance from the point to the triangulated plane during iteration, set to 1.40 m. The statistical software R (R Development Core Team, 2010) and in-house developed programs were used to further process the ALS point cloud.

A digital elevation model (DEM) with 0.5 m grid cell size, representing the ground level, was created for each data set. Canopy heights (CH) were calculated by subtracting the DEM from the z-value of each laser return. A digital surface model (DSM) with 0.5 m grid cell size was created by assigning each grid cell the maximum z-value for laser returns classified as non-ground. A normalized digital surface model (nDSM) was calculated by subtracting the DEM from the DSM. Metrics used in this paper were created only from CH. All data points above 15 m were omitted in the CH to avoid false reflections (the 15 m limit was chosen given that no trees taller than 10.5 m were found in the area).

A height threshold (Nilsson 1996) of 1.0 m was used when calculating metrics from the ALS point cloud above ground. The threshold was chosen to obtain reliable laser metrics for sample plots with trees around two meter tall. The height threshold was varied from 0 to 2.0 m in increments of 0.1 m. A threshold around 1 m gave the smallest root mean square error (RMSE) for the biomass prediction.

Height percentiles (H_{xx}) in increments of 10 were calculated using canopy heights above the height threshold. In addition, height percentiles were also calculated for H_{95} and H_{99} since these are strong indicators of the height of vegetation. Ten vertical canopy layers (D_x) were estimated following the example of Næsset and Gobakken (2008), where the height threshold is used as the lower limit and H_{95} as the upper limit. Standard deviation (H_{sd}) and mean (H_{mean}) of the canopy heights were calculated on values above the height threshold. A sum of squared canopy heights (H_{sum}) was calculated by taking the sum of squared laser points' heights divided by the number of laser points above the height threshold (Eq. (1)).

$$H_{sum} = \frac{\sum_{i=1}^n h_i^2}{n} \quad (1)$$

where n is all laser points within the sample plot above the height threshold.

Several ways to calculate the vegetation ratio (VR) from CH were considered. The common approach is based on dividing the number of hits above the height threshold with the total number of returns inside the sample plot. However, this approach may not be well suited to the data because

the laser points were not evenly spread on the sample plots due to roll and pitch of the aircraft during scanning. This was particularly evident in the TopEye scanning (Fig. 5). Vegetation ratio was calculated in three ways, either all returns were used (denoted VR^{all}), only first returns were used (denoted VR^{1st}), or it was calculated in an alternative way developed in this study. In the alternative way, the points were assigned a weight according to the density of surrounding points, in other words, points in areas containing more laser points were given lower weight. This alternative is called a point-weighted vegetation ratio (VR_{pw}), which was calculated as

$$VR_{pw}(r) = \frac{\sum_{i=1}^{n_{veg>1m}} p_i(r)}{\sum_{i=1}^{n_{all}} p_i(r)} \quad (2)$$

where $n_{veg>1m}$ are all vegetation classified points above 1 m and n_{all} are all points inside the sample plot. $p_i(r)$ is the weight of point i and calculated as

$$p_i(r) = \frac{d_{sp}}{d_i(r)} \quad (3)$$

where d_{sp} is the average point density on the sample plot and determined by

$$d_{sp} = \frac{n_{sp}}{A_{sp}} \quad (4)$$

where n_{sp} is the total number of points inside the sample plot and A_{sp} is the area of the sample plot. The density of all points in a radius r around the point i is denoted by $d_i(r)$, such that

$$d_i(r) = \frac{n_i(r)}{A(r)} \quad (5)$$

where $n_i(r)$ is the total number of points inside a cylinder parallel with the z-axis having radius r centered on the point i , and $A(r)$ is the area of a circle with radius r . If the circle with radius r falls partly outside of the sample plot's border, then $A(r)$ is set to the area that falls inside the sample plot. The radius r was varied from 0.01 to 5 m in 0.01 m increments. For each step, the point-weighted vegetation ratio (VR_{pw}) was calculated and models for biomass and vertical canopy cover were estimated. For each data set, the radius (r) resulting in the lowest RMSE for the two models was chosen.

2.5 Model development

Regression models were developed based on data from the sample plots to predict the following response variables: maximum tree height (MH), above ground tree biomass (AGB) and vertical canopy cover (VCC). The regression models were based only on field data from trees taller than 2 m. Candidates for explanatory variables (in total 29 variables calculated from CH) were height percentiles (H_{xx}), vegetation ratios (VR^{all} , VR^{1st} , VR_{pw}^{all} , VR_{pw}^{1st}), standard deviation of canopy heights (H_{sd}), mean canopy heights (H_{mean}), sum of squared canopy heights (H_{sum}), and height densities (D_x).

Robust models were required to support the objective of making predictions in the forest-tundra ecotone. For each response variable, the physical properties and robustness were analyzed, and therefore the following criteria for selecting explanatory variables were used:

- For maximum tree height – one explanatory variable preferably measuring the higher part of the canopies and eventually one explanatory variable measuring the density of vegetation. The reason for the latter variable is that percentiles could depend on vegetation density.

- For biomass – one explanatory variable measuring the height and one explanatory variable measuring the density of vegetation.
- For vertical canopy cover – one explanatory variable measuring the density of vegetation.

By using these rules, explanatory variables were selected using the leaps function (Lumley and Miller, 2009) in R, which is a best subset regression. Variance inflation factor (VIF) analysis was used to check multicollinearity and residuals were checked. Logarithmic transformation was tried on both response and explanatory variables; the model without logarithmic transformations was selected if the resulting RMSE was not improved and the residuals were well distributed.

The back-transformed response variables were corrected for bias (caused by the logarithmic transformation) by multiplying with a correction coefficient (Holm 1977):

$$q = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n \hat{y}_i} \quad (6)$$

where y_i is the field measured value and \hat{y}_i the predicted value on sample plot i . Root mean square error (RMSE), relative RMSE ($RMSE_r$), bias (Bias), and relative bias ($Bias_r$), given in Equations (7) – (10), were calculated using leave-one-out cross-validation on sample plot level:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (7)$$

$$RMSE_r = \frac{RMSE}{\bar{y}} \cdot 100 \quad (8)$$

$$Bias = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)}{n} \quad (9)$$

$$Bias_r = \frac{Bias}{\bar{y}} \cdot 100 \quad (10)$$

where \bar{y} is the mean of the field measured values and n is the number of observations.

2.6 Raster maps

Raster maps with a pixel size of 10×10 m were calculated from the estimated models for maximum tree height, biomass and vertical canopy cover. Each pixel's value was calculated from laser data based on a ten meter radius circular area, centered on the pixel. If less than five laser points within the ten meter radius circle were above the height threshold (1 m), the value of the pixel was set to zero. The models for biomass and vertical canopy cover were applied only when maximum tree height was predicted to be at least 2 m, since the models were only valid for trees taller than 2 m.

3 Results

Multiple linear regression was successfully used to estimate models to predict above ground tree biomass, maximum tree height and vertical canopy cover from ALS data in the forest-tundra ecotone. The results showed only a small improvement in the predictions when using the denser TopEye data with 6.1 points m⁻² compared to the sparse Optech data with 1.4 points m⁻² (Table 3).

Metrics used in this paper were created only from CH, as no improvement was achieved when metrics were calculated from the nDSM. For the biomass and the vertical canopy cover model, both response variables and explanatory variables were transformed by taking the natural logarithm. The model used to predict maximum tree height included both a percentile and the vegetation ratio.

The models on sample plot level for prediction of maximum tree height (*MH*), above ground tree biomass (*AGB*), and vertical canopy cover (*VCC*) were different in the two laser data sets.

Equations (11) - (13) describe the models used with data from the TopEye scanning.

$$MH = 1.09 + 0.89 \cdot H_{95} + 3.26 \cdot VR_{pw}^{all} \quad (11)$$

$$\ln(AGB) = 2.52 + 0.42 \cdot \ln(H_{sum}) + 0.87 \cdot \ln(VR_{pw}^{1st}) \quad (12)$$

$$\ln(VCC) = 0.32 + 0.88 \cdot \ln(VR_{pw}^{all}) \quad (13)$$

Equations (14) - (16) describe *MH*, *AGB* and *VCC* for the Optech scanning.

$$MH = 1.91 + 0.68 \cdot H_{99} + 3.61 \cdot VR_{pw}^{all} \quad (14)$$

$$\ln(AGB) = 2.56 + 0.42 \cdot \ln(H_{sum}) + 0.81 \cdot \ln(VR_{pw}^{1st}) \quad (15)$$

$$\ln(VCC) = 0.29 + 0.85 \cdot \ln(VR_{pw}^{all}) \quad (16)$$

The back-transformed values of Equations (12), (13), (15), and (16) were corrected with the logarithmic bias correction coefficient (Eq. (6) and Table 3).

Table 3 shows the significance of the selected explanatory variables and the accuracy statistics for the developed models. The point-weighted vegetation ratio (Eq. (2)) was used on the TopEye laser data with a radius of 0.59 m (Eq. (5)) which improved the relative RMSE for prediction of biomass and vertical canopy cover from 19.9% to 18.7% and from 19.1% to 16.8%, respectively. In the Optech data the improvement in RMSE was somewhat less using a radius of 1.00 m (Eq. (5)). For biomass, the relative RMSE improved from 22.1% to 21.2% and vertical canopy cover improved from 19.9% to 18.7%. There was also an improvement of the prediction of maximum tree height, but this was almost negligible because the influence of the vegetation ratio in the model was very small. A few outliers could be observed for all predicted variables (Figs. 3 and 4). Figure 5 shows a sample plot from the TopEye data with an uneven distribution of laser points.

The raster map with predicted values of biomass (Fig. 6) from TopEye laser data shows the mosaic pattern in the forest-tundra ecotone. This map is extrapolated outside the area calibrated with sample plots. The same map was created for the Optech laser data and there were only small differences in the predictions.

4 Discussion

In this study, maximum tree height, above ground tree biomass, and vertical canopy cover for sub-arctic mountain birch forest were successfully predicted from the ALS point cloud data. This is, to our knowledge, the first time that ALS-based biomass predictions for this type of forest have been reported, since earlier ALS studies of mountain birch concentrated on the existence of forested areas (Rees 2007), and the existence of pioneer trees (Næsset 2009; Næsset and Nelson 2007).

To monitor the wide spectrum of this forest type, from short trees (~2 m) up to tall trees (~10 m) often with sparse canopy cover and irregular branching, requires models that will not fail in areas with low and/or sparse vegetation. Therefore the models were developed to include few yet highly significant variables. Only higher height percentiles (95th and 99th) were used to avoid problems with sample plots with low and/or sparse vegetation. The vegetation ratio was included in the model for maximum tree height because without it, it was impossible to obtain good predictions in both low and high vegetations.

Næsset & Bjercknes (2001) used regression analysis with a stepwise selection procedure to predict dominant height of young boreal coniferous forest with average heights similar to those in our study. The final model included both the 90th height percentile and the canopy density derived from laser data with approximately 0.9 m between footprints. The standard deviation of the difference between predicted and observed dominant height was 0.57 m on plot level, which is similar to the RMSE of 0.60 m obtained for maximum tree height in our study. Næsset (2011) predicted above ground biomass in young forest in south-eastern Norway with metrics derived from ALS data having a point density of 0.9 to 1.2 m⁻². In Næsset's study, the mean tree height ranged from 1.77 to 9.76 m and the plot size was 200 to 232.9 m². The relative RMSE derived using cross-validation ranged from 28.2% to 48.1% for plot level predictions of biomass, which was less accurate than the results obtained in our study. Næsset tested different height thresholds (0.5, 1.3, and 2.0 m) but no clear effect could be observed on prediction accuracy, in contrast to the findings of our study. However, tree species had a significant influence on the models, something we not could observe because of the dominance of birch trees. Given the irregular shape and variable distribution of mountain birches, our results of 18.7% and 21.2% relative RMSE (TopEye and Optech, respectively) for the biomass predictions are surprisingly good. One positive factor in this study compared to several earlier studies in boreal forest is that we were able to locate the sample plots with sub-dm accuracy due to the use of RTK-GPS technology. The RTK-GPS technology is normally difficult to use in boreal forest because the tree canopies disturb the GPS-signal. However, the conditions are better in the sub-arctic forest where the trees are low-growing and have sparse canopies.

Korhonen et al. (2011) predicted vertical canopy cover for 30 plots in a boreal forest in southern Finland using the proportion of vegetation returns in first return airborne laser data and obtained highly accurate predictions with a relative RMSE of 3.7% to 7.0%. They used a grid approach proposed by Vauhkonen et al. (2008) to correct for effects of uneven distribution of laser points: the data were decimated by allocating each laser return to a grid cell based on its XY coordinate and a random return was selected for each grid cell. The predictions of *VCC* were improved by testing this grid approach with different grid cell sizes. In our study we propose a new method to correct for uneven distribution of laser points. One advantage of our proposed method is that the data set does not need to be decimated but instead the points are assigned a weight. This method of weighting the points improved all predictions. In our study, the height percentiles were not corrected for the uneven distribution of laser points. Therefore, the improvement of biomass prediction was not as large as for vertical canopy cover. There are several reasons why we not could obtain the same high accuracy of *VCC* predictions as was obtained by Korhonen et al. (2011). In our study, crown diameter was only measured for a sample of trees in the field inventory and predicted for the remaining trees. The RMSE for this prediction was 3.9 m² and relative RMSE was 54% using leave-one-out cross-validation on tree group level. Using the 14 sample plots with all tree groups measured, relative RMSE on sample plot level was 17%. It was also difficult to describe the canopy area of the irregularly shaped tree crowns using an ellipse. For future studies, another

field method should be considered, for example, photographing the crowns from ground or terrestrial laser scanning.

Næsset and Gobakken (2008) predicted above- and below-ground biomass in young and mature coniferous forest in Norway. They used similar techniques to reduce the number of variables and for the above-ground biomass they selected the 90th height percentile and height density 0 (i.e., the proportion of points above 2 m to total number of points). In this study we selected the normalized squared sum of vegetation hits above 1 m and the proportion of points above 1 m to total number of points.

In the maximum tree height model, logarithmic variables were not used since no improvement in RMSE was achieved and the residuals were better distributed without logarithmic variables. In comparison to the other parameters' models, the maximum tree height models had the lowest RMSE (8.8% for TopEye and 9.5% for Optech). This can be explained by the fact that heights are directly measured both by the laser scanner and in the field. The biomass predictions were based on the correlation between laser metrics and an allometric biomass model driven by field surveyed stem diameter and height measurements. These allometric relationships are not as well behaved for mountain birch forest as for normal boreal forest (Dahlberg et al. 2004). Another problem with the biomass function is how the tree cross-sectional area is measured. In this study, a tree trunk splitting off below 1.3 m was measured as two separate trees, but in Dahlberg et al. (2004) the sum of cross-sectional areas was measured. The field material in our study, however, included only a few percent of trees whose trunks split off below 1.3 m.

Visual inspection revealed that TopEye laser data points were more unevenly distributed than the Optech data. This was most likely the reason why the point weight correction had a larger improvement on the TopEye data. There can be many reasons why the TopEye data were more unevenly distributed than the Optech data, such as acquisition from different platforms (helicopter vs. fixed-wing aircraft), weather conditions and scanning patterns. Further research on methods compensating for unevenly distributed laser points is needed.

The field inventory method should be further developed before applied in operational use. In this study, trees taller than 2 m and DBH smaller than 30 mm were only counted in the field inventory. For these trees, the biomass had to be derived based solely on an average of sample trees. Another source of error could be a few standing dead trees that were not considered but could affect the laser metrics.

The grid cell size of the DEMs varied from 0.25 to 2 m in 0.25 m increments, with the smaller grid cell sizes resulting in lower RMSEs (for biomass). The 0.5 m grid cell size was then chosen to calculate the mean from several points in each grid cell. In the low density laser data (Optech), the 0.5 m grid was also used to enable direct comparisons. Since TIN was used to interpolate values where no laser points existed, there was no problem in using a smaller grid cell size than the distance between laser points. In the Optech data the average point density was 1.4 m⁻², which means there was always interpolation between grid cells. For the Optech scanning, using a smaller grid cell size also resulted in lower RMSE (for biomass).

The foot-print was about 0.5 m with both laser scanners regardless of the difference in flying altitude. This causes the Optech scanner with low point density to omit vegetation because the ground is not fully illuminated.

In June 2010 all sample plots were revisited with the knowledge of the predictions from the TopEye laser data. Sample plots with a large difference between field measured and laser predicted biomass, maximum tree height, and vertical canopy cover could then be studied to find an explanation for the error (Figs. 3 and 4). Some cases of overestimation of biomass and maximum tree height were observed. This was most likely due to tall trees occurring on the border of the sample plot, in some cases leaning inside the sample plot. By visual inspection of the ALS point cloud it could be verified that the highest laser points were located close to the border of the sample plot.

When creating the biomass map, each 10×10 m pixel was assigned a value calculated from data within a ten m radius circle centered on the pixel. The reason for using circles was to use exactly the same area and shape as the field sample plots. The pixel size was chosen to be small enough to depict the spatial pattern of the vegetation. Before using the raster map for aggregation or change detection, verification of this method to assign pixel values is needed to determine whether or not it causes bias.

This study demonstrates that biomass in the sub-arctic mountain birch forest can be accurately predicted on plot level from ALS data. Predictions were made using both low and high resolution laser data and the results show that high accuracy could be achieved even when using low resolution laser data ($1.4 \text{ points m}^{-2}$). A method to correct the vegetation ratio for unevenly distributed laser points was developed which improved the predictions considerably. A high resolution raster map of predicted biomass in the forest-tundra ecotone was created (Fig. 6), useful for future monitoring of potential changes in this area.

Acknowledgments

This study is part of the research program Environmental Mapping and Monitoring with Airborne laser and digital images (EMMA), financed by the Swedish Environmental Protection Agency. The laser data sets were acquired through cooperation with the University of Lund and the new national elevation model project at the Swedish National Land Survey. The Abisko Scientific Research Station is acknowledged for supporting the work with field data acquisition. We are also thankful to Peder Axensten, Heather Reese and Darius Culvenor for discussions about algorithms and comments on the manuscript. Finally we thank the two anonymous reviewers for their constructive comments.

References

- Axelsson, P.E. (1999). Processing of laser scanner data - algorithms and applications. *Isprs Journal of Photogrammetry and Remote Sensing*, 54, 138-147.
- Axelsson, P.E. (2000). DEM Generation from Laser Scanner Data Using Adaptive Models. *International Archives of Photogrammetry & Remote Sensing*, 33, 110-117.
- Babst, F., Esper, J., & Parlow, E. (2010). Landsat TM/ETM plus and tree-ring based assessment of spatiotemporal patterns of the autumnal moth (*Epirrita autumnata*) in northernmost Fennoscandia. *Remote Sensing of Environment*, 114, 637-646.
- Crawford, R.M.M., Jeffrey, C.E. and Rees, W.G. (2003). Paludification and forest retreat in Northern Oceanic environments. *Annals of Botany* 91, 213-226.
- Crookston, N.L., & Finley, A.O. (2008). yaImpute: An R package for kNN imputation. *Journal of Statistical Software*. 23(10):1-16.

- Dahlberg, U., Berge, T.W., Petersson, H., & Vencatasawmy, C.P. (2004). Modelling biomass and leaf area index in a sub-arctic Scandinavian mountain area. *Scandinavian Journal of Forest Research*, 19, 60-71.
- FAO (2004). Global forest resources Assessment 2005 – Terms and Definitions. *FAO Working Paper 83*. Rome: Food and Agriculture Organization of the United Nations.
- Hedenås, H., Olsson, H., Jonasson, C., Bergstedt, J., Dahlberg, U., Callaghan, T.V. (2011). Changes in Tree Growth, Biomass and Vegetation Over a 13-Year Period in the Swedish Sub-Arctic. *Ambio*, 40, 672-682.
- Heiskanen, J. (2006a). Estimating aboveground tree biomass and leaf area index in a mountain birch forest using ASTER satellite data. *International Journal of Remote Sensing*, 27, 1135-1158.
- Heiskanen, J. (2006b). Tree cover and height estimation in the Fennoscandian tundra-taiga transition zone using multiangular MISR data. *Remote Sensing of Environment*, 103, 97-114.
- Heiskanen, J., Nilsson, B., Mäki, A.-H., Allard, A., Moen, J., Holm, S., Sundquist, S., & Olsson, H. (2008). Aerial photo interpretation for change detection of treeline ecotones in the Swedish mountains. Swedish University of Agricultural Sciences, Department of forest resource management and geomatics, Working report 242, 59 p.
- Hill, R.A., Granica, K., Smith, G.M., & Schardt, M. (2007). Representation of an alpine treeline ecotone in SPOT 5 HRG data. *Remote Sensing of Environment*, 110, 458-467.
- Holm, S. (1977). Transformation av en eller flera beroende variabler i regressionsanalys. Swedish University of Agricultural Sciences, Stockholm, Hugin rapport nr 7, 21 pp. (in Swedish).
- Holmgren, J., Nilsson, M., & Olsson, H. (2003). Simulating the effects of lidar scanning angle for estimation of mean tree height and canopy closure. *Canadian Journal of Remote Sensing*, 29, 623-632.
- Korhonen, L., Korpela, I., Heiskanen, J., & Maltamo, M. (2011). Airborne discrete-return LIDAR data in the estimation of vertical canopy cover, angular canopy closure and leaf area index. *Remote Sensing of Environment*, 115, 1065-1080.
- Kullman, L. (1998). Tree-limits and montane forests in the Swedish Scandes: Sensitive biomonitors of climate change and variability. *Ambio*, 27, 312-321.
- Kullman, L. (2010a). One Century of Treeline Change and Stability - Experiences from the Swedish Scandes. In, *The Official Journal of the International Association for Landscape Ecology, Chapter Germany (IALE-D)*.
- Kullman, L. (2010b). A Richer, Greener and Smaller Alpine World: Review and Projection of Warming-Induced Plant Cover Change in the Swedish Scandes. *Ambio*, 39, 159-169.
- Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2, 18-22.
- Lämås, T. (2010). The Haglöf PosTex ultrasound instrument for the positioning of objects on forest sample plots. Swedish University of Agricultural Sciences, Department of Forest Resource Management, Umeå, Sweden. Working report no 296, 12pp.
- Lumley, T. using Fortran code by Miller, A. (2009). leaps: regression subset selection. R package version 2.9. <http://CRAN.R-project.org/package=leaps>.
- Næsset, E. (2009). Influence of terrain model smoothing and flight and sensor configurations on detection of small pioneer trees in the boreal-alpine transition zone utilizing height metrics derived from airborne scanning lasers. *Remote Sensing of Environment*, 113, 2210-2223.
- Næsset, E. (2011). Estimating above-ground biomass in young forests with airborne laser scanning. *International Journal of Remote Sensing*, 32, 473-501.
- Næsset, E., & Bjerknes, K.O. (2001). Estimating tree heights and number of stems in young forest stands using airborne laser scanner data. *Remote Sensing of Environment*, 78, 328-340.
- Næsset, E., & Gobakken, T. (2008). Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sensing of Environment*, 112, 3079-3090.

- Næsset, E., Gobakken, T., Holmgren, J., Hyypä, H., Hyypä, J., Maltamo, M., Nilsson, M., Olsson, H., Persson, Å., & Söderman, U. (2004). Laser scanning of forest resources: The Nordic experience. *Scandinavian Journal of Forest Research*, *19*, 482-499.
- Ørka, H.O., Wulder, M.A., Gobakken, T., & Næsset, E. (2012). Subalpine zone delineation using LiDAR and Landsat imagery. *Remote Sensing of Environment*, *119*, 11-20.
- Næsset, E., & Nelson, R. (2007). Using airborne laser scanning to monitor tree migration in the boreal-alpine transition zone. *Remote Sensing of Environment*, *110*, 357-369.
- Nilsson, M. (1996). Estimation of tree weights and stand volume using an airborne lidar system. *Remote Sensing of Environment*, *56*, 1-7.
- Payette, S., Fortin, M.-J. and Gamache, I. (2001). The subarctic forest-tundra: the structure of a biome in a changing climate. *BioScience* *51*, 709-718.
- R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org/>.
- Ranson, K.J., Sun, G., Kharuk, V.I., & Kovacs, K. (2004). Assessing tundra-taiga boundary with multi-sensor satellite data. *Remote Sensing of Environment*, *93*, 283-295.
- Rees, W.G. (2007). Characterisation of Arctic treelines by LiDAR and multispectral imagery. *Polar Record*, *43*, 345-352.
- Sjögersten, S., & Wookey, P.A. (2009). The Impact of Climate Change on Ecosystem Carbon Dynamics at the Scandinavian Mountain Birch Forest-Tundra Heath Ecotone. *Ambio*, *38*, 2-10.
- Soininen, A. (2010). TerraScan User's Guide. TerraSolid Ltd.
- Tenow, O., Nilssen, A.C., Bylund, H., & Hogstad, O. (2007). Waves and synchrony in *Epirrita autumnata*/*Operophtera brumata* outbreaks. I. Lagged synchrony: regionally, locally and among species. *Journal of Animal Ecology*, *76*, 258-268.
- Tømmervik, H., Johansen, B., Riseth, J.Å., Karlsen, S.R., Solberg, B., & Høgda, K.A. (2009). Above ground biomass changes in the mountain birch forests and mountain heaths of Finnmarksvidda, northern Norway, in the period 1957-2006. *Forest Ecology and Management*, *257*, 244-257.
- Van Bogaert, R. (2010). Recent treeline dynamics in sub-arctic Sweden: a multi-disciplinary landscape assessment. Doctoral Dissertation, Department of Geography, Universiteit Gent.
- Vauhkonen, J., Tokola, T., Maltamo, M., & Packalen, P. (2008). Effects of pulse density on predicting characteristics of individual trees of Scandinavian commercial species using alpha shape metrics based on airborne laser scanning data. *Canadian Journal of Remote Sensing*, *34*, S441-S459.
- Weiss, D.J., & Walsh, S.J. (2009). Remote Sensing of Mountain Environments. *Geography Compass*, *3*(1), 1-21. Doi:10.1111/j.1749-8198.2008.00200.x.
- Zhang, Y.J., Xu, M., Adams, J., & Wang, X.C. (2009). Can Landsat imagery detect tree line dynamics? *International Journal of Remote Sensing*, *30*, 1327-1340.



Fig. 1. The sub-arctic birch forest in the Abisko area is characterized by low, often leaning and crooked trees with umbrella-like canopies and slow growth.

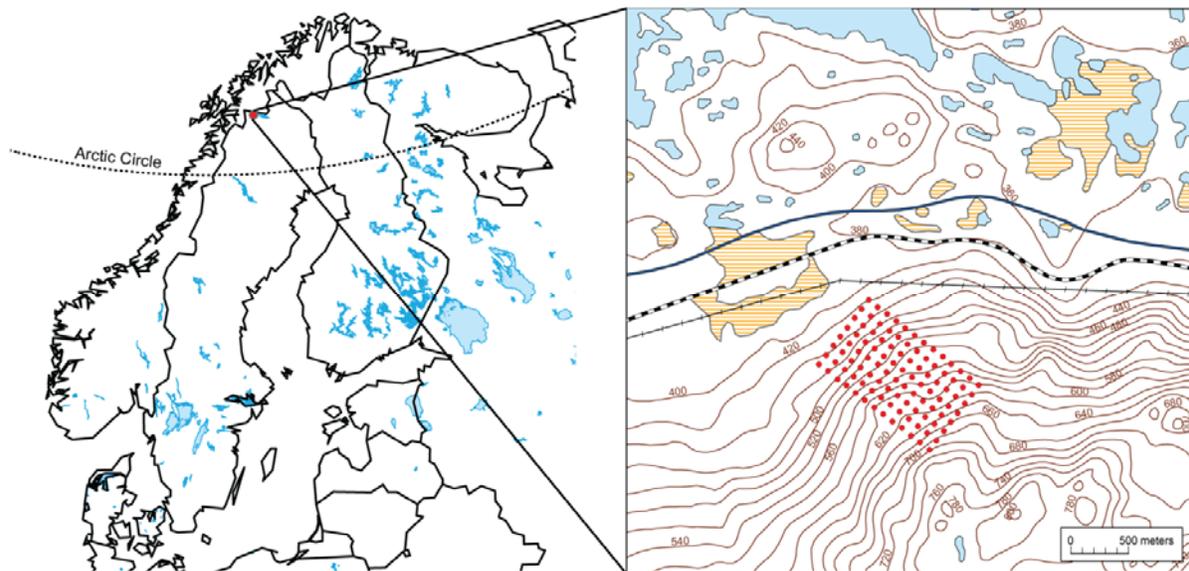


Fig. 2. Map of the Nordic countries, showing the study area in northern Sweden. The regular pattern of dots in the enlarged map is the sample plots. © Lantmäteriet, I 2010/0345.

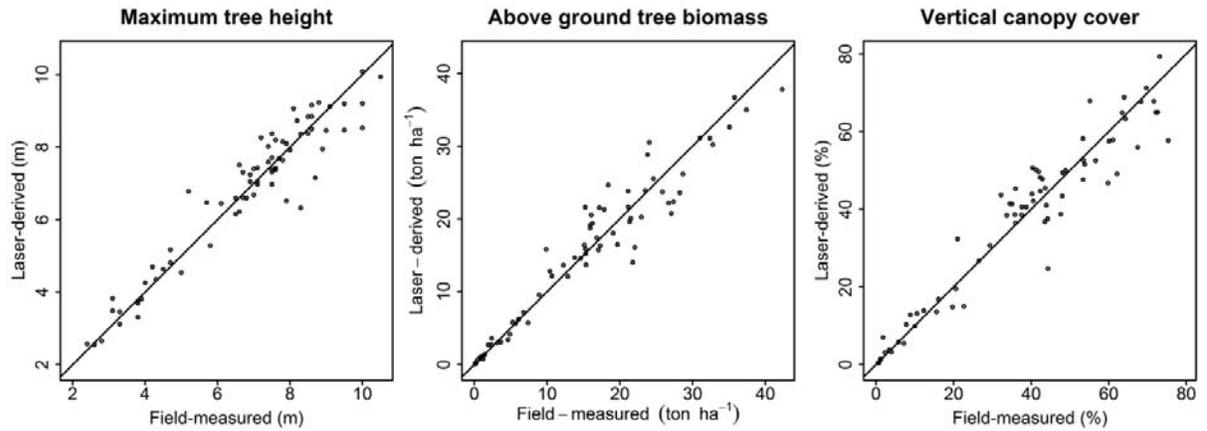


Fig. 3. Laser scanning performed in 2008 with TopEye. One-to-one plots of field based data versus laser prediction for plot level maximum tree height (left), above ground tree biomass (middle), and vertical canopy cover (right).

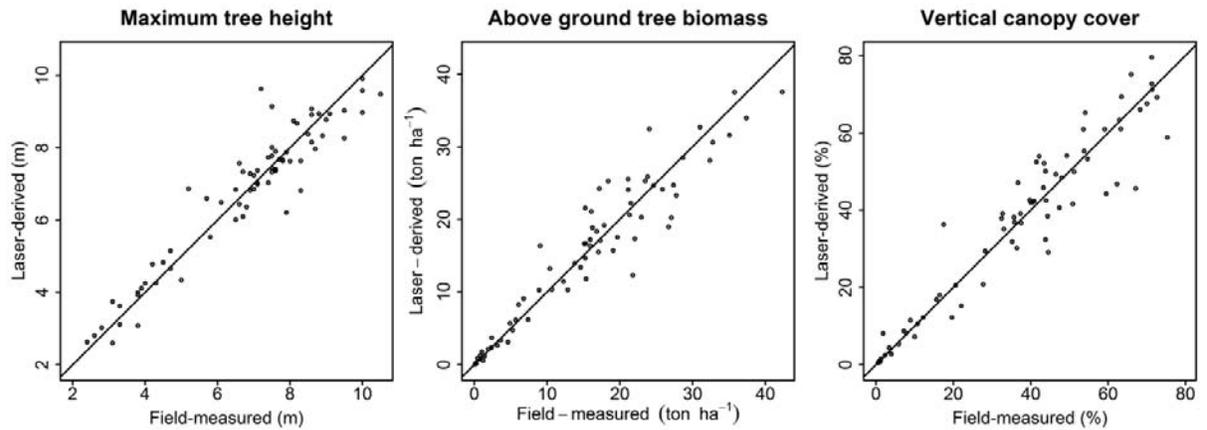


Fig. 4. Laser scanning performed in 2010 with Optech. One-to-one plots of field based data versus laser prediction for plot level maximum tree height (left), above ground tree biomass (middle), and vertical canopy cover (right).

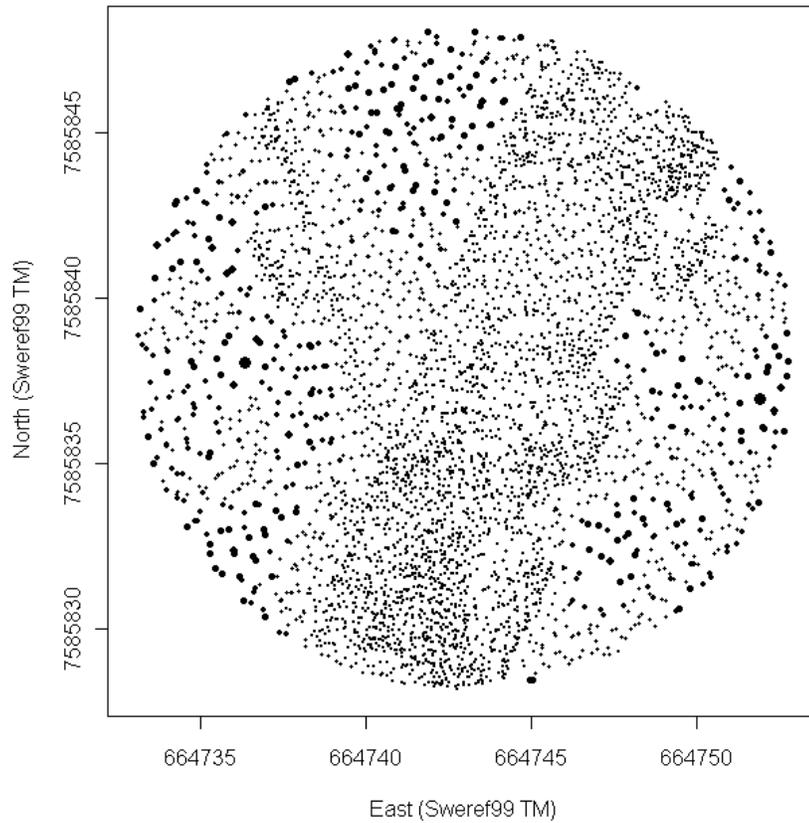


Fig. 5. Example of the distribution of laser points on a ten meter radius sample plot from the TopEye scanning. The points are unevenly distributed due to pitch and roll during scanning. The size of the dots corresponds to their calculated point weight (according to Eq. (5)). By using the point weights in the calculation of vegetation ratio, areas with high point density affect the total vegetation ratio on the sample plot in the same magnitude as points in low density areas of the sample plot. The values on the axes are the Swedish coordinate system, Sweref99 TM, with units in meters.

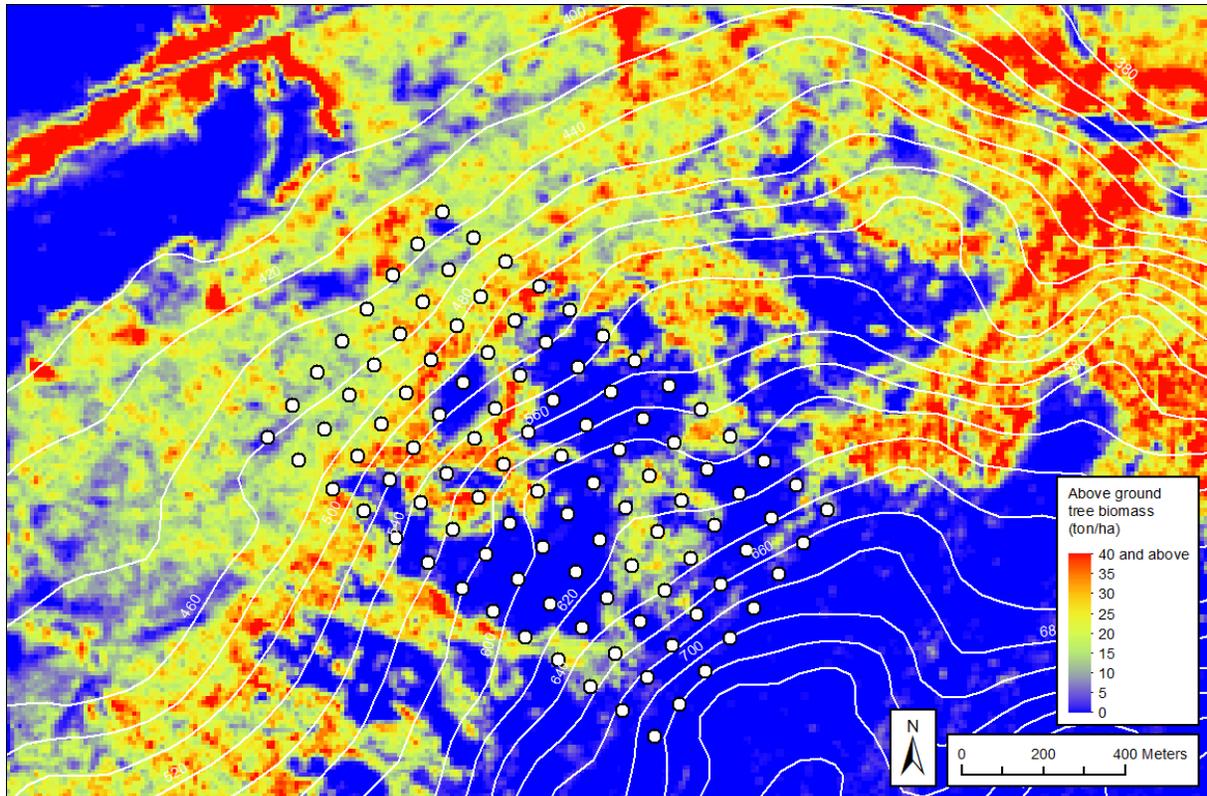


Fig. 6. Example of a raster map derived from the model (Eq. (12)) for above ground tree biomass ($t\ ha^{-1}$). The sample plots are marked with circles.

Table 1

Summary of the laser scanner's properties and flight parameters of the two laser data acquisitions.

Parameter	TopEye MkII	Optech ALTM Gemini
Flight altitude (above ground)	500 m	1740 m
Foot-print	0.5 m	0.5 m
Pulse repetition frequency	50 kHz	70 kHz
Scan frequency	35 Hz	37 Hz
Wave length	1064 nm	1064 nm
Pulse length	4 ns (1.2 m)	6.8 ns (2.0 m)
Scan type	Palmer	Oscillating mirror
Scan width (across flight dir.)	$\pm 20^\circ$	$\pm 20^\circ$
Scan width (along flight dir.)	$\pm 14^\circ$	0°
Point extraction	Up to 2 per pulse	Up to 4 per pulse
Minimum point density ^a	$2.5\ m^{-2}$	$0.7\ m^{-2}$
Average point density ^a	$6.1\ m^{-2}$	$1.4\ m^{-2}$
Maximum point density ^a	$17.3\ m^{-2}$	$2.2\ m^{-2}$

^a based on the laser data from the 73 sample plots used in development of the models.

Table 2

Summary statistics of the 73 forested sample plots (with trees taller than 2 m).

Definition	Unit	Abbr.	Min ^a	Mean	Max	N
Stem diameter ^b	Mm	DBH	4	69	266	2351 ^f
Stem density ^c	ha ⁻¹		32	1025	2165	73
Maximum tree height ^d	M	<i>MH</i>	2.4	6.8	10.5	73
Biomass ^e	t ha ⁻¹	<i>AGB</i>	0.08	15.0	42.3	73
Vertical canopy cover ^d	%	<i>VCC</i>	1	40	82	73

^a The lowest non-zero value.^b On individual tree level instead of sample plot level.^c Stems taller than 2 m and >30 mm DBH.^d On sample plot level, trees under 2 m are not measured.^e Above ground tree biomass on sample plot level for trees taller than 2 m.^f N=2347 for the Optech data but the other statistics are very similar.**Table 3**Summary of the models (Eqs. (11)–(16)). The last column gives the relative RMSE when the normal vegetation ratio is used instead of the point-weighted vegetation ratio. α is the intercept.

Dependent variable	Data	Indep. variables, significance ^a			q ^d	R ² adj.	BIAS _r	RMSE _r	RMSE _r without pw ^c
Maximum tree height	TopEye	α ***	H_{95} ***	VR_{pw}^{all} ***	-	0.92	0.0%	8.8%	8.9%
Maximum tree height	Optech	α ***	H_{99} ***	VR_{pw}^{all} ***	-	0.91	0.0%	9.5%	9.6%
Biomass ^b	TopEye	α ***	H_{sum} ***	VR_{pw}^{1st} ***	1.018	0.99	0.0%	18.7%	19.9%
Biomass ^b	Optech	α ***	H_{sum} ***	VR_{pw}^{1st} ***	1.021	0.98	0.0%	21.2%	22.1%
Vertical canopy cover	TopEye	α ***	VR_{pw}^{all} ***		1.023	0.97	0.0%	16.8%	19.1%
Vertical canopy cover	Optech	α ***	VR_{pw}^{all} ***		1.016	0.96	0.0%	18.7%	19.9%

^a Abbreviation of significance (p): *<0.1, **<0.01, ***<0.001^b Above ground tree biomass (for trees taller than 2 m).^c Without point-weighted vegetation ratio.^d Logarithmic bias correction coefficient according to Eq. (6).