

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:

Eva Lindberg, Kenneth Olofsson, Johan Holmgren, Håkan Olsson. (2012)
Estimation of 3D vegetation structure from waveform and discrete return
airborne laser scanning data. *Remote Sensing of Environment*. Volume: 118,
Number: March, pp 151-161.

<http://dx.doi.org/10.1016/j.rse.2011.11.015>.

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, 118, MARCH, (2012)

<http://dx.doi.org/10.1016/j.rse.2011.11.015>

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

Estimation of 3D vegetation structure from waveform and discrete return airborne laser scanning data

Eva Lindberg^{a,*}, Kenneth Olofsson^a, Johan Holmgren^a, and Håkan Olsson^a

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

* Corresponding author. Email: Eva.Lindberg@slu.se,
Telephone: +46 90-7868536, Fax: +46 90-778 116

Abstract

This study presents and compares new methods to describe the 3D canopy structure with Airborne Laser Scanning (ALS) waveform data as well as ALS point data. The ALS waveform data were analyzed in three different ways; by summing the intensity of the waveforms in height intervals (a); by first normalizing the waveforms with an algorithm based on Beer-Lambert law to compensate for the shielding effect of higher vegetation layers on reflection from lower layers and then summing the intensity (b); and by deriving points from the waveforms (c). As a comparison, conventional, discrete return ALS point data from the laser scanning system were also analyzed (d). The study area was located in hemi-boreal, spruce dominated forest in the southwest of Sweden (Lat. 58° N, Long. 13° E). The vegetation volume profile was defined as the volume of all tree crowns and shrubs in 1 dm height intervals in a field plot and the total vegetation volume as the sum of the vegetation volume profile in the field plot. The total vegetation volume was estimated for 68 field plots with 12 m radius from the proportion between the amount of ALS reflections from the vegetation and the total amount of ALS reflections based on Beer-Lambert law. ALS profiles were derived from the distribution of the ALS data above the ground in 1 dm height intervals. The ALS profiles were rescaled using the estimated total vegetation volume to derive the amount of vegetation at different heights above the ground. The root mean square error (RMSE) for cross validated regression estimates of the total vegetation volume was 31.9% for ALS waveform data (a), 27.6% for normalized waveform data (b), 29.1% for point data derived from the ALS waveforms (c), and 36.5% for ALS point data from the laser scanning system (d). The correspondence between the estimated vegetation volume profiles was also best for the normalized waveform data and the point data derived from the ALS waveforms and worst for ALS point data from the laser scanning system as demonstrated by the Reynolds error index. The results suggest that ALS waveform data describe the volumetric aspects of vertical vegetation structure somewhat more accurately than ALS point data from the laser scanning system and that compensation for the shielding effect of higher vegetation layers is useful. The new methods for estimation of vegetation volume profiles from ALS data could be used in the future to derive 3D models of the vegetation structure in large areas.

1 Introduction

Vertical vegetation structure can be defined as the amount of vegetation material as a function of height above ground. This information is useful for habitat modeling, for example for bird species (Lefsky et al.

2002), and for mapping of tree vegetation successions. Shugart et al. (2010) also found 3D vegetation structure useful for modeling of the carbon cycle.

Airborne laser scanning (ALS) measures both the height of vegetation elements and the ground. From such data, it is possible to derive variables related to vegetation height and vegetation density as a function of height above ground. The height data can be used to describe the vegetation structure and, to a certain extent, obtain information about the height and density of different layers of vegetation such as the field layer, shrubs and canopy layers. Most commercial ALS systems deliver discrete returns or point laser data. The point data represent high intensities in the reflected light corresponding to surfaces where the light has been reflected. Due to limitations in the electronics of most ALS systems, only sufficiently spaced surfaces are distinguished as separate returns. However, with the development of sensors and electronics, waveform laser data have also become available from commercial ALS systems. ALS waveform data are values of the intensity of the reflected laser light sampled at short, regular intervals. ALS waveform data describe the whole backscattered signal and allow for more detailed processing, for example, derivation of points from the waveforms (Persson et al. 2005). The intensity of the reflected light depends on the reflectance of the reflecting object and the distance from the scanner to the reflecting object. Additionally, the gain of the laser scanner might be adjusted depending on the conditions at the moment when the light is received. To estimate the absolute physical properties of the reflecting object, information about sensor gain and distance to the reflecting object is necessary. However, for nearby pulses the sensor gain and the distance to the reflecting objects are similar and an approximation is to use the measured intensity as proportional to the sum of the reflectance of the reflecting object. If the reflecting objects have similar reflectance, a further approximation is to use the intensity as proportional to the area of the reflecting object.

Leaf area index (LAI) can be estimated based on ALS point data (Korhonen et al. 2011; Morsdorf et al. 2006; Solberg et al. 2009). Since parts of the laser beams penetrate through a canopy, estimation of the understory is also possible to some degree (Hill and Broughton 2009; Martinuzzi et al. 2009). ALS data may be used to separate single-story and multi-story stand structures through the shape of the distribution of ALS point data (Maltamo *et al.* 2005) or by examination of the height variability of local maxima in a canopy surface model defined from the points (Zimble *et al.* 2003). Correlation has been reported between the vegetation cover in different height intervals and the number of points in the intervals which makes it possible to use the distribution of points to characterize forest ecological structure (Miura and Jones 2010). A quantitative measure of the vertical vegetation structure can be derived by dividing the canopy into key layers and fitting a Weibull distribution, for example, to each layer (Coops et al. 2007; Jaskierniak et al. 2011). Su and Bork (2007) determined the height of the herbaceous layer, understory shrub and the overstory tree layer in rangeland by calculating the average height of all ALS points falling in each interval. ALS waveform data from the experimental SLICER system at NASA have been used to estimate a canopy height profile (CHP) that quantitatively represented the relative vertical distribution of canopy surface area which seems to be correlated with a CHP measured in field (Harding *et al.* 2001). The effects of canopy structure on the ALS waveform can be described by a 3D radiative transfer model (RTM) (Ni-Meister et al. 2001; Yang et al. 2010). To characterize the canopy structure and physical properties, RTMs have also been used to invert the waveform (Koetz *et al.* 2006). The range, amplitude, width, and backscatter cross-section may be estimated for each echo by modeling the waveform as a series of Gaussian pulses to provide more accurate vegetation classification (Wagner *et al.* 2008).

Even if several previous studies have used ALS data to model the vertical vegetation structure, we have only found three studies that validate the results against detailed ground measurements. Harding *et al.* (2001) estimated canopy height profiles for four selected forest stands from ALS data from SLICER using measurements from a telephoto lens calibrated to measure distances as ground truth. Hilker *et al.* (2010) estimated canopy volume profiles for four forest stands from ALS point data and compared them with canopy volume profiles estimated from terrestrial laser scanning. Hosoi *et al.* (2010) estimated Leaf Area Density (LAD) in a forest plot from ALS data and compared with LAD estimated from TLS data. All three studies showed reasonably good agreement between the estimates from ALS data and ground measurements.

The aim of this study is to develop and evaluate automatic methods to detect vertical vegetation structure from ALS data to represent the shrub layer and one or several tree layers. The study compares different methods of estimating the vegetation volume profile from ALS waveform data. In one method the ALS waveform data is used directly (direct waveform (a)), in a second method care is taken to first compensate for the shielding effect of higher vegetation layers on reflections from lower layers (normalized waveform (b)), and in a third method points are derived from the ALS waveform (waveform points (c)). For comparison, the vegetation volume profile is also estimated from the vertical distribution of conventional, discrete return ALS points from the laser scanning system (system points (d)). The vegetation volume profile is defined as the volume of all tree crowns and shrubs in 1 dm height intervals in a field plot. The vegetation volume profile is used rather than biomass profiles or LAD (Morsdorf *et al.* 2006) since it is feasible to measure in a large number of field plots. The vertical distribution of ALS data is compared with the vegetation volume profile and the vegetation volume at different heights above the ground is estimated. The results from the different methods are evaluated with cross validation to find the methods which produce the most accurate results.

2 Field reference data

2.1 Study area

The study area is located in the southwest of Sweden (Lat. 58° N, Long. 13° E). Parts of the study area have been used in studies of remote sensing for forest management planning (Holmgren *et al.* 2003). The most common tree species and their fraction of the total basal area are Norway spruce (*Picea abies* (L.)) (38.5%), Scots pine (*Pinus sylvestris* (L.)) (28.0%), birch (*Betula pendula* (L.) and *Betula pubescens* (L.)) (18.0%), oak (*Quercus robur* (L.)) (6.0%), alder (*Alnus glutinosa* (L.)) (5.5%), maple (*Acer platanoides* (L.)) (2.0%), aspen (*Populus tremula* (L.)) (0.5%), rowan (*Sorbus aucuparia* (L.)) (0.5%), and other broadleaved trees (0.5%). Additional small trees and shrubs in the area include bird cherry (*Prunus padus* (L.)) (0.5%), hazel (*Corylus avellana* (L.)), juniper (*Juniperus communis* (L.)), rose (*Rosa spp.*), willow (*Salix spp.*), raspberry (*Rubus idaeus* (L.)), currant (*Ribes spp.*) and other shrubs.

2.2 Field data collection

The study area was stratified on crown coverage and shrub coverage. An 80x80 m grid of points was laid out in the study area. Forest land was identified by using an existing forest management plan. The forest management plan was based on manual photo interpretation and contained polygons with forest stands and other land use types. Points outside forest land were divided into five crown coverage strata based on

manual photo interpretation. Only points with crown and/or shrub coverage > 0 were included. Points inside forest land were divided into five volume density strata. Volume density is a function of the stem volume divided by the mean tree height (Jonson 1914). The stem volume was taken from the forest management plan and the 95th percentile of the ALS data represented the mean tree height in each point. Only points where the maximum height of the ALS data was ≥ 20 m were included.

The result was ten crown coverage strata: five from manual photo interpretation and five from volume density. Three hundred and twenty nine points were randomly selected; one point in each polygon in the forest management plan. During the field inventory, the selected points were visited and the shrub coverage was estimated by counting the number of trees and shrubs with height 0.3-3 m within 5 m from the point. The point was placed in a shrub coverage stratum for numbers 5-15, 16-25, 26-35, 36-45 and ≥ 45 . If the number of already measured field plots in the stratum was less than three, the point was chosen as the center of a field plot.

Table 1. Number of field plots with basal area of different species or species groups $\geq 70\%$ for different intervals of basal area weighted mean DBH.

Basal area $\geq 70\%$	Basal area weighted mean DBH (cm)									
	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	>90
Pine	0	0	11	6	2	0	0	0	0	0
Spruce	0	0	8	3	3	1	0	0	0	0
Birch	0	1	4	1	1	0	0	0	0	0
Oak	0	0	0	0	1	0	0	1	0	3
Other broadleaf	0	0	0	1	1	1	0	0	0	0
Mixed	0	3	3	11	2	0	0	0	0	0

Sixty-eight circular field plots with 12 m radius were allocated during July and August 2009 (Table 1, Figure 1). The positions of the center of the field plots were measured using a Differential Global Positioning System (DGPS). The accuracy of the DGPS has been validated for 18 points in forest with crown coverage $\geq 50\%$ and mean basal area weighted height 6 m in the north of Sweden. The root mean square error of the DGPS positions was 0.27 m after post processing. The positions of all trees and shrubs were measured relative to the center using an ultrasound instrument (Lämås 2010). Within the field plots, the diameter at breast height (DBH) of all trees and shrubs with $DBH \geq 40$ mm was measured using a caliper and the species was recorded. The width of the tree crown was measured in two perpendicular directions, one of which was the direction towards the field plot center. If the crown was overlapping with another crown, the whole diameter was still measured. The total number of trees and shrubs in this

category was 2263. For a random sub-sample of trees with inclusion probability proportional to basal area, the height was also measured using a hypsometer. The number of trees for which the height was measured was 174. For all trees and shrubs with $DBH < 40$ mm and height ≥ 0.3 m, the width of the crown was measured in two perpendicular directions (in the same way as for the tree crowns), the height was measured and the species was recorded. The total number of trees and shrubs in this category was 3547. In September 2010 a complementary field inventory was undertaken to measure the live crown height of a sub-sample of trees using a hypsometer. The sub-sample was every tenth tree with $DBH \geq 40$ mm. The live crown height was defined as the height from the ground to the location of the lowest green branch on the stem. A single green branch isolated from the rest of the tree crown with at least three layers of dead branches did not count as live crown height. The number of trees for which the live crown height was measured was 328. The height and live crown height of all trees were then estimated from DBH using regression models (Eq. 1) and (Eq. 2) based on the sub-sample where tree height and/or live crown height were measured.

$$h_i = \alpha_0 + \alpha_1 \ln(DBH_i) + \epsilon_i \quad (1)$$

$$c_i/h_i = \beta_0 + \beta_1 DBH_i + \epsilon_i \quad (2)$$

where h_i is the height of tree i , DBH_i is the DBH of tree i , and c_i is the live crown height of tree i . Separate models were used for spruce (RMSE 28 dm and 0.29 respectively), pine (RMSE 32 dm and 0.10 respectively), birch (RMSE 21 dm and 0.15 respectively), and other species (RMSE 37 dm and 0.17 respectively). For the trees where the height and/or the live crown height were measured in the field, the field-measured values were used.

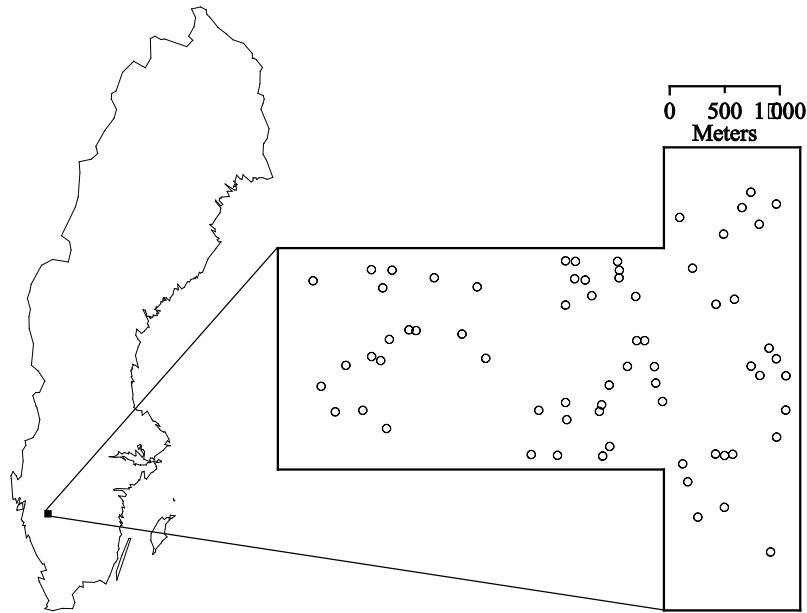


Fig. 1. Map of Sweden showing the location of the study area (left) and the study area with the extent of the ALS data and the location of the field plots (right).

2.3 Estimation of vegetation volume profiles from field data

The field data were used to derive 3D models of the trees and shrubs in the field plots. For trees and shrubs with DBH ≥ 40 mm, the crowns were modeled as ellipsoids (ellipses rotated around their vertical axis) by using the tree height minus the live crown height and the average width of the tree crown (Figure 2). Other trees and shrubs were modeled as half- ellipsoids by using the height and the average width of the crown. There is no standard choice for tree crown models although ellipses have been suggested by simulation studies (Nelson 1997)

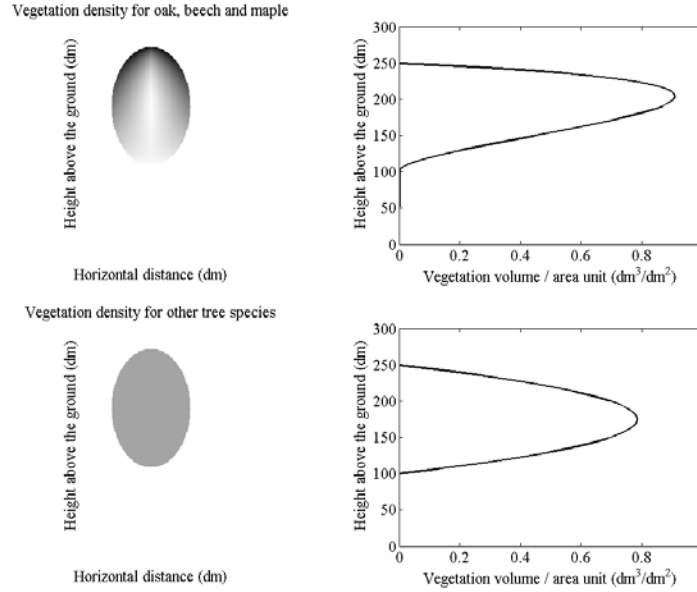


Fig. 2. Example of vegetation density and vegetation volume profiles for oak, beech and maple (top) and other tree species (bottom) for ellipsoid tree crown models.

For oak, beech and maple the density of the tree crowns $d(r, z)$ was set to be highest close to the surface of the tree crown and to decrease closer to the center of the tree crown (Eq. 3).

$$d(r, z) = W \times \frac{r+R_{max}-R}{R_{max}} \times \frac{z-c}{h-c} \quad r \leq R \leq R_{max}, c \leq z \leq h \quad (3)$$

where z is the height above the ground, r is the distance from the center of the tree stem, R is the radius of the tree crown at height z , R_{max} is the maximum radius of the tree crown, c is the live crown height, and h is the tree height. Although the density of the tree crown models for oak, beech and maple was not homogeneous, the weight, W , was selected and applied to make the total vegetation volume equal to that from a homogeneous tree crown. As a comparison, the tree crowns were also modeled using egg shaped curves rotated around their vertical axis and the shrubs were modeled using the upper part of an egg shaped curve (Eq. 4) (Figure 3).

$$R = R_{max} \times \sqrt{(1 - (\frac{z-m}{d})^2) \times (1 - S(\frac{z-m}{d}))} / C_S \quad m = \frac{h+c}{2}, d = \frac{h-c}{2} \quad (4)$$

The shape factor S was set to 0.5 and the constant C_S was set to 1.1009 to make the maximum of R equal to R_{max} .

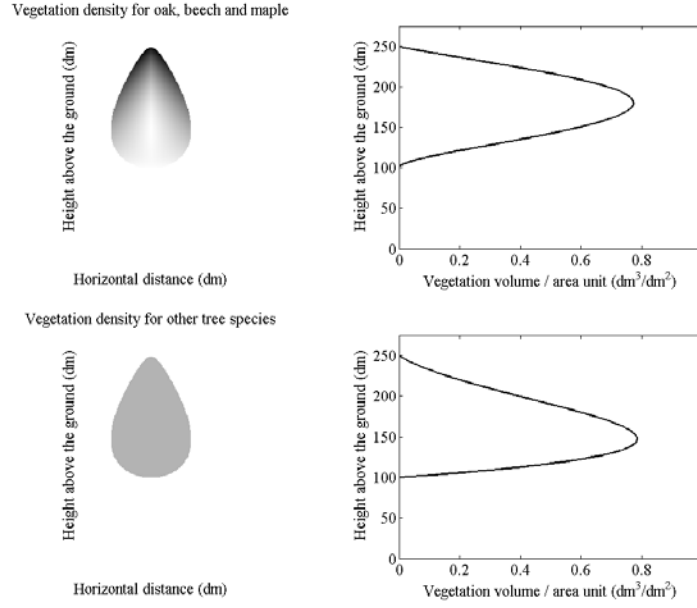


Fig. 3. Example of vegetation density and vegetation volume profiles for oak, beech and maple (top) and other tree species (bottom) for egg shaped tree crown models.

The measure used to characterize the vegetation structure was the volume of the tree crowns as derived from the field data. The vegetation volume profile $V_{veg}(z)$ was then modeled by summing the volume of all crowns in the field plot in intervals $\Delta z = 1$ dm above the ground (Eq. 5) (Figure 2).

$$V_{veg}(z) = \frac{\Delta z}{A} \sum_{i=1}^n \int_0^{R_i(z)} d_i(r, z) dr \quad (5)$$

where $d_i(r, z)$ is the density of tree i at the point (r, z) , $R_i(z)$ is the radius of the tree crown at height z , n is the total number of trees in the field plot and A is the area of the field plot. The value of the vegetation volume profile in each interval should be between 0 and 1 but could be greater than 1 if the tree crowns are overlapping. In such cases, the value was set to 1. The total vegetation volume V_{veg} in each field plot was calculated by summing the vegetation volume in each height interval ≥ 1 m above the ground (Eq. 6).

$$V_{veg} = \sum_{z=0}^{h_{max}} V_{veg}(z) \quad (6)$$

where h_{max} is the maximum height of all trees in the field plot.

3 ALS data processing

3.1 ALS data acquisition

The ALS data were acquired on September 4th 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, the footprint size was 25 cm, the average pulse density was 7 m^{-2} and the standard deviation of the pulse density was 3 m^{-2} for all field plots. The returned waveform was measured with a sampling frequency of 2 GHz and the first and last returns were also saved for each laser pulse. The accuracy of the ALS data was $< 0.1 \text{ m}$ in vertical direction and $< 0.3 \text{ m}$ in horizontal direction (communication with surveying company Blom)

3.2 Derivation of DEM from ALS data

System points were classified as ground or non-ground using a progressive Triangular Irregular Network (TIN) densification method (Axelsson 2000) in the TerraScan software (Soiainen 2004). A Digital Elevation Model (DEM) was derived as the mean value of the ground points in 0.5 m raster cells. TIN interpolation was used for raster cells with no data. A positive bias was observed for the waveform points relative to the system points. The reason for the bias is that the laser scanning system uses a different algorithm to derive the ALS points. A bias of 1 m was added to the DEM when processing the ALS waveforms based on the distribution of the waveform points and the assumption that the distribution should have a maximum at ground level.

3.3 Normalization of ALS waveforms

This processing was done to derive the normalized waveform (b). The laser scanner detects the intensity of the reflected light. If the emitted light intersects with the canopy, the intensity of the light that penetrates further into the vegetation will decrease exponentially (Chevalier *et al.* 2007). To compensate for this, an iterative normalization algorithm based on Beer-Lambert law was developed. Beer-Lambert law describes exponential decrement of light intensity and has been used to estimate LAI based on ALS data (Solberg *et al.* 2006). The algorithm assumed that each thin layer of vegetation which the light penetrates decreases the intensity according to (Eq. 7).

$$I_{k+1} = I_k \times e^{-\sigma_k N_k l} = I_k \times e^{-A_k} \quad (7)$$

where I is the intensity, σ is the cross-section of light attenuation by one particle, N is the density per volume of attenuating particles, and l is the path length along the light beam. A is the product of I , σ and N . In the case of laser scanning of vegetation, the attenuating particles are parts of the vegetation. The algorithm further assumed that the ratio between the reflectance and the attenuation is constant. The initial value of A_k was set to a multiple of R_k which is the value of the reflected intensity in bin k of the waveform (Eq. 8).

$$A_k = \frac{(A_{k,app})}{(R_k)} \times R_k \quad (8)$$

where $A_{k,app} = -\ln\left(\frac{R_{k+1}}{R_k}\right)$. For each iterative step, a multiple of A_k was approximated (Eq. 9).

$$CA'_k = \frac{R_k}{e^{-\sum_{m=1}^{k-1} A_m}} \quad (9)$$

where C is a constant valid during that step. An updated value of A_k was calculated (Eq. 10).

$$A'_k = \frac{(A_{k,app})}{(CA'_k)} \times CA'_k \quad (10)$$

3.4 Decomposition of ALS waveforms

This processing was done to derive the waveform points (c). Each ALS waveform y was decomposed into a sum of Gaussian components $N(\mu, \sigma)$ (Eq. 11).

$$\hat{y} = \sum_{s=1}^m p_s N(\mu_s, \sigma_s) \quad (11)$$

The components represented surfaces where the light had been reflected, which made it possible to analyze the distribution of those surfaces. Additionally it was necessary to identify reflections from the ground which was possible by identifying the components closest to the Digital Elevation Model (DEM). The decomposition was done using the Expectation-Maximization (EM) algorithm (Persson *et al.* 2005).

The EM algorithm was run with 25 iterations. Only components with $p_s \geq 0.05$ were saved. A solution was accepted if all components were sufficiently separated (Eq. 12).

$$\frac{|\mu_s - \mu_t|}{\sqrt{\sigma_s^2 + \sigma_t^2}} \geq 2 \quad \forall s \leq m, \forall t \leq m, s \neq t \quad (12)$$

The decomposition was done in three steps.

Step 0. The initial number of components m was the number of local maxima in the waveform, the initial values of μ_s were set to the bins of the local maxima of the waveform, the initial values of σ_s were set to the number of bins in the waveform divided by two times the initial number of components, and the initial values of the weight p_s were set to one divided by the initial number of components. The EM algorithm was run once in step 0.

Step 1. The number of components was decreased by removing the component with the smallest p_s . The EM algorithm was run again with the updated components as initial values. Step 1 was repeated as long as the minimum value of $p_s < 0.2$ and the number of components was greater than one.

Step 2. The number of components was increased by splitting the component with the largest local deviation from the waveform into two components and the EM algorithm was run again with the updated components as initial values. The local deviation for each component was calculated by summing the absolute value of the difference between the approximated waveform \hat{y} and the original waveform y in all bins that were closer to the component than to any other component. Step 2 was repeated ten times.

The solution with the minimum value of AIC as defined in Persson et al. (2005) was selected. The resulting components were saved in a LAS file for further processing as ALS points. The intensity of each point was set to the relative weight of the component multiplied by the total area under the waveform.

3.5 Estimation of ALS point profiles

This analysis was undertaken for both the waveform points (c) and for the system points (d). A local z value was calculated from the ALS points and the DEM (Eq. 13).

$$z_{local} = z - z_{DEM} \quad (13)$$

where z_{DEM} is the z value of the DEM for the same x,y location. ALS point profiles were derived by dividing the local z values into height intervals of 1 dm height and counting the number of points in each interval. The height intervals were chosen to correspond to the intervals used for the field data. The vegetation ratio r_{point} was calculated by dividing the number of points with local $z \geq 1$ m ($\sum(n_{veg})$) by the total number of points ($\sum(n_{tot})$) (Eq. 14). The analysis was carried out using the data from within the circular field plots.

$$r_{point} = \frac{\sum(n_{veg})}{\sum(n_{tot})} \quad (14)$$

3.6 Estimation of ALS waveform profiles

This analysis was undertaken for both the direct waveform (a) and for the normalized waveform (b). The DEM was used to calculate local z values for all ALS waveforms where some part of the waveform originated from the area inside the circular field plots. If a Gaussian component from the EM algorithm was < 1 m above the DEM, the Gaussian component was subtracted from the waveform. The remaining waveform was divided into cubic voxels with 1 dm sides and the maximum intensity of all waveforms in each voxel was saved. An ALS waveform vegetation profile was calculated as the sum of the intensities of all voxels in height intervals of 1 dm height inside the circular field plots. This is similar to the canopy volume profile defined by Lefsky et al. (1999) with the difference that Lefsky et al. set the value in a voxel to 1 if waveform intensity was present and 0 otherwise. The total ALS waveform profile was calculated in the same way as the waveform vegetation profile but including the whole waveforms. The vegetation ratio r_{wave} was calculated by dividing the area under the waveform vegetation profile ($\sum(I_{veg})$) by the total area under the waveform profile ($\sum(I_{tot})$) (Eq. 15) (Figure 4).

$$r_{wave} = \frac{\sum(I_{veg})}{\sum(I_{tot})} \quad (15)$$

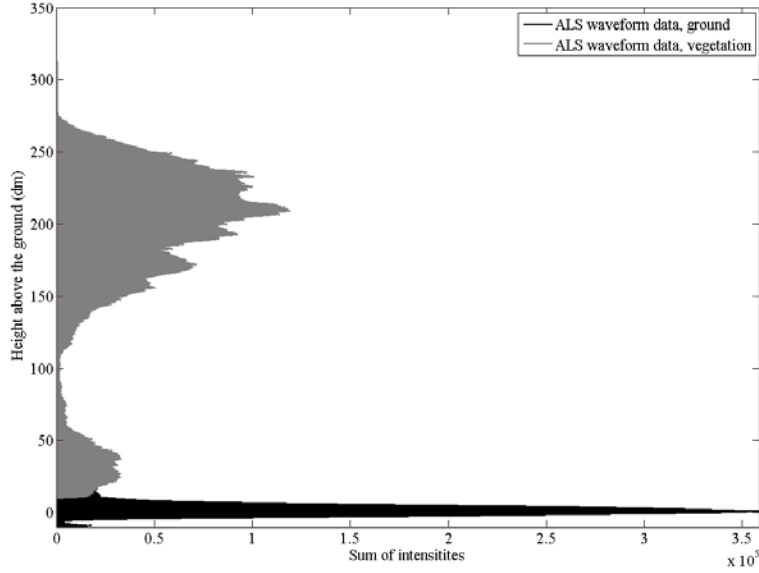


Fig. 4. Example of an ALS profile from ground (local $z \leq 1$ m) and vegetation (local $z \geq 1$ m).
The total ALS profile is the combination of the two ALS profiles together.

4 Estimation and validation

4.1 Estimation of total vegetation volume and vegetation volume profiles from ALS data

This analysis was undertaken using the vegetation ratio and the ALS profiles derived by all four methods. The total vegetation volume was estimated from the vegetation ratio with a log-linear model based on Beer-Lambert law (Solberg *et al.* 2009) (Eq. 16).

$$V_{veg,j} = \alpha_0 + \alpha_1 \ln(1 - r_j) + \epsilon_j \quad (16)$$

where $V_{veg,j}$ is the total vegetation volume in plot j and r_j is the vegetation ratio calculated from the points or the waveforms. The ALS profile in each field plot $I(h)$ was rescaled by dividing by the area under the ALS profile $\sum(I_{veg})$ and multiplied by the estimated total vegetation volume \hat{V}_{veg} to give the ALS profile the same area as the estimated total vegetation volume (Eq. 17).

$$\hat{V}_{veg}(z) = \hat{V}_{veg} \times \frac{I(z)}{\sum I_{veg}} \quad (17)$$

The area under the estimated vegetation volume profile was calculated for three different height intervals above the ground: local $z \leq 30$ dm, local $z \leq 100$ dm and local $z > 100$ dm. This was used as an estimate of the vegetation volume for the different height intervals.

4.2 Validation

The validation was done for all four methods. The accuracy of the total vegetation volume estimated from ALS data was validated with the total vegetation volume modeled from field measurements using the root mean square error (RMSE) and bias (Eq. 18) and (Eq. 19).

$$RMSE = \sqrt{\frac{\sum_{j=1}^n (V_{veg,j}^{\hat{}} - V_{veg,j})^2}{n}} \quad (18)$$

$$bias = \frac{\sum_{j=1}^n (V_{veg,j}^{\hat{}} - V_{veg,j})}{n} \quad (19)$$

where $V_{veg,j}$ is the total vegetation volume in plot j and n is the number of plots. The error index (EI) (Reynolds *et al.* 1988) was calculated for the vegetation profiles and the estimates from ALS data (Eq. 20).

$$EI = \frac{1}{N_T} \sum_{k=1}^m \left| \hat{F}_k - F_k \right| \quad (20)$$

where \hat{F}_k is the estimated value in interval k , F_k is the true value in interval k , m is the number of intervals, and N_T is the sum of F_k in all intervals. The size of the intervals was set to 1/10th of the maximum tree height or height of the ALS data in each field plot. The error index can be large for two histograms with similar shapes that are dislocated relative to each other. To compare the shapes of the profiles in such cases, the heights of the estimated vegetation volume profiles in each field plot were adjusted to have the same mean as the vegetation profile from field data. The mean height of the estimated vegetation volume profile was subtracted and the mean height of the vegetation profile from field data was added. The error index was calculated for the adjusted profiles in the same way as for the original profiles. To compare the shapes of the profiles independently of the estimated total vegetation volume, the error index as well as the mean height adjusted error index were also calculated for estimated vegetation volume profiles that were rescaled to have the same area as the total vegetation volume from field data (Figure 5). The validation was done for heights ≥ 1 m above the ground.

The accuracy of the vegetation volume for the different height intervals estimated from ALS data was validated with the vegetation volume modeled from field measurements using the RMSE and bias. The validation was done for heights ≥ 1 m above the ground.

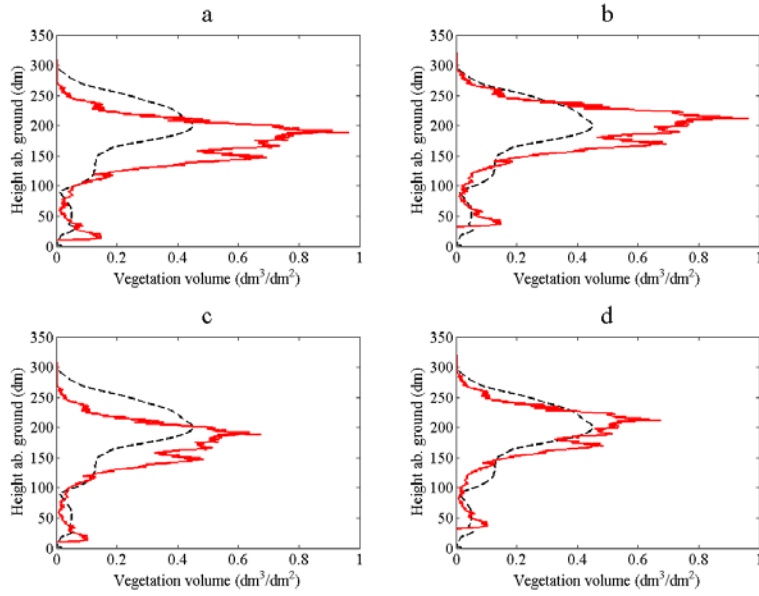


Fig. 5. Vegetation profiles from one field plot. Dotted line = field data, solid line = estimates from ALS data. The original estimated profile from ALS data (a), the estimated profile adjusted with the mean height from field data (b), the estimated profile with the area rescaled to field data (c), the estimated profile with the area rescaled to field data and adjusted with the mean height from field data (d).

5 Results

The RMSE of the total vegetation volume was smallest for the normalized waveform (b), second smallest for the waveform points (c), third smallest for the direct waveform (a) and largest for the system points (d). The bias of the total vegetation volume was also smallest for the normalized waveform (b) and largest for the system points (d) although the bias was very small (Figure 6 and Table 2).

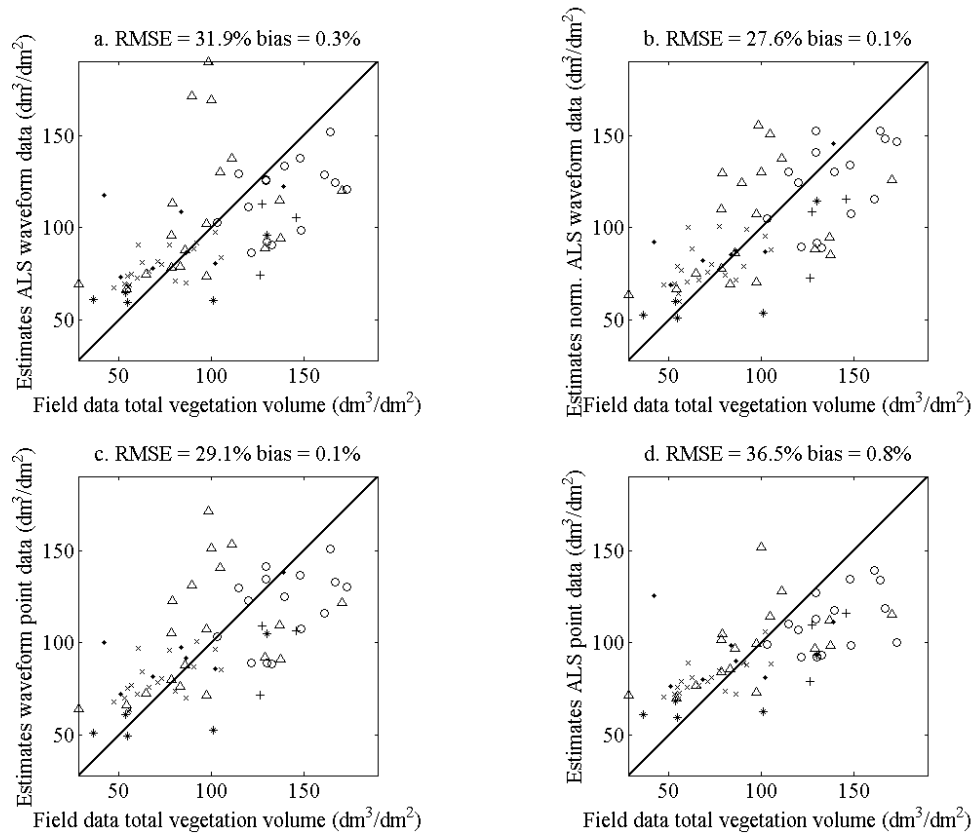


Fig. 6. Total vegetation volume estimated from direct waveform (a), normalized waveform (b), waveform points (c), and system points (d) for the 68 field plots. Field plots dominated ($\geq 70\%$ of basal area) by pine are marked 'x', spruce 'o', birch '.', oak '*', other broadleaf '+' and mixed forest ' Δ '.

Table 2. RMSE and bias of the estimates of vegetation volume.

	Direct waveform (a)		Normalized waveform (b)		Waveform points (c)		System points (d)	
	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
Total vegetation volume (dm ³ /dm ²)	30.9 (31.9%)	0.3 (0.3%)	26.8 (27.6%)	0.1 (0.1%)	28.3 (29.1%)	0.1 (0.1%)	35.4 (36.5%)	0.8 (0.8%)
Vegetation volume for local $z \leq 30$ dm (dm ³ /dm ²)	2.0 (76.0%)	0.5 (18.5%)	2.1 (77.2%)	0.2 (7.3%)	1.7 (64.0%)	0.3 (10.2%)	2.8 (105.5%)	1.1 (42.5%)
Vegetation volume for local $z \leq 100$ dm (dm ³ /dm ²)	13.7 (62.6%)	-3.8 (-17.2%)	11.8 (54.0%)	2.0 (9.3%)	12.1 (55.5%)	-1.4 (-6.5%)	16.1 (73.7%)	0.3 (1.5%)
Vegetation volume for local $z > 100$ dm (dm ³ /dm ²)	27.2 (36.0%)	4.0 (5.3%)	24.7 (32.7%)	-1.9 (-2.5%)	25.9 (34.3%)	1.6 (2.2%)	28.0 (37.1%)	0.5 (0.7%)

The bold numbers indicate the best results for RMSE and bias in each row.

Figure 7 shows examples of vegetation profiles and estimated vegetation volume profiles from the normalized waveform (b) for nine different field plots. The shape of the profiles was generally consistent between the field data and the ALS data, however, the mean height and the total area under the profiles sometimes differed. Also the amount of understory was often underestimated (as seen for field plots b, c and f).

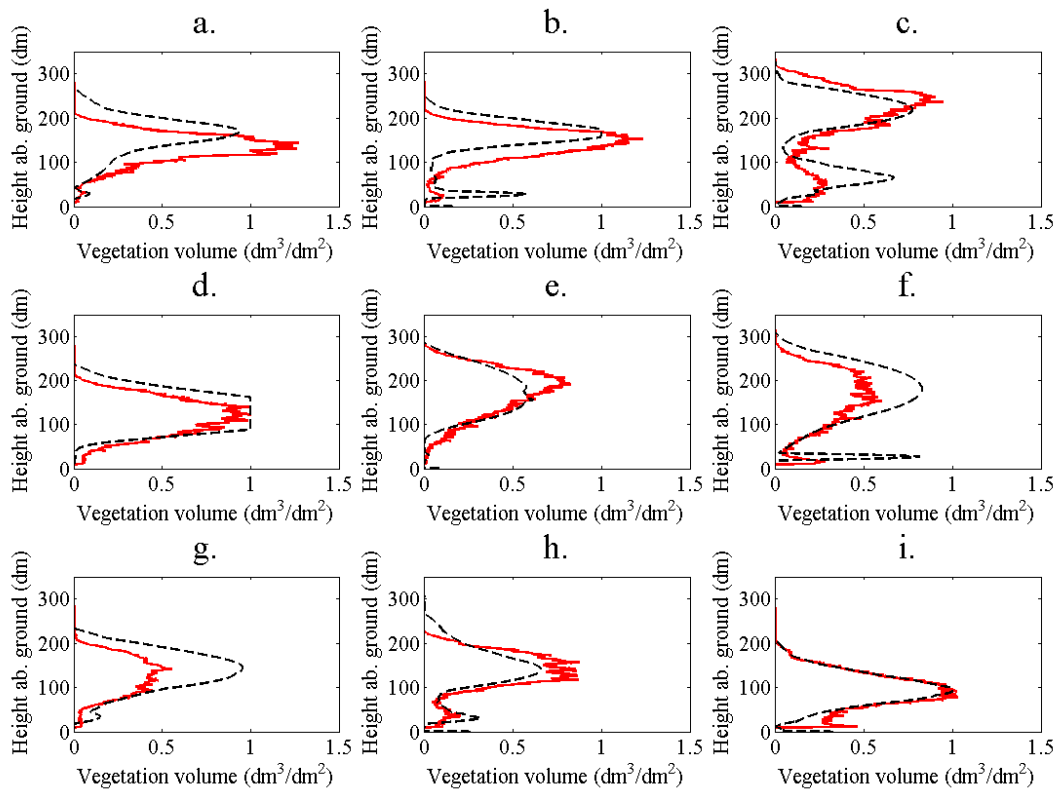


Fig. 7. Vegetation profiles from nine field plots. Dotted line = field data, solid line = estimates from normalized waveform (b). Gaussian components b1 m above the DEM have been subtracted from the ALS profiles. Pine forest with no undergrowth (a), pine forest with low undergrowth (b), pine forest with high undergrowth (c), spruce forest with no undergrowth (d), spruce forest with no undergrowth (e), spruce forest with low undergrowth (f), birch forest with medium undergrowth (g), birch forest with high undergrowth (h), birch/spruce forest, multilayered (i).

The mean error index was smallest for the normalized waveform (b) and waveform points (c), second smallest for the direct waveform (a) and largest for the system points (d) (Table 3).

The RMSE and bias for the height intervals ≤ 30 dm and ≤ 100 dm (Table 2) were much larger on a percentage basis than for the total vegetation volume. The RMSE was smallest for the waveform points (c) and largest for the system points (d). The bias for the height interval ≤ 30 dm was positive in all cases. However, for the height interval ≤ 100 dm both positive and negative bias was observed.

The RMSE of the vegetation volume for local $z > 100$ dm was smallest for the normalized waveform (b) and largest for the system points (d). The bias of the total vegetation volume for this height interval was smallest for the system points (d) and largest for the direct waveform (a) (Table 2).

The largest positive bias of the total vegetation volume was found for birch dominated field plots and field plots with mixed forest. For the birch dominated field plots, the bias was largest for waveform points

(c) and smallest for normalized waveform (b). For the field plots with mixed forest, the bias was largest for system points (d) and smallest for normalized waveform (b). The largest negative bias was found for field plots dominated by other broadleaf and spruce dominated field plots. For the field plots dominated by other broadleaves, the bias was largest for waveform points (c) and smallest for system points (d). For the spruce dominated field plots, the bias was largest for system points (d) and smallest for normalized waveform (b) (Table 4).

Table 3. Mean error index (unitless) for the different estimated vegetation volume profiles. Note that the error index is only comparable within each row.

	Direct waveform (a)	Normalized waveform (b)	Waveform points (c)	System points (d)
Error index original profiles	0.58	0.58	0.54	0.62
Error index mean height adjusted profiles	0.43	0.39	0.39	0.46
Error index profiles with area rescaled to field data	0.50	0.44	0.46	0.52
Error index profiles with area rescaled to field data and mean height adjusted	0.35	0.29	0.30	0.37

The bold numbers indicate the best result in each row.

Table 4. RMSE and bias of the estimates of total vegetation volume (dm^3/dm^2) for different species.

Basal area \geq 70%	Mean total veg. volume	Direct waveform (a)		Normalized waveform (b)		Waveform points (c)		System points (d)	
		RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
Pine	71.0	15.3 (21.5%)	8.0 (11.2%)	17.2 (24.3%)	8.8 (12.4%)	16.3 (23.0%)	8.9 (12.5%)	15.7 (22.1%)	9.4 (13.2%)
Spruce	138.8	29.5 (21.2%)	-21.4 (-15.4%)	26.4 (19.0%)	-14.9 (-10.7%)	28.6 (20.6%)	-18.9 (-13.6%)	33.1 (23.9%)	-27.0 (-19.4%)
Birch	81.7	33.1 (40.5%)	13.6 (16.7%)	21.9 (26.8%)	11.1 (13.6%)	25.4 (31.0%)	13.7 (16.8%)	36.2 (44.3%)	13.2 (16.1%)
Oak	75.1	26.7 (35.5%)	-6.7 (-9.0%)	23.6 (31.5%)	-8.8 (-11.7%)	25.6 (34.0%)	-11.2 (-14.9%)	27.2 (36.2%)	-6.0 (-8.0%)
Other broadleaf	133.2	38.9 (29.2%)	-35.7 (-26.8%)	37.1 (27.9%)	-34.2 (-25.7%)	40.1 (30.1%)	-37.3 (-28.0%)	33.8 (25.4%)	-31.5 (-23.7%)
Mixed	96.0	41.0 (42.7%)	12.5 (13.0%)	34.4 (35.8%)	6.9 (7.2%)	35.9 (37.4%)	10.3 (10.8%)	50.2 (52.4%)	16.4 (17.0%)

For the egg shaped tree crown models, the RMSE and bias were larger but the order between the methods was similar as for ellipsoids except for the height interval ≤ 30 dm where the order between the different methods differed slightly.

6 Discussion

This study has introduced new methods to estimate vegetation volume profiles from ALS data. The estimation was done for conventional, discrete return ALS point data from the laser scanning system and in three different ways for ALS waveform data. The study has shown that estimation of vegetation volume profiles from ALS data is feasible and may be used in the future to derive 3D models of the vegetation structure in large areas after calibration with total vegetation volume from field plots. Furthermore, it was shown that ALS waveform data produced the best results. Although the differences between the different methods using waveform data were small, they suggest that processing of waveform data may improve the accuracy and that the waveform data should preferably be compensated for the shielding effect of higher vegetation layers. The results from egg shaped tree crown models were similar except for the height interval ≤ 30 dm. The accuracy was low for the height interval ≤ 30 dm for all methods and it is probably not feasible to estimate such low vegetation below a canopy. The accuracy for the height interval ≤ 100 dm was slightly higher but was probably also affected by higher layers of vegetation.

The method requires field measurements of the total vegetation volume in a number of field plots to be used as training data although measurements of the vegetation volume profile are not needed. Before rescaling, the ALS profiles represent the relative amount of vegetation at different height above the ground. These profiles cannot be compared for different field plots since the distance from the laser scanner and the sensor gain varies. However, the profiles could be useful for manual interpretation. The vegetation volume profile is probably strongly correlated with the established ecological term LAD but measurements of LAD for a large number of field plots would be difficult.

An important source of error in this study is that the field data and the tree crown models have limitations. The tree height and the live crown height were measured for a sub-sample of trees and estimated from DBH for the rest. Due to this limitation, the trees in some field plots may be lower than the maximum height of the ALS data. Additionally, only simple tree crown models were used because of the limited measurements of tree crowns in the field. The results showed a tendency for overestimation of the total vegetation volume from ALS data for plots with tall birch trees; this may be due to these plots having denser than average tree crowns and higher reflectance, despite the fact that the same tree crown models were used. The structure of the tree crowns varies both between different species and between trees of the same species of different age. This is also valid for the reflectance, which typically is higher for deciduous trees than for coniferous trees. More detailed field measurements of the tree crowns would make it possible to create very detailed models of the tree crowns, but this was not realistic for a large number of field plots. This means that the vegetation volume profiles derived from the field data have considerable uncertainty and it can be argued that the ALS data describe the tree crowns more accurately than the measurements on the ground. However, the field data are collected with statistically sound methods since random sub-sampling was used and the diameter of the tree crowns was measured in directions parallel with the radius of the field plot. This could give rise to an overestimation or underestimation of individual tree crowns but it should not cause any systematic error which could happen if the tree crowns were measured in north-south and east-west directions. Therefore the data should be useful for training and comparison of the different methods. The similarity of the results from egg shaped tree crown models also implies that the validation is not sensitive to the exact choice of crown models. To obtain detailed ground measurements of the tree crowns, terrestrial laser scanning is a promising technique which is currently being developed for this purpose (Hilker et al. 2010; Hosoi et al. 2010; Lovell et al. 2003). Most of the field inventory was done less than one growth season after the laser scanning except for the measurement of live crown height which was done two growth seasons after the laser scanning. However, the live crown height does not undergo major changes for older trees. For younger trees however, this time lag could also be a source of error.

The normalization of the waveform data requires knowledge of how much light is lost when the light beam is reflected in higher layers of the vegetation. The algorithm used in this study is a simplification and approximation but there are few studies that have taken this into account. Wagner et al. (2008) decomposed waveforms and used calibration to find the cross-section for each component and the area of the corresponding reflecting object. The area of collision of all reflecting objects for one laser pulse were assumed to be equal to the footprint area which made it possible to calculate the area of collision for all objects. Calculation of the cross-section requires reference measurements for calibration. It also requires data on the width of the components which makes it impossible to use conventional, discrete return ALS data in the current LAS format for this purpose.

The analysis of the ALS waveform data in this study did not include deconvolution of the signal. Deconvolution would be necessary to distinguish surfaces separated by a smaller distance than the order of the length of the transmitted pulse. The duration of the transmitted pulse of the laser scanning system used in this study was 4 ns, which means that the length of the pulse is 1.2 m. Since the shrub layer and the tree layer of the forest are typically at least a few meters in height, this limitation most likely does not present a problem in this study. The current study has analyzed the vertical vegetation structure of the forest at an area level. To analyze the horizontal vegetation structure of the forest with a higher resolution, analysis at individual tree level would most likely be better, for example, by clustering of the ALS points in 3D (Reitberger *et al.* 2009). Analysis at individual tree level would also make it possible to obtain information about the number of trees in different height classes which could also be valuable for describing the vertical vegetation structure.

The new methods for estimation of vegetation volume profiles from ALS data could be used in the future to produce input for mapping of forest structure and modeling of habitats. The methods have been evaluated numerically and the results show good agreement with ground measurement in line with previous studies (Harding *et al.* 2001; Hilker *et al.* 2010; Hosoi *et al.* 2010). The results suggest that the vegetation structure can be described more accurately from ALS waveform data and that compensation for the shielding effect of higher vegetation layers is useful.

Acknowledgements

This study was financed by The Swedish Environmental Protection Agency through the research program EMMA. The field inventory and the acquisition of ALS data were financed by the Hildur and Sven Wingquist foundation. The field inventory was done by Axel Bergsten and Mårten Svensson. We would also like to thank the anonymous reviewers who helped us to improve our manuscript as well as Heather Reese and Neil Cory who reviewed the language in the manuscript.

References

- Axelsson, P. (2000). DEM generation from laser scanner data using adaptive TIN models. *International Archives of Photogrammetry and Remote Sensing*, 33, 111-118.
- Chevalier, T., Steinvall, O., & Larsson, H. (2007). Performance of laser penetration through forest vegetation. In, *Laser Radar Technology and Applications XII* (p. 65500Q). Orlando, FL, USA: SPIE.
- Coops, N.C., Hilker, T., Wulder, M.A., St-Onge, B., Newnham, G., Siggins, A., & Trofymow, J.A. (2007). Estimating canopy structure of Douglas-fir forest stands from discrete-return LiDAR. *Trees-Structure and Function*, 21, 295-310.
- Harding, D.J., Lefsky, M.A., Parker, G.G., & Blair, J.B. (2001). Laser altimeter canopy height profiles - Methods and validation for closed-canopy, broadleaf forests. *Remote Sensing of Environment*, 76, 283-297.
- Hilker, T., van Leeuwen, M., Coops, N.C., Wulder, M.A., Newnham, G.J., Jupp, D.L.B., & Culvenor, D.S. (2010). Comparing canopy metrics derived from terrestrial and airborne laser scanning in a Douglas-fir dominated forest stand. *Trees-Structure and Function*, 24, 819-832.

- Hill, R.A., & Broughton, R.K. (2009). Mapping the understory of deciduous woodland from leaf-on and leaf-off airborne LiDAR data: A case study in lowland Britain. *Isprs Journal of Photogrammetry and Remote Sensing*, *64*, 223-233.
- Holmgren, J., Nilsson, M., & Olsson, H. (2003). Estimation of tree height and stem volume on plots-using airborne laser scanning. *Forest Science*, *49*, 419-428.
- Hosoi, F., Nakai, Y., & Omasa, K. (2010). Estimation and Error Analysis of Woody Canopy Leaf Area Density Profiles Using 3-D Airborne and Ground-Based Scanning Lidar Remote-Sensing Techniques. *Ieee Transactions on Geoscience and Remote Sensing*, *48*, 2215-2223.
- Jaskierniak, D., Lane, P.N.J., Robinson, A., & Lucieer, A. (2011). Extracting LiDAR indices to characterise multilayered forest structure using mixture distribution functions. *Remote Sensing of Environment*, *115*, 573-585.
- Jonson, T. (1914). Om bonitering av skogsmark (in Swedish). *Skogsvårdsföreningens tidskrift, tolfte årgången*, 369 – 392.
- Koetz, B., Morsdorf, F., Sun, G., Ranson, K.J., Itten, K., & Allgower, B. (2006). Inversion of a lidar waveform model for forest biophysical parameter estimation. *Ieee Geoscience and Remote Sensing Letters*, *3*, 49-53.
- 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.
- Lefsky, M.A., Cohen, W.B., Acker, S.A., Parker, G.G., Spies, T.A., & Harding, D. (1999). Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. *Remote Sensing of Environment*, *70*, 339-361.
- Lefsky, M.A., Cohen, W.B., Parker, G.G., & Harding, D.J. (2002). Lidar remote sensing for ecosystem studies. *Bioscience*, *52*, 19-30.
- Lovell, J.L., Jupp, D.L.B., Culvenor, D.S., & Coops, N.C. (2003). Using airborne and ground-based ranging lidar to measure canopy structure in Australian forests. *Canadian Journal of Remote Sensing*, *29*, 607-622.
- Lämås, T. (2010). The Haglöf PostEx ultrasound instrument for the positioning of objects on forest sample plots. In (p. 10): Swedish University of Agricultural Sciences, Department of Forest Resource Management.
- Maltamo, M., Packalén, P., Yu, X., Eerikäinen, K., Hyypä, J., & Pitkänen, J. (2005). Identifying and quantifying structural characteristics of heterogeneous boreal forests using laser scanner data. *Forest Ecology and Management*, *216*, 41-50.
- Martinuzzi, S., Vierling, L.A., Gould, W.A., Falkowski, M.J., Evans, J.S., Hudak, A.T., & Vierling, K.T. (2009). Mapping snags and understory shrubs for a LiDAR-based assessment of wildlife habitat suitability. *Remote Sensing of Environment*, *113*, 2533-2546.
- Miura, N., & Jones, S.D. (2010). Characterizing forest ecological structure using pulse types and heights of airborne laser scanning. *Remote Sensing of Environment*, *114*, 1069-1076.
- Morsdorf, F., Kotz, B., Meier, E., Itten, K.I., & Allgower, B. (2006). Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction. *Remote Sensing of Environment*, *104*, 50-61.

Nelson, R. (1997). Modeling forest canopy heights: The effects of canopy shape. *Remote Sensing of Environment*, 60, 327-334.

Ni-Meister, W., Jupp, D.L.B., & Dubayah, R. (2001). Modeling lidar waveforms in heterogeneous and discrete canopies. *Ieee Transactions on Geoscience and Remote Sensing*, 39, 1943-1958.

Persson, Å., Söderman, U., Töpel, J., & Ahlberg, S. (2005). Visualization and analysis of full-waveform airborne laser scanner data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36, 103–108.

Reitberger, J., Schnorr, C., Krzystek, P., & Stilla, U. (2009). 3D segmentation of single trees exploiting full waveform LIDAR data. *Isprs Journal of Photogrammetry and Remote Sensing*, 64, 561-574.

Reynolds, M.R., Burk, T.E., & Huang, W.C. (1988). Goodness-of-fit tests and model selection procedures for diameter distribution models. *Forest Science*, 34, 373-399.

Shugart, H.H., Saatchi, S., & Hall, F.G. (2010). Importance of structure and its measurement in quantifying function of forest ecosystems. *Journal of Geophysical Research-Biogeosciences*, 115

Soininen, A. (2004). Terra Scan for MicroStation, user's guide. *Terrasolid Ltd., Jyväskylä, Finland*, 132

Solberg, S., Brunner, A., Hanssen, K.H., Lange, H., Næsset, E., Rautiainen, M., & Stenberg, P. (2009). Mapping LAI in a Norway spruce forest using airborne laser scanning. *Remote Sensing of Environment*, 113, 2317-2327.

Solberg, S., Næsset, E., Hanssen, K.H., & Christiansen, E. (2006). Mapping defoliation during a severe insect attack on Scots pine using airborne laser scanning. *Remote Sensing of Environment*, 102, 364-376.

Su, J.G., & Bork, E.W. (2007). Characterization of diverse plant communities in Aspen Parkland rangeland using LiDAR data. *Applied Vegetation Science*, 10, 407-416.

Wagner, W., Hollaus, M., Briese, C., & Ducic, V. (2008). 3D vegetation mapping using small-footprint full-waveform airborne laser scanners. *International Journal of Remote Sensing*, 29, 1433-1452.

Yang, W.Z., Ni-Meister, W., Kiang, N.Y., Moorcroft, P.R., Strahler, A.H., & Oliphant, A. (2010). A clumped-foliage canopy radiative transfer model for a Global Dynamic Terrestrial Ecosystem Model II: Comparison to measurements. *Agricultural and Forest Meteorology*, 150, 895-907.

Zimble, D.A., Evans, D.L., Carlson, G.C., Parker, R.C., Grado, S.C., & Gerard, P.D. (2003). Characterizing vertical forest structure using small-footprint airborne LiDAR. *Remote Sensing of Environment*, 87, 171-182.