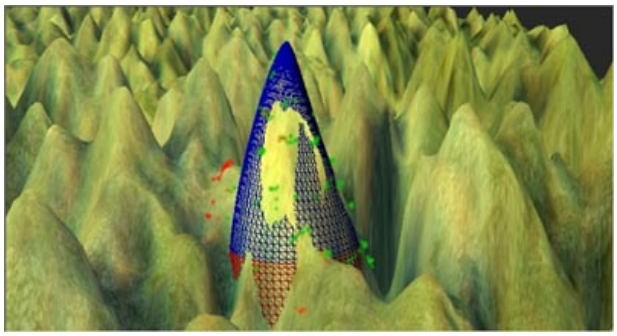
Proceedings SilviLaser 2010

The 10th International Conference on LiDAR Applications for Assessing Forest Ecosystems



Proceedings Silvilaser 2010

The 10th International Conference on LiDAR Applications for Assessing Forest Ecosystems

September 14th - 17th, 2010
Freiburg, Germany

Organised by

Sponsored by

Albert-Ludwigs-University of Freiburg, Department of Remote Sensing and Landscape Information Systems (FeLis)

Forstliche Versuch- und Forschungsanstalt Baden-Württemberg (FVA)

Prof. Dr. Barbara Koch, Dr. Gerald Kändler

Christelle Teguem (Editor)

Department of Remote Sensing and Landscape Information Systems (FeLis),

Albert-Ludwigs- University of Freiburg

Tennenbacherstr. 4
D-79106 Freiburg

Phone +49-(0)761-203-3694
Fax. +49-(0)761-203-3701

Email: ferninfo@felis.uni-freiburg.de

The Proceedings are available in digital form at: http://www.isprs.org/proceedings

Table of Contents

Session 1 - Forest and ecological applications for terrestrial Laser Scanning

Session 2 a – Poster Session 2

Session 2 b - Forest and ecological applications for terrestrial Laser Scanning

Session 2 c - Forest and ecological applications for terrestrial Laser Scanning

Session 2 d - Satelitte Based Lidar Systems and Data for Forest and environmental Applications

Session 2 e - Wood Quality Assessment with Laser

Session 3 a - Airborne Laserscanner- Ecological and landscape inventories

Session 3 b - Airborne Laserscanner-Single Tree Based Inventories

Session 3 c - Airborne Laserscanner-Area Based Inventories

Session 3 d - Airborne Laserscanner Biomass and forest fuel wood assessment

Session 3 e - Airborne Laserscanner - Ecological and landscape inventories

Session 3 f - Airborne Laserscanner -Single Tree Based Inventories

Session 3 g - Airborne Laserscanner -Area Based Inventories

Session 4 a - Multi-sensoral applications with Lidar data

Session 4 b - Multi-sensoral applications with Lidar data

Automatic Registration of Multi-scan Terrestrial LiDAR Data of Forests with Stem Features

ZHAN LI†, HUABING HUANG†, PENG GONG†‡*

gong@irsa.ac.cn

†State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Institute of Remote Sensing Applications, Chinese Academy of Sciences, and Beijing Normal University, Beijing, 100101, China

‡ Department of Environmental Science, Policy and Management, University of California, Berkeley, CA 94720-3114

Abstract

Terrestrial LiDAR has been used to make detailed assessments of forest structural parameters that are crucial to forest resource management, research of forest ecosystems and wildlife habitat. To estimate these parameters accurately by terrestrial LiDAR needs full views of trees. It requires registration to combine multi-scan terrestrial LiDAR data. However automatic registration methods suitable for forests are rarely seen but they are needed to improve the efficiency of applying terrestrial LiDAR in forests. This paper describes an automatic method to align multiple scans with stem features. Neither the placing of artificial reflectors nor the measurement of scanners' orientations is required by this method. First, stem centres near the ground namely tree locations were extracted as feature points by separating trees, estimating stem profiles and modeling stems as lines. Secondly, an iterative match algorithm based on the invariance of spatial distributions of points was introduced to search matched feature points. Thereafter these matched feature points were used to align scans together in sequence. We applied this method to 20 scans of forests collected by Riegl LMS-360i in Culai Mountain National Forest, Shandong Province, China. The results showed that the method extracted tree locations with lower vertical positions than the actual ones. The iterative match algorithm matched corresponding feature points correctly. Extracted tree locations and their correct match could give coarse alignments of scans that was acceptable as an initial input for further fine registration.

1. Introduction

Forest structural parameters, such as tree height, crown width and DBH (diameter at breast height), are essential data in forest inventory and they provide critical information for the study of forest ecosystems. Accurate forest inventory data are also crucial to forest resource management and wildlife habitat assessment (Gong et al. 1999). Recently, terrestrial LiDAR has been used to make highly efficient, accurate and detailed assessments of forest structural parameters (Watt and Donoghue 2005, Henning and Radtke 2006, Bienert et al. 2007, Maas et al. 2008, Huang et al. 2010).

The accurate extraction of forest structural parameters with terrestrial LiDAR needs full views of trees without shading zones. It requires registration to combine multiple scans of forests. Registration methods are broadly categorized as user-guided and automated (Henning and Radtke 2008). User-guided methods require manual selection and match of common feature points in overlapping scans. Such feature points are sometimes created by using artificial reflectors (Simonse et al. 2003, Huang 2008). However, placing artificial reflectors in forests is time-consuming and troublesome (Huang 2008), and sometimes even impractical (Henning and Radtke 2006) due to the complex natural environment. Automated methods rely on computers to detect and select features (points or other geometric representations) and then match them automatically. The most commonly used automated method is the iterative closest points (ICP) algorithm (Chen and Medioni 1991, Besl and McKay 1992) and its variants (Rusinkiewicz and Levoy 2001). The ICP algorithm and its variants do not cover the detection of features but match them according to their distances apart in Euclidian space (Besl and McKay 1992) or other characteristics such as points' normal distances from a destination surface, angles between normals and so on (Rusinkiewicz and Levoy 2001). ICP-based methods could give registrations with high accuracy but tend to fail if they are used without the detection and matching

of sufficiently reliable features to perform coarse registration beforehand. Many studies seek to fulfill coarse registration automatically before further fine registrations with ICP based algorithms. Mian et al. (2004) represent a surface of objects by fourth order tensors and use a correlation and verification technique to establish correspondences between surfaces that are used for coarse registration. Bae and Lichti (2008) propose a method using changes of geometric curvature and approximate normal vector of the surface formed by a point and its neighborhood to search for possible matched points. In addition some studies propose registration methods without ICP. He et al. (2005) introduced complete plane patches as features that were matched by searching an interpretation tree generated from features and then performed registration with matched features directly. Dold and Brenner (2006) extracted planar patches textured with image data and matched patches by correlations. However, most of these automated methods are developed based on buildings and engineered objects. Their applications in forests are limited due to more surface irregularity, spatial variability and occlusion than buildings and engineered objects. Recently a few studies have tried to develop automated registration methods suitable for forests. Bienert and Maas (2009) introduced an automatic method to align scans in forests via tree axes and only one artificial reflector. And the scanner has to be leveled when scanning. Henning and Radtke (2008) fulfilled registration in a forest with feature points extracted from the ground surface and estimated stem centres that have to be matched manually by visual analysis. Moreover the scanner's positions and orientations need be measured in the field for initial registration. However, manual visual analysis on forest scenes is still time-consuming and measurements of scanners' orientations may be hard to carry on in forests or sometimes not accurate enough even for coarse registration. Thus an automatic registration method based on features of forests without other manual intervention is desirable for applications of terrestrial LiDAR in forests.

In this paper, we develop a method to perform coarse registration automatically suitable for forests where artificial reflectors are not used and initial scanner positions are not known. Our objective is to detect and select feature points from scan data of forests and then match them automatically to align multi-scan terrestrial LiDAR data.

Study area and data

Study Area

We selected the Culai Mountain National Forest in Shandong Province of China (117°16′~117°20′E, 36°02′~36°17′N) as our study area. This site is about 7,164 hectares and covers nine forest management units, see Figure 1.

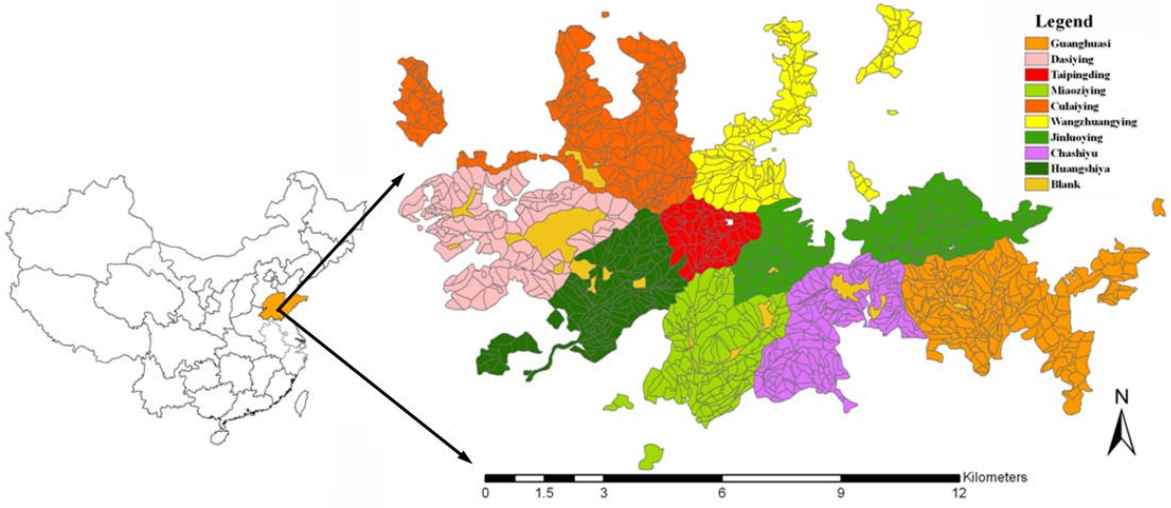


Figure 1. Location of Culai Mountain in Shandong Province, China.

A 200m by 40m subset area in Taipingding was chosen to acquire LiDAR data due to the relative flat terrain with low relief and there are only a few short slopes around it. The high resolution image of this subset area is shown in Figure 2(a) with sixteen scan stations marked with black dots. As shown in Figure 2 (b), at this

plot most trees had approximate straight stem. Few tree stems bent. The understory vegetation is sparse weed less than 30 cm in heights.

Terrestrial LiDAR data

We collected terrestrial LiDAR data in October 2007 using a Riegl LMS-360i with a data acquisition rate of 24,000Hz. The scanner measured the distance based on time-of-fight measurement with a 360° horizontal scanning range and 90° vertical scanning range. Its range of distance measurement is 1~200m. Four scan stations were chosen to acquire a full view of a 30m by 30m plot (red box in Figure 2(a)). Vertical and tilted with 90° scans as illustrated in Figure 2 (c) & (d) respectively were combined to obtain the complete vertical information of trees. The tilted scan was done four times at a 60° angular interval to achieve a complete circular scan at each station. Totally we collected 20 scans at this plot, including 4 vertical scans and 16 tilted scans. These scans were marked with scan ID as shown in Table 1. For example, “11V” denotes the vertical scan at the 11th station and “6T2” denotes the 2nd tilted scan with 90° at the 6th station. The angular scan resolution of the scanner was set 0.1°. The geographical positions of scan stations were recorded by differential GPS.

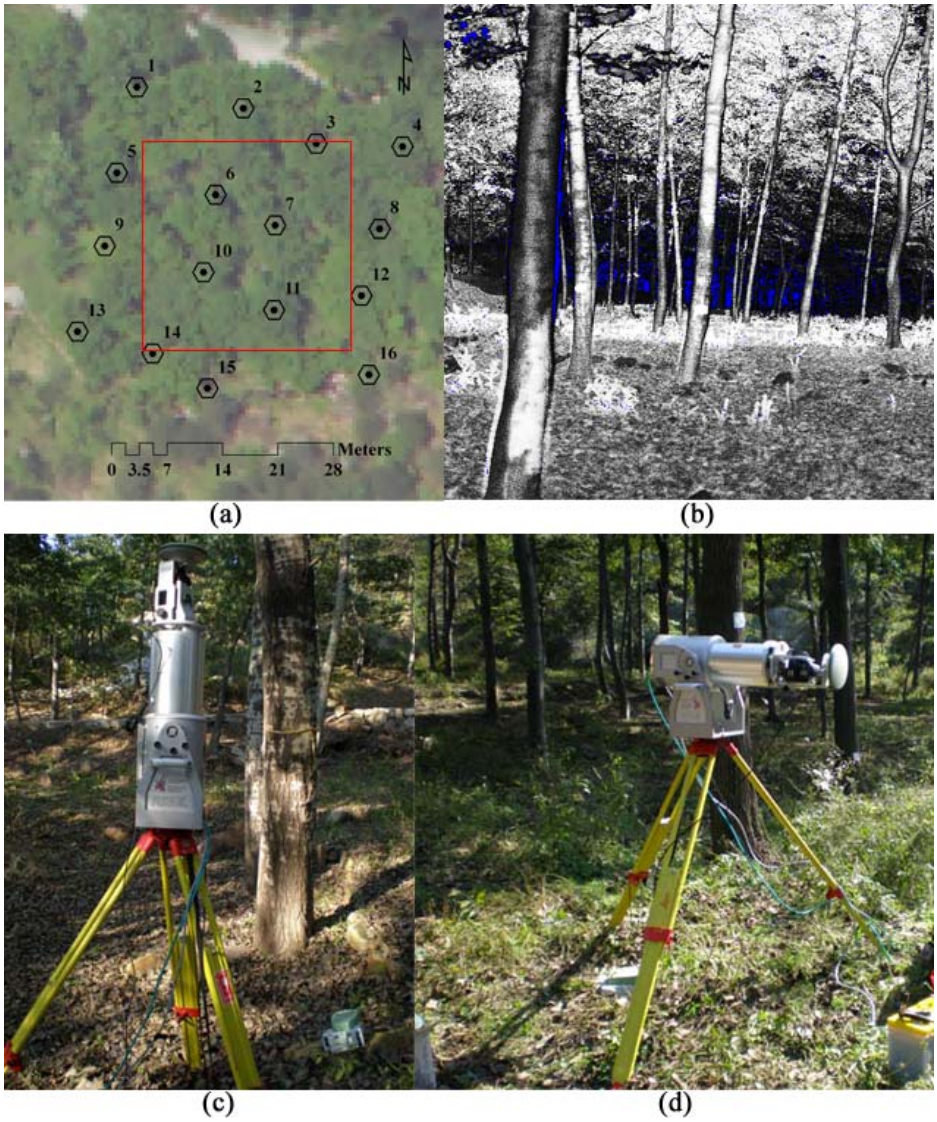


Figure 2. (a) Positions of scan stations. Black dots represent the scan positions of terrestrial LiDAR scanner. Four stations used in this study are within the red box. (b) Illustration of part of point clouds.

2. Methods

Outline of multi-scan registration

We align all scan data by in-station and between-station registration methods given by Huang *et al. (2010)*. First, tilted scans are aligned to the vertical scan within a station which is called in-station registration. Second, adjacent in-station scans are aligned to one another, which is called between-station registration. This registration method is used to align every two scans with an overlapping area and then to link them together in sequence.

In a pair of two scans, one has to be transformed into the coordinate system of the other and in this study we call the former one the source and the latter the destination. Feature points from the overlapping area are used to calculate transformations. Let $(x_{s,i}, y_{s,i}, z_{s,i})$ and $(x_{d,i}, y_{d,i}, z_{d,i})$ denote point *i*, in the source and destination, respectively. This transformation in 3-D space is described as following:

$$\begin{bmatrix} R & T \\ \vec{0} & 1 \end{bmatrix} \times \begin{bmatrix} x_{s,i} \\ y_{s,i} \\ z_{s,i} \\ 1 \end{bmatrix} = \begin{bmatrix} t(x_{s,i}) \\ t(y_{s,i}) \\ t(z_{s,i}) \\ 1 \end{bmatrix}, TM = \begin{bmatrix} R & T \\ \vec{0} & 1 \end{bmatrix} \qquad (1)$$

where $\left(t(x_{s,i}), t(y_{s,i}), t(z_{s,i})\right)$ are the transformed coordinates, R is a 3 by 3 matrix that is defined by three angles of rotation, T is a vector that relates origins of the two coordinate systems, $\vec{0}$ is a zero vector in three dimensions, TM is called a transformation matrix. We estimate TM by minimizing the error function as following:

$$MSS = \frac{1}{N}\sum_{i=1}^{N}\left[\left(x_{d,i} - t(x_{s,i})\right)^2 + \left(y_{d,i} - t(y_{s,i})\right)^2 + \left(z_{d,i} - t(z_{s,i})\right)^2\right] \qquad (2)$$

where i stands for *i*th points, MSS is the mean of squares of Euclidean distances between transformed points and their counterparts in the destination. To solve the minimization problem described by equation (2), we use the active-set optimization algorithm (MATLAB, 2010). For robustness of estimation, we empirically choose at least four pairs of matched feature points to estimate TM .

Extraction of feature points in a single scan

We extract tree locations defined by the centres of stems near the ground as feature points from each single scan with three steps as following:

1) Preprocessing scan data. We rotate the tilted scan about y axis to align stems so that they become perpendicular to XY plane in the transformed coordinate system. It is done by transforming the tilted scan via $TM_{tilted \to vertical}$ as following:

$$TM_{tilted \to vertical} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad (3)$$

Figure 3 (a) illustrates the 16 tilted scans in their original scanner's coordinate systems. Figure 3 (b) illustrates them after transformed by $TM_{tilted \rightarrow vertical}$ with showing vertical scans as well.

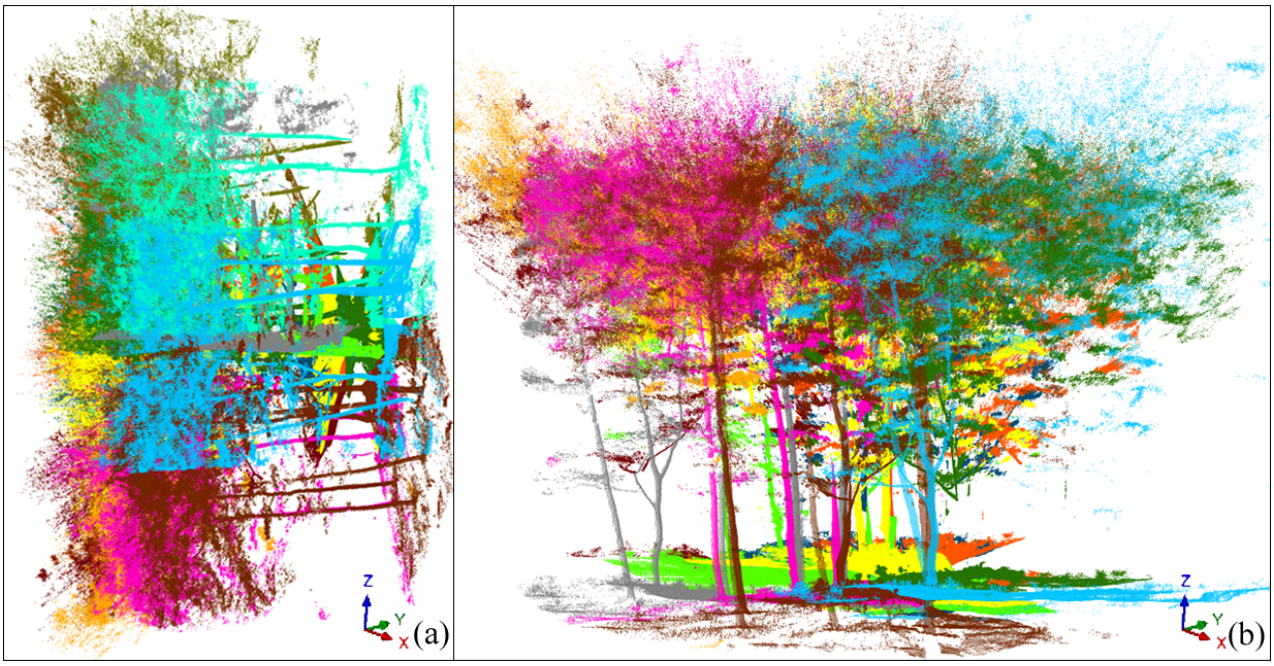


Figure 3. (a) 16 tilted scans in their original scanner's coordinate systems. (b) All 20 scans after tilted scans are transformed by

2) Detection and classification of individual tree stem. A method given by Huang *et al. (2010)* is used to identify points from stems and to mark each point with an ID number indicating the tree to which it belongs.

3) Determination of tree locations. With points of each classified individual tree, horizontal slices are intercepted with a predefined 0.1m thickness at a 0.1m height interval from the bottom point to the top. With the points of slices at different heights of each stem, we use a method given by Huang *et al. (2010)* to estimate centres of each stem. Thereafter we fit a line in 3D space with all the stem centres of a given stem by minimizing the sum of squares of distances from centres to the fitted line. During the fitting process, stem centres with distances to the fitted line greater than a given threshold (0.05m in this study) are removed as unreliable ones and then a new line is fitted again with the remaining stem centres. This fitting process is repeated until all the distances of the remaining stem centres to the fitted line are less than the threshold (0.05m). In addition no line is fitted if less than six stem centres remain. Each fitted line's intersection with the ground is used as the tree location. The ground is represented by a coarse DEM derived from scan data. First a grid is created to find the lowest point in each cell. The cell is 10m by 10m and big enough to eliminate non-ground objects such as trees and understory. Then a DEM of 1m by 1m resolution is generated by interpolation with these lowest points using the ordinary spherical Kriging method. This method for generating the DEM here is very simple and the generated DEM will be quite coarse.

Automatic matching of feature points

The automatic matching of feature points, i.e. tree locations in the source and destination, makes use of the similarity between spatial distributions of matched feature points based on the method of Bienert and Maas *(2009)*. We describe the spatial distribution of feature points with a so-called "distance pattern matrix", for instance the matrices in Figure 4.

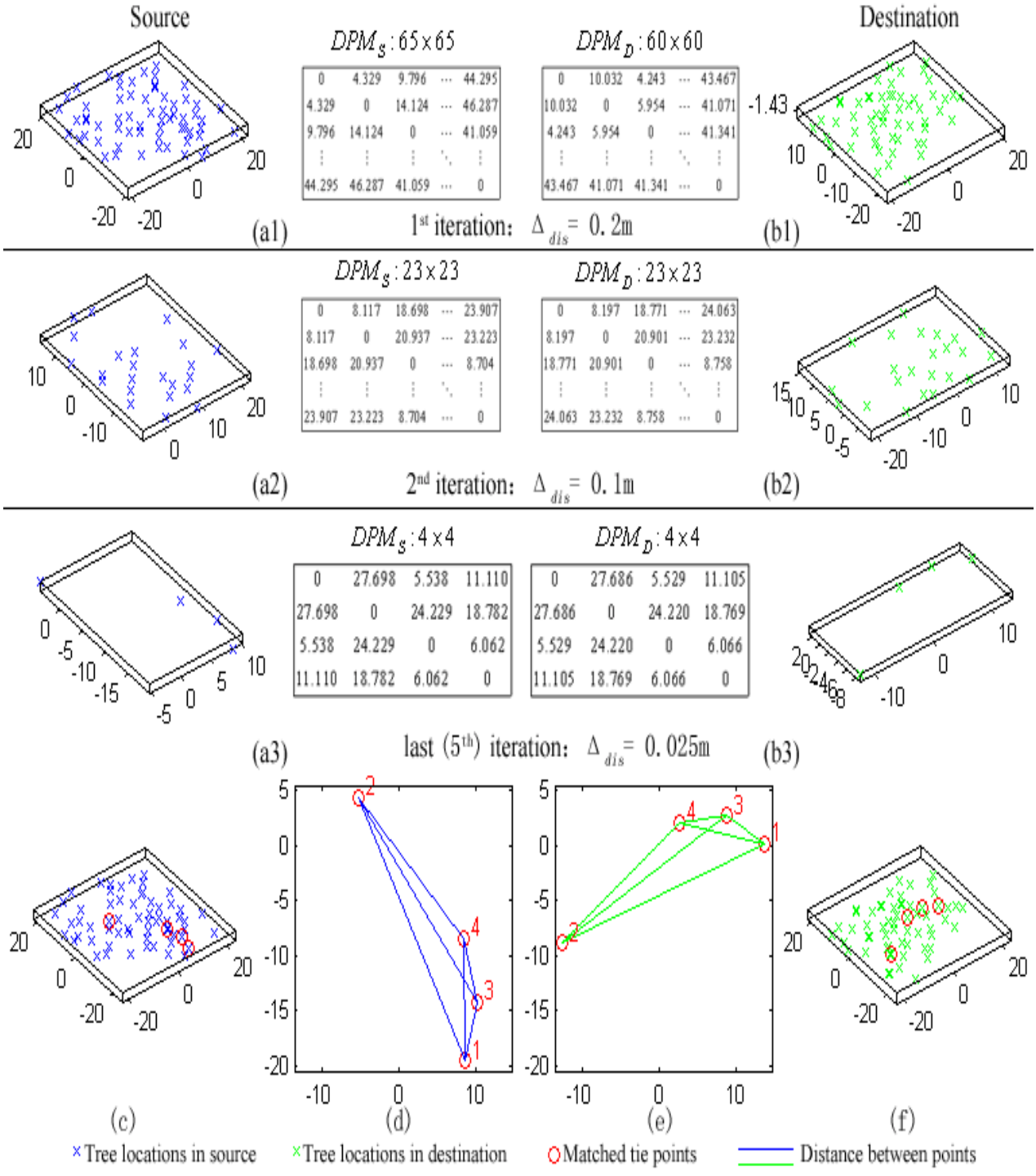


Figure 4. Automatic match of feature points. (a1, a2, a3) Feature points in S and corresponding DPM_S in 1st , 2nd and last (5th) iteration respectively. (b1, b2, b3) Feature points in D and corresponding DPM_D in 1st , 2nd and last (5th) iteration respectively. (c, f) All feature points are superposed with matched ones in S and D respectively. (d, e) Projections of matched feature points on the XY plane in S and D respectively, to view distance pattern of each feature point clearly.

The value in *i*th row, *j*th column represents the distance between *i*th and *j*th points in a point set. The *i*th row represents the distance pattern of *i*th point. Let S and D denote the feature point sets of source and destination respectively. DPM_S and DPM_D denote distance pattern matrices of S and D respectively. Ideally, the distance patterns of matched feature points in S and D are identical because the distance between two points is invariant when they are rotated and translated (Figure 4 (d) & (e); points are plotted on a plane to view clearly). However in practice, it is impossible for the distance between two points in S to be exactly the same as that of their matched points in D due to location errors. Consequently we treat two distances as 'identical' so long as their difference is less than a given threshold Δ_{dis} and call them 'matched' distances.

It is nearly impossible to ensure that there are equal numbers of tree locations in S and in D, for instance, there are 65 tree locations in S (Figure 4

Figure 4. Automatic match of feature points. (a1, a2, a3) Feature points in S and corresponding DPM_S in 1st , 2nd and last (5th) iteration respectively. (b1, b2, b3) Feature points in D and corresponding DPM_D in 1st , 2nd and last (5th) iteration respectively. (c, f) All feature points are superposed with matched ones in S and D respectively. (d, e) Projections of matched feature points on the XY plane in S and D respectively, to view distance pattern of each feature point clearly.(a1)) but 60 in D (Figure 4

Figure 4. Automatic match of feature points. (a1, a2, a3) Feature points in S and corresponding DPM_S in 1st , 2nd and last (5th) iteration respectively. (b1, b2, b3) Feature points in D and corresponding DPM_D in 1st , 2nd and last (5th) iteration respectively. (c, f) All feature points are superposed with matched ones in S and D respectively. (d, e) Projections of matched feature points on the XY plane in S and D respectively, to view distance pattern of each feature point clearly.(b1)). Some of the tree locations in S have no counterparts in D at all. Moreover, some tree locations may be false due to possible inaccuracy of extraction using the method described in section 3.2. Hence, we have to remove tree locations without counterparts or false tree locations and match correct feature points accurately. We achieve this goal by an iterative method as following.

Calculate DPM_S and DPM_D using tree locations in the inputs S and D respectively. The number of points in S and D are denoted as $Number_S$ and $Number_D$ and m, n represent two points ($m \in S, n \in D$).

Calculate $C_{m,n}$ and $MDD_{m,n}$ by equation (4) and (5).

$$\begin{aligned} SetA_{m,n} &= \left\{ (j,k) \middle| \left| DPM_S(m,j) - DPM_D(n,k) \right| < \Delta_{dis}, j \in S \text{ and } j \neq m, k \in D \text{ and } k \neq n \right\} \\ C_{m,n} &= card\left(SetA_{m,n} \right) \end{aligned} \quad (4)$$

$$MDD_{m,n} = \frac{\sum_{(j,k)} \left(\left| DPM_S(m,j) - DPM_D(n,k) \right| \right)}{C_{m,n}}, (j,k) \in SetA_{m,n} \quad (5)$$

,where $C_{m,n}$ and $MDD_{m,n}$ represent the count of matched distances under Δ_{dis} and the average difference of matched distances between the m^{th} row of DPM_S and the n^{th} row of DPM_D. For the same m, calculate $C_{m,n}$ and $MDD_{m,n}$ with all points in D, i.e. $n = 1, 2 \cdots Number_D$. A high level of similarity between distance patterns of point m and point n leads to a high value of $C_{m,n}$ and a low value of $MDD_{m,n}$ for the same inputs S and D. So we find point N_m in D such that C_{m,N_m} is the maximum of $C_{m,n}(n \in D)$. If more than one point is found, we choose the one with the lowest value of MDD_{m,N_m}. If there is still more than one point remaining, we remove point m from S. Otherwise, point m and point N_m are matched. After finding the matched points N_m from D for all points in S, i.e. $m = 1, 2 \cdots Number_S$, we obtain the set of matched feature point pairs denoted by $\{(m, N_m) | m \in S, N_m \in D\}$.

If some matched point pairs in $\{(m, N_m) | m \in S, N_m \in D\}$ have duplicate N_m, it indicates that multiple points in S corresponds to the same point in D which is unexpected. Thus we remove these point pairs from $\{(m, N_m) | m \in S, N_m \in D\}$ and also remove relevant points m and N_m from S and D respectively. If less than four point pairs are left, we relax the threshold of distance difference to $2\Delta_{dis}$ and go back to redo step 1 to step 3.

If the number of matched point pairs in $\{(m, N_m) | m \in S, N_m \in D\}$ is less than eight but no less than four and no point pair is removed in step 3, $\{(m, N_m) | m \in S, N_m \in D\}$ is output as the final matched tie points. If not, we generate new inputs S and D with points m (constituting S) and N_m (constituting D) respectively from half of the point pairs in $\{(m, N_m) | m \in S, N_m \in D\}$ with highest C_{m,N_m}. Then we update the threshold of distance difference to $\Delta_{dis}/2$. Using the new feature points and threshold, we repeat step 1 to step 4. If the condition of termination is not achieved when the count of iterations reaches a given maximum, no matched feature points are found.

Figure 4 illustrates this iterative process by giving changes of DPM_S, DPM_D and Δ_{dis} in 1st, 2nd and last (5th in this case) iteration.

Registration with matched feature points

Matched feature point pairs which are identified by the method of section 3.3 are used to calculate transformation matrix TM via the algorithm of section 3.1.

On the other hand, considering that the DEM is quite coarse, the Z values of these feature points are possibly inaccurate. In addition, stems in our study plot are generally perpendicular to XY plane. Hence we try performing registration with projections of feature points on the XY plane. That is to say we use $\{(x_{s,i}, y_{s,i}, 0) \leftrightarrow (x_{d,i}, y_{d,i}, 0)\}$ denoted by 'Prj' instead of $\{(x_{s,i}, y_{s,i}, z_{s,i}) \leftrightarrow (x_{d,i}, y_{d,i}, z_{d,i})\}$ denoted by 'NoPrj' in registration. After doing in-station and between-station registrations, we align all scans to the coordinate system of scan 6V in a sequential manner. For example, $TM_{7T1 \to 6V}$ denoting transformation matrix used to align scan 7T1 to scan 6V is calculated as the following:

$$TM_{7T1 \to 6V} = TM_{7V \to 6V} \times TM_{7T1 \to 7V} \quad (6)$$

3. Results and Discussion

Extraction of feature points in single scan

Table 1. Number of feature points in each scan, iteration counts, initial and final Δ_{dis} of automatic match registration, number of matched feature point pairs, square root of MSS

No.[1]	Source		Destination		# iterations	Initial Δ_{dis} (meter)	Final Δ_{dis} (meter)	# MFPP[4]	Square root of MSS (meter)	
	Scan ID[2]	# FP[3]	Scan ID[2]	# FP[3]					Prj[5]	NoPrj[5]
1	7T1	43	7V	65	4	0.2	0.025	4	0.010	0.065
2	7T2	50	7V	65	4	0.2	0.025	4	0.030	0.107
3	7T3	44	7V	65	5	0.2	0.05	4	0.035	0.063
4	7T4	45	7V	65	3	0.2	0.05	7	0.052	0.137
5	6T1	37	6V	60	5	0.2	0.05	4	0.063	0.098
6	6T2	44	6V	60	3	0.2	0.05	7	0.035	0.068
7	6T3	48	6V	60	4	0.2	0.025	4	0.018	0.043
8	6T4	38	6V	60	5	0.2	0.05	4	0.035	0.140
9	10T1	37	10V	63	3	0.2	0.05	5	0.051	0.056
10	10T2	48	10V	63	5	0.2	0.05	4	0.032	0.048
11	10T3	40	10V	63	4	0.2	0.05	6	0.074	0.081
12	10T4	41	10V	63	4	0.2	0.05	5	0.039	0.054
13	11T1	41	11V	62	4	0.2	0.05	6	0.076	0.172
14	11T2	46	11V	62	4	0.2	0.05	6	0.053	0.085
15	11T3	43	11V	62	5	0.2	0.05	4	0.056	0.057
16	11T4	41	11V	62	3	0.2	0.05	7	0.051	0.099
17	7V	65	6V	60	5	0.2	0.025	4	0.011	0.026
18	10V	63	6V	60	4	0.2	0.025	5	0.023	0.140
19	11V	62	6V	60	5	0.2	0.05	4	0.037	0.034
Minimum		37		60	3	0.2	0.025	4	0.010	0.026
Maximum		65		65	5	0.2	0.050	7	0.076	0.172
Mean		46.1		62.1	4.2	0.2	0.043	4.9	0.041	0.083

[1]Mark the 19 experiments of automatic match of feature points and registration.
[2]Mark scans at different stations and with different orientations.
[3]FP denotes Feature Points.
[4]MFPP denotes Matched Feature Point Pairs.
[5]Prj denotes registration using projections on XY plane of feature points while NoPrj denotes registration using feature points directly.

Table 1 lists the number of tree locations (feature points) extracted from each scan. A typical part of point cloud in the scan 6V (orthogonal view in Figure 5 (a), perspective view in Figure 5 (b)) is shown in Figure 5 (c) (orthogonal view) and Figure 5 (d) (perspective view). Most points of stems (showed with non-grey colors) were identified by the method of section 3.2. False stem centres are removed as unreliable (black squares in Figure 5 (d)) by the line fitting method of section 3.2.

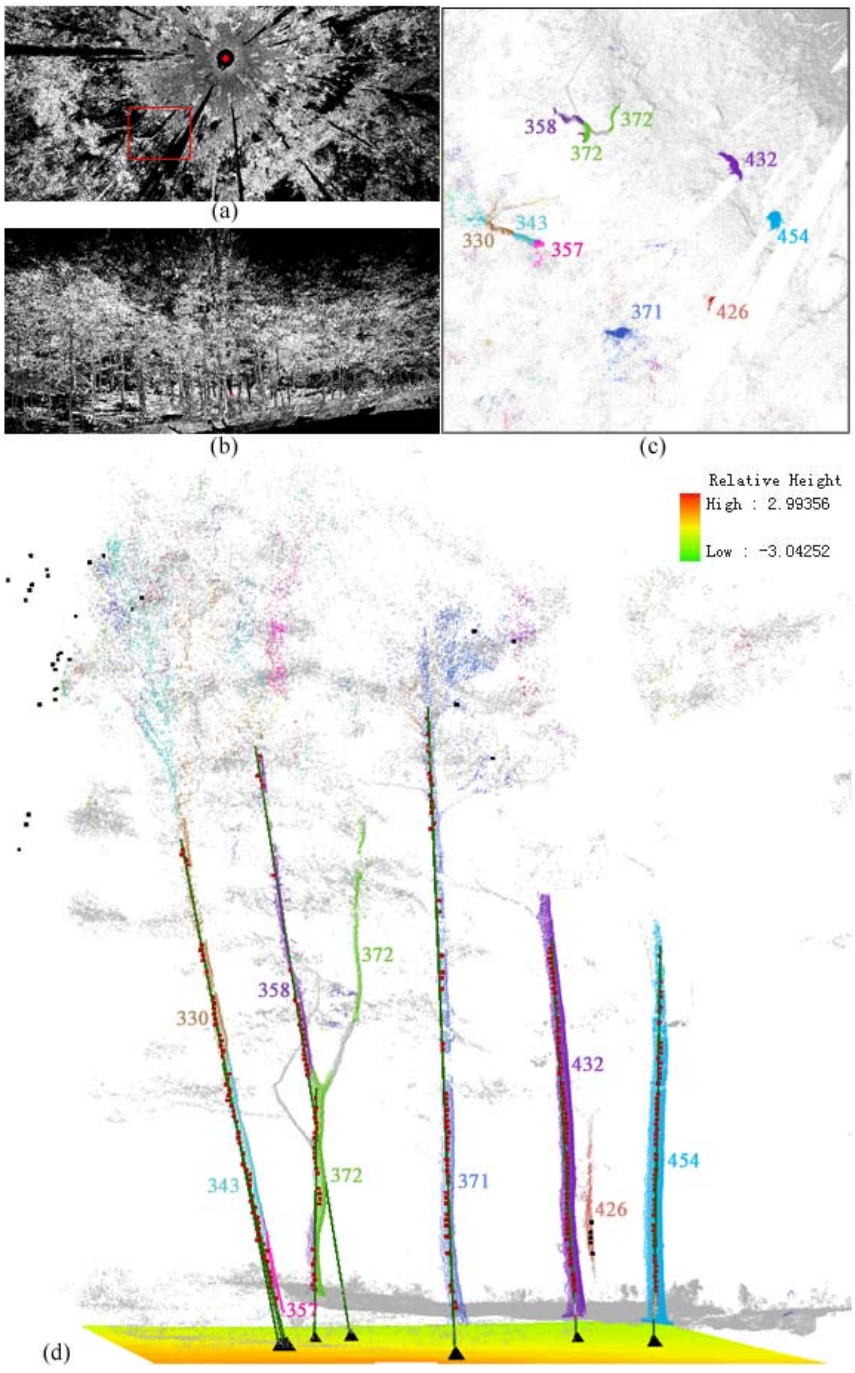


Figure 5. Extraction of feature points. (a) Orthogonal view of the scan 6V. (b) Perspective view of the scan 6V. (c) Orthogonal view of a part of the scan 6V marked by red box in (a). Different colors of points represent classified different stems. The numbers next to stems are the ID of stems. (d) Perspective view of part of the scan 6V. Different colors of points represent classified stems. The numbers next to stems are the ID of stems. The gray represent points classified as non-stem. The colorful surface under the point cloud is the DEM. Red and black cubes respectively represent reliable and unreliable stem centres. Dark green lines are the fitted lines. Black tetrahedrons represent tree locations estimated by the method of section 3.2

Tree locations determined by the intersections between fitted lines (dark green lines in Figure 5 (d)) and the DEM (colorful surface in Figure 5 (d)) were lower in vertical positions than the true locations due to the inaccurate DEM being lower than the actual ground (Figure 5 (d)). Selection of lowest points simply in each 10m by 10m cell as positions of the ground led to a lower DEM than the actual. A more accurate DEM was needed to give more accurate tree locations.

Automatic matching of feature points

We did 19 experiments of automatic match as listed inTable 1. Iteration counts were at most 5. Changes of eight parameters during iterations for all 19 experiments were analyzed to see how our algorithm works (Figure 6). The eight parameters were defined as following:

a) $Number_S$, number of input feature points in source.

b) $Number_D$, number of input feature points in destination.

c) $Number_{MFPP} = card\left(\left\{(m, N_m) \middle| m \in S, N_m \in D\right\}\right)$, number of matched feature point pairs.

d) $MeanC_{m,N_m} = \frac{\sum_{(m,N_m)} C_{m,N_m}}{Number_{MFPP}}, (m, N_m) \in \left\{(m, N_m) \middle| m \in S, N_m \in D\right\}$, average count of matched distances between distance patterns of points in S and D.

e) Δ_{dis}, threshold to determine matched distances.

f) $MeanMDD_{m,N_m} = \frac{\sum_{(m,N_m)} MDD_{m,N_m}}{Number_{MFPP}}, (m, N_m) \in \left\{(m, N_m) \middle| m \in S, N_m \in D\right\}$, average distance differences between distance patterns of points in S and D.

g) $Ratio_{MFPP} = \frac{Number_{MFPP}}{Number_S}$.

h) $Ratio_C = \frac{MeanC_{m,N_m}}{Number_{MFPP}}$.

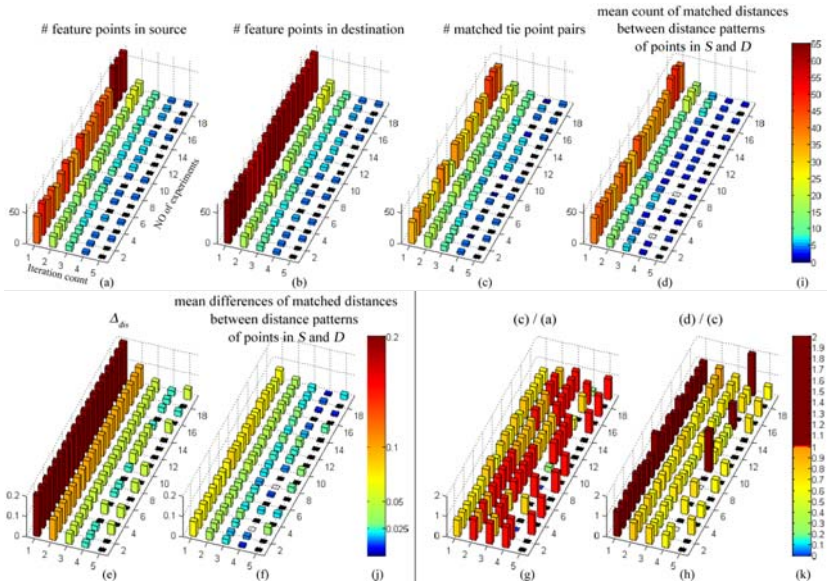


Figure 6. Parameters in each iteration of 19 experiments of automatic match. Black bars represent no iteration. Empty bars in (d), (f) and (h) represent no value due to no matched point pair at that iteration. (a) $Number_S$. *(b)* $Number_D$. *(c)* $Number_{MFPP}$. *(d)* $MeanC_{m,N_m}$. *(e)* Δ_{dis}. *(f)* $MeanMDD_{m,N_m}$. *(g)* $Ratio_{MFPP}$. *(h)* $Ratio_C$. *(i) Color bar of count, for (a), (b), (c) and (d). (j) Color bar of distance, for (e) and (f). (k) Color bar of ratio, for (g) and (h).*

Although $Number_S$, $Number_D$ and $Number_{MFPP}$ (Figure 6 (a), (b) & (c)) decreased with iterations, $Ratio_{MFPP}$ (Figure 6 (g)) generally increased to one (bright red bar) in the last iteration for all 19 experiments. $Ratio_{MFPP}$ told us whether there existed false feature points without counterparts. $Ratio_{MFPP} = 1$ meant an expected result that false feature points without counterparts were removed and all points in the input S found correct matched points in D. Note that $Ratio_{MFPP}$ was much less than one in the 3rd iteration of No. 3, 8, 10, 15 and 19 experiments with $\Delta_{dis} = 0.025\text{m}$ while $Ratio_{MFPP}$ increased to one in next iteration with $\Delta_{dis} = 0.05\text{m}$. And the matched point pairs in the last iteration for these five experiments were found to be correct (not shown here). We inferred that 0.025m was too strict for these inputs to give enough matched point pairs and our algorithm adjusted Δ_{dis} to a larger value.

$MeanC_{m,N_m}$ and $MeanMDD_{m,N_m}$ described the mean level of similarity between distance patterns of matched feature points in S and D. A value of $MeanC_{m,N_m}$ closer to $Number_{MFPP}$, i.e. $Ratio_C$ closer to one and smaller $MeanMDD_{m,N_m}$, indicated a higher level of similarity and thus a better match of feature points. $Ratio_C$ mostly came to 0.75 in the last iterations (Figure 6 (h)). It indicated that distance patterns of points in S and D did not match with each other completely due to errors of point positions. Figure 6 (h) showed some values even greater than one (dark red bar) because a larger value of Δ_{dis} in those iterations led to too many matched distances, some of which were falsely matched by our algorithm. Consequently this caused abnormally larger $MeanC_{m,N_m}$ than $Number_{MFPP}$. In addition, too small Δ_{dis} led to not enough matched point pairs, i.e. too small $Number_{MFPP}$ that also caused an $Ratio_C$ abnormally greater than one, such in No. 10, 15 and 18 experiments. The two situations both indicated possibly incorrect matches of feature points in S and D. On the other hand, $MeanMDD_{m,N_m}$ generally decreased with more iterations, which indicates that differences of matched distances became smaller. It implied more accurate positions of matched feature points.

Δ_{dis} was adjusted during iterations automatically by the matching algorithm. The algorithm should select as low Δ_{dis} as possible so long as it could give enough matched feature points pairs with this low Δ_{dis} and found those pairs with as low $MeanMDD_{m,N_m}$ as possible.

Registration with matched feature points

We did 19 experiments of registrations, including in-station registrations of the four scan stations (No.1 ~ No.16 in Table 1) and between-station registrations (No.17 ~ No.19 in Table 1) using the matched feature point pairs found by the automatic match algorithm. Two manners of using feature points i.e. denoted by 'Prj' and denoted by 'NoPrj', gave different results of registrations illustrated in Figure 7 (a) and (b). Square root of calculated by equation (2) were listed in Table 1 and shown in Figure 8 with final and in last iterations for 19 experiments. As we could see from Figure 7, registration result by Prj was better than by NoPrj and root of by Prj were smaller than that by NoPrj (Table 1, Figure 8). It was because Prj did not use inaccurate Z values of feature points. However, the two manners of using feature points both only gave coarse registration result. Common stems and braches aligned together generally but not accurately (Figure 7). The ground did not align together, especially along the Z axis. Square root of by Prj and NoPrj were 0.041m and 0.083m respectively that were sufficient for coarse registration but not for fine registration. Further fine registration either automatically or manually could be done on the basis of this result.

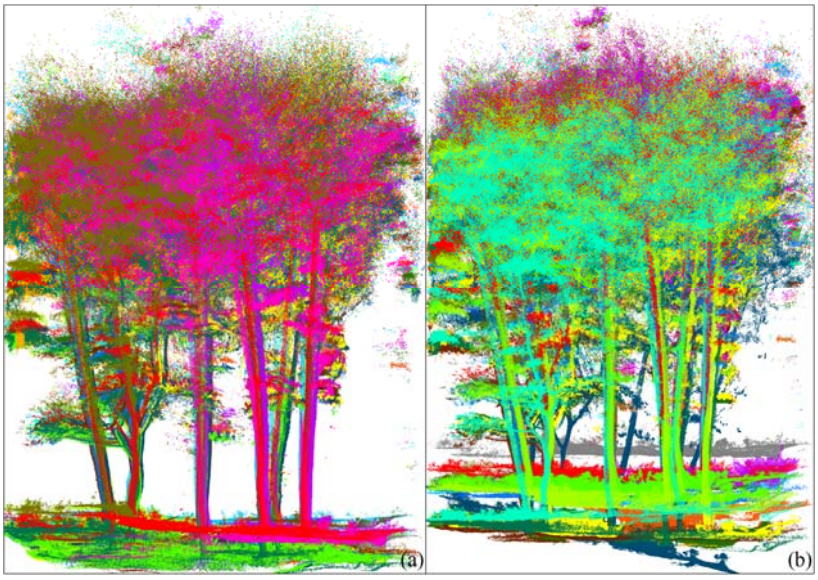


Figure 7. (a) All 20 scans after registration using feature points in the manner 'Prj'. (b) All 20 scans after registration using feature points in the manner 'NoPrj'.

In addition, we compared square root of MSS, final Δ_{dis} and $MeanMDD_{m,N_m}$ (Figure 8) and found that low root of MSS such as No. 1, 7, 14 and 17 corresponded to low Δ_{dis} and $MeanMDD_{m,N_m}$. Low Δ_{dis} and low $MeanMDD_{m,N_m}$ both implied small differences of matched distances between distance patterns of points, which indicated minor error of feature point positions. Feature points with more accurate positions gave better registration.

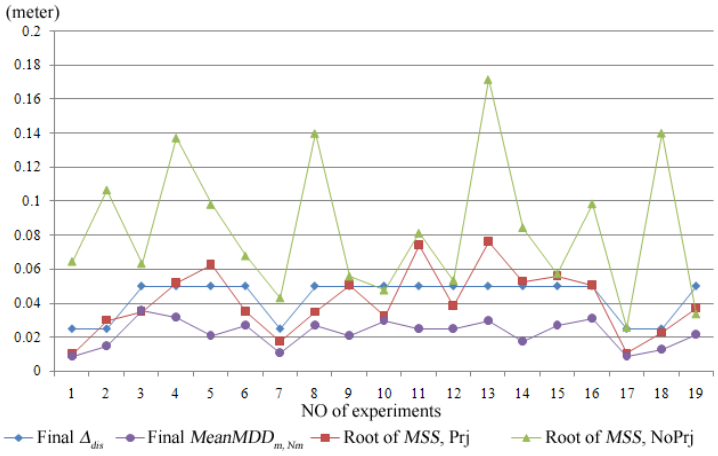


Figure 8 Line chart of square root of MSS, final Δ_{dis} and $MeanMDD_{m,N_m}$ in last iterations for 19 experiments.

4. Conclusion

An automatic registration method suitable for forests is described in this paper. It is shown that stem centres near ground, namely tree locations could be extracted and used as reliable feature points for forests. An iterative match algorithm with varying threshold is used to search matched feature points automatically based on their spatial distribution described by distance pattern matrix. It can be used to align scans without using artificial reflectors or measuring scanners' orientations. This will improve the efficiency of applying terrestrial LiDAR in forests.

It should be noted that the use of inaccurate DEM created from point clouds limits the accuracy of tree locations and registration results. A DEM with high accuracy would be helpful to improve registration. Although the automatic method here gives a coarse registration, it is acceptable to be the initial input for further fine registration. Completely automatic methods for further fine registration of scans in forests should be sought in future.

Acknowledgements

The study was supported by NSFC *'Extraction of Forest Structure Parameters Using Ground-based LiDAR'*. The authors would thank Professor Arthur Cracknell for his help with this paper.

References

BAE, K.-H., and LICHTI, D. D., 2008, A method for automated registration of unorganised point clouds. ISPRS Journal of Photogrammetry &Remote Sensing, 63, pp. 36-54.

BESL, P. J., and MCKAY, N. D., 1992, A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14, pp. 239-256.

BIENERT, A., SCHELLER, S., KEANE, E., MOHAN, F., and NUGENT, C., 2007, Tree detection and diameter estimations by analysis of forest terrestrial laserscanner point clouds. In Proceedings of ISPRS Workshop on Laser scanning 2007 and SilviLaser 2007, 12-14 September 2007, Espoo, Finland, pp. 50-55.

BIENERT, A., and MAAS, H.-G., 2009, Methods for the automatic geometric registration of terrestrial laser scanner point clouds in forest stands. In Proceedings of ISPRS Workshop on Laser scanning 2009, 1-2 September 2009, Paris, France, pp. 93-98.

CHEN, Y., and MEDIONI, G., 1991, Object modeling by registration of multiple range images. In Proceedings of the 1991 IEEE International Conference on Robotics and Automation, 9-11 April 1991, Sacramento, CA, pp. 2724 - 2729.

DOLD, C., and BRENNER, C., 2006, Registration of terrestrial laser scanning data using planar patches and image data. In ISPRS Commission V Symposium 'Image Engineering and Vision Metrology', 25-27 September 2006, Dresden, Germany, pp. 78-83.

GONG, P., BIGING, G., LEE, S., MEI, X., SHENG, Y., PU, R., XU, B., SCHWARZR, K., and MOSTAFA, M., 1999, Photo ecometrics for forest inventory. Geographic Information Sciences, 5, pp. 9-14.

HE, W., MA WEI, and ZHA, H., 2005, Automatic registration of range images based on correspondence of complete plane patches. In Fifth International Conference on 3-D Digital Imaging and Modeling, 13-16 June 2005, Ottawa, Canada, pp. 470-475.

HENNING, J., and RADTKE, P., 2006, Detailed stem measurements of standing trees from ground-based scanning lidar. Forest Science, 52, pp. 67-80.

HENNING, J. G., and RADTKE, P. J., 2008, Multiview range-image registration for forested scenes using explicitly-matched tie points estimated from natural surfaces. ISPRS Journal of Photogrammetry and Remote Sensing, 63, pp. 68-83.

HUANG, H., 2008, Forest structural parameters extraction using LiDAR [unpublished Ph. D. dissertation thesis]: Graduate school of the Chinese Academy Sciences, Beijing, 158 pp.

HUANG, H., LI, Z., GONG, P., CHENG, X., CLINTON, N., CAO, C., NI, W., and WANG, L., 2010, Forest structural paramter extraction using terrestrial LiDAR. Photogrammetric Engineering & Remote Sensing, In press.

MAAS, H., BIENERT, A., SCHELLER, S., and KEANE, E., 2008, Automatic forest inventory parameter determination from terrestrial laser scanner data. International Journal of Remote Sensing, 29, pp. 1579-1593.

MATLAB, http://www.mathworks.com/access/helpdesk/help/toolbox/optim/ug/fmincon.html, accessed in May 2010.

MIAN, A., BENNAMOUN, M., and OWENS, R., 2004, Matching tensors for automatic correspondence and registration. In 8th European Conference on Computer Vision, 11-14 May, 2004, Prague, Czech, pp. 495-505.

RUSINKIEWICZ, S., and LEVOY, M., 2001, Efficient Variants of the ICP Algorithm. In Third International Conference on 3D Digital Imaging and Modeling (3DIM '01), May 28 - June 1, 2001, Québec City, Canada, pp. 145-152.

SIMONSE, M., ASCHOFF, T., SPIECKER, H., and THIES, M., 2003, Automatic determination of forest inventory parameters using terrestrial laser scanning. In Proceedings of the ScanLaser Scientific Workshop on Airborne Laser Scanning of Forests, Umea, Sweden, pp. 251-257.

WATT, P., and DONOGHUE, D., 2005, Measuring forest structure with terrestrial laser scanning. International Journal of Remote Sensing, 26, pp. 1437-1446.

Volume calculations of single trees based on terrestrial laser scanning

GÉZA KIRÁLY† and GÁBOR BROLLY†*

kiraly.geza@emk.nyme.hu

†Department of Surveying and Remote Sensing, University of West Hungary, Sopron, Hungary

Abstract

Terrestrial laser scanning is an efficient tool for collecting metric data even in forests. While numerous papers are about to derive classical forestry and single tree parameters, such as DBH, tree height, etc, the method records much more information, it records the actual state of the forest. We would like to exploit the possibilities better and compare the different methods through volume calculations of single trees.

An overview on the stem metrics can be derived from terrestrial laser scanning is presented and discussed. The most common volume calculation formulas, from the single cylinder through truncated cones and tapering numbers and functions to the more complex free-form polygons, correlating to the metric parameters are presented.

This research was carried out on a near natural complex and uneven aged forests. The following tree parameters were determined for group of species: DBH, diameter at crown base, height of crown base, tapering, diameter of slices, free-form polygons of slices. The different volume calculations are presented and evaluated afterwards. The accuracy and the complexity of the models are discussed in detail. The visualisations of the different models are also presented, and give a very demonstrative way to compare them easily.

The results can help to choose the right method for volume calculations, assortments on standing trees, and visualisation purposes.

1. Introduction

Terrestrial laser scanning (TLS) has widely enlarged the possibilities on forest mensuration. Some researches utilises the more accurate determination of the classical parameters, such as diameter at breast height (DBH), tree and stand height, etc.; and some others define some novel parameters which were unavailable before. The volume calculations are one of the very classical determinations, and primarily based on the DBH and the height of the single tree. There are several studies on (semi-) automatic estimations of DBH. Hopkinson *et al*. (2004) demonstrated the potential of TLS in semi-automatic circle-based DBH estimations. Aschoff and Spieker (2004) used a pattern recognition based on Hough-transformation, and calculated a lot of important parameters on stems' shape. Pfeifer and Winterhalder (2004) applied free-form curves, Wezyk *et al*. used polygons for modelling of the cross section of the trees. There have been also several researches on the tree height as well, but hardly any research deal with complex volume calculation based on TLS.

The methods are presented here, try to define the crucial parameters for the volume calculation such as the DBH, the height of the monopodial trunk and the form number of it, with more reliability and accuracies and combining the results into volume calculations.

Stem metrics

Defining the volume of a stem requires horizontal and vertical measures of it. The most important horizontal measures of a stem are: 1) diameter (*d*), 2) perimeter or circumference (*c*) and 3) basal area (*g*). The most important vertical measure is the length of the tree or height (*h*).

Modelling is essential for volume calculation of a standing tree. The horizontal extension, the cross section of a stem can be modelled by several ways. The circle is the most commonly used model. The ellipse is also common, but usually does not give much more accuracy against the extra parameter. The real cross section of a tree can be very far from the circular or even from the elliptical one, which can be

described by a free form polygon (FFP hereafter) easily (Figure 1Figure 1), can be defined by 1) polar or by 2) Cartesian coordinates (Figure 1).

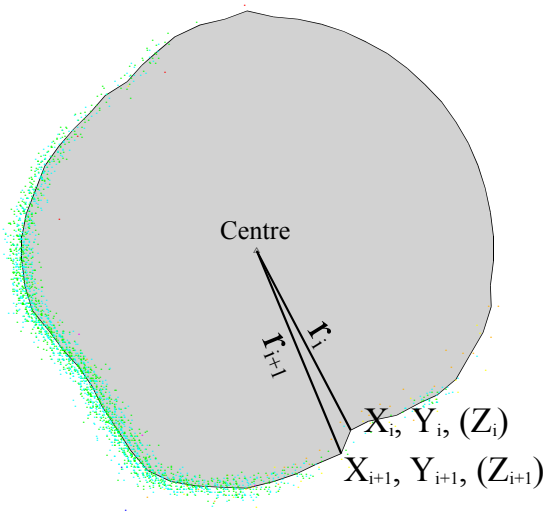


Figure 1: The Free Form Polygon and its possible descriptions

The area of the cross section is the most important horizontal measure for volume calculations, which is usually referred as basal area (*g*). Calculation of the basal area in any of the above mentioned models is well known

(equation ($g = \frac{d^2 \cdot \pi}{4}$ $g = \frac{d_1 \cdot d_2 \cdot \pi}{4}$ $g = \pi \cdot \sum_{i=1}^{n} \frac{r_i^2}{n}$ (1))

$$g = \frac{d^2 \cdot \pi}{4} \qquad g = \frac{d_1 \cdot d_2 \cdot \pi}{4} \qquad g = \pi \cdot \sum_{i=1}^{n} \frac{r_i^2}{n} \qquad (1)$$

cylinder ellipse FFP

The area of a free form polygon can be calculated from the Cartesian coordinates, too. This kind of area calculation is the most common in GIS software (equation ($g = \sum_{i=1}^{n} \frac{(x_i + x_{i-1}) \cdot (y_{i+1} - y_i)}{2}$ ()).

$$g = \sum_{i=1}^{n} \frac{(x_i + x_{i-1}) \cdot (y_{i+1} - y_i)}{2} \qquad (2))$$

The most common group for trunk modelling is the conoids. Accepting the circular symmetry of the trunk, a conoid is a kind of body of revolution, where the relative area of cross section can be expressed by a power function of the relative length measured from the top (equation $\frac{g(h)}{g_0} = \left(\frac{H-h}{H}\right)^r$ (3))

$$\frac{g(h)}{g_0} = \left(\frac{H-h}{H}\right)^r \qquad (3)$$

The most common conoids applied for trunk modelling according to index *r* are follows:

The applications of the form numbers (*f*) are very common in volume calculations. We used the special form number for breast height ($f_{1.3}$), which expresses the ratio of the trunks' volume to the volume of the base cylinder, which is a cylinder with the diameter of DBH and with the length of the height in this case. The form numbers of the conoids are also indicated in Table 1.

Table 1: The overview of the conoids

type of conoid	*r*	$f_{1.3}$	*Volume*
		$\frac{1}{r+1}$	
Cylinder	0	$\frac{1}{1}$	$v = d_0^{\,2} \cdot h \cdot \frac{\pi}{4}$
Paraboloid	1	$\frac{1}{2}$	$v = \frac{1}{2}\left(d_0^{\,2} + d_t^{\,2}\right) \cdot h \cdot \frac{\pi}{4}$
Cone	2	$\frac{1}{3}$	$v = \frac{1}{3}\left(d_0^{\,2} + d_0 d_t + d_t^{\,2}\right) \cdot h \cdot \frac{\pi}{4}$
Neiloid	3	$\frac{1}{4}$	$v = \frac{1}{4}\left(d_0^{\,2} + d_0^{\frac{4}{3}} d_t^{\frac{2}{3}} + d_0^{\frac{2}{3}} d_t^{\frac{4}{3}} + d_t^{\,2}\right) \cdot h$

2. Materials and methods

The study area and its surveying

The study area is situated in the Pilis Mountains, Hungary. The forest has been treated based on the Pro Silva rules here since 1999. This was the first of this kind area in the country, which was certified as demonstrational area (category 'B' – PSB) by the Pro Silva Hungary in 2001.

The laser scanning campaign was performed in April, 2009. The point cloud was surveyed from 38 scanning position by a Riegl LMS-Z420i terrestrial laser scanner with two different tilt angle, resulted approximately 360 million points. The co-registration and transformation of the points were done by the surveying company with the utilisation of a diploma work (Kiss 2009), which founded the geometric base for the laser scanning.

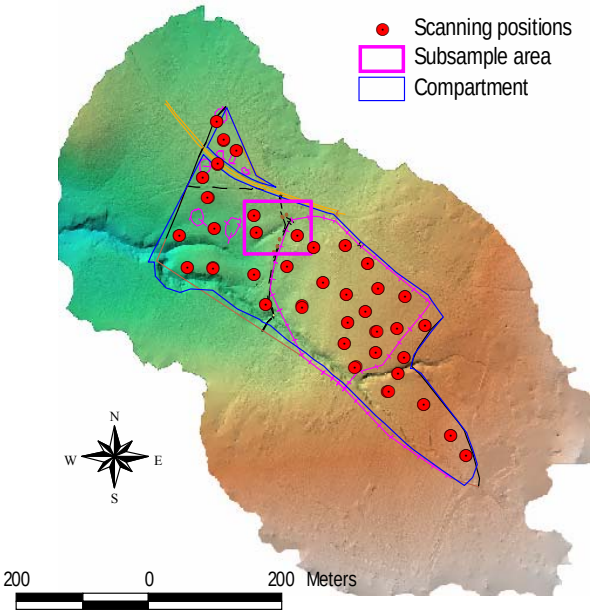


Figure 2: The study area with the scanning positions and the subsample area

The pre-processing

The digital terrain model of the study area was calculated according to Király and Brolly (2007), supplemented by manual filtering. The points at breast height (1.25-1.35 m) were queried for the stem's location. Our previous 'crescent moon' method (Király and Brolly 2007) has been developed further for creating a semi-automatic stem map. The areas without understory vegetation have to be delineated first, where the stems were identified automatically, and the best scanning position was determined, too, in the case multiply scans. The trees within the understory were identified manually approximately, and the algorithm determined the exact location and calculated the diameter.

Determination of the basal area

The basal areas were calculated for the whole area (9.5 ha) at breast height (1.3 m) in four different ways:

1. crescent moon
2. free form polygon determination
3. least squares circle fitting
4. least squares cylinder fitting

The first one was circles fitted by the semi-automatic crescent moon method.

A new method was developed for determination of the free form polygons. The circle – which was determined by crescent moon method first – was the starting point for the FFP. The points within a certain tolerance were selected, counted and sorted by the azimuth from the centre. The tolerances for azimuth-step and averaging were calculated based upon the number and angular distribution of the points. The radial distance of a group of points is continuously recalculated and smoothed, resulting an incessant and event polygon (see Figure 3.

The DBH were calculated from the area of the polygon backwards here. The centroid of the FFP were taken as the improved location of the tree.

The third one was a least squares circle fitting based on the locations and starting DBH from the results of the FFP method on the points above breast height (1.25-1.35 m). The cylinders were fitted to the points from 0.8 m to 1.8 m by least squares according to Brolly and Király (2009). The points were selected by the results of the FFP in this fourth method, too.

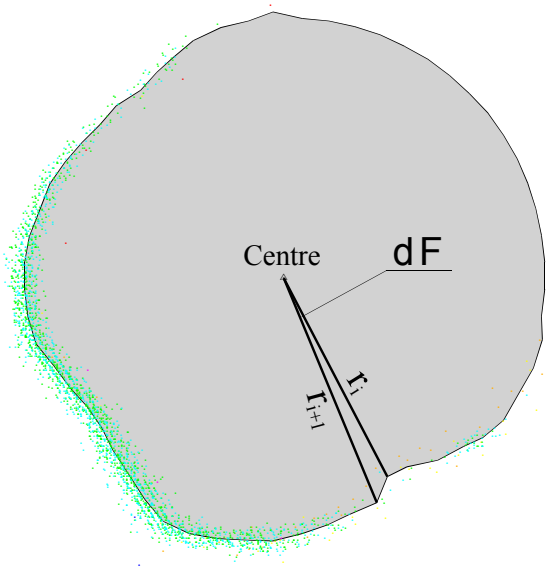


Figure 3: The mechanism of the FFP algorithm

Determination of the heights

The lengths of the monopodial trunk were calculated for the subsample area (see Figure 2), which is 100 m by 80 m large with exactly 100 trees and nicely representing the whole area.

A tracing algorithm based on the FFP method was developed to trace the trunk from the root up to the first or second significant arm. The basic idea was very similar to our previous method (Király and Brolly 2007), to slice the trunk, but we used the free form polygons here and watched the shape of it, to determine the junctions. The changes of the polygons' area and its shape factor were the essential parameters to control here. The form number (*f*) was also continuously monitored to separate the trunk to parts with different properties. The current section was signed by the *f*, which was derived from the nearest integer of *r.*

Modelling the trunks and calculating the volumes

Constructing the timber tree from the slices requires modelling. Modelling is the way to create the model of the trunk from the components previously generated and this makes the volume calculation available. The volumes have been calculated based on all four types of conoid (see Table 1) and all four types of basal area calculations (see Chapter 3.3). The volume calculations were referenced based upon the previously determined form number (*f*) of the current section, and the free from polygon area determinations and the appropriate functions, can be seen in Table 1.

3. Results and discussions

The basal area

There were 1669 trees identified in the whole compartment, which means an 176 trees/ha average. The basal areas – calculated with all four methods – of the compartment can be seen in table 2

Table 1: The comparison of the basal area calculations

G (m2)	CM	**FFP**	Circle	Cylinder
Count	118	**118**	118	118
Average	0.2365	**0.2091**	0.2066	0.2092
Sum	**25.0732**	**24.6718**	**24.3788**	**24.6909**
Minimum	0.00769769	**0.00781201**	0.01056832	0.00723823
Maximum	0.55022561	**0.52129926**	0.48274969	0.46566257
Std	0.1043	**0.0931**	0.0933	0.0976
var	0.0109	**0.0087**	0.0087	0.0095

The areas of the FFPs were used as reference. A very typical example for the deviations is presented on Figure 4.

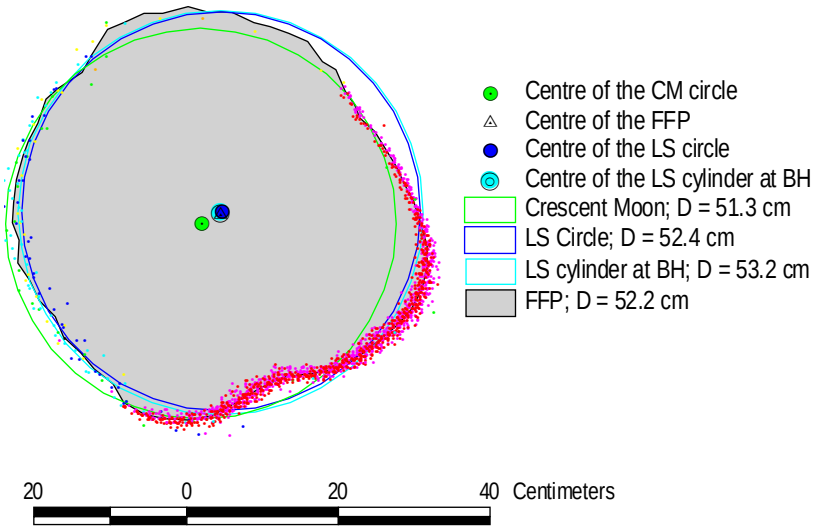


Figure 4: The different determinations of the basal area

The distributions of the points along the perimeter of the tree are rarely even. The most extreme case when only a part was measured from one scanning positions, but 1/3 dense, 1/3 sparse and 1/3 no points is a very typical case in multi-scanning situations.

The further developed CM method choose the scan position with the highest number of points and/or closest scanning distance, nevertheless the DBH and the basal area is the worst here. The determined positions and diameters are still very good as starting points for the other methods, which requires some input parameters.

The free form polygons have not given much more accurate results then, e.g. the traditional least square adjustment circle fitting, but the recalculated centroid and the shape of the polygon is very precious for querying and selecting points are belonging to a trunk.

The cylinder fitting usually gives a very similar results as the circle fitting, but this method is much more reliable in the case of low point numbers.

The height of the monopodial trunk

The automatic determination of the junctions based on free form polygons is far from perfect. It determined the junction and its height in 57% correctly. There were 11 trees (11%) where the determination produced a false positive. Most of these were understory vegetation in the neighbourhood, which bewildered the algorithm. Four trees were only with junction without recognition. The free form polygon method can trace the trunk much better, then the 'traditional' CM method, because it is less sensitive to noise and much more robust. The determination of the accurate height of the junction is ambiguous. The most promising way was the tracing back and force several times with different parameters, but it does not work automatically currently (see Figure 5).

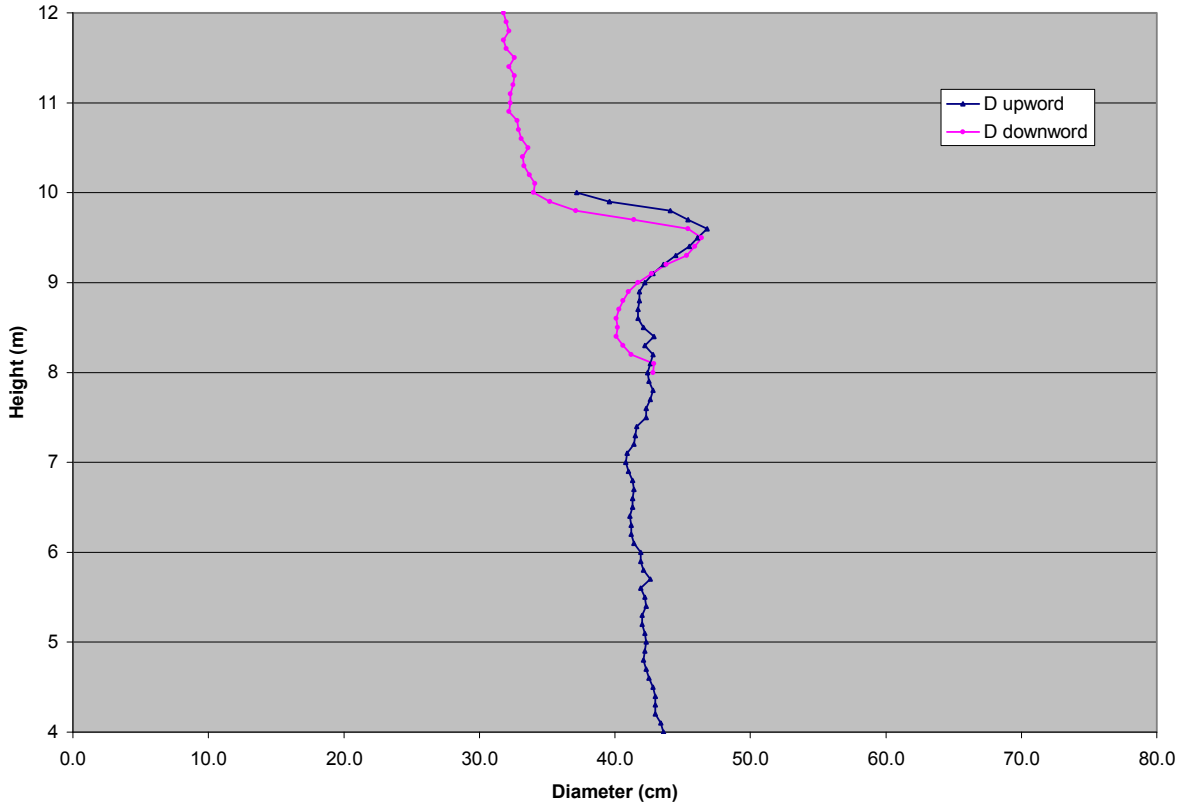


Figure 5: The comparison between tracing up- and downwards

The volumes

The volumes of the single trees are almost the same with the different modelling approach. The discrepancies between the different kinds of conoid's modelling approach were always under 0.1%. This result was very surprising at first sight, but if we consider the isotropy of the model – the vertical dimension is usually 20 times bigger than the horizontal one –, and the fine vertical samplings (usually we applied 10 cm , or sometimes 5 cm saplings), then there is not any astonishing in it. The application of the FFP resulted much more realistic root (see Figure 6)

Figure 6: 3D model of a trunk by FFPs, and details.

4. Conclusions

The calculation of the volume of single trees requires some horizontal and vertical measures. The basal area (*g*) and the height (*h*) were the most important horizontal and vertical parameter in this study. The improved crescent moon method served as a semiautomatic tool for the stem map and for the initial value of the basal area. A newly developed algorithm for free form polygon creation was applied for more accurate basal area calculation and more realistic basal area shape for 3D visualisation. The FFP algorithm was implemented for the determination of the first junctions and its height. The form factor of the 2D shape was monitoring to detect the junctions, and the 3D form number (*f*) was continuously calculated and the volumes were calculated according to the form.

The improvement of the DBH or the height was not very significant, but with the application of the free form polygons, the robustness and flexibility of the algorithm are much better. This portends the possible direction of the further improvements: the real automatic tracing of the trunks.

Acknowledgements

We thanks the good cooperation and the lot of helps for the piLine surveying company, who performed the laser scanning surveys, and for the Pilisi Parkerdő Forestry Company for their efforts in the field.

References

ASCHOFF, T. and SPIECKER, H., 2004, Algorithms for the automatic detection of trees in laser scanner data. In Proceedings of the ISPRS working group VIII/2, ”Laser-Scanners for forest and Landscape assessment”, 3-6 October 2004, Freiburg, Germany. pp. 71-75.

BROLLY, G. and KIRÁLY G., 2009, Algorithms for stem mapping by means of Terrestrial Laser Scanning. Acta Sylvatica et Lignaria Hungarica, 5, pp. 119-130.

HOPKINSON, C., CHASMER, L., YOUNG-POW, C. and TREITZ, P., 2004, Assessing forest metrics with a ground-based scanning LIDAR. Canadian Journal of Forest Research, 34, pp. 573-583.

KIRÁLY, G. and BROLLY, G., 2007, Tree height estimation methods for terrestrial laser scanning in a forest reserve. In Proceedings of ISPRS Workshop on laser scanning 2007 and SilviLaser, 12-14 September, 2007, Espoo, Finland. pp. 211-215.

KISS, B. 2009, Geodetic preparation survey for laser scanning on Pilisszentlélek 25A compartment, a Pro Silva Demonstrational Area. Master Thesis. University of West Hungary, Sopron.

PFEIFER, N. and WINTERHALDER, D. (2004): Modelling of tree cross sections from terrestrial laser scanning data with free-form curves. In: Proceedings of the ISPRS working group VIII/2, ”Laser-Scanners for forest and Landscape assessment”. Freiburg, Germany. 3-6 October 2004. 76-82.

VEPERDI, G. 2008, Erdőbecsléstan (Forest mensuration) Educational Aid. University of West Hungary, Sopron. Available online at: http://www.nyme.hu/fileadmin/ dokumentumok/emk/moi/TantKovetelm/Erd_becslestan_2008.pdf (accessed 12/03/2010).

WEZYK, P., KOZIOL, K., GLISTA, M. and PIERZCHALSKI, M., Terrestrial laser scanning versus traditional forest inventory first results from the polish forests. In Proceedings of ISPRS Workshop on laser scanning 2007 and SilviLaser, 12-14 September, 2007, Espoo, Finland. pp. 424-430.

Certification of carbon credit using airborne lidar

A. KATO†, T.KOBAYASHI†, L.M.MOSKAL‡, P.SHIESS‡, D.CALHOUN§

†Graduate School of Horticulture, Chiba University, 648 Matsudo Matsudo-shi, Chiba 2718510 Japan.

‡College of Environment, Precision Forestry Cooperative, Box 352100 University of Washington Seattle, 98195 WA USA

§Laboratoire d'Etudes des Transferts et de Mecanique des Fluides, Commissariat l'Energie Atomique F-91191 Gif-sur-Yvette Cedex France

Abstract

The Kyoto Protocol recognizes afforestation, reforestation, and forest management as forest sink activities. There has been, however, no standard and accurate certification method established for carbon credit. Lidar (Light Detection and Ranging) can be a good certification tool to prove the amount of carbon stock of forest, because lidar measurement is the most accurate to detect carbon change among remotely sensed data. The certification of forest carbon stock needs to monitor the change of stem volume and forest management activities such as thinning and pruning. Airborne lidar can provide the accurate tree parameters as well as forest management activities. Our study area is located at Sanmu city in Chiba Prefecture, Japan. The research area is used to be a good timber production place of Japanese cedar (*Cryptomeria japonica*). But most area is abandoned for a long time and is infected by tree disease, because timber production is not profitable and stands are not managed well. The infected trees have cracks on the stem causes the degradation of timber quality and prevent the normal growth of stem volume. To certify the carbon credit in this region, identification of damaged trees is required. Previous research shows the crown properties are significant variables to estimate stem volume. In this study crown volume is quantified by airborne lidar and used to get the relationship with stem volume. The healthy trees are better suited for generating carbon credit more. Then monitoring the tree health condition is desirable from airborne lidar analysis. The estimation of crown volume in this study is given by a computer graphic technique using radial basis function and isosurface. As the result, there is high correlation between field and airborne lidar measurement to get R^2 of 0.91 and R^2 of 0.75 for tree height and crown base respectively. Crown volume estimated by the wrapped surface is compared with stem volume derived from field measured tree height and stem diameter. The result shows that crown size is diminished after the crack appears on the stem surface f infected trees. Crown volume estimation derived by lidar has a good potential for identifying tree disease infection.

1. Introduction

Forest sink activities recognized by the Kyoto Protocol are afforestation, reforestation, and forest management. These activities are accounted as carbon credit. The carbon stock from forest is becoming a major credit for emission trading in the future framework. Moreover, United Nation has the collaborative program named Reducing Emissions from Deforestation and Forest Degradation (REDD) in developing countries. Monitoring forest activities are required for generating carbon credits. Active remote sensing technologies are expected to quantify and monitor carbon stocks and change.

The country which set the target of carbon reduction under the Protocol is expected to reduce the carbon emission during the period. Japanese target is 6% and 3.8 % of it is allowed to count from forest carbon sink (around 13 million CO_2 tons). Since 69% of the total Japanese land is classified as forest and the lands are limited to create a new space for afforestation and reforestation, the forest management of the natural and commercial plantation forest is considered to achieve the target. Furthermore, Japanese government set the target of 25% carbon reduction by 2020. To accomplish such a high percentage of the carbon reduction, domestic carbon trading should be introduced to trade the carbon credits among industries. And the carbon credit should be supplied well to fit the increasing demand from big emitter such as utilities. Forest carbon

sink will play a more important role in the carbon trading. Monitoring these forest change activities are, therefore, required for generating carbon credit.

Single tree based parameters have been derived from airborne Light Detection and Ranging (lidar) as the followings: tree height (Andersen et al., 2006, Yu et al., 2004), crown width (Persson et al., 2002, Popescu and Zho, 2007), crown base height (Næsset & Økrand, 2002, Holmgren & Persson, 2004, Popescu & Zhao, 2007), and crown volume (Kato et al., 2009). Patenaude et al (2003) compared approaches to quantify the forest carbon sink among several remotely sensed data include the both passive and active sensors. They concluded that lidar is the most accurate and appropriate tool for the estimation of forest carbon stock. Forest carbon stock is calculated by the multiplication among expansion factors (for branches, leaves, and roots), density factor (depends on species), conversion factor from cellulose to carbon (0.5), and stem volume given by an allometric equation using stem diameter and tree height. The allometric equations have been established using field measured stem diameter and tree height. Technically, tree height and stem diameter are only required to calculate carbon stock. Eventually, the certification of forest carbon stock needs to monitor the change of stem volume and forest management activities such as thinning and pruning. Airborne lidar can provide the accurate tree parameters as well as forest management activities.

While the general algorithm and methodology to analyze lidar has been developed, Japanese local problems using airborne lidar have been raised by previous research (Hirata.2004, Omasa et al., 2003, Takahashi et al., 2005a, 2005b, 2006, Yone et al., 2002). Hirata (2004) and Takahashi et al (2005a) found the terrain slope effect to identify trees and measure tree height. And Takahashi et al (2006) discussed the penetration rate of closed-canopy and middle-aged (40–50 years old) hinoki cypress (*Chamaecyparis obtuse*) and Japanese cedar (*Cryptomeria japonica*) plantations. Takahashi et al (2005b) made various regression models for stem volume based on different slope condition. They found crown properties such as crown area and sunny crown mantle volume as a significant predictor variable for stem volume. Omasa et al. (2003) and Yone et al. (2002) made Lidar derived model to fit the field based allometric equations to estimate carbon stocks from densely planned Japanese cedar and pine (*Larix letolepis Gordon* & *Picea jezoensis Carr.*) forest.

All of these previous approaches and techniques were required to solve Japanese local situation. There is another problem besides these conditions of Japanese forest. Most Japanese natural and commercial forest is located in rural mountainous area and the trees have been densely planted and have not been managed well so that tree disease is spread out for the wide area in our study site. We chose our study site over the moderate slope area to mainly focus on the tree disease issue in this region.

Since Takahashi et al. (2005b) shows the crown properties are significant variables to estimate stem volume, in this study crown volume is quantified by airborne lidar and used to get the relationship with stem volume. Certified carbon credit should be generated from good stand accumulates carbon stock well. The healthy trees are better suited for generating carbon credit more. Then the tree health condition is anticipated from airborne lidar analysis.

Crown volume is also a good indicator for forest management activities such as pruning branches. To certify the forest carbon credit, crown volume can be a key predictor variable. The estimation of crown volume in this study is given by a computer graphic technique to quantify the crown volume of any irregular shape of trees. In this study Japanese cedar trees are only used to assess because of the influence of tree disease. The purpose of this paper, therefore, covers the following two topics;

1) relationship between crown volume and stem volume.

2) stem volume estimation influenced by the damage of tree disease in this region.

Research site and field data

The research area is located in Sanmu City, Chiba prefecture at the east of Chiba city, Japan (figure 1). The total area is around 9 km^2. The area is a secondary forest and is used to be a good production area of

Japanese cedar. But the most area has been abandoned for a long time, since the forest operation was not economically feasible. The main coniferous species of our study is the most common species in Japan; Japanese cedar and hinoki cypress. The terrain of our study area is moderate in slope. Stands are mostly homogenous and partially heterogeneous, multi-layered stand age and mixtures of tree species. Summary of field measurement is shown in Table 1. Most Japanese cedar trees in this region are infected by the fungi (*Cercospora sequoiae*). The fungi are carried by wind and specially infect Japanese cedar trees. When trees are infected by the fungi, the stem has a crack which degrades the quality of timber and interrupts the normal and healthy growth of trees. Height of trees is influenced by the damage level of the infection. Two categories (damage or no damage) were, therefore, observed during the field work. The deeply cracked tree is shown in figure 2 to show the visual influence of the infection of the tree disease.

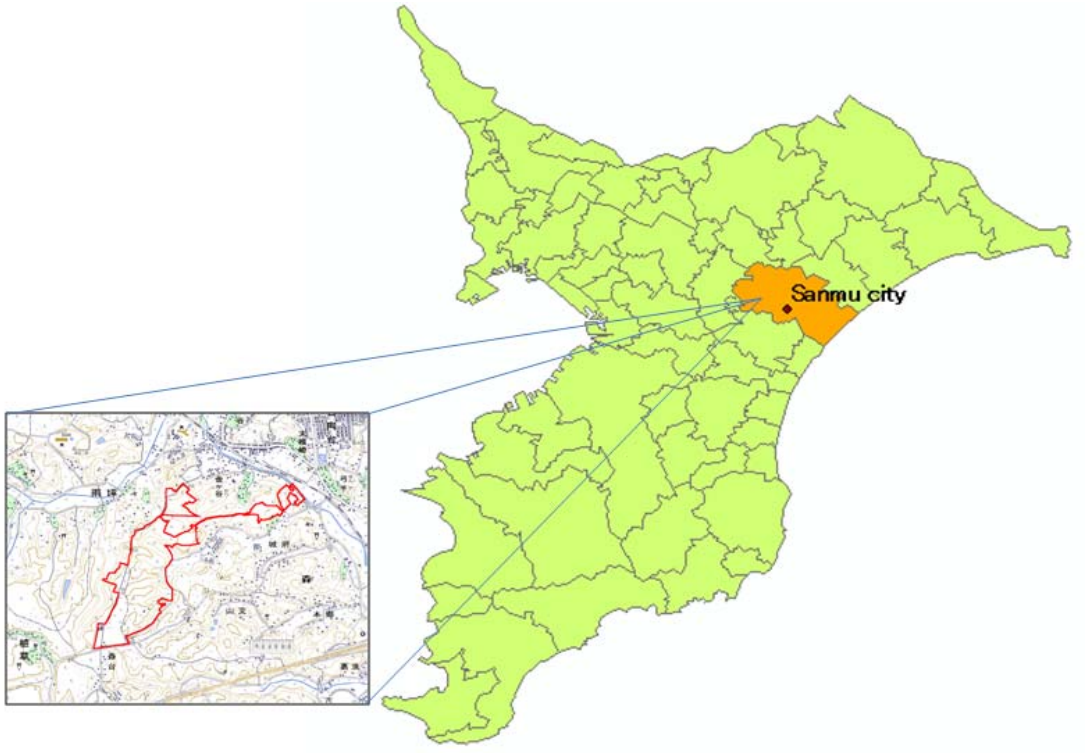


Figure 1. Study site located at Sanmu city, Chiba Prefecture in Japan. The red boundary area was taken by airborne lidar data.

Table 1. Summary of field measurement (n = 154)

	No damage	Damage
DBH (cm)	12.2 ~ 61	15.7 ~ 58.3
Tree height (m)	16.6 ~ 32.4	18.3 ~ 31.6
Crown base height (m)	10.1 ~ 26.6	11.6 ~ 25.6

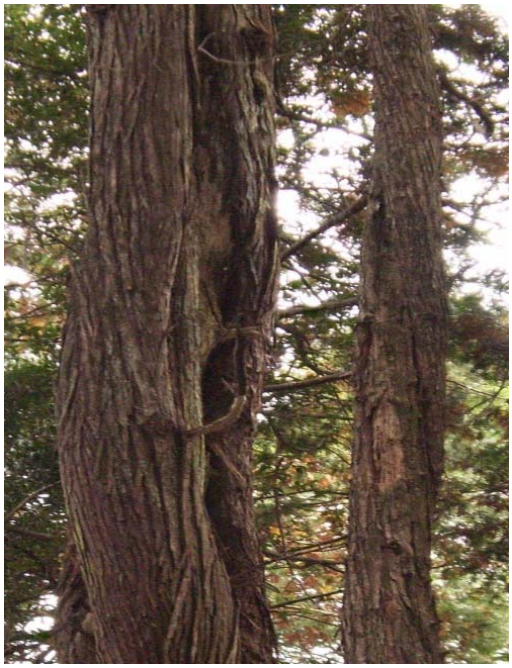


Figure 2. The damage of crack on the stem due to the infection of the fungi (Cercospora sequoiae)

We used subjective random sampling to collect data for the total 154 trees to measure tree height and crown base height using the clinometers Vertex Ⅲ(Haglöf Inc., USA). The tree species name and the damage were observed for all trees. The field-measured tree stem location was captured by the compact total station (Ushikata TEO Ray-130). In closed canopy, GPS location was not accurately collected. The vantage points under the open sky area were set using Trimble GeoXT handy GPS unit (Nikon-Trimble Inc., USA). All trees were georeferenced based on the location of vantage points. Stem volume equation of Japanese cedar tree is given by locally available allometric equation as follows;

DBH between 6 cm to 10 cm: Log10V = 1.810505 log10 D + 1.041345log10 H – 4.205648 (1)

DBH between 11 cm to 20 cm: Log10V = 1.787554log10 D + 1.164989log10 H – 4.288017 (2)

DBH between 21 cm to 30 cm: Log10V = 1.564747log10 D + 1.103501log10 H – 3.907737 (3)

DBH above 31 cm: Log10V = 1.741999log10 D + 0.981423log10 H – 3.995470 (4)

where V = stem volume (m3), D = stem diameter and diameter at breadth height (DBH) (cm),

and H = tree height (m).

In this study, Japanese cedar tree samples are only used to get the relationship between crown volume and the damage of tree disease infection.

Lidar Data

A small-footprint waveform airborne lidar dataset was acquired over this research site in August 2009 by Nakanihon Air Service, Japan. A small-footprint full waveform sensor is a new sensor and more points reflected from the inside of canopy can be received by this sensor. Figure 3 shows the effect of penetration. It is clear to see more points are collected relative to conventional sensor. The lidar sensor setting is shown on Table 2. The coordinate system of data associated with this lidar system is Japanese local coordinate, 9 zones of rectangular plane coordinate. The flight campaign was conducted during leaf on season with

average altitude 450 m and 30 degrees scan angle. It is the same season of the fieldwork. Average point density is 20 points per square meter.

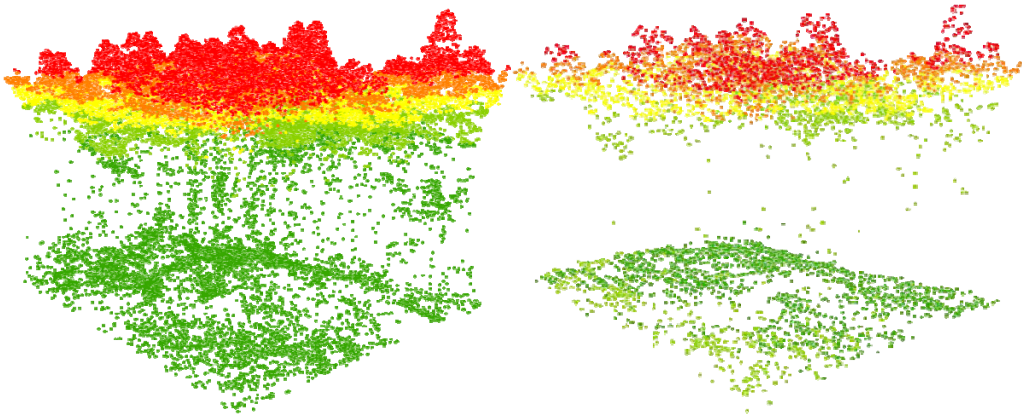


Figure 3. More penetration of the inside of tree canopy using a small footprint full waveform sensor.The right figure is a new sensor data and the left figure is the data taken during leaf-off condition using airborne Optec ALTM 1223 sensor. The data was taken with average point density of 8 points per square meter at the same year of the new sensor. The points are colored by height.

Table 2. A new airborne sensor setting

Acquisition date	August 14th, 2009
Laser sensor	Riegl LMS-Q560
Laser wavelength	1,550 nm (Near infrared red)
Average laser point	20 points/m^2

2. Methodology

The detail algorithm used for lidar processing is described by Kato et al (2009). Firstly, Digital Canopy Model (DCM) is generated to identify tree location and measure tree height Secondly watershed segmentation algorithm (Sollie. 2003, Chen et al. 2006, Yone et al., 2002) is applied to DSM. Based on the segmented region, lidar points are classified horizontally. Lastly, a median filter (Holmgren & Persson., 2004) is applied to separate lidar points vertically between upper and lower canopy or understory vegetation. Classified point clusters are used for the following wrapped surface reconstruction to calculate crown volume.

DCM generation

Creating a Digital Canopy Model (DCM) comes from the difference between Digital Terrain Model (DTM) and Digital Surface Model (DSM). To create DTM, ground returns were classified from all returns by the vendor. Triangulated Irregular Network (TIN) is utilized to interpolate linearly among classified ground returns. TIN surface is smoothed by the natural neighborhood algorithm provided by ArcGIS ver 9.3 (ESRI Inc.) to create 50 cm by 50 cm DTM. Average point density (20 points per square meter) used in this study is high enough to generate 50 cm resolution of DTM. To fill the holes created by no returns on DTM, mean spatial filter was applied to DTM and to make a continuous DTM. DSM is also created by local maximum points within 50 cm by 50 cm grids. A 3 x 3 Gaussian filter (Hyyppä, et. al. 2001) was convolved over the DCM. DCM is used for segmentation labeling for lidar points in the following procedure. Elevation value of

each lidar point was subtracted by the elevation values of DTM to remove the slope effect for the following process.

Watershed algorithm

In order to segment lidar points, a marker controlled watershed algorithm (Chen et al. 2006) is used. In this study, a level set method is used to identify the local peaks of the smoothed DCM surface. In this approach, the plane continued progressively through DCM from top to bottom and 'slices' the DCM by 20 cm interval. For each sliced plane, a value of 0 is assigned for pixels whose height was less than the height of plane and 1 for all others to create a binary image. For each binary image, a connected component labeling is implemented to label and classify the pixels. To identify the peaks, one sliced image at a certain height is compared with the other sliced image of the next height to see the difference between them. If the total number of labels increased from one image to the other, the marching sliced plane passed some local peaks of the surface and the locations of the missed local peaks are collected as treetops.

At the same time, local peaks are set as markers and the marker-controlled segmentation (Chen et al., 2006; Sollie, 2003) and a gradient flow analysis in eight neighboring pixels is used to determine which peak the surrounding pixels belonged to. All pixels are classified based on the number labeled on the local peak. From the classified image, the georeferenced pixels are used to assign all discrete lidar points into point clusters.

Vertical separation of lidar points

To separate between lower and upper canopy laser returns, we adapted a similar technique used by Holmgren & Persson (2004). The vertical height bin is generated to the maximum height. For each bin, zero is assigned to the bin which has less than 1% of the total point density and one is assigned for the others. Over the resulting binary bin, a one-dimensional 3 by 3 median filter (Holmgren and Persson 2004) is applied on the array of vertical height bin. The height of bin is generated from 0.1 to 1m by 0.1 m interval. Appropriate height of bin is influenced by the vertical profile of point distribution. If the profile is clear bimodal between lower and upper vegetation, any height of bin is acceptable. But if it is not clear and skewed towards upper height, smaller bin size is applied. As smaller bin is used, higher the height of separation is. The appropriate height of the bin is determined by the field measured crown based height. It is a semi-automatic approach.

Wrapping process and crown volume estimation

Wrapped surface is the fitted surface over discrete Lidar points. Kato et al. (2009) used this technique for isolated trees to get R^2 of 0.84 and 0.89 for coniferous and deciduous trees respectively. In this research, the same technique is applied to various stand conditions.

The process of wrapping uses convex hull algorithm to extract crown surface points only, radial basis function to calculate 4th dimensional attribution using the normal vector of each point, and isosurface is applied to create three dimensional contour represents wrapped surface over points. The detail description is shown on Kato et al. (2009). And crown volume is calculated by calculus divergence theorem. Tree height is measured by the maximum height of lidar points within the segmented point cluster. And crown base height is measured at the bottom of the wrapped surface around the stem location (the mean coordinates of the wrapped surface).

3. Result

Out of all sampled trees, co-dominant tree segments are excluded due to poor segmentation result from watershed segmentation. Tree height and crown base height measurement are validated by field measurement, because the both variable is a significant variable to calculate stem volume and crown base height is also a significant value to estimate crown volume. And the relationship between the crown volume derived by the wrapped surface and stem volume given by the field based allometric equation is obtained.

The relationship with stem volume is classified by damage or no damage. Visualization of the wrapped surface is shown in figure 3.

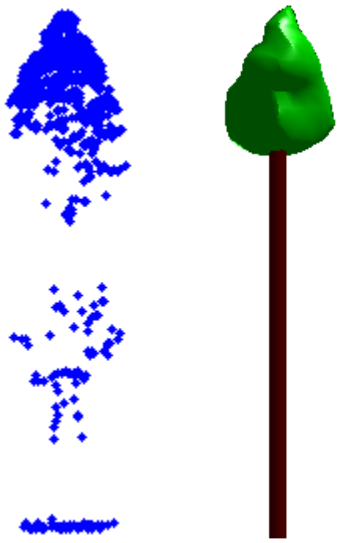


Figure 3. Visualization of wrapped surface reconstruction

Validation of tree height and crown base height

Tree height and crown base height derived by the wrapped surface are validated by the field measured tree height and crown base height (figure 4 and figure 5). Figure 4 shows R^2 of 0.91 for tree height measurement, even tree height measurement is conducted under closed canopy site and it was difficult to see the tips of treetops. Figure 5 shows high R^2 of 0.75 for crown base measurement.

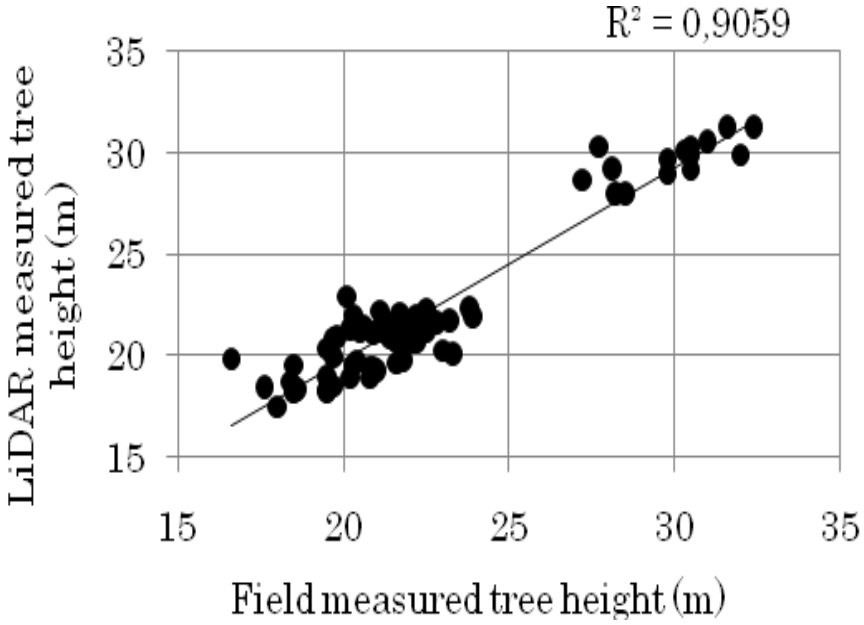


Figure 4. Correlation of tree height between Lidar and field measurement. p-value is 0.00 (p < 0.01).

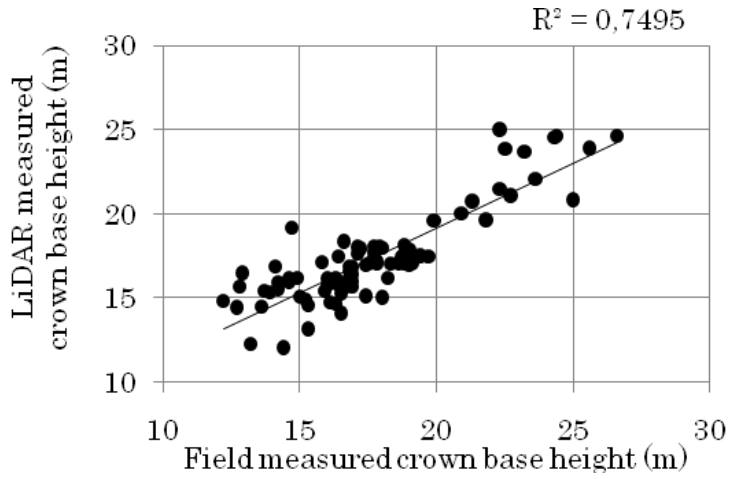


Figure 5. Correlation of crown base height between Lidar and field measurement. p-value is 0.00 (p < 0.01).

The relationship between crown volume and stem volume

Crown volume estimated by the wrapped surface is compared with stem volume (figure 6). Tthe left side of figure 6 shows no damage samples and the right side of figure 6 shows samples has at least one crack on the stem.

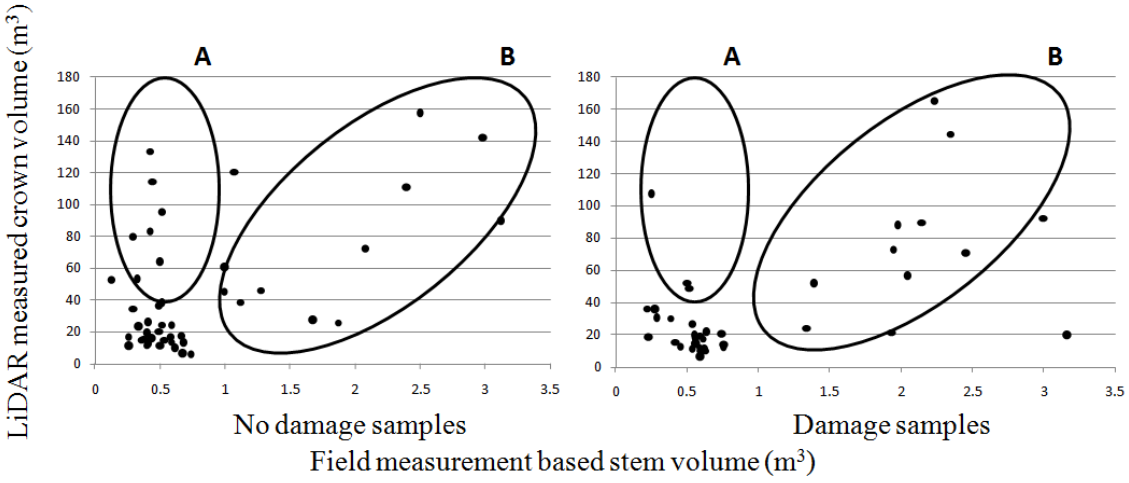


Figure 6. Crown volume estimation based on the damage of tree disease.

4. Discussion

Validation of tree height and crown base height

Correlation coefficient (R^2) of tree height and crown base height is high enough to derive these parameters from airborne lidar data in this research. From figure 5, R^2 of 0.75 of crown base height shows that various height of bin appropriately separate between upper and lower canopy of vertically overlapped segments. Holmgren and Persson (2004) also got R^2 of 0.71 using the same technique and Popesucu and Zhao (2007) used a voxel-based segmentation to get R^2 from 0.73 to 0.78. R^2 of 0.75 in this study is reasonable accuracy

compared with previous studies. Over homogeneous stands, the co-dominant trees cannot be segmented well enough to separate lidar points horizontally. If convex shapes of tree are not distinct due to high tree density, there is some limitation to segment lidar points.

The relationship between crown volume and stem volume

Since the new sensor (a small footprint waveform lidar) can receive laser returns reflected from the inside of canopy more, the wrapped surface reconstructed using the segmented lidar points can provide better crown shape. The crown base height was accurately obtained by airborne lidar. It indicates that airborne laser can reach the degree of crown base height.

It is difficult to observe crown edge in densely planted area. In this study, the field measurement of crown volume was not feasible to validate the lidar derived crown volume estimation. Then crown volume estimation between no damage and damage trees is relatively compared in this study. Figure 6 shows the two different A and B region for the both no damage (left) and damage (right) figures. The figures show the change in A region when two figures are compared. The x axis shows the stem volume and y axis shows the crown volume. It is expected to have linear relationship between stem volume and crown volume like B region. But in A region, no damage samples have various crown size with the same stem volume. But the damage samples have constant smaller crown size. It means that the crown size is diminished after the crack appears on the stem. And it is reasonable to say that the damage of this tree disease causes to diminish the size of tree crown. It takes some time interval between the infection and the appearance of crack on the stem. There is an enough time interval between them to change tree crown size. Airborne lidar has a good application to detect the change caused by tree disease infection.

This region is abandoned for a long time so that the pruning activities have not been conducted. The change shown in this study is, therefore, not related with forest management activities and more related with natural change of crown shape. But the tree crown shape is also influenced by tree density. In dense forest, trees stand close each other to interrupt the expansion of branches. In order to identify the influence of tree density, more samples are required to separate between the changes caused by tree density and by tree disease.

The stem volume equations were established around 1960. The equation is too old to fit the current stand condition of this region. The spread of tree disease was not considered when the equations were made. Update of these equations is needed, because carbon sink from the growth of the infected trees is low. In monitoring the location of the poor growth of stem, the change of crown volume from multi-temporal data can provide better estimates. As the next step of this research, lidar oriented stem volume equation is established to monitor the carbon change for the certification of carbon credit.

5. Conclusion

A small footprint waveform sensor is used to estimate crown volume over densely planted area. The new sensor can receive more laser returns from the inside of canopy than the conventional sensors can. It is a good accurate tool to certify carbon credit in this region. To certify carbon credit, forest management activities and forest carbon stock are required to monitor efficiently. Crown volume is a good indicator to monitor the both. Since the tree disease was spread out in this region without any treatment, the crown health is mainly analyzed for the certification carbon credit. Tree height and crown base height measured by airborne lidar have high correlation with field measurement. And crown volume shows the change caused by the tree health status. As the next step of this research, more samples are collected to identify the crown volume changes comes purely from the tree disease. Then the goal of this research is to establish an updated allometric equation using airborne lidar data.

Acknowledgement

This research was supported by the Environment Research and Technology Development Fund (RF-1006) of the Ministry of the Environment, Japan. We thank Akira Nemoto (Graduate School of Horticulture, Chiba University) for collecting field data of this research.

References

Andersen, H. E., Reutenbuch, S. E., & McGaughey, R. J., 2006. A rigorous assessment of tree height measurements obtained using airborne LIDAR and conventional field methods. Canadian Journal of Remote Sensing, 32, 355−366.

Chen, Q., Baldocchi, D., Gong, P., & Maggi, K., 2006. Isolating individual trees in a savanna woodland using small footprint LIDAR data. Photogrammetric Engineering and Remote Sensing, 72, 923−932.

Hirata, Y., 2004. The effects of footprint size and sampling density in airborne laser scanning to extract individual trees in mountainous terrain. Proc. ISPRS WG VIII/2 “Laser-scanners for forestry and landscape assessment”, Vol. XXXVI, Part 8/W2, 3–6 October 2004, Freiburg, Germany.

Holmgren, J., & Persson, Å., 2004. Identifying species of individual trees using airborne laser scanner. Remote Sensing of Environment, 90, 415−423.

Hyyppä, J., Kelle, O., Lehikoinen, M.,&Inkinen, M. 2001 A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. IEEE Transactions on Geoscience and Remote Sensing, 39, 969−975.

Kato, A., Moskal, L., Schiess, P., Swanson, M., Calhoun, D., & Stuetzle, W. 2009. Capturing tree crown formation through implicit surface reconstruction using airborne lidar data. Remote Sensing of Environment, 113, 1148−1162.

Næsset, E. & Økrand, T., 2002. Estimating tree heights and tree crown properties using airborne scanning laser in a boreal nature reserve. Remote Sensing of Environment,79, 105−115.

Omasa, K., Qiu, G. Y., Watanuki, K., Yoshimi, K., & Akiyama, Y., 2003. Accurate estimation of forest carbon stocks by 3-D remote sensing of individual trees. Environmental Science and Technology 37, 1198–1201.

Patenaude G., Milne R., and Dawson T.P., 2003. Synthesis of remote sensing approaches for forest carbon estimation: reporting to the Kyoto Protocol. Environmental Science & Policy 8, 161-178.

Persson, Å., Holmgren, J., & Söderman, U., 2002. Detecting and measuring individual trees using an airborne laser scanner. Photogrammetric Engineering & Remote Sensing, 68, 925−932.

Popescu, S. C., & Zhao, K., 2007. A voxel-based LIDAR method for estimating crown base height for deciduous and pine trees. Remote Sensing of Environment, 112, 767−781.

Sollie, P., 2003. Morphological image analysis principles and applications (pp. 105−137). NY: Springer.

Takahashi T, Yamamoto K, Senda Y, Tsuzuku M., 2005a Estimating individual tree heights of sugi (*Cryptomeria japonica* D. Don) plantations in mountainous areas using small-footprint airborne LiDAR. J For Res 10:135–142

Takahashi T, Yamamoto K, Senda Y, Tsuzuku M., 2005b Predicting individual stem volumes of sugi (*Cryptomeria japonica* D. Don) plantations in mountainous areas using small-footprint airborne LiDAR. J For Res 10:305–312

Takahashi T., Yamamoto K, Miyachi Y., and Tsuzuku M., 2006 The penetration rate of laser pulses transmitted from a small-footprint airborne LiDAR: a case study in closed canopy, middle-aged pure sugi

(*Cryptomeria japonica D. Don*) and hinoki cypress (*Chamaecyparis obtuse Sieb. et Zucc.*) stands in Japan

J For Res: 11:117–123

Yone Y, Oguma H, Yamagata Y., 2002 Development of measurement system for the carbon sinks under the Kyoto protocol (in Japanese with English summary). J Rem Sens Soc Jpn 22:531–543

Yu, X., Hyyppä, J., Kaartinen, H., & Maltomo, M., 2004. Automatic detection of harvested trees and determination of forest growth using airborne laser scanning. Remote Sensing of Environment, 90, 451−462.

SINGLE TREE INVESTIGATIONS IN URBAN AREAS USING AIRBORNE LASER DATA

Alvand Miraliakbari, Afua Abasa†, Detlev Wagner, Michael Hahn, Johannes Engels, Gerald Kändler‡

† afuaabasa@googlemail.com; ‡ gerald.kaendler@forest.bwl.de

Department of Geomatics, Computer Science and Mathematics, University of Applied Sciences Stuttgart

Schellingstraße 24, D-70174 Stuttgart, Germany

(alvand.miraliakbari,detlev.wagner, michael.hahn, johannes.engels)@hft-stuttgart.de

‡ Forest Research Institute of Baden-Württemberg,Department of Biometry and Information Sciences
Wonnhaldestrasse 4, D - 79100 Freiburg i. Br.

Abstract

Nature and scope of features collected in tree cadastres of cities and municipalities differ widely. Among the basic parameters of such cadastral databases are the location of each single tree and other geometric features like stem diameter, crown diameter, height and sometimes basal area and volume. Field methods are applied in collecting those parameters but strategies employing aerial photographs and airborne LiDAR data promise a much higher efficiency. The latter ones are often combined with field methods which then may focus on vitality checks and assessment of possible damages for every tree.

In this paper airborne laser data is analyzed to determine geometric tree parameters. A major advantage of airborne laser scanning is that the vertical tree structure can be directly measured. The approach chosen for processing the laser data is to employ filtering and classification methods mainly used in airborne LIDAR processing. Classification distinguishes ground and tree. Three approaches are proposed for single tree investigation (1) filed work to measure the diameter at the breast height (DBH) and tree position (2D coordinates) and identify the tree species, (2) semi-automatic approach for measurement of the crown diameter and height of the trees, (3) automatic approach for detection of the individual tree based on the 2D watershed algorithm. Regression analysis is applied for estimating the DBH. Proposed formulae for DBH and biomass estimation which are proposed in the literature are utilized and the related coefficients are estimated.

1. Introduction

Investigations on trees in urban areas are of great interest in forestry and city planning. For describing a tree with respect to its 3D structure or other characteristics, geometrical features of the single trees can be extracted in particular from recorded laser scanning point clouds. Terrestrial and airborne laser scanning is nowadays playing a significant role in recording 3D information about urban vegetation. Geometrical parameters of the trees may serve as proxies for tree properties like biomass and leaf mass. Proposing methods for representing the vertical canopy structures is one of the main issues in single tree investigation. “The major task of vertical canopy structure analysis is to detect the number of main canopy layers and the higher range of each canopy layer.” (Wang et. al. 2008) The important factor of single tree investigation using LiDAR data is that the measurements should be based on both upper and lower part of the canopy.

The main objective of this research is to investigate the geometrical information of the trees in the urban area. Tree investigation is not only focused on the rural area but also the forest. Due to the different climates in different countries of the world, the amount of the trees in the urban and rural area will be different. In Germany, because of the high percentage of the forest cover (31% of the territory of Federal Republic of Germany), investigation of the trees is an important issue. Thus, after agriculture, forestry is the second largest land use form. The forest cover in Baden-Württemberg is the 38.1 % of the whole state (Roering, 2004). In this regard, definitely, the single tree investigation outside of the forest like in the urban area is of special interest.

LiDAR is an active remote sensing technology that determines ranges by measuring the time of flight of an emitted laser to travel to a target and back to the detector. Accurate positioning and attitude data of the LiDAR system allows for retrieving 3D positions of the target points. Major application is the production of DTMs, but in recent years forestry applications have become more and more important. A major advantage of LiDAR is the provision of detailed information on both the horizontal and vertical distribution of vegetation in forests. LiDAR is widely used to measure forest and tree parameters in recent years; main reasons are (1) measuring forest structures directly, (2) range measurements of high accuracy and (3) reliable repeatability of surveys.

Discrete return LiDAR systems have smaller footprint and very good horizontal resolution but require high point density. They are suitable for collecting parameters on an individual-tree base, such as crown characteristics. Waveform LiDAR systems produce more accurate height data but usually come with a larger footprint. Hence, they are indicative of multiple forest elements (Lim et al., 2003). Accuracy of LiDAR derived attributes is comparable to field enumeration techniques (Holmgren et al., 2004).

Individual-tree detection and segmentation include neighborhood information of point clouds and pixels of Digital Surface Model (DSM) to derive physical features. Extraction, modeling and analysis of individual trees requires modern high-rate scanning systems with pulse rates of 100 000 pulses per second or more and a density of at least 1 point/m^2 (Vosselman & Maas, 2010). Calibration applied with (field) reference data is often limited. In many cases, individual-tree based approaches have not been calibrated at all. However, it is recommended to calibrate the data in order to avoid systematic errors in the applied models and to improve the volume and diameter estimation.

Trees can be segmented by detecting local maxima in the first pulse image, using watershed algorithms. Region growing in a gradient image has a similar effect. Overlapping of trees or under-storey trees can not be detected. In dense forests the discrimination of individual trees can be problematic. Point clouds of lower density might cause the same problems. For characterization of under-storey trees the original point clouds must be used instead of DSMs or CHMs. Multi-storey forest stands can be classified using the height distribution of multiple returns. Number and sizes of suppressed trees can be predicted with estimated regression models. The accuracy of this result is dependent on the density of the dominant tree layer. Using wave-form signals, the number of extracted points can be increased by 18% to 57%, depending on the type of vegetation (Persson et al., 2005).

Wang et al., (2008) have investigated fully automatic 3D single tree modelling in forest. In their research procedure, individual trees are extracted in different ways: from top canopy layer and sub canopy layer. The authors have reconstructed the 3D shape of the extracted crowns of the trees. For modelling of single tree crowns, first morphological opening and closing process is utilized in a hierarchical way in order to delineate the tree crown regions. In the next step they have used DSM-based single tree delineation algorithms which are based on morphological pouring or watershed. The most important factor in this part is that the region growing should be stopped when the neighbouring regions touch each other.

Rahman et al., (2008) have investigated the tree crown delineation based on the density of the high points(DHP). In this research they have compared DHP and CHM for tree detection. In both cases for tree crown delineation they have used inversed watershed segmentation. They conclude that, compared to the CHM based approach, the DHP based tree detection and tree crown delineation performs better. The exceptions are coniferous trees. For this species the CHM based approach performs on a small scale better than the DHP approach. It is also mentioned that post-processing of tree crown segments is essential particularly for CHM based method to improve the accuracy of tree detection procedure and also tree crown delineation.

For the experiments an area called Stadtgarten is used which is located in the city of Stuttgart, Germany. Stadtgarten covers an area of 456m*406 m. The data which is used in this research is mainly the airborne laser data as irregular and interpolated data, ground truth provided by field surveys, terrestrial laser data which is used only for testing the result of the automatic measurements.

2. Measurement Procedure

The workflow encompasses three parts:

Field work

Field work is carried out to get reliable ground truth data for the test area. The following tree attributes have been collected: The diameter at breast height (DBH) of each tree is captured with the help of a tape measure and the coordinates of the trunks is measured using GPS and tachometry. During this field inspection the species of all trees is recorded as well.

Semi automatic height and crown diameter measurement

In this part, the main focus is to detect the height and crown diameter using semi automatic methods.

First, the position of the tree is selected. The utilized data is the difference of first and last pulse of the interpolated airborne laser scanning (ALS) data. The next step is to draw a circle which surrounds the tree (the diameter of this circle is larger than crown diameter (CD) of the tree). Finally, the last step is to draw the profiles of four diameters of the circle (0°, 45°, 90° and 135°) and to calculate the mean value of the length of the profile to estimate the crown diameter. To calculate the height, obviously, the largest value of the pixels inside the circle is taken into account as top height.

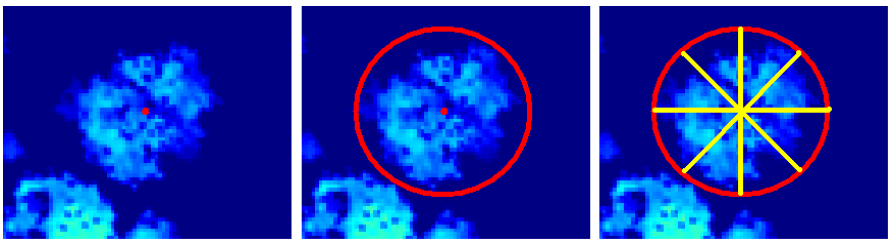


Figure 1 (a). Manual selection of the middle of the crown, (b) drawing neighbouring circle around the tree. (c) drawing four diameter of the neighbouring circle

Proposed workflow of automatic height and crown diameter measurement

We have used the difference of first and last pulse of the interpolated ALS data. The following workflow describes the automatic measurement of height and crown diameter:

(1) Running 2D watershed algorithm. In this case, the data will be converted to a set of segments.

(2) Plotting position of the trunks on the dataset (segmented image).

(3) Labelling the trunk position and the segments.

(4) Searching for the segments in which more than 1 trunk are located.

(5) Subdividing these segments into more parts in which only one trunk is located into every sub-segment.

(6) Associating segments in the neighbourhood of the individual trunks which have no segments.

(7) The next step is grouping of the segments which are associated with the individual trunk positions.

(8) Finally, the crown diameter and height of each tree will be calculated based on the segment(s) which is/are associated with the trunk.

Experimental Investigation

For the experiments an area called Stadtgarten is used which is located in the city of Stuttgart, Germany. The size of the rectangular test area is 456*406 m^2. The diameter at breast height (DBH) of 116 trees was recorded manually in the field, together with the tree species. Aesculus hippocastanum (chestnut) is most prevalent with 57 individual trees, followed by *Acer* (10), *Betula* (7) and *Taxus* (6).

Another field work was to measure the 2D position of the trunks. There are always some challenges in order to estimate the correct coordinates of the stem using only DGPS, because it is difficult to measure the trunk

coordinate under branches of leaves. Thus, apart from DGPS, laser range finder device is utilized. The rover of DGPS is Trimble Geo XH and the master can be connected to this hand-help GPS via receiver called Geo beacon (Trimble website, 2010). The laser range finder called TruPulse 360° B, sends the laser pulse and using the internal compass calculates the bearing (Lasertech website, 2010). The first result of the measured points is the polar coordinates of the target from the rover station. The coordinates of each target is transferred to the UTM with a reasonable accuracy (0.3 to 1 m).

In semi-automatic measurement of the CD and height (H), as discussed, after selecting the point in the middle of the tree and drawing the profiles on the diameters of the circles, length of the profiles (red lines) is measured and the mean values of these lengths is assign as the CD of each tree. Table 1, shows some samples of measured CD based on the semi automatic approach.

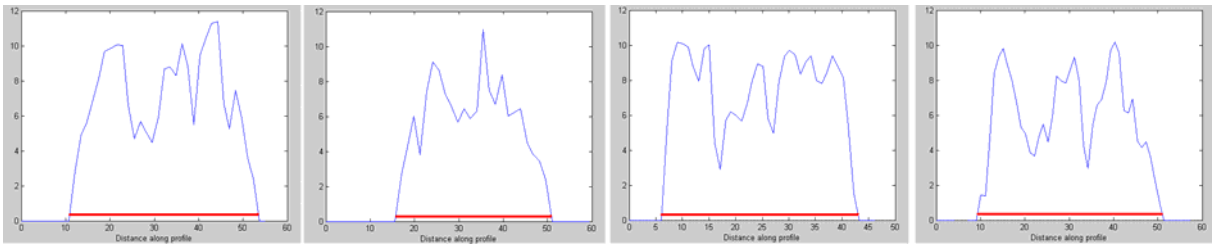


Figure 2. Example of four profiles over the tree, mean value of the horizontal length is assigned as CD[m].

In automatic measurement, vegetation area is extracted from the difference of the interpolated raster data sets of first and last return. Gray scale dilation of the last return data is applied to remove locations along building edges. Only points more than 1 m above ground are considered in the last step for removing small objects. The segmentation with a watershed algorithm produces tree outlines, which are displayed in Figure 3(b). Some trees are over segmented, few are not detected as trees.

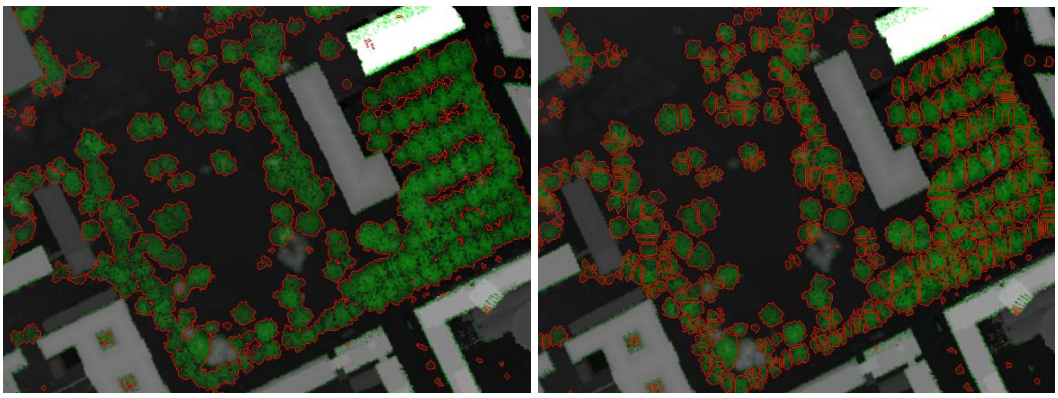


Figure 3. Single tree segmentation before running (a) and after running (b) watershed algorithm,

The segmentation is based on binary image. It means after producing the raster image, there should be a threshold taken into account to extract off-terrain binary image of vegetation. The pixel value of ground is assigned as 0 and the vegetation area is classified into1. By using this binary image, we have applied the distance transform to produce the catchment basin based on the distance from every pixel to the nearest nonzero pixel. Then watershed algorithm was applied to segment the trees in order to estimate the crown diameter (CD). As we see in the figure 3(b), the segments are more separable than the figure 3(a). But, there are automatically segmented and due to be sure if each segment is a single tree, more investigation is necessary. We may face some problems. There are either some segments which together show a single tree, or on the other side some segments are individually composed of more than a single tree. That is the reason of measuring position of the trunks. Figure 4(b) shows some sample position of the trunks (red dots) measured by the GPS and the segmented vegetation area.

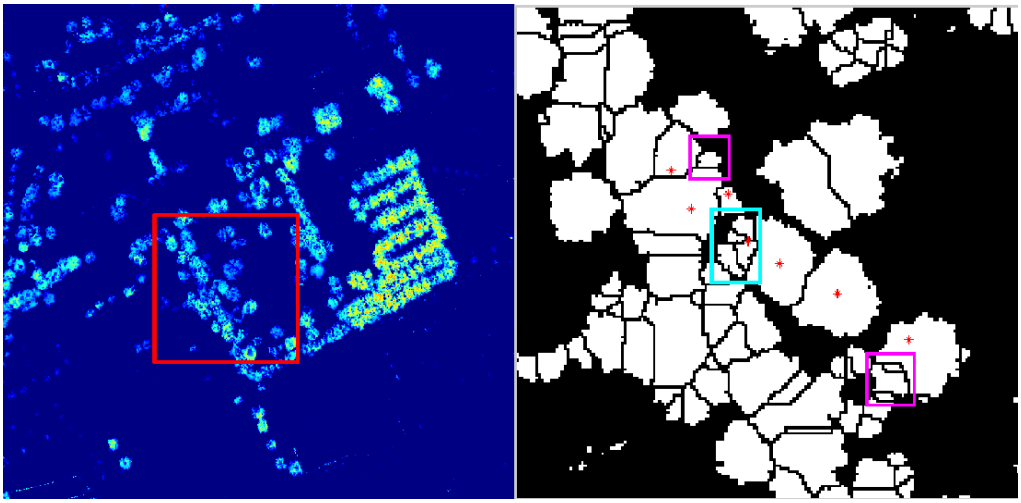


Figure 4 (a, b). Seven samples of the trunk coordinate on the segments.

There are seven samples of the trunks coordinates on the segments. The figure shows that the segmentation for most of the trees is appropriate. There is one exception in this figure belonging to the trunk in the middle of the samples points (displayed inside the cyan rectangle). The appropriate segment of this point cannot be identified because the point is outside of the segments.

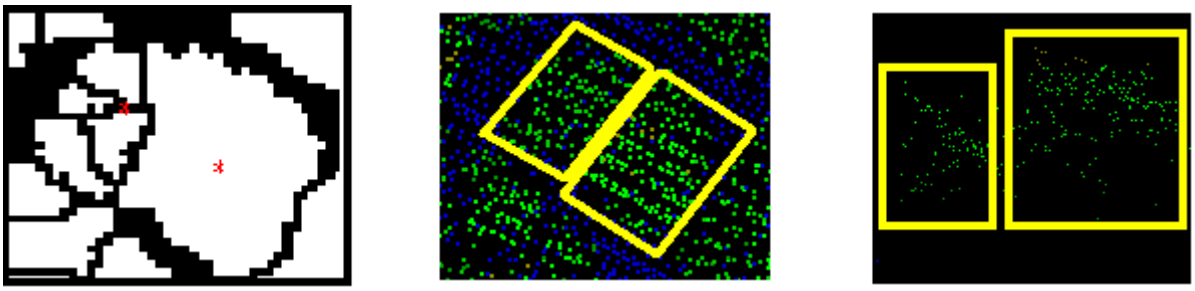


Figure5 a. Position of two trunks on the segmented area, b. top view of two trees, c. cross section of two trees

In this case, there is more than one segment which can be associated with the trunk position because of the smallest distance of the trunk to the segments (figure 5(a)). The biggest neighboring segment is associated with one of the trunks, therefore, it is not taken into account. The remaining segments are appropriate to the distance threshold and will be associated with the isolated trunk. As it is shown by figure 5(b and c), more investigation on the irregular point cloud shows that the trunk should be associated with more than one segment. The point is surrounded by three small segments and by applying a certain distance from the point, these three segments will be assigned to the isolated trunk.

Figure 6. Classified three segments as one segment

Regression Analysis

The single tree attributes measured in the field as ground truth were used for fitting a regression equation to predict DBH with ALS-derived variables as predictors. The variable crown diameter (CD) was derived from ALS data by drawing profiles, height (H) were measured as the highest pixel value within the individual tree area in the ALS data.

Table 2: Measured values of Aesculus hippocastanum trees

Tree	Species	DBH [cm]	CD [m]	H [m]
1	Aesculus	56,7	8,4	14,2
2	Aesculus	55,1	8,47	13
3	Aesculus	41,4	8,27	13
4	Aesculus	46,8	7	16
5	Aesculus	50,0	7,3	16
....	....	...	...	...
57	Aesculus	46.8	12.3	10.9

Table 3: Estimates of regression coefficients for DBH

Estimated parameters	a_0	a_1	a_2
Aesculus	22.65	1.16	1.35
Acer	9.07	0.55	1.79
Betula	8.46	0.24	1.88

In order to gain insight into the dependency of the diameter at breast height from the crown diameter and the tree height the linear model proposed by Popescu (2007) is used.

$$DBH = a_0 + a_1 CD + a_2 H \qquad (2)$$

The data are the CD derived from ALS data by employing profile views, the H measured as the highest value within the individual tree area in the ALS data. Aesculus hippocastanum is used in this regression as most of the trees in the test field are from this species. The results for Aesculus hippocastanum are given in Tabble 1 and 2. The standard deviations of the DBH (σ_0) are 10.3 cm, 9.8 cm and 6.6 cm for Aesculus hippocastanum, *Acer* and *Betula*, respectively. However, only the results for Aesculus hippocastanum are based on a sufficient number of observations (57 trees), whereas the others are only indicative.

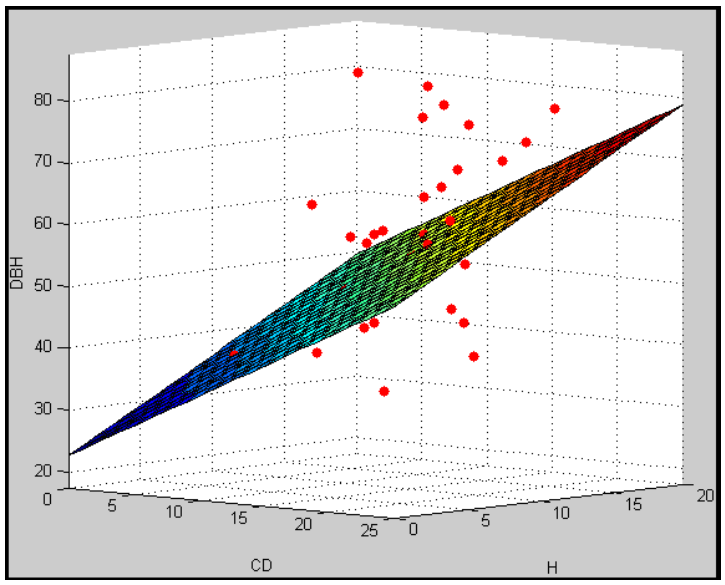


Figure 7. Fitting a surface on the DBH[cm] values with respect to the CD[m] and H[m]

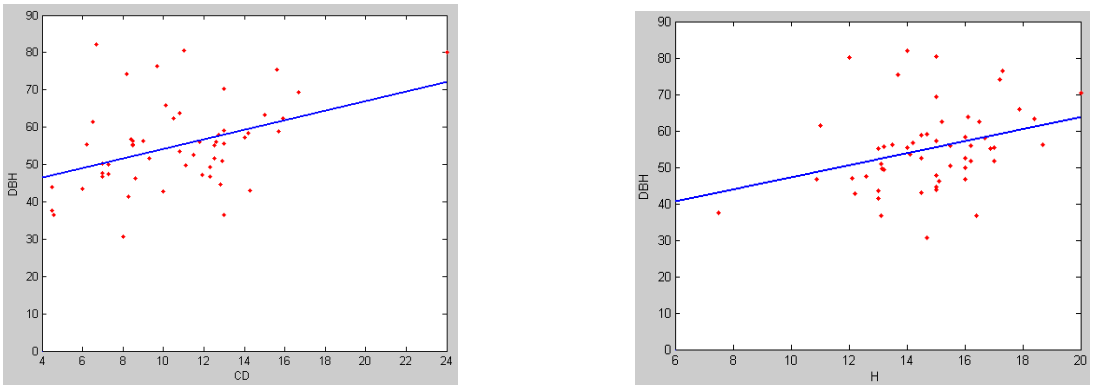


Figure8 . Linear regression of DBH[cm] with respect to the CD[m] (a) and H[m] (b) for Aesculus hippocastanum.

Biomass estimation

Biomass estimation using the ALS data is one of the most important aspects of forestry. There are quite a lot of formulas in this regards. The formula which we have used in this research for above ground biomass (dry weight in kg) is as follows:

$$B = b_0 d_{1.3}{}^{b1} h^{b2} \qquad (3)$$

where b_0, $b1$ and $b2$ are the coefficients, $d_{1.3}$ is the DBH (1.3 m above ground) and h is the height of the trees (H). The function was basically calibrated for the trees in the city of Karlsruhe. The trees were sampled using randomized branch sampling (Gregoire & Valentine., 2008). Based on the measurements, the entire above-ground tree volume was measured. The biomass equation was fitted as a multi-level model with tree species as grouping effect which acts only on the coefficient b2. In addition, the residual variance was modeled as a power function of the fitted values. Because of the large number of Aesculus hippocastanum trees, we used the species-specific coefficients as shown in Table 3.

Table3. Coefficients of the allometric biomass functions

	b0	b1	b2
Fixed effects	0.08313406	1.80334833	0.98060397
Specific coefficients for Aesculus	0.08313406	1.80830354	0.98060397

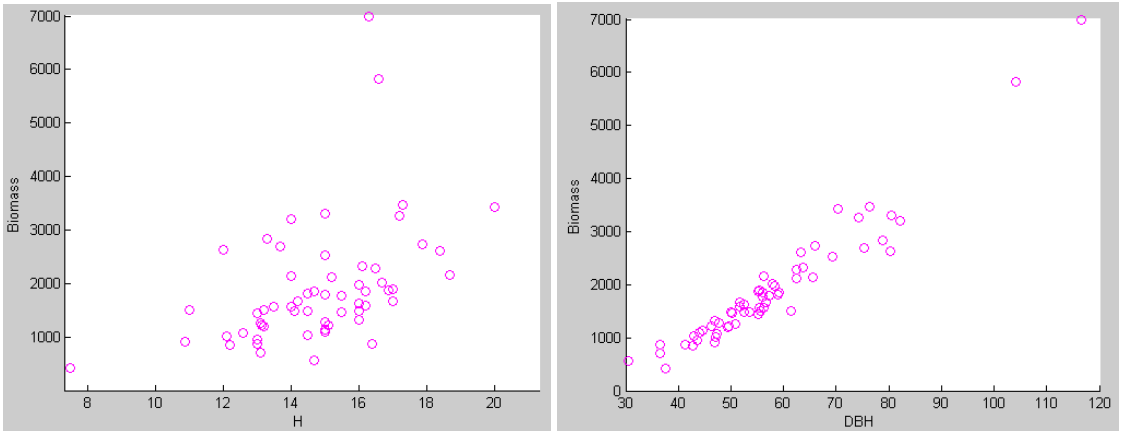


Figure 9 (a,b). Scatter plot of calibrated biomass[kg] with respect to the H[m] (a) and DBH (b) for Aesculus

3. Conclusion

Trees in urban areas have a significant impact on human health and the environment. Comprehensive modeling assists planning and management of natural resources. The studies to be carried out focus on collecting 3D data by airborne laser scanning to derive detailed estimates of important tree parameters (DBH, CD, H and biomass).In general, the LiDAR-derived estimates of canopy variables are of the order of 10 times more precise than those used in conventional forest models to date (Roth et al., 2007).

The proposed method of semi- automatic measurement is simple and straight forward. The only challenge is to detect the central part of the tree in order to make a circular buffer with a special diameter. There might be some mismeasurement due to have close trees whose crowns are connected to each other.

The automatic procedure is relatively faster than the previous method. Segmentation is more than necessary. Some are over – segmented. In this case, using the trunk position with respect to the proposed method is useful for grouping the segments which belong to an individual tree.

References

Gregoire, T.G. and H.T. Valentine 2008. Sampling Strategies for Natural Resources and the Environment. Applied Enviromental Statistic. Chapman & Hall. 474p.

Holmgren, J., & Jonsson, T. (2004). Large scale airborne laser-scanning of forest resources in Sweden. international Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36(Part 8/W2) , 157-169.

Lim, K., Treitz, P., Wulder, M., St-Onge, B., & Flood, M. (2003). LiDAR remote sensing of forest structure. Progress in Physical Geography 27,1 , 88-106.

Oono, K., Numata, Y., & Hirano, A. (2008). An improved method of individual tree detection using airborne LiDAR. Proceedings of Silvilaser 2008, September 15-17 , pp. 508-516.

Persson, A., Söderman, U., Töpel, J., & Ahlberg, S. (2005). Visualization and analysis of full-waverform airborne laser scanner. International Archives of Photogrammetry, Remote Sensing and Spatial Informatioin Sciences, XXXVI, Part 3/W19 , 103-108.

Popescu, Sorin C. (2007). Estimating biomass of individual pine trees using airborne lidar, Spatial Sciences, Laboratory, Department of Ecosystem Science and Management, Texas A&M University, 1500 Research Parkway, Suite B 223, College Station, TX 77845, USA,Received 11 August 2006; Accepted 17 June 2007.Available online 13 August 2007.

Rahman, M.Z.A. B. G. H. Gorte (2008). Tree Crown Delineation From High Resolution Airborne LiDAR Based on Densities of High Points, Delft Institute of Earth Observation and Space Systems (DEOS), Delft University of Technology, the Netherlands.

Roering, Hans-Walter (2004). Study on Forestry in Germany, Arbeitsbericht des Instituts für Ökonomie 2004/16, Hamburg, October 2004

Roth, B. E., Slatton, K. C., & Cohen, M. J. (2007). On the potential for high-resolution lidar to imrpove rainfall interception estimates in forest ecosystems. Frontiers in Ecology and the Environment, 5(8) , 421-428.

Vosselman, G. and Maas, H. G. (2010). Airborne and Terrestrial Laser Scanning. Boca Raton, Fl, USA: CRC Press.

Wang, Y. Weinacker, H., Koch, B., Stereńczak. K. (2007) LiDAR Point Cloud Based Fully Automatic 3D Single Tree Modelling in Forest and Evaluations of the Procedure, Dept. of Remote Sensing and Landscape Information Systems(FeLis), University of Freiburg, Tennenbacher Str. 4, 79106 Freiburg, Germany.

Wang, Y., Weinacker, H. and Koch, B. 2008, A Lidar Point Cloud Based Procedure for Vertical Canopy Structure Analysis And 3D Single Tree Modelling in Forest, Dept. of Remote Sensing and Landscape Information Systems, University of Freiburg, Germany,

http://www.trimble.com/geoxh.shtml (accessed July 2010)

http://www.lasertech.com/TruPulse-Laser-Rangefinder.aspx (accessed July 2010)

Adapting α-shapes for forest delineation using ALS Data

LOTHAR EYSN† , MARKUS HOLLAUS†, MICHAEL VETTER‡, WERNER MÜCKE†, NORBERT PFEIFER†, BRUNO REGNER§

† Institute of Photogrammetry and Remote Sensing

‡ Centre for Water Resource Systems

Vienna University of Technology, Gußhausstraße 27-29, 1040 Vienna, Austria

§ Department of Forest Inventory at the Federal Research and Training Center for Forests, Natural Hazards and Landscape, Seckendorff-Gudent-Weg, 1130 Vienna, Austria

Abstract

The objective of this paper is to evaluate a new approach for the automatic delineation of forested areas based on airborne laser scanning (ALS) and national forest inventory (NFI) data. In the Austrian NFI a forest area is mainly defined with four fundamental criteria. One of these criteria, the so called “crown coverage”, is the most complex variable and therefore the main focus of this paper is on defining and implementing this criterion in an automatic process to delineate forested areas. Based on Austrian NFI data functions were determined for two different test sites in Austria, describing the criterion crown coverage as a relation between tree height and the distance between trees. Based on the ALS data an automatic method on the basis of adapting □-shapes was developed to link these functions to the ALS data. The approach was tested for two different test sites in Austria. For the first test site a tree species independent function was applied. The results of the delineated forest mask are validated with a reference forest mask which was manually delineated based on orthophotos. The derived forest mask differ less than 1.6% from the reference forest mask and shows a very high accuracy. For the second test site tree species dependent functions were applied for the assessment of the crown coverage. The presented approach shows promising results and shows the high potential for the automatic forest area delineation based on ALS data.

1. Introduction

Acquiring topographic data of the Earth's surface is widely realized with airborne laser scanning (ALS). Especially for forestry applications height information is a fundamental input to derive different forest parameters (Means et al., 2000). There are several economical and ecological applications like e.g. the estimation of tree heights (Næsset and Bjerknes, 2001), growing stock estimations (e.g. Hollaus et al., 2009b) or forest condition monitoring (Rutters et al., 1992). The determined results of these applications are highly correlated with the fundamental input parameters size and position of the delineated forest areas. Related to forest condition monitoring, the increasing need of a regional and global monitoring of e.g. deforestation requires an automatic determination of forested areas by means of remote sensing data since a manual delineation is a very time- and cost-intensive task.

In the past mainly aerial images were used for manual or semi-automated extraction of forested areas. Shadowing effects limit this task especially for the detection of small forest clearings and the exact delineation of forest borders. Additionally the quality of the results of a manual delineation is highly correlated with the experience of the human analyzer and may lead to inhomogeneous, maybe even incorrect datasets. However, an automated method for forest delineation based on ALS data can overcome these limitations in most instances and shows objective and reproducible results at short evaluation times. Both, manual and automated methods to delineate forested areas require an exact geometric forest definition. Depending on the different locations worldwide many different national forest definitions are available (Lund, 2010) beside a global definition of Food an Agriculture Organization of the United Nations (FAO) (Zhu and Waller, 2003).

The forest definition of the Austrian national forest inventory (NFI) is mainly based on the four criteria (1) minimum height depending on an in situ reachable tree height, (2) minimum crown coverage, (3) minimum area size and (4) minimum area width (Gabler and Schadauer, 2006). Additionally the criterion of land use has to be considered. The criteria of minimum area and minimum tree height can easily be considered, whereas the parameter crown coverage is not clearly defined in the NFI. Therefore, this study aims at developing a generic, automated approach for delineating forested areas from ALS data using these three criteria with the focus on defining and implementing the crown coverage. The remaining parts of this paper are organized as follows: Section 2 describes the selected study areas and the used data. Section 3 describes the methodology and implementation whereas Section 4 shows results and their discussions. Finally, concluding remarks are given in Section 5.

Study area and data set

In this contribution the approach for an automated delineation of forested areas is applied for two different study areas in Austria. For the first study area Ötscher an approach depending on tree species (coniferous and deciduous trees) is applied. The tree species are automatically extracted from full-waveform ALS data. For the second study area Zillertal a tree species independent approach is applied to discrete ALS data. A manual delineated forest mask is used for validation purposes within the study area Zillertal.

Study area Ötscher

The study area Ötscher covers an area of 2.2 x 1.5 km and is located in the southern part of the federal state of Lower Austria (Figure 1b). The predominant tree species are red beech (Fagus sylvatica), spruce (Picea abies) and larch (larix decidua) and cover about 80% of the trees in this area. For a previous study the BFW has installed a local forest inventory (FI) for this region. The tree species information from the sampled trees is used as reference for the validation of the derived tree species map. Further information about this study area can be found in (Hollaus et al., 2009a). The used full-waveform ALS data was acquired using a RIEGL LMS-Q560 full-waveform laser scanner during a flight campaign in January 2007 under leaf-off conditions. The mean flying height above ground was 620 m. The mean point density is about 30 echoes/m². For a knowledge-based classification of coniferous and deciduous trees the 3D point cloud with their observables echo width and the calibrated quantity backscatter cross section (Briese et al., 2008) as well as the distribution of the echoes in vertical direction were used. The achieved overall accuracy was 83% (Hollaus et al., 2009a).

Study area Zillertal

The study area Zillertal is located in the eastern part of the federal state of Tyrol and covers an area of 2.5 x 2.5 km (Figure 1a). The lowest elevation of the study area is 620 m above sea level up to 1500 m above sea level at the highest point. The dominant tree species is spruce (*Picea abies*). Beside the forested areas buildings, cable cars and power lines can be found in the study area. The used ALS data was acquired using an Optech Inc. ALTM 3100 laser scanner during multiple flight campaigns in 2008 under leaf-off and leaf-on canopy conditions. The mean flying height above ground was 1200 m. The mean point density is about 5 echoes/m². For the validation of the delineated forest areas a forest mask, which was manually derived from orthophoto interpretations, is used as reference. This forest mask was provided by the Amt der Tiroler Landesregierung, Abteilung Forstplanung.

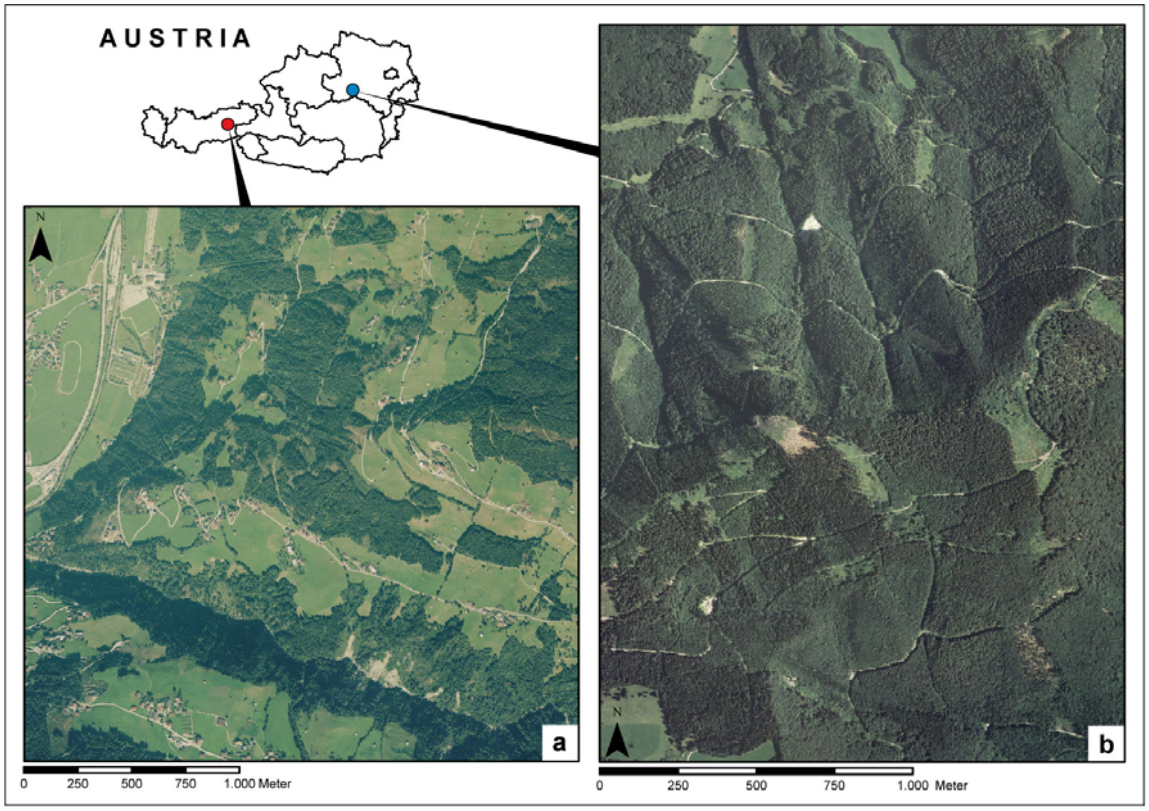


Figure 1: Observed study areas (a) orthophoto of the study area Zillertal (b) orthophoto of the study area Ötscher (sources: Bing maps).

Derived base products

For both study areas the ALS data has been processed and filtered using the hierarchic robust filtering approach (Kraus and Pfeifer, 1998). As a digital terrain model (DTM) and a digital surface model (DSM) was derived. By subtracting the DSM from the DTM a normalized digital surface model (nDSM) was created as a fundamental base product for delineating forested areas. Additionally a slope adaptive echo ratio (*sER*) map, as a measure for local transparency and roughness, was derived (section 3.1). The derived products have a spatial resolution of 1 x 1 m.

2. Methodology and Implementation

As described in section 1 the criteria of the Austrian NFI (area, height and crown coverage) are used for the delineation of forested areas in this study. Based on the NFI the minimum area is defined with 500 m² and is applied by using GIS tools. The minimum height was set to 3.0 m and is considered by applying a height threshold on the nDSM heights. Artificial objects i.e. buildings, power lines and cable cars, which have similar objects heights as forests, are removed from the nDSM in a pre-processing step (section 3.1). The parameter crown coverage defines the projected crown area of trees within a reference area. Current automatic methods calculating crown coverage maps are commonly based on a moving window approach. The kernel size of the moving window, which defines the reference area, is a fundamental parameter. Since there is no exact definition of the size and the shape of the reference area available in the NFI, different results are derived if different kernel sizes and shapes (square, circle, and irregular polygons e.g. forest stands) are applied. Another limitation of the moving window approach is, that especially with circle- or square-shaped kernels smoothing effects occur at the border of a forest and at small clearings. To overcome these problems a method is investigated, which is currently used for the manual delineation of forested areas at the BFW. This method describes the crown coverage by the relation between the tree crown size and the distance between trees (section 3.2) and is originally based on the work of (Hasenauer, 1997). The transfer of this approach to the ALS data is described in section 3.3. For the derivation of the final forest mask all

previously describe processing steps are combined, the minimum area criterion is applied and a final check of the crown coverage within each delineated forest polygon is done (section 3.4).

Removing artificial objects

As all elevated objects e.g. buildings, forests, power lines and cable cars are present in the nDSM a pre-processing step is required to extract a vegetation mask that represents the potential forested area. As shown in previous studies (Höfle et al., 2009; Hollaus et al., 2009a) the slope adaptive echo ratio (*sER*) can be used to differentiate between buildings and forested areas. The *sER* is defined as the ratio between the number of neighboring echoes in a fixed search distance of 1.0 m measured in 3D (a sphere) and all echoes located within the same search distance in 2D (a cylinder) (Höfle et al., 2009; Rutzinger et al., 2008). An *sER* value of 100% means, that the echoes within the 2D search radius describe a planar surface (e.g. roofs), whereas a *sER* value <100% means that the echoes are vertically distributed within the 2D search area and thus indicating transparent objects i.e. forests, building borders and power lines. An empirically determined *sER* threshold of *sER* less than 85% is used to extract the vegetation mask. Finally, morphological operations (open, close) are applied to remove the remaining building borders and power lines from the vegetation mask.

Crown coverage calculation based on NFI data

Based on the defined minimum crown coverage the maximum distances between trees are calculated, which are used as weights for the adapting α-shapes (see section 3.3). To derive the maximum distance the following two steps are necessary:

- A statistical relationship between tree height and crown radius and
- a mathematical relation between crown radii and the maximum distance between two neighboring trees to fulfill the crown coverage threshold.

For the first step measurements of crown radii from the Austrian NFI were used. A subsample of measured trees was chosen for describing the crown radii for trees with low competition according to a border situation. Thus dense stands and trees of lower social classes according to (Kraft, 1884) were excluded. A short description of the data used is given in table 1.

Table 1: NFI data description used for the statistical models between crown radius, tree height and elevation.

	Coniferous trees			Deciduous trees		
	n	mean	Std.dev.	n	mean	Std.dev.
Crown radius (m)	1972	3.14	0.97	242	4.56	1.76
Tree height (m)		26.5	7.2		23.4	6.7
Elevation (m)		1137	426		686	318

For the second step a simple linear approach is chosen leading to the following model:

$$C_r = a + b*H + c*E \quad \text{Eq. (1)}$$

whereas C_r is the crown radius (m), H is the tree height (m), E is the elevation above sea level (m) and a, b, c are factors that are different for coniferous and deciduous trees.

The mathematical relation between crown radii and the maximum distance between the trees to fulfill the crown coverage threshold of the forest definition is a complex problem. For our study a simple solution is to restrict the problem to two trees and define the forest area of these two trees as the area between the trees including the crown projections as shown in figure 2a. The mathematical derivation of the forest area is simplified (figure 2c) by neglecting the exact solution as shown in figure 2b. According to this approximated solution the relation between the maximum distance and the crown radii for variable crown coverage thresholds can be derived as:

$$d = f*(C_{r1}^2 + C_{r2}^2) / (C_{r1} + C_{r2}) \quad \text{Eq. (2)}$$

whereas f is a constant for different crown coverage thresholds.

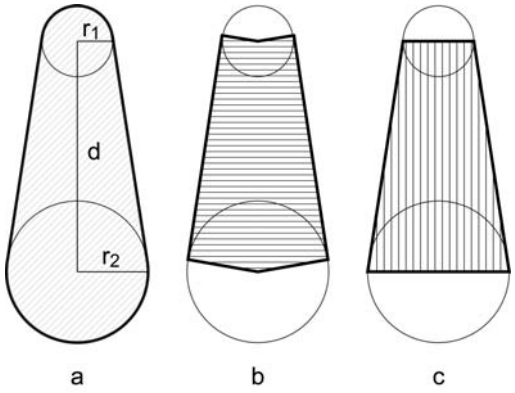


Figure 2: Forest area of two neighboring trees at the borderline. d is the distance between the trees, r_1, r_2 are crown radii. (a) General solution, (b) exact solution and (c) approximated solution for the area between the trees.

Crown coverage calculation based on ALS data

The determined relation between tree height, sea level, main tree species and crown diameter as well as the relation between crown diameter and crown coverage, both derived from NFI data (section 3.2), are applied to the ALS data.

In the first processing step the tree species specific crown diameters are assessed for the coniferous and deciduous trees. The input parameters are the tree height, the sea level and the tree species (see section 3.2). For the study area Zillertal only coniferous trees and for the Ötscher test site both coniferous and deciduous trees are considered. The derivation of the tree species map for the Ötscher test site is described in section 2.1. For the extraction of the tree heights the positions of single trees have to be determined. This is done by using a local maxima filter based on a circular kernel with a diameter of three pixels. Furthermore, the criterion of the minimum tree height is considered. For each detected tree position the sea level is determined by the DTM and the tree height is extracted from the pre-processed nDSM (see section 3.1). Applying Eq. (1) the crown diameter for each tree is computed.

In the second processing step adapting α-shapes are used to consider the maximum allowed distance between trees taking into account the required minimum crown coverage of 30%, which is defined in the NFI. In detail the distances between trees are calculated using a Delaunay triangulation of the detected tree positions. The Delaunay triangulation is calculated using the Open Source software CGAL. The α-shape of a set of 2D-points is a "shape" following the outline of the given points. Depending on the value of α, the shape follows cavities or displays inner holes to a larger or lesser extent (Edelsbrunner and Mücke, 1994) . In our case, the length of the maximum allowed distance between trees (see equation 1) defines the value α and is adapted for each tree pair, depending on tree species and crown diameter. Each triangle is validated if all edges are shorter than the maximum length calculated based on Eq. (1). If this criterion is not fulfilled the triangle is deleted. The remaining triangles are combined to connected areas and provide a potential forest mask.

Forest area delineation

For the delineation of the final forest mask, additional post-processing steps on the derived potential forest mask (section 3.3) are required. As the borderlines of the derived potential forest mask represent the tree stem axis, the potential forest mask is buffered by the half of the maximum available crown diameter found in the study area. In order to prevent an overestimation of the derived forest mask the buffered area is intersected with the vegetation mask (section 3.1). The expanded forest mask is vectorized and the minimum area criterion is applied by deleting single polygons and by filling forest gaps with an area less than 500 m².

3. Results and discussion

Removing artificial objects

The results of the method described in section 3.1 are shown in figure 3 for the study Area Zillertal. Figure 3a shows the original *sER*-map with colored markers pointing to selected artificial objects. Figure 3b shows the processed *sER*-map without the artificial objects. The method shows suitable results for the elimination of artificial objects and deriving a vegetation mask. Buildings, power lines, etc. are removed from the *sER*-map in most instances while vegetated areas with a *sER* value <85 are retained within the processing. The so derived vegetation mask is used to eliminate the nDSM heights from the artificial objects.

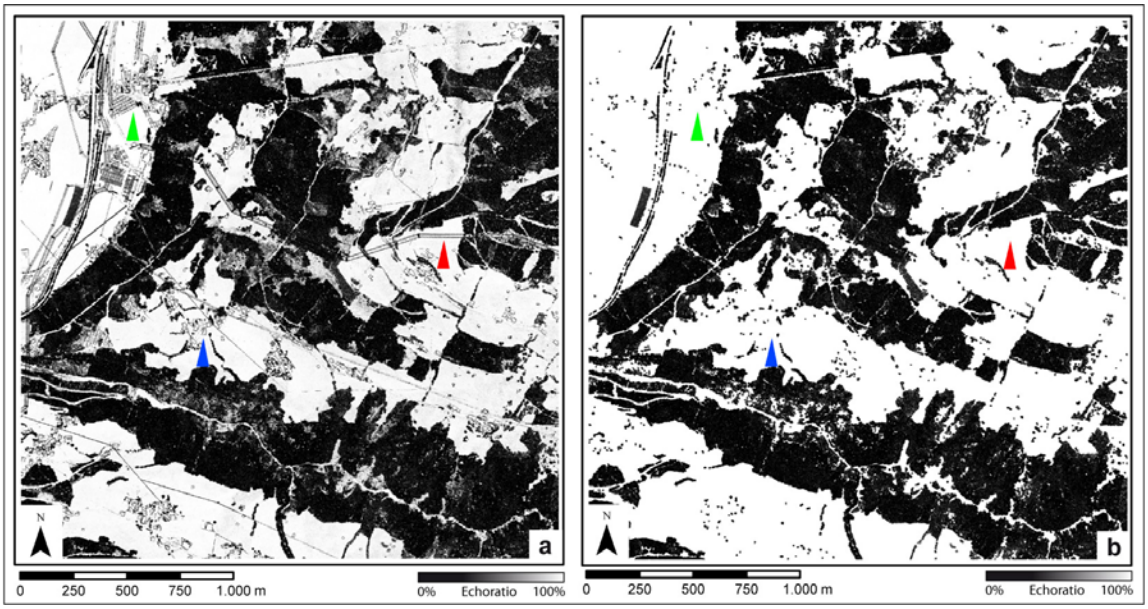


Figure 3: Preliminary mask of vegetated areas: (a) Echo ratio map and (b) adjusted Echo ratio map with eliminated man-made objects of the study area Zillertal. The red arrow shows a power line, the blue arrow shows a building and the green arrow shows a cable car station.

Crown coverage calculation based on NFI data

Eq. (1) was calibrated for coniferous and deciduous trees, whereas the crown radii, tree heights and elevations were taken from NFI sample trees (see table 1). The following equations for calculating the crown radii were found:

Coniferous trees: $C_r = 1.02 + 0.0625*H + 0.000416*E$

Deciduous trees: $C_r = 1.34 + 0.1331*H + 0.000164*E$

Based on Eq. (2) the constant *f* was calculated for coniferous and deciduous trees and is 29.8 and 8.9 for a crown coverage of 10% and 30% respectively.

Crown coverage calculation based on ALS data

Because of the small kernelsize of 3 x 3 pixels multiple local maxima were found within the area of single tree crowns. Especially within dense forested areas the detected local maximums do not represent the exact tree stem positions. Therefore, the amount of detected local maximums is highly correlated with the kernelsize. In the case of multiple local maxima within one tree crown the error of the distance between trees could be the crown diameter at the maximum. However this limitation plays a minor role for the delineation of forests along the timberline or along the forests borderlines, where sparse forests and clear separable single trees are present. For the detected local maxima the corresponding tree crowns were calculated based on the calibrated formulas (see section 4.2) and serve an input for the adapting α–shape method. As shown in figure 4, triangles with edges larger than the maximum possible distance between trees are reliably eliminated from the final triangulation result.

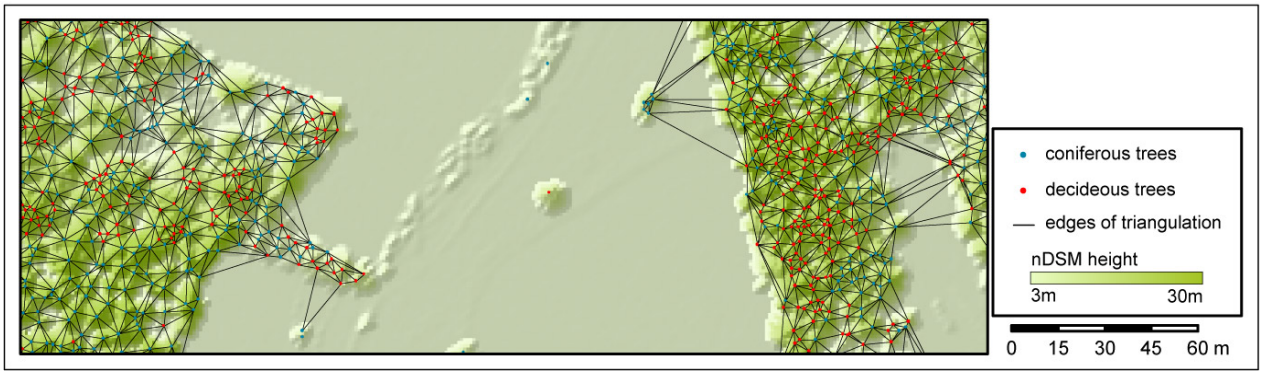


Figure 4: result of the tree species specific adapting α–shape method

Forest area delineation

In Figure 5 the results for the automatic delineation of the forested areas are presented for the study areas Zillertal (figure 5a) and Ötscher (figure 5b). Single trees at the forest's border as well as small forest clearings within a sparse forest are considered as forested area if the maximum possible tree distance is not exceeded (figure 5b). Due to the applied area criterion small forest patches with an area less than 500 m² are removed and forest clearings (<500 m²) are assigned to the forest area. As the preliminary output of the α–shape approach represents the forests borderline along the tree axis the expanded forest area delineate the real forest area with high accuracy (figure 5). The final validation of the crown coverage, which is averaged for each forest polygon, fulfills the required threshold of 30%.

For the study area Zillertal the automatically delineated forest mask was visually validated with the manual delineated forest mask, which is based on an orthophoto interpretation. As shown in figure 5b and 5c the high potential of the ALS based forest delineation is especially within shadowed areas. Based on the clearly defined geometric criteria the result is objective and repeatable. This can be shown in figure 5b where the manual delineation of single trees near the forests border is not comprehensible. Finally the areas of the manually and automatically detected forests were calculated and show a very good agreement. For example for the Zillertal study site the total forest area is 316.375 ha and 311.327 ha for the manual and the automatic detected forest area respectively.

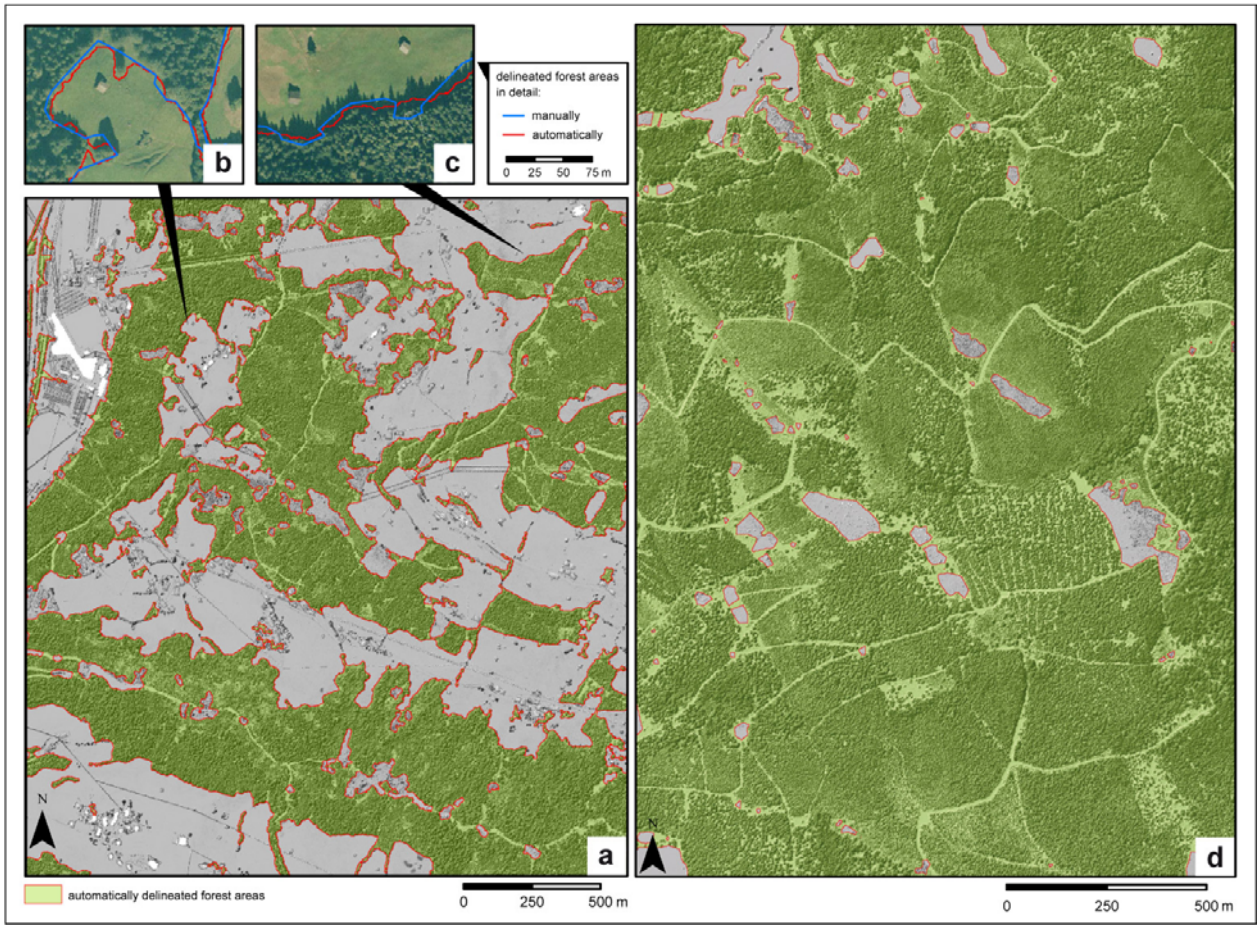


Figure 5: Final forest mask for the study areas Zillertal (a) and Ötscher (d) overlaid over the sER-map. (b) and (c) show the comparison of the automatically delineated forest mask (red line) against the manually delineated forest mask (blue line) as an overlay of an orthophoto.

4. Conclusion and Outlook

The results of the presented approach show the high potential of an automatic delineation of forested areas based on ALS- and NFI-data. The presented method delivers repeatable and objective results with high accuracy along the forests borderline. Since fragments of artificial objects remain in the vegetation mask further steps on eliminating these fragments need to be done. Three critera of the NFI were used and therefore additionally the two criteria minimum area width and land use need to be considered in further investigations.

Acknowledgements

The ALS data for the test area Ötscher was kindly provided by the *Amt der Niederösterreichischen Landesregierung, Gruppe Baudirektion, Abteilung Vermessung und Geoinformation* and for the test area Zillertal by the *Amt der Tiroler Landesregierung, Gruppe Landesbaudirektion, Abteilung Geoinformation*. This study is done within the project LASER-WOOD (822030), funded by the *Klima- und Energiefonds* in the framework of the program "NEUE ENERGIEN 2020".

References

BRIESE, C., HÖFLE, B., LEHNER, H., WAGNER, W., PFENNIGBAUER, M. AND ULLRICH, A., 2008. Calibration of full-waveform airborne laser scanning data for object classification. In: SPIE: Laser Radar Technology and Applications XIII, Orlando; 03-19-2008, Vol. 6950: 8.

EDELSBRUNNER, H. AND MÜCKE, E.P., 1994. Three-dimensional Alpha Shapes. ACM Transactions on Graphics

GABLER, K. AND SCHADAUER, K., 2006. Methoden der Österreichischen Waldinventur 2000/02 - Grundlagen, Entwicklung, Design, Daten, Modelle, Auswertung und Fehlerrechnung. BFW-Berichte; Schriftenreihe des Bundesforschungs- und Ausbildungszentrum für Wald, Naturgefahren und Landschaft, Nr. 135, 132 S.

HASENAUER, H., 1997. Dimensional relationships of open-grown trees in Austria. Forest Ecology and Management, 96, 197-206.

HÖFLE, B., MÜCKE, W., DUTTER, M., RUTZINGER, M. AND DORNINGER, P., 2009. Detection of building regions using airborne LiDAR – A new combination of raster and point cloud based GIS methods. GI_Forum 2009 - International Conference on Applied Geoinformatics, Salzburg

HOLLAUS, M. et al., 2009a. Tree species classification based on full-waveform airborne laser scanning data. 9th International Silvilaser Conference, October 14-16, 2009 – Texas A&M University, College Station, TX, USA, 54-62.

HOLLAUS, M., WAGNER, W., SCHADAUER, K., MAIER, B. AND GABLER, K., 2009b. Growing stock estimation for alpine forests in Austria: a robust lidar-based approach. Canadian Journal of Forest Research, 39(7), 1387-1400.

KRAFT, G., 1884. Beiträge zur Lehre von Durchforstungen, Schlagstellungen und Lichtungshieben.

KRAUS, K. AND PFEIFER, N., 1998. Determination of terrain models in wooded areas with airborne laser scanner data. ISPRS Journal of Photogrammetry & Remote Sensing, 53(4), 193-203.

LUND, H.G., 2010. Gainesville, VA: Forest Information Services.

MEANS, J.E., ACKER, S.A., FITT, B.J., RENSLOW, M., EMERSON, L. AND HENDRIX, C., 2000. Predicting forest stand characteristics with airborne scanning lidar. Photogrammetric Engineering and Remote Sensing, 66(11)

NÆSSET, E. AND 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.

RUTTERS, K.H., LAW, B.E., KUCERA, R.C., GALLANT, A.L., DEVELICE, R.L. AND PALMER, C.J., 1992. A selection of forest condition indicators for monitoring. Environmental Monitoring and Assessment, 20(1), 21-33.

RUTZINGER, M., HÖFLE, B., HOLLAUS, M. AND PFEIFER, N., 2008. Object-Based Point Cloud Analysis of Full-Waveform Airborne Laser Scanning Data for Urban Vegetation Classification. Sensors, 8, 4505-4528.

ZHU, Z. AND WALLER, E., 2003. Global forest cover mapping for the United Nations Food and Agriculture Organization forest resources assessment 2000 program. Forest Science, 49(3)

Terrestrial LiDAR scanning for forest inventory and forest growth monitoring

Taye Mengesha, Maarten Nieuwenhuis & Michael J. Hawkins

taye.mengesha@ucdconnect.ie

UCD Forestry, School of Agriculture, Food Science and Veterinary Medicine, University College Dublin, Ireland

Abstract

The application of terrestrial laser scanning (TLS) technology in forestry has received special attention in recent years for facilitating 3D data acquisition, such as individual tree geometry parameters in forest stands. The objective of the research presented in this paper is to investigate and validate the potential of TLS in monitoring forest growth by attempting to estimate individual tree parameters and plot level growth in a stand of Sitka spruce (*Picea sitchensis* (Bong.) Car.) in Ireland. The data acquisition for all the plots in the stand of different tree size and different slope was carried out using a terrestrial laser scanner in November 2007 and November 2009, using the same plot centres and measurement procedures. The point cloud data were processed with commercially available software (AutoStem) and the derived results were analyzed and compared with the actual values measured using conventional forest inventory techniques. The forest inventory parameters were estimated from the point cloud data with acceptable accuracy, except in the upper stem due to the occurrence of occlusion by branches, and in the lower stem (i.e. the first 1.5 m) due to interference of ground unevenness and buttressing. The preliminary results didn't show a significance difference for the two approaches we used to estimate forest growth. Future work will investigate the potential of TLS for producing accurate data for sustainable forest management planning in stands to be clearfelled by calibrated harvester, and supported by additional information from aerial laser scanning.

Keywords: *Terrestrial laser scanner, LiDAR, Point cloud data, forest inventory, tree growth*

1. Introduction

Terrestrial Laser Scanners (TLS) allow the capture of forest inventory data, such as diameter at breast height (DBH), tree height and stem density, which are essential for modern forestry management including timber harvest forecasting, growth monitoring, as well as biodiversity monitoring. Its capacity for collecting 3D data clouds (X, Y, Z) of several million data points within a few minutes is of special interest for forest inventory data collection. With the advent of terrestrial laser scanning, an active measurement technology independent of the sun or an artificial light source capable of providing millions of points on highly irregular surfaces, is now available for measuring inside forests including forest crops and their surrounding environment. The current forest inventory estimation process, which is subjective and time consuming, can be replaced with an objective and efficient industry standard with the use of this technology.

Several studies have been carried out to automatically determine forest inventory parameters from point cloud data, and have reported on the retrieval of vegetation canopy structure data, including height, basal area, stem counts, and branching parameters, as well as information on standing woody and green biomass (Hopkinson et al., 2004, Watt and Donghne 2005, Bienert et al., 2007, Russo et al., 2007, Mass et al., 2008. Tansey et al., 2009). Henning and Radtke (2006) reported their findings in using ground based LiDAR scanning for the retrieval of stem diameters from loblolly pine (*Pinus taeda L.*) in central Virginia. The results indicated that the LiDAR derived diameter measurements and field measurements were closely in agreement, with an average error < 1 cm for measurements below the base of the live crown, and < 2 cm for heights up to 13 m.

In this study we used tree parameters like DBH and height measured by convectional forest inventory methods in order to investigate and validate the potential of TLS in monitoring forest growth by attempting to

estimate individual tree parameters and plot level growth in a stand of Sitka spruce (*Picea sitchensis* (Bong.) Car.) in Ireland.

2. Materials and Methods

The data acquisition was done with the terrestrial laser scanner FARO LS 800 HE80. The scanner has a field of view of 360° horizontal and 320° vertical and a range up to 80 meter with a distance accuracy of +/- 3mm and a data rate of 120 000 points per second scan speed, and point clouds with more than a million accurate measured surface points can be obtained with a wave length of 785 nm. The FARO LS 800 HE80 uses phase based principle range determination technology where a mirror rotates and directs the laser pulses.

A total of 9 plots with a radius of 15 meters were selected in the forest stand owned by Coillte, the Irish State Forestry Board. All trees in each plot were clearly numbered sequentially. This was necessary in order to reconcile scan data with manually measured data for each stem. As each tree was numbered, a line was painted around the stem at 1.3 m above ground level with a recognizable permanent spray before the scanning was carried out. The trees in each plot are scanned before and after applying a low pruning up to 6 meters height. Two scan measurements were taken in each plot to capture the hidden trees during the first round scan. The initial scan was acquired from the central position of the sample plot and the second scan was positioned close to the first scan but in a position that enabled it to view the hidden trees. Field validation data, which includes DBH, number of stems per plot and total tree height were collected manually using conventional forest inventory methods at the same time. A total of three randomly selected trees were felled in each plot and their diameter was measured at half meter intervals for validation purposes. The same process in the same stand using the same scan positions was repeated after two years for monitoring growth.

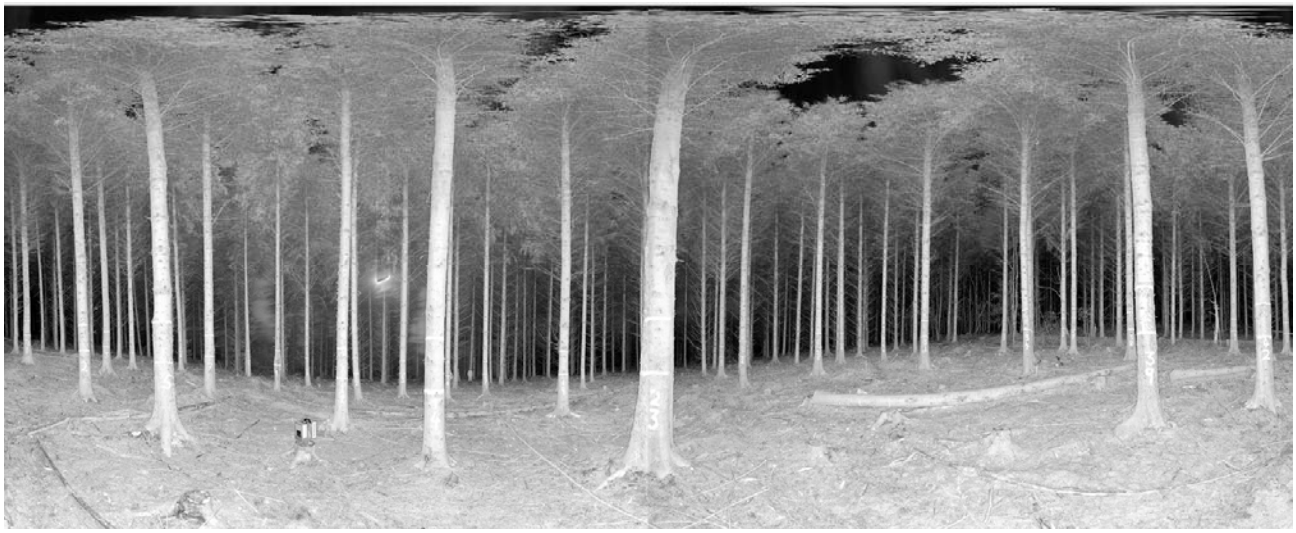


Fig 1. Intensity image of terrestrial scanner data in the sample plots of 15 meters radius

Pre-processing

During the acquired point cloud data pre-processing, the data were flittered and ghost points which can be caused by ambiguity of phase measurement technology were removed. The Autostem software clipped the parts of the point cloud data within 15 meters radius for further processing. The software uses a density allocation along the Z axis of a DTM patch for extracting the DTM without ghost points. This enables it to determine the ground level or the lowest point of a tree.

The stem profile along tree height at half meter intervals was derived from the point cloud data using the Autostem starting from a reference point of DBH for selected tree numbers. The principle behind this processing was based on a least square circle fitting algorithm (Bienert et al., 2006). The reliability factor

introduced by Bienert et al. (2007) was implemented in this processing and the circular fitting algorithm with a reliability factor greater than 85 percent was used to avoid the impact of unreliable data during model fitting.

3. Preliminary Result

A number of measurements were extracted from the point cloud data and compared with field measurements. The preliminary results after processing all 9 plots and the felled trees are summarized as follows. The removal of the lower branches up to 6 meter didn't show a statistical significant difference in diameter estimation. This can be seen in Figure 2, which depicts the range of residuals of diameter from field measurement along the stem compared to the diameter derived from the point cloud data for the scans acquired before and after pruning. The smaller residual values are recorded for the pruned stands, especially up to the height of 10.7 m. The majority of the tree diameters between 10.7 m and 15.3 m are underestimated in a scanning acquired before and after pruning. This could be due to the effect of branches as pruning is only done up to 6 m. Overestimation of diameter in the stem profile is observed between 19.9 m and l24.5 m. The effect of extrapolation applied to estimate the top part of the trees based on the DBH of individual trees might be the reason for the overestimation.

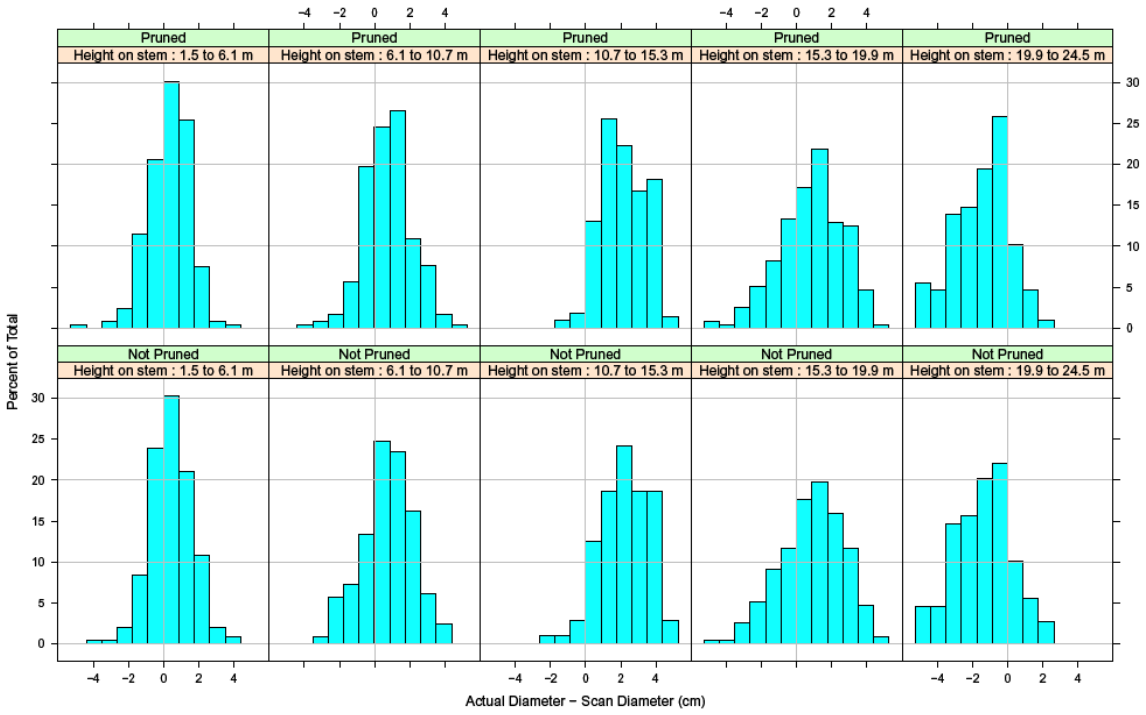


Figure 2. Distribution of percentage diameter difference along stem profile, between actually measured diameter data and diameter derived from point cloud data, within 95 percent confidence interval. The residuals at the upper histograms are for plots after pruning and the bottom histograms show the residuals for plots before pruning.

Volume is an important parameter for the price of a stem as well as for the investigation of tree growth. By using repeated scanning in the same plot for several periods of time, forest growth can be calculated for various fields of application such as ecological monitoring and carbon budgeting.

Accordingly, diameter increment over the period of two years, from November 2007 till November 2009, was calculated after deriving a diameter for each and every tree in 7 plots in the forest stand. Table 1 shows the summary of average DBH increment observed from field measurement data using callipers and diameter derived from point cloud data over the two years period. An overall difference (i.e. growth) of 0.05 cm can be observed.

Table 1.Average DBH increment per plot from field measurements and from TLS data over a period of two years.

4.

Plot	DBH increment from actual measurement(cm)	DBH increment from derived data (cm)	STDEV of actual DBH	STDEV of derived DBH
	1.87	1.83	0.67	1.02
2	2.10	1.90	0.74	1.17
3	1.98	2.07	0.61	1.60
4	2.14	2.33	1.49	0.71
5	2.68	2.12	0.80	1.17
6	1.92	1.98	0.74	0.88
7	2.17	2.19	0.74	1.10

5. Conclusions and future work

Based on the preliminary result of the pilot study, we are able to show that terrestrial LiDAR enables the acquisition of forest stand parameters, such as diameter profile and diameter increment, with an acceptable accuracy. The multi- temporal terrestrial LiDAR data can be used for precise volume (growth) calculation if supported with stand and site specific field information.

Future works will focus on cost benefit analysis of TLS for forest inventory and on the determination of cost effective and efficient sampling techniques by using different plot numbers and setups in stands scheduled for clearfelling.

Acknowledgement

We are grateful to COFORD (National Council for Forest Research and Development) for funding the FORESTSCAN project. PTR Ltd. and Treemetrics Ltd. are our research partners in this sudy.

Reference

Bienert , A., Scheller, S., Kean, E., Mohan F.,and C.Nugent. 2007, Tree detection and Diameter estimations by analysis of forest terrestrial laser scanners point clouds International Archives of Photogrametry, Remote Sensing and Spatial Information Sciences, XXXVI, Part 3/W52.

Bienert, A., Scheller , S., Kean, E., Mullooly, G., and Mohan F., 2006, Application of terrestrial laser scanners for the determination of forest inventory parameters. International Archives of Photogrametetry, Remote Sensing and Spatial Information Sciences, Vol. 36, Part 5.

Danson, F.M., Hetherington, D., Morsdorf, F., Koetz, B., and Allgower, B. 2007. Forest canopy gap fraction from terrestrial laser scanning. IEEE Geoscience and Remote Sensing Letters 4(1): 157-160.

Henning, J.G., and Radtke, P.J. 2006a. Ground-based laser imaging for assessing three-dimensional forest canopy structure. Photogrammetric Engineering and Remote Sensing 72(12): 1349-1358.

Henning, J.G., and Radtke, P.J. 2006b. Detailed stem measurements of standing trees from ground-based scanning lidar. Forest Science 52(1): 67-80.

Hopkinson, C., Chasmer, L., Young-Pow, C., Treitz, P., 2004. Assessing forest metrics with a ground-based scanning LiDAR. Can.J. For Res, 34, 573-583.

Maas, H.G., Bienert, A., Scheller, S., and Keane, E. 2008. Automatic forest inventory parameter determination from terrestrial laser scanner data. International Journal of Remote Sensing 29(5): 1579-1593.

Thies, M., Pfeifer, N., Winterhalder, D., and Gorte, B.G.H. 2004. Automatic determination of forest inventory tile. In Scandlaser Workshop 2003. Taylor & Francis As, Umea, SWEDEN. pp. 571-581.

Tansey, K., Selmes, N., Anstee, A., Tate, N.J., and Denniss, A. 2009. Estimating tree and stand variables in a Corsican Pine woodland from terrestrial laser scanner data. International Journal of Remote Sensing 30(19): 5195-5209.

Single Tree Canopy Projection Area Extraction from Single-scan Terrestrial Laser Scanner Data

PUTTONEN, E.†, LITKEY, P., LIANG, X., KAARTINEN, H. and KUKKO, A.*

eetu.puttonen@fgi.fi

†Finnish Geodetic Institute, Geodeetinrinne 2, 02431 Masala, Finland

Abstract

In this paper, a novel method is presented for canopy projection area extraction from terrestrial laser scanning (TLS) data. While merged multi-scan TLS data provides whole canopy coverage and gives an accurate input for canopy studies, single-scan data is of a great practical interest for its low collecting and processing cost. In single-scan TLS data, laser points cover precisely only the tree side facing the scanner. Therefore, a compensating mechanism is needed to calculate the canopy projection area for the whole tree in single-scan TLS data.

Individual tree crowns are manually delineated using their trunk positions as starting points. The proposed method densifies the scanned and delineated point cloud in a mirroring process where tree crown points are copied with respect to a mirror plane. The mirror plane is defined with an orthogonal vector base that is set to the tree trunk centre. The orthogonal base is constructed from the laser point distribution and the scanner location. The boundaries of the densified point cloud are searched by filtering its two-dimensional Delaunay triangulation. The canopy projection area estimation is then calculated from the filtered triangulation. Merged multi-scan point clouds of each tree are used as references for canopy projection area estimate. The results of point cloud densification, the boundary search procedure, and their effects on the canopy projection area estimation are presented and discussed.

The canopy projection area estimates were calculated from both original and mirrored single-scan mode laser point data. Canopy projection area estimates measured with the single-scan mode had clear variations for the same tree depending on the scanner location and its field of view. The results showed that canopy projection area estimates calculated with the presented mirroring and boundary search steps correspond with the results from reference data, but that careful parameter selection is needed.

The presented method for canopy projection area estimation with mirroring and point cloud outline search steps allows fast data collection of individual trees on a plot level.

1. Introduction

Canopy projection area of a single tree crown, also known as dimensionless canopy cover, is a basic parameter collected in forest studies as an end or an intermediate product. For example, in forest inventory it is used in timber volume estimation (Maltamo *et al.* 2004) and in forest fuel studies (Riano *et al.* 2003). In ecological studies, it is used in natural habitat mapping (Hyde *et al.* 2005) and in physically based radiation models (e.g. Kuusk and Nilson (2000), Huemmrich (2001), Jensen *et al.* (2008)). It is also used in urban forest inventory and planning (for example, McPherson *et al.* (1999) and Walton *et al.* (2008)).

The canopy projection area is traditionally measured in a wide variety of methods on ground level. These methods include a line-intersection method, various viewing tube measurements, (fish-eye) cameras, densiometers, and ocular inspection. Several studies have compared and discussed traditional ground measurement methods and their performance (Jennings *et al.* 1999, Korhonen *et al.* 2006, Fiala *et al.* 2006). These studies have reported similar results. Conventional methods, such as the line intersection method, were found to be the most accurate but also laborious and time-consuming. Faster methods (e.g. digital photography) were prone to biasing errors and sensitive to the viewing geometry. Jennings *et al.* (1999) and

Fiala *et al.* (2006) pointed out that the selection of the measurement method is situational and that there is no single method that would excel others in most circumstances. All studies expected remote sensing solutions to become more important in canopy projection area and cover assessment studies when suitable methods are developed.

More recently, terrestrial laser scanning (TLS) has been shown to be a practical technology for forest parameter retrieval. Applications have been developed to extract parameters, such as tree location, height and diameter, and canopy structure and gap fraction (Aschoff *et al.* 2004, Bienert *et al.* 2006, 2007, Liang *et al.*2008, 2009, Litkey *et al.* 2008, Maas *et al.* 2008, Simonse *et al.* 2003, Thies *et al.*2004, Wezyk *et al.* 2007). Overall, TLS provides high accuracy spatial data thatscales from a single tree to a plot level. TLS data collection is both labour- and cost-effective, which adds to its appeal.

TLS data collection can be done in both single- and multi-scan modes. Multiply scanned datasets have comprehensive coverage over the scanned area and a high level of detail. However, the measurement of a multi-scan dataset takes additional time in the field and extra work during data processing when different single-scan datasets are co-registrated and combined (Bienert *et al.* 2006). Single-scan data collection gives a clear improvement in both measurement and processing times (Bienert *et al.* 2006, Liang *et al.* 2008, Litkey *et al.* 2008). Single-scanned data can be also handled more automatically than multi-scan data. The drawback of single scanned datasets is their high sensitivity to occlusions that are present even in sparse forests.

This study presents a novel method for calculating a canopy projection area from single-scanned TLS data. The canopy projection area calculation is implemented in two steps. First, a single-scanned point cloud is completed in a mirroring procedure which creates a mirrored image of the original point cloud across a mirror plane. The mirror plane is placed in the tree trunk and aligned with it. Second, a 2D Delaunay triangulation of the point cloud is calculated. Then, the triangulated area's outline is searched and filtered with a threshold and a layer index parameter. The canopy projection area is calculated from the outlined triangulation.

The effectiveness of the canopy projection area calculation is tested by comparing the calculated areas of both the original and the mirrored laser point clouds with each other. The surface areas of single-scan tree point clouds are compared with the one calculated from a combined single-scan point cloud. The combined point cloud is used as an evaluation reference. Surface area calculated from mirrored single-scan point cloud is supposed to correspond with a relatively small variance to that combined from multiple scans.

2. Dataset

Tree data and field measurements

The TLS data were collected in Koli, Eastern Finland, in June 2006. The study plot was located in a managed Scots pine forest with homogenous and even-aged tree specimen. The study plot was open with little understory and young trees. The data consisted of 11 Scots pine (*Pinus sylvestris*) trees and were scanned from four positions. Each tree was scanned and registered from 3 to 4 locations depending on their visibility to the scanning locations. Same measurement resolution was used for all scans, producing a point spacing of 6 mm at the distance of 10 m. Data were collected with a FARO LS 880HE80 laser scanner (FARO, Lake Mary, USA). Individual scans were georeferenced to a local coordinate system using spherical reference targets. Coordinates for the reference targets were measured using a Trimble 5602 DR 200+ total station (Trimble Navigation Limited, CA,USA), which was set up using the known coordinates of the rectangle-shaped test plot corners. Tree and scanner locations are illustrated in figure 1.

The scanned point clouds were filtered after georeferencing to reduce the amount of outlier points. Outliers occur in phase-based measurement systems when the measuring beam does not

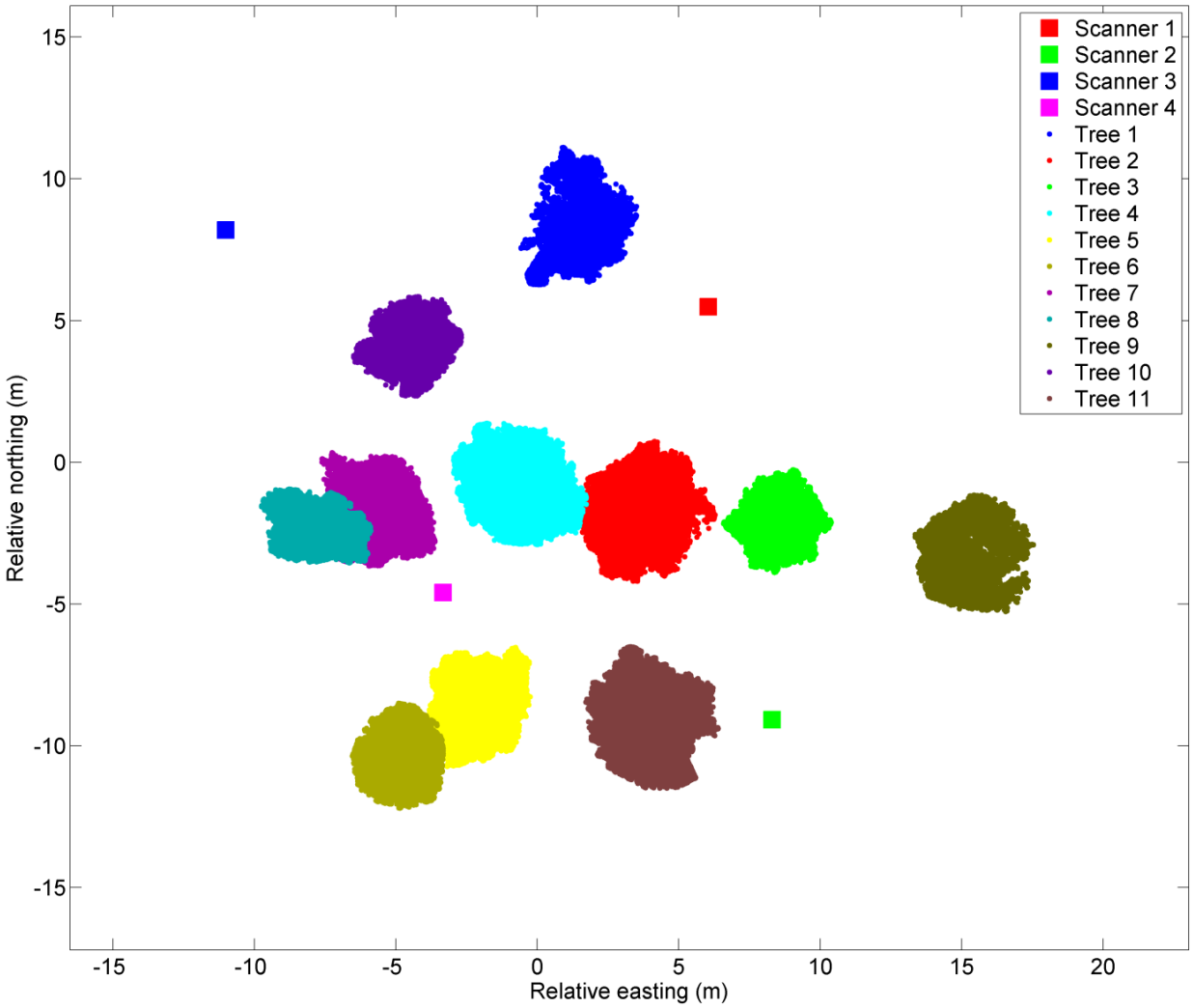


Figure 1. Location of the scanning points and the combined point clouds of all studied trees. Distances in the figure are measured from the centre point set in the middle of all scanning locations.

hit anything or when it hits several objects simultaneously. Filtering was carried out by removing points that had a greater distance than 20 cm to half of its 3 by 3 neighbouring points in scanner's row-column system. All points with a returning intensity level below 800 out of 2044 were also removed.

Tree delineation

An individual tree point cloud was separated from the original point cloud by detecting first the trunk of the tree automatically and then manually delineating other canopies around it. The tree trunk detection method was presented in Liang *et al.* (2009). The method searches for three features for a single laser point from its neighbourhood, namely the neighbouring point density, flatness of the neighbouring region, and the neighbouring region's normal vector direction.

The flatness of the point neighbouring region was calculated from its covariance values and the respective eigenvectors derived from them. Points, whose neighbouring regions had a high local point density, two large principal components in point coordinate system and a close-to-horizontal normal vector in world coordinates, were most likely candidates to belong into a tree trunk.

Possible trunk points were clustered and identified to remove misclassifications and noise points. The clustering was done in two parts. In the first part, possible trunk points within a given distance from each other were grouped together. If the formed point group was mainly vertically distributed, it was labelled as a tree trunk. The second clustering step consisted of aggregating the trunk point groups in the horizontal plane with each other. This ensured that the point groups belonging to the same tree were merged together in the case they were separated by shadows.

The clustered trunk points were further modelled with a set of stacked circles placed along the tree trunk. Possible noise points, e.g. from branches, were filtered from trunk point clouds before fitting. After filtering, the tree trunk was divided into 20 cm thick slices and a single circle was fitted for each slice. The circle-fitting was performed as a least squares fit, where the sum of Euclidean distances between each point in a trunk slice and the searched circle was minimized. The validity of each fitted circle was inspected by comparing the circle centre location and it's radius with other fitted circles. If the fitted circle's radius or centre point location did not fit into the statistical variance of all fitted circles, it was discarded.

The tree canopies were manually delineated from individually scanned tree point clouds. The delineation was carried out by drawing a curve around the chosen tree in several viewing planes. First, all points at maximum distance of 4 m from the modelled trunk centre line were found and collected. Second, planes parallel to the modelled trunk centre line were constructed so that they covered semicircle. Scanned laser points that were at most at 30 cm distance from a plane were projected on it and a curve enclosing the tree points was drawn manually. All points that were selected from any projection plane were included in the delineated tree point cloud. The delineated point cloud was completed with an outlier filtration. Outliers were removed from the final point cloud based on their projected locations in the ground plane.

Tree point cloud mirroring

A single scanned point cloud cannot give complete information from the back side of a tree, as the laser beam is unable to penetrate through whole canopy. Also, it is not always possible to create a combined and georeferenced tree point cloud from multiple TLS scans, e.g. due to measurement time limitations. This complicates the canopy projection area calculations. Therefore, an estimating procedure for completing a single scanned tree canopy point cloud was developed and tested. The procedure is based on mirroring the delineated tree points over a predefined mirror plane.

The mirroring procedure is started with defining a mirror point for the delineated tree point cloud. The mirror point is set in the centre of the tree trunk. Then, a new orthonormal basis is constructed in three steps: First, a new principal axis is calculated from the trunk point covariance matrix, where the principal component is the largest. Second, another base vector is chosen so that it points towards the scanner in the local xy-plane while being orthogonal with respect to the previously defined base vector. The second eigenvector represents the normal of the mirror plane and the mirroring is done with respect to it. Third, the new orthonormal base is completed by setting a third base vector that is orthogonal with the previous two.

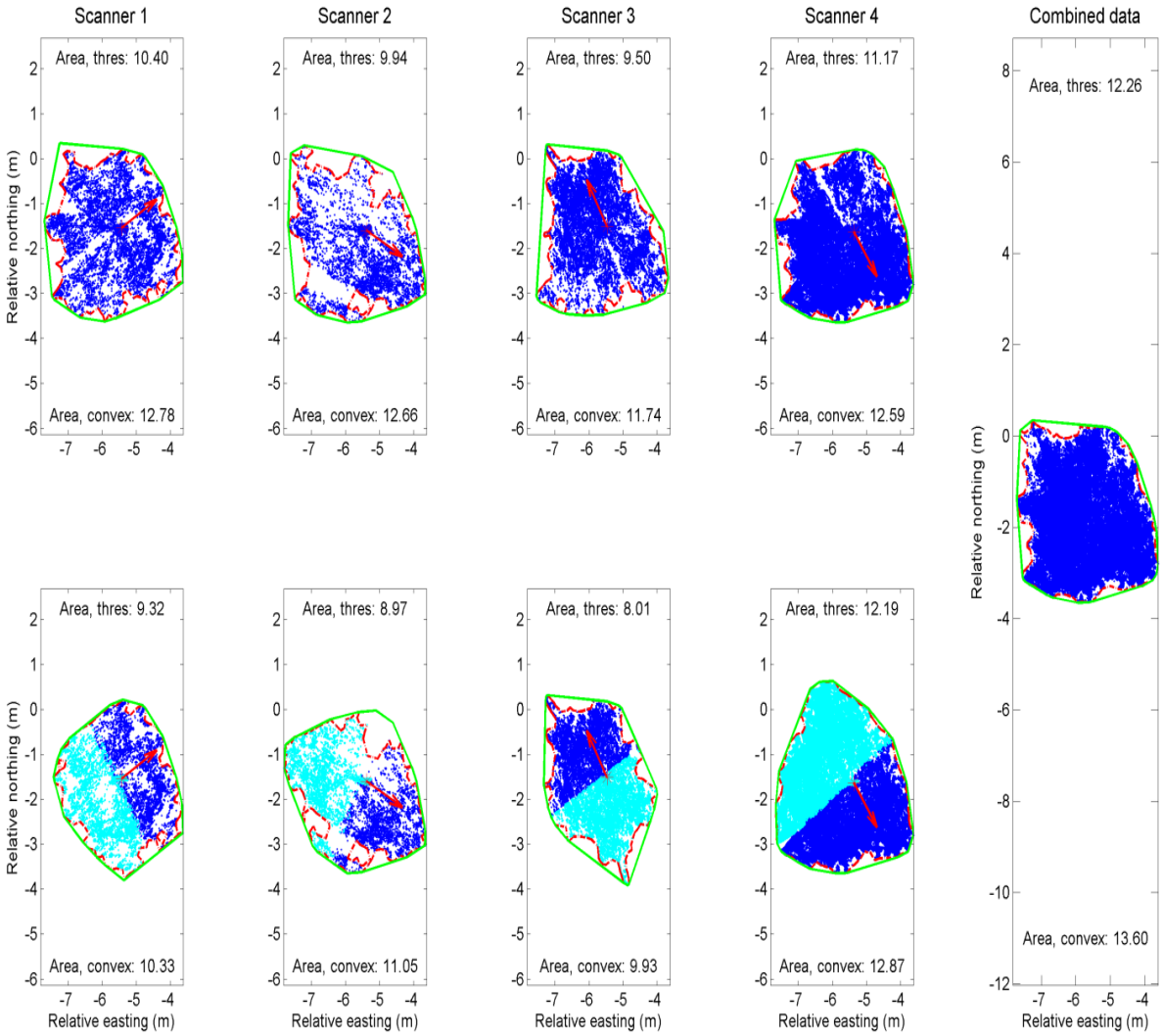


Figure 2. Results of the point cloud mirroring and outline search of the tree 7. Upper row Delineated point clouds of the tree 7 illustrated from above. The red asterisk and the arrow show the location of the tree trunk centre and the direction to the scanning point. The red, dash-dotted line illustrates how the edge search algorithm has defined the projection plane outline. The green line depicts the outline of the convex hull. Tree projections areas calculated for both outline cases are written above and below of the point clouds. Lower row Blue points are the original points in front of the mirror plane. Cyan coloured points are copied across the mirror plane to the back side. The rightmost subfigure Combined point cloud consisting of the original points of all scans. All distances are related to the centre of all scanning points.

After the orthogonal base is formed, dot products between the mirror plane normal and vectors drawn from the chosen tree centre point towards the measured laser points are calculated. The points that have a dot product that is larger than or equal to zero are on the front side with respect to the mirror plane and are retained. All other points are behind the mirror plane and they are removed. The mirroring procedure is finalized by creating mirrored copies of the retained laser points. The mirroring can be written as

$$\mathbf{P}_m = (\mathbf{P}_{orig} - \mathbf{P}_{plane}) - 2 \cdot ((\mathbf{P}_{orig} - \mathbf{P}_{plane}) \cdot \mathbf{n}_{plane}) \cdot \mathbf{n}_{plane}, \quad (1)$$

,where $\mathbf{P}_m$ is a location of a mirrored point, $\mathbf{P}_{orig}$ is the location of the original point, $\mathbf{P}_{plane}$ is a point located in the mirror plane, and $\mathbf{n}_{plane}$ is the mirror plane normal vector. The copied points are shifted to

the opposite side of the mirror plane against the normal vector direction. Figure 2 illustrates the effect of the mirroring procedure on tree point clouds.

Tree crown surface area calculation

The calculation of tree crown surface areas was done by doing a 2D Delaunay triangulation on the delineated point cloud in the xy-plane. The definition of the 2D Delaunay triangulation is that no points exist within any circumcircle of any triangle in a triangulated point set (de Berg *et al.* 2008). The definition is generalizable to n dimensions.

The initial Delaunay triangulation connected all points in the delineated, and possibly mirrored, tree point cloud. To outline the point cloud precisely, the triangulated points were labelled in concentric layers, where the outermost point layer was labelled as the first layer. After layering, the triangulation was filtered to remove possible outlier points left after the delineation. Then, a new search was made, where triangle side lengths were compared against a preset threshold length. If a single side of the triangle was longer than the set threshold length, the whole triangle was removed from the triangulation. The triangle side length search was limited to the few outermost point layers to prevent too excessive triangle removal caused by internal shadowing.

The canopy projection area was then calculated by summing over areas of all accepted triangles. If the triangle side threshold was set arbitrarily large, the resulting area corresponded the one calculated for the convex hull of the same point cloud. This correspondence gave another reference to compare the areas calculated from different scan locations with each other. Figure 2 shows the results of point cloud outline search for one of the studied trees with both short and very long threshold lengths.

3. Results

Point clouds of eleven Scots pines were selected to test the effects of the mirroring and the outlining procedures in canopy projection area estimation. Each selected pine was registered at least in three different scanning locations. Projection plane areas were calculated for every selected tree from all scanning locations with both a short 25 cm threshold length and with the convex hull. Calculated projection plane areas were compared with the reference area. The reference area consisted of merged single-scan point clouds. Area comparison results with the best and the worst estimations of all scans are presented in table 1.

In general, two distinctive results are seen: the mirrored point clouds give the best projection area estimates in more than half of the cases. Another obvious result is the variability between the best and the worst area estimations that are relatively large in both original and mirrored point clouds. Large variances between different scans are dependent on the visibility of a tree to different scanning points. When the tree of interest has been occluded, the amount of laser hits on the canopy has decreased to very low level making area estimates very unreliable. Table 1 illustrates that the area estimates with the most variance are usually derived from scans where the point count for the tree has been less than 5000 hits. Low point count usually results in a severe area underestimation. On the other hand, the best area estimates are more dependent on how comprehensively a single scan has captured the whole form of the tree canopy. Thus, the scan with the highest point returns does not necessarily give the best projection area estimate. The results show that the best results are usually obtained with over ten thousand or more returns in the dataset.

The effect of the laser point cloud mirroring

The best area estimates calculated from mirrored laser point clouds corresponded reference areas better than the areas calculated from original point clouds. Variance between the single-scan area estimations was

still large. The most influential reason for large variances in area estimates was due to the mirroring assumption itself: shapes of tree canopies growing in a forest are not symmetric in respect to any particular plane in general.

Another variance inducing factor was that the mirror plane was placed along the tree trunk direction. This plane alignment was chosen because it corresponded a real tree shape better than a mirror plane set along the world vertical axis. The mirror plane selection affected the results because tree trunks were tilted a little from the world vertical axis. Depending on the scanner location and the direction of the tree tilt, there were additional over- or underestimates to the canopy projection area. If the tree was tilted towards the scanner, the mirrored point cloud looked like it was folding on itself in the middle when the viewing direction was chosen to be along the world vertical axis. The folding caused underestimation to the canopy area. The folding case is illustrated in figure 2, under 'Scanner 1'. In an opposite case, the mirroring caused point cloud to spread on edges leading into a possible area overestimation.

Point cloud mirroring cannot remove shadows caused by other tree trunks or by dense parts of canopies that occlude the scanner's line of sight to the studied tree. This error type was impossible to correct with mirroring as it happened in perpendicular to the mirror plane. Figure 2, 'Scanner 2' illustrates this situation. The mirroring procedure should be used when there are no reliable data available from the backside of a scanned tree, which is common when there are dense foliage and low branches blocking the line of sight to the scanner. Mirroring should give reliable projection area estimates when a close symmetry across the mirror plane is expected, which is the case in open areas with little understory, such as managed forests or parks.

The effect of the triangulation threshold

Canopy projection area estimates were calculated with two different outline thresholds. The shorter threshold was set to 25 cm that was found to cover the most of the canopy areas and shaded gaps, while still following the canopy outline in detail. The long threshold was set very large, in order of several meters, which made it correspond the convex hull of the point cloud.

The best and worst canopy projection area estimates varied less with the convex hull outlining than with the short outline threshold. This is because the convex hull includes all delineated points, minimizes the number of edge points, and connects them directly to each other. Thus, the estimated canopy projection area is insensitive to possible shadows from other vegetation and relatively sparse point clouds as long as there are any hits returning from the tree canopy. The convex hull estimate works with a relatively good accuracy, if the returns are coming from the whole tree area. In such a case, the number of returned laser hits can be low. However, the use of convex hull estimate requires that the canopy delineation is accurate as any outlier points left after the delineation will be included in the area estimate. This leads to large overestimates in projection area calculations.

Canopy projection area estimates calculated with the 25 cm threshold outline point clouds in detail (figure 2). The detailed outlining produces systematically smaller canopy projection areas than the convex hull. The chosen threshold length is long enough to cover possible trunk shadow

Table 1. Projection plane area estimations. Tree areas and their estimates are italicized. The best and the worst estimates out of all individual scans are given in relative units.

Areas of scanned trees						
		Combined	Original		Mirrored	
	Thres	Area (m^2)	Best	Worst	Best	Worst
Tree 1	0.25	*Nov 79*	*-10.4*	*-71.7*	*02. Mrz*	*-14.8*
	Points	61311	47570	2276	19102	3376
	Convex	*14.00*	*-5.1*	*-38.9*	*01. Jan*	*08. Sep*
	Points	61311	47570	2276	3376	75345
Tree 2	0.25	*15.26*	*-14.2*	*-60.0*	*06. Apr*	*-35.2*
	Points	101368	41814	11065	48336	4239
	Convex	*16.94*	*-6.6*	*-30.6*	*-7.5*	*-22.6*
	Points	101368	32238	11065	32251	45334
Tree 3	0.25	*Aug 39*	*-9.3*	*-20.0*	*-1.4*	*-7.9*
	Points	96887	42281	11065	70070	48034
	Convex	*Sep 24*	*-3.2*	*-8.0*	*-0.8*	*-8.0*
	Points	96887	42281	11065	16033	48034
Tree 4	0.25	*14.62*	*-7.3*	*-35.0*	*-5.9*	*-52.8*
	Points	200129	63645	5264	21591	4977
	Convex	*15.74*	*-3.9*	*-16.9*	*-3.6*	*-21.4*
	Points	200129	63645	16397	21591	4977
Tree 5	0.25	*Nov 47*	*-4.8*	*-30.3*	*-8.1*	*-38.4*
	Points	209175	164230	10160	200785	4315
	Convex	*Dez 46*	*-3.1*	*-19.4*	*-2.6*	*-19.7*
	Points	209175	164230	3462	37819	4315
Tree 6	0.25	*Aug 90*	*-4.7*	*-79.7*	*-21.2*	*-74.2*
	Points	104994	83318	1465	92632	1519
	Convex	*Sep 19*	*-2.6*	*-41.5*	*-13.9*	*-44.4*
	Points	104994	83318	1465	20249	1519
Tree 7	0.25	*Dez 26*	*-8.9*	*-22.5*	*-0.6*	*-34.7*
	Points	175745	134246	20949	199157	24768
	Convex	*13.60*	*-6.0*	*-13.7*	*-5.4*	*-27.0*
	Points	175745	12880	20949	199157	24768
Tree 8	0.25	*Jul 57*	*-3.6*	*-43.9*	*10. Jul*	*-44.6*
	Points	63780	50459	2917	80234	10200
	Convex	*Aug 26*	*-0.4*	*-32.8*	*10. Aug*	*-37.9*
	Points	63780	50459	2917	80234	3625
Tree 9	0.25	*Dez 19*	*-11.1*	*-77.8*	*-0.7*	*-45.1*
	Points	46477	28019	2054	21115	3216
	Convex	*13.32*	*-2.9*	*-41.6*	*0.2*	*15. Mai*
	Points	46477	28019	2054	21115	44249
Tree 10	0.25	*Aug 55*	*-10.8*	*-72.7*	*-10.5*	*-71.0*
	Points	80641	27148	1724	55851	1884
	Convex	*Sep 29*	*-4.6*	*-19.1*	*-2.3*	*-32.3*
	Points	80641	27148	1724	55851	1884
Tree 11	0.25	*15.48*	*-5.8*	*-99.7*	*-9.8*	*-99.7*
	Points	165722	121972	782	47713	869
	Convex	*17.14*	*-2.1*	*-72.9*	*2.0*	*-71.9*
	Points	165722	121972	782	47713	869

regions. The short threshold is sensitive to the low return point densities. The sensitivity leads to the exclusion of point returns that are located on the backside of the point cloud, which causes an

underestimated canopy projection area. Such underestimations usually happen when a scanned tree is located a long distance away from the scanner or when its canopy is dense and the number of point returns coming from the backside of the tree is low.

The results between different threshold lengths show that the threshold value should be defined according to the available data. Data with high laser return density and some noisiness are handled best with a relatively short threshold value combined to an outermost layer filtration. The convex hull outlining is better suited for datasets with fewer returns and noise points. Convex hull routines are full-fledged and efficient which favours them in studies where a single tree projection area does not need to be measured in high detail.

4. Conclusions

A new method for canopy projection area estimation for a single-scanned tree point cloud was introduced in the study. The area estimation was calculated in two steps. First step involved a single-scan point cloud completion that was done by mirroring the original point cloud in respect to a chosen mirror plane. The mirror plane was placed in the tree trunk and aligned with it. The point cloud completion step was introduced because it was expected to correspond the actual canopy shape better than the self-shaded side of the original canopy. The second data processing step used a 2D Delaunay triangulation of the laser point cloud. All triangulated points were sorted in layers. The layered triangulation was then filtered with a predefined triangle side threshold and a layer order parameter. The canopy projection area was calculated by summing over all preserved triangle areas.

The developed data processing steps were tested on a dataset consisting of 11 Scots pines that were scanned from three to four different locations. Canopy projection areas were calculated for each single scan with mirrored and original point clouds using a short and a long threshold length. Tree-wise canopy projection areas calculated from combined multiple scan data were used as a reference. The results showed that canopy projection area estimates calculated with mirrored single-scan data agreed with the references in the best cases. However, calculated projection areas of a single tree displayed clear variation between different scans. Large variances in tree canopy cover have been also reported in other studies where different canopy cover estimation methods were compared with each other (Fiala *et al.* 2006, Korhonen *et al.* 2006). Thus, a careful parameter selection is needed as the projection area calculations are sensitive to available point number and density, tree crown delineation precision, and occlusions in the scanner's field of view.

The method performed the best against canopy projection areas calculated from original point clouds in cases where some level of tree crown symmetry can be expected. Canopy projection area estimates for canopies with significant internal shading are expected be also more accurate with mirroring.

In future, the presented study can be expanded in two ways: first, by improving the precision of both the mirroring and the outline search procedures, and second, by doing comprehensive projection area measurements with large datasets and comparing them with results obtained from ALS and traditional forest surveys.

The point cloud outlining routine can be extended straight forward into three dimensions. In a three-dimensional case, the triangulation is represented with tetrahedra instead of triangles and the canopy and stem volumes can be calculated in addition to the surface area. However, TLS data does not cover tree top in plots where there are extensive canopy overlap and occlusion. This is likely to make volume estimates based on the introduced method more sensitive to point cloud density than area estimates in the two-dimensional case.

References

ASCHOFF, T., THIES, M. and SPIECKER, H., 2004, Describing forest stands using terrestrial laser-scanning. In Proceedings of the XXth ISPRS Congress: Geo-Imagery Bridging Continents, pp. 12–23.

BIENERT, A., SCHELLER, S., KEANE, E., MOHAN, F. and NUGENT, C., 2007, Tree detection and diameter estimations by analysis of forest terrestrial laser scanner point clouds. In Proceedings of the International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Proceedings of ISPRS Workshop on Laserscanning, XXXVI-3/W52, pp. 50–55.

BIENERT, A., SCHELLER, S., KEANE, E., MULLOOLY, G. and MOHAN, F., 2006, Application of terrestrial laser scanners for the determination of forest inventory parameters. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36.

de BERG, M., CHEONG, O., van KREVELD, M. and OVERMANS, M., 2008, Computational Geometry: Algorithms and Applications, 3rd. rev. (Springer-Verlag).

FIALA, A., GARMAN, S. and GRAY, A., 2006, Comparison of five canopy cover estimation techniques in the western Oregon Cascades. Forest Ecology and Management, 232, pp. 188–197.

HUEMMRICH, K.F., 2001, The GeoSail model: a simple addition to the SAIL model to describe discontinuous canopy reflectance. Remote Sensing of Environment, 75, pp. 423 – 431.

HYDE, P., DUBAYAH, R., PETERSON, B., BLAIR, J., HOFTON, M., HUNSAKER, C., KNOX, R. and WALKER, W., 2005, Mapping forest structure for wildlife habitat analysis using waveform lidar: Validation of montane ecosystems. Remote Sensing of Environment, 96, pp. 427 – 437.

JENNINGS, S., BROWN, N. and SHEIL, D., 1999, Assessing forest canopies and understory illumination: canopy closure, canopy cover and other measures. Forestry, 72, pp. 59.

JENSEN, J.L., HUMES, K.S., VIERLING, L.A. and HUDAK, A.T., 2008, Discrete return lidar-based prediction of leaf area index in two conifer forests. Remote Sensing of Environment, 112, pp. 3947 – 3957.

KORHONEN, L., KORHONEN, K., RAUTIAINEN, M. and STENBERG, P., 2006, Estimation of forest canopy cover: a comparison of field measurement techniques. Silva Fennica, 40, pp. 577–588.

KUUSK, A. and NILSON, T., 2000, A Directional Multispectral Forest Reflectance Model. Remote Sensing of Environment, 72, pp. 244 – 252.

LIANG, X., LITKEY, P., HYYPPÄ, J., KAARTINEN, H., VASTARANTA, M. and HOLOPAINEN, M., 2009, Automatic stem location mapping using TLS for plotwise forest inventory. In Proceedings of the Silvilaser 2009, Oct. 14 16 College Station, Texas, US, S. C. Popescu, R. F. Nelson, K. Zhao, & A. Neuenschwander (Ed.), p. 9.

LIANG, X., LITKEY, P., HYYPPÄ, J., KUKKO, A., KAARTINEN, H. and HOLOPAINEN, M., 2008, Plot-level trunk detection and reconstruction using one-scan-mode terrestrial laser scanning data. In Proceedings of the Earth Observation and Remote Sensing Applications, 2008. EORSA 2008. International Workshop on, pp. 1–5.

LITKEY, P., LIANG, X., KAARTINEN, H., HYYPPÄ, J., KUKKO, A. and HOLOPAINEN, M., 2008, Single-scan TLS methods for forest parameter retrieval. In Proceedings of the Silvilaser 2008 proceedings, R. Hill, J. Rossette and J. Suarez (Eds), 4132, pp. 295 – 304.

MAAS, H., BIENERT, A., SCHELLER, S. and KEANE, E., 2008, Automatic forest inventory parameter determination from terrestrial laser scanner data. International Journal of Remote Sensing, 29, pp. 1579–1593.

MALTAMO, M., EERIKÄINEN, K., PITKÄNEN, J., HYYPPÄ, J. and VEHMAS, M., 2004, Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. Remote Sensing of Environment, 90, pp. 319–330.

McPHERSON, E., SIMPSON, J., PEPER, P. and XIAO, Q., 1999, Benefit-cost analysis of Modesto's municipal urban forest. Journal of Arboriculture, 25, pp. 235–248.

RIANO, D., MEIER, E., ALLGÖVER, B., CHUVIECO, E. and USTIN, S., 2003, Modelling airborne laser scanning data for the spatial generation of critical forest parameters in fire behaviour modelling. Remote Sensing of Environment, 86, pp. 177–186.

SIMONSE, M., ASCHOFF, T., SPIECKER, H. and THIES, M., 2003, Automatic determination of forest inventory parameters using terrestrial laser scanning. In Proceedings of the ScandLaser Scientific Workshop on Airborne Laser Scanning of Forests, 2003, pp. 252–258.

THIES, M., PFEIFER, N., WINTERHALDER, D. and GORTE, B., 2004, Three-dimensional reconstruction of stems for assessment of taper, sweep and lean based on laser scanning of standing trees. Scandinavian Journal of Forest Research, 19, pp. 571–581.

WALTON, J., NOWAK, D. and GREENFIELD, E., 2008, Assessing urban forest canopy cover using airborne or satellite imagery. Arboriculture and Urban Forestry, 34, pp. 334–340.

WEZYK, P., KOZIOL, K., GLISTA, M. and PIERZCHALSKI, M., 2007, Terrestrial Laser Scanning versus Traditional Forest Inventory. First Results from the Polish Forests. In Proceedings of the ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007, September 12-14, 2007, Espoo, Finland, Volume XXXVI, Part 3/W52, pp. 424–429.

Crown transparency assessment based on terrestrial LiDAR

Nicole Legner†, Stefan Fleck‡, Dominik Seidel†, Christoph Leuschner†*

nlegner@gwdg.de,

† Department of Plant Ecology and Ecosystem Research, Albrecht von Haller Institute for Plant Sciences, University of Göttingen, Untere Karspüle 2, 37073 Göttingen, Germany;

‡ Department of Intensive Environmental Monitoring, Nordwestdeutsche Forstliche Versuchsanstalt, Grätzelstr. 2, 37079 Göttingen, Germany;

Abstract

The aim of this study is to compare crown transparency of mature forest trees with the help of high-resolution terrestrial LiDAR measurements of whole tree crowns. The space occupied by single branches and their appending leaf area was cut from the 3D-point cloud and their volume was approximated with a voxel model. Volumetric leaf area densities of 10 mid-size branches of the species *Fagus sylvatica*, *Carpinus betulus*, *Acer pseudoplatanus*, *Fraxinus excelsior* and *Tilia cordata* were determined based on hand measured leaf area of each branch. Analogously, volumetric area densities of woody tissue were calculated using the projected area of all woody material of each branch that was harvested in diameter classes and weighed. Transparency of the branches was calculated according to the Lambert-Beer equation for a light beam that travels 1m through the branch. The highest branch transparencies were found on *Carpinus betulus* trees, while most branches of beech (*Fagus sylvatica*) were of medium transparency. The four lowest measured branch transparencies were found on *Acer pseudoplatanus* trees. This species also showed the highest variability between branches. In order to enable the calculation of crown transparencies based on 3D-point clouds of the crown, we established allometric relationships between branch basal area and appending leaf area for each species.

1. Introduction

Crown transparency, often given as volumetric leaf area density, is a very sensitive parameter in canopy light models and also an important indicator in forest tree vitality assessment (Eichhorn *et al.*, 2006; Innes, 1998). Contrary to its relevance, it is rarely measured in terms of percentage of light transmitted, since it is hardly possible to attribute light transmission measurements in a forest to single crowns or parts of them. Though crown transparency appears to be highly variable especially in complex multi-species canopies, 3D-light and canopy photosynthesis models must still rely on average values for the whole canopy (Falge *et al.* 2000).

The common measurement method in tree vitality assessment is an intersubjective visual assessment by experienced experts (Eichhorn *et al.* 2006). This method is also subject to criticism with regard to comparability (Dobbertin 2006) and would not be sufficient for parameterization of a light model. As an alternative, crown transparency can be calculated based on the Beer-Lambert law and volumetric leaf area density of trees or branches, which requires volume and leaf area and projected area of woody elements of these units. These values are usually derived from extensive biomass harvests, which are laborious and destructive and require access to the upper canopy. 3D-laser-scanning provides means to facilitate measurements of leaf area density and tree structure (Hosoi & Omasa 2006; Van der Zande *et al.* 2006).

Allometric relationships can be used to quantify the amount of leaf area in parts of tree crowns and are well studied for some species (Bartelink 1997; Grace & Fownes 1998; Boardman 1977; King 1991). In this study we try to combine analysis of the allometric relationships with 3D-laserscanning of trees and hence calculated the crown transparency for different tree species in an old-grown forest.

2. Materials and Methods

2.1 Study site and plant description

The study took place in the old-grown forest Hainich located in Thuringia, central Germany. This forest is the largest continuous deciduous forest area of Germany with a size of 16,000ha from which 7,600ha is registered as Hainich National Park. The study site is a transect of approximately 200m length located in the south-east of the National Park between Bad Langensalza and Eisenach. The site has a mean annual temperature of 7.5°C and a mean annual precipitation of 670mm (Meteomedia, station Weberstedt/Hainich; 51°06´N, 10°31´E; 270m asl).

We studied branches of the following tree species: European beech (*Fagus sylvatica)*, European ash (*Fraxinus excelsior)*, small-leaved lime (*Tilia cordata)*, sycamore maple (*Acer pseudoplatanus)* and European hornbeam (*Carpinus betulus)*. We used adult, co-dominat trees which represented the phylogenetic composition trees of the study site (Table 1).

Table 1: General characteristics of the sampled trees. Diameter at breast height (DBH); n = number of tree individuals; n_b = number of branches sampled.

species	n	DBH [cm]	n_b	branch length [cm]	branches diameter [cm]
F. sylvatica	9	14 – 62	21	72 - 555	0.8 - 6.5
C. betulus	1	62	14	127 - 510	1.2 - 4.1
F. excelsior	21	saplings	26	20 - 510	0.6 - 5.5
T. cordata	7	12 – 35	20	90 - 696	0.7 - 5.4
A. pseudoplatanus	1	80	13	85 - 397	0.9 - 4.4

2.2 Branch sampling

The branches were sampled between July and September 2009. We used a mobile canopy lifter (model DL30, Denka-Lift A/S, Denmark) with a maximum height of 30m in order to enable the canopy access. For each species, we sampled 8 to 13 large (diameter: 1.5 to 6.5cm) and 5 to 8 smaller branches (diameter up to 1.5cm, Table 1). Sampling was done on one individual in case of European hornbeam and sycamore maple, from which only few individuals were present at the study site, whereas different tree individuals were chosen for branch harvesting in case of small-leaved lime and European beech. The crowns of the European ash could not be accessed with the canopy lifter. In this case, we used saplings. All smaller branches were cut from the sun crown. The larger branches grew in the shade.

2.3 3D-laserscanning

Prior to harvesting, the larger branches were marked for laserscanning for which we used the 3D-laserscanner Imager 5006 (Zoller+Fröhlich, Germany). The scan data were registered (Z+F laser control version 7.3, Zoller+Fröhlich, Germany) and all the target branches were afterwards manually separated from the rest of the scan with the help of a graphic program (Pointools, UK). The volume of the branches was calculated from the separated 3D-point clouds of the branches (Mathematica Version 3.0, Wolfram Research, USA). For precise volume calculations we used a voxel model with a voxel edge length of 3cm. The smaller branches were not used for scanning because, firstly, the resolution of the scanner is reduced in the upper canopy layer due to the dense canopy structure and secondly, because the target branch would not be recovered in the scan data because of the small size.

2.4 Leaf area and branch area

After cutting, the branches were divided into 1m-sections and leaves were separated from the branch material. From each 1m-section we randomly selected 10 leaves to measure the leaf area. All leaves were dried at 70°C for 48h and then weighted. From the known weight and area of the 10 leaves and the weight of the rest of the leaves we calculated the leaf area of the 1m-section (equation 1) and the entire branch (equation 2).

(1) $$area\ of\ leaves\ [m^2] = \left[\left(\frac{weight_{rest\ leaves}\ [g] * area_{10\ leaves}\ [cm^2]}{weight_{10\ leaves}\ [g]} \right) + area_{10\ leaves}\ [cm^2] \right] * 0.0001$$

(2) $$leaf\ area\ (branch)\ [m^2] = \sum area\ of\ leaves\ [m^2]$$

The branch material per meter was separated into different categories: branches <0.5cm basal diameter, branches ≥0.5 and <1cm basal diameter, branches ≥1 to <2cm basal diameter, branches ≥2 cm to <3cm basal diameter etc. For all branches larger than 0.5cm basal diameter we measured the exact length and the exact basal diameter of each branch. All branch material was weighted. Using the basal diameter and the length, we calculated the projected branch area (PA) according to equations (3) and (4).

(3) $$PA\ (twig)\ [m^2] = (radius\ 1\ \ [cm] + radius\ 2\ \ [cm]) * length\ \ [cm] * 0.0001$$

(4) $$PA\ \ [m^2] = \sum PA\ (twig)\ \ [m^2]$$

The leaf area density (LAD) and the woody area density (WAD) were calculated from the leaf area which originated from the harvested branches and the volume of these branches (data from the scanner) (equation 5 and 5).

(5) $$LAD\ [m^2 * m^{-3}] = \frac{leaf\ area\ [m^2]}{volume\ [m^3]}$$

(6) $$WAD\ [m^2 * m^{-3}] = \frac{PA\ [m^2]}{volume\ [m^3]}$$

Crown transparency was calculated as the ratio of incoming light (I_0) to transmitted light (I) based on the Beer-Lambert law (equation 7). We calculated crown transparency for a light beam that travels 1m through the leaf cloud (distance d=1m) and assumed the extinction coefficient (c) of the leaves of all species to be the same (c=1).

(7) $$\frac{I_0}{I}\ [\%] = Exp[-(LAD + WAD) * c * d] * 100$$

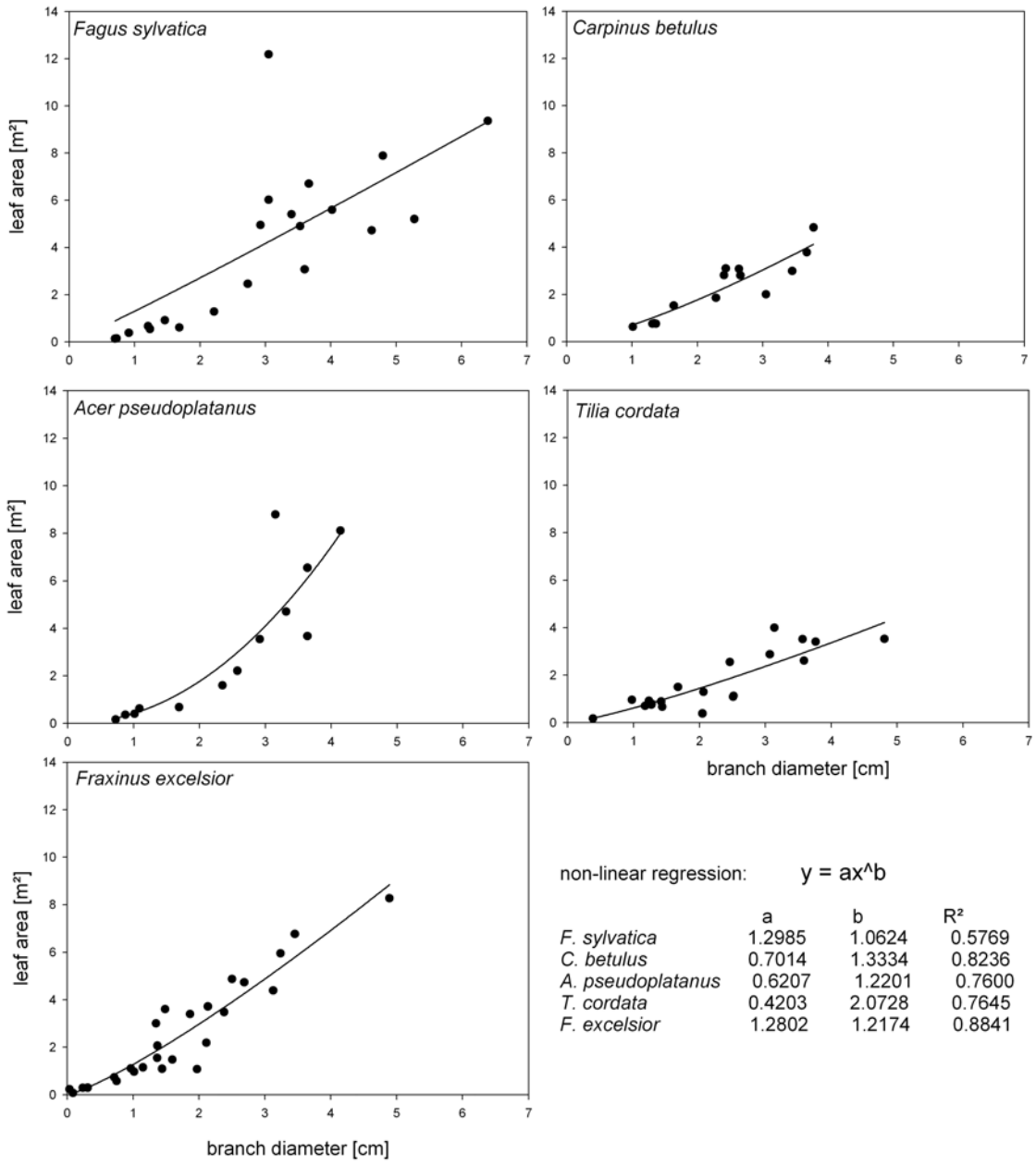
2.5 Statistical analysis

Relationships were determined with non-linear regressions and linear regressions (SigmaPlot 10.0.0.54, Systat Software Inc., Richmond, CA, USA).

3. Results and discussion

3.1 Allometric relationship between branch diameter and associated leaf area

For all species, we derived an allometric relationship (Fig. 1). The non-linear regressions fitted best to our data. The leaf area was similar among species at small diameter sizes (e.g. 2cm, Fig. 1), whereas the leaf area differed between the species at bigger diameter size (e.g. 4cm). Small-leaved lime had the lowest leaf area with approximately 3m³ at a size of 4cm and sycamore maple showed the highest value (approx. 7.5m³) at a size of 4cm. The difference could be based on different growing strategies of the species. The allometric relationships are consistent to earlier work (Grace & Fownes 1998; Westoby & Wright 2003)



non-linear regression: $y = ax^b$

	a	b	R^2
F. sylvatica	1.2985	1.0624	0.5769
C. betulus	0.7014	1.3334	0.8236
A. pseudoplatanus	0.6207	1.2201	0.7600
T. cordata	0.4203	2.0728	0.7645
F. excelsior	1.2802	1.2174	0.8841

Figure 1: Allometric relationship between branch diameter and associated leaf area for the species F. sylvatica, C. betulus, A. pseudoplatanus, T. cordata and F. excelsior. For each species a non-linear regression was fitted to the data.

3.2 Allometric relationship between the measured and the calculated volume of the branches

The branches recovered from the scan were manually separated (Fig. 2b) and a 3D-voxel model was calculated (Fig. 2c). During manually separation some branches couldn't unerringly been kept apart from the neighbor branches. Then these branches were sorted out from further analysis. But with successful separation the reproduction of the branch (Fig. 2b, 2c) was identical with the real branch (Fig. 2a).

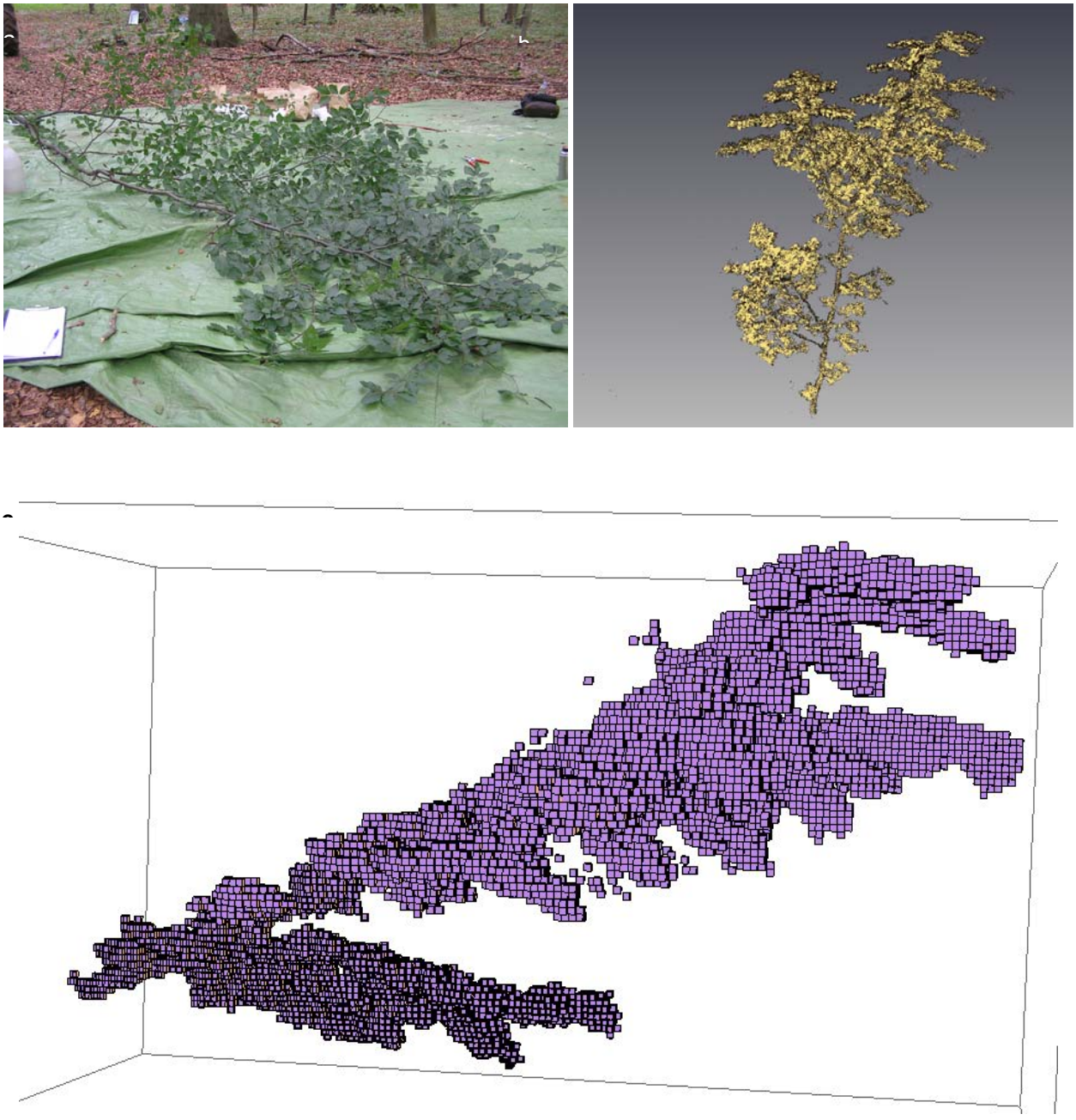


Figure 2: Photo of a selected Europeen beech (F. sylvatica) branch (a), and the same branch presented as 3D-point cloud (b) and as voxel-picture (c).

Due to the non-optimal refinding process of the branches in the scan, the number of the samples decreased for two species tremendously to n=4 (*T. cordata*) and n=3 (*F. excelsior*) so that we excluded this data from further calculations.

The relationship between leaf area from the harvested branches and the volume of their leaf cloud revealed that branch leaf area densities were not constant among and for each species (Fig. 3). While *F. sylvatica* and *A. pseudoplatanus* showed high variability with large leaf clouds having low leaf area as well as small and very dense leaf clouds, the branches of *C. betulus* showed a more or less constant relationship between leaf area and the volume that the leaves occupy.

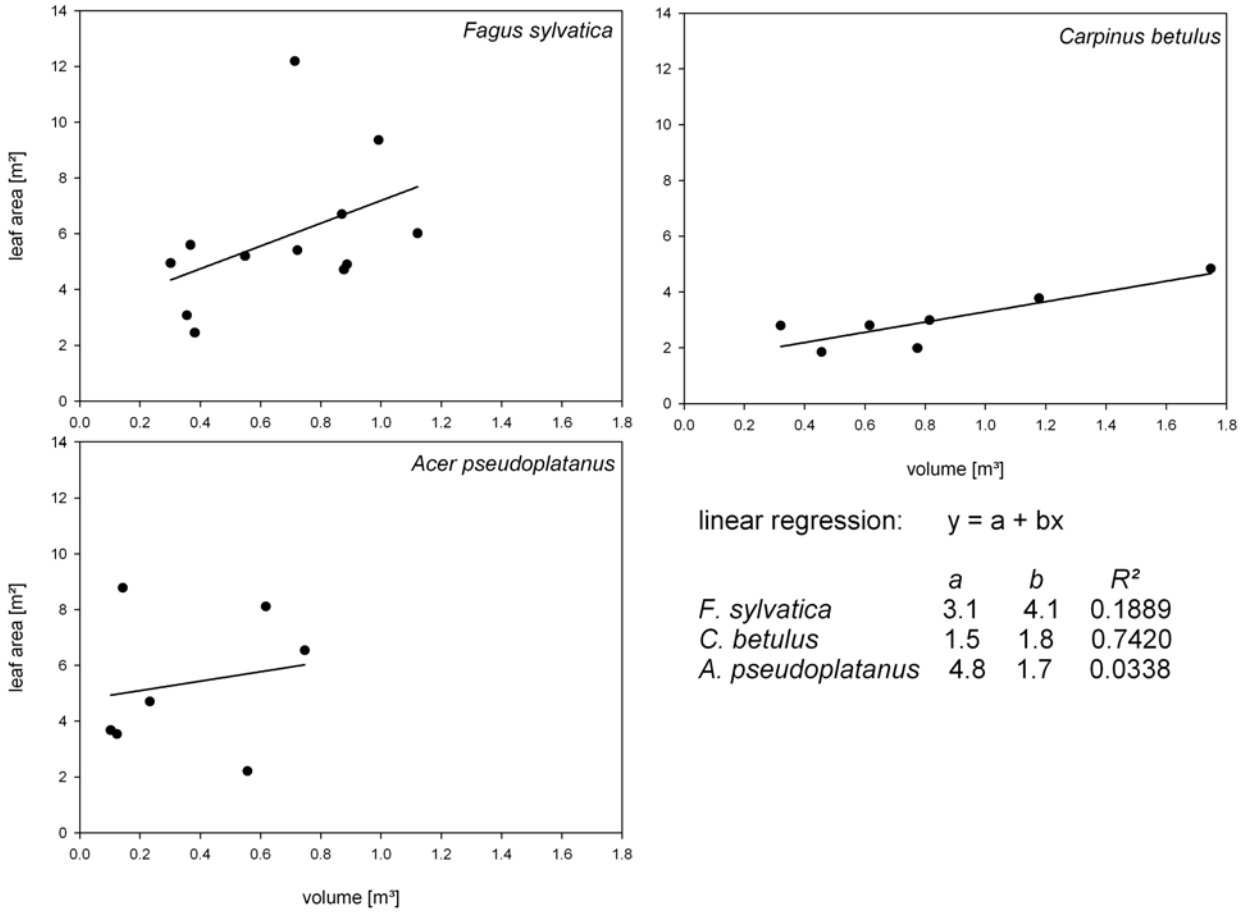


Figure 3: Relationship between leaf area and volume for the species F. sylvatica, C. betulus and A. pseudoplatanus.

3.3 Leaf area density

The same trend was directly visible in the leaf area densities for the three investigated species. Plotted against both the branch length and the branch diameter, no clear trend was visible so the difference could not be attributed to these quantities (Fig. 4). European hornbeam had the lowest leaf area densities, lying in the range of 2-10m² m^{-3}. The leaf area densities of beech was usually higher (4-20m² m^{-3}) and those of sycamore maple exhibited the widest range (4-65m² m^{-3}), including the highest values measured in this study.

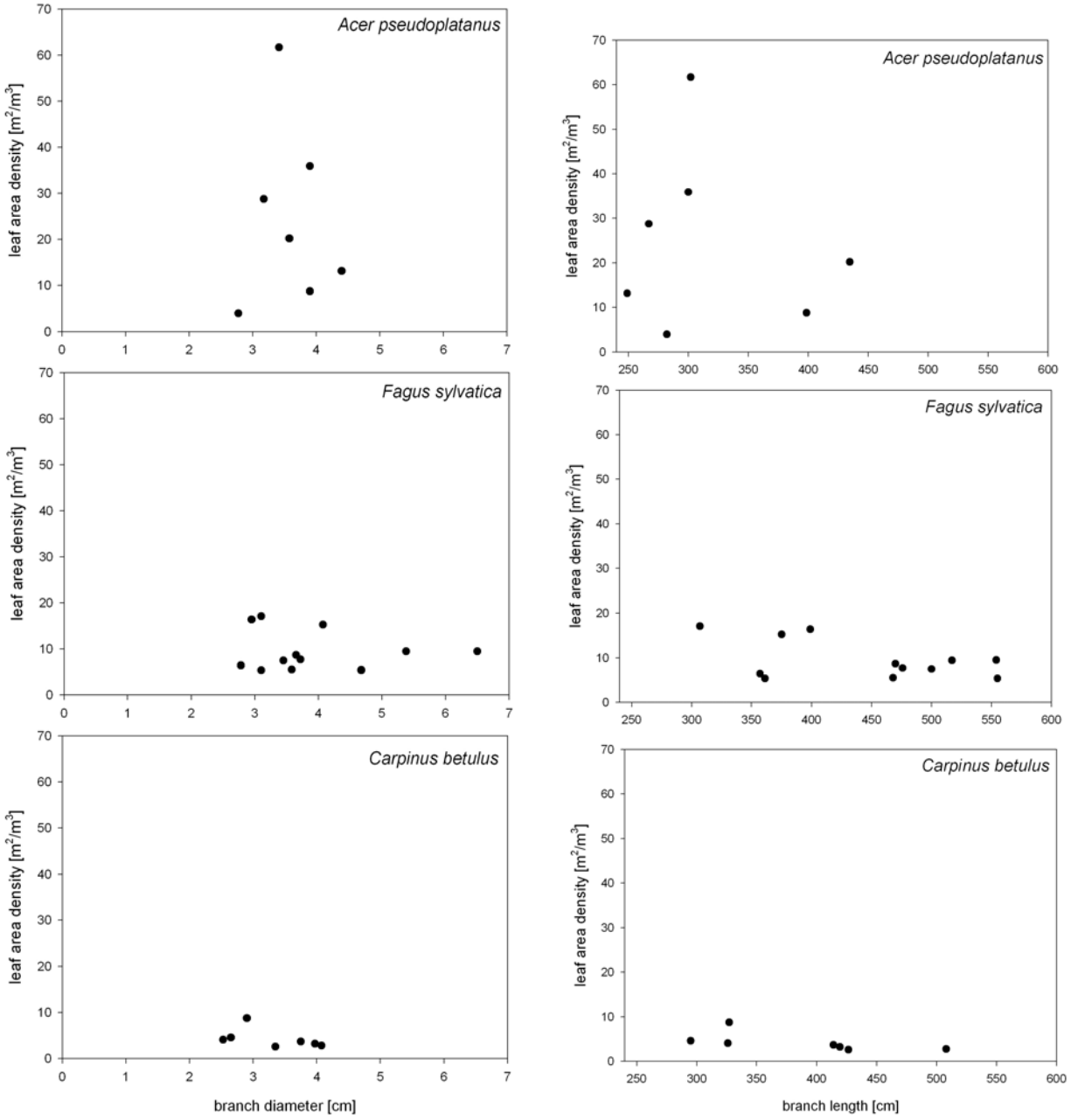


Figure 4: Leaf area density applied to the branch diameter (left side) and the branch length (right side).

For sycamore maple high values of leaf area densities occurred at low volume (Fig. 5). This allows the interpretation that they were not representative for the whole crown but profit from an eventual higher light availability. The branches of sycamore maple originated all from a height between 13 and 23m and from the shade crown, therefore they were not real sun branches. But there is still a possibility that the branch and the leaf area density were influenced by gaps in the canopy which caused more light at that particular spot. The other species showed no clear trend in the leaf area density but they didn't vary in such amplitude like sycamore maple. The leaf area densities were similar to those of beech and oak found by Fleck (2001).

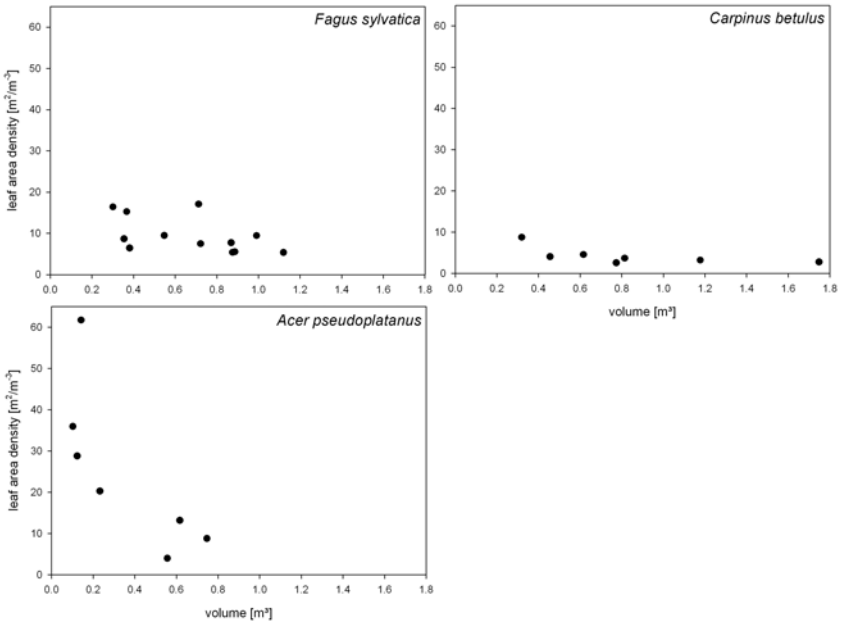


Figure 5: Leaf area densities for F. sylvatica, C. betulus and A. pseudoplatanus.

3.5 Crown transparency

The crown transparency was calculated for a light beam of 1m and plotted as percentage of incoming light (Fig. 6). European hornbeam showed the highest values which allow the interpretation that these branches let more light through than those of European beech and sycamore maple. Sycamore maple revealed low values which were in agreement with the high values of leaf area density. The crown transparency appeared to be highly variable.

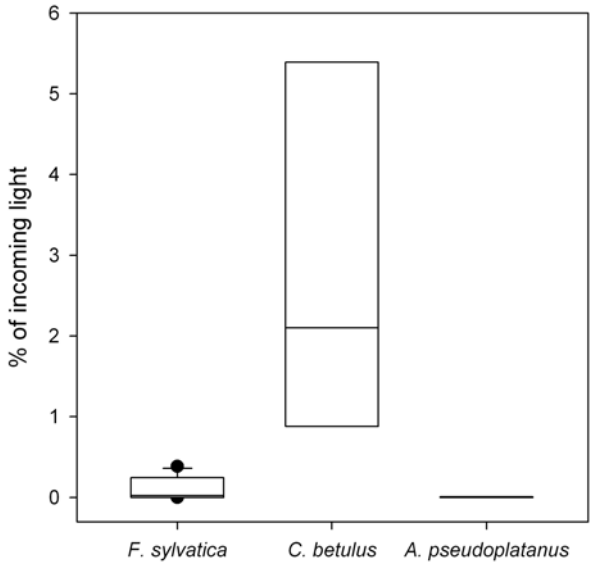


Figure 6: Crown transparency for F. sylvatica, C. betulus and A. pseudoplatanus. Crown transparency displayed as percentage of incoming light.

4. Conclusions

Our results gave first insights to that topic, but to get morereliable results, we suggest conducting a slightly different experimental set-up. First the whole tree should be scanned prior to harvesting and the branches should be scanned an additional time after harvesting on the ground and finally re-scanning the tree after

harvesting. Scanning the branches on the ground increases the resolution of the scanner and, more important, makes it easier to differentiate between the branches and reduces therefore the loss of branches (due to not re-finding them in the scan). Scanning the tree before and after harvesting makes an extrapolation of the branch associated data to the canopy more likely.

Acknowledgements

We thank the research assistants Marlene Iwanowski, Yulan Jin, Dennis Werhahn, David Burger and Sandra González Monge and Christina Langenbruch, Andreas Jacob and Paul Köcher of the Graduiertenkolleg 1086 who helped with the sampling of the branches and the field work and Jasmin Seven (Graduiertenkolleg 1086) for her help in the lab. Special thanks to Alexandra Zach and Christina Langenbruch who helped us with this manuscript. We also thank the DFG for funding the project.

References

Bartelink,H.H. (1997) Allometric relationships for biomass and leaf area of beech (Fagus sylvatica L). Annales des Sciences Forestieres 54, 39-50.

Boardman,N.K. (1977) Comparative Photosynthesis of Sun and Shade Plants. Annual Review of Plant Physiology and Plant Molecular Biology 28, 355-377.

Dobbertin,M. (2006) Tree growth as indicator of tree vitality and of tree reaction to environmental stress: a review. European Journal of Forest Research 125, 89.

Eichhorn,J., Szepesi,A., Ferretti,M., Durrant,D., Roskam,P. (2006) Manual on Methods and Criteria for Harmonized Sampling, Assessment, Monitoring and Analysis of the Effects of Air Pollution on Forests - Part II: Visual Assessment of Crown Condition. United Nations Economic Commission for Europe Convention on Long-Range Transboundary Air Pollution 1-68.

Falge,E., Tenhunen,J.D., Ryel,R., Alsheimer,M., Kostner,B. (2000) Modelling age- and density-related gas exchange of Picea abies canopies in the Fichtelgebirge, Germany. Annals of Forest Science 57, 229-243.

Fleck,S. (2001) Integrated analysis of relationships between 3D-structure, leaf photosynthesis, and branch transpiration of mature Fagus sylvatica and Quercus petraea trees in a mixed forest stand. 2001. PhD Thesis, University of Bayreuth.

Grace,K.T., Fownes,J.H. (1998) Leaf area allometry and evaluation of non-destructive estimates of total leaf area and loss by browsing in a silvopastoral system. Agroforestry Systems 40, 139-147.

Hosoi,F., Omasa,K. (2006) Voxel-based 3-D modeling of individual trees for estimating leaf area density using high-resolution portable scanning lidar. Transactions on Geoscience and Remote Sensing 44, 3610-3618.

Innes,J.L. (1998) An assessment of the use of crown structure for the determination of the health of beech (Fagus sylvatica). Forestry 71, 113-130.

King,D.A. (1991) Tree Allometry, Leaf Size and Adult Tree Size in Old-Growth Forests of Western Oregon. Tree Physiology 9, 369-381.

Van der Zande,D., Hoet,W., Jonckheere,L., van Aardt,J., Coppin,P. (2006) Influence of measurement set-up of ground-based LiDAR for derivation of tree structure. Agricultural and Forest Meteorology 141, 147-160.

Westoby,M., Wright,I.J. (2003) The leaf size-twig size spectrum and its relationship to other important spectra of variation among species. Oecologia 135, 621-628.

Monitoring Forest Defoliation with Terrestrial Laser Scanning

A. Krooks a, J. Hyyppä a, P.Lyytikäinen-Saarenmaa b, A. Jaakkola a, S. Kaasalainen a, M. Holopainen b*

anssi.krooks@fgi.fi

a Finnish Geodetic Institute, Department of Remote Sensing and Photogrammetry, P.O. Box 15, 02431 Masala, Finland

b Department of Forest Sciences P.O. Box 27 FI-00014 University of Helsinki, Finland

Abstract

The current knowledge on forest biomass and particularly the changes related to it are almost entirely based on subjective ground measurements and therefore remote sensing methods are called for. Airborne and terrestrial laser scanning is a promising method for efficient biomass detection because of their capability of direct measurement of vegetation structure and stand attributes. We have investigated a forest defoliation hazard caused by the European pine sawfly (*Neodiprion sertifer*) in a Scots pine (*Pinus sylvestris*) dominated forest in Eastern Finland. Results from terrestrial laser scanner based change detection are compared with visual estimation of the defoliation intensity as percentage of needle loss in the living crown. Clear trends of defoliation were observed in both visual and TLS-based analysis of the tree canopies. The accuracy was affected by, e.g., the time gap between the visual estimation and TLS experiments, and the upward scanning direction preventing some parts of the canopy from being measured.

Keywords: *terrestrial laser scanning, mobile mapping, forest monitoring, defoliation*

1. Introduction

The international interest in biomass detection is strongly related to forest health and carbon cycle monitoring, as well as biodiversity and forest management sciences (Sexton et al., 2009, Hawbaker et al., 2009). The need for improved tools for, e.g., carbon monitoring applications, and the shortage of data for accurate biosphere and climate models has been internationally recognized. As the current knowledge on biomass, and particularly the changes related to it are almost entirely based on subjective ground measurements, remote sensing methods are called for. Airborne laser scanning (ALS) is a promising method in biomass detection (e.g. Lim and Treitz, 2004), because of its capability of direct measurement of vegetation structure (Zimble et al., 2003) and stand attributes (Hyyppä, 2004).

Measuring forest canopy height is crucial for biomass detection (Sexton et al., 2009). In comparisons of LiDAR, radar, and passive optical data, laser scanner has been found to be the best single sensor for canopy height and biomass detection in many studies (e.g., Hyde et al., 2006, Sexton et al., 2009, Kenyi et al., 2009). As canopy height and biomass are functionally related with carbon storage (Balzter et al., 2007, Bortolot et al., 2009), LiDAR studies of forest biomass have been active in the recent years (Bortolot et al., 2009, Hawbaker et al., 2009, García et al., 2010), and LiDAR data have also been used in the modelling of photosynthesis to investigate the carbon exchange (Thomas et al., 2009).

Terrestrial laser scanning (TLS) has been increasingly applied in environmental studies (Barber & Mills, 2007, Côté et al., 2009, Luzi et al., 2009). Terrestrial laser scanner has been found to be an effective and low-cost monitoring method, and the information on TLS performance and range data accuracy is constantly increasing. Static measurements are however limited in range and area coverage. Laser scanner based mobile mapping systems (MMS) can be used to collect detailed data in a cost-effective manner from large spatial areas (Kukko et al. 2007). Vehicle mounted mobile mapping systems have also been deployed to gather information from changing biomass and defoliation (Rossel Polo et al., 2009).

This study was a part of ongoing monitoring campaign for forest defoliation caused by European pine sawflies (*Neodiprion sertifer*). The traditional monitoring methods have been based on field sampling (e.g., manual collection of different life stages) and subjective visual observation of tree condition. The test site was located near the city of Outokumpu, Eastern Finland (N 62.72, E 29.01), in a Scots pine (*Pinus sylvestris*) dominated forest. Reference data were collected from 20 field plots (consisting of 526 trees in total) in June 6-9, 2009. The visual assessment of defoliation was carried out simultaneously with tree-wise measurements in the field plots, and an additional visual assessment was done after defoliation by sawfly larvae (July 26-28) (see Lyytikäinen-Saarenmaa et al., 2006, for more details).

2. Methods

Data Collection and Processing

We used a Leica HDS6000 which is a 685nm phase-based continuous wave terrestrial laser scanner with a 360°×310° field-of-view. The distance measurement accuracy is 4-5mm and the angular resolution is selectable from full 0.009° down to 0.288°. Resulting point clouds have maximum resolution of 40 000 x 17 000 pixels. In this study 1/4 resolution was used to speed up the measurements. Full dome scan time is 3 minutes and 22 seconds with point spacing of 6.3 x 6.3 mm at the range of 10 metres. The scan files were processed with Z+F LaserControl 7.45 by Zoller+Fröhlich GmbH. Basic noise reduction was performed for all files as phase difference based distance measurement method causes ambiguity problems and points that are outside of the maximum measuring distance of the scanner (79 metres) are incorrectly placed in the first ambiguity interval. These points however have lower intensity than points closer to the scanner due the distance effect so removing them is easy with proper intensity threshold. Phase difference based distance measurement also causes mixed pixel noise when laser beam hits multiple objects at the same time and the resulting scan point lies in between. Even with the small laser spot diameter (3 mm at 1 m) and low beam divergence (0.22 mrad) this is likely to happen in tree crowns. Removing mixed pixel noise is simple task but on the other hand it also provides information from the canopy. Scans from different measurement dates were registered in to same local coordinate system for each test plot using tree trunk details as common reference points.

TLS Laboratory case studies

The capability of TLS of deriving changes on the standing tree biomass and defoliation degree was verified by destructive, consecutive defoliation operations by two different laboratory case studies. In the first case study, one Scots Pine tree was defoliated in the laboratory in 7 steps. The biomass of the tree was measured simultaneously with the TLS, and defoliation was also estimated visually. Three simple tree-wise parameters were derived to represent the tree quality and defoliation: number of hits coming from the tree, ratio of tree hits (i.e., number of hits coming from the tree divided by the total number of hits), and the number of hits coming from the ground. These parameters were compared with the tree biomass measured with a 2-g accuracy for each step. The Pearson correlation coefficients derived were 0.996, 0.977 and 0.929, respectively, which implies that the number of points reflected from the tree represent accurately the standing biomass of the tree. The reason why the visual estimation of defoliation did not correlate that well with the number of laser hits is that trunks and branches affect the biomass but are ignored in the defoliation estimation. Therefore it can be expected that the visual estimation produces errors larger than those in terrestrial laser scanning.

In the second laboratory experiment (Fig. 1), five pine (*Scots Pine*) and spruce (*Norway spruce*) trees were measured with TLS from above and below. Here too the biomass of the tree was measured simultaneously with 2-g accuracy. The coefficient of determination with a linear regression model was 0.92 for total biomass and 0.98 for needle and branch biomass, which implies that the relative number of points reflected from the tree represent accurately the relative standing biomass of the tree.

Figure 1: Defoliation time series from laboratory test derived from TLS scan files

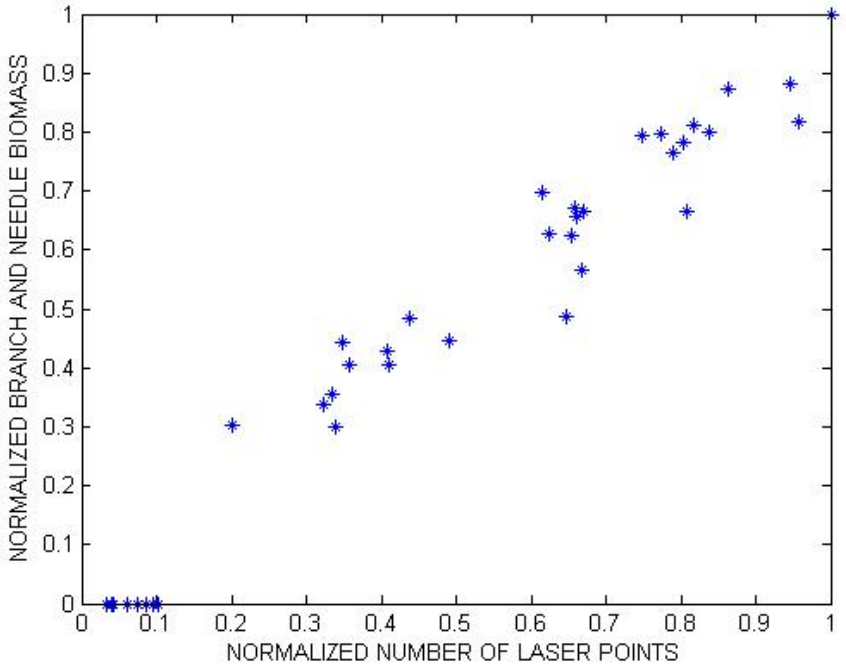


Figure 2: Normalized laser point number versus normalized total biomass of the trees from second laboratory experiment. R2=0.95.

TLS Field Experiment

Two sets of TLS measurements were carried out in the study site in June 25 and July 26-27 during the active period of the pine sawfly hazard. The first measurement was made at the early phases of defoliation and the second one after defoliation period. The scanning was performed from the centre of each field plot with similar scanning parameters and resolution at both dates. The trees that were directly visible (i.e., not obscured by other trees) were extracted from the resulting laser point clouds, and the change in the number of laser returns from each tree was compared to the visually estimated defoliation intensity given as percentage of the lost needle biomass of the living crown.

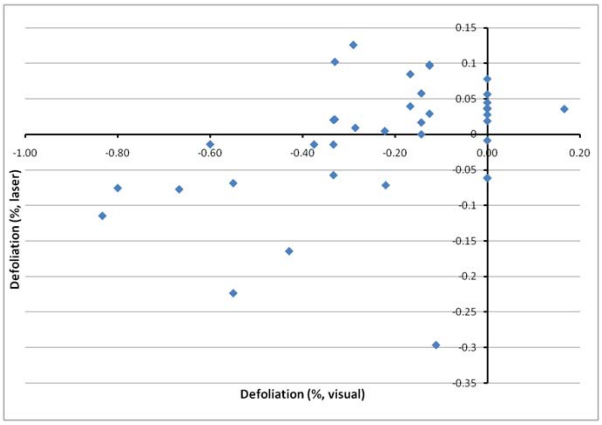


Figure 3: Visual and TLS-based measurement of forest defoliation (in percentages).

The change in laser returns from the tree canopy (in percentages) between the measurements in June 25 and July 26-27 is presented in Figure 3 and compared with visual observation. Clear trends of defoliation are visible in both visual and TLS-based analysis of the tree canopies. There are however several factors that affect the accuracy, e.g.

- Deviation in change detection from TLS point clouds (e.g. mixed pixel effect with phase shift based laser scanners)
- The first visual estimation was carried out in June 6-9, i.e. two weeks before the first TLS measurements.
- The visual estimation of defoliation may be subjective, and it is not based on change detection rather than comparing the present situation with an ideal one (the crown is healthy by soil condition is taken into account).
- The ground-based TLS measurement is made in the upward direction, preventing some parts of the canopy from being measured. Airborne measurements facing downwards would improve the accuracy.

MMS Field Experiment

The mobile mapping system used in this study was a low cost approach combining Time-of-Flight (TOF) based laser scanner ibeo LUX and Novatel SPAN inertial measurement unit (IMU) + a GPS -unit. The ibeo scanner uses pulsed laser beam at 895-910 nm and has field of view 85° x 3.2° with 4 parallel beams. The scanner was mounted on the roof of a car (Fig. 4.) and the sample data was collected by driving along the test plots. GPS-IMU solution was post calculated using virtual reference station data (VRS-GPS).

The MMS data (Fig. 5.) comparison to TLS data and visual interpretation was done with three sample plots. Change in laser returns from the canopy was determined comparing canopy returns to trunk returns as laser point density in target depends on vehicle movement. Early results from the data point out that defoliation trends were similarly visible as with TLS measurements. The accuracy was affected by, e.g., quality of GPS-IMU solution, large laser footprint and shadowing effect from non-test sample trunks makes it difficult to detect small changes in test tree canopy biomass.

Figure 4: The MMS system mounted on car

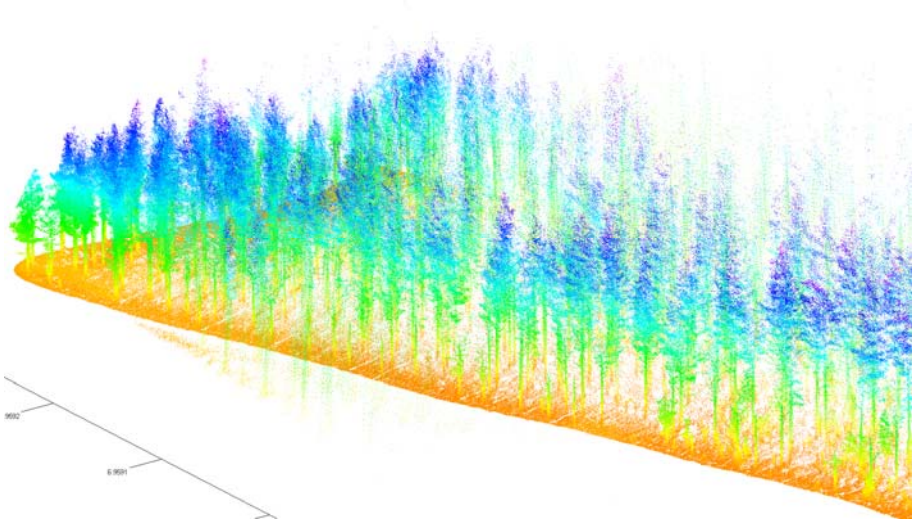


Figure 5: Point cloud derived from MMS data

3. Conclusions

In the laboratory experiments the point cloud agreed with standing biomass with 92-99% coefficient of determination implying that terrestrial laser can be a powerful tool for biomass change reporting, and thus, usable for defoliation measurement. The defoliation trend was also visible in TLS and MMS field experiments, although correlation between point cloud and visual estimation of defoliation was not strong.

Future tests will show whether the effect of the trunk is important. Better field references are also needed as visual estimation may be subjective and is not based on change detection rather than comparing the present situation with an ideal one. The results explain why airborne laser scanning (ALS) is effective for stem volume estimation, since the number of hits recorded by ALS is most probably highly correlated with the biomass, especially needle and branch biomass, which in turn correlated highly with the stem volume. Future tests are needed to verify this, but it can be assumed that a laser scanner measures tree height, crown area and biomass with a high accuracy. In this study we used also mobile laser scanner parallel with TLS and visual defoliation estimation. First results from the MMS data point out potential of using mobile laser scanner as defoliation trends were visible in data and with mobile platform data collection can be done in very efficient manner. In the future studies, more accurate synchronization of different methods is essential, as well as the further development of mobile and UAV-based approach.

References

Balzter, H., Rowland, C.S., Saich, P. Forest canopy height and carbon estimation at Monks Wood National Nature Reserve, UK, using dual-wavelength SAR interferometry. Remote Sensing of Environment, Volume 108, Issue 3, 15, 224-239

Barber, D.M., and Mills, J.P. Vehicle based waveform laser scanning in a coastal environment, Int. Arch. Photogramm. Remote Sens., 36, C55, 2007.

Côté, J.-F.; Widlowski, J.-L.; Fournier, R.A.; Verstraete, M.M. The structural and radiative consistency of three-dimensional tree reconstructions from terrestrial lidar. Remote Sens. Environ. 2009, 113, 1067-1081.

Hyde, P., Ralph Dubayah, Wayne Walker, J. Bryan Blair, Michelle Hofton, Carolyn Hunsaker. Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) synergy. Remote Sensing of Environment, Volume 102, Issues 1-2, 30, Pages 63-73

Hyyppä, J. 2004. Method for Determination of Stand Attributes and a Computer Program to Perform the Method, U.S. Patent No. 6,792,684.

Jaakkola, A., Hyyppä, J., Hyyppä, H., Kukko, A., 2008. Retrieval Algorithms for Road Surface Modelling Using Laser-Based Mobile Mapping. *Sensors*, 8(9), pp. 5238-5249.

Kenyi, L.W., Dubayah, R., Hofton, M., Schardt, M. Comparative analysis of SRTM-NED vegetation canopy height to LIDAR-derived vegetation canopy metrics. International Journal of Remote Sensing. Vol. 30, Issue 11, 2009, Pages 2797-2811

Kukko, A., Constantin-Octavian, A., Salminen, V-M., Kaarinen, H., Chen, Y., Rönnholm, P., Hyyppä, H., Hyyppä, J., Chen, R., Haggrén, H., Kosonen, I., Čapek, K. 2007. Road environment mapping system of the Finnish Geodetic Institute – FGI Roamer. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences,* 36(3/W52), 241-247.

Lim, K.S. and P.M. Treitz, 2004. Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. Scandinavian Journal of Forest Research, 19(6): 558-570.

Luzi, G.; Noferini, L.; Mecatti, D.; Macaluso, G.; Pieraccini, M.; Atzeni, C.; Schaffhauser, A.; Fromm, R.; Nagler, T. Using a ground-based SAR interferometer and a terrestrial laser scanner to monitor a snow-covered slope: Results from an experimental data collection in Tyrol (Austria). IEEE Trans. Geosci. Remote Sens. 2009, 47, 382-393

Lyytikäinen-Saarenmaa, P., Varama, M., Anderbrant, O., Kukkola, M., Kokkonen, A.-M., Henderström, E., and Högberg, H.-E., 2006. Monitoring the European pine sawfly in maturing Scots pine stands with pheromone traps. *Agricultural and Forest Entomology*, 8, pp. 7-15.

Rosell Polo, J., Sanz, R, Llorens, J., Arnó, J., Escolà, A., Ribes-Dasi, M., Masip, J., Camp, F., Gràcia, F., Solannes, F., Pallejà, T., Val, L., Planas, S., Gil, E., Palacín, J. 2009. A tractor-mounted scanning LIDAR for the non-destructive measurement of vegetative volume and surface area of tree-row plantations: A comparison with conventional destructive measurements. *Biosystems Engineering,* 102, pp. 128-134

Sexton, J.O., Bax, T., Siqueira, P., Swenson, J. J., and Hensley, S., 2009. A comparison of lidar, radar, and field measurements of canopy height in pine and hardwood forests of southeastern North America. *Forest Ecology and Management* 257(3), pp. 1136-1147.

Zimble, D.A., Evans, D.L., Carlson, G.C., Parker, R.C., Grado, S.C. and P.D. Gerard, 2003. Characterizing vertical forest structure using small-footprint airborne LiDAR. RS of Environ. 87(2-3): 171-182.

An approach to combine two novel sensing-techniques for quality assessment of single trees and logs of branch diameters obtained by terrestrial laser scan data

URSULA KRETSCHMER†, AXEL WINKING‡, FRANKA BRÜCHERT‡, JÖRG STAUDENMAIER‡, UDO HANS SAUTER‡, and HEINRICH SPIECKER†

ursula.kretschmer@iww.uni-freiburg.de

†Institute for Forest Growth, Albert-Ludwigs-University Freiburg,
Tennenbacher Str. 4, 79106 Freiburg, Germany

‡Department of Forest Utilisation, Forest Research Institute Baden-Württemberg,

Wonnhaldestr. 4, 79100 Freiburg, Germany

Abstract

In the Flexwood project a system should be established to provide information on trees and logs that will match customer demands in subsequent processing stages. Besides air and space born measurement techniques, terrestrial laser and detailed CT scanning will be tested as a complete chain from single tree to sawlogs, to verify measurement results. In this paper the combination of exterior information from tree and stem characteristics will be compared and merged with terrestrial laser scan data and interior information by use of a CT for timber quality assessment. Single trees will be manually measured and then stem wise scanned by a terrestrial laser scanner for complete 3-D representation. After harvesting the logs are described by an innovative CT scanning device. By applying geometrical surface models, the external features of a tree like dimension, curvature, ovality and humps can be detected in terrestrial laser scan data. These external patterns are then compared to the internal characteristics of the tree. Using the CT reconstruction external features such as humps should become explicably by the underlying internal characteristics. Combining the exterior and interior pattern, knowledge about the interior quality of the tree can be derived. This approach aims in the long-term to develop a decision system, which can provide reliable information also related to internal wood properties already for the standing tree with respect to allocation and further production lines.

1. Introduction

In the past decades the European forest based sector, has been increasingly challenged by global operating competitors. Those, provided with low labour costs, high availability of natural resources and/or possibilities to cultivate fast growing plantations, are entering the European market with low price timber products. To overcome this problem and to enhance competitiveness of Europe's forest industry, many investigations have been undertaken to detect weak-points and value adding potential along the forest-wood-supply chain. As an outcome, the lack in information flow between the forestry sector and wood processing industry on availability and accessibility of wood raw material in forest stand in contrast to information on quality and quantity demands of processing industries was identified as one of the main drivers that determine waste of wood raw material and value adding potential.

To tackle that problem, the newly started European Project Flexwood (Flexible Wood Supply Chain) was initialized. Beside the development of improved harvesting and sorting decisions, as well as upgraded bucking, logistic and decision support systems, one of the main targets of this project is to provide high quality forest stand inventory information collected with novel sensing technologies, which is easily accessible to wood consuming industries. Comparable data on the forest resource in Germany is currently only available through cost intensive and time consuming operational and national forest inventories, whose results are mainly dedicated for enterprise internal use only, or are too aggregated for a commercial operative application. Data collection here is usually taking place at observation grid determined fixed

sample plots, with a small radius of 12 m. Further results are extrapolated to the area and therefore always containing uncertainties in accuracy issues.

Novel sensing technologies like air borne (ALS) and terrestrial laser scanning (TLS) allow to overcome identified disadvantages by inventory methodology as described. Advantages can be seen in an easier and more precise area-covering data acquisition (ALS) as well as in the ability to provide quality assessment of standing trees prior to harvesting (TLS). If combined, these sensing technologies have great potential to deliver a consistent and highly valuable information flow on destined harvesting areas.

In this study terrestrial laser scanning will be used to detect external stem quality features in order to assess the future quality of round wood. The TLS's ability to detect external features will be complemented with non-invasive computer tomography imaging (CT), which allows explaining external features from internal stem properties and the other way round. This gained knowledge of derivable features will improve quality assessment of stem parts, to specify bucking instructions at harvesting and thus raise production yield at consecutive conversion stages.

This paper reports on preliminary results to online TLS and CT data. Working steps are conceptual elaborated but have not been tested on feasibility and effectiveness so far.

Test site

The test site is located in the southwest of Germany, the communal forest of Karlsruhe. The area covers about 10 km² and is stocked with beech (*Fagus sylvatica* L.*)*, pine (*Pinus sylvestris*) and oak (*Quercus petraea* and *Quercus rubra)* as dominant tree species. The vegetation height reaches up to 40 m. The terrain is plain and situated in 101-123 m height above sea level.

Within the project about 15 beech trees will be chosen as sample trees for this case study. They are stocked close-by georeferenced inventory plots, to assure the possibility to match TLS with ALS data, sampled in previous flights. Those trees will be sampled in a diameter at breast height (dbh) range from 35-40 cm, due to CT limitations in the ability to radiograph reliably through hardwood material thicker than that given range. Tree ages are estimated to 80 years.

Prior to TLS scanning, the sample trees will be measured manually in terms of dbh and height. The TLS scanning procedure for each plot will be done from multiple positions to get a thorough image of external sample tree stem structures. After scanning the trees will be felled and cut to three logs each, which are brought to the Forest Research Institute of Baden-Württemberg in Freiburg (FVA) for further CT investigations.

Terrestrial laser scanning

Terrestrial laser scanning has become a research method to obtain information about the shape of tree stems. Several research units as well as already a private company have developed algorithms to process information about tree position, stem diameter, taper, and ovality as well as predictions of the tree height and the crown contour. The corresponding software works full or at least semi-automatically. This allows gathering of data of standing trees in the forest by use of a terrestrial laser scanner and the processing of the data off-site, i.e. later in the office.

Projective of the Flexwood project for the use of terrestrial laser scanning data is the detection of bark characteristics. To do so, the different scans of the single trees are registered and the exterior stem information is then mapped onto a plane. There are two different kinds of data the laser scanner gathers. When the beam returns from an object, the intensity of the beam is measured. The range information contains the three dimensional information, i.e. the distance of the object point to the scanner. Both signals can be visualized. An example is shown in

Figure 1. It is selected from a feasibility study in which different tree species were analysed to gather species characteristics which may be detected by a terrestrial laser scanner. The intensity image looks similar to a greyscale photo. The range image reflects the distance of each point to the scanner. Bright areas are closer to the scanner than dark areas. These image data facilitate the processing of the laser scan data. Image processing algorithms can be applied to the image data and bark heterogeneity can be detected automatically.

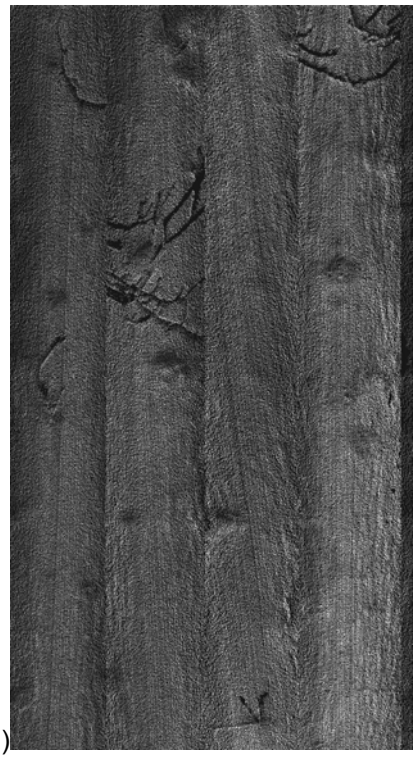
a)  b) 

Figure 1: Intensity (a) and range (b) image of an oak tree (Quercus spec.) by terrestrial laser scanner

To extend this approach to combine the position and size of these characteristics with the data of the CT, a coordinate transformation between the terrestrial laser scanning data and the CT information needs to be established.

CT analysis

There is a large interest of the wood consuming industry in the detection and localization of internal wood defects, prior to conversion processes. The sooner the information is gained and used along the forest wood supply chain, the more flexible can wood raw material be allocated and most effectively used. Aim of this study is to provide quality assessment of logs at stand level, by combining external stem properties from TLS with internal wood properties detected with CT scanning.

X-ray technology is world wide the most applied non-invasive investigation technology and has therefore logically found entrance to wood processing industries also. At present, one or two-directional scanners are in use at sawmills to scan logs upstream to conversion processes. By generating perpendicular projections only, the obtained information gain from this method is limited. Exemplary applications are the detection of heartwood borders, localization of metallic contaminants and average density measurements before cutting. More advanced, but currently not in commercial use for forestry applications, computed tomography scanning technology is able to provide a complete three dimensional image representation. Internal wood properties are hence not only detectable, but also locatable in their geometrical behaviour and dimensions. Future sawing patterns, assisted with this knowledge on internal wood properties will improve sawing

efficiencies and therefore complete productivity to percentage rates which are not achievable with state of the art technologies. Concurrently the same knowledge will serve to assess the value of silvicultural concepts and tending operations, and hence improve pre-harvest actions at forest stand.

Figure 2: 3-D stem surface reconstruction of an oak tree (Quercus spec.)

The physical principle of CT-tomography is based on the varying attenuation of radiation by objects of different dimension and material properties. An x-ray source rotates around an object and materials attenuation profiles are recorded, measuring the difference between emitted radiation intensity and measured intensity at the detector rows. To compute attenuation for each picture element (pixel) per cross section, CT is using up to 1200 projections per 360° scanner rotation from a large number of angular positions. Internal structures of a log like branches, splits, checks, year ring boundaries, heartwood and sapwood can be identified by their varying attenuation due to density differences. For a 80 cm reconstruction field and N=768 x 768 pixels displayed, a resolution of 1.1 mm per pixel allows to distinguish, identify and measure internal log characteristics and relate them to external features in the x/y plane. The attendant depths of volume elements (voxel), which is equivalent to the resolution in the z-plane, can be varied from 1-5 mm by parameter settings like table feed (mm s^{-1}), rotor speed (N 360° rotations per minute) and amount of simultaneously scanned slices. Each pixel is supplied with its geometrical location (X, Y, Z - coordinate), as well as its computed attenuation coefficient. Summarized to a general view, a complete three dimensional image reconstruction is the result, with the feasibility to determine surface coordinates as well as internal quality properties.

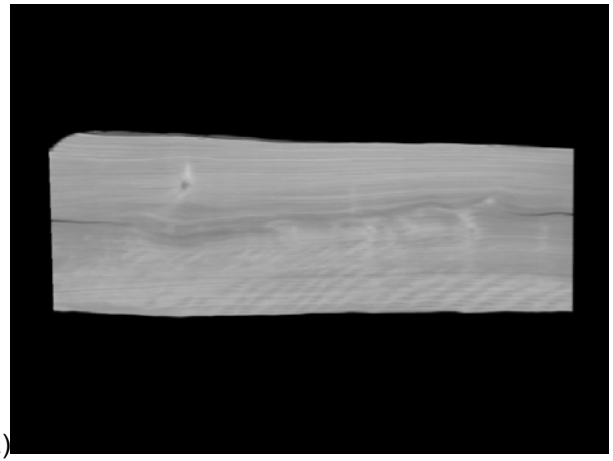
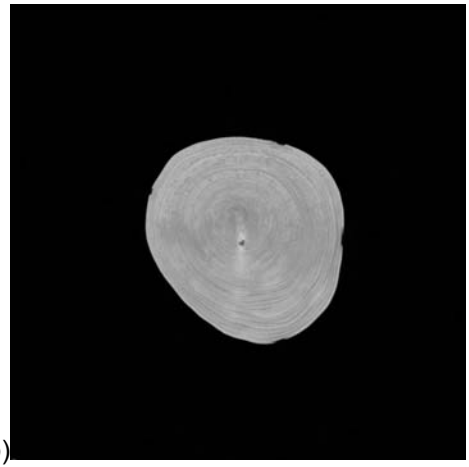
a)  b) 

Figure 3: lateral (a) and transversal (b) CT slice image of a beech log (Fagus sylvatica L.)

By means of gained TLS and CT surface structures these two datasets will be matched, to validate TLS scanning precision as well as the capability to reliably assess internal log quality from external parameters.

Decision system

The Decision support system to be developed in the Flexwood project, will in its final form qualitatively classify parts of standing trees as sawlogs by means of their external stem properties. Terrestrial laser scanning will serve as the data source on which the log classification will be done by automated algorithms in practical applications.

Aim of this case study is to check the feasibility to collect log quality predicting stem surface characteristics by TLS in combination with the ability to verify predictions on internal wood qualities by CT scanning technology. CT scanning itself is then tested on its ability to serve as a non-invasive data source for statistical analysis.

Before an external property will be used as a modelling variable for log quality assessment, it has to fulfil a cascade of demands. In the beginning, it must be detectable in TLS data at stem or bark surface and be in interaction with internal wood properties. Further on, it must be able to visualize the affected or vice versa affecting internal wood property by CT data. At last, the metrical correlation between external expressivity to internal wood composition must be statistically significant with a high correlation coefficient to be used as an explanatory variable. Due to a small sampling size of 15 beech trees, the statistical affirmation of these relations won't be able to deliver within this study.

Providing the scientific base within this project, it must be the target of continuing investigations to supply the decision system with a meaningful statistically proven background on specific log quality determining correlative external to internal relations described in Flexwood.

Acknowledgements

The Flexwood project has been funded within the 7th framework programme of the European Commission (Grant agreement no.: 245136).

Estimating vertical canopy cover with terrestrial and Airborne laser scanning

LAURI KORHONEN†, HARRI KAARTINEN‡, ANTERO KUKKO‡, SVEIN SOLBERG§ and RASMUS ASTRUP§

lauri.korhonen@uef.fi

† University of Eastern Finland, School of Forest Sciences, P.O Box 111, FI-80101 Joensuu, Finland

‡Finnish Geodetic Institute, Geodeetinrinne 2, P.O. Box 15, FI-02431 Masala, Finland

§Norwegian Forest and Landscape Institute, P.O. Box 115, N-1431 Ås, Norway

Abstract

Forest canopy cover (CC) is an important ecological variable and also a basis for the international definition of forest. CC is defined as the proportion of forest floor covered by the vertical projection of the tree crowns, i.e. unbiased CC measurements should be made using vertical observations. Only gaps between the crowns should be considered. If instruments having a non-zero angle of view are used to map the canopy, sides of the crowns will also be observed, which leads to overestimation of CC. Both airborne (ALS) and terrestrial (TLS) scanning lasers measure the canopy with non-vertical laser beams, i.e. CC estimates are likely to be biased.

We measured CC at 16 plots located in Eastern Finland and Southern Norway with a sighting tube to obtain an unbiased field CC, and compared these results to ALS and TLS-based estimates. In case of ALS, the simple proportion of single and first canopy echoes estimated CC very well with a small overestimation (absolute RMSE 3.7–7.6%, absolute bias -3.4–-4.4%) due to relatively narrow nadir angle. The TLS scanners used phase comparison method and had a hemispherical field of view, so instead of trying to calculate the proportion of canopy echoes, we used the points above the height threshold (1.3 m) to create a raster map of the canopy. In the initial image, the brightness of the 4-cm pixel was related to the number of echoes at the pixel. The image was first median filtered, then processed with morphological operations to reduce noise and remove within-crown gaps, and finally binarized to separate covered and open pixels. CC was then estimated as the proportion of canopy pixels. Although TLS created a detailed canopy map close to scanning points, just a small number of echoes were received from more distant crowns, which led to underestimation of CC by 42%–0.1% (absolute RMSE 8.0–17.9%, absolute bias 6.8–13.1%). We conclude that ALS can be safely used in CC estimation despite a minor bias. TLS allows detailed canopy mapping, but a dense network of scan points is required to cover the entire plot.

1. Introduction

Forest canopy cover (CC), defined as the vertical projection of tree crowns ignoring within-crown gaps (Jennings *et al.* 1999, Gschwantner *et al.* 2009), is a forest characteristic that has recently become important in forest inventories. It is commonly used as an ecological indicator (Jennings *et al.* 1999, Gill *et al.* 2000) and also forms a basis for the international definition of forest (FAO 2000). However, its reliable estimation in the field using traditional methods is usually either too laborious or too inaccurate (Korhonen *et al.* 2006). Airborne (ALS) and terrestrial (TLS) laser scanning offer an effective and precise alternative for measuring the percentage of canopy gaps (Lovell *et al.* 2003, Danson *et al.* 2007, Solberg *et al.* 2009), and thus have a great potential to replace the less accurate field methods that are currently used in forest inventories (Korhonen *et al.* 2006).

The traditional methods for *in situ* canopy cover measurements include ocular estimates, sighting tubes (Rautiainen et al. 2005), spherical densiometers (Lemmon 1956), canopy photography (Korhonen and Heikkinen 2009), and modelling based on stem dimensions (Gill *et al.* 2000). Of these methods, the use of

vertically balanced sighting tubes to estimate the proportion ground that is covered by the canopy is usually considered to give an accurate estimate of CC (Rautiainen *et al.* 2005). In practice, sampling transects are established so that they cover the entire plot. The measurer walks the lines and at each grid point looks upwards through the tube to determine whether the point is under a crown or not. Because such measurements require plenty of time, cameras or other instruments with non-zero angle of view are often used to reduce the number of sampling points. However, measuring a larger area from each point makes the estimated CC biased, because the sides of the crowns are also observed (Jennings *et al.* 1999, Paletto and Tosi 2009). When large view angles are used, the trees seem to fall towards the centre of the observed area, and CC is overestimated. This bias increases along with the angle of view, becoming significant at around 40° (20° from zenith) (Korhonen and Heikkinen 2009).

Airborne laser scanning provides data that is very similar to the dot counts with the sighting tube: all that is needed is to calculate the proportion of first and single echoes that are above a specified height threshold (i.e. canopy echoes) (Holmgren *et al.* 2008). The main difference is that the beams of a scanning laser are not exactly vertical (except at nadir), so it is likely that this method overestimates CC similarly to field measurements with non-zero angle of view. However, in typical ALS surveys the off-nadir angles are at maximum 20°, so this effect should remain relatively small. In addition, however, the penetration ability of an ALS pulse may be limited through small canopy gaps, and vary somewhat with the technical acquisition settings.

The potential of terrestrial lasers in mapping canopy gaps and density has also been demonstrated (Danson *et al.* 2007, Huang and Pretzsch 2010). Many TLS systems have hemispherical field of view, so CC estimated as the proportion of canopy gaps within the entire hemisphere will overestimate CC considerably. However, allocating the canopy echoes to a grid based on their XY coordinates should reduce this effect. Our aim in this study is to compare ALS and TLS-based estimates of vertical CC to accurate field data.

Materials

Our first study site is the Koli National park in Eastern Finland (63˚04´ N, 29˚51´ E). Seven 30 x 30 metre sample plots were scanned with both terrestrial and airborne lasers. All plots were located at Scots pine (*Pinus sylvestris* L.) dominated stands (Table 1). More fertile plots (16 and 17) also had a significant number of other species, mainly birches (*Betula spp.* L.) and higher understory vegetation. The rest of the plots were barren pine stands with very low understory vegetation. CC was measured using a sighting tube to determine if the sample point was covered or open. The density of the dot count grid was 1 x 2.5-metres, resulting in 403 points per plot. The methodology was similar to Korhonen *et al.* (2006), except that vegetation lower than 1.3 metres was ignored. Also typical forest parameters were recorded.

Table 1. Koli plot data.

Plot-ID	SP	CC	G	DgM	HgM	CBHgM
16	Pine 69%	80.9	25.8	31.4	23.7	11.0
17	Pine 58%	78.2	24.3	33.2	22.0	10.5
19	Pine 93%	63.0	22.4	23.9	18.4	11.3
20	Pine 99%	43.2	18.4	24.8	18.2	11.5
25	Pine 96%	66.5	29.3	31.2	24.4	12.1
26	Pine 98%	59.8	24.5	29.2	24.2	12.6
27	Pine 98%	59.3	22.9	23.8	18.5	13.3

Abbreviations: SP, dominant species; CC, canopy cover (%); G, basal area (m^2/ha), DgM, mean diameter at breast height (cm); HgM, mean height (m); CBHgM, mean crown base height (m). The mean values are given for the basal area median tree.

The Koli ALS data were obtained in July 2005, while other measurements were made during May– June 2006. Optech ALTM 3100 scanner (Optech Inc., Vaughan, Ontario, Canada) was flown at one kilometres altitude. Half scan angle was 11°, footprint size 25 cm, and mean pulse density 4.6 m^{-1}.

The Koli TLS data were acquired using a FARO LS 880HE80 scanner in June 2006. It is a continuous wave, 785 nm scanner, which uses phase modulation technique (Petrie and Toth 2009, p. 18) for distance measurement with three different carrier wavelengths and has an unambiguity range of approximately 76 m. Point measurement frequency was 120000 points per sec, vertical field of view 320° and horizontal field of view 360°. Beam size was 3 mm at exit and beam divergence was 0.25 mrad (0.014°). Distance measurement error was 3 mm at 25 m (84% reflectivity). Five to eight scans were made in each plot. Same measurement resolution was used for all scans, producing a point spacing of 6 mm at the distance of 10 m. Individual scans were georeferenced to local coordinate system using spherical reference targets. Coordinates for the reference targets were measured using a Trimble 5602 DR 200+ total station, which was setup using the known coordinates of the rectangle-shaped test plot corners. Because the data was meant for timber volume studies, scan points were selected subjectively so that the stems could be viewed from several directions. Thus many scan points were located near the edges of the plot, which was not ideal for CC estimation.

The second study site is located in Lardal, Southern Norway (59˚23´ N, 9˚58´ E). The stands were fertile mixed forests, usually dominated by Norway spruce (*Picea abies* (L.) Karst). Circular sample plots with 12.5 metres radius were measured during summer 2009. CC measurements were made similarly to Koli, resulting in 195 sample points per plot. The ALS data were gathered with the Optech ALTM05SEN180 and ALTM04SEN161 scanners, 690 m above ground. Half scan angle was 12°, footprint size 13 cm, and mean pulse density 10.0 m^{-1}. Four TLS scans were achieved from six of the Lardal plots. One scan was taken at the plot centre while the other three were obtained at six metres distance (N, SE, and SW). Otherwise the setup was similar to Koli.

Table 2. Lardal plot data.

Plot-ID	SP	CC	G	DgM	HgM	CBHgM
3664_2	Spruce 93%	76.9	40.8	29.0	21.9	10.0
3684_7	Spruce 98%	84.1	45.5	30.9	18.9	5.7
3721_5	Spruce 100%	87.2	48.9	31.4	23.7	9.1
3726_1	Spruce 91%	79	42.3	25.7	19.2	6.7
3731_12	Spruce 88%	82.6	41.5	27.3	22.0	4.8
4373_3	Birch 37%	63.6	14.6	32.5	18.3	2.3
4373_6	Spruce 98%	76.9	27.3	22.1	19.1	8.9
4375_3	Spruce 99%	73.8	33	25.9	25.8	10.1
4375_5	Spruce 73%	88.7	45.2	29.1	25.8	8.1

Abbreviations: SP, dominant species; CC, canopy cover (%); G, basal area (m^2/ha), DgM, mean diameter at breast height (cm); HgM, mean height (m); CBHgM, mean crown base height (m). The mean values are given for the basal area median tree.

2. Methods

The ALS data from both sites were processed similarly: the height of the echoes above the digital terrain model was calculated and the percentage of first and single echoes that were above a 1.3 metres threshold was used as the CC estimate. In addition, the high density ALS data were decimated to a density of 1 pulse

per m^2, which is typical to practical forest inventories. This was done using a grid-based method similar to Vauhkonen *et al.* (2009), and CC was re-estimated using the decimated data.

Processing of the TLS data involved several steps. The scanners used the phase comparison method, and therefore the analysis was different from earlier studies in which time-of-flight scanners have been utilized (Danson et al. 2007, Huang and Pretzsch 2010). First, the scanners' own software was used to filter most of the noise and create georeferenced images depicting the density of points above 1.3 m (Fig. 1). Image resolution was 4 cm and (8-bit) image brightness was scaled according to the point density within the cell. After georeferencing, the scanned point clouds were filtered to reduce the amount of outlier points, which occur in phase-based measuring system when the measuring beam hits more than one target or no target at all. Filtering was done by removing all points that had a greater distance than 20 cm to half of its 3 by 3 neighbouring points in scanner's row-column system. Also all dark points, i.e. points with low returning intensity, were removed. FARO Scene software was used for point cloud georeferencing and filtering.

All ground points were deleted and point clouds were processed to create a map of the canopy. Here the analysis differed slightly between Koli and Lardal plots. For the Koli plots, this was done by applying the 'Clear view mode' directly in Scene software: it adds transparency to the otherwise completely opaque point cloud rendering. This allows for viewing through very dense point clouds and gives a better impression of the spatial structure of the underlying point cloud. Settings for the clear view mode and laser point size were selected so that the orthogonal top view of the 3D-point cloud was visually optimized. Image crops were taken and image corner coordinates and pixel size were determined using test plot corner coordinates and ArcGIS-software. Lardal TLS data were pre-processed by the contractor (Treemetrics Ltd., Cork, Ireland) and the point cloud was delivered in ASCII format. This time the Scene software was not available, so instead the raw number of above-ground echoes was calculated for each cell. Because the echo numbers were very high in some places (especially at the stems) and low elsewhere, the echo counts were log-transformed to reduce the intensity range of the images. In this way the Lardal canopy images became comparable to the Koli canopy images.

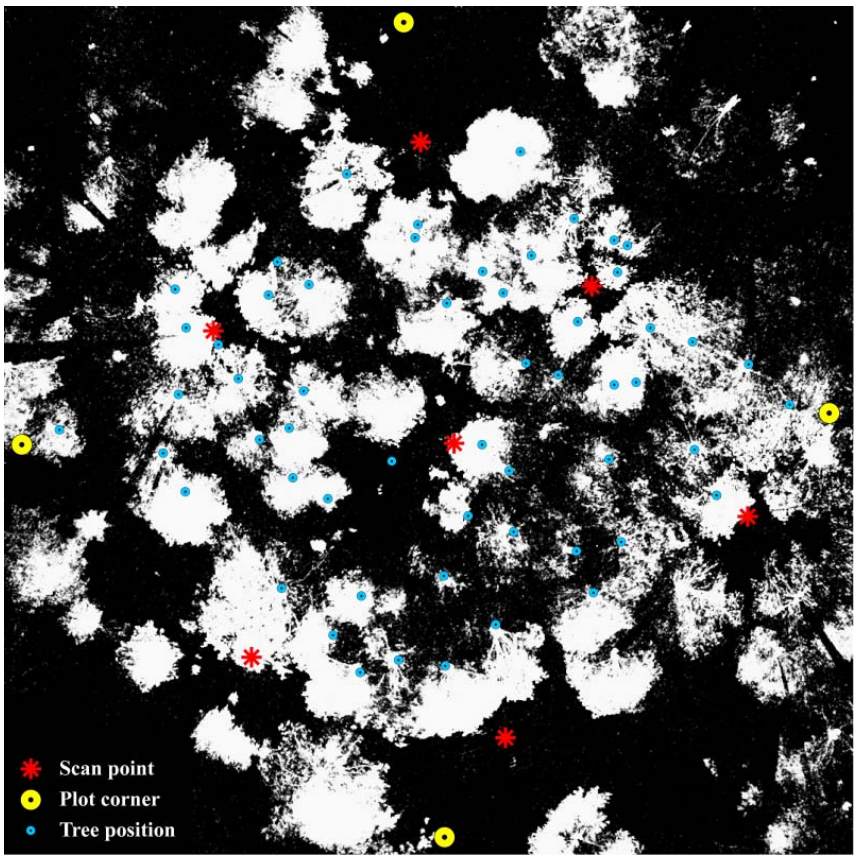


Figure 1. TLS-based canopy map from Koli plot 25 was made with FARO Scene software and imported to ArcGIS for visualization. Tree positions come from field measurements.

These images were analysed further using MATLAB® 7.9.0 numerical computing environment and programming language (MathWorks Inc. 2010) and image processing toolbox extension. The processing chain was as follows:

1. Median filtering to reduce noise, repeated twice.

2. Removal of remaining small peaks and gaps with mathematical greyscale morphology (Soille 2003; see also Wikipedia title on mathematical morphology). Bright peaks were smoothed by morphological reconstruction (Vincent 1993): the image was first eroded, and the eroded image was used as marker and the median filtered image as a mask in Matlab's function *imreconstruct*. Gaps were removed similarly by using the negative of the resultant image.

3. Binarization to separate canopy and empty pixels. Threshold values were selected so that the crowns could be separated as well as possible. In Koli the 8-bit brightness limit was 25 DN. With log-transformed Lardal images the limit was 2.1, corresponding to eight individual echoes.
4. CC was calculated as the percentage of canopy pixels of all pixels that were inside the plot.

The final maps (Fig. 2) were generalized versions of the initial maps (e.g. Fig 1.) where processing had eliminated small within-crown gaps, so that CC estimated from the map was equivalent to the estimates obtained with the sighting tube. In addition, most irrelevant details between the continuous canopy areas were removed in the process.

The accuracy of the results was examined by calculating root mean squared error (RMSE) and bias between the estimates (ŷ) and field reference (y) (Eqs. 1–2).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \qquad \text{(Eq. 1)}$$

$$Bias = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{n} \qquad \text{(Eq. 2)}$$

3. Results

A general view of the quality of our data can be obtained from Fig. 2, where ALS canopy hits and TLS-based canopy map are compared. The ALS point cloud clearly shows where the crowns are, and the TLS map has a good agreement with the ALS results near the scan points. However, the TLS map has large shadowed areas especially outside the plot borders, but also between the scan points. There echoes were often received only from the stem and lowest braches, while most of the crown stayed hidden. In addition, parts of the canopy were shadowed by nearby stems, leaving black strips into otherwise continuous canopy (Fig. 1).

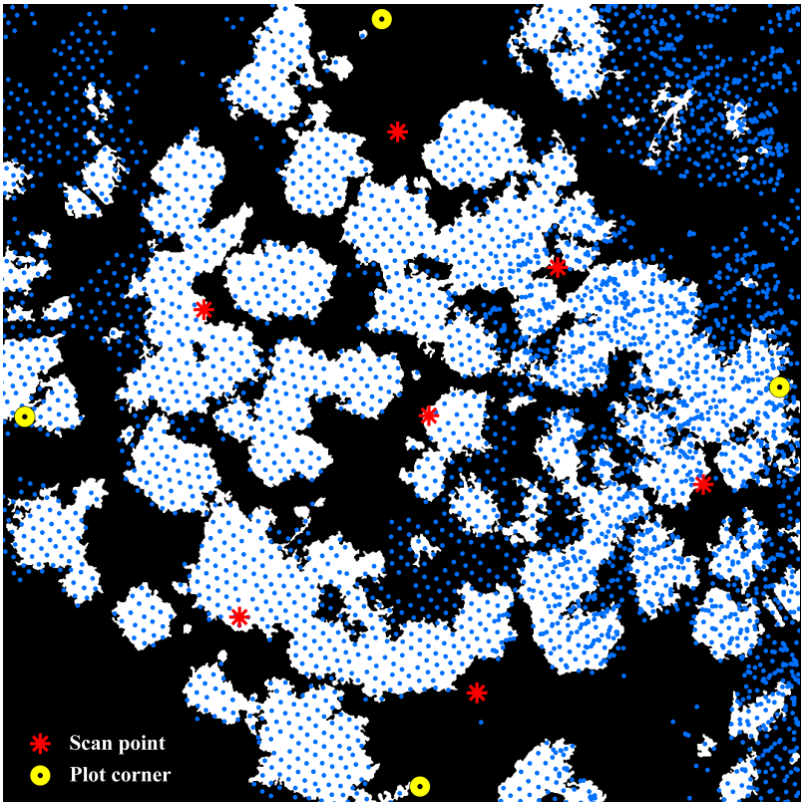


Figure 2. TLS-based binary canopy map from plot 25 overlaid by ALS vegetation hits. Some of the crowns visible in ALS point cloud were poorly detected by the TLS.

Table 3. Koli results.

Plot	CC	TLS scans	TLS-CC	ALS-CC	Decimated ALS-CC
16	80.9	8	75.6	82.6	80.0
17	78.2	8	66.4	82.7	80.9
19	63.0	6	54.2	65.5	62.5
20	43.2	6	32.7	44.0	40.5
25	66.5	7	54.1	72.3	68.9
26	59.8	5	18.0	63.9	61.4
27	59.3	7	58.0	63.5	57.1

Table 4. Lardal results.

Plot	CC	TLS scans	TLS-CC	ALS-CC	Decimated ALS-CC
3664_2	76.9			82.5	80.4
3684_7	84.1	4	77.9	89.0	85.9
3721_5	87.2	4	87.1	91.6	91.5
3726_1	79.0	4	72.7	80.2	83.3
3731_12	82.6	4	76.8	88.5	87.1
4373_3†	63.6			76.7	74.3
4373_6†	76.9	4	62.4	88.8	85.1
4375_3	73.8	4	65.8	75.9	79.0
4375_5†	88.7			79.1	73.3

†Field CC measured by different person.

The comparison of the results with the field-measured CC confirmed what could be seen visually from the images (Tables 3 and 4). At Koli, raw ALS results were very close to sighting tube estimates (RMSE 3.7%, bias -3.4%), and when the data were decimated to typical inventory density, the two methods yielded an even higher agreement (RMSE 2.0%, bias -0.1%). In Lardal the ALS and field inventory results did not agree as much as in Koli (RMSE 7.6%, bias -4.4%). Still, however, the general view was very similar – ALS overestimated field-measured CC by a few percent. One thing to be noted was that if three Lardal plots where field-CC was measured by a less experienced field-worker were removed from the analysis, RMSE decreased to 4.4%. The bias, however, remained the same due to removal of an outlier plot (no. 4375_5), where ALS underestimated field-CC. The point cloud decimation reduced the bias from -4.4 to -3.0%, i.e. it did not reduce the bias as much as with the less dense Koli data.

The TLS results were opposite to ALS: CC was always underestimated. In Koli RMSE and bias were 17.9% and 13.1%, respectively, and in Lardal 8.0% and 6.8%, respectively. This result is in line with the visual observation that more distant crowns remained shadowed in the TLS maps. This problem was especially evident at Koli plot 26, where field CC was 63.9% but TLS estimate only 18.0%. One reason for this unacceptable error was that only five scans were made, four of which were from the corners, which was not enough to cover the entire plot.

4. Discussion

Our results indicate that a simple ALS-based vegetation index can produce accurate canopy cover data, at least if the scan angle is kept small. The results also improved when the data density was reduced to approximately one pulse per square meter. The slight overestimation of CC may be explained by the shadowing effect, i.e. when pulses are arriving at an angle then their probability of having a first or single echo at the ground is lower as compared to vertical pulses. Grid-based decimation of the ALS data reduced the bias. Most likely this resulted from the decimation procedure being most pronounced in the crowns where the density of echoes are high, leaving a relatively higher share of ground-echoes in the remaining data set. ALS data seems to be the best available solution for acquisition of reliable canopy cover data for large areas. As ALS is increasingly used in practical forest inventories due to its ability to produce high-precision estimates on growing stock, the availability of data for CC estimation should also increase.

TLS-based canopy mapping also produced accurate maps of the canopy near the scan locations, but crowns further away remained occluded. Thus TLS is better suited for mapping canopy gaps in the hemispherical perspective projection than vertical map projection. However, in small plots where several scans are made, and in open canopies with good visibility, TLS can produce a very accurate description of the horizontal and vertical structure of the canopy. The methods used to generate the canopy maps functioned very well and can be used safely if such maps are required. However, in cases where a simple estimate of CC is enough and TLS measurements are not otherwise available, traditional field techniques such as sighting tubes or point-and-shoot canopy photography may be more convenient alternatives.

Acknowledgements

Authors wish to thank Espen Martinsen, Jussi Peuhkurinen, and Maria Villikka and for their help during the field work. James Hurley and Garret Mullooly from Treemetrics kindly provided the Lardal TLS data. This study was supported by the Finnish graduate school in forest sciences (GSForest) and Metsämiesten säätiö.

References

DANSON, F.M., HETHERINGTON, D., MORSDORF, F., KOETZ, B. and ALLGÖWER, B., 2007, Forest canopy gap fraction from terrestrial laser scanning. IEEE Geoscience and Remote Sensing Letters, 4, pp. 157–160.

FAO, 2000, On definitions of forest and forest change. Forest Resources Assessment Programme, Working Paper, 33, 15 p.

GILL, S.J., BIGING, G.S. and MURPHY, E.C., 2000, Modeling conifer tree crown radius and estimating canopy cover. Forest Ecology and Management, 126, pp. 405–416.

GSCHWANTNER, T., SCHADAUER, K., VIDAL, C., LANZ, A., TOMPPO, E., DI COSMO, L., ROBERT, N., DUURSMA, D.E. and LAWRENCE, M., 2009, Common tree definitions for national forest inventories in Europe. Silva Fennica, 43, pp. 303–321.

HOLMGREN, J., JOHANSSON, F., OLOFSSON, K., OLSSON, H. and GLIMSKÄR, A., 2008, Estimation of crown coverage using airborne laser scanner. In Proceedings of SilviLaser 2008: 8th International Conference on LiDAR Applications in Forest Assessment and Inventory, R.A. Hill, J. Rosette and J. Suárez (Eds.), pp. 50–57 (Heriot-Watt University: Edinburgh).

HUANG, P. and PRETZSCH, H., 2010, Using terrestrial laser scanner for estimating leaf areas of individual trees in a conifer forest. Trees, doi:10.1007/s00468-010-0431-z.

JENNINGS, S.B., BROWN, N.D. and SHEIL, D., 1999, Assessing forest canopies and understorey illumination: canopy closure, canopy cover and other measures. Forestry, 72, pp. 59–74.

KORHONEN, L. and HEIKKINEN, J., 2009, Automated analysis of in situ canopy images for the estimation of forest canopy cover. Forest Science, 55, pp. 323–334.

KORHONEN, L., KORHONEN, K.T., RAUTIAINEN, M. and STENBERG, P., 2006, Estimation of forest canopy cover: a comparison of field measurement techniques. Silva Fennica, 40, pp. 577–588.

LEMMON, P.E., 1956, A spherical densiometer for estimating forest overstory density. Forest Science, 2, pp. 314–320.

LOVELL, J.J., JUPP, D.L.P., CULVENOR, D.S. and COOPS, N.C., 2003, Using airborne and ground-based ranging lidar to measure canopy structure in Australian forests. Canadian Journal of Remote Sensing, 29, pp. 607–622.

MATHWORKS INC., 2010, MATLAB – The language of technical computing. Available online at: www.mathworks.com/products/matlab/ (accessed 4.5.2010).

PALETTO, A. and TOSI, V., 2009, Forest canopy cover and canopy closure: comparison of assessment techniques. European Journal of Forest Research, 128, pp. 265–272.

PETRIE, G. and TOTH, C.K., 2009, Introduction to laser ranging, profiling, and Scanning. In Topographic laser ranging and scanning, J. Shan and C.K. Toth (Eds.), pp. 1–27 (Boca Raton: Taylor & Francis).

RAUTIAINEN, M., STENBERG, P., and NILSON, T., 2005, Estimating canopy cover in Scots pine stands. Silva Fennica, 39, pp. 137–142.

SOILLE, P., 2003, Morphological image analysis, 2nd edition, pp. 183–218 (Berlin: Springer).

SOLBERG, S., BRUNNER, A., HANSSEN, K.H., LANGE, H., NÆSSET, E., RAUTIAINEN, M. and STENBERG, P., 2009, Mapping LAI in a Norway spruce forest using laser scanning. Remote Sensing of Environment, 113, pp. 2317–2327.

VAUHKONEN, J., TOKOLA, T., MALTAMO, M. and PACKALÉN, 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, pp. S441-S459.

VINCENT, L., 2003, Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. IEEE Transactions on Image Processing, 2, pp. 176-201.

Effect of airborne laser scanning accuracy in forestry yield value calculations

Markus Holopainen1, Antti Mäkinen1, Jussi Rasinmäki2, Kari Hyytiäinen3, Saeed Bayazidi1, Mikko Vastaranta1, Ilona Pietilä1 & Juha Hyyppä4*

markus.holopainen@helsinki.fi

1University of Helsinki, Department of Forest Sciences, Finland, markus.holopainen@helsinki.fi, antti.makinen@helsinki.fi, saeed.bayazidi@helsinki.fi, mikko.vastaranta@helsinki.fi, ilona.pietila@helsinki.fi

2 Simosol Oy, Finland, jussi.rasinmaki@simosol.fi

3MTT, Economic Research, Finland, kari.hyytiainen@mtt.fi

4Finnish Geodetic Institute, juha.hyyppa@fgi.fi

Abstract

Accuracy of Airborne Laser Scanning (ALS) inventory has a decisive impact on the success of determination of forest yield value (net present value, NPV) calculated by forestry decision support systems (DSSs). Errors in inaccurate inventory data increase in magnitude during the execution of long model chains of the simulation systems and causes significant output errors. The longer the reference period, the larger the output errors; thus, inaccurate input data are especially problematic in the case of forestry yield value determination throughout the rotation period. In this paper we give a short summary of the two studies in which we have investigated the effect of the statistical area-based ALS-inventory accuracy in forestry yield value calculations. Based on our results, it is possible to develop an ALS inventory-based methodology for estimating forestry yield value in which it is also possible to derive statistics depicting the reliability of the derived estimate.

1. Introduction

The economic value of forests is crucial information for landowners and various forestry organizations. Estimates of the value of forest property are needed for many purposes, e.g. in the real estate business, land divisions and exchanges and for considering forestry investment. The need for determining the value and the value development of forests has become more important, since forests are increasingly considered as one possible investment outlet amongst other real or financial assets. The International Financial Reporting Standards (IFRS) require that forest enterprises present systematically computed estimates of the value of their forested land annually.

One method for deriving the economic value of a forest stand or property is to calculate the difference between the present values (net present value, NPV) of all future expected revenues and expenses. This approach is referred to as the forestry yield value method (Holopainen & Viitanen 2009) and it is based on the fundamental ideas of forest economics (Faustmann 1849). The NPV of forested land is subject to various uncertainties. The sources of uncertainty include growth and yield models used in the simulators, development of timber prices, the rate of interest used and uncertainties in the input data.

In practice, the determination of a forestry yield value is currently carried out as a part of the forest-planning procedure, during which the scenario best fulfilling the forest owner's objectives is selected by optimization from a set of alternative scenarios produced by simulating tree- or standwise development. Forest-planning simulations are based on a vast number of models depicting the development of single trees or forest stands (In Finland, 400-600 models depending on the forest-planning software system). These models are never complete, notwithstanding the manner in which they were derived. The complexity of the problem is further

heightened by the fact that uncertainty factors tend to accumulate due to the long chains of models and long observation periods used.

If the input data of the forestry yield value calculations are inaccurate, they will significantly affect the functionality of the simulation models, resulting in unrealistic simulation outputs and correspondingly wrong optimization solutions. Reliable inventory data are thus essential for forestry yield value simulations. In assessing the state of a stand, the estimates may differ significantly from the real situation, due to the inventory method used.

Acquisition of forest-planning data is currently in a phase of radical change. In Finland, operative forest planning is evolving into a methodology by which stock characteristics are estimated by means of tree-wise measured sample plots and area-based statistical features of airborne laser scanning (ALS) data and digital aerial photographs. Estimation will be performed, using the nonparametric k-nearest neighbour (k-NN) or k-most similar neighbour (k-MSN) method (Packalén 2009). With respect to the estimation of stand mean characteristics (e.g. Næsset 1997, 2002, 2004, Holmgren 2003) and tree species- or timber assortment-specific characteristics (Packalén & Maltamo 2006, 2008, Peuhkurinen et al. 2008, Holopainen et al. 2010a), it has become possible to achieve at least the same level of accuracy using low-pulse ALS data as that found in traditional standwise field inventory (SWFI).

In this paper, we give a short summary of the two previous studies (Holopainen et al. 2010a, 2010b) in which the effect of the statistical area-based ALS-inventory accuracy in forestry yield value calculations have been investigated.

2. Materials and methods

Holopainen et al (2010a) examined uncertainty factors related to inventory methodologies and forest-planning simulation computings in the estimation of logging outturn assortment volumes and values. The uncertainty factors investigated were (i) forest inventory errors, (ii) errors in generated stem distribution, (iii) effects of generated stem distribution errors on the simulation of thinnings and (iv) errors related to the prediction of stem form and simulation of bucking. Regarding inventory errors, standwise field inventory (SWFI) was compared with statistical area-based airborne laser scanning and aerial photography inventorying (A_ALS).

Holopainen et al. (2010a) utilized timber assortment outturn data gathered by logging machine were applied as the study's ground truth data. These data were acquired from 31 logging sites (5950 trees) logged in the Evo area during winter 2008. The data consisted of timber assortment volumes measured by the logging machines. For clear-cutting sites (12), accurate timber assortment outturn volumes were known for each stand. With respect to thinning sites (19) determination of stand-level figures was based on the measurement of 90 circular plots that were measured before and after the logging operation. Uncertainty caused by the investigated sources of error was analysed, using simulation computations performed by the SIMO forest management planning simulation and optimization system (Rasinmäki et al. 2009). A_ALS feature selection was based on the genetic algorithm method presented by Holopainen et al. (2008) and estimation of forest characteristics on the nonparametric k-NN method.

Holopainen et al. (2010b) compared uncertainty related to inventory data, growth models and fluctuation in raw timber prices in forest yield values (NPV) computed throughout the rotation period. With respect to inventory data, uncertainty analyses were carried out for methodologies currently in operative use only, i.e. the SWFI and A_ALS inventory methodologies. The study was performed, using Monte Carlo simulations (SIMO software) and interest rates of 3%, 4% and 5%. The study material consisted of a simulated forest property of 40 stands having an even representation of various tree species and development classes. The uncertainty figures concerning inventory data accuracy were adopted from two southern Finland A_ALS pilot projects. The growth model uncertainty figures were derived, using the respective error models. Raw timber price fluctuation was predicted using a Geometric mean-reverting (GMR, Insley 2002) price model and the realized price development of the period January 1986 and August 2008.

3. Results

Holopainen et al. (2010a) compared SWFI with A_ALS inventory with respect to assessing timber assortment-level forest data. They focused on the estimation of timber assortment quantities, economic values and related sources of uncertainty. In addition to inventory errors associated with the generation of stem distributions, estimation of single-tree stem form and simulation of bucking during the forest-planning simulation computations were scrutinized.

Table 1. Effects (bias%^{NPV} and rmse%^{NPV}) of different error sources on predicted stock value (€/m^3) at stand level. Field reference measured by logging machine: 5400 trees within 12 clear-cutting stands. A_ALS = area-based ALS inventory, SWFI = standwise field inventory. The active error source is marked with X.

ERROR SOURCES				Predicted stock value	
A_ALS Inventory error	SWFI inventory error	Stem distribution generation	stem form prediction & bucking simulation	$bias\%^{NPV}$	$rmse\%^{NPV}$
X				0.6	24.7
	X			-0.1	29.1
		X		-1.2	2.6
	X	X	X	2.5	33.4
X		X	X	4.2	23.8

The results showed that the most significant source of error in the prediction of clear-cutting assortment outturns was inventory error. The bias and root-mean-squared error (RMSE) of inventory errors varied between -11.4 and 21.6 m^3/ha and 6.8 and 40.5 m^3/ha, respectively, depending on the assortment and inventory methodology. The effect of forest inventory errors on the value of logging outturn in clear-cuttings was 29.1% (SWFI) and 24.7% (A_ALS). The respective RMSE values related to thinnings were 41.1% and 42%. The errors related to stem distribution generation, stem form prediction and bucking simulation were significant but considerably lower in magnitude than the inventory error (see Table 1). Timber assortment-level errors were greater than those related to mean characteristics (e.g. mean stem volume) estimation. This is a consequence of errors stemming from stem distribution generation and stem form prediction, bucking simulation and also uncertainty in tree species recognition. Effects of A_ALS inventory error, SWFI error, stem distribution generation error, stem form prediction and bucking simulation error and their combined effects to the predicted stock value (bias% and rmse%) are summarized in Table 1.

Table 2. Averages of the relative biases ($bias\%^{NPV}$) and sds ($sd\%^{NPV}$) of the simulated NPV distributions with given source of uncertainty with 3% rate of interest. The active uncertainty sources in each combination are marked with X. (Table modified from Holopainen et al. (2010b))

SOURCES OF UNCERTAINTY				Interest rate 3%	
U_{PRICE}	U_{FIELD}	U_{ALS}	U_{GROWTH}	$bias\%^{NPV}$	$sd\%^{NPV}$
X				-6.1	8.2
	X			-6.8	28.8
		X		1.7	26.5
			X	-9.5	33.2
X	X			-9.1	29
X		X		-1	27.4
X			X	-5.7	34.9
	X		X	-12.5	46.9
		X	X	-2.1	46.5
X	X		X	-9.2	47.4
X		X	X	0.1	46.5

Holopainen et al. (2010b) showed that the growth models used in forest-planning simulation computations were the greatest source of uncertainty with respect to NPVs computed throughout the rotation period (see Table 2). Uncertainty almost as great was caused by input data uncertainty, while the uncertainty caused by fluctuation of raw timber prices was considerably lower in magnitude. The interest rate used in the computations did not affect the relative importance of the various sources of uncertainty in any meaningful manner. Each single source of uncertainty caused an uncertainty in NPVs ranging from 8% to 32%. The joint effect was appr. 50% at most. Interestingly, the joint effects turned out to be significantly less than the sum of the respective sources of uncertainty.

4. Discussion & Conclusions

The economic value of the stand or estate can be estimated, using forest management planning simulations. Forest planning is based on stand-level or tree-level models and simulators. When tree-level forest-planning simulators are used, stem distributions can be formed, based on mean stock characteristics, by determining and measuring the inventory unit (e.g. compartment) median trees and utilizing theoretical stem distributions, such as the Weibull distribution, in the modelling and derivation phases, or it can be derived by enumerating all trees present in the inventory unit, resulting in true stem distribution series.

A_ALS interpretation provides new opportunities but also challenges with respect to forest-planning computation procedures, e.g. it offers several alternatives for forming stem distributions. One alternative is to first estimate the mean stand characteristics and then apply stem distribution models based on theoretical distributions (e.g. the Weibull distribution). Another alternative is to use ALS features to directly estimate stem distribution parameters in a manner proposed and studied by Gobacken & Næsset (2004), Maltamo et al. (2006) or by Breidenbach et al. (2008). Gobacken & Næsset (2004) and Maltamo et al. (2006) showed that size distributions of trees can be estimated, using locally modelled distribution functions in which ALS-based canopy height metrics are used as predictors. It is also possible to utilize stem distribution series measured for field plots as reference in k-NN or k-MSN method (Packalén & Maltamo 2008).

Forest organizations involved in the planning of privately owned forests in Finland are currently adopting a new inventory methodology in which A_ALS interpretation is implemented in a grid consisting of 16-m x 16-m cells. Forest characteristics are estimated for each cell, using the nonparametric k-NN or k-MSN algorithms. In this case stem distributions could be formed separately for each grid cell, which would enable better analysis of intra stand variation. However, stand-level mean values will be used for forming stem distributions at least for the time being, probably because of the heavy computations involved. Instead of using single grid cells as computation units, rudimentary compartments formed by automatic segmentation of grid cells could be used as well (e.g. Tuominen & Haapanen 2009).

Single-tree (or individual tree detection, ITD) ALS inventorying may result in more accurate stem distributions than the application of A_ALS inventorying. This also leads to more accurate timber assortment volume estimates and consequently to lower levels of inventory-related uncertainty. However, current ITD inventory algorithms require further development in automatic location of trees and automatic tree species recognition. In several studies (e.g. Peuhkurinen et al. 2007, Maltamo et al. 2009, Vastaranta et al. 2009), allometric models describing the relationships between tree crown size, height and DBH were applied as a part of the ITD methodology. These models are, however, highly sensitive to errors in their input data. The automatic measurement results of tree crown size in particular tend to be error-prone. Nonparametric algorithms applicable to single-tree interpretation are, therefore, currently under development (e.g. Maltamo et al. 2009, Yu et al. 2010a,b) to part with the use of allometric models.

Our results show that forest inventory methodologies, growth models applied in forest-planning simulation computations and fluctuation of timber prices cause significant uncertainty in the computation of stand-level forestry yield value. Uncertainty resulting from inventory methodology and growth model functioning is greater than that caused by timber price fluctuation, given that future timber price fluctuation will be predicted with a GMR price model. With respect to forest inventory methodology, uncertainty results addition to the actual inventory error by generation of stem distribution, prediction of stem form and simulation of bucking.

A_ALS inventorying result similar or lesser degrees of uncertainty in computed stand-level NPV than traditional SWFI. Based on these results, it is possible to develop an ALS inventory-based methodology for estimating forestry yield value in which it is also possible to derive statistics depicting the reliability of the derived estimate. In further studies, developed methodology can be linked into market valuation of the forest properties.

Acknowledgements

This study was made possible by financial aid from the Finnish Academy project Improving forest supply chain by means of advanced laser measurements.

References

Faustmann, M. 1849. Berechnung des wertes welchen Waldboden sowie noch nicht haubare Holzbestände für die Waldwirtschaft besitzen. Allgemeine Forst- und Jadg-Zeitung 15: 441-455.

Gobakken, T & Næsset, E. 2004. Estimation of diameter and basal area distributions in coniferous forest by means of airborne laser scanner data. Scandinavian Journal of Forest Research 19:529-542.

Breidenbach, J., Gläser, C. & Schmidt, M. 2008. Estimation of diameter distributions by means of airborne laser scanner data. Canadian Journal of Forest Research 38:1611-1620.

Holmgren, J. 2003. Estimation of forest variables using airborne laser scanning. PhD Thesis. Acta Universitatis Agriculturae Sueciae, Silvestria 278, Swedish University of Agricultural Sciences, Umeå, Sweden.

Holopainen, M., Haapanen, R., Tuominen, S. & Viitala, R. 2008. Performance of airborne laser scanning- and aerial photograph-based statistical and textural features in forest variable estimation. In Hill, R., Rossette, J. and Suárez, J. 2008. Silvilaser 2008 proceedings:105-112.

Holopainen, M. & Viitanen, K. 2009. Käsitteistä ja epävarmuudesta metsäkiinteistöjen taloudellisen arvon määrittämisessä. Metsätieteen aikakauskirja 2/2009:135-140. (About concepts and uncertainties in economical valuation of forests, In Finnish).

Holopainen, M., Vastaranta, M. Rasinmäki, J., Kalliovirta, J., Mäkinen, A., Haapanen, R. Melkas, T., Yu, X., Hyyppä, J. 2010a. Uncertainty in timber assortment estimates predicted from forest inventory data. European Journal of Forest Research, in press. Doi 10.1007/s10342-010-0401-4.

Holopainen, M., Mäkinen, A., Rasinmäki, J., Hyytiäinen, K., Bayazidi, S. & Pietilä, I. 2010b. Comparison of various sources of uncertainty in stand-level net present value estimates. Forest Policy and Economy, Doi 10.1016/j.forpol.2010.02.009.

Insley, M. 2002. A real option approach to the valuation of a forestry investment. J. Environ. Econ. Manage 44, 471-492.

Maltamo, M., Peuhkurinen, J., Malinen, J., Vauhkonen, J., Packalén, P. & Tokola, T. 2009. Predicting tree attributes and quality characteristics of Scots pine using airborne laser scanning data. Silva Fennica 43(3): 507–521.

Næsset, E. 1997. Estimating timber volume of forest stands using airborne laser scanner data. Remote Sens. Environ. 61:246-253.

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

Næsset, E. 2004. Accuracy of forest inventory using airborne laser-scanning: Evaluating the first Nordic full-scale operational project. Scandinavian Journal of Forest Research, 19, 554-557.

Packalén, P. 2009. Using airborne laser scanning data and digital aerial photographs to estimate growing stock by tree species. Dissertationes Forestales 77. PhD thesis, Faculty of Forest Sciences, University of Joensuu.

Packalén, P. & Maltamo, M. 2008. Estimation of species-specific diameter distributions using airborne laser scanning and aerial photographs. Canadian Journal of Forest Research 38: 1750–1760.

Peuhkurinen, J., Maltamo, M. & Malinen, J. 2008. Estimating species-specific diameter, distributions and saw log recoveries of boreal forests from airborne laser scanning data and aerial photographs: a distribution-based approach. Silva Fennica, 42:625-641.

Rasinmäki, J., Kalliovirta, J., Mäkinen, A. 2009. An adaptable simulation framework for multiscale forest resource data. Computers and Electronics in Agriculture 66, 76-84.

Tuominen, S. & Haapanen, R. 2009. Comparison of a grid-based and segment-based estimation of forest attributes using airborne laser scanning and digital aerial imagery. Will appear in the proceedings of IUFRO Division 4: Extending

forest inventory and monitoring over space and time, May 19-22, 2009, Quebec City, Canada. 5 p. http://blue.for.msu.edu/meeting/proc2/Tuominen_Haapanen.pdf

Vastaranta, M, Holopainen, M., Haapanen R., Yu, X., Melkas, T., Hyyppä, J. & Hyyppä, H. 2009. Comparison between an area-based and individual tree detection method for low-pulse density als-based forest inventory. In Bretar, F, Pierrot-Deseilligny, M. & Vosselman, G. 2009. LaserScanning 2009 proceedings:147-151.

Yu, X, Hyyppä, J., Holopainen, M. & Vastaranta, M. 2010a. Comparison of area based and individual tree based methods for predicting plot level attributes. Remote Sensing, 2010/2:1481-1495; Doi:10.3390/rs2061481.

Yu, X, Hyyppä, J., Holopainen, M. Vastaranta, M. & Viitala, R. 2010b. Predicting individual tree attributes from airborne laser point clouds based on random forest technique. Submitted to ISPRS Journal of Photogrammetry and Remote Sensing.

Is it possible to predict gap age in North American temperate deciduous forest using LiDAR data?

JEAN-FRANÇOIS SENÉCAL†‡, FRÉDÉRIK DOYON‡§ and BENOÎT ST-ONGE†*

†Département de Géographie, Université du Québec à Montréal, Montréal, Québec, H2X 3R9, Canada

‡Institut québécois d'Aménagement de la Forêt feuillue, Ripon, Québec, J0V 1V0, Canada

§Département de travail social et de sciences sociales, Université du Québec en Outaouais, Gatineau, Québec, J8X 3X7, Canada

Abstract

Knowledge on fine spatio-temporal gap phase dynamics of the temperate deciduous forest of Eastern North America is still poorly documented, particularly at its northern range. Traditional methods used to sample gaps, like recording information along transects, are time consuming and expensive. Also, studies in temperate forests have usually been limited to recent gaps (i.e. gaps with low regeneration within). However, gap structure can change rapidly over time in response to resources availability. With this study, we hypothesised that within-gap vegetation can present a distinct heterogeneous 3D structure that should be detectable by LiDAR and could be linked to gap age. To test this hypothesis, we have investigated the forest gap regime of three old-growth temperate deciduous forests in southern Québec, Canada with a gap detection algorithm using LiDAR and field surveyed gap data. Field sampling of a random subset of detected gaps was carried out during summer 2009. Dendrochronological samples were taken within gaps in structurally different sapling groves to date the canopy opening events using a growth release detection algorithm. Gaps were clustered using height structural parameters derived from LiDAR data. Average gap area was 20 to 23 m^2, cumulatively occupying 6.2% to 7.5%. New growth releases occurred each year in 2.7% of the sampled saplings on average between 1960 and 2005. Gap structure clusters were not strongly related to gap age, but presented a distinct synchronous and asynchronous growth releases history. However, they were aggregated by study site. Local external factors thought to have an influence on these results are a pre-established shade-tolerant regeneration, ungulate browsing and differential growth rate due to site conditions.

1. Introduction

The gap phase regime is the main regeneration mechanism of the temperate deciduous forest (Runkle 1985, Platt and Strong 1989). Characterizing the canopy gap regime using traditional methods raises a number of challenges, including cost, time and reproducibility (Runkle 1992). For these reasons, studies on this subject are usually limited to small study areas, with an emphasis on large gaps (Seymour et al. 2002). This is a problem since most temperate deciduous forest gaps are small and to not take it into account may have a significant influence on our perception of the gap regime dynamics (Brokaw 1982, McCarthy 2001). An ideal method for characterizing the gap regime would do so at a spatial scale fine enough to detect small gaps for an area large enough to be representative of the phenomenon. At a finer scale, the analysis of the internal structure of gaps due to differential mortality of trees also presents an interesting challenge. Changes in the internal structure of gaps often come from their expansion and this component is usually not considered in studies on gaps because of costs (Runkle 1982). Light Detection And Ranging (LiDAR) is particularly well suited to these tasks (Koukoulas and Blackburn 2004, Vepakomma *et al.* 2007, Gaulton and Malthus 2010). LiDAR remote sensing instruments can acquire high resolution vertical and horizontal information and have been used in forestry applications for a number of years (Lim *et al.* 2003). Such an instrument could thus be used to extract gap structural information related to gap dynamics.

1.1 Gap age and structure

Gaps are opening in the canopy caused by the death of part of a tree, a complete tree or a small number of trees (Runkle 1985, Barden 1989). In literature, the size of gaps varies somewhat depending on the study. To our knowledge, the minimum size of a gap to have been considered in a study was 5 m^2 (Gaulton and Malthus 2010). There are also many definitions of maximum gap sizes, such as a gap formed by a few dead trees (Runkle 1992) or an area of less than 200 m^2 (McCarthy 2001). Larger openings in the canopy are usually not considered as gaps but as patches. Processes that characterize their life cycle are more like the dynamics of a new stand resulting from a small-scale disturbance (McCarthy 2001). Gaps can disappear in two ways: either by the growth of trees located in the gap to canopy height or lateral filling by trees bordering the gap (Runkle 1982, Runkle 1985). The contribution of these mechanisms to gap closure depends heavily upon some of the gaps properties (e.g. gap surface or presence of advance regeneration). Lateral closure in large gaps might not be fast enough to prevent some saplings from reaching the canopy. On the other hand, small gaps will probably be closed laterally before a sapling has a chance of reaching the canopy. In this case, the shade-tolerant species stay suppressed under the canopy, awaiting the next gap opening. Thus, it is possible for advance regeneration to reach the canopy even in small gaps, through multiple canopy openings (Runkle and Yetter 1987). A number of studies have reported higher growth rate in gaps (Runkle 1981, Canham 1985, Canham 1990, McClure *et al.* 2000). As such, canopy opening-closure has an incidence on tree growth in gaps which can be observed in growth rings (McCarthy 2001). Canopy opening also has an effect on sapling height growth and stem density (Runkle 1998), suggesting that sapling groves structure could be regarded as a marker of gap age. However, gap age is pretty tricky to define, since gaps can be formed by the death of more than one tree and because they can expand and contract over time. To further complicate matters, multiple gap-forming tree deaths can be asynchronous, i.e. for a single gap the deaths are distinct events usually years apart (Runkle 1992). Each gap history can thus integrate a number of disturbances that have shaped its structure resulting from their timing and intensity.

The goal of this study was to predict canopy gap age from three dimensional structural parameters obtained through LiDAR. Our hypothesis was that some measurable features of the three dimensional structure of gaps are linked to their age. Gap age can be defined as the number of years since the first canopy opening disturbance or since the most recent disturbance event (Runkle 1992). In this study, we chose release detection from dendrochronological analysis as the simplest and most accurate method for finding gap age. This method relies on detecting the acceleration of tree growth due to the increased availability of resources in a newly opened gap and has been used in many studies to get information on forest disturbances (Runkle 1982, Poage and Peart 1993, Abe *et al.* 1995, Bräker 2002, Rubino and McCarthy 2004, Fraver and White 2005). Dendrochronological signal from sampled sapling in the gap's sapling groves should provide us with information on gap history.

1.2 Study sites

The study area consists of three sites with an area of approximately 1 km^2 each, separated by less than 20 km. The first two sites studied were located in the Ecological Reserve of the Forêt-la-Blanche. The reserve is a protected area of 2 052 hectares dedicated to the conservation of ecosystems and research (45° 44' N, 75° 16' W). The third site was located in the Lac-de-l'Écluse Exceptional Forest Ecosystem (45° 52' N, 75° 24' W), also under a conservation status. The dominant species in the three sites is sugar maple (*Acer saccharum* Marsh.), with American beech (*Fagus grandifolia* Ehrh.) as the codominant species. Other notable species in gaps or in the canopy are red maple (*Acer rubrum* L.), yellow birch (*Betula alleghaniensis* Britt.), basswood (*Tilia americana* L.), ironwood (*Ostrya virginiana* (Mill.) K. Koch), striped maple (*Acer pensylvanicum* L.), Eastern hemlock (*Tsuga canadensis* (L.) Carrière) and white ash (*Fraxinus americana* L.). The average annual temperature varies from 2.5 to 5°C, with a growing season ranging from 180-190 days. Mean annual precipitation ranges from 900 to 1000 mm (Robitaille and Saucier 1998). The three study sites are old growth forests that have not suffered catastrophic disturbances for hundreds of years. However, the three sites studied were affected by the North American ice storm of 1998, receiving between 40 and 100 mm of freezing rain (Olthof *et al.* 2004). Furthermore, beech bark disease (BBD) has been present for

several years, causing mortality and defects, particularly on large beeches (Houston 1994). Indeed, during our 2009 survey, many of the large beeches in our study area were dead or dying.

1.3 Data

Airborne discrete return LiDAR data was acquired in September 2007 using an Optech ALTM 3100 instrument at an average altitude of 1300 m (Hopkinson 2007). Average point density was over two returns per m^2, maximum half-scan angle was 20°, scan rate was 41 Hz and line spacing was 750 m (for a targeted 50% overlap between strips). Preprocessing of the LiDAR data was accomplished with the Terrascan software package (Terrasolid, Finland). This includes data cleaning, bird hits removal and classification of LiDAR points as being ground hits using the morphological properties of the point cloud.

2. Methodology

2.1 LiDAR data processing and gap detection

Treatments to produce height models and to detect gaps were done in ArcGIS Desktop (version 9.1, Environmental research institute Inc.). A digital elevation model (DEM) was initially created from LiDAR points classified as ground (Lim *et al.* 2003). The DEM was produced by interpolation with a Triangulated irregular network (TIN) to a resolution of 50 cm. A digital surface model (DSM) was produced in the same way with the first LiDAR returns. The canopy height model (CHM), a representation of the height of vegetation above ground, was then obtained by subtracting the DEM from the DSM.

Gaps were obtained from an adaptive thresholding algorithm applied on local height. This algorithm has been created to detect height changes in the CHM that are steep enough to represent a break in the canopy. This method of gap detection gives conservative gap boundary compared to other methods based on Lidar (Koukoulas and Blackburn 2004). Using this algorithm, a CHM pixel is classified as a gap if the height it represents is smaller than one of the two height thresholds (see equations 1 and 2) computed using the surrounding pixels in a circular neighbourhood of 0.25 ha:

$$H_i < [H_{¼ha} - (1.5 * SD_{H¼ha})] \quad (1)$$

$$H_i < (0.20 * H_{max¼ha}) \quad (2)$$

Where H_i is the evaluated CHM pixel value, $H_{¼ha}$ is the height average of the CHM in an area of 0.25 ha around the pixel, $SD_{H¼ha}$ and $H_{max¼ha}$ are the standard deviation and the maximum value of the CHM in the same area respectively. Adjacent gap pixels in all directions were then grouped to form a gap. Gaps with less than 4 m^2 were eliminated to remove artefacts due to the porosity of the canopy and those caused by the variation in point density.

2.2 Field data collection

Fieldwork was conducted during summer 2009. LiDAR-detected gaps were chosen randomly in the three landscapes. Very small gaps, i.e. less than 20 m^2, were not sampled because of difficulties in positioning and sampling them. Most of these very small gaps were in high canopies and appeared to have been created by broken branches, while some others were tree gaps in the process of closing laterally. The position of a central point in each gap was obtained with a GPS (ProMark3, Thales Navigation Inc.). Differential corrections were applied afterward using data from the Natural resources Canada's Ottawa GPS reference station, located about 50 km from the study sites. Post-processing of the GPS gave us an estimated average precision of 5 m. Other elements of gaps where positioned relative to the central GPS point with a distance rangefinder (Haglof Vertex 3 Rangefinder /Hypsometer) and a compass (Sunnto KB-14/360R). Woody debris that could be associated with stumps inside or close to the gap border were identified as gap makers. Gap makers with diameter at breast height (DBH) ≥ 250 mm were located with a distance rangefinder and

identified to the species if possible. Inside the gap, sapling groves with distinct height were identified. The grove position and characteristics, such as mean height, density, and composition, were recorded. For each distinguished sapling grove, a sapling with a DBH ≥ 50 mm was selected and a core was taken to the heart using a Pressler's probe.

2.3 Detection of growth release

The cores were air-dried, glued onto wooden boards and sanded to make the growth rings appear. The rings were measured using a Leica MZ125 stereomicroscope (40 X) coupled to a Velmex electronic digitizing table that has an accuracy of 0.01 mm. The cores were not cross-dated since our assumption is that the dendrochronological signature of each core is the result of differences in gap's history, whether it is local to the gap or global to the landscape. The detection of growth release was performed with a method adapted to species and conditions of our study sites (Gravel *et al.* 2009). A significant growth release is defined as an increase in radial growth of 100% sustained for at least four years, compared to the previous year. The year of each release were recorded for each core.

2.4 Fieldwork data validation

Some sampled gaps were removed from the dataset because the GPS point didn't fall within the detected gap boundary or because the tree core was unusable. This left 172 gaps in our dataset, on a total of 228 sampled gaps. Also, a number of sapling groves were removed from the dataset because they were not located within the gap boundaries detected by LiDAR, i.e. the gap detection method didn't include in the gap boundary some regeneration groves that were recorded as part of the gap during fieldwork. This problem has affected 25% of the 172 gaps in the dataset. Among those, 80% had only one sapling grove within the gap boundary detected by LiDAR and 20% had two to four groves. Those were probably formed from multiple gap-forming tree falls (Barden 1981) or by gap edge expansion (Runkle 1998). Gaps with more than one sapling grove were removed from the statistical analysis, which left us with 138 gaps for the statistical analysis.

2.5 Cluster analysis

For each gap, the 20^{th}, 50^{th} (median value), and 80^{th} percentiles values of the CHM were computed (R statistical package, version 2.10.0) and gaps were classified in groups using these three CHM percentiles values. Cluster analysis was conducted for distinguishing gap structure using Ward's hierarchical clustering method with Euclidean distances (hclust package in R statistical package, version 2.10.0). The number of clusters was chosen interactively from observation of the dendrogram.

2.6 Statistical analysis

Analysis of variance (ANOVA) was performed with the gap structure clusters as factors and each of the 20^{th}, 50^{th}, 80^{th} percentiles of CHM height and the mean sapling grove height estimated during fieldwork (mentioned as sapling grove height from here on) as variables using R (aov function, R version 2.10.0). When significant ANOVA differences were found, Tukey's contrasts post-hoc mean comparison test was performed. Pearson's product-moment correlation was calculated on the median of CHM height and the sapling grove height. Analyses were also done to find a relationship between gap age and structure. To do so, analysis of variance (ANOVA) was performed with the gap structure clusters as factors and gap age as variable. Pearson's product-moment correlation was performed on the CHM height percentiles values and gap age.

3. Results and discussion

Gaps detected from LiDAR occupied 6.2%, 7.5% and 7.0% of study sites 1, 2 and 3 respectively. Mean gap area was 21.2 m^2 with standard deviation (SD) of 5.1 m, 20.4 m^2 with SD of 6.3 m, and 23.1 m^2 with SD of 5.5 m in sites 1, 2 and 3 respectively. Proportion of the study sites occupied by gaps is lower than found in the literature (Runkle 1982). This is surprising, considering that the presence of BBD usually increases total gap area (Krasny and DiGregorio 2001). The proportion of gaps with more than one sapling grove is also higher than in other studies (Barden 1981). This difference probably comes from the differences in the

definition of gaps (Barden 1989) or from divergences in multiple-treefalls gap definition between authors (Runkle 1992).

Growth release detection showed that each year an average of 2.7% of the saplings started a new period of growth release between 1960 and 2005 (figure 1).

Figure 1(a)

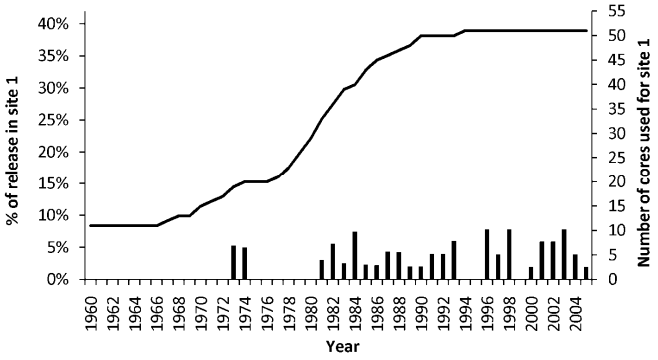


Figure 1(b)

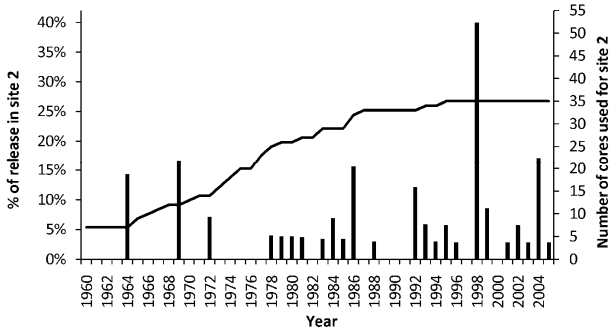


Figure 1(c)

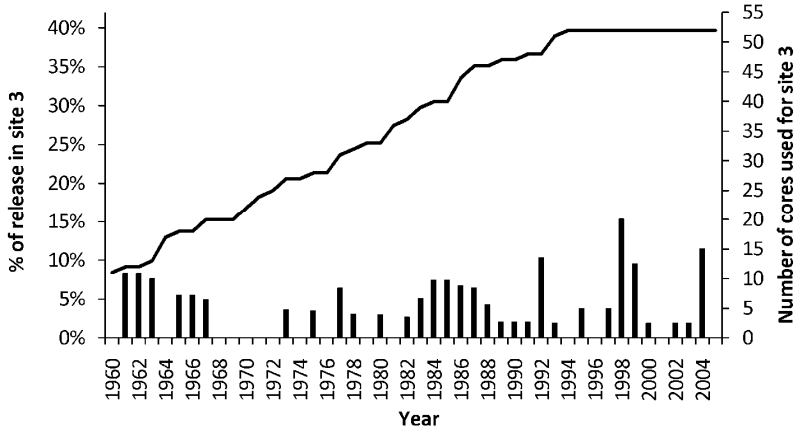


Figure 1. Proportion of cores starting a growth release (growth increase >100% sustained for at least four years, compared to the previous year) and number of cores used between 1960 and 2005 in site 1 (figure 1(a)), site 2 (figure1(b)) and site 3 (figure 1(c)). Proportion of cores starting a growth release is represented as bars and is associated to the left Y axis. Number of cores used is represented as a line and is associated to the right Y axis.

Some years exhibit a higher than average amount of new releases for sites 2 and 3. The largest was in the late 1990's, which corresponds to the growing season following the ice storm of 1998. Other noteworthy years for large number of growth releases are 1992 and 2004. Some differences can be observed in those years between study sites. Site 1 does not have years with more release than usual. That site has a lower mean elevation than the other two sites, which might explain the difference in 1998. Other growth releases differences between study sites are probably caused by local variations of disturbances intensity or site characteristics.

Six gap structure clusters were obtained from the CHM percentiles values. The ANOVA performed on the clusters and the height percentiles showed that they have distinct structural characteristics (table 1).

Table 1. Mean and standard deviation values of the 20th, 50th and 80th CHM height percentiles for each gap structure clusters. Below are the ANOVA and correlation between the six gap structure clusters, height percentiles, sapling grove height and minimum gap age. Below that are the ANOVA and correlation between gap structure clusters subset (clusters 1, 2, 3 and 5), height percentiles, sapling grove height and minimum gap age.

Gap structure clusters	20th percentile	50th percentile	80th percentile	Sapling grove height
1 – Mean ± SD	5.4 ± 0.91 [a*]	6.8 ± 0.59 [a]	8.1 ± 0.70 [a]	6.6 ± 1.72 [a]
2 – Mean ± SD	3.5 ± 1.04 [b]	5.0 ± 0.59 [b]	6.4 ± 0.81 [b]	5.3 ± 1.69 [b]
3 – Mean ± SD	7.3 ± 0.59 [c]	9.1 ± 0.79 [c]	10.6 ± 1.24 [c]	7.5 ± 2.52 [a]
4 – Mean ± SD	3.4 ± 1.55 [a]	7.4 ± 1.52 [d]	11.4 ± 1.34 [c]	3.2 ± 2.14 [c]
5 – Mean ± SD	1.3 ± 0.62 [d]	3.0 ± 0.63 [e]	5.1 ± 1.11 [d]	3.5 ± 2.39 [bc]
6 – Mean ± SD	10.4 ± 1.81 [e]	13.4 ± 0.67 [f]	15.0 ± 0.64 [e]	2.30 ± 1.5 [bc]
Analysis using all clusters (1-6)				
ANOVA with clusters as factors				
F	93.24	155.48	162.25	18.04
P	< 0.001	< 0.001	< 0.001	< 0.001
Correlation with minimum gap age	0.06	0.09	0.12	0.12
Correlation - P value	0.243	0.137	0.087	0.078
Analysis using clusters 1, 2, 3 and 5				
ANOVA with clusters as factors				
F	158.33	310.75	131.02	14.19
P	< 0.001	< 0.001	< 0.001	< 0.001
Correlation with minimum gap age	0.16	0.14	0.18	0.15
Correlation - P value	0.048	0.071	0.035	0.058

* Clusters with the same letter are non-significantly different from post-hoc means comparison with Tukey's contrasts at the 95% confidence level.

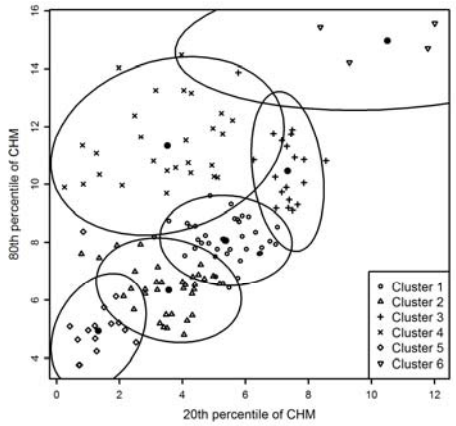


Figure 2. Scatterplot of the clustered gaps (Cluster 1 (o), cluster 2 (Δ), cluster 3 (+), cluster 4 (x), cluster 5 (◊), cluster 6 (▽) with the 20th and 80th percentiles values of the CHM. 95% confidence ellipse and its centre are drawn for each cluster.

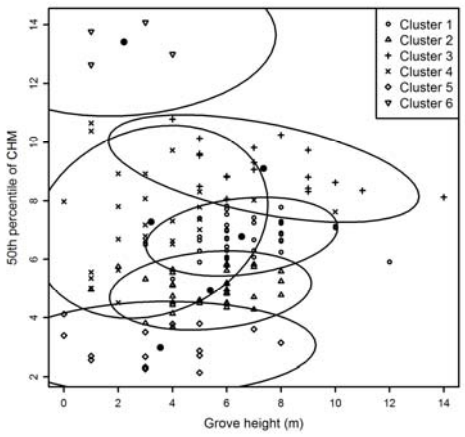


Figure 3. Scatterplot of the clustered gaps (Cluster 1 (o), cluster 2 (Δ), cluster 3 (+), cluster 4 (x), cluster 5 (◊), cluster 6 (▽)) with the estimated grove height and the median height values of the CHM. 95% confidence ellipse and its centre are drawn for each cluster.

Tukey's contrasts shows that most clusters are significantly different for each height percentile parameter. Clusters 1 and 4 in the 20th height percentile and clusters 3 and 4 in the 80th height percentiles are not significantly different. Gap structure clusters are arranged along a height gradient with clusters 5, 2, 1, 3 and 6 in ascending order, except for cluster 4 (figure 2).

Gap structure clusters have a compact distribution around their centre, except for clusters 4 and 6. Cluster 4 has a larger 95% confidence ellipse and its height distribution is more skewed toward higher values than the other clusters.

Clusters show a comparable pattern of height gradient with the sapling grove height and CHM median values (figure 3).

The correlation between the median CHM values and the sapling grove heights was performed on data from all clusters and on a subset composed of clusters 1, 2, 3 and 5. Significant correlation is achieved using the subset (R^2 = 0.23), but isn't when clusters 4 and 6 are included (R^2 = 0.01). Correlation between height parameters is much lower than anticipated. This might be a consequence of the way gaps are detected. Detected gap borders can sometimes be at a high height if the canopy is very high. It was also found that the clusters are strongly associated to the study sites (table 2).

Table 2. Number of gaps in each gap structure cluster by study sites.

Cluster	Site 1	Site 2	Site 3
1	8	11	20
2	6	11	17
3	5	3	11
4	25	1	1
5	3	9	3
6	4	0	0

Site 1 has most of the cluster 4 and all of the cluster 6 gaps, site 2 has most of the cluster 5 gaps (low gap heights) and site 3 has at least 50% of clusters 1 to 3 (intermediate gap heights). Detected gap structure was associated to study sites characteristics, which probably indicates that it was significantly affected by other factors. Such local factors might be border trees height, pre-established shade-tolerant tree regeneration, ungulate browsing and differential growth rate due to site conditions and/or sapling grove composition.

Growth release frequencies do show some differences between clusters, but not as much as anticipated (figure 4). However, some patterns of synchronous growth releases can be observed. They would be indicative of synchrony in gap openings. Synchronous growth release events are defined as a higher than normal number of release events happening during one year. On the other hand, gaps form at a continuous rate even without major disturbance events. Those would be asynchronous release events. Clusters 5 and 2, the two lowest clusters in height, are the ones with the most numerous recent synchronous growth releases in 1998 and 2004 (figure 4 (*b)* and *(e*)).

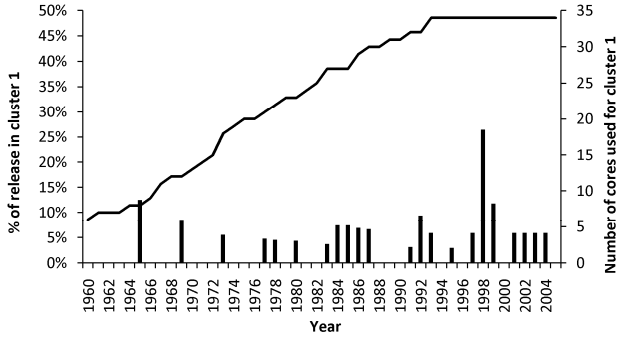


Figure 4 (a)

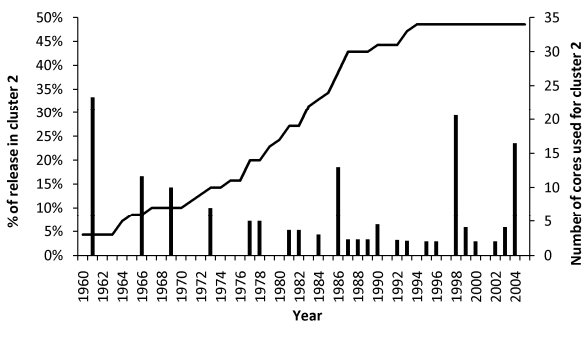


Figure 4 (b)

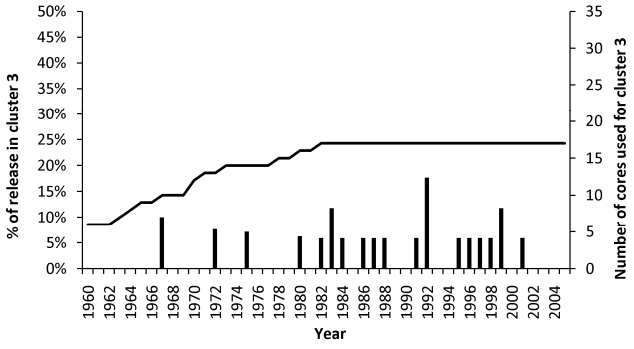


Figure 4 (c).

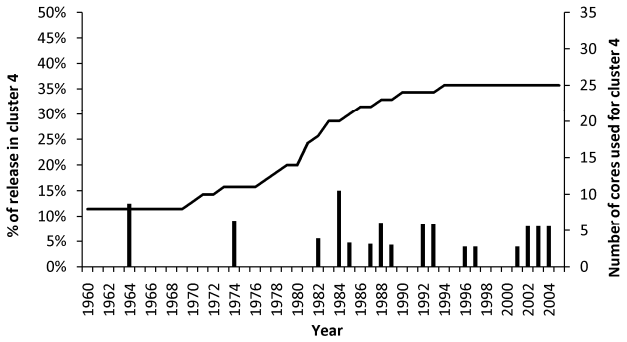


Figure 4 (d).

Figure 4 (e).

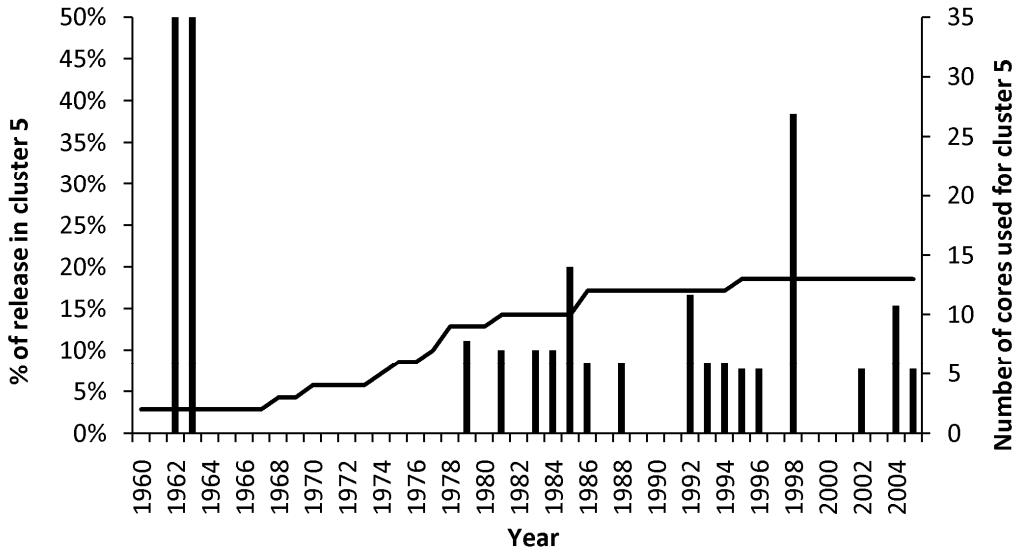


Figure 4. Proportion of cores starting a growth release (growth increase >100% sustained for at least four years, compared to the previous year) and number of cores used between 1960 and 2005 for cluster 1 (figure 4 (a)), cluster 2 (figure 4 (b)), cluster 3 (figure 4 (c)), , cluster 4 (figure 4 (d)) , cluster 5 (figure 4 (e)). Proportion of cores starting a growth release is represented as bars and is associated to the left Y axis. Number of cores used is represented as a line and is associated to the right Y axis.

Cluster 1 is similar, but the last synchronous growth release was only in 1998 (figure 4 (*a*)). Cluster 3, the highest in heights percentiles, is the only one without recent growth releases (figure 4 (*c*)) and has a weak synchronous growth release in 1992. Cluster 4 has a few recent growth releases, but doesn't have any around the 1998 ice storm. It has no synchronous growth release (figure 4 (*d*)), a result that clearly contrast with the other clusters. This could be why this cluster has a structure quite different from the others. These differences between clusters are indicative of structure response to major synchronous or asynchronous disturbances events.

We decided to use the most recent detected release as the gap age since the structure clusters heights percentiles seemed to be lower for clusters with recent disturbances and higher for clusters without any. ANOVA between gap age for all gap structure clusters gave no significant result (df=5, F= 1.45 and P = 0.2091). ANOVA between gap age for clusters 1, 2, 3 and 5 gave close to significant results (df=3, F= 2.35

and P = 0.0768). Tukey's contrasts indicate that only clusters 2 and 3 are different to the 95% confidence level. Correlation between gap age and height percentile values was performed on data from all clusters and on a subset composed of clusters 1, 2, 3 and 5 (table 1). Even though cluster 5 had low height values and had seen some recent disturbances, its mean gap age was still high and thus did not fit with the general gradient of height and age observable with the other clusters (table 3).

Table 3. Gap age mean and standard deviation for each structure cluster

Cluster	Mean	Standard deviation
1	9.3	11.7
2	6.9	6.87
3	14.8	11.8
4	10.7	14.1
5	9.1	12.2
6	4.5	4.20

4. Conclusion

LiDAR derived data was found to be useful in detecting gaps in the forest cover. Gap properties like area and structure were also extracted from LiDAR derived data. Gap structure clusters appeared to have distinct history of synchronous disturbances. However, relationship between gap age and structure was not as strong as expected, even though growth releases distribution differences between clusters were apparent. One of the reasons of such weak relationship might be the insufficient number of sampled saplings in the groves to show the gap disturbance history. Taking multiple cores from each sapling grove would have given us a more precise gap age structure, but would have required much more resources and time. Gap dynamics is a much more complex process than was assumed at first. Even though we're only in the beginning of its widespread use, LiDAR has made some significant contributions to our understanding of gap dynamics and will continue to do so in the future.

Acknowledgements

We would like to thank the Centre d'étude de la forêt (CEF) and the human sciences faculty of the Université du Québec à Montréal for providing funds for the principal author's participation to the conference. We are very grateful to Patsy Teske, Michel Staniforth (Lauzon ressource forestières), SEPAQ (Reserve faunique Papineau-Labelle) and MDDEP (Réserve écologique de la Forêt-la-Blanche) for providing access authorisation and help during fieldwork. Particular thanks are due to Audrey Girard-Miron for assistance during and after fieldwork. LiDAR data was acquired with funds from the Canadian Foundation for Innovation. We thank the Applied geomatics research group and Dr. Chris Hopkinson for providing the LiDAR data. Finally, too many IQAFF personnel provided help that proved invaluable during fieldwork to thank them all by name.

References

ABE, S., MASAKI, T. and NAKASHIZUKA, T., 1995, Factors influencing sapling composition in canopy gaps of a temperate deciduous forest. Plant Ecology, 120, pp. 21-32.

BARDEN, L., 1981, Forest development in canopy gaps of a diverse hardwood forest of the southern Appalachian Mountains. Oikos, 37, pp. 205-209.

BARDEN, L. S., 1989, Repeatability in forest gap research: Studies in the Great Smoky Mountains. Ecology, 70, pp. 558-559.

BRÄKER, O. U., 2002, Measuring and data processing in tree-ring research - a methodological introduction. Dendrochronologia, 20, pp. 203-216.

BROKAW, N. V. L., 1982, The definition of treefall gap and its effect on measures of forest dynamics. BIOTROPICA, 14, pp. 158-160.

CANHAM, C. D., 1985, Suppression and release during canopy recruitment in Acer saccharum. Bulletin of the Torrey Botanical Club, 112, pp. 134-145.

CANHAM, C. D., 1990, Suppression and release during canopy recruitment in Fagus grandifolia. Bulletin of the Torrey Botanical Club, 117, pp. 1-7.

FRAVER, S. and WHITE, A., 2005, Identifying growth releases in dendrochronological studies of forest disturbance. Canadian journal of forest research, 35, pp. 1648-1656.

GAULTON, R. and MALTHUS, T. J., 2010, LiDAR mapping of canopy gaps in continuous cover forests: A comparison of canopy height model and point cloud based techniques. International Journal of Remote Sensing, 31, pp. 1193 - 1211.

GRAVEL, D., BEAUDET, M. and MESSIER, C., 2009, Large-scale synchrony of gap dynamics and the distribution of understory tree species in maple–beech forests. Oecologia, 162, pp. 1-9.

HOPKINSON, C., 2007, Gatineau LiDAR survey data report, Applied Geomatics Research Group, Centre of Geographic Sciences, NSCC Annapolis Valley Campus, pp. 1-8.

HOUSTON, D., 1994, Major new tree disease epidemics: beech bark disease. Annual Review of Phytopathology, 32, pp. 75-87.

KOUKOULAS, S. and BLACKBURN, G. A., 2004, Quantifying the spatial properties of forest canopy gaps using LiDAR imagery and GIS. International Journal of Remote Sensing, 25, pp. 3049-3072.

KRASNY, M. E. and DIGREGORIO, L. M., 2001, Gap dynamics in Allegheny northern hardwood forests in the presence of beech bark disease and gypsy moth disturbances. Forest Ecology and Management, 144, pp. 265-274.

LIM, K., TREITZ, P., WULDER, M., ST-ONGE, B. and FLOOD, M., 2003, LiDAR remote sensing of forest structure. Progress in Physical Geography, 27, pp. 88.

MCCARTHY, J., 2001, Gap dynamics of forest trees: A review with particular attention to boreal forests. Environmental reviews, 9, pp. 1-59.

MCCLURE, J. W., LEE, T. D. and LEAK, W. B., 2000, Gap capture in northern hardwoods: patterns of establishment and height growth in four species. Forest Ecology and Management, 127, pp. 181-189.

OLTHOF, I., KING, D. and LAUTENSCHLAGER, R., 2004, Mapping deciduous forest ice storm damage using Landsat and environmental data. Remote Sensing of Environment, 89, pp. 484-496.

PLATT, W. J. and STRONG, D. R., 1989, Special Feature: Gaps in Forest Ecology. Ecology, 70, pp. 535-535.

POAGE, N. J. and PEART, D. R., 1993, The Radial Growth Response of American Beech (Fagus grandifolia) to Small Canopy Gaps in a Northern Hardwood Forest. Bulletin of the Torrey Botanical Club, 120, pp. 45-48.

ROBITAILLE, A. and SAUCIER, J.-P., 1998, Paysages régionaux du Québec méridional, pp. 213 (Québec: Gouvernement du Québec, Ministère des ressources naturelles).

RUBINO, D. L. and MCCARTHY, B. C., 2004, Comparative analysis of dendroecological methods used to assess disturbance events. Dendrochronologia, 21, pp. 97-115.

RUNKLE, J. R., 1981, Gap Regeneration in Some Old-growth Forests of the Eastern United States. Ecology, 62, pp. 1041-1051.

RUNKLE, J. R., 1982, Patterns of Disturbance in Some Old-Growth Mesic Forests of Eastern North America. Ecology, 63, pp. 1533-1546.

RUNKLE, J. R., 1985, Disturbance regimes in temperate forests. In The ecology of natural disturbance and patch dynamics, S. T. A. Pickett and P. S. White (Ed.), (San Diego: Academic Press, Inc.).

RUNKLE, J. R., 1992, Guidelines and Sample Protocol for Sampling Forest Gaps (Portland, Or: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station), pp. 44.

RUNKLE, J. R., 1998, Changes in southern Appalachian canopy tree gaps sampled thrice. Ecology, 79, pp. 1768-1780.

RUNKLE, J. R. and YETTER, T. C., 1987, Treefalls revisited: gap dynamics in the southern Appalachians. Ecology, 68, pp. 417-424.

SEYMOUR, R. S., WHITE, A. S. and DEMAYNADIER, P. G., 2002, Natural disturbance regimes in northeastern North America—evaluating silvicultural systems using natural scales and frequencies. Forest Ecology and Management, 155, pp. 357-367.

VEPAKOMMA, U., ST-ONGE, B. and KNEESHAW, D., 2007, Spatially explicit characterization of boreal forest gap dynamics using multi-temporal lidar data. Remote Sensing of Environment, 112, pp. 2326-2340.

Estimation of forestry parameters in mountainous coppice stands using airborne laser scanning

Jean-Matthieu Monnet†, Eric Mermin†, Jocelyn Chanussot‡ & Frédéric Berger†

jean-matthieu.monnet@cemagref.fr

†UR EMGR, Cemagref, France

‡Gipsa-Lab, Grenoble Institute of Technology, France

Abstract

An area-based method is implemented to predict forest stand parameters from airborne laser scanning data. Multiple regression models are calibrated with 31 field plots inventoried in a hillside dominated by coppice stands, located in the French Alps. Cross-validated prediction accuracies are respectively 13.2, 14.3, 19.4 and 25.5% for dominant height, mean diameter, basal area and stem density. Reducing calibration plot radius greatly influences prediction results. Although median values of forest parameters remain unchanged, field observations variability is higher for smaller plots. Depending on the forest parameter, prediction accuracy is significantly correlated with some distribution statistics (minimum and maximum values) of field observations computed with various radii. Indeed, greater variance or outliers may degrade the degree of fit of regression models.

1. Introduction

In alpine environments with high topographical constraints, evolutions in forestry practises and labour costs resulted in a progressive neglecting of mountainous stands. Near the footslopes, coppice stands which used to provided local inhabitants with fuelwood are quite frequent in the French Alps. They were among the first to be left over as a combination of the development of other energy sources and of the impossibility to grow high value timber products due to poor site quality. However, fossil fuels rarefaction and global warming alarms prompted public authorities to set ambitious objectives of increased woody biomass harvesting (Ginisty et al. 2007). Unfortunately, information about forest stands characteristics is now missing or outdated. Due to accessibility constraints, conventional field inventory methods can not provide the information required to forecast harvesting operations. Prospecting difficulties are all the more critical since forest stands display a high spatial heterogeneity linked with complex landform patterns encountered in mountainous areas.

High hopes have been set on airborne laser scanning (ALS) as this remote sensing technique was shown to be successful in stand parameters estimation and tree detection for coniferous forests (see review by Hyyppä et al. 2008). Following works also demonstrated its accuracy in others contexts such as tempered deciduous forests (Popescu et al. 2002, Patenaude et al. 2004) and alpine environments (Heurich and Thoma 2008, Hollaus et al. 2009). To our knowledge, the case of mountainous coppice stands has not been investigated so far.

The aim of this paper is to evaluate the efficiency of ALS for forest parameters estimation in mountainous coppice stands. The area-based method proposed by Næsset (2002) is implemented. The effects of forest spatial heterogeneity are investigated by examining the influence of calibration plot size on field observations and on regression models accuracy.

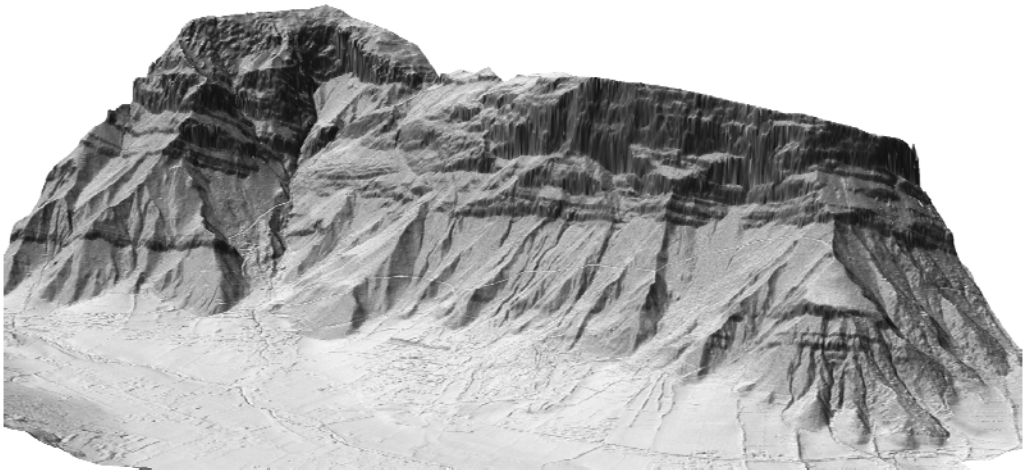


Figure 1. Shaded digital terrain model of the study area

2. Material

Study area

*t*he study area is a 4 km^2 hillside situated in the French Alps (town of Saint Paul de Varces, 45° 04'17"N, 05°38'25"E)(figure 1). The forest is mainly constituted of coppice stands and deciduous stands on poor quality sites, dominated by Italian maples (*Acer opalus*) and downy oaks (*Quercus pubescens*). Downslope, old chestnut (*Castanea sativa*) coppice stands are frequent. Common whitebeam is present in all the area, especially at the foot of cliffs. In thalwegs or in upper parts with better site quality, ash (*Fraxinus excelsior*) and beech (*Fagus sylvatica*) are common. Some areas have a dense understory of holly (*Ilex Aquifolium*), common hazel (*Corylus Avellana*) or box (*Buxus sempervirens*). Altitude ranges from 330 to 1270 m above sea level. High limestone cliffs overhang the area and rockfall events are frequent. No major silvicultural or harvesting operations have been performed in the area for more than fifty years.

Field data

From September to November 2009, *N* = 31 circular field plots were inventoried. Plots were distributed every 400 m along the 550, 750, 950 and 1150 m height contours, resulting in an irregular sampling scheme where horizontal distances between neighbouring plots ranged from 180 to 412 m with a mean value of 302 m. Plot centres were georeferenced using a Trimble GPS Pro XRS receiver. After differential correction with the Pathfinder®; software, position precision (95% confidence radius) ranged from 0.6 to 1.5 m. All trees with diameter at breast height larger than 5 cm and located within 10 m horizontal distance from the plot centre had their diameter measured with a tape. Their positions to the plot centre were recorded using a Suunto KB-14 compass and Suunto PM-5 clinometer mounted on a tripod, and a Vertex III hypsometer. Maples (mainly *Acer opalus*), downy oak (*Quercus pubescens*) and common whitebeam (*Sorbus aria*) represented nearly 60 % of the stems. Ten tree heights were also measured on each plot with the hypsometer. Height sampling probability was proportional to stem basal area to ensure that dominant trees would be represented.

Laser data

Laser data was acquired on August 27th, 2009 over 8.6 km^2 with a fullwave RIEGL LMS-Q560 scanner. Acquisition parameters are summarised in table 1[1]. Echoes were extracted from the binary acquisition files and georeferenced with the RIEGL software suite. The contractor also classified the resulting point cloud into ground and non-ground echoes using the TerraScan software, which implements an algorithm based on

iterative surface reconstruction by triangulated irregular network (Axelsson 2000). Final echo density was 10 m-2.

Table 1. Laser scanner acquisition parameters

Item	Value
Wavelength	1550 nm
Pulse repetition rate	200 kHz
Scan frequency	111.1 Hz
Scan angle	± 30°
Flight height	550 m
Laser footprint	0.29 m
Theoretical point spacing	0.47 m

3. Methods

The area-based, two steps method proposed by Næsset (2002) is implemented to predict forest parameters from airborne laser scanning data. To evaluate the effect of calibration plot size on prediction accuracy, plot radius is reduced by excluding trees situated further than various distance thresholds before reiterating the method. Tested radii are $r \in \{5,5.5,...,10\}$.

Forest parameters

The following stand parameters are computed for each plot $j \in \{1,2,...,31\}$ and each radius r: basal area (G_r^j: surface occupied by the horizontal section of tree stems at 1.30 m height), stem density (N_r^j) and mean diameter at breast height (D_r^j). Dominant height (H_{10}^j: mean height of the 30 highest trees per hectare) is calculated for r = 10 m only, as the sampling scheme does not ensure that enough measured trees are included within each radius. Wilcoxon signed-rank tests are performed to compare forest parameters observations obtained with different plot sizes. Correlation between field observations statistics and plot radius is evaluated by computing Spearman ρ.

Extraction of laser metrics

For each plot j, laser points within r meters horizontal distance form the plot centre are extracted. Their relative heights are computed by subtracting the terrain height at their orthometric coordinates. Terrain surface is estimated by bilinear interpolation of points classified as ground points. Points with relative height lower than 2 m are excluded to avoid influence of dense shrubs understory. Three point groups are then constituted according to return positions: single echoes (only one echo for a given pulse), first echoes and last echoes. For each group two types of laser metrics are calculated. Height metrics correspond to the breakpoints of four height bins containing an equal number of points: minimum ($h_{g,0}$), first quartile ($h_{g,0.25}$), median ($h_{g,0.5}$), third quartile ($h_{g,0.75}$) and maximum ($h_{g,1}$) values, plus mean height ($h_{g,mean}$). Subscript $g \in \{s,f,l\}$ refers to the point groups: single, first or last echoes. Three density metrics are computed as the proportion of echoes recorded below height thresholds corresponding respectively to 0.75, 0.5 and 0.25 of the maximum echo height recorded on the plot ($d_{g,0.75}$, $d_{g,0.5}$ and $d_{g,0.25}$ respectively). For each radius r, the predictors set $P_r = (P_r^i)_{i \in \{1,...,27\}}$ consists of 31 observations of 3 × 9 = 27 laser metrics.

Multiple regression models

For each predictors set P_r and each corresponding dependent variables $y_r \in \{H_r , G_r , N_r, D_r\}$, a multiple regression model is fitted by ordinary least squares.

$$y_r = b_r + \sum_{i=1}^{27} a_r^i \times P_r^i \qquad \text{with } ((a_r^i)_{i\in\{1,\dots,27\}}, b_r) \text{ the model parameters} \tag{1}$$

Models including a maximum of three predictors are tested by exhaustive search among possible combinations. Models which do not fulfil the linear model assumptions or including a predictor with a partial *p*-value greater than 0.05 are discarded. The model with highest adjusted coefficient of determination (*adj-*R^2) is selected.[2]

Prediction accuracy is evaluated in leave-one-out cross validation by computing the root mean square error (*RMSE*) and its coefficient of variation CV_{RMSE}.

$$RMSE = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(y_j - \hat{y}_j)^2} \qquad \text{with } \begin{cases} y_j \text{ the observed values} \\ \hat{y}_j \text{ the predicted values} \end{cases} \tag{2a}$$

$$CV_{RMSE} = \frac{RMSE}{\bar{y}} \qquad \text{with } \bar{y} = \frac{1}{N}\sum_{j=1}^{N} y_j \tag{2b}$$

Differences between predicted values and field observations are evaluated by Wilcoxon signed-rank tests. Spearman ρ is computed to assess correlation between prediction accuracy and plot radius for each stand parameter.

4. Results

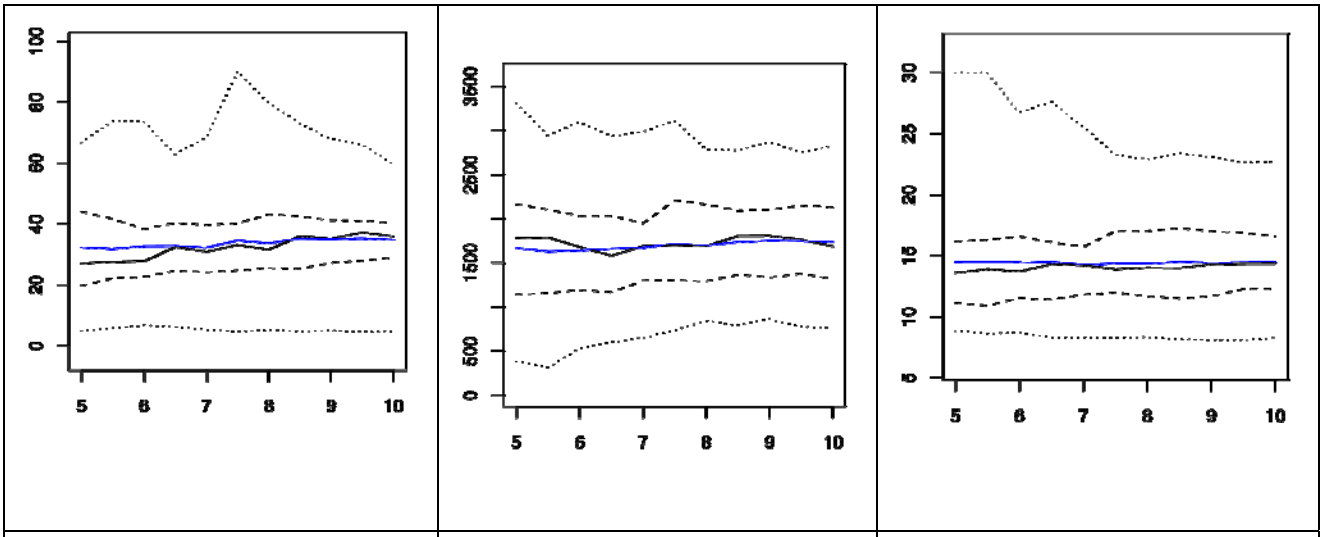
Field observations

Forest plots statistics for 5, 7.5 and 10 m radii are displayed in table 2[3]. As site quality is rather poor in the area, average dominant height is only 17.8 m. Generally, stands with high values for basal area, dominant height and mean diameter are located on a few good quality sites in thalwegs with deep soil, such as the ash-dominated plot #14 with H_{10} = 28.5 m, G_{10} = 59.7 $m^2.ha^{-1}$ and D_{10} = 21.6 cm. Small values are encountered on steep slopes at the bottom of cliffs with rockfall activity. For example, plot #21 is located on a scree and has H_{10} = 13.7 m, G_{10} = 4.6 $m^2.ha^{-1}$ and D_{10} = 8.3 cm. In such areas stem density is highly variable.

Table 2. Dominant height (H), basal area (G), stem density (N) and mean diameter (D) field observations statistics for calibration plot radius r ∈{5, 7.5, 10} (N = 31 plots).

	H (m)	*G* ($m^2.ha^{-1}$)			*N* (ha^{-1})			*D* (cm)[4]		
Radius	10	5	7.5	10	5	7.5	10	5	7.5	10
Mean	17 .8	32 .3	34 .6	34 .8	1668	1714	1735	14 .5	14 .4	14 .5
Min	8 .1	4.9	4.7	4.6	382	736	764	8.9	8.3	8.3
Max	28 .5	66.8	90.1	59.7	3310	3112	2833	30.0	23.3	22.7
Sd	5 .3	16.7	15.9	11.4	724	560	577	4.6	3.9	3.6

Figure 2 plots field observations statistics of basal area, stem density and mean diameter as functions of plot radius. Although mean basal area seems to increase with plot radius, Wilcoxon signed-rank tests indicate that the medians of the observations differences (y_{r1} - y_{r2} with $r1 \neq r2$ and $y_r \in \{H_r, G_r, N_r, D_r\}$) are not significantly different from zero at the $p < 0.05$ level.



(a) Basal area ($m^2.ha^{-1}$) | *(b)* Stem density (ha^{-1}) | *(c)* Mean diameter (cm)

Figure 2. Influence of calibration plot radius (x-axis, meters) on field parameters observations. Dotted lines are the minimum and maximum values, dashed lines the first and third quartiles, black solid line is the median and blue solid line the mean value.

Spearman correlation tests (table 3) show that minimum values of basal area and mean diameter are negatively correlated with plot radius, whereas stem density minimum is positively correlated. First quartile (all tested parameters) and mean values (basal area and stem density) are also positively linked to plot radius. Negative ρ values are found for maximum stem density and diameter, and standard deviation of the basal area and mean diameter. The median is correlated to plot radius for basal area and mean diameter, whereas the third quartile yields no significant correlation.

Table 3. Spearman correlation coefficient ρ between field observations statistics and plot radius for basal area (G), stem density (N) and mean diameter (D).

	Min	1st quart.	Median	3rd quart.	Max	Mean	Sd
G	-0.68*	0.98***	0.91***	-0.02	-0.31	0.85**	-0.89***
N	0.84**	0.90***	0.12	0.11	-0.76**	0.87***	-0.54
D	-0.87***	0.77**	0.75*	0.57	-0.93***	-0.52	-0.98***

$***p < 0.001$, $**p < 0.01$, $*p < 0.05$ (two-sided test)

Multiple regression models

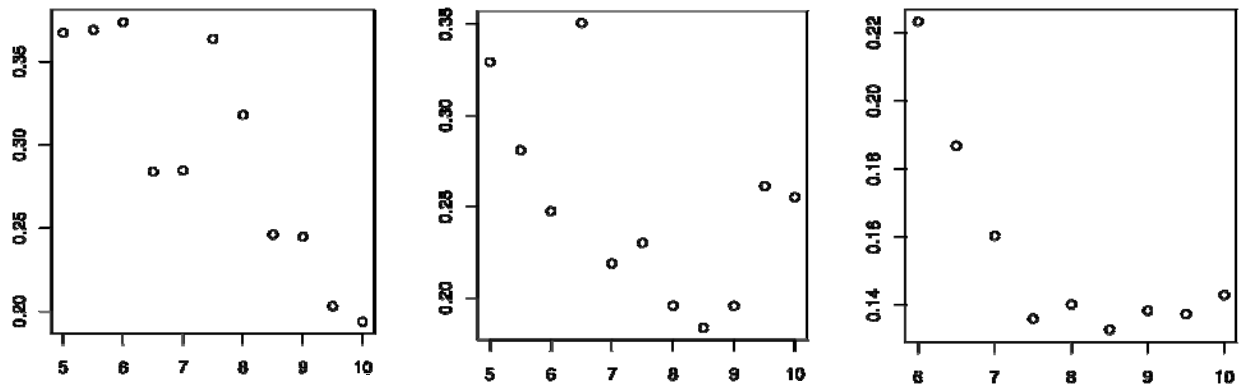
Selected multiple regression models for $r \in \{5, 7.5, 10\}$ are detailed in table 4. With 10 m radius calibration plots, prediction accuracies (cross validated coefficient of variation of the RMSE) are respectively 13.2, 14.3,

19.4 and 25.5 % for dominant height, mean diameter, basal area and stem density. For r = 7.5 m results slightly improve for mean diameter and stem density (respectively 13.6 and 23.0 %), but are worse for basal area (36.4 %). The lowest accuracies are achieved with 5 m radius: 32.9 % for stem density and 36.7 % for basal area. Wilcoxon signed-rank tests indicate that the median of the differences between predicted and observed values is not significantly different from zero at the $p < 0.05$ level for any forest parameter. Mean diameter models do not fulfil linear model assumptions for $r < 6$. It is noteworthy that laser metrics included in the models depend both on the forest parameter and on calibration plot radius.

Table 4. Selected multiple regression models for radius $r \in \{5, 7.5, 10\}$.

Variable	Radius	Laser metrics in the model	*adj-*R^2 (%)	*CV RMSE*(%)
Basal area (*G*)	5	$h_{s,0} + h_{s,1} + h_{f,mean}$	55.5	36.7
	7.5	$h_{f,0.75}$	46.0	36.4
	10	$h_{f,0.25} + h_{f,0.5} + d_{s,0.5}$	70.8	19.4
Stem density (*N*)	5	$h_{l,mean} + d_{f,0.75} + d_{l,0.25}$	51.0	32.9
	7.5	$h_{f,0.75} + h_{l,1} + d_{l,0.25}$	59.1	23.0
	10	$h_{l,0.5} + d_{f,0.25}$	58.0	25.5
Mean diameter (*D*)	5	linear model assumptions not satisfied		
	7.5	$h_{f,0} + h_{l,0.25} + d_{l,0.75}$	76.6	13.6
	10	$h_{s,0.75} + h_{f,0.75} + d_{l,0.25}$	71.4	14.3
Dominant height (*H*)	5, 7.5	no relevant due to sampling scheme		
	10	$h_{s,0.5} + h_{s,0.75} + h_{f,0}$	84.1	13.2

Figure 3 displays prediction accuracy of multiple regression models as a function of calibration plot radius for basal area, stem density and mean diameter. Spearman correlation coefficient ρ between the forest parameter and plot radius is significantly different from zero for basal area only (ρ = -0.87, $p < 0.001$, two-sided test). Stem density displays a similar decreasing trend but the coefficient of variation of the RMSE increases again when $r > 9$. For mean diameter the coefficient of variation of the RMSE decreases when plot radius increases from 6 to 7.5 m, and then remains stable around 14%.



(a) Basal area ($m^2.ha^{-1}$) *(b)* Stem density (ha^{-1}) *(c)* Mean diameter (cm)

Figure 3. Influence of calibration plot radius (x-axis, meters) on the cross-validated coefficient of variation of the RMSE (y-axis) of multiple regression models.

5. Discussion

The ALS based method produces accurate estimates of forest stand parameters, showing that it is suitable for deciduous forests such as coppice stands. Results obtained with 10 m radius calibration plots are consistent with those obtained by Heurich and Thoma (2008) with 34 deciduous plots located in the Bavarian Forest National Park (Germany). In their study a multiple regression was also performed with forest parameters as dependent variables and laser metrics as predictors. Accuracy was similar for basal area (20.3%) and mean diameter weighted by basal area (13.2%). Stem density also yielded the greatest error with 29.8%, whereas a better result was achieved for dominant height with 8.1%.

Even tough correlation is significant for basal area only, it is noteworthy that prediction accuracy varies greatly with plot calibration radius (figure 3). As the median differences between observed and predicted values are not significant for any of the plot radii, plot size may influence residuals variance or some higher order statistics.

Comparison of field parameters values obtained with different radii show that the median of differences are not significantly different from zero. However the mean and standard deviation, as well as several distribution quantiles, are significantly correlated with plot radius for some of the forest parameters. Even tough special attention is paid to ensure that field inventories are not biased, operational constraints and stand characteristics in such areas may explain such patterns, e.g.:

1 need for a small stable platform to install and operate the tripod;
2 minimum distance to big tree trunks to ensure visibility from the tripod and acceptable GPS signal;
3 minimum distance to small trees and understory to unfold the 4 m GPS antenna;
4 presence of compact groups of several stems in coppice stands.

Presence of outliers in field observations is all the more likely since small plots exhibit higher variability of forest parameters. Such data points may affect the coefficient of determination and prediction accuracy of multiple regression models. Besides, this effect is enhanced by GPS positioning errors. Differential GPS allows sub meter accuracy in open areas, but in mountainous forests its precision depends on canopy cover density and on the elevation mask. Indeed, propitious time intervals for acquisition are short and fragmented due to topographic conditions. Real precision is more likely to be around two to four meters. For small radii, there are high chances that the extracted point cloud is only partially located within the actual field plot, resulting in erroneously fitted linear models.

6. Conclusion

With an area-based method, forest stand parameters such as dominant height, basal area, stem density and mean diameter can be precisely estimated from airborne laser scanning data for mountainous, coppice stands. Comparison of regression models obtained with plot radii ranging from five to ten meters show that prediction accuracy depends on calibration plot size. This effect may be due to biases in field data collection or to high spatial variability of mountainous coppice stands.

However, another major factor that may greatly influence prediction models is GPS precision. Indeed, topographic constraints make GPS acquisition quite uncertain in alpine environments. Further investigation is required to quantify the influence of positioning accuracy on prediction results and optimise operational field protocols for calibration of airborne laser scanning models.

Acknowledgements

J.-M. Monnet held a doctoral fellowship from Cluster de Recherche Energies (région Rhône-Alpes, France)

References

AXELSSON, P., 2000, DEM generation from laser scanner data using adaptive TIN models. In *Proceedings of the XIXth ISPRS Congress, IAPRS*, XXXIII, pp. 110–117.

GINISTY, C., VALLET, P., CHABÉ FERRET, S., LEVESQUE, C. and CHAUVIN, C., 2007, Disponibilités en biomasse forestière pour des usages énergétiques et industriels en France. Note de synthèse. Technical report, Cemagref.

HEURICH, M. and THOMA, F., 2008, Estimation of forestry stand parameters using laser scanning data in temperate, structurally rich natural European beech (*Fagus sylvatica*) and Norway spruce (*Picea abies*) forests. *Forestry*, **81**, pp. 645–661.

HOLLAUS, M., DORIGO, W., WAGNER, W., SCHADAUER, K., HÖFLE, B. and MAIER, B., 2009, Operational wide-area stem volume estimation based on airborne laser scanning and national forest inventory data. *International Journal of Remote Sensing*, **30**, pp. 5159–5175.

HYYPPÄ, J., HYYPPÄ, H., LECKIE, D., GOUGEON, F., YU, X. and MALTAMO, M., 2008, Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing*, **29**, pp. 1339–1366.

NÆSSET, E., 2002, Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, **80**, pp. 88–99.

PATENAUDE, G., HILL, R.A., MILNE, R., GAVEAU, D.L.A., BRIGGS, B.B.J. and DAWSON, T.P., 2004, Quantifying forest above ground carbon content using LiDAR remote sensing. *Remote Sensing of Environment*, **93**, pp. 368–380.

POPESCU, S.C., WYNNE, R.H. and NELSON, R.F., 2002, Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size. *Computers and Electronics in Agriculture*, **37**, pp. 71–95.

Model-based variance estimation for aggregated predictions of forest attributes to stand level based on airborne laser scanning data

*Johannes Breidenbach *†, Erik Næsset†, Terje Gobakken†*

Johannes.Breidenbach@umb.no

†Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences, 1432 Ås, Norway

Keywords: Forest inventory, small area estimation, model-based inference, ALS

A large number of studies have proven the usefulness of airborne laser scanning (ALS) data for the estimation of biophysical forest attributes. Both area-based and single tree approaches may be applied and have already reached operational status in several countries (Næsset 2004; Lindberg et al. 2010). The aim of both methods is often a wall-to-wall estimate of the response variables. Therefore, a model is estimated using sample units where both field and remote sensing data are available. Field sample plots are typically used for the model development. The complete area of interest is then tessellated into non-overlapping tiles for which only ALS data are available and the statistical model is subsequently used to produce a prediction for each tile. Since forest stands usually are the basic geographical units of forest management planning, the ALS predictions for the tiles are typically aggregated to stand level. This results in an estimate of the mean (or total) of a certain response variable. Along with the mean, a measure of uncertainty of this estimate is important, for example when choosing between different potential forest management alternatives. A common measure of uncertainty is the standard error of the estimate which can be derived from the estimate's variance.

However, since the tiles are not independent samples, common variance estimators are not applicable. Consequently, most studies have so far only reported on mean stand predictions without associated measures of uncertainty. It should be noted, however, that most studies report on root-mean-squared-errors (RMSE), which is a measure of the performance of the statistical model. Nonetheless, the RMSE does only provide a single number for all samples used for validation and does not give any measure of uncertainty of the estimate provided for each individual target unit (e.g., forest stands).

In statistics, the field of small area estimation (SAE) provides methods for derivation of variances for predictions aggregated to larger units that contain no or only few samples (Rao 2003). Although SAE is common in ALS-supported area-based forest inventories to provide estimates of mean values (Næsset 2002), studies also providing variance estimates using ALS data are yet rare (Andersen and Breidenbach 2007; Breidenbach et al. 2010). Methods for inference exist though even for nonparametric estimation techniques (McRoberts et al. 2007) applicable to ALS-assisted forest inventory. Results of the studies by Andersen and Breidenbach (2007) and Breidenbach et al. (2010) showed smaller model-assisted and model-based variances, respectively, of the ALS predictions at stand level compared with probability-based variances derived solely from field sample plot inventories. This proves the great potential of ALS data as auxiliary information for stand-based predictions. In addition, a model-based approach allows estimation of means and variances for units without any sample plots which also merits for the application of ALS data. In a new study, we compare model-based variance estimators following the SAE approach to derive a measure of uncertainty for predictions aggregated to stand level. First results based on nonparametric regression techniques for a study area in Norway will be presented.

References

Andersen, H. E. and Breidenbach, J. (2007). Statistical properties of mean stand biomass estimators in a lidar-based double sampling forest survey. In: IAPRS Volume XXXVI, Part 3 / W52. ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007, September 12-14, 2007, Espoo, Finland.

Breidenbach, J., Nothdurft, A. and Kändler, G. (2010). Comparison of nearest neighbour approaches for small area estimation of tree species-specific forest inventory attributes in central Europe using airborne laser scanner data. European Journal of Forest Research, (published online), doi: 10.1007/s10342-010-0384-1.

Lindberg, E., Holmgren, J., Olofsson, K., Olsson, H. and Wallerman, J. (2010). Estimation of tree lists from airborne laser scanning by combining single-tree and area-based methods. International Journal of Remote Sensing, 31(5), 1175-1192.

McRoberts, R., Tomppo, E., Finley, A. and Heikkinen, J. (2007). Estimating areal means and variances of forest attributes using the k-Nearest Neighbors technique and satellite imagery. Remote Sensing of Environment, 111(4), 466-480.

Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sensing of Environment, 80(1), 88- 99.

Næsset, E. (2004). Accuracy of forest inventory using airborne laser scanning: evaluating the first Nordic full-scale operational project. Scandinavian Journal of Forest Research, 19(6), 554-557.

Rao, J. (2003). Small area estimation, Wiley-Interscience.

The role of Airborne Laser Scanning in updating and revision of GIS databases – a case study in the Polish State Forest (Chojna District)

PIOTR WEZYK†, MARTA SZOSTAK†, PIOTR TOMPALSKI†, GRZEGORZ ZAJACZKOWSKI††*

rlwezyk@cyf-kr.edu.pl, http://geo.ur.krakow.pl

†Laboratory of GIS&RS, Department of Forest Ecology, Faculty of Forestry, Agricultural University of Krakow; Al. 29 Listopada 46, 31-425 Kraków;

†† Department of Forest Recourses Management, Forest Research Institute, Warsaw, Braci Lesnej 3, 05-090 Sekocin Stary

Keywords: ALS, Digital Forest Map (LMN), descriptive database (SILP), GIS, Polish State Forest Holding.

1. Introduction

Polish forests grow on approx. 9 mln ha (28% of whole country; Scots pine – dominate on 71%), which places Poland in the sixth place in Europe in terms of forest cover. The structure of the forest property in Poland is different than in other EU countries. Over 78.4% (7.5 million ha) belongs to the State Forest National Holding (PG Lasy Panstwowe – *pl.*). For about 15 years a modern information system called SILP (System Informatyczny Lasow Panstwowych - *pl.*; State Forests Informatics System) has been implemented as a standard in every of 431 Forest Districts in Poland. This comprehensive database consists of detailed information about each of 8 million forest compartments and can be joined with geometrical database called LMN (Lesna Mapa Numeryczna - *pl*; Digital Forest Map), which contains polygons (borders of forest compartments with unique ID), lines (e.g. forest roads) and points (e.g. survey station). Together, they create a topologically correctly, geometric layer with many attributes derived by the forest inventory, which can be managed by GIS user. Regarding internal rules of Polish State Forest (IUL 2003), the descriptive (SILP) and geometrical database (LMN) as well, should be updated periodically using information collected during fieldwork (e.g. powered by GPS data). Once every 10 years a management plan should be prepared for each forest district, and should be verified using remote sensing data like aerial or satellite orthophotos and national geo-data as well.

During forest inventory measurements, aerial photos and ortophotomaps are often acquired simultaneously to support changes recognition in the forest (fires, clear cuts etc.). Besides National State Forests also National Parks and areas within European Ecological Network Natura 2000 are subjects of such inventories. Unfortunately,besides pilot projects (Wężyk, Solecki 2008; Wężyk et al. 2008a, 2008b) regarding ALS usage in forest inventory and management, there is still lack of law regulations and executive internal instructions that would allow implementing this GI technology on a large scale in Poland.

Airborne laser scanning is a modern remote sensing technology used for collecting 3D information (point cloud) and provides precise information about terrain elevation and vegetation structure for large areas (Andersen et al., 2006; Holmgren and Jonsson 2004; Hyyppä et al., 2004). Filtration and classification of point clouds leads to creating accurate models like: DTM, DSM or nDSM, representing terrain and objects on it (Axelsson 2000; Wężyk et al. 2008a). ALS data have many applications in forest management but can also be used in ecological research on biomass or carbon sequestration (Gołuch i in. 2009; Wężyk et al. 2008b). Many forest taxation parameters can be retrieve or predicted using ALS technology. Especially the height of trees is important parameter, playing a crucial role during determination of forest stand volume. The height of tree is determined by: age, health status, site fertility, simultaneously affected by management operations and random events occurring in the stand. Measuring all tree heights in the forest was up to now not possible to perform, because the process would be too time consuming. On the other hand, the error an operator

makes during traditional height measurement by hypsometer often exceeds 5% and is caused by subjectivity in pointing the tree top and when the tree is leaned.

The aim of presented study was an elaboration of automatic processing of ALS point cloud data for Digital Forest Map (LMN) revision (geometric errors or lack of information) and an update of attributes stored in the descriptive database (SILP). Such elements like: openings, gaps, clear cuts, bio-groups, dead trees, areas of low canopy closure, and the borders of forest compartments are possible to be mapped and checked using ALS. Some raster and vector GIS spatial analyses based on nDSM or even raw ALS cloud point can give an important feedback to operator of GIS system where the update of geometry is needed and topology of Digital Forest Map should be rebuild. To reach the goal, the comparison of automated method based on ALS data with the manual work of trained operator (on-screen digitizing) was performed.

Another goal to reach was to elaborate a method of automated stand height verification based on ALS data. The height was compared with attributes from SILP descriptive database. A vector shape file containing compartment border was used to specify the region of accuracy analysis.

Study Area

The study area – Piasek management unit in the Chojna Forest Disctrict (RDLP Szczecin; NW Poland; 52°56'45" N; 14°13'18" E) was chosen due to its average size (6380.26ha), regular shape, age of the stands (mean ~59 years), diversity of species, site categories distribution and topographical feature, and, what is most important, updated in year 2005 forest inventory database (Zajączkowski, Wężyk 2007). Study area consisted of 1565 forest compartments of total area equal to 5747ha. Scots pine was the dominant species, covering 75% (1032 compartments) of study area. The 25% of the forest consist of species like: Beech - 7% (111 compartments), Oak – 10% (183 compartments) and Black Alder – 2.31% (85 compartments) and other 6% (154 compartments).

Age diversity of the Scots pine stands in the study area is large (Table 2). The youngest (<20 years) and oldest (100÷120 years) age classes consist of about 6% of pine stands area. The age classes with highest number of pine stands were: II-V (20÷120 years) covered together approx. 73%. The most numerous (approx. 31%) was the III age class (60÷80 years). Stands older than 120 years were in minority (0.5% of area).

2. Methods

ALS mission

In September 2006 (one year after the forest inventory) the ALS data was collected with RIEGL LMS-Q560 (full waveform data) together with airborne images from small format camera (Rolleiflex 6008) with ground resolution of 7cm. The flight was performed 500m above the terrain by Eurocopter in NE-SW directions. The width of single scan was equal to around 600m (60% overlap). Together 44 scans were acquired. DGPS correction was performed with reference stations from SAPOS network. The accuracy was equal to 0.5m (XY) and 0.15m vertically, with the mean point density of 4 points/m^2. Based on first and last echo information DSM and DTM models were generated with 1m spatial resolution.

ALS data analysis

Fusion (USDA Forest Service; McGaughey 2007) and Terrasolid software were used to perform filtering, classification and analysis of the ALS point cloud. ArcGIS 9.3 (ESRI) was used for further raster analysis. To obtain correct DTM, the Axelsson (2000) method was used (Terrascan, Terrasolid Ltd.). DSM was generated using FUSION (USDA Forest Service) with resolution of 1.0m. This two models where then substracted in order to create nDSM representing the tree heights (H_{ALS}).

The reference compartment outlines were based on the existing Digital Forest Map from 2005 year terrain revision. The accuracy of the border was however not sufficient and needed improvement by GIS operator. This was done manually on digital orthophoto and nDSM as well (Tompalski et al. 2009). During this process also outlines of canopy gaps where defined and the topology of vector layer was rebuild.

To define the proper compartment borders in semi-automatic way, nDSM was classified into two categories: below and above 1.0m. This simple categorization resulted in creating forest mask and areas with no tree canopy cover. According to Polish forest law (IUL 2003) areas smaller than 200m^2 are not treated as gaps and therefore were generalized. The automatic process was constructed using model builder application (ESRI; Figure 1).

After gaining accurate and up-to-date vector of compartments, revision of height attribute was performed on the descriptive database, which was the main goal of the study. The SILP database was directly joined with the Digital Forest Map (LMN) through unique compartment ID. The mean height (H_{ALS95}) of the homogenous compartment was determined as 95th percentile of the ALS point cloud within borders of each forest stand.

Using the descriptive forest database SILP, the raster representing compartment height was generated (based on forest inventory data from year 2005). Then this height model (GRID) was compared with nDSM and the database information was automatically updated (model builder ESRI) Two additional fields in the SILP database were created containing: stand height based on ALS data (H_{ALS95}.) and the difference between ALS and inventory data (H_{diff}).

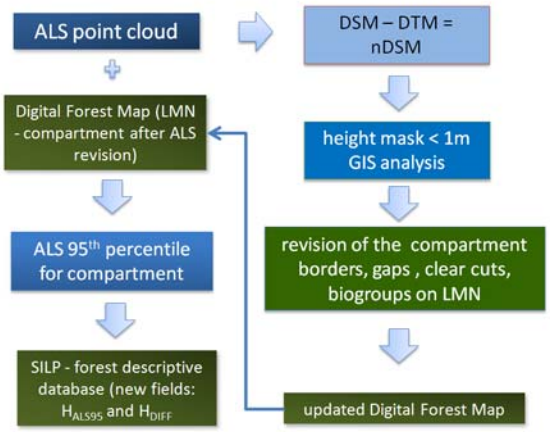


Figure 1. Working flow of the map revision (LMN) and descriptive SILP database update.

3. RESULTS

The manual on-screen delineation of test area resulted in 244 gaps of total area of 68.80ha (mean 0.28ha). Automatic analysis based on nDSM resulted in lower total area (58.58ha) but higher number of gaps (342; +98; mean area equal to 0.17ha). The difference of two described methods was equal to 10.22ha (14.8%). Only 206 from 342 automatically delineated objects, spatially corresponded to the reference gaps (Figure 2, Table 1).

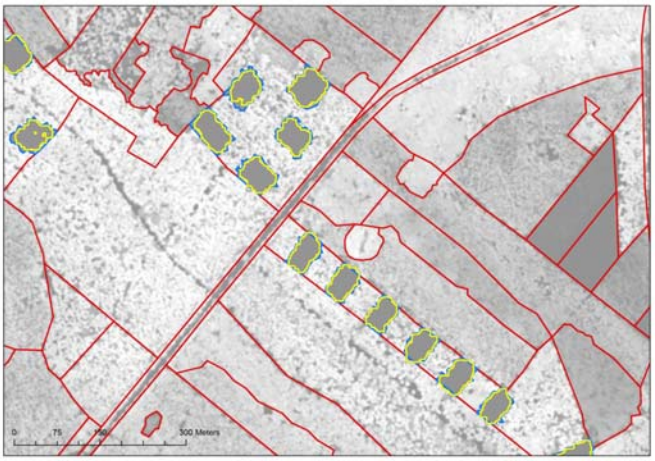


Figure 2. Forest gaps over 200m2 (in blue - based on manual delineation; in yellow - derived automatically on ALS data. Compartment borders showed in red, nDSM in grey as background)

Table 1. Results of the update performed using manual interpretation of nDSM and orthophotomap and automatic ALS data processing

Type of map revision	Area [ha]	No. of gaps	Mean area [ha]	Correct localization
manual orthophoto – reference	68.80	244	0.28	244
automatic nDSM (ALS)	58.58	342	0.17	206
Difference	-10.22	98	-0.11	-38

In order to verify and update the height information of stands stored in SILP database, the 95th percentile derived for each compartment was compared to the SILP database value. Software Statistica 9 was then used to calculate if the differences were statistically significant. Non parametrical Wilcoxon test showed that differences for most of stands categories (Table 2) and for whole area as well, are in most cases significant ($p<0.05$). Only for stands older than 60 years the differences were not significant.

*Table 2. Significance of differences between H_{ALS95} and descriptive database SILP (** $p < 0.05$; * $< 0.01 < p < 0.05$; n - differences not significant)*

Tree species		Significance of the differences
All		**
Scots pine (Pinus silvestris)		**
age class	I	**
	II	**
	III	**
	IV	n
	V	*
	VI	n
	others	n
Beech (Fagus silvatica)		**
Oak (Quercus sp.)		**
Black Alder (Anlus glutionosa)		**

The stand height difference (H_{Diff}) between the ALS method (H_{ALS95}) and the height from forest traditional forest inventory (H_{SILP}) was equal to +0.90m (2.06m for absolute differences). The 935 (60%) from 1565 analyzed forest compartments showed H_{Diff} higher than zero (the SILP values were lower) and the others 627 (40%) below zero (SILP values too high). In the H_{Diff} range between -1m and +1m, the 661 forest compartment (42%) was found. Introducing the compartment area as weight in differences calculation, the results slightly changed to +0.60m (1.52m for absolute differences). This indicates that the height values in the SILP descriptive database are slightly understated compared to the ALS data (Figure 3).

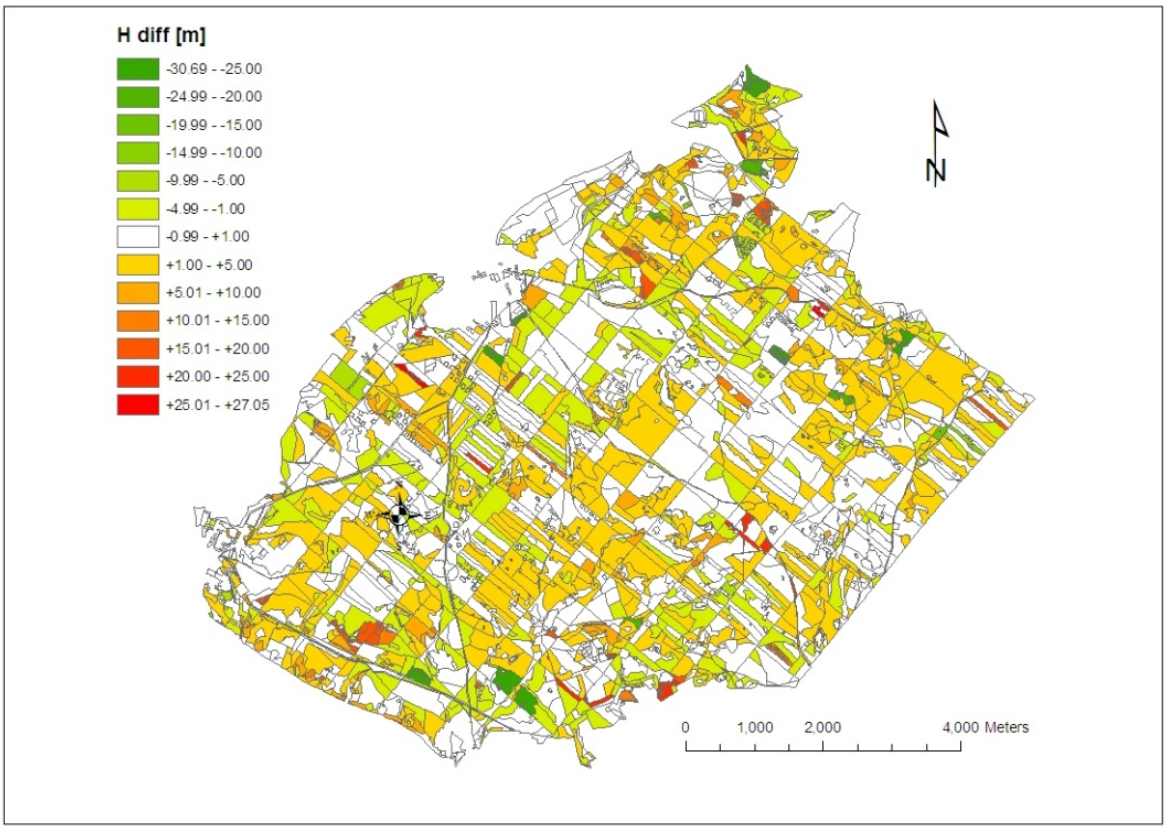


Figure 3. The spatial distribution of the forest compartment height difference (HDiff) in Piasek management unit in Chojna forest district (RDLP Szczecin).

In order to explain in details these differences, additional calculation was made based on main tree species criteria. For Scots pine the difference between mentioned two methods (SILP as reference) was equal to +0.57m (1.76m for absolute differences). Weighted with the pine stands area, the mean differences were equal to +0.41 (1.34m for absolute differences).

Considering the deciduous tree species, the results showed following differences: for Beech +1.63m (2.51m absolute value), Oak +1.67m (+2.89 m) and Black Alder +1.00m (2.31m). Using area of the stand as the weight, the difference (H_{Diff}) resulted in values equal to: Beech +1.37m (2.21m), Oak +1.09m (2.10m) and Black Alder +1.43m (2.20m).

The differences of stand heights were higher for deciduous stands than for coniferous stands, but due to their small total area (20% of study area), their influence was rather limited.

Consecutive analysis lead to define differences in various stand age classes (Table 3). They showed that for classes below 60 years height values in the descriptive database SILP are lower than those derived from ALS data. Mean differences occurred to be highest for I age class (0÷20 years) and were equal to +3.18m (Std. dev. 6.80m). They were smaller considering the II age class (20÷40 years) reaching +1.14m (Std. dev. +2.42m) and III age class (40÷60 years) +0.41 m (Std. dev. 1.58 m). For older stands (>60 years) the differences were of opposite signs which means that they were lower for ALS_{95} method. On average their values were: -0.31 m (Std. dev. 1.67m) for IV age class (60÷80 years); -0.53 m (Std. dev. 2.64m) for V age class (80÷100 years) and -1.51 m (Std. dev. 5.68m) for VI age class (100÷120 years). Similar tendency is described in other publications regarding the accuracy of determining stand height with ALS data (Abraham et al., 2006; Andersen et al. 2006; Hyyppä et al., 1999; Persson et al., 2002; Maltamo et al., 2004; Næsset and Okland 2002; Yu et al., 2004).

Table 3. Differences between height values derived on ALS data (H_{ALS95}) and values from the descriptive database [m].

Tree species		H_{Diff} mean [m]	H_{Diff} absolute value [m]	H_{Diff} area weighted mean [m]	H_{Diff} absolute value of weighted mean (area) [m]
all		0.90	2.06	0.60	1.52
Scots pine		0.57	1.76	0.41	1.34
age class	I	3.18	3.65	1.09	1.67
	II	1.14	1.80	1.00	1.53
	III	0.41	1.27	0.56	1.30
	IV	-0.31	1.35	-0.09	1.09
	V	-0.53	1.38	-0.23	1.24
	VI	-1.51	2.33	-0.47	1.45
Beech		1.63	2.51	1.37	2.21
Oak		1.67	2.89	1.09	2.10
Black alder		1.00	2.31	1.43	2.20

4. CONCLUSIONS

The results of the study have proven the possibility of using Airborne Laser Scanning in automated updating of forest GIS databases - descriptive (height attribute) and geometrical (forest compartment border correction). The usage of automated analysis based on nDSM model can accelerate and replace expensive traditional measurements. It can also provide accurate clear-cut or gap borders. However, till this moment there are no internal regulations existing about defining localization of gaps or biogroups within stands of Polish State Forest.

Additionally the lack of radial shift on nDSM model in comparison to orthophotos usually based on DTM models, places it as the better base for update of forest geometrical layers. The shift of treetops and crowns on orthophotos can reach tens of meters and in such situations it cannot be treated as a reference image.

Automatic height analysis based on ALS data gives the opportunity of objective choice during setting inventory plot localization as it is very rare situation that the compartment is homogeny in height, age or species composition. It is possible in next step to create maps representing spatial distribution of volume.

One of the main problem considering the stand height determination is the height measurement. It is very risky to say that a small number of trees measured in the field with hypsometer during forest inventory are the good representation of the whole stand. The analysis of the ALS point cloud of the whole compartment is unquestionably and statistically more reliable. The results clearly show that the highest differences occur for the youngest stands, which can be explained in their very fast growth and big deviation of single tree height. Moreover, the possibility of correct measurements inside young stands is limited due to dense canopy and lack of tree-top visibility.

The automated analysis of ALS data can result in extraction of another important forest parameter which currently is assessed very subjectively - canopy cover. The integration of descriptive information about stand vertical structure from the database with ALS analysis can lead to new solutions in the future.

Very important especially for forestry in mountainous regions is the possibility of generating accurate terrain and surface models from ALS data, than can be utilized in many forest management aspects.

Periodical monitoring of the stands based on ALS technology and digital photogrammetry guaranties keeping forest GIS - descriptive and geometric databases updated, without the need of constant on-site measurements that are time and cost consuming.

Refences

Abraham, J., Adolt, R. 2006. Stand height estimations using aerial images and laser scanning data. In: International Workshop "3D Remote Sensing in Forestry". Wien, pp. 24-31.

Andersen H. E., Reutebuch S. E., Mcgaughey R. J., 2006. A rigorous assessment of tree height measurements obtained using airborne lidar and conventional field methods. Canadian Journal of Remote Sensing, Vol 32 (5), pp. 355-366.

Axelsson P., 2000. DEM generation from laser scanner data using adaptive TIN models. International Archives of Photogrammetry and Remote Sensing, Vol. XXXIII/4B, Amsterdam., pp. 203-210.

Gołuch P., Borkowski A., Jóźków G., Tymków P., Mokwa M. 2009. Application of Digital Terrain Model generated from Airborne Laser Scanning data in Hydrodynamic Modelling. Studia Geotechnica et Mechanica, Vol. XXXI No. 3, pp. 61-72.

Holmgren, J., Jonsson, T. 2004. Large Scale Airborne Laser Scanning of Forest Resources in Sweden. Proc. of the ISPRS working group VII/2 "Laser-Scanners for Forest and Landscape Assessment" . Freiburg, Germany: International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVI, part 8/W2. pp. 157-160

Hyyppä J., Hyyppä H., Litkey P., Yu X., Haggrén H., Rönnholm P., Pyysalo U., Pitkanen J., Maltamo M. 2004. Algorithms and methods of airborne laser-scanning for forest measurements. In: Thies M., Koch B., Spiecker H. i Weinacker H. (eds.): Laser-Scanners for Forest and Landscape Assessment: Proceedings of the ISPRS Working Group VIII/2. Freiburg, Germany. International Archives of Photogrammetry, Remote Sensing, and the Spatial Information Sciences. XXXVI-8/W2, pp. 82-89

IUL, 2003. Forest Management Guide. Part 3. Polish State Forest Holding, Warsaw 2003

Maltamo M., Mustonen K., Hyyppä J., Pitkänen J., Yu X. 2004. The accuracy of estimating individual tree variables with airborne laser scanning in boreal nature reserve. Absract. Canadian Journal of Forest Research, pp. 1791–1801

McGaughey R. J. 2007. Fusion/ldv: Software for lidar data analysis and visualization. Software manaul. USDA Forest Service. Pacific Northwest Research Station.

Næsset, E., Okland, T. 2002. Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. Remote Sensing of Environment, Vol. 79, pp. 105–115.

Persson, A., Holmgren, J., Söderman, U. 2002. Detecting and measuring individual trees using an airborne laser scanner. Photogrammetric Engineering and Remote Sensing, Vol. 68, No. 9, pp. 925–932

Tompalski P., Wężyk P., de Kok R., Kukawski M., 2009. Determining the number of trees using airborne laser scanning and true orthoimagery. Annals of Geomatics, Vol. VII, 2 (32), pp. 133-141.

Wężyk P., Borowiec N., Szombara S., Wańczyk R. 2008a. Generowanie numerycznych modeli powierzchni oraz terenu w Tatrach na podstawie chmury punktów z lotniczego skaningu laserowego (ALS). Geoinformacja obrazowa w świetle aktualnych potrzeb. Archiwum Fotogrametrii, Kartografii i Teledetekcji. Vol. 18, pp. 651-661.

Wężyk P., Solecki K., 2008. Określanie wysokości drzewostanów nadleśnictwa Chojna w oparciu o lotniczy skaning laserowy (ALS). In: Geoinformacja obrazowa w świetle aktualnych potrzeb. Archiwum Fotogrametrii, Kartografii i Teledetekcji. Vol. 18. pp. 663-672.

Wężyk P., Tompalski P., Szostak M., Glista M., Pierzchalski M. 2008b. Describing the selected canopy layer parameters of the Scots pine stands using ALS data. 8th international conference on LiDAR applications in forest assessment and inventory. Sept. 17-19. 2008 – Edinburgh, UK. ISBN 978-0-85538-774-7 (CD), pp. 636-645

Yu, X., Hyyppä, J., Hyyppä, H., Maltamo, M. 2004. Effects of flight altitude on tree height estimation using airborne laser-scanning. In: Thies M., Koch B., Spiecker H. i Weinacker H. (eds.): Laser-Scanners for Forest and Landscape Assessment: Proceedings of the ISPRS Working Group VIII/2, 3–6 October 2004. Freiburg, Germany: International Archives of Photogrammetry, Remote Sensing, and the Spatial Information Sciences, Vol. XXXVI-8/W2.

Zajączkowski, G., Wężyk, P. 2007. Ocena przydatności skaningu laserowego oraz cyfrowych obrazów multi- i hiperspektralnych do określania miąższości drzewostanów. Sękocin Stary, Raport z tematu BLP 302. Instytut Badawczy Leśnictwa, Sękocin Stary

Fusion of multispectral and LIDAR remote sensing data for the estimation of forest attributes in an Alpine region

MICHELE DALPONTE†, SERGIO TONOLLI‡, LORIS VESCOVO†, MARKUS NETELER†, and DAMIANO GIANELLE†

† IASMA Research and Innovation Center, Fondazione E. Mach, Environmental and Natural Resources Area, Via E. Mach, 1, I-38010 S. Michele all'Adige (TN), Italy

‡ Dipartimento Risorse Forestali e Montane, Provincia Autonoma di Treno, via Trener 3, 38121 Trento, Italy

Abstract

This paper presents an analysis on the integration of airborne LIDAR and satellite multispectral data (IRS 1C LISS) for the prediction of forest stem volume at plot level. A set of variables has been extracted from both LIDAR and multispectral data and some models have been defined considering data source (LIDAR, multispectral and a combination of both) and the species composition of the plot areas. The analyzed data set comprises 799 ground-truth plots within the forested areas of the Trento Province, Italy (about 3 000 km²), in the Italian Alps. This area is characterized by a large heterogeneity in terms of ecological environments, species composition, morphology, and altitude.

Experimental results show that the combination of LIDAR and IRS 1C LISS data for the estimation of forest attributes is effective. The best model developed comprises variables extracted from both these dataset, even if variables derived from LIDAR data provide the most important contribution.

1. Introduction

Remote sensing of biophysical variables is a key step for the quantification of the carbon fluxes of forests. Forests are rather challenging targets, due to their architectural heterogeneity, understory vegetation effect, plant shadows, etc. Nowadays many studies exist that use remote sensing data over forest areas and almost all existing remotely sensed data have been investigated for the estimation of forest parameters. In this study we focused our attention on LIDAR and multispectral data. These data have different characteristics and they provide very different information. LIDAR represent the “best” source of information for the study of forest parameters, providing detailed information on the vertical structure of the canopy allowing the estimation of many different forest parameters: volume, basal area, height, etc. Satellite multispectral data, like IRS 1C LISS data provide a spectral information of the ground cover, allowing a detection of the tree species and the estimation of some species-related parameters.

In the literature, many studies exist on the use of LIDAR data to study forest environment (e.g., Naesset 2009, Neasset and Gobakken 2008, Coops *et al*. 2007, Andersen *et al*. 2005). As an example, Naesset 2009 analyzed the effects of different sensors (Optech ALTM1233 and ALTM3100), flying altitudes (1100, 1200 and 2000 m), and pulse repetition frequencies (PRF; at 33, 50 and 100 kHz) on the estimation of stem volume and mean height at stand level using LIDAR-derived variables. All the datasets acquired in different conditions appear to be suitable for the estimation of volume (the “best” model developed has a R^2 of 0.92) and mean height, with a mean error of up to 10.7% for stem volume and 2.5% for mean height. Coops *et al*. 2007 estimated the canopy structure of a Douglas-fir forest with 1st return LIDAR data and found high correlations between field data and LIDAR derived data (R^2 = 0.85 for the mean height, and R^2 = 0.65 for basal area). Neasset and Gobakken 2008 estimated above ground and below ground biomass with airborne LIDAR data. They used a regression model made up of variables describing both height and coverage of the canopy. The final model explained 88% and 85% of the variance for the above and below ground biomass respectively. Andersen *et al*. 2005 developed some regression models starting from LIDAR-derived variables

for the estimation of crown fuel weight, crown bulk density, canopy base height, and canopy height. They obtained good results for all the estimations with R^2 ranging from 0.77 to 0.98.

Multispectral data are also widely used, especially in the past. These data are much less expensive than LIDAR ones and in some cases they allow one to obtain quite good results. As an example, Muukkonen and Heiskanen 2005 used ASTER data for the estimation of forest biomass in Finland, obtaining predictions significantly close to the municipality-level mean values provided by the National Forest Inventory of Finland. Hall *et al*. 2002 exploited Landsat ETM+ data to estimate forest biomass in Canada. They found that Landsat derived forest biomass was statistically and moderately correlated to the inventory-derived biomass with values of adjusted R^2 of 0.63, 0.68, and 0.70 for conifer, deciduous, and mixed species, respectively.

In the last years, the possibility to have multiple data over forest areas has allowed the researchers to study the fusion of multi sensors data for the estimation of forest parameters. This is a challenging and interesting task. In fact the combination of different data sources can potentially provide a better explanation of forest characteristics and thus to have more precise estimations. Regarding the fusion of LIDAR and multispectral data some studies are available in the literature (e.g., Erdody *et al*. 2010, Hudak *et al*. 2006). Erdody *et al*. 2010 analyzed the fusion of LIDAR and aerial imagery data for the estimation of forest canopy fuels. They developed three separated models considering these data separately and combined. For all the parameters considered (height, canopy base height, canopy bulk density and available canopy fuel) the model that combines LIDAR and imagery data provided the highest correlation (the increase of R^2 respect to the use of LIDAR data alone is between 2 to 5%). Moreover, it emerged that the use of imagery data alone provide good results only on two out of four of the considered parameters. Hudak *et al*. 2006 compared the estimations of basal area and tree density obtained using LIDAR and ALI satellite data along with their combination. As in the previous cited study, the most informative variables emerged to be LIDAR ones. The combination of the two information sources allowed one to slightly increase the estimations accuracy.

From this introduction, it is clear that at the moment many papers exist on the use of LIDAR and multispectral data for the estimation of forest parameters. Despite that in our analysis of the literature we have identified some lacks: i) no papers exist that analyzes the fusion of LIDAR and IRS 1C LISS data for the estimation of stem volume; and ii) no papers exists that applied these kinds of data on the estimation of stem volume over a large and complex area characterized by many ecological environments.

Thus, the goal of this paper is the fusion of LIDAR and IRS 1C LISS multispectral data for the estimation of stem volume in the forested areas of Trento Province. This is an area of about 3 000 km2 characterized by the presence of many different ecological environments (it ranges from Mediterranean to Alpine environments), a complex morphology (it ranges from about 60 to 3 700 m over the sea level) and the presence of many different tree species.

2. Material and method

Data set description

The study area selected is the territory of the Autonomous Province of Trento in Italy in the Alps. The total area of the province is 6 212 km², of which about 50% are forested areas. The morphology is quite complex as the altitude over the sea level ranges from 65 m to 3 764 m, with almost half (49.9%) of the territory that ranges between 1 000 and 2 000 m.

The field data reference were collected through a forest inventory during the growing seasons 2004 and 2005. 799 sample plots were distributed over the forest areas of the Province of Trento. These plots were randomly distributed with the constraint to have an homogeneous distribution over the whole province. In these plots we measured some trees according to the relascopic technique and we estimated the overall volume of each plot area for the trees with a Diameter at Breast Height (DBH) higher than 17.5 cm (i.e. U.S. EPA guidance QA/G-5S 2002).

The stem volume was calculated using species–specific models that need both the height and the diameter. The models were provided by the Forest Service of the Province of Trento.

A summary of the field references, processed as sampling design of forest inventory, is shown in table 1. We divided the plot areas into three sets according to the species present in the areas: i) deciduous; ii) evergreen; and iii) mixed.

Table 1. Descriptive statistics for the stem volume of the sample points collected over the Trento Province used in our experiments

	Plots number	Mean (m^3 ha^{-1})	SD	Max (m^3 ha^{-1})	Min (m^3 ha^{-1})
Training					
All	534	241.9	148.8	874.9	6.9
Deciduous	169	157.3	118.8	630.9	6.9
Evergreen	199	317.3	157.6	874.9	20.2
Mixed	166	237.6	114.7	698.3	19.7
Validation					
All	265	240.6	150.0	749.6	9.3
Deciduous	83	137.7	77.8	383.5	18.4
Evergreen	99	326.7	158.3	749.6	9.3
Mixed	83	240.9	129.6	508.9	10.2

The LIDAR data have been acquired between October 2007 and December 2008. The used sensor was an Optech ALTM 3100C with pulse repletion frequency of 100 kHz and a laser wavelength of 1 064 nm. For the acquisition it was used a PARTENAVIA P68 airplane flying between 1 000 and 1 800 m above ground with a flight speed of 250 km h^{-1}. The mean point density was of about 0.48 per square meter. For each emitted pulse both first and last pulse have been recorded.

As multispectral data we considered a IRS 1C LISS image acquired 18th of July, 2003. This multispectral data comprises four bands: green (520 - 590 nm), red (620 - 680 nm), Near-Infrared (770 - 860 nm), and Medium-Infrared (1 550 - 1 700 nm) The ground spatial resolution is 25 m.

Data Processing

Figure 1 shows a flowchart of the processing steps carried out in this paper. Starting from the raw data we defined some regression models for stem volume estimation for each source data: i) LIDAR models; ii) IRS 1C LISS models; and iii) LIDAR + IRS 1C LISS models.

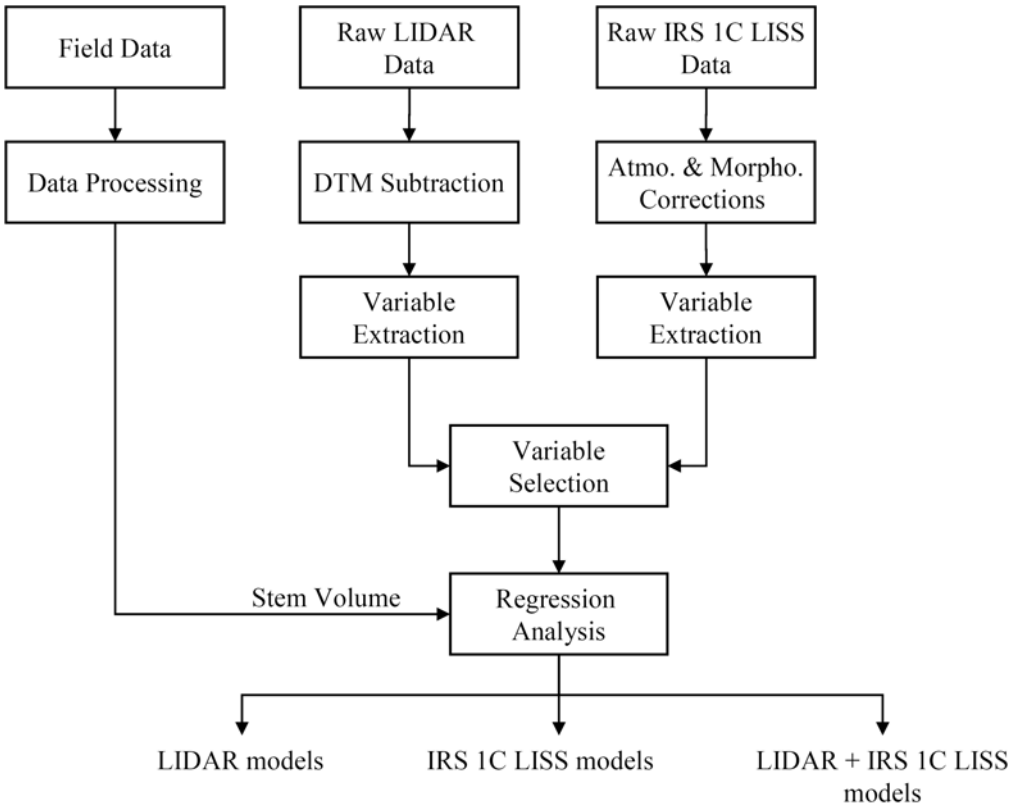


Figure 1. Flowchart for model building methods.

Ground data preprocessing

The stem volumes estimated with the height-DBH relationships were Box-Cox transformed (i.e., Box and Cox 1964) in order to improve the subsequent estimation results. This transformation is used to stabilize the variances of the sample volumes, and to reduce or eliminate the correlation between the means and standard deviations.

LIDAR data preprocessing

The preprocessing that we applied to the LIDAR data consisted in the subtraction of the Digital Terrain Model (DTM) from the elevation of each LIDAR pulse. The DTM used was derived from the original LIDAR acquisition; it has a ground resolution of 1 m. The process of DTM extraction was carried out by the company that acquired the data.

IRS 1C LISS preprocessing

Multispectral data have been atmospherically corrected with the 6S model (Second Simulation of Satellite Signal in the Solar Spectrum) implemented as module in GRASS GIS V6.4 (Neteler and Mitasova 2008). For this purpose, the sensor specifications of IRS 1C LISS had to be implemented first. The correction was performed with the midlatitude summer atmospheric model, the continental aerosols model and an estimated visibility of 60 km.

LIDAR variables extraction

A circle of 30 m radius was placed on the centre of each plot area and LIDAR pulses falling inside it were selected. From these pulses a series of variables have been extracted from each plot. In particular we extracted variables belonging to three categories: i) height; ii) coverage; and iii) other variables. Table 2

shows a summary of the variables extracted. In the computation of these variables we considered only LIDAR pulses with an elevation higher than 2 m (*lhc*).

Height variables describe the height distribution of laser pulses. These variables are widely used in the previous literature (e.g., Naesset and Økland 2002, Andersen *et al.* 2005, Naesset and Gobakken 2008) and they showed to be effective for tree height characterization. Coverage variables are related to the description of the canopy density and coverage. Also these variables have been derived from the literature (e.g., Naesset *et al.* 2005). The variable "*N_LIDAR*" represents the number of dominant trees in the considered plot. It has been extracted according to a tree top identification algorithm (see Hyyppä *et al.* 2001 for more details).

Table 2. Variables extracted from LIDAR pulses in each plot area

Variable ID	Variable description
Height	
H_{mean}	mean value of *lhc*
H_{max}	max value of *lhc*
H_{CV}	coefficient of variation of *lhc*
H_{q20}	the 20th quintile of *lhc*
H_{q50}	the 50th quintile of *lhc*
H_{q90}	the 90th quintile of *lhc*
$H_{q90}^{2.5}$	the 90th quintile of *lhc* at the power of 2.5
H_{q95}	the 95th quintile of *lhc*
Coverage	
C_{2m}	the canopy density as C_{lhc}/C_{lh} where C_{lhc} is the number of *lhc* values and C_{lh} is the total number of pulses
C_{mean}	the ratio C_{lhm}/C_{lh} where C_{lhm} is the number of canopy hits with *lhc* values majors of H_{mean}
C_{q20}	the ratio C_{lh20}/C_{lh} where C_{lh20} is the number of canopy hits with *lhc* values majors of H_{q90}
C_{q50}	the ratio C_{lh50}/C_{lh} where C_{lh50} is the number of canopy hits with *lhc* values majors of H_{q90}
C_{q90}	the ratio C_{lh90}/C_{lh} where C_{lh90} is the number of canopy hits with *lhc* values majors of H_{q90}
Other variables	
N_LIDAR	Number of trees extracted from LIDAR data

IRS 1C LISS variables extraction

Starting from the central point of each plot we defined a circle of 30 m radius and we considered the pixels of the multispectral image falling inside this circle. We use the average value of these pixels for the extraction of some variables from each plot. In table 3 these variables are summarized. We considered the four original bands and some vegetation indexes derived from the existing literature.

Table 3. Variables extracted from ISR 1C LISS data in each sample plot.

Variable ID	Variable description
B1N	Green band normalized between 0 and 1
B2N	Red band normalized between 0 and 1
B3N	NIR band normalized between 0 and 1
B4N	SWIR band normalized between 0 and 1
RR	Band ratio (B4N/B2N)
NDWI	Normalized Difference Water Index (Lymburner *et al*. 2000)
GNDVIgreen	Green Normalised Difference Vegetation Index (Gitelson *et al*. 1996)
SRc	Corrected Simple Ratio (Brown *et al.* 2000)
NDVIc	Corrected NDVI (Nemani *et al.* 1993)
SLAVI	Specific Leaf Area Vegetation Index (Lymburner *et al*. 2000)
NCI	Normalized Canopy Index (Vescovo and Gianelle, 2008)

3. Results

Table 4 shows the models obtained with the three data sources. The variable selection has been performed using a stepwise selection. No predictor variable was left in the model with a significance value of the F statistic greater than 0.05. The standard least squared method was used for the model definition.

In the sets considering LIDAR or LIDAR + IRS 1C LISS datasets the first variable selected in all the models ("all" and three main vegetation classes) is a variable related to the canopy height. In particular, the H_{med} variable is the first variable selected in six models out of eight, while in the remaining models H_{q20} has been selected. Considering only LIDAR variables in six out of eight models the second variable selected is a variable that describes the canopy coverage (C_{2m}, C_{med} and C_{q50}), while in the remaining two models the second variable selected is the number of trees extracted from LIDAR data. In the models derived from the combination of both the analyzed remote sensing data, the multispectral variables are selected as third position for "All" and "Deciduous", while as forth position for "Evergreen". No multispectral variable has been selected for the "Mixed". This is interesting as seems that for the areas where there is the mixed presence of deciduous and evergreen species LIDAR data are enough for explaining volume variability. In particular this model is the one that comprises the lowest number of variables (only two). These two variables has different physical meanings: height and coverage. In these models the vegetation indexes were not considered significant for the volume estimation.

Among the multispectral models in three out of four the first variable selected is the Medium-Infrared band (in the deciduous model it was selected at the second step). Only three of the vegetation indexes considered were used in the models (i.e., RR, NDWI, GNDVIgreen).

*Table 4. Regression coefficients for the models obtained with the three variable set (LIDAR, IRS 1C LISS and LIDAR + IRS 1C LISS) and the four species sets (all species, deciduous, evergreen and mixed). Variable selection order: i) **first**; ii) second; iii) third; and iv) fourth.*

	All		Deciduous		Evergreen		Mixed	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
LIDAR								
Intercept	0.0	1.0	1.2	1.1	0.1	2.2	7.9	1.4
C_{2m}	*17.4*	1.4						
C_{q50}							*37.7*	4.9
C_{q90}					104.5	30.7		
C_{med}			*32.5*	4.5				
H_{q20}							**0.9**	0.1
$H_{q90}^{2.5}$	0.0	0.0						
H_{med}	**1.5**	0.1	**1.2**	0.1	**1.2**	0.1		
N_LIDAR					*0.0*	0.0		
IRS 1C LISS								
Intercept	40.7	1.6	39.4	3.9	56.2	3.6	17.1	8.0
B1N					*-76.7*	28.7		
B2N			**-99.8**	50.1			127.8	64.6
B4N	**-70.8**	4.5	*-41.7*	12.6	**-40.3**	11.5	**-87.0**	21.3
RR	*2.6*	0.6					*7.4*	2.4
NDWI	9.3	4.6						
GNDVIgreen			23.4	9.1				
LIDAR + IRS 1C LISS								
Intercept	16.6	2.6	11.1	2.8	10.1	4.5	7.9	1.4
C_{2m}	*15.9*	1.4						
C_{q50}							*37.7*	4.9
C_{q90}					84.5	31.3		
C_{med}			*28.3*	4.4				
H_{q20}							**0.9**	0.1
H_{med}	**0.9**	0.1	**1.0**	0.1	**1.1**	0.1		
H_{cv}	-0.1	0.0						
N_LIDAR					*0.0*	0.0		
B1N	-39.1	8.4			-41.6	16.2		
B2N			-79.2	20.7				

Table 5 shows the results in terms of adjusted R^2 and RMSE% obtained on the training sets with the 12 models considered. It is worth noting that the results are quite good and they are in line with previous works in the literature. The models that provide the best performances are models that includes LIDAR variables,

and among them the ones that contains deciduous species are the most effective. Additionally, the model that comprises all the species provide good correlations (Adj-R^2=0.7). Multispectral data alone does not seem effective in volume estimation with an adjusted R^2 that ranges from 0.21 to 0.40. The combined used of these data allow one to have a slight increase in the estimation performances, especially for the deciduous species.

Table 5. Results in terms of Adj-R^2 and RMSE% obtained on the training set.

	LIDAR variables		IRS 1C LISS		LIDAR + IRS 1C LISS	
	Adj-R^2	RMSE (%)	Adj-R^2	RMSE (%)	Adj-R^2	RMSE (%)
All	0.69	19.3	0.40	26.7	0.70	18.9
Deciduous	0.68	23.7	0.35	33.9	0.71	22.7
Evergreen	0.60	17.1	0.28	23.1	0.61	16.8
Mixed	0.57	17.6	0.21	23.6	0.57	17.6

Table 6 shows the results in terms of RMSE% obtained on the validation sets with the 12 models considered. The performances are comparable to that obtained on the training set. This is important as it shows that the models developed have a good generalization ability and they are affective for the estimation of volume over all the Autonomous Province of Trento.

Table 6. Results in terms of RMSE% obtained on the validation set.

	LIDAR	IRS 1C LISS	LIDAR + IRS 1C LISS
	RMSE (%)	RMSE (%)	RMSE (%)
All	20.0	30.2	20.1
Deciduous	25.2	34.5	25.7
Evergreen	17.2	22.7	16.8
Mixed	20.3	29.4	20.3

Figure 2 shows the observed versus the predicted volumes (Box-Cox transformed) for the twelve models extracted. This figure confirms the results of Table 5. The “best” models are the ones derived from the

combination of LIDAR and IRS 1C LISS data, while the models derived only from multispectral variables do not provide good results.

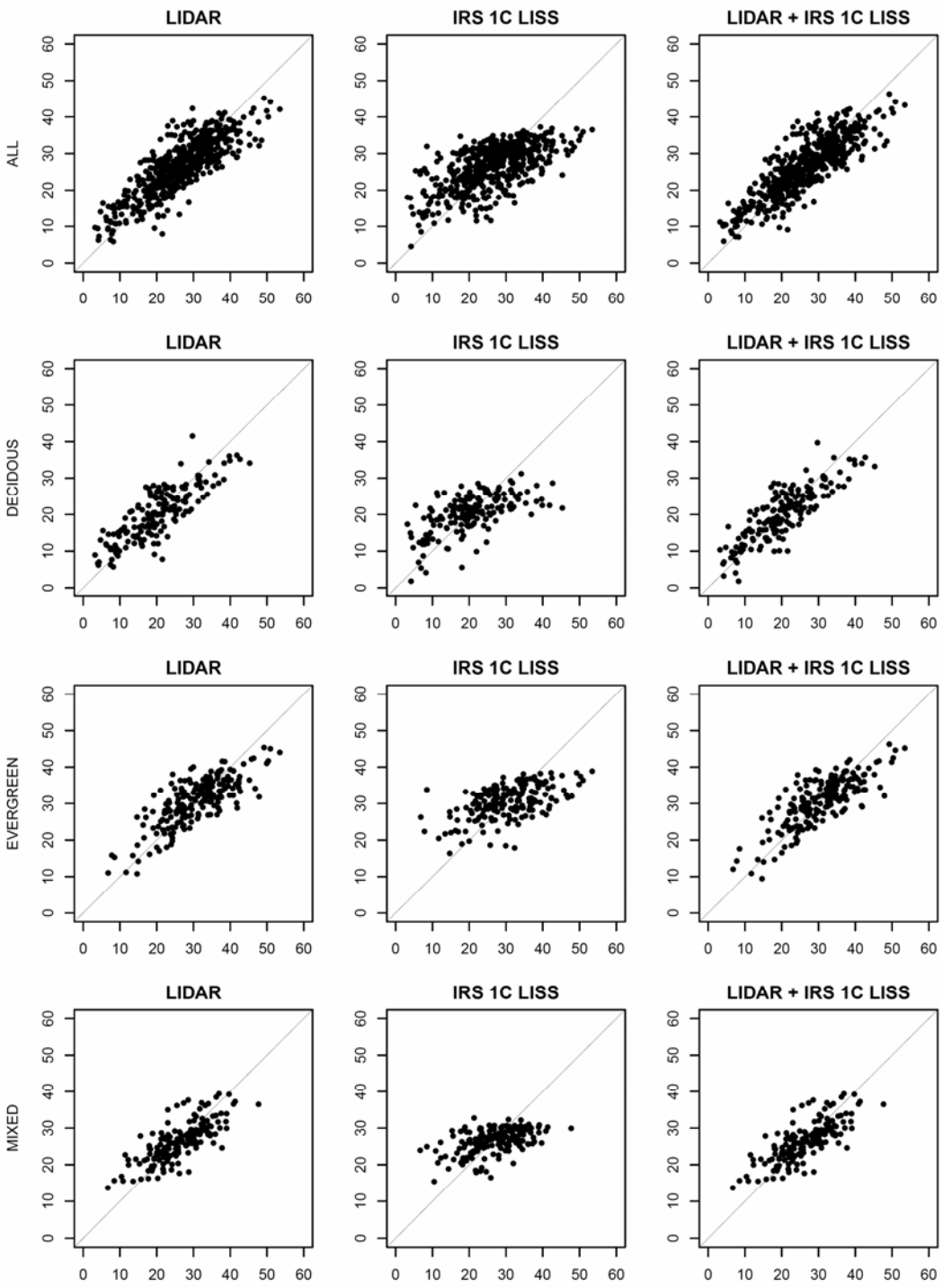


Figure 2. Observed versus predicted volumes (Box-Cox transformed) for the twelve models considered. Solid line show 1:1 relationship.

4. Conclusions

In this paper, an analysis on the fusion of LIDAR and multispectral data for the estimation of forest stem volume has been presented. Our analysis pointed out that: i) the combination of LIDAR and multispectral data can be useful even if it provides a slight increase of estimation accuracy (about 2%), especially for deciduous areas; ii) models derived only from multispectral data are not effective for the estimation of stem

volumes in an area like the Trento Province; iii) LIDAR variables provide the majority of the explanative contribution; and iv) the presented models can be effectively used for the estimation of stem volume over the whole Trento Province.

As future developments of this analysis we plan to consider other target variables (such as LAI, basal area) and to carry out an analysis on the single tree species.

References

ANDERSEN, H., E., MCGAUGHEYB, R., J., and REUTEBUCH, S., E., 2005, Estimating forest canopy fuel parameters using LIDAR data. Remote Sensing of Environment, 94, pp. 441-449.

BOX, G. E. P. AND COX, D. R., 1964, An Analysis of Transformations, Journal of the Royal Statistical Society, pp. 211-243.

BROWN, L., CHEN, J.M., LEBLANC, S.G., and CIHLAR, J., 2000, A shortwave infrared modification to the simple ratio for LAI retrieval in boreal forests: and image and model analysis. Remote Sensing of Environment, 71, pp. 16-25.

COOPS, N. C., HILKER, T., WULDER, M. A., ST-ONGE, B., NEWNHAM, G, SIGGINS, A., and TROFYMOW, J.A. , 2007, Estimating canopy structure of Douglas-fir forest stands from discrete-return LiDAR. Trees, 21, pp. 295-310.

DIEDERSHAGEN, O., KOCH, B., AND WEINACKER, H., 2004, Automatic estimation of forest inventory parameters based on LIDAR, multi-spectral and FOGIS data. International Archives Of Photogrammetry, Remote Sensing And Spatial Information Sciences, XXXVI, pp. 208-212.

ERDODY, T. L., and MOSKAL, L. M., 2010, Fusion of LiDAR and imagery for estimating forest canopy fuels. Remote Sensing of Environment, 114, pp. 725-737.

GITELSON, A.A., KAUFMAN, Y.J., and MERZLYAK, M.N., 1996, Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS. Remote Sensing of Environment, 58, pp. 289-298.

HYYPPÄ, J., KELLE, O., LEHIKOINEN, M., and INKINEN, M., 2001, A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. IEEE Transactions on Geoscience and Remote Sensing, 39, pp. 969–975.

HUDAK, A. T., CROOKSTON, N. L., EVANS, J. S. FALKOWSKI, M. J., SMITH, A. M. S., GESSLER, P. E., and P. MORGAN, 2006, Regression modeling and mapping of coniferous forest basal area and tree density from discrete-return lidar and multispectral satellite data. Canadian Journal of Remote Sensing, 32, pp. 126-138.

LYMBURNER, L., BEGGS, P.J., and JACOBSON, C.R., 2000, Estimation of canopy-average surface-specific leaf area using Landsat TM data. Photogrammetric Engineering and Remote Sensing, 66, pp. 183-191.

MUUKKONEN, P., HEISKANEN, J., 2005. Estimating biomass for boreal forests using ASTER satellite data combined with standwise forest inventory data. Remote Sensing of Environment, 99, pp. 434-447.

NAESSET, E., and ØKLAND, T., 2002, Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. Remote Sensing of Environment, 79, pp. 105-115.

NAESSET, E., BOLLANDSÅS, O. M., and GOBAKKEN, T., 2005, Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data. Remote Sensing of Environment, 94, pp. 541-553.

NAESSET, E., and 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, pp. 3079-3090.

NAESSET, E., 2009, Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. Remote Sensing of Environment, 113, pp. 148-159.

NEMANI, R., PIERCE, L., and RUNNING, S., 1993, Forest ecosystem process at the watershed scale: Sensitivity to remotely-sensed leaf area index estimates. International Journal of Remote Sensing, 14, pp. 2519–2539.

NETELER, M. and MITASOVA, H., 2008, Open Source GIS: A GRASS GIS Approach. Third edition. Springer, New York.

U.S. ENVIRONMENTAL PROTECTION AGENCY, 2002, Guidance on choosing a sampling design for environmental data collection. EPA QA/G-5S. Washington, D.C.

VESCOVO, L., and GIANELLE, D., 2008, Using the MIR bands in vegetation indices for the estimation of grasslands biophysical parameters from satellite remote sensing in the Alps region of Trentino (Italy). Advances in Space Research, 41, pp. 1764-1772.

Combining optical satellite data and airborne laser scanner data for vegetation classification

KARIN NORDKVIST†, ANN-HELEN GRANHOLM†, MATS NILSSON†, and JOHAN HOLMGREN †*

karin.nordkvist@srh.slu.se

† Swedish University of Agricultural Sciences, Department of Forest Resource Management, SE-90183 Umeå, Sweden

Abstract

Optical satellite imagery is often used for large area mapping of land cover, but is usually limited to providing two-dimensional data. Airborne laser scanning (ALS) provides three-dimensional data about vegetation cover, such as canopy height and structure, but lacks multispectral information useful for species discrimination. The aim of this study is to investigate if the accuracy of vegetation classification could be improved by combining optical satellite data and ALS data compared to using only one of the data sources. Data used in the study are SPOT 5 HRG XS scenes from 2009 and TopEye Mk II ALS data from 2008 covering the test area Remningstorp in southern Sweden. SPOT HRG XS data are used, both separately and in combination with ALS height percentiles and percentage of canopy hits, in a maximum likelihood classification. Interpretation of tree heights, tree species composition and canopy cover in DMC digital air photo stereo models is used as training data. Using the 70th percentile from ALS canopy hits together with SPOT data gave an overall accuracy of 78% for 6 forest classes, compared to 67% using only satellite data.

1. Introduction

The combined use of satellite imagery and airborne laser scanner (ALS) data in vegetation mapping is promising, since it makes use of both the spectral information in the optical satellite image, and the three-dimensional information in the ALS data. The EU Habitats Directive has set new and high standards for how protected sites and valuable habitats should be selected and monitored (EC 2000, 2003). In Sweden, with large areas of relatively unexploited natural habitats (Sohlman 2008), this requires efficient methods for identifying valuable sites over large areas at low costs. The currently available vegetation maps, produced by aerial photo interpretation, cover less than 50% of the land area in Sweden, and it is questioned if a national coverage with entirely manual methods is feasible.

Automated satellite image classification is an established method for producing large area land cover maps and estimations of forest variables, for example in the Swedish national version of the CORINE land cover data base, the nation-wide forest data bases kNN 2000 and kNN 2005 (Reese *et al.* 2003, Engberg 2005), and a state-wide land cover mapping of Wisconsin (Reese *et al.* 2002), just to mention a few. The accuracy obtained for such products based on 2-dimensional optical data only is however limited. ALS data has proved useful in vegetation mapping studies (Genc *et al.* 2004, Korpela *et al.* 2009). Several studies have shown the benefits of combining ALS data with different kinds of imagery in for example estimations of forest variables (Hudak *et al.* 2002, Hill and Thomson 2005, Hyde *et al.* 2006, Holmgren *et al.* 2008, Erdody and Moskal 2010, Ke *et al.* 2010), and vegetation or habitat mapping in rangeland and coastal zones (Lee and Shan 2003, Bork and Su 2007, Chust *et al.* 2008).

Beginning in 2009, the Swedish National Land Survey (NLS) is collecting laser scanner data for the whole country. Although the main purpose is the production of a new, national DEM, the nation-wide coverage of laser data might also be a resource for future vegetation mapping. SPOT images are freely available from the Saccess database, which is updated yearly since 2007 with a dataset covering the entire country during the vegetation period (http://saccess.lantmateriet.se/map_viewer?map=29&maplevelindex=0).

The main objective of our study is to investigate the usefulness of combining ALS data and SPOT images in a vegetation classification. A SPOT scene from August 2009, covering the test area Remningstorp in southern Sweden, is used together with TopEye Mk II ALS data from September 2008. Training data consists of interpretation of tree heights, tree species composition and canopy cover in DMC digital air photo stereo models. Focus is on mapping forest vegetation according to the classification system of the Swedish national version of CORINE (Engberg 2005).

2. Material and methods

2.1 Study area

The Remningstorp test area is located in southern Sweden (lat. 58° 30' N, long 13° 40' E). It is mainly covered by managed, hemiboreal forests, dominated by Scots pine (*Pinus Sylvestris*), Norway spruce (*Picea Abies*) and birch (*Betula spp*). Figure 1 shows an orthophoto and a SPOT 5 image of the 1.0 km × 1.5 km test site used in this study.

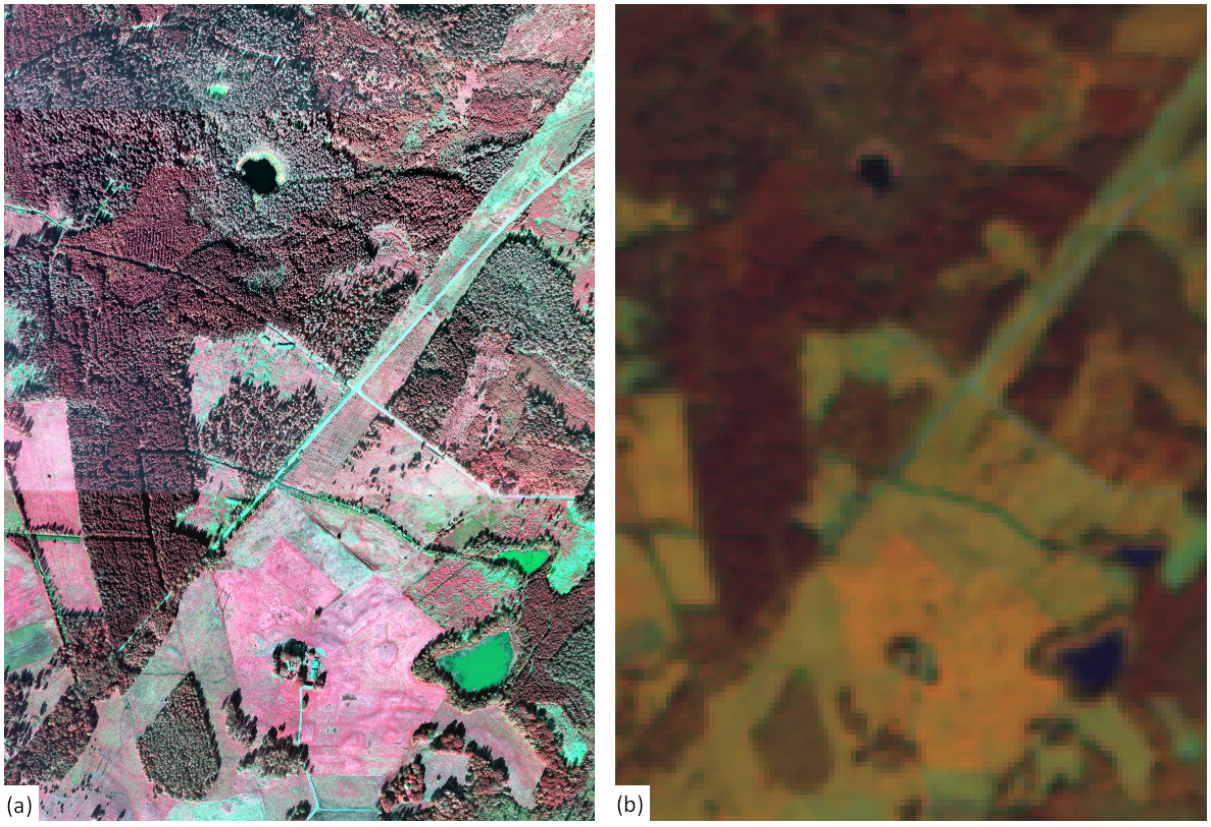


Figure 1. (a) Ortophoto and (b) SPOT 5 image of the test site at Remningstorp used in this study.

2.2 Remote sensing data

Optical satellite data used in the study was a SPOT 5 HRG XS scene from August 20, 2009. The pixel size is 10 m × 10 m for the green, red and near infrared bands, and 20 m × 20 m for the shortwave infrared band. The image had been geometrically precision corrected to the Swedish grid system SWEREF 99, with an error of less than 0.5 pixels. A laser scanning of the test area was done on September 4, 2008, using a

TopEye MkII system carried by a helicopter. The wavelength was 1024 nm, the flying altitude 250 m above ground and the average point density 26 m^{-2}. First and last returns were recorded for each pulse.

2.3 Vegetation reference data

Ground truth samples were collected by aerial photo interpretation in color infrared DMC images acquired at 1200 meters above average ground level in September 2009. 999 circular sample plots with 10 meter radius were distributed in a regular grid with 40 meter spacing. The following data were registered per plot: mean basal area weighted tree height (m), tree species composition (percentage of canopy cover), diffuse canopy cover (percentage) and vegetation class. Canopy cover is defined as the area of the ground covered by a vertical projection of the canopy (Jennings *et al.* 1999). Diffuse canopy cover means the total vertical projection of the tree crowns on the ground, including any gaps within the crowns (Allard *et al.* 2003). The classification scheme has 6 classes and is based on the GSD Marktäcke data base, which is the refined Swedish national version of the CORINE land cover data base (Engberg 2005). Class definitions are given in table 1. Only plots that fall within the forest and clear cut masks in the Swedish terrain map were used. Plots that contained two or more classes were excluded, which left a set of 519 points.

Table 1. Classification scheme.

Class	Canopy cover (%)	Tree height (%)	Species composition (%)		N° of sample plots	Proportion (%)
			Coniferous	Deciduous		
Clear cut†	0-100	<3	0-100	0-100	57	11.0
Young	0-100	3-5	0-100	0-100	26	5.0
Conif. 5-15	≥30	>5, ≤15	≥70	<30	74	14.3
Conif. >15	≥30	>15	≥70	<30	234	45.1
Decid.	≥30	>5	<30	≥70	69	13.3
Mixed	≥30	>5	<70	<70	59	11.4

†The plot should show traces of felling, e.g. stumps, machine tracks and debris.

2.4 Processing of remote sensing data

Height distributions of laser returns were used to create a raster covering the entire laser scanned area. First, the laser returns were classified as ground or vegetation returns using a progressive triangular irregular network densification (TIN) method (Axelsson 1999, 2000) implemented in the TerraScan software (Soininen 2004). A digital elevation model (DEM) was estimated by linear TIN interpolation with the laser returns classified as ground hits. The height value (dz) of a laser return was computed as the difference between the z-value of the laser return and the z-value of the DEM. A height threshold of 10% of the maximum laser height and ≥ 1.0 m was applied in order to separate canopy returns from returns of ground, stones, and low vegetation. Several variables were extracted from laser data within each raster cell, based on the *dz* distribution of laser returns above the height threshold, and used to calculate raster cell values, one band for each derived variable. The variables were 10th percentile (h10), 20th percentile (h20),.... , 90th percentile (h90), 95th percentile (h95), and 100th percentile (h100). A vegetation-ratio (V_r) was calculated as the ratio between number of laser returns above the height threshold and total number of returns.

The ALS grid cells coincided with the pixels in the SPOT scenes, and each ground truth sample plot was centered in the corner between four adjacent pixels (figure 2). A training data set was generated by extracting the mean value of these four pixels or grid cells, which corresponds to a resampling of the data to 20 m × 20 m pixels.

Classification was done with the Minitab 15 statistical software, using the maximum likelihood method. The accuracy was estimated through cross validation by excluding one plot at the time. A first classification was made using the four SPOT bands, then different combinations of height percentiles and/or vegetation ratio from the ALS data were used as additional ”bands”. Classifications with only ALS data were also made in order to better understand the influence of these data on the result. Care was taken not to combine height percentiles that are too close and thus strongly correlated. An additional approach which was tested, was to do the classification in several steps, using the ALS data to create height classes that were then treated separately and classified according to the scheme.

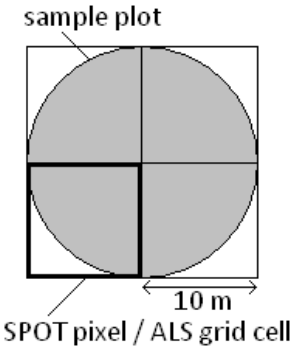


Figure 2. Schematic sketch of a circular sample plot (shaded), centered over four SPOT pixels/ALS grid cells.

3. Results

Over-all classification accuracy for different band combinations is shown in table 2, together with producer’s accuracies for each class.

Table 2. Over-all accuracy and classwise producer’s accuracies (%) for different band combinations.

Bands†	Clear cut	Young	Conif. 5-15	Conif. >15	Decid.	Mixed	Over-all
S1-4	84.2	50.0	51.4	83.8	53.6	27.1	67.1
S1-4, h70	89.5	61.5	74.3	89.3	72.5	37.3	77.6
S1-4, V_r	82.5	61.5	41.9	86.3	58.0	27.1	67.8
S1-4, h70, V_r	89.5	69.2	70.3	88.0	69.6	37.3	76.5
h30, h70	17.5	84.6	29.7	75.6	31.9	11.9	50.1
h30, h70, V_r	61.4	69.2	58.1	75.2	26.1	27.1	59.0

† S1-4 are the four SPOT bands, h10, h20, are height percentiles 10, 20 and so on, and V_r is the vegetation ratio

Classification using the four SPOT bands gave an overall accuracy of 67%. The highest overall accuracy, 78%, was obtained by combining the SPOT bands with percentile 70 from the laser data. The clear cut class had a high accuracy in both cases. Adding h70 resulted in a better separation between the classes coniferous 5-15 m and coniferous >15 m. The result for young forest was also improved, with fewer plots being classified as deciduous and vice versa. The accuracy for mixed forest was clearly improved, with less confusion with the deciduous class, but was still low. In the case S1-4,h70, 27% of the deciduous and mixed plots are classified as either coniferous 5-15 m or coniferous >15 m, compared to only 9% of the coniferous being classified as deciduous or mixed. S1-4,h70,V_r gave results similar to S1-4,h70. When using h30,h70, 18% of the clear cut pixels were correctly identified. The number increased to 61% when V_r was added.

Table 3. Error matrix for classification using S1-4.

Classification data	Training set data						User's accuracy (%)
	Clear cut	Young	Conif. 5-15	Conif. >15	Decid.	Mixed	
Clear cut	48	3	0	0	1	0	92.3
Young	7	13	2	2	10	2	36.1
Conif. 5-15	0	4	38	21	1	9	52.1
Conif. >15	1	1	22	196	3	18	81.3
Decid.	1	4	8	5	37	14	53.6
Mixed	0	1	4	10	17	16	33.3
Producer's accuracy (%)	84.2	50.0	51.3	83.8	53.6	27.1	

Tables 3-5 show error matrices for the cases S1-4, S1-4,h70 and h70,h30,V_r. In the first two cases, the clear cut and coniferous >15 m classes had the highest accuracies, while mixed forest had the lowest. Preliminary results (not shown) indicate that a classification made in several steps can increase the accuracy further.

Table 4. Error matrix for classification using S1-4 and h70.

Classification data	Training set data						User's accuracy (%)
	Clear cut	Young	Conif. 5-15	Conif. >15	Decid.	Mixed	
Clear cut	51	3	0	0	0	0	94.4
Young	4	16	5	0	3	2	53.3
Conif. 5-15	1	5	55	7	3	14	64.7
Conif. >15	0	1	4	209	2	15	90.5
Decid.	1	0	3	7	50	6	74.6
Mixed	0	1	7	11	11	22	42.3
Producer's accuracy (%)	89.5	61.5	74.3	89.3	72.5	37.3	

Table 5. Error matrix for classification using h30, h70 and V_r.

Classification data	Training set data						User's accuracy (%)
	Clear cut	Young	Conif. 5-15	Conif. >15	Decid.	Mixed	
Clear cut	35	2	1	8	2	1	71.4
Young	12	18	7	0	4	0	43.9
Conif. 5-15	0	5	43	3	14	24	48.3
Conif. >15	8	0	1	176	16	15	81.5
Decid.	0	0	10	25	18	3	32.1
Mixed	2	1	12	22	15	16	23.5
Producer's accuracy (%)	61.4	69.2	58.1	75.2	26.1	27.1	

4. Discussion

The main objective of our study was to investigate how a vegetation classification using optical satellite data can be improved by integrating ALS data. It was found that the combined use of these two data sources increases the classification accuracy, compared to using only satellite data. The laser feature giving the greatest improvement of the result was h70, probably because it is more strongly correlated with the mean

height than for example h100. The advantage of ALS data can be seen in the case of young forest, which often contains lots of birch before the pre-commercial thinning has been done. Several plots from this class were mistakenly classified as deciduous forest when only satellite data were used. When h70 was added, the young forest class with its low mean height was easily separated from the higher deciduous class. As could be expected, also the separation of coniferous forest into two height classes worked better when height information was used in the classification.

One might expect clear cuts to be easily identified when ALS data are used alone, as they are characterized by tree heights lower than 3 m. However, this was not the case – the accuracy was below 18% for classification using h30,h70. This might be explained by the fact that some retention trees are left on the clear cuts. A grid cell containing such a tree will get high values on the percentiles, as virtually all of the vegetation hits originate from this tree. It is also possible that clear felling of some areas occurred after the laser scanning but before the acquisition of aerial photos and satellite images. The vegetation ratio will remain low even if one or a few trees are present on the clear cut, and the accuracy increased from 18% to 61% when this feature was added.

Among the error sources can be mentioned the aerial photo interpretation, since it is based on subjective estimations and the performance of the interpreter is likely to vary. This will affect the quality of the training data, and thus the classification. As an attempt to improve the accuracy in the estimation of canopy cover, the interpreter used a calibration software (Gallegos 2005). The maximum likelihood classification assumes normally distributed training data for each class, which is not always the case in our data set. The distance between sample plots is small, which may lead to auto-correlation in the training data set.

Further studies will be performed using data from the national laser scanning (approx. 0.5 pulses/m^2) to evaluate the potential of this method for forest mapping on a national scale. Mire classes, that were to rare in the Remningstorp test site, will then be included. One way to improve separation of the mixed forest class could be to use two satellite scenes over the same area, acquired under leaf on and leaf off conditions, respectively. Other classification methods, such as decision tree, will be tested. A single SPOT scene measuring 60 km × 60 km does not contain enough NFI plots to be used as training data, and methods for merging several scenes should be investigated.

Acknowledgements

This work was financed by the Swedish National Space Board and the Environmental Protection Agency. We thank the Swedish National Land Survey for acquiring and processing the DMC images, and Håkan Olsson for valuable comments on the manuscript.

References

ALLARD, A., NILSSON, B., PRAMBORG, K., STÅHL, G., and SUNDQUIST, S., 2003, Manual for aerial photo interpretation in the National Inventory of Landscapes in Sweden. Swedish University of Agricultural Sciences, Umeå, Sweden.

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 TIN models. International Archives of Photogrammetry and Remote Sensing, 33:110-117.

BORK, E. W. and SU, J. G., 2007, Integrating LIDAR data and multispectral imagery for enhanced classification of rangeland vegetation: A meta analysis. Remote Sensing of Environment, 111:11-24.

CHUST, G., GALPARSORO, I., BORJA, A., FRANCO, J., and URIARTE, A., 2008, Coastal and estuarine habitat mapping, using LIDAR height and intensity and multi-spectral imagery. Estuarine Coastal and Shelf Science, 78:633-643.

ENGBERG, A., 2005, Produktspecifikation av Svenska CORINE Marktäckedata.

ERDODY, T. L. and MOSKAL, L. M., 2010, Fusion of LiDAR and imagery for estimating forest canopy fuels. Remote Sensing of Environment, 114:725-737.

GALLEGOS, Å., 2005, Design and evaluation of a computer aided calibration program for visual estimation of vegetation cover. Swedish University of Agricultural Sciences, Umeå, Sweden.

GENC, L., DEWITT, B., and SMITH, S., 2004, Determination of wetland vegetation height with LIDAR. Turkish Journal of Agriculture and Forestry, 28:63-71.

HILL, R. A. and THOMSON, A. G., 2005, Mapping woodland species composition and structure using airborne spectral and LiDAR data. International Journal of Remote Sensing, 26:3763-3779.

HOLMGREN, J., PERSSON, A., and SODERMAN, U., 2008, Species identification of individual trees by combining high resolution LIDAR data with multi-spectral images. International Journal of Remote Sensing, 29:1537-1552.

HUDAK, A. T., LEFSKY, M. A., COHEN, W. B., and BERTERRETCHE, M., 2002, Integration of lidar and Landsat ETM plus data for estimating and mapping forest canopy height. Remote Sensing of Environment, 82:397-416.

HYDE, P., DUBAYAH, R., WALKER, W., BLAIR, J. B., HOFTON, M., and HUNSAKER, C., 2006, Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM plus , Quickbird) synergy. Remote Sensing of Environment, 102:63-73.

JENNINGS, S. B., BROWN, N. D., and SHEIL, D., 1999, Assessing forest canopies and understorey illumination: canopy closure, canopy cover and other measures. Forestry, 72:59-73.

KE, Y. H., QUACKENBUSH, L. J., and IM, J., 2010, Synergistic use of QuickBird multispectral imagery and LIDAR data for object-based forest species classification. Remote Sensing of Environment, 114:1141-1154.

KORPELA, I., KOSKINEN, M., VASANDER, H., HOLOPAINEN, M., and MINKKINEN, K., 2009, Airborne small-footprint discrete-return LiDAR data in the assessment of boreal mire surface patterns, vegetation, and habitats. Forest Ecology and Management, 258:1549-1566.

LEE, D. S. and SHAN, J., 2003, Combining lidar elevation data and IKONOS multispectral imagery for coastal classification mapping. Marine Geodesy, 26:117-127.

REESE, H. M., LILLESAND, T. M., NAGEL, D. E., STEWART, J. S., GOLDMANN, R. A., SIMMONS, T. E., CHIPMAN, J. W., and TESSAR, P. A., 2002, Statewide land cover derived from multiseasonal Landsat TM data - A retrospective of the WISCLAND project. Remote Sensing of Environment, 82:224-237.

REESE, H., NILSSON, M., GRANQVIST PAHLÉN, T., HAGNER, O., JOYCE, S., TINGELÖF, U., EGBERTH, M., and OLSSON, H., 2003, Countrywide estimates of forest variables using satellite data and field data from the National Forest Inventory. Ambio, 32:542-548.

SOHLMAN, A. (ED.), 2008, Arter & naturtyper i habitatdirektivet - tillståndet i Sverige 2007. [Species and habitats in the EU habitats directive - the conservation status in Sweden 2007]. The Swedish Species Information Centre, Uppsala, Sweden.

SOININEN, A., 2004, Terra Scan for MicroStation, user's guide. Terrasolid Ltd., Jyväskylä, Finland.

Tree Detection from Low Resolution LIDAR Data

NUSRET DEMIR † and EMMANUEL BALTSAVIAS1‡

demir@geod.baug.ethz.ch

1 Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland

Abstract

In this paper, we present an approach to detect trees in low resolution LIDAR data. Trees are important objects for the landing and the take-off of airplanes, forestry applications, canopy modeling problems, etc. The density of LIDAR data affects directly the quality of the results but usually National LIDAR data don't have high density and it is not possible to use the first and last pulse information for tree detection. We made our investigations using Swiss national data which have 2m resolution which include four pulses LIDAR data. In this study, we investigate two different methods, the first one uses LIDAR DSM & DTM and CIR images, the other method uses only LIDAR data. The performance evaluation of the results has been done using the vector data of the tree regions. Omission and commission errors have been computed and reported.

1. Introduction

In this work, we focus on the extraction of trees by using low resolution LIDAR data. We present two methods, the first one is using LIDAR DSM&DTM and CIR image data, and the other one is based only on LIDAR data.

It is not always possible to have LIDAR data which provide all pulses separately and have a high density. Because of these constraints, some methods are needed to overcome this kind of problems. We have developed two different methods which are based on the data that are available. The first method applies a simple subtraction of DTM from DSM and intersection of the blobs with the classification result of NDVI image, and the second one is a two-step method and it uses only LIDAR DSM and raw point cloud data for tree detection. These two steps are, first a filtering step for the extraction of all the off-terrain points, and secondly, detection of trees in the off-terrain points.

Both results have been compared with vector data and a quality assessment has been performed.

Previous Work

Aerial images and Lidar data are common sources for object extraction. In digital photogrammetry, features of objects are extracted using 3D information from image matching or DSM/DTM data, spectral, textural and other information sources. Pixel-based classification methods, either supervised or unsupervised, are mostly used for land-cover and man-made structure detection. For the classical methods e.g. minimum-distance, parallelepiped and maximum likelihood, detailed information can be found in Lillesand and Kiefer (1994).

In general, the major difficulty in using aerial images is the complexity and variability of objects and their form, especially in suburban and densely populated urban regions (Weidner and Foerstner, 1995).

Regarding object extraction from LIDAR data, it has been defined as a filtering problem of the DSM (raw or interpolated) data by several researches. Some algorithms use raw data (Sohn and Dowman, 2002; Roggero, 2001; Axelsson, 2001; Vosselman and Maas, 2001; Sithole, 2001; Pfeifer et al., 1998), while others use interpolated data (Elmqvist et al., 2001; Brovelli et al., 2002; Wack and Wimmer, 2002).

For tree and canopy extraction, Meng et al. (2008) apply a morphological filter to filter off-terrain objects, and use images to detect trees. They assume that trees do not have linear shapes and additionally they use

morphological indices such as roundness or compactness. Lo and Chen (2008) first filter the digital surface model to derive the terrain model and then they extract the canopy using vertical profiles.

Clark et al. (2004) use the elevation threshold between DSM and DTM to detect tree height and sub-canopy using inverse distance weighting and ordinary kriging techniques. Wulder et al. (2007) combined LIDAR data and satellite imagery to identify tree crown changes.They used images to locate the species in spectral domain. On the other hand, they analyzed vertical profiles using LIDAR data to measure the changed areas. Lang et al. (2006) developed a local maxima algorithm to identify tree tops. They incorporated a region growing algorithm to delineate the respective tree crowns.

Koch et al. (2006) use rasterized laser data and tree tops are detected with a local maximum filter. Afterwards they delineate the crowns with a combination of a pouring algorithm, knowledge-based assumptions on the shape of trees, and a final detection of the crown-edges by searching vectors starting from the trees' tops. Secord and Zachor (2006) propose a two-step method for tree detection consisting of segmentation followed by classification. The segmentation is done using a simple region growing algorithm using weighted features from aerial image and LiDAR, such as height, texture map, height variation, and normal vector estimates. The classification is done using weighted support vector machines (SVM), allowing us to control the miss-classification rate. Kwak et al. (2007) present an approach for delineating individual trees and estimating tree heights using LiDAR. For delineation of individual trees, they apply watershed segmentation to the distance-transformed image from the detected tree tops. They extract the tree heights using the maximum value within the segmented crown boundary by LIDAR data. Rottensteiner and Briese (2002) separate the tree points from the points on buildings by evaluating the surface roughness using LIDAR data.

Few commercial software packages allow automatic terrain, tree and building extraction from Lidar data. In TerraSCAN, a TIN is generated and progressively densified, the extraction of off-terrain points is performed using the angles between points to form the TIN facets and the other parameter is the distance to nearby facet nodes (Axelsson, 2001). In SCOP++, robust methods operate on the original data points and allow the simultaneous elimination of off-terrain points and terrain surface modeling (Kraus and Pfeifer, 1998).

In summary, most approaches try to find the trees using high resolution data. In our work, we try to show the results in different cases by using national low resolution LIDAR data. The first case is based on DSM, DTM, and CIR images, and the second one is based on only LIDAR DSM data is available.

Input Data

The test area is Zurich Airport. CIR aerial images with 8.7 cm GSD, LIDAR raw and interpolated DSM and DTM data, 2D vector data are available. Vector data has only been used for quality assessment purposes.

LIDAR data were acquired with the characteristics given in table 1.

Table 1. LIDAR data characteristics (Zurich Airport).

Lidar Data	DSM	DTM
Provider	Swisstopo	Swisstopo
Type	Raw & grid	Raw & grid
Raw point density & Grid Spacing	1 pt / 2 sqm & 2m	1 pt / 2 sqm & 2m
Acquisition Date	Feb. 2002	Feb. 2002

The images have been first radiometrically preprocessed (noise reduction and contrast enhancement), then the DSM was generated with the software package SAT-PP, developed at the Institute of Geodesy and Photogrammetry, ETH Zurich (Zhang, 2005).The NIR band was selected for DSM generation. The final DSM was generated with 50cm grid spacing. Using this DSM, CIR orthoimages were produced with 12.5cm ground sampling distance. LIDAR raw data (DTM-AV and DSM-AV) have been acquired with "leaves off" in

February 2002 by Swisstopo. The DSM-AV point cloud includes all Lidar points (including points on terrain, tree branches etc.) and has an average point density of 1 point per 2 m^2. The DTM-AV data includes only points on the ground, so it has holes at building positions and less density at tree positions. The height accuracy (one standard deviation) is 0.5 m generally, and 1.5 m at trees and buildings. The 2m spacing grid DSM and DTM were generated by Swisstopo with the Terrascan commercial software from the original raw data.

2. Tree Detection Using LIDAR DSM & DTM and CIR Orthophoto

The above-ground objects have been detected by subtracting the LIDAR DTM from the LIDAR DSM, the blobs include buildings and trees. The height threshold has been selected as 1.5 m (see figure 1).

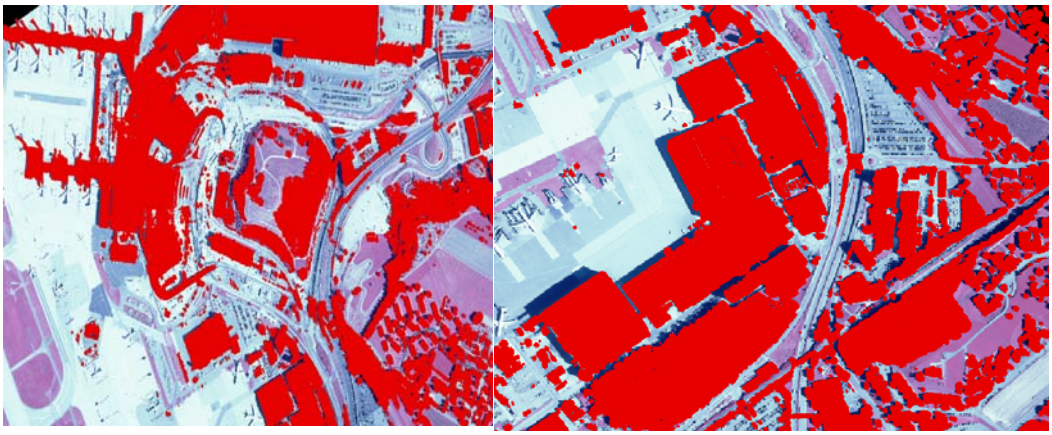


Figure 1. Detected blobs (red) overlaid on orthoimage.

A standard unsupervised (ISODATA) classification of the CIR orthoimage was used to compute an NDVI image, containing vegetation (mainly trees and grass) (see figure 2). The intersection of the nDSM with NDVI corresponds to mainly trees. The errors are mainly vegetation on the roofs. (See figure 3*(a)*- 3(*d*)).

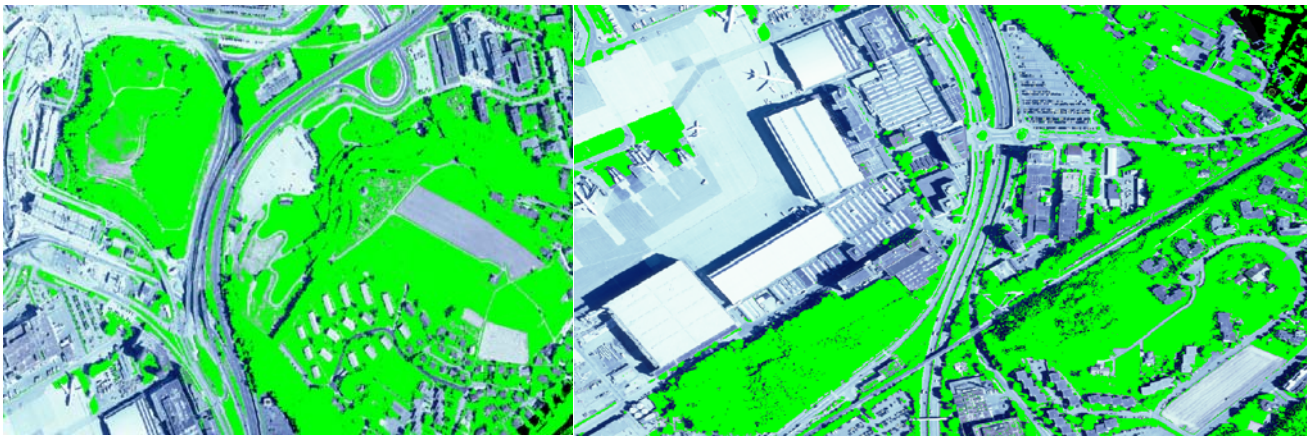


Figure 2. Detected vegetation (green) overlaid on orthoimage.

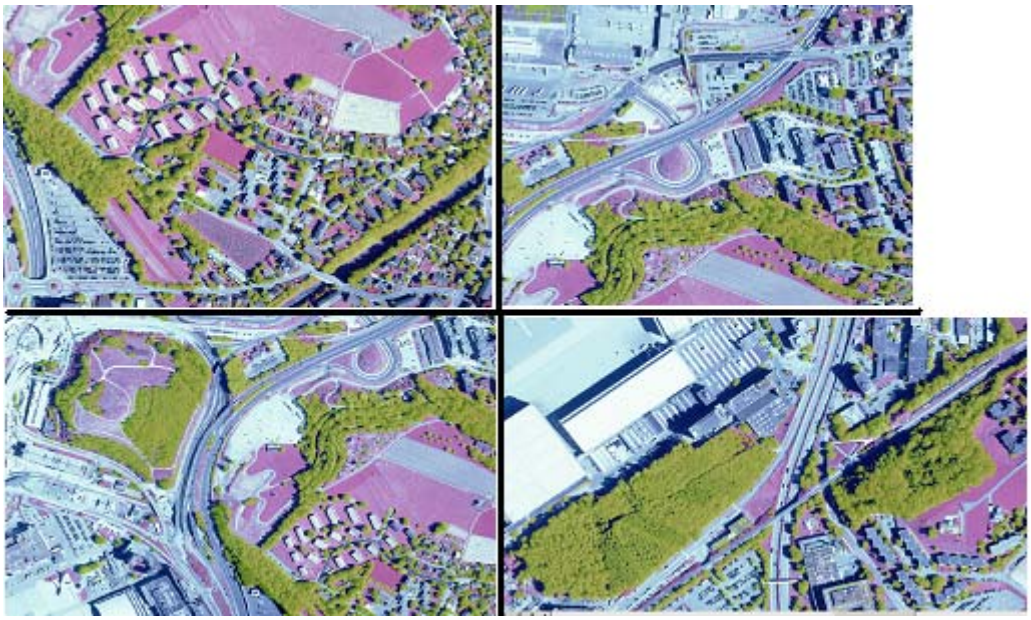


Figure 3(a)- 3(d).Tree detection result overlaid on orthoimage.

After the quality analysis with the reference data, the correctness has been calculated as 82% and the omission error is 5%.

3. Tree Detection Using only LIDAR Data

Reduction of DSM to DTM and detection of the blobs

Since the data do not allow using the pulses for tree detection, a filtering approach is needed to detect terrain points and off-terrain objects. A progressive morphological filtering method has been used for blob detection. For the filtering approach, an input interpolated DSM data has been used. Then a morphological filter has been used to detect all off-terrain objects (see fig. 4) which include buildings, trees and other objects. We perform an opening (erosion + dilation) operation on the interpolated surface to derive a secondary surface. The elevation difference of a grid between the previous and current surface is compared to a threshold to determine if the grid points are non-ground measurements. The height difference threshold (dh) has been computed using the predefined maximum terrain slope(s). The size of filtering windows (w) has been increased and the derived surface has been used as an input for the next operation (Zhang et al., 2003).

$$dh = s(wi - wi-1)c + dh_0$$

$$\text{If } dh > dh_{max} \quad dh = dh_{max}$$

dh is the height difference threshold

dh_0 is the initial elevation difference threshold which approximates the error of DSM measurements (0.2-0.3 m),

dh_{max} is the maximum elevation difference threshold (m)

c is the grid size (m)

s is the predefined maximum terrain slope (percent slope)

wi is the filtering window size (in number of cells) at ith iteration.

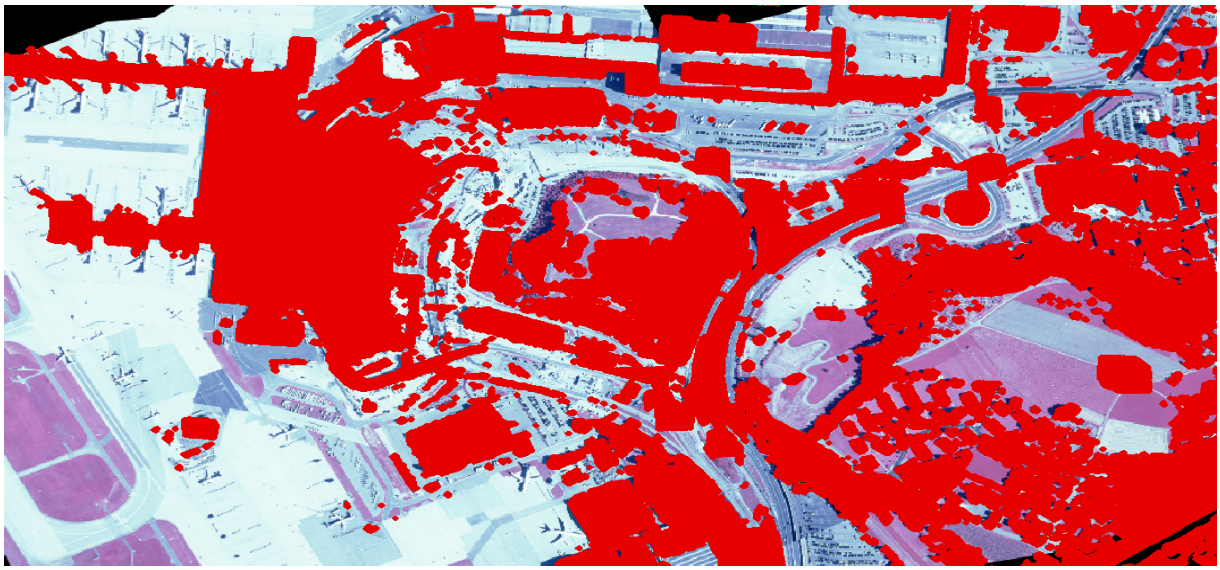


Figure 4. Detected blobs (red) overlaid on orthoimage.

After detection of the blobs, terrain regions have been interpolated and a DTM has been generated with 2 m resolution.

Tree detection

As mentioned above, in the raw DSM data the point density is generally much higher at trees than at open terrain or buildings (see figure 5). We start from the blobs that are generated by reduction of DSM, and we have intersected the blob regions with raw LIDAR data.

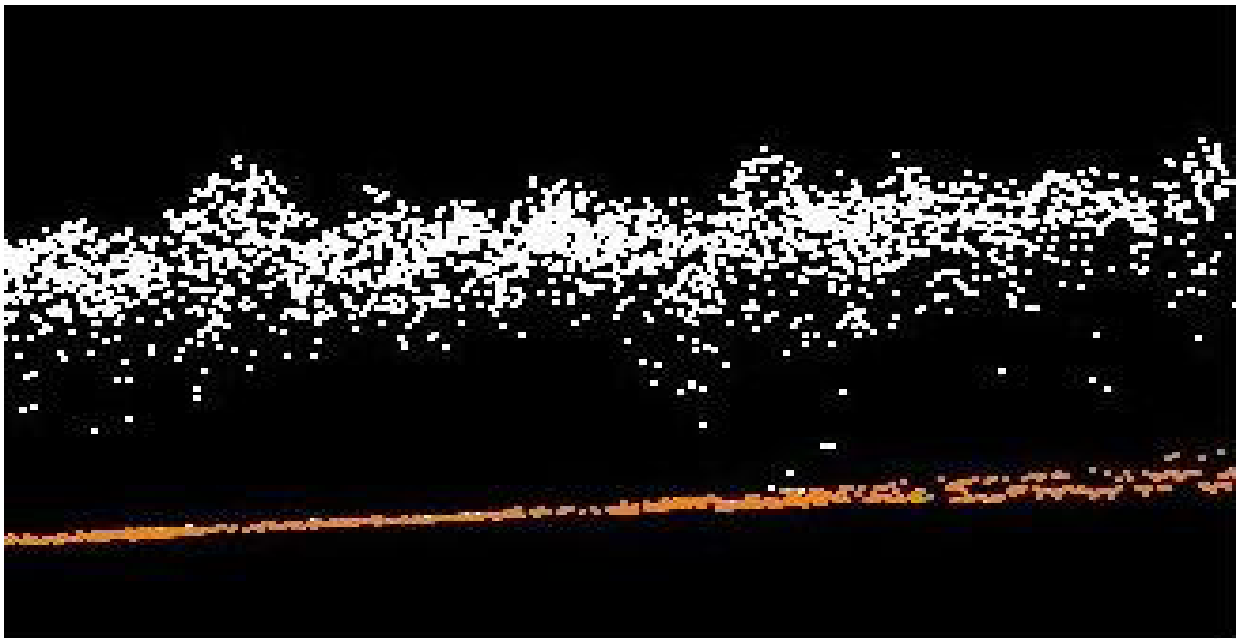


Figure 5. Trees in LIDAR data, profile view.

In the detection step, RANSAC method has been applied on all off-terrain points to find all planar surfaces which are assumed that they belong to the buildings. After the removal of all the planar objects, all the

remaining points consist of the unfiltered terrain, the building roofs which do not have planar surfaces, other objects e.g. cars and the tree regions.

First, raw LIDAR data have been overlaid on the blob detection result and later the plane detection process has been applied. Schnabel et al. (2007a, 2007b)'s RANSAC method has been used for fitting of the roof surfaces into the planes. RANSAC generates a large amount of hypotheses of primitive shapes by randomly selecting minimum subset of sample points that each uniquely determine the parameter of a primitive. The scoring mechanism is employed to detect the best primitive. The process starts with calculating the surface normal vectors of each point with selection of neighboring points. They use localized sampling strategy using octree data structure for the random selection of minimal subset of points. The score of the candidate shape is evaluated by using the parameters which are the tolerance distance of shape, minimum deviation of surface normal and connectivity of points.

After detection of the points which belong to roof surfaces, all these points, which belong to plane shapes, have been removed, and remaining points which include the tree regions, cars, and other objects have been used for further processing.

In the next step, we used a search window over the raw LIDAR DSM data with a size of 5 m x 5 m. Neighboring windows have an overlap of 50%. The window size has a relation with the number of points in the window and the number of the points in the search window affects the quality of the detection result. The method uses all points in the window and labels them as tree if all parameters below have been met. The size of $25m^2$ has been agreed to be enough to extract one single tree. A bigger size may result in wrong detection especially in areas where the buildings are neighboring with single trees.

The points in each search window are projected onto the xz and yz planes and divided for each projection in eight equal sub-regions using x_{min}, x_{mid}, x_{max}, z_{min} z_{mid1} z_{mid2} z_{mid3} z_{max} as boundary values of sub-regions, with $x_{mid} = x_{min} + 2.5m$, $x_{max} = x_{mid} + 2.5m$, $z_{mid1}=z_{min}+(z_{max}-z_{min})/4$, $z_{mid2} =z_{min}+2*(z_{max}-z_{min})/4$, $z_{mid3}=z_{min}+3*(z_{max}-z_{min})/4$ and similarly for the yz projection. The density in the eight sub-regions is computed. The first step is the detection of trees and the second the subtraction of tree points from all off-terrain points. The trees have been extracted by three different parameters. Since the estimation of the parameters is based on the LIDAR points which are in the tree regions, the parameters have been calculated using the statistics of previous results.

The first parameter (vd) is the number of the eight sub-regions which contain at least one point. The trees have high Lidar point density vertically. Thus, at trees more sub-regions contain Lidar points. Using the results of previous method, we have observed that at least 4 out of the 8 sub-regions contain points. Thus, the parameter (vd) has been selected as $vd \geq 4$. The third parameter (z) is the tree height. Using the results from the previous method, we calculated the minimum tree height as 3m. The fourth parameter (d) is the point density. The minimum point density has been calculated for the tree areas as 15 points/ $25m^2$. By applying these four parameters to the raw DSM Lidar data, the tree points have been extracted (see figure. 7). The workflow can be seen in Figure 6.

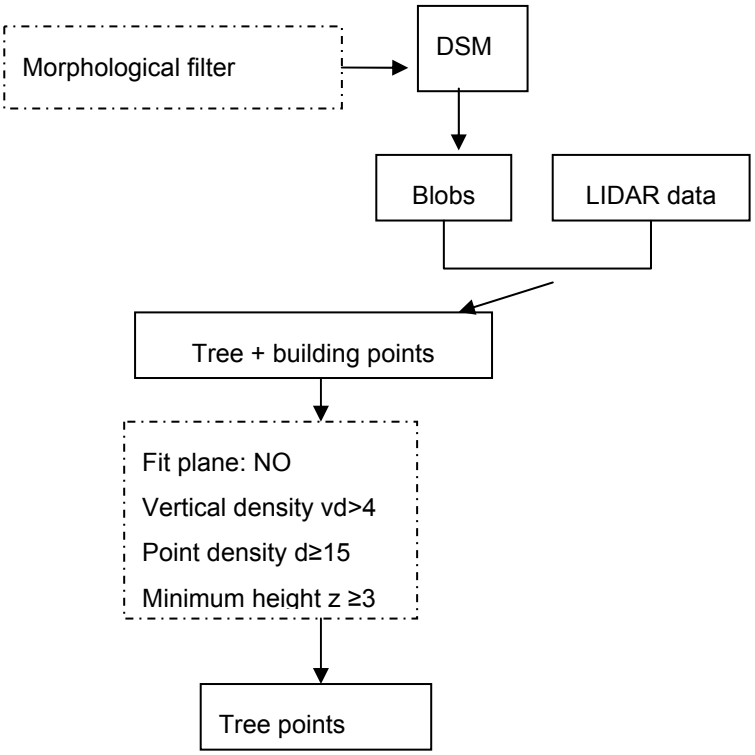


Figure 6. Workflow of detection of buildings.

Figure 7(a)- 7(d). Detected trees(green) overlaid on orthoimage.

After the quality analysis with the reference data, the correctness has been calculated as 75% and the omission error is 15%.

4. Conclusions

In this paper, two methods for tree detection in Lidar data and aerial images have been presented. National LIDAR data often have low density and using only LIDAR data have limitations. In our method which is based on only LIDAR data, we developed several steps to overcome the problems which are related to the data itself. First we tried to eliminate non-tree points as much as possible, then we tried to extract trees using vertical density analysis. Very dense tree areas which do not have high vertical density and some single small trees which do not have sufficient density could not be extracted, but the accuracy of all detection is reasonable with 75% correctness. We also showed the results in the case when image data is available. In this case, with usage of image data, the correctness becomes 82% with 5% omission error.

Acknowledgments

This work has been supported by the EU FP6 project Pegase and KTI. We acknowledge data provided by Swisstopo and Unique Company (Airport Zurich), and we would like to thank Mr. Schnabel from University of Bonn, for providing us the algorithm library for the RANSAC approach.

References

AXELSSON, P., 2001, Ground estimation of laser data using adaptive TIN-models. *In Proc. of OEEPE workshop on airborne laserscanning and interferometric SAR for detailed digital elevation models*, 1-3 March, Stockholm, Sweden, pp. 185-208.

CLARK, M.L., CLARK, D.B., AND ROBERTS, D.A., 2004, Footprint Lidar Estimation of Sub-canopy Elevation and Tree Height in a Tropical Rain Forest Landscape, *Remote Sensing of Environment*, 91, pp.68-89.

ELMQVIST, M., JUNGRT, E., LANTZ, F., PERSSON, A. AND SODERMAN, U., 2001, Terrain modeling and analysis using laser scanner data. IAPRS, Vol. 34, Part 3/W4, pp. 219-227.

KRAUS, K. AND PFEIFER, N., 1998, Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry & Remote Sensing* 53(4) ,193-203.

KOCH, B., HEYDER, U. AND WEINACKER, H., 2006, Detection of Individual Tree Crowns in Airborne Lidar Data , *Photogrammetric Engineering & Remote Sensing* Vol. 72, No. 4, pp. 357–363.

KWAK, D.A., W.K. LEE, J.H. LEE, G.S. BIGING AND P. GONG. 2007, Detection of individual trees and estimation of tree height using LIDAR data. *Journal of Forest Research* 12(6), pp. 425-434.

LANG, S., TIEDE, D., MAIER, B.AND BLASCHKE,T., 2006, 3D Forest structure analysis from optical and LIDAR data, *Revista Ambiência, Guarapuava, v.2 Edição* Especial 1, pp.95-110.

LILLESAND, M.,Tt. AND KIEFER, W., R., 1994, Remote Sensing and Image Interpretation, Third Edition, John Wiley & Sons, Inc., New York, 750 pp.

LO, Y. AND CHEN, L.C., 2008, Canopy Extraction Using Airborne Laser Scanning Data in Forestry Areas, IAPRS Vol. 37, Part 3B, pp. 367 – 372.

MENG, X., CURRIT, N. AND WANG, L., 2008, Morphology-based Building Detection from Airborne LIDAR Data, *ASPRS 2008 Annual Conference,* Portland, Oregon, April 28 – May 2, 2008, pp.432-442.

PFEIFER, N., KOSTLI, A. AND KRAUS, K., 1998. Interpolation and filtering of laser scanner data - implementation and first results IAPRSVol. 32, Part 3/1, pp. 153 – 159.

ROGGERO, M., 2001, Airborne Laser Scanning: Clustering in raw data. IAPRS Vol. 34, Part 3/W4, pp. 227-232.

ROTTENSTEINER, F. AND BRIESE, C., 2002, A New Method for Building Extraction in Urban Areas from High-Resolution LIDAR Data. IAPRS Vol 34/3A , pp. 295 – 301,

SCHNABEL, R., WAHL, R. AND KLEIN, R., 2007a. Efficient RANSAC for Point-Cloud Shape Detection. *Computer Graphics Forum*. 26(2). pp.214-226.

SCHNABEL, R., WAHL, R., WESSEL, R. AND KLEIN, R., 2007b. Shape Recognition in 3D Point Clouds. *Computer Graphics Technical report*, No : CG-2007-1, University of Bonn, Germany

SECORD, J. AND ZAKHOR, A., 2006, Tree detection in LiDAR data, *Proceedings of the 2006 IEEE Southwest Symposium on Image Analysis and Interpretation*, pp.86-90.

SITHOLE, G., 2001, Filtering of laser altimetry data using a slope adaptive filter. *IAPRS (*International Archives of The Photogrammetry, Remote Sensing and Spatial Information Sciences*),* Vol. 34, Part 3/W4, pp. 203-210.

SOHN, G. AND DOWMAN, I., 2007. Data fusion of high-resolution satellite imagery and LiDAR data for automatic building extraction. *ISPRS Journal of Photogrammetry and Remote Sensing* 62(1), pp.43-49.

VOSSELMAN, G. AND MAAS, H., 2001, Adjustment and filtering of raw laser altimetry data. *In Proc. of OEEPE workshop on airborne laserscanning and interferometric SAR for detailed digital elevation models*, 1-3 March, Stockholm, Sweden,pp.62-72.

WACK R. AND WIMMER A., 2002, Digital Terrain Models from Airborne Laser Scanner data – A grid based approach. *IAPRS,* Vol 34, Part 3B, pp. 293-296.

WEIDNER, U. AND FOERSTNER, W., 1995, Towards automatic building extraction from high-resolution digital elevation models. *ISPRS Journal of Photogrammetry and Remote Sensing* 50(4), pp.38 – 49.

WULDER, MA., HAN, T.,WHITE, J.C.,SWEDA AND TSUZUKI, H.,2007, Integrating profiling LIDAR with LANDSAT Data for Regional Boreal Forest Canopy Attribute Estimation and Change Characterization, *Remote Sensing of Environment*, pp.123-137.

ZHANG, K., CHEN, S., WHITMAN, D., SHYU, M., YAN, J. AND ZHANG, C., 2003. A Progressive Morphological Filter for Removing Non-ground Measurements from Airborne LIDAR data. *IEEE Transactions on Geoscience and Remote Sensing,* 41, pp. 872-882.

ZHANG, L., 2005. Automatic Digital Surface Model (DSM) generation from linear array images. *Ph.D. thesis*, Institute of Geodesy and Photogrammetry, ETH Zurich, Report No. 90.

The potential of airborne Lidar for forest stand map generation

Yong Pang †, Zengyuan Li†, Mingze Li‡, Kaiguang Zhao§, Tian Fu†, Qingwang Liu†, Erxue Chen†, Bingxiang Tan†, Yiming Feng†

†.Institute of Forest Resource and Information Technology, Chinese Academy of Forestry, Beijing 100091, China

‡. Northeast Forestry University, Harbin 15004

§. Center on Global Change, Duke University, Durham, NC 27705, USA

ABSTRACT

Forest stand map contains many attributes for each sub-compartment (the basic forest management unit). Traditional methods for forest stand map generation are based on high space resolution remote sensing image, topographic map and field inventory, and generally are very tedious and time-consuming. The airborne lidar technology brings new possibilities for forest stand map generation, especially in forest stand parameters estimation. The purpose of this study is to demonstrate the potential of lidar-derived fine-resolution forest information as a complement or even substitute to forest stand maps generated with traditional inventory methods. We used airborne lidar data and field data of concomitant plots in a forest of the Northeastern China to estimate several typical forest parameters, including dominate height, mean height, volume density, biomass, and canopy closure. The results showed the R^2 for Lorey's height, canopy closure and volume density estimations were of 0.85, 0.46 and 0.84. The results indicated that these forest parameters could be estimated accurately using airborne laser scanner data with comparable accuracy to forest inventory based measurements.

1. Introduction

Forest stand map, which is generated by forest management inventory, is the most fundamental map in forest management. In a stand forest map, the smallest unit is forest stand subcompartment which is also the smallest operational unit forest management planning in many countries such as China, Russia, and Finland. The soil, relief, tree status are at the same level in a subcompartment. Each forest stand subcompartment has an inventory database which contains the information of soil, tree, understory vegetation, disturbance history and management methods, etc. A forest stand subcompartment has uniform management plan and method. Detailed forest management plans cannot be produced without subcompartment-level estimates of site characteristics and growing stock and without silvicultural treatment proposals for each subcompartment. Models describing forests and forestry in forest management planning packages are merely simplifications of reality (Siitonen et. al. 1996).

Traditional method for forest stand map generation is based on topographic map and field inventory, which is very tedious and time-consuming. With the advent of high spatial resolution remote sensing images, the boundary of subcompartment could be divided with aid of topographic map in laboratory. But the attributes of each subcompartment still need to be collected with field measurements.

Laser altimeter systems provide high-resolution geo-located measurements of the vertical structure of vegetation and the ground elevations beneath the canopies, which can characterize the vegetation and terrain surfaces with high accuracy (Lefsky et al., 2002). It brings new possibilities for forest stand map generation with its capability of highly accurate height measurements and high spatial resolution photographs collected simultaneously.

Early researches found that canopy closure was most strongly related to the penetration capability of the laser pulse, and indicated that the pulsed laser system may be used to remotely sense the vertical forest canopy profile and assess tree height. Then the forest biomass and volume were estimated in related

researches which also obtained good results (Nelson et al., 1984; 1988). In the past decades, many researches exposed the enormous potential of laser applications in the forestry, forest variables such as tree height, basal area, biomass, etc. can be estimated accurately using airborne laser scanner data (Nelson et al., 1997; Popescu et al., 2002; Holmgren et al., 2004; Zhao et al., 2009). Till now, aerial photographs and low density lidar data have already been used to estimate Lorey's height with the aid of field data (Magnussen & Boudewyn, 1998; Lim et al., 2003; Næsset, 2004; Pang et al., 2008b). High density lidar data have been used successfully to estimate tree location, height, crown size and crown length at the level of individual trees (Hyyppä et al., 2001; Popescu et al., 2003; Falkowski et al., 2006; Solberg et al., 2006; Zhao & Popescu, 2007; Pang et al., 2008a, Popescu & Zhao, 2008).

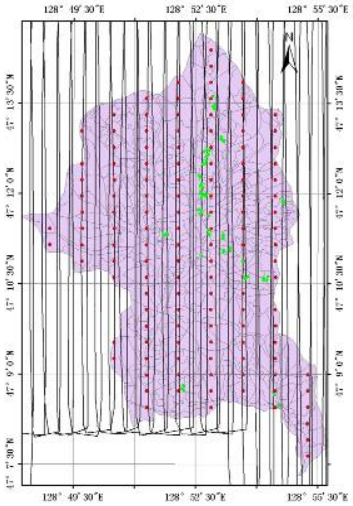
The purpose of this study was to investigate the potential of airborne lidar data for estimation of mean height, volume density, and canopy closure to generate forest stand maps. The data acquisition and processing will be introduced first. Then the lidar indices calculation and forest parameters estimation algorithms and inventory based forest map will be described. The usefulness will be analyzed in comparison to field measurements.

Study area and Data acquisition

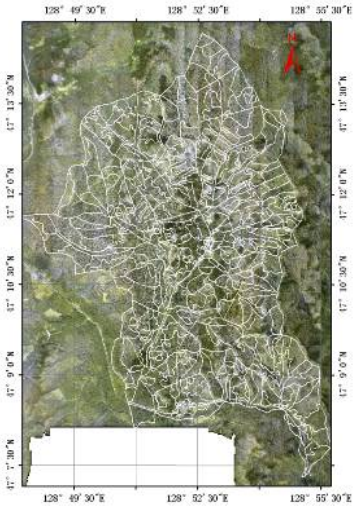
Study area

The study area is Liangshui National Natural Resever (centered at 47˚10'N, 128˚53'E) (Fig. 1) in the south part of Xiaoxinganling, Northeast of China. This area is intended to maintain and restore broad-leaf Korean pine forest conditions since 1980. Its annually mean, maximum and minimum temperatures are of -1°C, 37.4°C and -47.7°C respectively. Mean precipitation is 676.0 mm. Elevation range in the study area is approximately 280~707 m above sea level.

The main forest species in this region are coniferous, including *Pinus koraiensis, Picea jezoensis, Abies nephrolepis, Larix gmelin.* There are also several deciduous species such as *Betula platyphylla*, *Populus ussuriensis, Fraxinus mandshurica, Phellodendron amurense*.



(a) Study area and experiment design. Solid filled polygons are digitized forest stand maps. Black boundaries are lidar flight lines. Red and gree dots are plots by Northeast Forest University and Chinese Academy of Forestry.



(b) Acquired air-photos overlayed by subcompartment polygons (white color)

Fig. 1 Research area and project design

Field data

According to forest types and tree height, volume density levels, 78 circular plots (radius of 13.82 m) data were collected from July 26th to August 14th in 2009. The trees with a diameter at breast height (DBH) ≥5cm were recorded. DBH were measured using PI-tape. Tree height was measured using TruPulse laser altimeter. Crown widths were measured using the tape measure. The center location of each plot was measured using a differential global positioning system (DGPS). Then the Lorey's mean height (i.e., mean height weighted by basal area) and volume density were calculated.

The canopy closure was measured over another 32 plots. Two 20 meters transects were sampled with 1 m interval stop for each plot. Sky or canopy was marked for each stop from observer's vertical above. And canopy closure was calculated by total marks from these observations.

ALS data

Airborne waveform data was collected using the LiteMapper 5600 system in September of 2009 using a Yun-5 aircraft. The Riegl LMS-Q560 laser was used. This system has a wavelength of 1550 nm, with a 0.5 mrad beam divergence and 3 ns pulse length. It operated at a 50 kHZ pulse rate at 650 m AGL. The resulting footprint size was 35 cm with a point density of approximately 2 points/m^2. The digital air-photos with nominal spatial resolution of 18 cm were collected simultaneously.

Combining the range estimates from the lidar itself with platform position from Global Positioning System (GPS) receivers, platform attitude (roll, pitch and yaw) from an Inertial Navigation System (INS), and angle encoders for the orientation of the scan mirror, yields the absolute position of the reflecting surfaces for each laser pulse. Accuracy reports compared to GCPs from RTK measurements indicated that the RMS errors of the x, y, z coordinates for each point were within acceptable tolerances (vertical error < 15 cm, horizontal error < 50 cm, as estimated on smooth level surfaces).

2. Methods

ALS data processing

The acquired lidar waveform data were transformed into point cloud, key points from Gaussian decomposition. In cases which the lidar shots do not reach the actual ground surface (e.g. when dense vegetation is present), the minimum elevation of the lidar shots within a grid cell is not a good estimate of the actual terrain elevation. In these cases, the lowest elevation recorded most often reflects a dense layer of vegetation which the lidar cannot penetrate. The processing of the initial estimates of elevation to remove this effect is referred to as retrieving the "bare-earth surface" digital terrain model (DTM). In this work, the DTM was developed using the Terrascan software system and a digital terrain model was provided as a deliverable by the vendor.

Then the estimates of height for each lidar point were calculated as the difference between the point's elevation and the elevation of the DTM at that location. Then the lidar data were subset for each field plots and 1 km by 1 km tiles for further processing. The plot subsets will be used for forest parameters estimation with field measurements. And the tiled data will be used for parameters estimation for the whole study area.

Lidar indices calculation

Two sets of lidar indices were tested in this study. One is height indices and the other is density indices. The height indices evaluated included the maximum height of all points, mean height of all points, the quadratic mean height (the square root of the mean squared height of each lidar point) as well as height percentiles. Height percentiles are defined as the height at which a certain percent of data fall below - we evaluated 5% intervals from 5% to 95% (denoted as h_5, …, h_{95}). Those points with a height value > 2 m were considered to belong to the tree canopy and used for indices calculation (Pang et al., 2008; Næsset & Gobakken, 2008).

The percentiles of the canopy height distributions for 5% (h_5), 10% (h_{10}), ..., and 95% (h_{95}) were computed. Canopy density was then computed as the proportions of laser points above each percentile height to total number of points.

The DGPS coordinates and exact plot radius were used when calculating lidar indices plot subsets. Considering the plot size and the number of lidar points for lidar indices calculation, 10 m square cell was used when calculating lidar indices for the whole study area. This resulted to a 40 bands remote sensing dataset.

Forest parameters estimation using airborne lidar data

According to previous studies (Næsset, 2004; Næsset & Gobakken, 2008), the following linear regression models were selected

$$h_{\mathrm{L}} = \beta_0 + \beta_1 h_5 + \beta_2 h_{10} + ... + \beta_{19} h_{95} + \beta_{20} h_{\max} + \varepsilon \quad (1)$$

$$CC = \beta_0 + \beta_1 h_5 + \beta_2 h_{10} + ... + \beta_{19} h_{95} + \beta_{20} h_{\max} + \beta_{21} \mathrm{d}_5 + \beta_{22} \mathrm{d}_{10} + ... + \beta_{39} \mathrm{d}_{95} + \beta_{40} \mathrm{c} + \varepsilon \quad (2)$$

$$\ln V = \beta_0 + \beta_1 \ln h_5 + \beta_2 \ln h_{10} + ... + \beta_{19} \ln h_{95} + \beta_{20} \ln h_{\max} + \beta_{21} \ln \mathrm{d}_5 + \beta_{22} \ln \mathrm{d}_{10} + ... + \beta_{39} \ln \mathrm{d}_{95} + \beta_{40} \ln \mathrm{c} + \varepsilon \quad (3)$$

where h_L is the Lorey's mean height (mean height weighted by basal area) (m); CC is the canopy closure; and V is the forest volume density per hectare. The h_5, h_{10}, …, h_{95} are percentiles corresponding to 5, 10, …, 95% of laser canopy heights (m); h_{max} is maximum of the laser canopy heights (m); $d_0, d_1, \ldots, d_9$ are the canopy densities corresponding to the proportions of laser echoes above fraction 0,1,…,9 to total number of echoes; c is the canopy density corresponding to the proportions of laser echoes with a height value > 2 m to total number of echoes; ε = a normally distributed error term [$\varepsilon \sim N(0, \sigma^2)$].

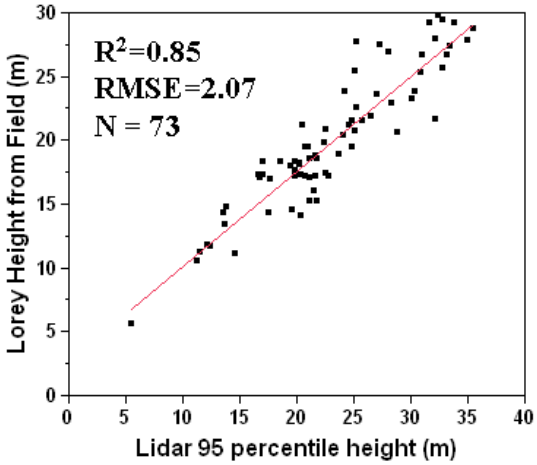
Stepwise regression was used for variable selection and the maximum R^2 improvement variable selection techniques were applied to select the ALS-derived variables to be included in the models (Næsset & Gobakken, 2008). Independent variables could be removed if the value of statistic F value is too small and the result of T-test do not reach the level of significance (P >0.1). On the contrary, independent variables are allowed to entered regression model (P < 0.05). The least squares method was used generally, and repeated until all the independent variables of the regression equation are accord with the requirements of entering models.

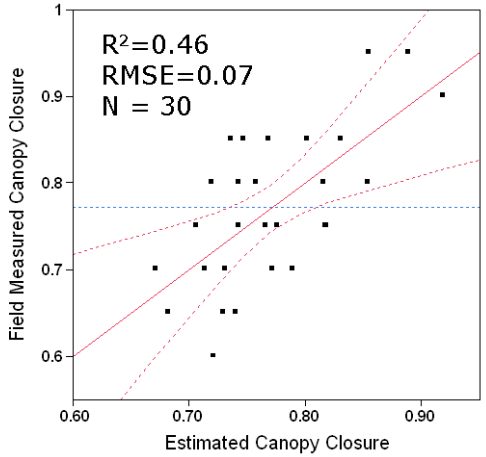
After the relationships between the Lorey height of each of the field plots and the airborne lidar data coincident with them have been developed, the trained equations were used to the lidar indices of each cell.

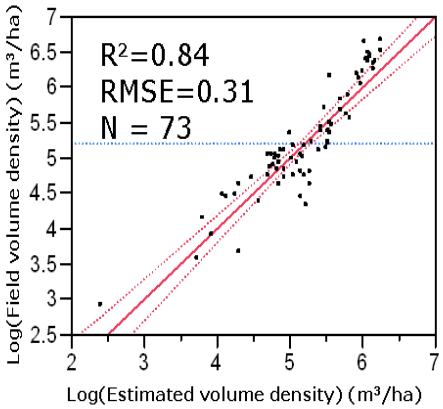
3. Results and Discussions

Comparison of Lidar estimated forest parameters with field measurements

Appropriate variables of laser indices were selected by stepwise regression







$$\hat{h}_{\mathrm{L}} = 0.75h_{95} + 2.63$$

(a) Estimation of Lorey's Height

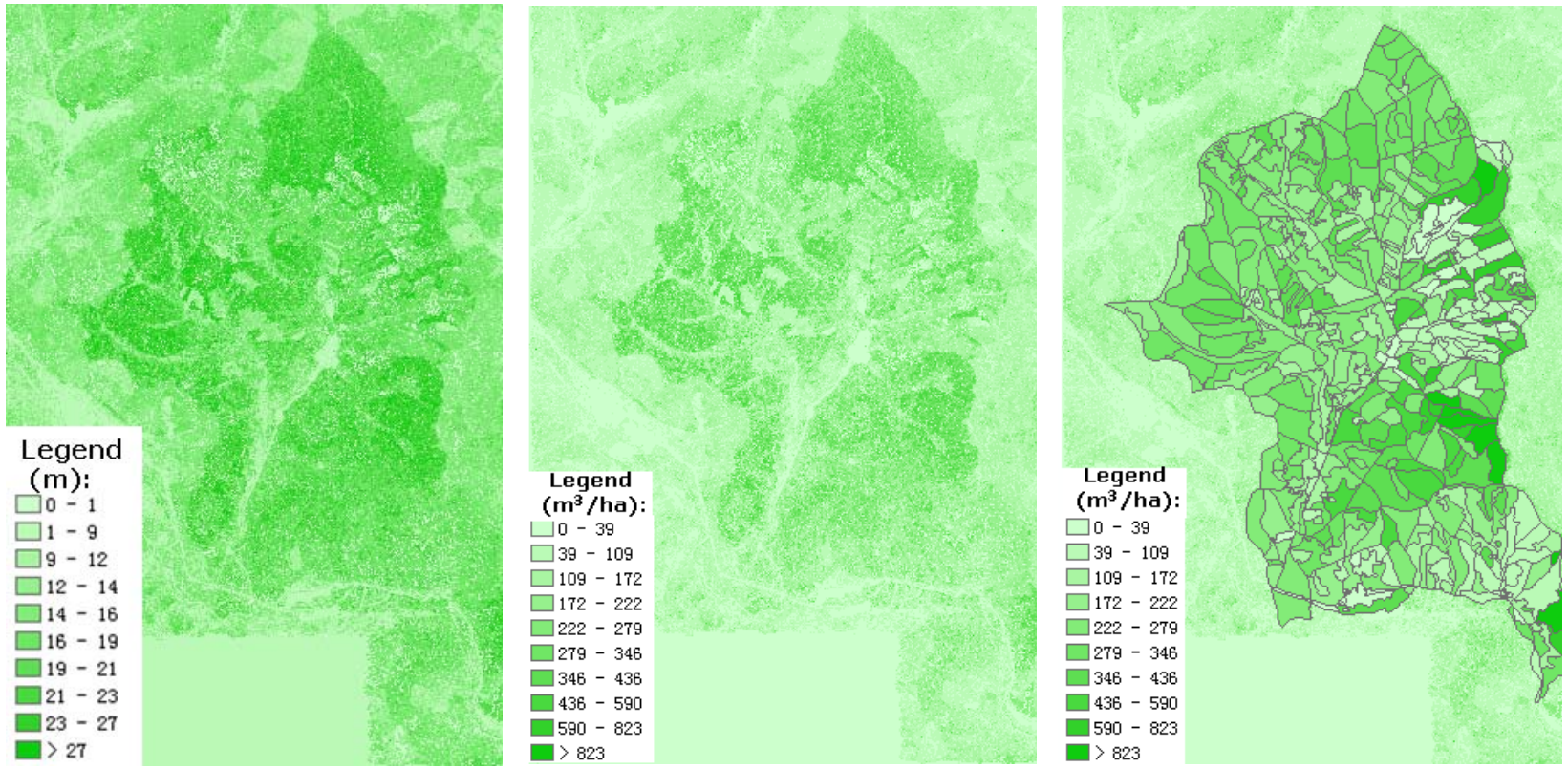
$$\hat{C}C = 0.87c + 0.12$$

(b) Estimation of Canopy Closure

$$\hat{V} = 0.4025\mathrm{c}^{0.18}{h_{95}}^{2.02} \cdot \exp(\frac{0.31^2}{2})$$

(c) Estimation of Volume Density

Fig. 2 Estimation of forest parameters using Lidar indices with field plot measurements



(a) Lidar Estimated Stand Mean Height

(b) Lidar Estimated Forest Volume Density

(c) Stand Volume Density from Inventory-based Forest Stand Map at Subcompartment Level

Fig. 3 Lidar estimated of forest parameters and comparison with inventory-based forest stand map

analysis method for each forest variables. The 95th percentile height (h_{95}) was selected for the Lorey's height (h_L) estimation. There was high correlation (r = 0.92) between h_{95} and h_L at plot level (Fig. 2a). 85% of the variability could be explained, and the root mean square error (RMSE) was 2.07 m, accounting for 6% of the average mean tree height. In order to test the stability of system, 50 plots were sampled randomly for regression analysis; another 23 plots were sampled for validation. The RMSE was 1.86 m. The average R^2 was 88%.

The lidar calculated canopy density at 2 m level (c), h_{15} and d_{95} were selected when we did stepwise variable selection for canopy closure estimation. But h_{15} and d_{95} are not significant statistically. Only c was used for CC estimation (Fig. 2b). The correlation is 0.68 between c and CC at plot level (Fig. 2b). 46% of the variability could be explained, and the RMSE was 0.07, accounting for 11% of the average canopy closure. This accuracy was less than height and volume density (below) estimation. The selection of different height threshold when calculating density indices might affect this. And the canopy closure from field measurement also affected by the experiences of observers.

The 95 percentile height (h_{95}) and density variable c were selected for the volume density (V) estimation. The correlation is high (r = 0.92) at plot level (Fig. 2c). 84% of the variability could be explained, and the root mean square error (RMSE) was 1.36 m^3ha^{-1}. When we did similar validation schema as height estimation, the RMSE was 1.32 m^3ha^{-1} and R^2 was 82%.

Forest parameters estimation for forest stand map using ALS

As there were good relationship between lidar indices and forest variables, the regression equations as shown in Fig. 2 were used over lidar metrics over each tiled cells as described in section 3.3. The lidar estimated forest mean height and volume density for the whole study area were shown in Fig. 3a~b.

When compared with volume density from the inventory-based forest stand map (Fig. 3c), the general patterns shown good consistent. The forest stands still shown some heterogeneous for some subcompartments (Fig. 3b~c). There were several inconsistent areas between lidar-based and inventory-based volume density (e.g., the central left corners), which might be caused by the selection of angle-gauge plots. This might bring new chances for more accurate characterize the attributes of subcompartments in stand maps, even for more intensive and automotive subcompartment segmentation using Lidar technology.

4. Conclusions

This study investigated the possibility of estimation forest variables from airborne lidar data for stand map generation. The forest mean height and volume density were estimated with high accuracy (R^2 was 0.85 and 0.84 separately). The accuracy of canopy closure estimation was less (R^2 was 0.46), which might be caused the selection of height threshold when lidar calculation and field measurement uncertainties. The quantitative continuous forest parameter maps from lidar data provided new chances for characterizing the attributes of subcompartments in stand maps more accurately, and more intensive, automotive subcompartment segmentation.

Acknowledgement

The authors would like to thank all the participants of 2009 Forest Airborne Remote Sensing Campaign from the Chinese Academy of Forestry and the Northeast Forest University for help to investigate these forest plots. This study was funded by the State Forest Administration of China (200704019) and the Ministry of Science and Technology of China 863 program (2007AA12Z173, 2009AA12Z142).

References

Falkowski, M.J., Smith, A.M.S., Hudak, A.T., Gessler, P.E., Vierling, L.A., and Crookston, N.L. 2006. Automated estimation of individual conifer tree height and crown diameter via Two-dimensional spatial wavelet analysis of lidar data. *Canadian Journal of Remote Sensing*, Vol. 32, No. 2, pp. 153-161.

Gobakken, T. & Næsset, E., 2004. Estimation of diameter and basal area distributions in coniferous forest by means of airborne laser scanner data. *Scandinavian Journal of Forest Research*, 19, pp. 529-542

Holmgren, J., 2004. Prediction of tree height, basal area and stem volume in forest stands using airborne laser scanning. *Scandinavian Journal of Forest Research*, 19, pp.543-553

Hyyppä, J., Kelle, O., Lehikoinen, M., and Inkinen M. 2001. A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. *IEEE Transactions on Geoscience and Remote Sensing,* Vol. 39, No. 5, pp. 969-975.

Koivuniemi J. and Korhonen K.T., 2006. Inventory by compartments. In: A. Kangas and M. Maltamo, Editors, *Forest Inventory. Methodology and Applications*, Springer, Dordrecht, pp. 271–278.

Lefsky, M.A., Cohen, W.B., Parker, G.G., & Harding, D.J. 2002. LiDAR remote sensing for ecosystem studies. *Bioscience*, 52(1), 19–30.

Lim, K., Treitz, P., Baldwin, K., Morrison, I., & Green, J., 2003. Lidar remote sensing of biophysical properties of tolerant northern hardwood forests. *Canadian Journal of Remote Sensing*, 29(5), pp. 658-678

Magnussen, S. & Boudewyn, P. 1998. Derivations of stand heights from airborne laser scanner data with canopy based quantile estimators. *Canadian Journal of Forestry Research*, Vol. 28, No. 7, pp. 1016-1031.

Næsset, E. (2004). Practical large-scale forest stand inventory using small-footprint airborne scanning laser. *Scandinavian Journal of Forest Research*, 19(2), 164-179.

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(6), pp. 3079-3090

Nelson, R., Krabill, W., & MacLean, G.., 1984. Determining forest canopy characteristics using airborne laser data. *Remote Sensing of Environment*, 15(3), pp. 201-212

Nelson, R., Krabill, W., & Tonelli, J., 1988.Estimating forest biomass and volume using airborne laser data. Remote Sensing of Environment, 24(2), pp. 247-267

Nelson, R., Oderwald, R., & Gregoire, T. G.., 1997. Separating the ground and airborne laser sampling phases to estimate tropical forest basal area, volume, and biomass. *Remote Sensing of Environment*, 60(3), pp. 311-326

Pang Y., Lefsky M., Andersen H., Miller M., Sherril K. 2008a. Automatic Tree Crown Delineation Using Discrete Return Lidar Data and its Application in ICEsat Vegetation Product Validation. *Canadian Journal of Remote Sensing*, Vol. 34, Suppl. 2, pp. 471–484.

Pang Y., Zhao F., Li Z., Zhou S., Deng G., Liu Q., Chen E. 2008b. Forest Height Inversion using Airborne Lidar Technology. *Journal of Remote Sensing*, Vol. 12 No.1, 152-158

Popescu, S.C., Wynne, R.H., & Nelson, R.F., 2002. Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size. *Computers and Electronics in Agriculture*, 37(1-3), pp. 71-95

Popescu, S.C., Wynne, R.H., and Nelson, R.H. 2003. Measuring individual tree crown diameter with LIDAR and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote Sensing*, Vol. 29, No.5, pp. 564−577.

Popescu, S. C., & Zhao, K. (2008). A voxel-based lidar method for estimating crown base

height for deciduous and pine trees. Remote Sensing of Environment, Vol. 112, pp. 767−78.

Solberg, S., Næsset, E., and Bollandsås, O.M. 2006. Single tree segmentation using airborne laser scanner data in a structurally heterogeneous spruce forest. *Photogrammetric Engineering & Remote Sensing*, Vol. 72, No.12, pp. 1369-1378

Zhao, K., and Popescu, S.C. 2007. Hierarchical Watershed Segmentation of Canopy Height Model for Multi-scale Forest Inventory. In Rönnholm, P., Hyyppä, H., Hyyppä, J. (eds). *Proceedings of the ISPRS working group "Laser Scanning 2007 and SilviLaser 2007", ISPRS Volume XXXVI*, Part3/W52. Espoo, September 12-14, 2007, Finland, pp. 436-442.

Zhao, K., Popescu, S., & Nelson, R. (2009). Lidar remote sensing of forest biomass: a scale-invariant approach using airborne lasers. Remote Sensing of Environment, Vol. 113, pp. 182−19.

Zhao, K., and Popescu, S. 2009. Lidar-based mapping of leaf area index and its use for validating GLOBCARBON satellite LAI product in a temperate forest of the southern USA, Remote Sensing of Environment, Vol. 113, pp. 1628–1645

Estimation of aboveground biomass using airborne LiDAR data

ANDREAS JOCHEM †, MARKUS HOLLAUS ‡, MARTIN RUTZINGER §, BERNHARD HÖFLE ¶, KLEMENS SCHADAUER |, BERNHARD MAIER ¤

† alpS-Centre for Natural Hazard Management, Grabenweg 3, 6020 Innsbruck, Austria, (jochem@alps-gmbh.at)

‡ Vienna University of Technology, Institute of Photogrammetry and Remote Sensing, 1040 Vienna, Austria

§ Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, 7500 AE Enschede, The Netherlands

¶ University of Heidelberg, Department of Geography, 69120 Heidelberg, Germany

| Department of Forest Inventory at the Federal Research and Training Center for Forests, Natural Hazards and Landscape, Seckendorff-Gudent-Weg, 1130 Vienna, Austria

¤ Stand Montafon Forstfonds, Montafonerstraße 21, 6780 Schruns, Austria

Abstract

In this study a semi-empirical model that was originally developed for stem volume estimation is used for aboveground biomass (AGB) estimation. The semi-empirical model is based on the relative heights of first echo LiDAR point cloud data and assumes a linear relationship between AGB and canopy volume. However, the usage of point cloud data leads to a computationally demanding task when processing large point cloud datasets for the generation of area-wide AGB maps. In the presented study the effects of using rasterized LiDAR data as input for the AGB model are investigated in order to speed up processing and to make use of the model on large spatial datasets. The canopy volumes are calculated from a Canopy Height Model (CHM). The optimum resolution of the CHM is determined by analyzing the effects of varying cell sizes (1.0 m, 1.5 m, 2.0 m, 3.0 m) on the achievable accuracies. Calibrating the model with rasterized input data having a spatial resolution of 2.0 m instead of using first echo point cloud data leads to a slight increase of the coefficient of determination (R^2 = 0.70 to R^2 = 0.72) and a slight decrease of the standard deviation of the prediction errors. For calibrating the model reference AGB is calculated per sample plot from local forest inventory data by means of averaged weighted (according to tree species and age class composition) extension factors. The influence of using rasterized LiDAR input data on the achievable accuracy of the assessed AGB is investigated for a coniferous dominated study area in Vorarlberg, Austria.

1. Introduction

Aboveground biomass (AGB) is defined as the total amount of aboveground oven dry mass of a tree that is expressed in tons per unit area (Brown 1997). Accurate estimation of AGB, also referred to as dry total biomass, in forested areas provides an indication of the potential energy that is stored in cellulosic material. Gaining knowledge about the spatial distribution of the potential bioenergy is essential for developing sustainable low carbon climate friendly strategies such as the optimization of timber harvesting chains. In contrast to time consuming and expensive field methods remote sensing is capable for mapping area-wide forest inventory (FI) data in a less expensive, fast and accurate way. This procedure is mainly based on the extrapolation of FI data measured at stand or plot level. In recent years Airborne Laser Scanning (ALS), also referred to as Light Detection And Ranging (LiDAR), has been established as a standard technology for the acquisition of high precision topographic data and has been widely used for mapping vegetation and forest inventory data, respectively (Lim et al. 2004, Næsset et al. 2004, Hyyppä et al. 2008). In contrast to multi-spectral satellite

imagery or arial photographs ALS data represent the horizontal and vertical distribution of the forest canopies and does not suffer from saturation in spectral response to dense canopies with high biomass (Zhao et al. 2009).

There are two methodological approaches for utilizing LiDAR data for AGB assessment. The (i) single-tree-based approaches and the (ii) area-based approaches. Both approaches mainly involve the use of empirical or semi-empirical models by using linear or non-linear regression analysis (e.g. Næsset 2004a, Popescu 2007, García et al. 2010). Single tree based approaches require LiDAR data with high point densities (>5 points/m²) and are mostly based on regression models focusing on a relationship between LiDAR derived individual tree parameters (e.g. tree height, crown dimensions) and field based estimates of AGB. Area-based methods can also be used for lower point densities but require an extensive set of reference data. Such methods analyze the vertical distribution of the laser echoes at stand or plot level by deriving various statistical quantities and estimate area-based forest inventory parameters (e.g. mean tree height, basal area, stem volume) and AGB, respectively.

LiDAR based estimations of AGB can be performed by means of both rasterized and point cloud data. Using rasterized data requires an aggregation of the 3D point cloud to 2.5D raster cells, meaning that the canopy surface is represented by a single-valued function. This procedure is accompanied with an irreversible loss of the 3rd dimension but makes processing less time consuming and reduces the amount of the storage size drastically. Hence, using rasterized data for the generation of area-wide digital high resolution maps is computationally more efficient.

In this paper a semi-empirical model (Hollaus et al. 2009) that was originally developed for stem volume estimation is used for AGB estimation. Furthermore, the model is investigated concerning the effects on the achievable accuracies of using rasterized instead of point cloud input data in order to make processing a less computationally intensive task and to apply the model on large spatial datasets. The optimum resolution of the rasterized input data is determined by analyzing the effects of different cell sizes on the accuracy of the assessed AGB. For this study an area of about 560 km² in the western part of Austria is analyzed.

This paper is structured as follows: In Section 2 the study area, the available ALS datasets as well as the estimation of the reference AGB are presented. The methodology including a short description of the semi-empirical model is subject of Section 3. In Section 4 the results are presented and discussed. A conclusion and outlook on future studies are given in Section 5.

Study area and datasets

The coniferous dominated study area is located in the southern part of the Federal State of Vorarlberg (Austria) in the so-called Montafon region, and covers an area of about 560 km². The main tree species in the area are Norway spruce (*Picea abies*) with 96% and fir (*Abies alba*) with 3%. The used ALS data are provided by the Land Survey Adminstration Feldkirch and were retrieved during several flight campaigns in the framework of a Vorarlberg-wide terrain mapping project. The ALS data were acquired under snow-free conditions in the years 2002 to 2004 using Optech Mapper systems (ALTM 1225, ALTM 2050) and Leica ALS-50 scanner. The Optech sensors have a beam divergence of 0.3 mrad and the ALS-50 scanner of 0.33 mrad, which resulted in a mean footprint diameter of 0.33 m and 0.36 m, respectively for the average flying height of 1100 m above ground. The mean point density within the study area varies between 0.9 point/m² and 2.7 points/m².

Besides the original ALS point cloud data a Digital Surface Model (DSM) and a Digital Terrain Model (DTM) with a spatial resolution of 1 m are available. The DTM was generated by using last echoes only and applying the hierarchic robust filter technique as described e.g. in Kraus and Pfeifer (1998). By subtracting the DTM from the DSM a Canopy Height Model (CHM) is produced. The relative height value of each laser point is derived by subtracting the underlying DTM elevation from each laser point.

For the investigated forests, FI data from 500 sample plots, which are regularly distributed in a 350 m grid are available. They were provided by the forest administration Stand Montafon Forstfonds and were collected in the year 2002 using the angle count sampling method (Bitterlich 1948), meaning that the plot areas and number of

sampled trees may vary strongly from sample plot to sample plot. For each sample plot the stem volume per unit area [m³ ha^{-1}] was determined from tree specific parameters, such as tree species, tree height and diameter at breast height (DBH). The possible inaccuracies in the spatial positions between the LiDAR data and the FI data are corrected by performing co-registration as described in Dorigo et al. (2009).

The reference AGB that is used as target variable in this study and is estimated by means of averaged weighted (according to tree species and age class composition) extension factors taken from Weiss et al. 2000. The extension factor is determined for each sample plot separately. The AGB is estimated from the stem volume that is assessed from FI data for each sample plot as described in Hollaus et al. (2009). The first step contains the transformation of the estimated stem volume into dry stem biomass by using tree specific average raw density factors (Weiss et al. 2000, p.29). In the following step the dry stem biomass is converted in dry total tree biomass by means of extension factors given in Weiss et al. (2000, p.31).

2. Methods

Semi-empirical model

The semi-empirical model is based on the assumption that there is a linear relationship between AGB and the canopy volume (V_{can}) that is defined as the entire volume between the terrain surface and the topmost tree surface. V_{can} is determined for circular areas A around the center coordinates of the FI sample plots and is based on the relative heights of the first echoes. The relative heights are derived by subtracting the DTM height from the absolute heights of the first echoes. In order to take the height-dependent differences in canopy structure into account the relative height above terrain surface of each first echo point is used to classify the points into m different height classes, whereas all points having a relative height value of less than 2.0 m are classified as points being reflected from the ground, stones or bushes (Naesset, 2004b) and are not included into the canopy volume calculation. Based on the findings of the study from Hollaus et al. (2009) four canopy height classes having a canopy height interval of 10 m are most suitable for the calculation of the canopy volume. $V_{can,1}$ ranges between 2 m and 12 m, $V_{can,2}$ ranges between 12 m and 22 m, $V_{can,3}$ ranges between 22 m and 32 m and $V_{can,4}$ contains all first echoes having a relative height greater than 32 m. $V_{can,\,i}$ is calculated as:

$$V_{can,i} = p_{fe,i} * ch_{mean} \qquad (1)$$

where ch_{mean} is the mean canopy height of all first echoes within the corresponding canopy height class. $p_{fe,i}$ is the relative portion of first echo points (between 0 and 1) within the canopy height range i. The linear regression model is formulated as:

$$AGB = \sum_{i=1}^{m} \beta_i * V_{can,i} \qquad (2)$$

where β_i are the unknown model coefficients that can be interpreted as the fraction of the corresponding canopy volume that is occupied by AGB.

Determining of the optimum sample plot size

The estimation of the reference AGB is based on FI data that was collected using the angle count sampling method (Bitterlich 1948), meaning that the reference AGB is not related to a defined sample plot size. The challenge is to find the optimum sample plot size in order to allow a proper comparison of the ALS data with the FI data. Therefore, an approach introduced by Hollaus et al. (2007) is chosen that analyzes different sample plot

sizes with radii ranging from 8.0 m to 16.0 m. This procedure is based on all co-registered sample plots containing at least 90% coniferous trees. The 90% coniferous trees threshold is introduced to avoid effects of different flight dates (winter/summer) on the calibration of the model for different tree species i.e. coniferous and deciduous trees. The sample plot size leading to the highest coefficient of determination (R^2) and the lowest standard deviation (SD) of the residuals derived from cross-validation (Section 3.4) is taken for further analyses.

Effects of rasterizing input data

Generating biomass maps for large spatial datasets based on first echo point cloud data is very time consuming and computationally intensive. Using rasterized data as input for the semi-empirical model speeds up processing and overcomes the disadvantages related to point cloud data (huge amount of storage size, computationally intensive task when performing spatial queries within the point cloud) when computing large area AGB maps. The effects of using rasterized input data on the achievable accuracies are analyzed by using a CHM for deriving the canopy volumes. The CHM is generated by aggregation of all first echo points into a regular grid, whereas the maximum relative elevation is chosen as cell value. Cells containing no laser point at all, obtain an elevation value of zero and consequently are not considered for the calculation of the canopy volumes. The optimum spatial resolution of the CHM is determined by investigating the effects of different cell sizes (1.0 m, 1.5 m, 2.0 m, 3.0 m) on R^2 and SD of the prediction errors, respectively.

Validation of the semi-empirical model

The predictive accuracy of the calibrated model is assessed by performing a leave one out cross validation procedure. This means that the model is fitted *n* times (*n* is the number of available sample plots), whereas for each step one sample plot is excluded and serves for the calculation of the prediction error. The remaining sample plots are used for the calibration of the model. Finally, one gets *n* prediction errors that are used for the calculation of statistical parameters such as minimum, maximum, mean and standard deviation.

3. Results and discussion

Determination of the optimum sample plot radius

The 90% coniferous trees threshold resulted in 450 out of 488 successfully co-registered sample plots that are taken as input for the determination of the optimum circular sample plot size. A sample plot radius of 12.0 m results in the highest R^2 (0.66) and the lowest SD of the prediction errors (109.0 tha^{-1}).

Table 4: Determination of the optimum circular sample plot size. Various radii are analyzed according to their R^2 and SD of the prediction errors.

Sample plot radius [m]	8.0	10.0	12.0	14.0	16.0
R^2	0.60	0.64	0.66	0.64	0.61
SD [tha^{-1}]	120.2	111.4	109.0	111.2	115.7

This procedure is followed by the selection of sample plots containing all trees measured in the field for reference AGB estimation within a radius of 12.0 m. This results in a selection of 196 out of 450 sample plots that are taken for the calibration of both the point cloud and the raster based semi-empirical model.

Calibration of the point cloud based model

The calibration of the point cloud based model is based on the 196 selected sample plots (Section 4.1) using the optimum sample plot radius of 12.0 m and four canopy height classes having a height interval of 10.0 m. The calibrated model achieved an R^2 of 0.70 and a SD of the prediction errors of 87.6 tha^{-1} (37.0 %). The scatter plot of the reference AGB versus the AGB estimated from LiDAR data is shown in figure 1.

Hollaus et al. (2009) achieved R^2 values up to 0.86 for the estimation of stem volume. The deviations to the presented approach can be explained by several reasons: (i) they used stem volume instead of AGB, (ii) the conversion of stem volume per sample plot to the reference AGB is accompanied with several uncertainties and (iii) the different sizes of the study areas. The study area in Hollaus et al. (2009) is about three times less in size and results in the usage of 103 reference sample plots. Investigating a larger study area means both an increase of the heterogeneity of the LiDAR data (e.g. varying flying heights, acquisition dates, different point densities) and an increase of the spatial variability of the forest stand properties. The latter is due to the alpine topography of the Montafon region that influences forest growing conditions.

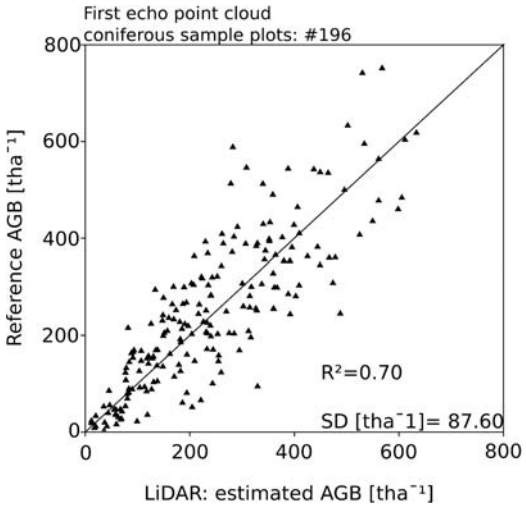


Figure 1: Scatter plot of reference AGB versus AGB estimated from first echo LiDAR data.

Effects of rasterizing input data

Calibrating the model with canopy volumes derived from rasterized instead of point cloud data does not change the achievable accuracy significantly. A slight increase of R^2 (0.72) as well as a slight decrease of SD of the prediction errors (84.11 tha-1) can be observed when using a CHM having a cell size of 2.0 m. This is considered to be the optimum spatial resolution when generating AGB maps for the investigated study area.

The effects of varying cell sizes, ranging from 1.0 m to 3.0 m on the accuracy statistics and the β coefficients are shown in table 2. Concerning R^2 and SD of the prediction errors the AGB model seems to be very robust against the varying cell sizes of the CHM. Analyzing the β coefficients confirmed the findings of Hollaus et al. (2009) where canopy heights between 22 m and 32 m are the highest contributors to growing stock and AGB, respectively. This applies to both the first echo LiDAR based (Section 4.2) and the CHM based AGB model. However, the coarser the spatial resolution of the CHM gets, the lower the values of the β coefficients $\beta 1, \beta 2, \beta 3$.

This can be explained by the overestimation of the canopy volume with increasing cell size as small gaps in the forest canopy are neglected. The variation of the corresponding $\beta 4$ values is less distinct. The CHM is generated by using the maximum relative elevation within one cell as the corresponding cell value, meaning that points of lower canopy height classes are not considered anymore if one or more points within a cell fall into a upper canopy height class. Hence, the bigger the cell size the lower the fraction of the lower canopy height classes to the reference AGB. The minor change of the $\beta 4$ values might be due to the small amount of first echo points falling into the canopy height class greater than 32 m.

Table 5: Accuracy statistic of the AGB model when using rasterized and point-based input data. R², SD of the prediction errors and the estimated ß coefficients are shown.

	Point Cloud	Resolution of CHM			
Parameters	**(0.9 – 2.7 p/m²)**	**1.0 m**	**1.5 m**	**2.0 m**	**3.0 m**
R^2	0.70	0.70	0.70	0.72	0.71
SD [tha^{-1}]	87.60 (37.0%)	88.84 (37.5%)	88.60 (37.2%)	84.11 (34.4%)	84.6 (34.6%)
$\beta 1$	$7.71*10^{-4}$	$6.97*10^{-4}$	$4.56*10^{-4}$	$2.70*10^{-4}$	$1.38*10^{-4}$
$\beta 2$	$19.91*10^{-4}$	$19.19*10^{-4}$	$17.34*10^{-4}$	$14.62*10^{-4}$	$10.98*10^{-4}$
$\beta 3$	$29.75*10^{-4}$	$28.36*10^{-4}$	$25.40*10^{-4}$	$22.80*10^{-4}$	$18.87*10^{-4}$
$\beta 4$	$15.87*10^{-4}$	$15.85*10^{-4}$	$15.84*10^{-4}$	$15.46*10^{-4}$	$15.48*10^{-4}$

The scatter plots of the reference AGB versus the AGB estimated from the canopy volumes derived from a CHM are shown in figure 2.

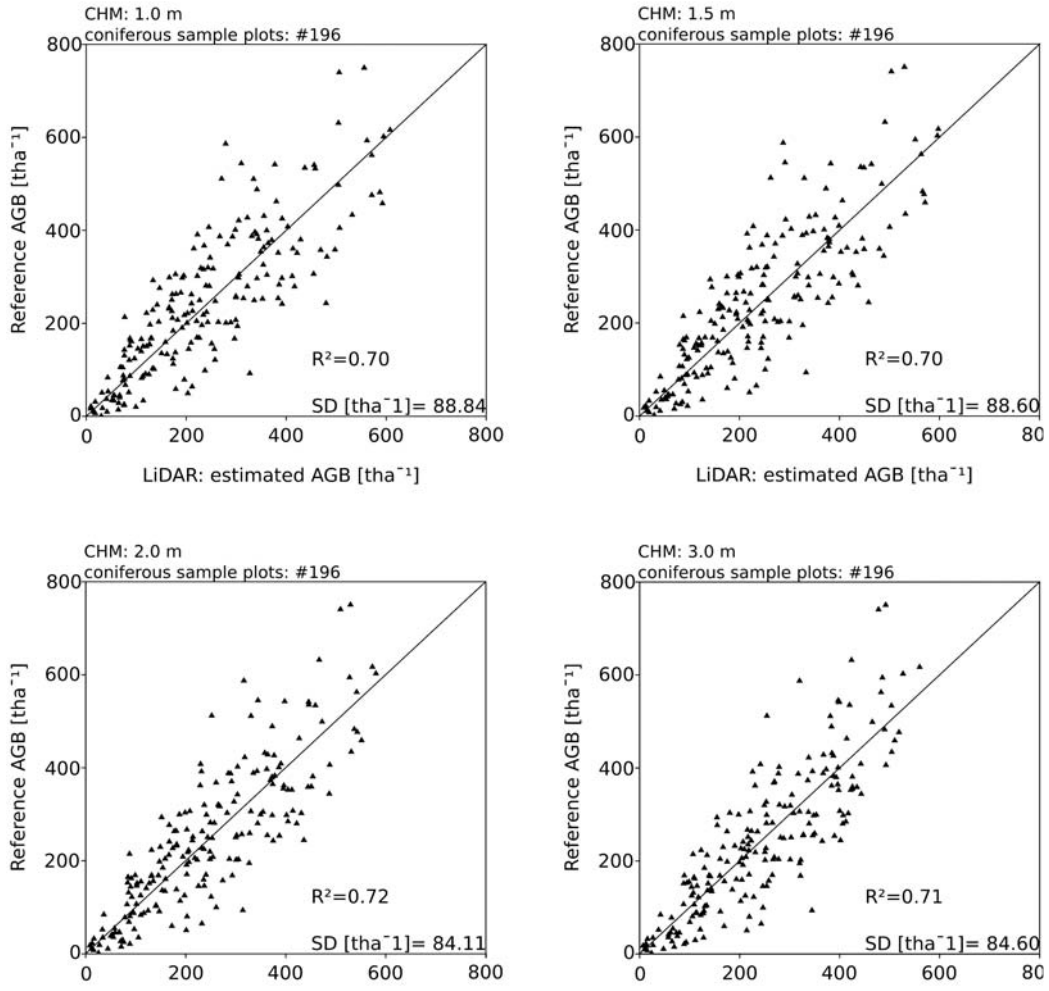


Figure 2: Scatter plots showing AGB derived from in-situ measurements versus AGB estimated from canopy volumes derived from rasterized input data. The x-axis represents the AGB of the sample plots estimated by the AGB model. The y-axis shows the reference AGB calculated from in-situ measurements.

4. Conclusion and outlook

In the presented study a semi-empirical model that was originally developed for stem volume estimation is investigated concerning its reliability for AGB estimation. The semi-empirical model assumes a linear relationship between AGB and canopy volume that is derived from first echo LiDAR point cloud data. Furthermore, the effects of using rasterized LiDAR data as input for the AGB model are analyzed in order to make processing a less computationally demanding task when applying the model on large spatial datasets. Therefore, the canopy volume is calculated from the 3D first echoes and the CHM, respectively. The effects of varying cell sizes on the achievable accuracy are investigated to find the optimum spatial resolution of the CHM-based canopy volume calculation. The results show that the semi-empirical model can also be used for AGB estimation of a spruce-dominated alpine forest and that AGB maps can be generated by means of rasterized input data. The usage of first echo point cloud data with leads to a R^2 of 0.70 and a SD of the prediction errors of 87.6 tha^{-1}. Calibrating the model with canopy volumes derived from a CHM does not change R^2 and SD of the prediction errors significantly. A spatial resolution of 2.0 m leads to the highest R^2 (0.72) and the lowest SD of the prediction errors (84.11 tha^{-1}).

Future studies will concentrate on the application of the AGB model on regions that are characterized by both a wider range of tree species and higher point densities. Due to the different crown properties of deciduous and coniferous trees it is expected that the consideration of tree species (deciduous versus coniferous trees) in the AGB model will increase the accuracy of the assessed AGB. Furthermore, the improvement of the calculation of the reference AGB will be in the focus of future research including the consideration of the biomass compartments (stem, branches, needles, foliage).

Acknowledgements

The authors would like to thank the Landesvermessungsamt Feldkirch (Austria) for granting the use of the ALS data, and the Stand Montafon Forstfonds for supplying the forest inventory data. This study was carried out within the project LASER-WOOD (822030) which is funded by the Klima- und Energiefonds in the framework of the program NEUE ENERGIEN 2020.

References

Bitterlich, W., 1948. Die Winkelzählprobe. *Allgemeine Forst und Holzwirtschaftliche Zeitung* 59, pp. 4–5.

Brown, S., 1997. Estimating biomass and biomass change in tropical forests: a primer. In *FAO Forestry Paper,* Vol. 134.

Dorigo, W., Hollaus, M., Schadauer, K., Wagner, W., 2009. An application-oriented automated approach for co-registration of forest inventory and airborne laser scanning data. *International Journal of Remote Sensing*, 31(5), pp. 1133 – 1153.

García M., Riaño, D., Chuvieco, E., Mark Danson, F., 2010. Estimating biomass carbon stocks for a mediterranean forest in central spain using LiDAR height and intensity data. *Remote Sensing of Environment* 114(4), pp. 816–830.

Hollaus, M., Wagner, W., Maier, B., Schadauer, K., 2007. Airborne laser scanning of forest stem volume in a mountainous environment. *Sensors* 7(8), pp. 1559–1577.

Hollaus, M., Wagner, W., Schadauer, K., Maier, B., Gabler, K., 2009. Growing stock estimation for alpine forests in Austria: a robust lidar-based approach. *Canadian Journal of Forest Research* 39(7), 1387-1400.

Hyyppä, J., Hyyppä, H., Leckie, D., Gougeon, F., Yu, X., Maltamo, M., 2008. Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing* 29(5), pp. 1339–1366.

Kraus, K., Pfeifer, N., 1998. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing* 54(4), pp. 193–203.

Lim, K., Treitz, P., 2004. Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. *Scandinavian Journal of Forest Research* 19, pp. 558– 570.

Næsset, E., Gobakken, T., Holmgren, J., Hyyppä, H., Hyyppä, J., Maltamo, M., Nilsson, M., Olsson, H., Persson, A., Söderman, U., 2004. Laser scanning of forest resources: the nordic experience. *Scandinavian Journal of Forest Research* 19(6), pp. 482–499.

Næsset, E., 2004a. Estimation of above- and below-ground biomass in boreal forest ecosystems. In *International Archives of Photgrammetry, Remote Sensing and Spatial Information Sciences*. Vol. XXXVI (8/W2). Freiburg, Germany, pp. 145–148.

Næsset, E., 2004b. Practical large-scale forest stand inventory using a small foot print airborne scanning laser. *Scandinavian Journal of Forest Research* 19(2), pp. 164–179.

Popescu, S., 2007. Estimating biomass of individual pine trees using airborne LiDAR. *Biomass and Bioenergy* 31, pp. 646–655.

Weiss, P., Schieler, K., Schadauer, K., Radunsky, K., Englisch, M., 2000. Die Kohlenstoffbilanz des Österreichischen Waldes und Betrachtungen zum Kyoto Protokoll. *Tech. Rep.* 106, Umweltbundesamt, Wien.

Zhao, K., Popescu, S., Nelson, R., 2009. LiDAR remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. *Remote Sensing of Environment* 113(1), pp. 182–196.

Effect of sampling intensity on the accuracy of species-specific volume estimates derived with aerial data: A case study on five privately owned forest holdings

EVELIINA KALLIO†, MATTI MALTAMO† & PETTERI PACKALÉN†*

†School of Forestry, Faculty of Science and Forestry, University of Eastern Finland, P.O. Box 111, FI-80101 Joensuu, Finland

Abstract

Aerial data based forest inventories should be able to provide accurate species-specific information on forest resources for practical forest management purposes. In Finland forest inventories based on aerial images and airborne laser scanning are usually carried out with nearest neighbor imputation of field sample plots. This study examines the effect of reduced sampling intensity on the errors of species-specific volume estimates on five forest holdings. The results indicate that the accuracy of the estimated *Picea abies* volume does not decrease until reducing the sample size under 200. The estimation accuracy of *Pinus sylvestris* and deciduous tree species volumes decreases remarkably when the sample intensity is decreased. The results stress the importance of careful sample design when conducting aerial data based species-specific forest inventories on privately owned forest holdings.

1. Introduction

Forest inventory augmented with airborne laser scanning (ALS) and aerial images is in operational use in Nordic countries. It can provide significant advantages when accurate information on forest resources is needed. Besides being accurate, a practical forest inventory tool should also be cost-efficient in providing tree species-specific forest information.

In forest planning the gathered data is typically used as an input to a decision support system. Incorrectly gathered, measured or predicted data may lead to non-optimal decisions, errors in the optimization of forest treatment schedules and thus, a non-optimal forest plan. Generally, the more precisely the present state of a forest is described the more accurately the future development can be estimated. On the other hand, the more precisely the description is wanted to be done the more measurements should be carried out. The cost-efficiency of the inventory method can be considered as an optimal sampling intensity which minimizes the errors in the data and the costs of the inventory but maximizes the accuracy of the information.

A sample of accurately measured field plots required in ALS based practical forest inventories forms a significant portion of the total inventory costs. Field measurements are utilized with variables calculated from height and density distributions of low-resolution ALS data (pulse density < 1 pulse/m^2) to predict stand-level attributes (e.g. volume) for the whole inventory area. This is usually done with regression models.

In Finland ALS based forest inventories are usually carried out with a Nearest Neighbor (NN) imputation in which aerial images and ALS data are used in the estimation of species-specific attributes (see Packalén and Maltamo (2006, 2007)). Species-specific growing stock estimates are the basis for Finnish forest management decision support systems. Traditionally the forest information is acquired with a stand-wise inventory method, in which

forest characteristics are estimated using angle count sampling and visual assessment. The gathered data is utilized in forest holding specific forest planning. Privately owned forest holdings and stands covered by them may vary in their size and characteristics a lot due to varying forest treatment histories. Species-specific estimation accuracy obtained with ALS based methods being better than that obtained with traditional inventory methods, it is possible to get growing stock estimates accurate enough for practical forestry purposes (Packalén 2009). The accuracy and the costs of the method are still dependent on the sample design, the amount of field observations and also on the characteristics of the forest in question.

Sampling design of carrier data in a species-specific estimation should be carefully considered since the discrimination of tree species is often burdensome (e.g. Packalén 2009). Species-specific estimation is especially difficult in mixed stands in which several tree storeys exist. To the very best of our knowledge, no studies have yet investigated the impact of the sample intensity on the precision of species-specific forest inventory augmented with ALS and aerial image data. Neither has any study focused on the precision of these estimations on privately owned forest holdings. These are relevant interests in Finland where forest inventories and management plans on privately owned forest holdings are being carried out with the aid of the remote sensing techniques and NN imputation.

This study examines the effect of reduced sampling intensity on the errors of species-specific growing stock volume estimates in ALS and aerial image based forest inventory on five privately owned forest holdings.

2. Material

2.1 Study area

Experiments were carried out at about 46 000 hectares area at Eastern Finland. The forests at the study area are mainly owned by private forest owners. They are mostly Norway spruce (Picea abies) dominating fertile forests whereas Scots pine (Pinus sylvestris) dominates at the dryer forest sites. Also birches (mainly Betula pendula) may be dominating but usually deciduous trees (Betula pubescens, Alnus incana, Alnus glutinosa, Sorbus aucubaria, Populus tremula) occur as admixtures.

2.2 Modeling data

Totally 431 field sample plots were measured in summer 2008. Sampling was carried out as it is designed in the Finnish aerial data based forest inventories on privately owned forests. The design attempts to mimic Finnish NFI with varying cluster and shorter plot distances, stratification and subjectively allocated additional measurements. The stratification is done on the basis of a priori information on the forest characteristics and the spatiality of the stands locating on the study area. Strata were (1) forest site type, (2) development class (young growing stand, mature growing stand, regeneration maturity stand), (3) dominating tree species (pine, spruce, deciduous trees), (4) basal area and (5) mean diameter. The used design aimed to get a good non-probability sample which also includes the extreme kind of forests (table 1).

The circular sample plots (radius 9 m) were prior-located but when measuring the plots in the field the actual center of each plot was located accurately with differential GPS. Diameter at the breast height (dbh), tree storey class (dominating or dominated) and tree species were measured for each tree with a dbh greater than 5 cm. Height was measured for the basal area median tree of each species and storey class by plots. The height measurements were used to construct linear mixed-effect height models (see details below). Species-specific volumes were calculated with Laasasenaho (1982) taper curves which are based on the tree height and the diameter at the breast height. Individual tree volumes were aggregated on the plot-level.

Table 1. *Statistics of the volumes (m^3ha^{-1}) of the sample plots by dominating tree species and development classes. N = number of observations, SD = standard deviation, Min = minimum, Max = maximum.*

		Development class			
		Young growing stand	Mature growing stand	Regeneration maturity stand	All
	N	154	150	127	431
Pine					
	Mean	43.66	47.57	76.21	54.61
	SD	51.33	72.47	119.19	84.57
	Min	0.31	0.27	6.57	0.27
	Max	193.37	332.18	588.92	588.92
Spruce					
	Mean	53.45	98.64	206.62	114.31
	SD	58.6	104.2	146.71	123.56
	Min	0.23	0.25	0.25	0.23
	Max	231.19	340.17	637.75	637.75
Deciduous trees					
	Mean	28.81	59.04	46.56	44.57
	SD	34.74	70.04	79.68	64.59
	Min	0.29	0.31	0.29	0.29
	Max	168.82	256.39	390.10	390.10
Total					
	Mean	125.92	205.25	329.40	213.49
	SD	43.25	67.59	120.63	115.19
	Min	33.29	57.32	94.21	33.29
	Max	248.45	384.96	754.10	754.10

2.3 Validation data

Independent stand-level validation data was collected in summer 2009. Data was gathered from an area confining to about 2 000 hectares in the Southern border of the whole study area. The network of 183 field plots distributed over five privately owned forest holdings and over 30 stands. Sampled stands represent all development classes and tree species admixtures existing on the area. A systematic sample plot inventory was carried out within each stand at intervening distances of 20 m to 100 m depending on the stand size and shape. The stand size being rather small (average 1.3 ha) the number of the plots within a stand varied between 3 to 10 plots. The plots were circular with radius of 8 m or 10 m if the stand was near to the rotation age. The stands eventually used in the study consist of the accurately measured plots alone. The size of the stands vary between 0.06 ha and 0.2 ha.

Table 2. *Statistics of the volumes (m^3ha^{-1}) of the post-stratified validation stands by dominating tree species and development classes. n_h = number of stands from stratum h included in the sample, P = number of strata, $Mean_{ST}$ = sample mean of the population, SD_{ST} = weighted mean of the standard deviations among sampling units within each stratum, Min = minimum of the observed stands, Max = maximum of the observed stands.*

		Development class			
		Young growing stand	Mature growing stand	Regeneration maturity stand	All
	n_h	9	11	10	30
	P	6	7	7	20
Pine					
	$Mean_{ST}$	29.9	50.61	125.88	78.88
	SD_{ST}	9.31	3.39	35.23	19.28
	Min	1.5*	0.16*	27.83*	0.16*
	Max	94.95	177.88	356.63	356.63
Spruce					
	$Mean_{ST}$	51.67	138.71	155.49	122.98
	SD_{ST}	12.28	5.46	11.80	10.13
	Min	1.04	1.08	8.88	1.04
	Max	113.67	349.86	267.80	349.86
Deciduous trees					
	$Mean_{ST}$	40.49	47.33	35.87	40.35
	SD_{ST}	3.40	12.55	3.44	6.01
	Min	0.99	0.69	2.61*	0.69*
	Max	116.71	93.08	141.65	141.65
Total					
	$Mean_{ST}$	122.07	236.66	317.24	242.21
	SD_{ST}	5.21	9.06	41.47	22.59
	Min	76.21	95.62	195.93	76.21
	Max	148.95	389.50	529.10	529.10

*The true minimum is zero

The location of the centre of each sample plot was determined with differential GPS. Diameter at the breast height (dbh), tree storey class (dominating or dominated) and tree species were measured for each tree with a dbh greater than 5 cm. Height was measured for the first observed tree of each species and storey class by plots. The height measurements were used to construct linear mixed-effect height models for each species and storey class with Näslunds (1937) formulation. The height models were done with one dataset consisting both the modeling and the testing measurements. Species-specific volumes were calculated with Laasasenaho (1982) taper curves. Individual tree volumes were aggregated on the stand-level.

The validation stands were post-stratified. Totally 21 stratum were defined according to the tree species and development classes. Tree species classes are: stand has only (a) pine, (b) spruce or (c) deciduous trees, stand has either (d) spruce and pine, (e) pine and deciduous or (f) spruce and deciduous so that dominated tree species covers at least 10% of the total volume, or (g) stand has all the tree species so that the two dominated tree species cover at least 10% and 5% of the total volume of the stand. Development classes are (i) young growing forests, (ii) mature growing forest and (iii) regeneration maturity forests. Measured stands distribute into 20 strata as such that each strata include one or two stands. The basic statistics of the stratified data can be seen in table 2.

Stratum estimates for the measured species-specific volumes are calculated with the following equations:

$$W_h = \frac{N_h}{N} \tag{1}$$

$$Mean_{ST} = \sum_{h=1}^{P} \sum_{i=1}^{n_h} W_h y_{ih} n_h^{-1} \tag{2}$$

$$SD_{ST} = \sum_{h=1}^{P} W_h S_{y_h}^2 \tag{3}$$

$$S_{y_h} = \frac{1}{n_h} \left[\sum_{i=1}^{n_h} y_{ih}^2 - \frac{\left(\sum_{i=1}^{n_h} y_{ih}\right)^2}{n_h} \right] \tag{4}$$

where P = number of strata into which the population is divided; now P=20, W_h = weight of stratum h; now h=1, …,20, N_h = number of the sampling units in stratum h within the whole population, N = number of the sampling units in the whole population; $N = \sum_{h=1}^{P} N_h$, n_h = number of the sampling units included in the sample from stratum h, $Mean_{ST}$ = sample mean of the population, y_{ih} = an observed value of the variable y on sampling unit i in stratum h; i=1, …,n_h, SD_{ST} = weighted mean of the standard deviations among sampling units within each stratum, S_{yh} = standard deviation among sampling units within stratum h.

2.4 Remote sensing data acquisition

The ALS data from Karttula study area were collected in August 2009 using Optech 3100 laser scanning system operating at an altitude of 2000 m above ground level (agl) using a half-angle of 15 degrees and side overlap of about 20 percent. This resulted in a swath width of 1070 metres and a nominal sampling density of about 0.64 measurements per square metre. The divergence of the laser beam (1064 nm) was 0.3 mrad, which produced a

footprint of 60 cm at ground level. The used laser scanning system produces four types of echoes which were re-classified as the first and last pulse data so that only echoes were in both classes. The ALS data were further used to generate a digital terrain model (DTM) using one-metre pixel size by the method explained in Axelsson (2000).

The aerial photographs were captured with Vexcel UltraCamD digital aerial camera during 5 different days on July and September 2009. Images were taken at an altitude of about 5630 m agl with the sidelap of 30%. As the camera actually consists of eight independent cameras it captured four high-resolution panchromatic images and four lower-resolution multi-spectral images from red, green, blue and near-infrared (NIR) portions of the spectrum. The final image is produced by applying image fusion technique called pan-sharpening, which combines these separate images to one. Pan-sharpened images were orthorectified to a pixel size of 0.5 m using the DTM generated from the ALS data.

3. Methods

3.1 Independent aerial data variables

Independent variables were calculated from the ALS data and the aerial images. ALS metrics were determined by using the pulses falling inside plot boundaries and hitting on tree canopy (height at agl > 0.5 m). Laser pulse height and density distributions were created separately with the first and the last pulses for each plot. Height and density percentiles for 10%, 30%, 50%, 70%, 90% (f_h_{10},…,f_h_{90}; f_p_{10},…,f_p_{90}) and for 20%, 40%, 60%, 80%, 95% (l_h_{20},…,l_h_{95}; l_p_{20},…,l_p_{95}) were computed with the first and the last pulses, respectively. Also the proportions of the hits on the forest canopy and ground (limit 0.5 m) were calculated for both the first and the last pulse data. All the calculated laser variables were included in the model formulation phase.

The aim of using aerial image features was to improve the separation of the tree species. Aerial photo pixels that centre fall inside the plot boundaries of a plot were used to calculate spectral statistics and grey-level co-occurrence matrix for that plot according to the principles presented by Haralick et al. (1973). The calculation of textures on the basis of grey-level co-occurrence matrix was done with varying rescaling classes and lag distances using a pixel size of 0.5 m, and as an average of all directions (0°, 45°, 90°, and 135°). All the variables were calculated by means of four image bands. The features were further analysed with regard to the correlations between the variables themselves and between them and the tree species. The ability of the aerial image variables to discriminate spruce and pine, and on the other hand detect deciduous trees, was tested also with the aim of discriminant-analysis. Variables selected on the basis of correlations and discriminant-analyses were imported into the NN model formulation phase.

3.2 Dependent variables

Volume (m^3ha^{-1}) estimates for pine, spruce and deciduous trees were predicted by means of the independent aerial data features. The estimation was done on the plot-level and further aggregated on the stand-level.

3.3 K-MSN model for volume estimation

Modeling and variable selection were done as explained in details in Packalén and Maltamo (2006). Model was of multivariate type i.e. volumes were estimated simultaneously with basal areas, stem numbers, mean heights and diameters by tree species. Complete plot-level model consisted of totally 19 independent variables. Six of the variables were calculated on the basis of the aerial images and the rest with the ALS data (table 3). Because

the variable selection and the model formulation is quite burdensome and time-consuming the variable selection was executed only once with the complete carrier data.

Table 3. *Predictor variables in the final k-MSN model and their explanations.*

Variables based on aerial images	Variables based on laser pulse distributions
$mean_{BLUE}$, $p40_{BLUE}^2$, $p50_{NIR}$, *ln*($dent_{RED}$), $sentro_{GREEN}$, $\sqrt{svar}_{NIR}$	*ln*(f_veg), 1/f_h_{avg}, *ln*(f_h_{90}), 1/f_p_{30}, $\sqrt{f\_p_{50}}$, *ln*(l_veg), *ln*(l_h_{std}), $\sqrt{l\_h_{40}}$, l_h_{60}, *ln*(l_h_{95}), l_p_{20}, l_p_{95}
The subscript denotes the band, mean = mean intensity, dent = difference entropy, sentro = sum entropy, svar = sum of squares or variance, p40 refers to the pixel value corresponding to the percentile at which 40% of the value distribution has accumulated.	In the variable names f or l denote to first and last pulses, respectively, veg = proportion of the pulses hit on the vegetation, h_{avg} = average of the pulse heights, h_{90} refers to the height at which 90% of the height distribution has accumulated, p_{30} refers to the canopy density corresponding to the proportion of pulses hits above the 30% height quantile.

3.4 Reduction of the sample intensity and validation

Several sample intensities were tested with Monte Carlo simulations with 1000 iterations. Tested sample intensities were 35, 50, 100, 150, 200, 250, 300, 350, 400, 430 and 431 plots. Plots for each sample and iteration were selected randomly from the complete data set. Selected sample plots were used as carrier data in the k-MSN imputation to search for the weights for the model predictors and to estimate the species-specific volumes. Estimation was carried out at the plot-level and further aggregated at the stand-level by calculating the arithmetic mean of the plots inside each stand. Accuracies of the estimation with the reduced sample intensities are calculated on the stratified sampling basis:

$$\mathrm{RMSE}_{ST} = \left(\sum_{j=1}^{M} \sum_{h=1}^{P} \sum_{i=1}^{n_h} W_h (\hat{y}_{ihj} - y_{ihj})^2 \, n_h^{-1} \right) M^{-1} \tag{5}$$

$$\mathrm{bias}_{ST} = \left(\sum_{j=1}^{M} \sum_{h=1}^{P} \sum_{i=1}^{n_h} W_h (\hat{y}_{ihj} - y_{ihj}) \, n_h^{-1} \right) M^{-1} \tag{6}$$

, where m = number of simulations, p = number of stratum, n_h = number of stands inside stratum h, W_h = weight of stratum h, y_{ihj} = observed volume of stand i in stratum h in simulation j, $\hat{y}_{ihj}$ = estimated volume of stand i in stratum h in simulation j.

4. Results

The results show that the accuracy of the volume of the main tree species, spruce, doesn't decrease remarkable until reducing the sample size under approximately 200 (see figure 1(a) and table 4). The RMSE of the total volume behave similarly (figure 1(a)) but is slightly lower with all the tested sample intensities. The errors of the volume of the other tree species increases, on the other hand, quite rapidly when the sample intensity is

decreased. The error level of the estimated volumes is surprisingly high: RMSEs of pine and deciduous trees volumes are even in the most optimum case about 74% and 91%, respectively (table 4). The RMSE of the spruce and total volumes are about 33% and 21% on their best, respectively.

Table 4. *Root mean square error (RMSE, m^3ha^{-1}), bias (Bias, m^3ha^{-1}) and their relative values (RMSE%, Bias%) for the species-specific volume estimates on the forest holding level.*

Sample intensity	RMSE		Bias		RMSE		Bias	
	m^3ha^{-1}	%	m^3ha^{-1}	%	m^3ha^{-1}	%	m^3ha^{-1}	%
	Spruce				*Deciduous trees*			
431	44.26	35.99	-14.97	-12.17	36.82	91.24	15.04	37.26
430	43.55	35.41	-14.71	-11.96	36.56	90.60	15.06	37.32
400	42.37	34.46	-14.22	-11.57	36.87	91.38	15.26	37.83
350	41.38	33.65	-13.19	-10.72	38.72	95.96	15.64	38.75
300	41.10	33.42	-11.84	-9.63	40.78	101.07	16.01	39.67
250	41.72	33.56	-11.05	-8.98	42.23	104.64	16.24	40.25
200	41.85	34.03	-10.00	-8.14	43.30	107.30	16.05	39.76
150	43.16	35.09	-8.62	-7.01	43.70	108.31	15.47	38.34
100	46.78	38.04	-7.17	-5.83	43.98	108.98	14.44	35.79
50	69.62	56.61	-6.46	-5.25	44.87	111.19	9.51	23.57
35	97.85	79.57	-6.52	-5.30	46.67	115.65	6.48	16.05
	Pine				*Total*			
431	58.66	74.36	-21.87	-27.72	50.80	20.97	-21.80	-9.00
430	60.11	76.20	-22.44	-28.45	52.09	21.51	-22.09	-9.12
400	62.75	79.55	-23.05	-29.23	53.69	22.17	-22.02	-9.09
350	66.32	84.07	-24.62	-31.21	55.10	22.75	-22.17	-9.15
300	70.01	88.75	-25.94	-32.88	56.80	23.45	-21.77	-8.99
250	74.10	93.93	-27.03	-34.26	59.29	24.48	-21.83	-9.01
200	77.36	98.07	-27.67	-35.08	61.25	25.29	-21.63	-8.93
150	81.51	103.33	-28.33	-35.91	63.69	26.29	-21.47	-8.86
100	87.04	110.67	-29.49	-37.38	68.77	28.39	-22.22	-9.17
50	96.14	121.87	-28.05	-35.55	78.07	32.23	-24.99	-10.32
35	105.07	133.20	-24.78	-31.42	102.86	42.46	-24.83	-10.25

The biases of the estimates are significant (table 4 and figure 1(b)). The biases of the spruce and pine volumes are between -6.52 and -14.97 m^3ha^{-1}, and between -21.87 and -29.49 m^3ha^{-1}, respectively. The biases of the deciduous trees volume are between 16.05 and 40.25 m^3ha^{-1}. The biases imply that the volumes of spruce and pine are averagely underestimated and the volume of deciduous trees is overestimated on the forest holdings. The biases of the total volume are between -21.02 and -24.99 m^3ha^{-1}. As such, the total volume is also underestimated. An unexpected result was that the amount of bias increases when the sample intensity increases (see figure 1(b) and table 4).

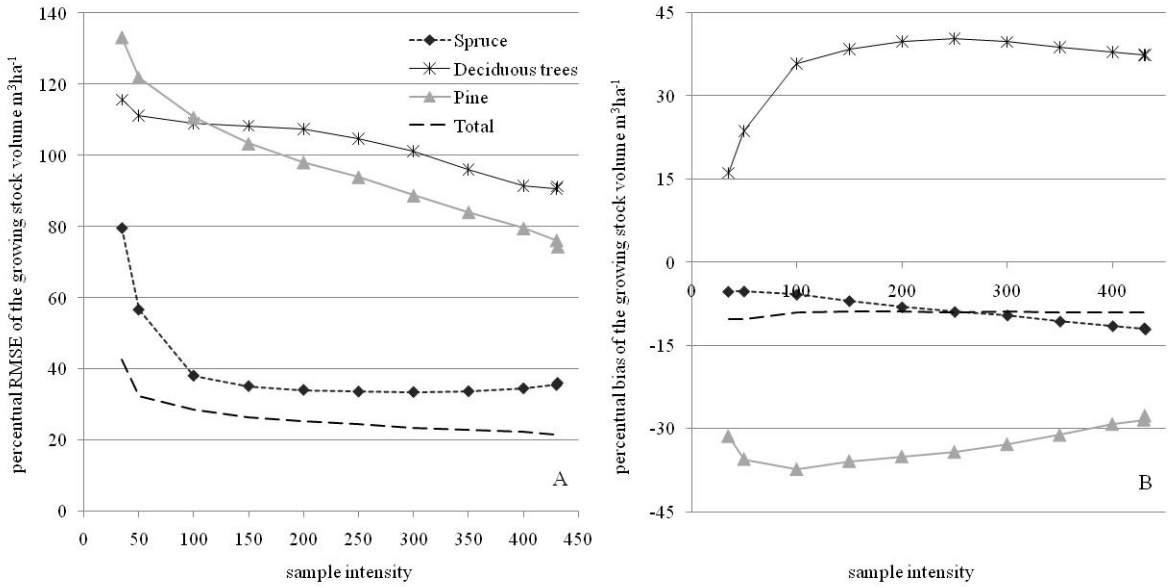


Figure 1. *Relative root mean square error (RMSE%) (a) and relative bias (b) of the species-specific volume estimates (m^3ha^{-1}) on the forest holding level.*

5. Discussions

Timber volume calculations and growth models, for example, are based on species-specific measurements in Finnish forest management decision support system. Thus, correct tree species recognition and species-specific estimation of forest characteristics is essential when conducting a forest inventory. In the present study we tested the effect of reduced sampling intensity on the errors of species-specific volume estimates in ALS and aerial image based forest inventory on five privately owned forest holdings.

In previous studies the k-MSN imputed species-specific estimates have been rather accurate. Packalén and Maltamo (2006) estimated volume for different tree species at plot-level. In their study relative RMSEs for pine, spruce and deciduous trees were 45.50%, 61.98% and 92.30%, respectively. Error of the total volume estimate was 23.86%. In the study of Packalén and Maltamo (2007), on the other hand, the relative errors of the plot-level estimates of total, spruce, pine and deciduous trees volumes were 20.51%, 55.72%, 51.55% and 102.84%, respectively. In the same study the corresponding errors on stand-level were 10.36%, 32.64%, 28.08% and 62.33%, respectively. Stand-level estimation accuracies in Maltamo *et al.* (2009) for pine, spruce, deciduous trees and total volumes were about 30%, 130%, 115% and 19%, respectively. The biases of the estimates introduced in the both studies mentioned above were significantly smaller when compared to the biases presented in our study. The stand-level estimation errors obtained with the traditional inventory methods for pine, spruce, deciduous trees and total volumes are about 29%, 43%, 65% and 25%, respectively (Haara and Korhonen 2004).

When comparing between the previous and the present studies it has to be kept in mind that the estimation levels and the amount of the plots in the carrier data differ. Larger carrier data and wider estimation unit should produce more accurate results. As such, the errors in the present study appear to be of a high quantity with all the tested sample intensities if the change in the dominating tree species (spruce or pine) is considered. Now, however, the stands from which the forest holding level was averaged with post-stratification were very small (0.06–0.2 ha) since they consisted of the measured plots alone. We prospect that the error level would have been substantially lower if the stand-level estimation was done with e.g. grid-based approach adopted by Næsset *et al.* (2004).

Species-specific estimation is always rather burdensome in heterogeneous forests and the accuracy of the estimates depends on how well the modeling data covers the variation in the forests of interest. With the nearest neighbor methods, as k-MSN, an extrapolation is impossible and thus, extremely high values might be underestimated and extremely small values overestimated. In the present study the volume estimates for spruce and pine were clear underestimations and the volume of deciduous trees overestimation. It seems that the modeling data didn't covered enough observations comparable to the forest of interest and as such, k-MSN model could not find proper nearest neighbors for them. Privately owned forests are generally rather diverse: even a forest holding may cover both extremely small and high values of the forest characteristics.

The selection of the sample design and the intensity of the carrier data should be considered carefully when conducting a remote sensing based forest inventory by tree species in privately owned forests. Sample design used in this study was non-probability sample, designed as it is currently carried out in the Finnish aerial data based practical forest inventories. The design attempts to mimic the Finnish NFI with stratification and subjectively allocated additional measurements. Our results indicate that the use of the non-probability sample should be weighed up and the use of probability sample considered again.

Multivariate type of the modeling might have had its own affect on the results. In this approach of modeling the weighted average of the RMSEs of the estimated variables is minimized instead of minimizing the RMSE of one variable only. Predictor variable selection for the k-MSN model was conducted identically to Packalén and Maltamo (2007) who presented rather accurate species-specific estimates with the method. Even though, it might have been that the aerial image-based metrics were not that useful in discriminating between tree species in our study. Also the assumption of achieving some modeling synergy by combining aerial image and ALS metrics might have retracted.

References

AXELSSON, P., 2000, DEM generation from laser scanner data using adaptive TIN models. In *XIXth ISPRS Conference*, Vol. XXXIII, IAPRS, Amsterdam, The Netherlands, pp. 110–117.

HAARA, A., AND KORHONEN, K.T., 2004, Kuvioittaisen arvioinnin luotettavuus. *Metsätieteen aikakauskirja* 4/2004, pp. 489–508. (in Finnish)

HARALICK, R.M, SHANMUGAM, K., AND DINSTEEIN, J., 1973, Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics,* 3, pp. 610–621.

LAASASENAHO, J., 1982, Taper curve and volume function for pine, spruce and birch. *Communicationes Instituti Forestalis Fenniae*, 108, pp. 1–74.

MALTAMO, M., PACKALÉN, P., SUVANTO, A., KORHONEN, K.T., MEHTÄTALO, L., AND HYVÖNEN, P., 2009, Combining ALS and NFI training data for forest management planning : a case study in Kuortane, Western Finland. *European Journal of Forest Research*, 128, pp. 305–317.

MOEUR, M., AND STAGE, A.R., 1995, Most similar neighbor : An improved sampling inference procedure for natural resource planning, *Forest Science*, 41, pp. 337–359.

NÄSLUND, M., 1937, Skogsförsöksanstaltens gallringsförsök i tallskog. *Meddelanden från Statens Skogsförsöksanstalt*, 27, pp. 1–169. (in Swedish)

NÆSSET, E., GOBAKKEN, T., HOLMGREN, J., HYYPPÄ, H., HYYPPÄ, J., MALTAMO, M., NILSSON, M., OLSSON, H., PERSSON, Å., AND SÖDERMAN, U., 2004, Laser scanning of forest resources : The nordic esperience. *Scandinavian Journal of Forest Research*, 19, pp. 482–499.

PACKALÉN, P., AND MALTAMO, M., 2006, Predicting the plot volume by tree species using airborne laser scanning and aerial photographs. *Forest Science*, 52, pp. 611–622.

PACKALÉN, P., AND MALTAMO, M., 2007, The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. *Remote Sensing and Environment*, 109, pp. 328–341.

PACKALEN, P., 2009, Using airborne laser scanning data and digital aerial photographs to estimate growing stock by tree species. *Dissertationes Forestales* 77. PhD thesis, University of Joensuu, Finland.

Integrating airborne laser scanner data and ancillary information for delineating the boreal-alpine transition zone in Hedmark County, Norway

H. O. ØRKA†, M. A. WULDER‡, T. GOBAKKEN† and E. NÆSSET†

hans-ole.orka@umb.no

†Norwegian University of Life Sciences, Department of Ecology and Natural Resource Management, P.O. Box 5003, NO-1432 Ås, Norway

‡Canadian Forest Service, Pacific Forestry Centre, Natural Resources Canada, 506 West Burnside Road, Victoria, Canada British Columbia V8Z 1M5, Canada

Abstract

The boreal-alpine transition zone represents the gradient, or ecotone, from boreal forest to open alpine tundra. At present, resource inventories and/or systematic monitoring of mountainous areas are not commonly undertaken in many regions and countries. Current plot-wise national forest inventories typically focus on monitoring of the productive forests which are more valuable from an economic perspective. There is an increasing demand for information concerning high altitude forests and full capture of treed areas to support national and international biomass and carbon reporting. Further, impacts of climate change are expected to be most pronounced over this transition zone, leading to a need for methods for monitoring the status and dynamics of vegetation over the ecotone. We propose a method for integrating airborne laser scanner (ALS) data collected as a strip sample and ancillary information to delineate the boreal-alpine transition zone. In this study, the boreal-alpine transition zone is defined according to international definitions based on tree heights and crown coverage to provide the basis for reporting according to established international standards. The 3-dimensional measurements of forest structure obtained from an airborne laser scanner provided the basis for detecting the boreal-alpine transition zone. We establish, validate, and discuss an heuristic method to delineate the boreal-alpine transition zone using ALS data. The method was implemented using 53 ALS sample strips in Hedmark County, Norway, and validated with field measurements of the transition zone represented by forest and tree lines at 26 locations. The ALS delineation of the boreal-alpine transition was accurate when compared to field measurements. Furthermore, a non-parametric method was used to upscale the ALS estimates to the entire area of Hedmark County (27 400 km^2) using Landsat images and information derived from a digital terrain model as ancillary data. The size of the estimated boreal-alpine transition zone in Hedmark was 3750 km^2.

1. Introduction

The boreal-alpine transition zone is the shift between the forest and alpine vegetation communities (Kimmins 1997). In the current study, the boreal-alpine transition was defined to be the area between the forest line and the tree line. Until now inventory and systematic monitoring of such mountainous areas are not undertaken in many regions and countries, including Norway. Current plot-wise national forest inventories typically focus on monitoring of the productive forests which are more valuable from an economic perspective. There is an increasing demand for information concerning high altitude forests and trees, to support national and international biomass and carbon reporting activities that are intended to be inclusive of all treed areas. Furthermore, impacts of climate change are expected to be most pronounced over this transition zone, leading to a need for methods for monitoring the status and dynamics of vegetation over the ecotone.

Forest- and tree lines are expected to advance (e.g., increased altitude and latitude) as a result of a warmer climate (Harsch et al. 2009). In addition, changes in human use and activities in mountain areas will affect the boreal-alpine transition zone. Additionally, the presence of grazing domestic animals has over the last centuries

pushed the tree line below the expectation due to climatic constraints alone; diminished grazing would result in an advance of forest- and tree lines (Hofgaard 1997, Cairns and Moen 2004). The effect of land use change may override the responses of climate change on the vegetation in the boreal-alpine transition zone (Hofgaard 1997). A meta-analysis, showed that half of the studied tree lines worldwide had advanced during the last century (Harsch et al. 2009). Hence, there is an urgent need for inventory and monitoring systems to characterize the boreal-alpine transition in systematic, transparent, and repeatable manner.

Airborne laser scanning (ALS) are the most promising technique for measuring biophysical properties in forested areas (e.g. Hyyppä and Hyyppä 1999, Hyde et al. 2006). Tree heights and canopy cover, which are essential for delineating the boreal-alpine transition zone, are frequently derived using ALS data (c.f. Andersen et al. 2006, Hopkinson and Chasmer 2009). Hence, a wall-to-wall acquisition of ALS data over a region would probably provide the most accurate monitoring system of these transition zone areas. However, a full coverage ALS scanning will be costly. To minimize cost and achieve a high accuracy in regional forest inventory using ALS as a strip sampling tool has been suggested and is under development (Næsset et al. 2009). In the proposed inventory protocol ALS strips will sample boreal-alpine transition zones where they occur. Gathering field information is costly in these mountainous and often remote areas, ALS provides accurate measurements of these transition zones with the possibility to detect, e.g. forest lines (Rees 2007). However, full coverage maps are often required, such as for providing separation of land areas into strata for carbon and biomass inventories. In such cases, remote sensing data covering the entire region of interest are needed. Medium spatial resolution satellite images that provide wide-area coverage with a sufficient level of spatial detail are ideal for stratification. The spatial, spectral, temporal, and radiometric resolutions of Landsat, combined with a large image footprint, making Landsat among the best sources of information (Wulder 1998, Cohen and Goward 2004, Wulder et al. 2008). In addition, other spatial data layers with full-coverage are known to increase the accuracy of maps created from satellite images (Franklin 1995).

The current study utilize ALS data from 53 flight-lines collected in support of testing ALS as a strip sampling tool for regional forest inventories (Næsset et al. 2009). Canopy cover are thought to be nearly unbiased when derived from ALS data (Lucas et al. 2008). There are limitations in direct canopy cover estimates from ALS data related to the underestimation of tree height (Gaveau and Hill 2003), the need of nadir measurements and inclusion of small canopy gaps in the canopy cover required by the definitions described above (Gschwantner et al. 2009). However, ALS derived canopy cover are known to be a good approximation (McLane et al. 2009). A heuristic approach was used to delineate the boreal-alpine transition zone in the ALS data. This heuristic approach used the United Nations Food and Agriculture Organization (FAO) definitions of forest and other wooded land to define the boreal-alpine transition zone. The ALS estimates provide a large sample of transition zones suitable for use to produce a full coverage map (mask) of the boreal-alpine transition zone and to estimate the area of this zone in the study area. The specific objectives of the current study are to 1) delineate the boreal-alpine transition zone in ALS data, 2) upscale ALS information from sample strips to full coverage using Landsat and digital terrain data, and 3) estimate the area of the boreal-alpine transition in Hedmark, Norway.

2. Materials and Methods

2.1. Study area

The study was conducted in Hedmark County in south-eastern Norway. The total land area of Hedmark is approximately 27 400 km^2. The county is covered by boreal and alpine vegetation zones with a slightly continental climate (Moen 1999) . Elevations range from 120 to 2180 m above sea level.

2.2. Field data collection

During the summer of 2008 the boreal-alpine transition zone was mapped at 26 locations in Hedmark. Locations were selected subjectively based on the following criteria: located in the ALS transects, orthophoto available, accessibility, and contributing to suitable spatial distribution of plots. The boreal-alpine transition zone was

manually digitized by applying common practices following the forest- and tree-lines. Digitizing was conducted with a Bluetooth GPS receiver (Holux M-1000) connected to a Personal Data Assistant with a Geographical Information System.

2.3. Airborne laser scanner data

ALS data were acquired during summer 2006 with two different Optech ALTM 3100 sensors (Table 1). Flight lines were flown in an east-west direction with 6 km spacing between flight lines. The total length of all flight lines was over 4500 km. When considering the scanned swath of 490 m, the areal coverage is approximately 8% of the study area. The initial processing of the data was accomplished by the contractor (Blom Geomatics, Norway). Ground returns were determined using the Terrascan software (Terrasolid Ltd., 2004) and a triangulated irregular network (TIN) was created from the echoes classified as ground returns. Heights above the ground surface were calculated for all echoes by subtracting the respective TIN heights from the height values of all echoes recorded.

Table 1. ALS sensor and acquisition settings.

Parameters	
Sensor	ALTM 3100
Mean flying altitude AGL (m)	800
Pulse repetition frequency (kHz)	100
Scanner frequency (Hz)	55
Half scan angle (deg.)	17
Mean pulse density (m^{-2})	2.7
Footprint diameter (cm)	21

2.4. Landsat images and digital elevation data

Four different Landsat-5 TM images were obtained from USGS to cover the study area. The scenes used were WRS Path 197 / Row 18 and 16 acquired at 3 June 2007, and Path 198 / Row 18 and 17 acquired 10 June 2007. The images were georeferenced, orthorectified, and converted to top of atmospheric reflectance (TOA) following the procedure of Han et al. (2007). The TOA corrected images were mosaicked together. From the TOA corrected Landsat mosaic the normalized difference vegetation index (NDVI) and the brightness, greenness and wetness from the tasselled cap transformation was derived (Kauth and Thomas 1976). From the DTM elevation, slope, curvature, and solar radiation were derived and used as ancillary support in the subsequent classification.

2.5. Processing of airborne laser data

The point cloud obtained from the ALS sensors can be considered as a sample of the forest canopy where each echo (x, y, z) is a sample point. Classifying each point according to presence or absence of canopy makes the point cloud a sample of a binomial distribution (canopy or not) (Cochran 1977). Hence, the canopy cover can be computed as the number of echoes in the canopy over the total number of echoes:

$$C = \frac{N_c}{N_t} \quad (1)$$

where C is canopy cover, N_c is number of first returns in canopy, and N_t is total number of first returns. Similar approaches have previously been used (c.f. Hopkinson and Chasmer 2009). The canopy coverage was computed using two height thresholds, 0.5 and 5m, to represent canopy cover for shrubs and trees and for only trees, respectively. The boreal-alpine transition zone was indentified where canopy cover of shrubs and trees were above 10% and/or canopy cover of only trees was between 5 and 10%. The thresholds used for canopy cover and heights was according to the FAO definitions of forest and other wooded land (FAO 2006).

The accuracy of the implemented procedure is validated with the field measured forest- and tree lines. At the 26 field locations three classes (boreal, alpine, and transition zone) were determined. The class map was validated to the field measured tree- and forest lines, separately. In the accuracy assessment an image gradient based method was utilized (Pitas 2000, Wulder et al. 2007). In the ALS class map the rate of change in a local neighbourhood was computed as the gradient:

$$|\nabla f(x,y)| = \sqrt{[f(x+1,y) - f(x,y)]^2 + [f(x,y+1) - f(x,y)]^2} \quad (2)$$

, where $|\nabla f(x,y)|$ is the gradient and x and y are row and column in the raster file created at each location.

The gradient is aggregated with information related to the distance to the field measured line and the average gradient values are summarized over all locations for both the forest- and tree lines.

2.6. Processing of ancillary data

To estimate the boreal-alpine transition zone a non-parametric random forest method was used (Breiman 2001). Random forest is a development of regression and classification trees that is subject to increasing usage due to a capacity to handle a large amount of classification features, mixing of variable types, high predictive performance, and an absence of over fitting (Falkowski et al. 2009, Ørka et al. 2010). The spectral indexes (NDVI, brightness, wetness, and greenness), elevation, slope, solar radiation, curvature, and location variables (latitude and longitude) were used as classification features. Since, transition zones are best represented with soft classifiers (Foody 1996) the random forest classification was used to predict the probability of boreal using only boreal and alpine classes from the ALS as input. A feature selection algorithm was used to select relevant features (Díaz-Uriarte and Alvarez de Andrés 2006). The reference data include the plots with size equal to the image pixel size, laid out every three kms on the ALS transects. Reference data were pre-stratified according to the potential boreal-alpine transition area using the DTM.

To determine the three classes; boreal, alpine, and transition zone, we estimated the probability of boreal zone for the reference plots. However, we now also included the locations indicated by the ALS data to be in the transition zone. Furthermore, the probability density function was estimated separately for all three classes. Then the alpha-cuts were set for the upper and lower boundary where the transition zone had higher density than boreal and alpine.

The random forest probability classification and the alpha-cut were validated using a test dataset which represent the plots located ± 1 km in the east-west direction of the calibration plots.

3. Results and discussion

The ALS derived classes of boreal, alpine, and transition zones using the simple heuristic approach fits the field data with a high accuracy (see figure 1). The highest gradient values appear at the forest line and tree line. There was however, a presence of high gradient values in some other areas, such as transitions with within some forested areas. The forest below the forest line is diverse and include mires, peatland and also patches of

alpine meadow. In the current study only the upper forest and tree lines were mapped in field. In some locations the forest or tree lines could be characterized as diffuse; it was obvious that these “lines” not are strict boundaries. This situation is also known from the vector representation of transitions zones in geographical information systems, since transitions cannot be readily represented using a vector model (Franklin 1995, Wang and Hall 1996). The ALS classes developed here using the heuristic classification are represented with a raster model in an attempt to mitigate this limitation / consideration.

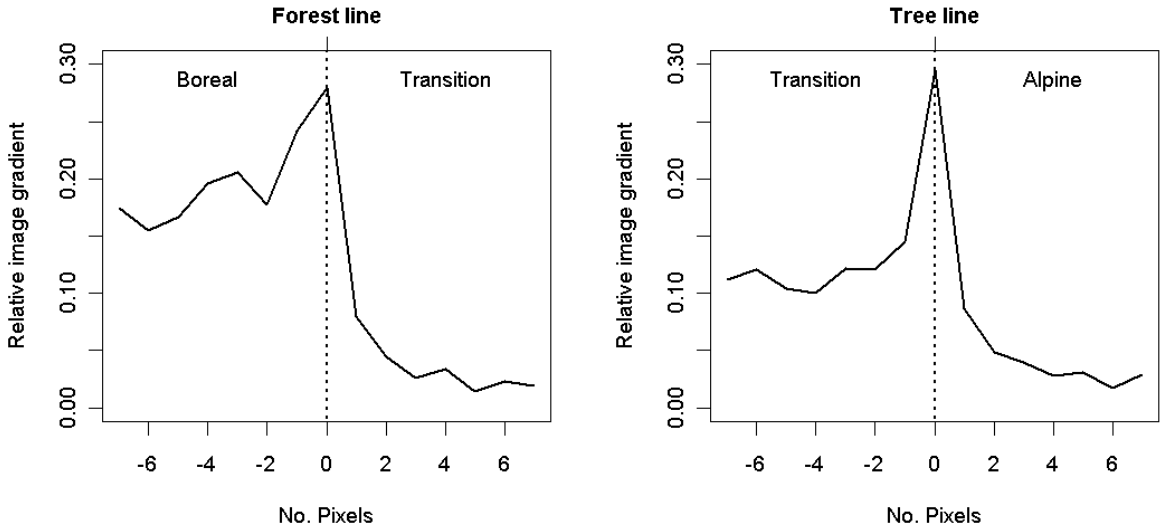


Figure 1. Accuracy assessment using an image gradient analysis of 26 field locations against forest line (left) and tree line (right). Relative gradient in ALS classes against distance to forest and tree line. Dotted lines indicate forest and tree lines, respectively.

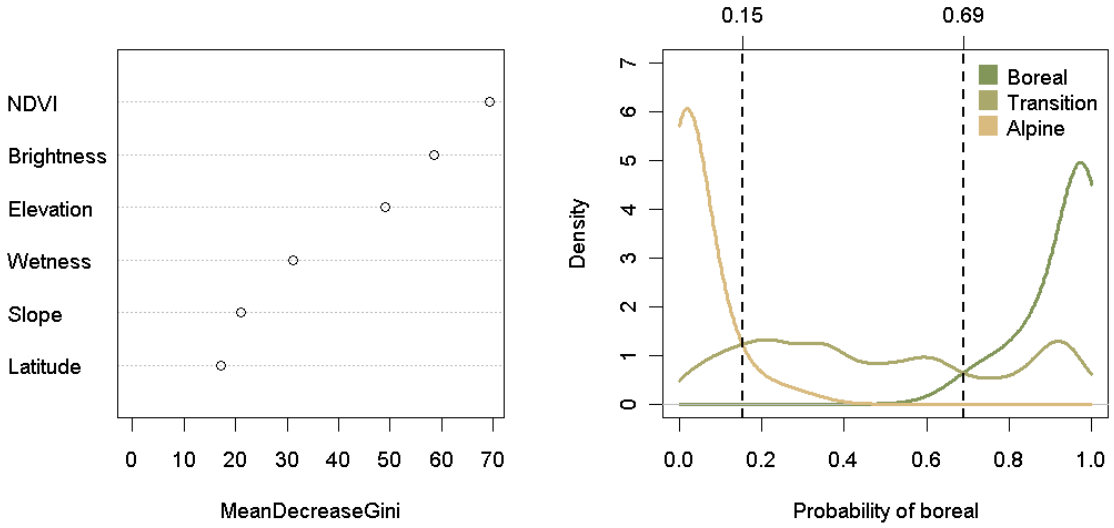


Figure 2. LEFT: The mean decrease in Gini coefficient for the five features in classification. RIGHT: Alpha-cuts derived using the density functions for probability computed from the binary random forest classification for all calibration plots and for separate classes.

The feature selection resulted in five features used in the classification. The selected features included the tree spectral features; brightness, wetness, and NDVI; plus elevation and slope derived from the DTM. The importance of the features appears in figure 2. NDVI and brightness represent different objects in the images. NDVI values are related to vegetation and brightness is usually related to soil, both important in the boreal-alpine transition zone where vegetation is decreasing and soil intensity is increasing from the boreal- to the alpine environment. Contrary to expectation, solar radiation was not selected, even though it is known to relate environmental factors related to slope among others. Latitude is related to the regional trend as altitude and the altitude of forest- and tree lines are increasing further north in the county. The alpha-cuts derived from the density functions was 0.15 to alpine areas and 0.69 to forested areas (see figure 2). Hence, areas with a probability predicted between 0.15 and 0.69 belong to the boreal-alpine transition zone. The calibrated random forest classification and alpha-cuts were applied on the whole region. The error matrix for the three classes - boreal, alpine and transition zone, for the test dataset, appears in table 2. The hard classification is accurate when taking into account the problems of mixing between classes that occurs in the transition. Table 2 shows highly accurate characterizations of the distinct boreal and alpine classes (especially when excluding the confusion related to the transition class), with variable accuracy and across class confusion by the transition class. The confusion evident in depicting the transition class is expected as the category is a mixture of the conditions dominant in the boreal and alpine classes.

The total area of the boreal-alpine transition zone in Hedmark was 3750 km^2. Hence, about 14 % of the area in Hedmark is a transition between the boreal forest and the alpine environment. From the known impacts of climate change (Harsch et al. 2009) this area could change to a forest class. The full coverage maps are also an important source to enable monitoring of possible changes. Both the soft classification represented by the probability for a boreal class and the hard classification produced by the alpha-cuts can support different needs. For monitoring purposes the soft classification is most useful (Foody 2001), but for some needs, such as area estimation, a hard classification is preferred.

Table 2. Error matrix of classification of boreal, alpine, and the transition zone based on the test dataset using binary random forest classification and alpha-cuts.

	Reference				
Classification	Boreal	Alpine	Transition	Sum	User accuracy
Boreal	302	20	75	397	76.1
Alpine	6	488	42	536	91.0
Transition	88	173	118	379	31.1
Sum	396	681	235	1312	
Producer accuracy	76.3	71.7	50.2		
Overall accuracy					69.2
Kappa					0.523

4. Conclusion

In this research, we apply a heuristic classification of ALS data which enables an accurate depiction of the boreal-alpine transition zone over a large region. The information derived from ALS data can further be combined with ancillary information to produce full coverage maps when needed. The described method does not use any field data for calibration. Hence, there is no need to gather expensive field data from remote

mountainous areas. The high accuracy of a non-field calibrated method is also promising for similar approaches when field data is available for calibration. An improved capacity to capture an entire forested area, rather than limited to managed forest areas, is increasingly desired and aided by the approach presented here. The ability to portray boreal, alpine, and related transitional areas augments our ability to monitoring and report on carbon stocks and change and to ensure that all applicable forested areas are included. Studies of climate change are also aided by the ability to map the boreal-alpine transition zone over large areas. Changes found over time will be important on describing the processes operating and the rates of transition among classes.

Acknowledgements

This research was funded by the Norwegian University of Life Sciences and the Research Council of Norway (research grant #192792/I99). We also wish to thank Blom Geomatics, Norway, for providing and processing the airborne laser data.

References

ANDERSEN, H. E., REUTEBUCH, S. E., and MCGAUGHEY, R. J., 2006, A rigorous assessment of tree height measurements obtained using airborne lidar and conventional field methods. Canadian Journal of Remote Sensing, 32, 355-366.

BREIMAN, L., 2001, Random forests. Machine Learning, 45, 5-32.

CAIRNS, D. M., and MOEN, J., 2004, Herbivory Influences Tree Lines. Journal of Ecology, 92, 1019-1024.

COCHRAN, W. G., 1977, Sampling Techniques (New York: John Wiley & Sons).

COHEN, W. B., and GOWARD, S. N., 2004, Landsat's role in ecological applications of remote sensing. Bioscience, 54, 535-545.

DÍAZ-URIARTE, R., and ALVAREZ DE ANDRÉS, S., 2006, Gene selection and classification of microarray data using random forest. BMC bioinformatics, 7, 3.

FALKOWSKI, M. J., EVANS, J. S., MARTINUZZI, S., GESSLER, P. E., and HUDAK, A. T., 2009, Characterizing forest succession with lidar data: An evaluation for the Inland Northwest, USA. Remote Sensing of Environment, 113, 946-956.

FAO, 2006, Global Forest Resources Assessment 2005 - Progress towards sustainable forest management. FAO Forestry Paper, 147.

FOODY, G. M., 1996, Fuzzy modelling of vegetation from remotely sensed imagery. Ecological Modelling, 85, 3-12.

FOODY, G. M., 2001, Monitoring the magnitude of land-cover change around the southern limits of the Sahara. Photogrammetric Engineering and Remote Sensing, 67, 841-848.

FRANKLIN, J., 1995, Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. Progress in Physical Geography, 19, 474.

GAVEAU, D. L. A., and HILL, R. A., 2003, Quantifying canopy height underestimation by laser pulse penetration in small-footprint airborne laser scanning data. Canadian Journal of Remote Sensing, 29, 650-657.

GSCHWANTNER, T., SCHADAUER, K., VIDAL, C., LANZ, A., TOMPPO, E., DI COSMO, L., ROBERT, N., DUURSMA, D. E., and LAWRENCE, M., 2009, Common Tree Definitions for National Forest Inventories in Europe. Silva Fennica, 43, 303-321.

HAN, T., WULDER, M. A., WHITE, J. C., COOPS, N. C., ALVAREZ, M. F., and BUTSON, C., 2007, An Efficient Protocol to Process Landsat Images for Change Detection With Tasselled Cap Transformation. Geoscience and Remote Sensing Letters, 4, 147-151.

HARSCH, M., A., HULME, P., E., MCGLONE, M., S., and DUNCAN, R., P., 2009, Are treelines advancing? A global meta-analysis of treeline response to climate warming. Ecology Letters, 12, 1040-1049.

HOFGAARD, A., 1997, Inter-Relationships between Treeline Position, Species Diversity, Land Use and Climate Change in the Central Scandes Mountains of Norway. Global Ecology and Biogeography Letters, 6, 419-429.

HOPKINSON, C., and CHASMER, L., 2009, Testing LiDAR models of fractional cover across multiple forest ecozones. Remote Sensing of Environment, 113, 275-288.

HYDE, P., DUBAYAH, R., WALKER, W., BLAIR, J. B., HOFTON, M., and HUNSAKER, C., 2006, Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM plus , Quickbird) synergy. Remote Sensing of Environment, 102, 63-73.

HYYPPÄ, H., and HYYPPÄ, J., 1999, Comparing the accuracy of laser scanner with other optical remote sensing data sources for stand attributes retrieval. Photogramm. J. Finland, 16, 5 -15.

KAUTH, R. J., and THOMAS, G. S., 1976, The tasseled cap—a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat In: Final proceedings: 2nd international symposium on machine processing of remotely sensed data

KIMMINS, J. P., 1997, Forest Ecology, A foundation for sustainable management, 2nd ed. (Upper Saddle River, New Jersey: Prentice-Hall Inc.).

LUCAS, R. M., LEE, A., ARMSTON, J., and BREYER, J., 2008, Advances in forest characterisation, mapping and monitoring through integration of LiDAR and other remote sensing datasets. In Proceedings of SilviLaser 2008, edited by R. Hill, J. Rosette, and J. Suárez (Edinburgh, UK.

MCLANE, A. J., MCDERMID, G. J., and WULDER, M. A., 2009, Processing discrete-return profiling LiDAR data to estimate canopy closure for large-area forest mapping and management. Canadian Journal of Remote Sensing, 35, 217-229.

MOEN, A., 1999, National atlas of Norway: Vegetation (Hønefoss: Norwegian Mapping Authority, Hønefoss).

NÆSSET, E., GOBAKKEN, T., and NELSON, R., 2009, Sampling and mapping forest volume and biomass using airborne LIDARs. Proceedings of the Eight Annual Forest Inventory and Analysis Symposium, Monterey, CA, USA, 297-301.

ØRKA, H. O., NÆSSET, E., and BOLLANDSÅS, O. M., 2010, Effects of different sensors and leaf-on and leaf-off canopy conditions on echo distributions and individual tree properties derived from airborne laser scanning. Remote Sensing of Environment, 114, 1445-1461.

PITAS, I., 2000, Digital image processing algorithms and applications (New York: Wiley).

REES, W. G., 2007, Characterisation of Arctic treelines by LiDAR and multispectral imagery. Polar Record, 43, 345-352.

WANG, F., and HALL, G. B., 1996, Fuzzy representation of geographical boundaries in GIS. International Journal of Geographical Information Systems, 10, 573 - 590.

WULDER, M., 1998, Optical remote-sensing techniques for the assessment of forest inventory and biophysical parameters. Progress in Physical Geography, 22, 449-476.

WULDER, M. A., HAN, T., WHITE, J. C., BUTSON, C. R., and HALL, R. J., 2007, An approach for edge matching large-area satellite image classifications. Canadian Journal of Remote Sensing, 33, 266-277.

WULDER, M. A., WHITE, J. C., GOWARD, S. N., MASEK, J. G., IRONS, J. R., HEROLD, M., COHEN, W. B., LOVELAND, T. R., and WOODCOCK, C. E., 2008, Landsat continuity: Issues and opportunities for land cover monitoring. Remote Sensing of Environment, 112, 955-969.

Assessment of the performance of eight filtering algorithms by using full-waveform LiDAR data of unmanaged eucalypt forest

GIL GONÇALVES†‡ and LUÍSA GOMES PEREIRA§¶

gil@mat.uc.pt

†INESCC, Instituto de Engenharia de Sistemas e Computadores de Coimbra,

‡Faculty of Sciences and Technology of the University of Coimbra, PORTUGAL

§Higher School of Technology and Management of Agueda, University of Aveiro,

¶Research Centre for Geo-Spatial Sciences, University of Porto, PORTUGAL

Abstract

In this study the strengths and weaknesses of eight filtering algorithms are evaluated by using the mean, standard deviation and RMSE metrics. Seven of these algorithms are implemented in the freeware software ALDPAT (Airborne LiDAR Data Processing and Analysis Tools) and the eighth, known as the Axelsson filter, in the commercial software Terrascan. The referred metrics are calculated by using DTM of topographic surfaces with quite different morphologies and vegetation covers. Forty-three of these surfaces, on circular plots of 400 m^2 each, are covered by brushwood and unmanaged eucalypt forest with different stand characteristics. The mean tree density is around 1600 trees per hectare. The reference DTM for assessing the DTM produced by filtering full-waveform LiDAR data using the eight filtering algorithms are created with the help of a total station and geodetic GNSS receivers. The results show that the Axelsson and the so-called Polynomial Two Surface Fitting filters give the best results in terms of RMSE. Nonetheless, the results also show that all the tested filters are suitable for the filtering of full-waveform LiDAR data used in forestry related work, and collected over areas with great amount and high brushwood, chaotic eucalypt tree distribution and high tree density. The results obtained for a forest area with such characteristics – among which it should be mentioned a RMSE of 15 cm - are quite surprising.

1. Introduction

The characteristics of data collected by small footprint LiDAR systems are adequate to derive estimates of several 3D vegetation structure metrics (such as canopy height and crown diameter). The estimation of these metrics requires the separation of the ground surface from the vegetation on it. The ground surface, represented by a Digital Terrain Model (DTM), is obtained by the filtering of the laser-point cloud. This filtering process is a key issue for the computation of vegetation heights. Errors in DTM may result in erroneous vegetation structure metrics, which may have unforeseen repercussions.

During the last decade several filtering algorithms have been proposed to cope with different types of landscapes (urban and forest), and terrain morphologies. While a general understanding of the accuracy of the LiDAR systems has been achieved, the accuracy of the derived DTM from LIDAR data in forest environments has not been thoroughly evaluated (Hodgson and Bresnahan, 2004; Reutebuch et al., 2003), mainly in unmanaged eucalypt forests. Indeed, while in the recommendations of the work of (Hyyppä et al., 2008) it is said that the extraction of DTM for forest areas is well established, this conclusion is based on a list of works using other forests than eucalypt. Moreover, the full-waveform data have, in comparison to conventional pulsed LiDAR data, the advantage of echo detection being done in post processing making the ranging process more robust.

While this is expected to lead to higher accuracy of the derived distances and thus to a more accurate DTM (Ullrich et al., 2008), it remains to be proved in areas with the characteristics of the one here studied.

The few published works on the assessment of the performance of LiDAR filtering algorithms mainly address the statistics of omission and commission errors of the filtered data and not the geometric quality of the derived DTM with respect to an external reference. Huising and Gomes Pereira (1998) mentioned an error of 15 cm for the standard deviation of height differences on flat, bare terrain and identified some problems that may exist in the filtering process. Sithole and Vosselman (2004) conducted an experimental comparison of eight filters to evaluate their performance. The performance was assessed mainly by generating error matrices and by spatial representation of these error matrices in 15 subsets of the dataset. Zhang and Whitman (2005) compared three filters by using three LiDAR datasets collected on urban flat areas, coastal areas on smooth terrain and on mountainous areas and by testing their sensitivity to the filters parameters. Seo and O'Hara (2008) compared three filter algorithms that exploit morphological operations, progressive TIN densification and kriging, and by computing the omission and commission errors and RMSE for three LiDAR datasets (collected on residential areas on smooth terrain, residential areas on hilly terrain and commercial areas on moderate terrain relief). In order to compute the RMSE values they used reference surfaces interpolated from the manually filtered ground points. This so-called reference surface is bias since it is derived from the filtered LiDAR data.

Although the comparison of the performance of several filter algorithms has been assessed quantitatively by using the omission and commission errors, this procedure becomes impractical when the data are collected in unmanaged forested areas with high point densities (>1 pts/m^2). This is because the manually classification of the millions of points involved in a single survey is an unfeasible task.

In this paper it is assessed the performance of eight filtering algorithms by using full-waveform high density LiDAR data (> 10pts/m^2) of an unmanaged eucalypt forest. Seven of these algorithms are implemented in the freeware software ALDPAT (Airborne LIDAR Data Processing and Analysis Tools) and the eighth, the Axelsson filter, in the commercial software Terrascan. Their strengths and weaknesses are investigated by using DTM produced for topographic surfaces with quite different morphologies and vegetation covers. Forty-three of these surfaces, on circular plots of 400 m^2 each, are covered by brushwood and unmanaged eucalypt forest with different stand characteristics. The mean tree density is around 1600 trees per hectare. The DTM used to assess the DTM obtained with the eight filtering algorithms are produced with the help of a total station and geodetic GNSS receivers. The results show that the Axelsson and the so-called Polynomial Two Surface Fitting filters give the best results in terms of RMSE. Nonetheless, the results also show that all the tested filters are suitable for filtering of full-waveform LiDAR data used in forestry related work, and collected over areas with great amount and high brushwood, chaotic eucalypt tree distribution and high tree density. The results obtained for a forest area with such characteristics – among which it should be mentioned a RMSE of 15 cm - are quite surprising.

Study area and data

The study area, with 900 ha, was selected nearby the city of Águeda, in the district of Aveiro, situated in the Northern part of Portugal (Figure 1-a). Its topography varies from gentle to steep slopes, with altitudes varying from 27 to 162 m (Figure 1-b). Being the area dominated by eucalypt plantations, it also includes some pine stands and few built-up areas. The mean tree density is around 1600 trees per hectare. The forest stands in the area comprise regular and irregular spacing plantations, both even and uneven-aged stands, and stands with various undergrowth characteristics (Figure 1-c).

The LiDAR data were acquired on the 14th of July of 2008. The laser system utilized was the Litmapper 5600, operating with a pulse repetition frequency of 150 KHz, an effective measurement rate of 75 KHz and using a half-angle of 22.5º. Thirty overlapping strips (70% of sidelap) were flown from an average flying height above the ground of 640 m with an average single run density of 3.3 pt/m^2. The full-waveform laser data were processed

with the RiAnalyze software from Riegl. A maximum of 5 returns were obtained with a minimum vertical separation of 50 cm and the average values of laser footprint and point density were 30 cm and 10 pts/m^2 respectively.

Reference data are needed to verify, in terms of precision and reliability, the DTM produced by means of the laser data and a filtering algorithm. The strategy for the reference data collection was not straightforward. In forest areas, the collection of these data is time consuming, mainly in plots with a high density of shrubs and trees. Furthermore, because the data were georeferenced, geodetic GNSS receivers had to be used. The reference DTM is represented by the coordinates of terrain points located aside trees, which give also the locations of the trees, and by the coordinates of prominent terrain points, like those on breaklines. This information was collected by means of a topographic survey. The coordinate system in which the LiDAR data were collected is the WGS84 UTM zone 29, for X and Y coordinates, and the WGS84 ellipsoidal height for the Z coordinate. Because this is not a local system, the geographic information collected in the field had to be converted to that system by using the Global Positioning System (GPS). To this end, it was decided to attach to each plot two points, named GPS base, whose coordinates were measured with two GNSS receivers. These two points were placed as close as possible to the plot and as much as possible in an opened space. This criterion turned out to be difficult to fulfil in the study area. Finally, 3 174 points were measure on 43 circular plots, of radius 11.28 m, using this methodology.

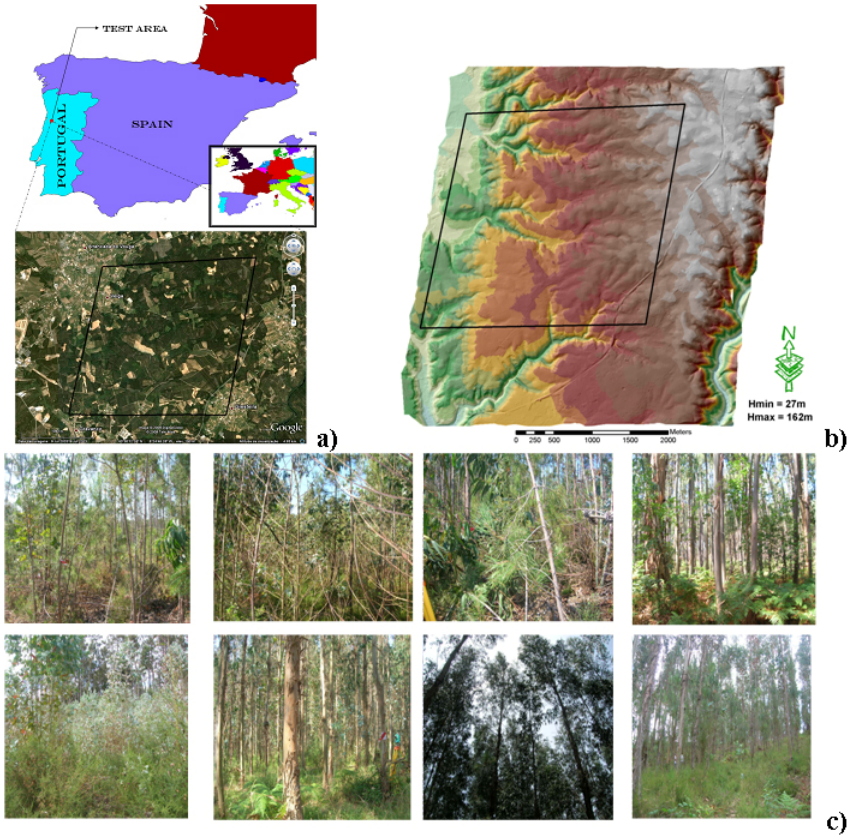


Figure 1: (a) Localization of study area within Portugal and its delimitation; (b) DTM of the study area; (c) Examples of vegetation covers inside plots.

2. Filtering methods

As stated above, seven of the eight filters tested are implemented in the free software ALDPAT®. The eighth filter is the well-known Axelsson filter (ATINT) implemented in the TerraScan® software. A short description of each filter and of its basic parameters is presented underneath.

1. Elevation threshold with expand window (ETEW) – To label the points as ground, this filter uses the concept of minimum height value inside square windows with growing size. Its main parameters are:

Initial cell size (c) for gridding the point cloud.

Slope factor (s) used for calculating the threshold value (th_i) of the elevation differences between the point with minimum elevation value (Z_{min}) and any other Z_j point located inside the window. The threshold is computed from $th_i = s \times c_i$, where c_i is the cell size at iteration i (which is twice that of the previous iteration). For the i^{th} iteration the point j is labelled as ground point if:

$$\left[Z_j - Z_{\min} \right]_i \leq th_i \qquad (1)$$

Number of iterations (i).

2. Iterative polynomial fitting (IPF) - This filter classifies the LiDAR points into ground points by selecting iteratively the ground measurements from the original point cloud inside a size decreasing moving window. A candidate point (lowest point inside the moving window) is added to the set of ground points if the elevation difference between the elevation of this point and that at the same planimetric location given by the polynomial fitted ground surface is less than a threshold (th). In each iteration, the ground points used for the surface fitting with piecewise polynomials are selected as the lowest points within the moving window. For the sake of simplicity, the cloud point is converted to a grid format and all the filtering process is done over this minimum elevation grid. The main parameters of this filter are:

Cell size (c) used in each iteration for the fitting of the piecewise polynomial surface to the ground points.

Maximum vertical difference (th) between the elevation of a point and that at the same planimetric location given by the fitted ground surface.

Maximum vertical difference (to) between the elevation of a point and that computed at the same planimetric location by using the final interpolated ground surface.

Initial size of the moving window (wi) for selection of the ground points. In the first iteration the points with minimum elevation falling within the window are selected as ground points. For the remaining iterations, the moving window is centred over each grid node and the minimum elevation point within the window is selected as a ground candidate.

Number of windows (or iterations) (wn).

Window sizes (ws) used in iteration i. Although ws at iteration i can be set automatically as half of its size at iteration i-1, ws may be chosen.

3. Polynomial two surface fitting (P2Surf) - This filter is an extension of the IPF filter and uses two polynomial fitting surfaces. In order to remove the omission errors (ground points classified as non-ground) the difference in elevation between that of a candidate point (or cell) and that of the current surface is calculated and compared with a threshold. To remove the commission errors (non-ground points classified as ground) the fitness of the current and of the previous surface to the ground points are compared to a threshold. A candidate point that falls within a given interpolation window (iw) is added to the set of ground points if the fitness of the current surface is better than that of the previous fitting surface within this fitting window. The filter parameters are similar to those of IPF filter adding the following:

Maximum value for the fitness (sig) of the current and previous polynomial surfaces to the ground points.

Radius of the neighbour search window (nr).

Size of the window (iw) used in the fitting operation.

Degree of the polynomial function (p) used for fitting the ground surfaces (current and previous).

4. Maximum local slope (MLS) - This filter uses the assumption that the terrain slope is different from the slope between the ground and the tree or building tops. It is similar to the filter described by Vosselman in [6]. A point p_i of a given point cloud V is classified as ground if the maximum value of slopes between this point and any other point p_j located in the circular neighbourhood of p_i with radius r ($v_{p_i}(r)$) is less than the predefined slope threshold (s). That is, denoting by d_{ij} the Euclidian distance between the two points p_i and p_j:

$$Ground = \left\{ p_i \in V \left| \forall p_j \in v_{p_i}(r), \frac{Z_i - Z_j}{d} < s \wedge d_{ij} > d \right. \right\} \quad (2)$$

The main parameters of this filter are:

Cell size (c) for creating the minimum elevation grid.

Minimum separation (d) between points allowed in slope computation.

Slope threshold (s).

Size of the search window (r).

5. Progressive morphology 1D (PM1D) - This filter is the progressive morphological (PM) filter presented in [7], and uses successive morphological openings for removing, from the ground non-ground objects of varying sizes. At each iteration (i) an opening (erosion + dilation) is performed on the previous opened surface. A cell j is classified as ground if the elevation difference between the previous (i-1) and current surface (i) is smaller than the height threshold:

$$th_i = s \times (ws_{i-1} - ws_i) \times c + th_0 \quad (3)$$

where th_0 is an initial elevation threshold to allow for small ground variations, s is the predefined maximum terrain slope, c is the cell size and ws_i is the size of the structuring element (window size) at i^{th} iteration. This structural element used by the PM filter can be either 1D (line segments) or 2D (squares or circles). The main parameters of the PM1D filter are:

Size of the cells (c) into which the point cloud is subdivided. All points, except those with minimum elevation are discarded.

Maximum terrain slope (equation (3)).

Maximum elevation difference (th) between terrain and ground objects.

Number of iterations (wn).

Sizes of windows (ws) used in the successive openings. For iteration i this size is calculated as $ws_i = 2b^i + 1, i = 0,1,\ldots,N$, where b is the base of an exponential function.

Search radius (r) used in the initial nearest neighbour interpolation for filling the empty cells.

6. Progressive morphology 2D (PM2D) - In this case, the structural element used to perform the morphological erosion and dilatation operations is a square window of size (ws). The main parameters of this filter are similar to those of the PM1D filter.

7. Adaptive TIN (ATIN) - The main concept used in this filter was originally developed by Axelsson (see the ATINT filter) and slightly modified as follows: i) the point cloud is first reduced by projecting it onto a grid of cell size (c) and keeping for each cell the point with minimum z value; ii) in this new point cloud a coarse TIN is obtained from this points using the Delaunay triangulation algorithm. The TIN is progressively densified (see below).

8. Adaptive TIN in TerraScan® (ATINT) – The basic assumption of this method is that the terrain relief can be locally and globally approximated by triangular facets. As mentioned above, this algorithm starts by creating a coarse TIN based on seed points (those with minimum elevation within a given square neighbourhood (wi)). Adding one ground point to each Delaunay triangle successively densifies this TIN. A point is added to the triangle if its distance to the triangle surface is less than a predefined threshold (th). An angle threshold (ta) may be also used to control the inclusion of a point close to a ground point with a steep slope. In this case it is necessary to calculate for each point the maximum of three angles between the triangular surface and the lines connecting the point and the vertices of the triangle. The main differences between this filter and the ATIN are: i) it works with the original point cloud; ii) in each iteration statistics from the points classified as ground are collected in the form of histograms of surface normal angles and elevation differences. These histograms are used to update the values of the thresholds used in the iterative process. The main parameters of this filter are:

Window size (wi) (size of the largest structure).

Steepest allowed slope (s).

Maximum distance (th) between the candidate point and the triangle plane.

Maximum angle (ta) between the candidate point, its projection on a triangle plane and the closest triangle vertex.

Procedure to assess the performance of the filters

The filters performances are assessed by estimating the accuracy of the DTM produced by filtering the LiDAR data. This accuracy assessment relates to the estimation of the mean, standard deviation and RMSE of the residuals or differences (dz;) between the Z values of the reference points (Z_{Ref}) and those at the same locations (x, y) of the LiDAR terrain points (Z_{LiDAR}).

$$dz(x,y) = Z_{\mathrm{LiDAR}}(x,y) - Z_{\mathrm{Ref}}(x,y) \tag{4}$$

These LiDAR terrain points are interpolated from a TIN computed with the filtered LiDAR terrain points. It was decided to create a TIN from the filtered LiDAR data instead of from the reference data because the density of LiDAR terrain points is higher than that of reference points (3.8 versus 0.2 points/m^2). The higher density of LiDAR terrain points per m^2 implies that points on the terrain are represented in the LiDAR filtered data but not in the reference data. Therefore, it is more adequate to interpolate the laser data to the planimetric positions of the reference data. The TIN format is also more appropriate than the grid format once the interpolated height is influenced only by the unaltered height values of its neighbours (3 neighbours when linear interpolation is used). The accuracy of the DTM obtained with the LiDAR data was computed for each plot individually and collectively.

Table 1 shows the values of the main parameters used in the filtering process for each of the filters above listed. These values were tuned experimentally by trial-and-error and by evaluating the quality of the resulting surface visually using shaded relief techniques.

Table 1: Parameters for each of the eight filters.

	c	s	th	ta	to	sig	wi	wn	ws	r	iw
ETEW	0.2	0.25	0.2					5	[0.2 0.4 0.8 1.6 3.2]		
IPF	0.2		0.1		0.05		10	4	[10 5 2 1]		
P2Surf	0.4		0.1		0.05	0.01	10	4	[10 5 2 1]	10	5
MLS	0.2	0.25	0.2					5			
PM1D	0.2	0.01	0.1				1	5	[0.2 0.4 0.8 1.6 3.2]		
PM2D	0.2	0.2	0.2				1	5	[0.2 0.4 0.8 1.6 3.2]		
ATIN	0.4		0.1	10			10				
ATINT			1.4	6			10				

3. Results and final considerations

Figures 2, 3 and 4 illustrate, respectively, the estimated values for the mean, standard deviation and RMSE, of the residuals obtained in the 43 circular plots and by using the eight LiDAR filters.

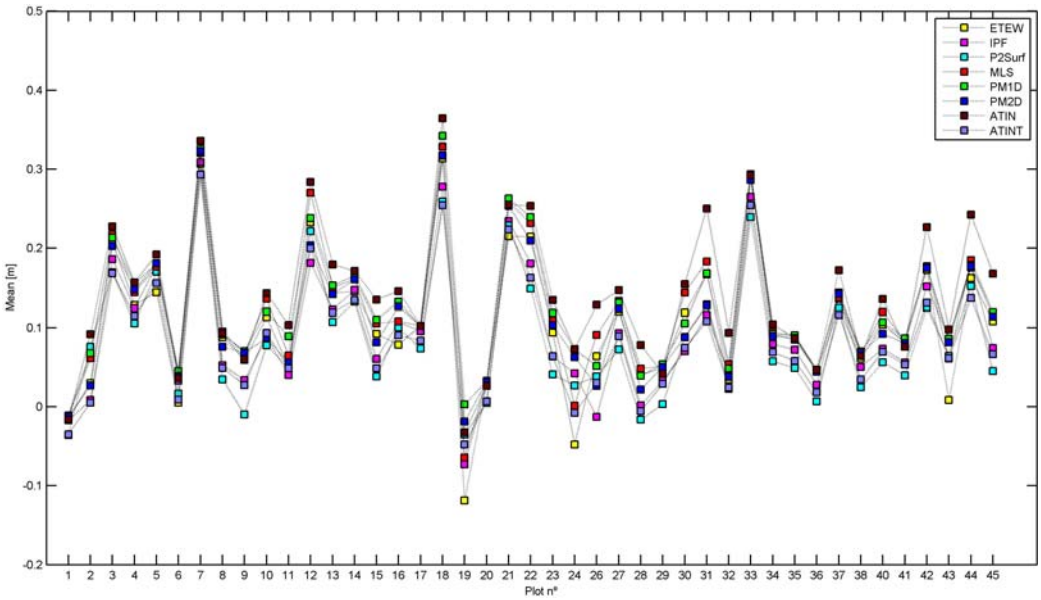


Figure 2: Values of the Mean of residuals per plot for the eight filters.

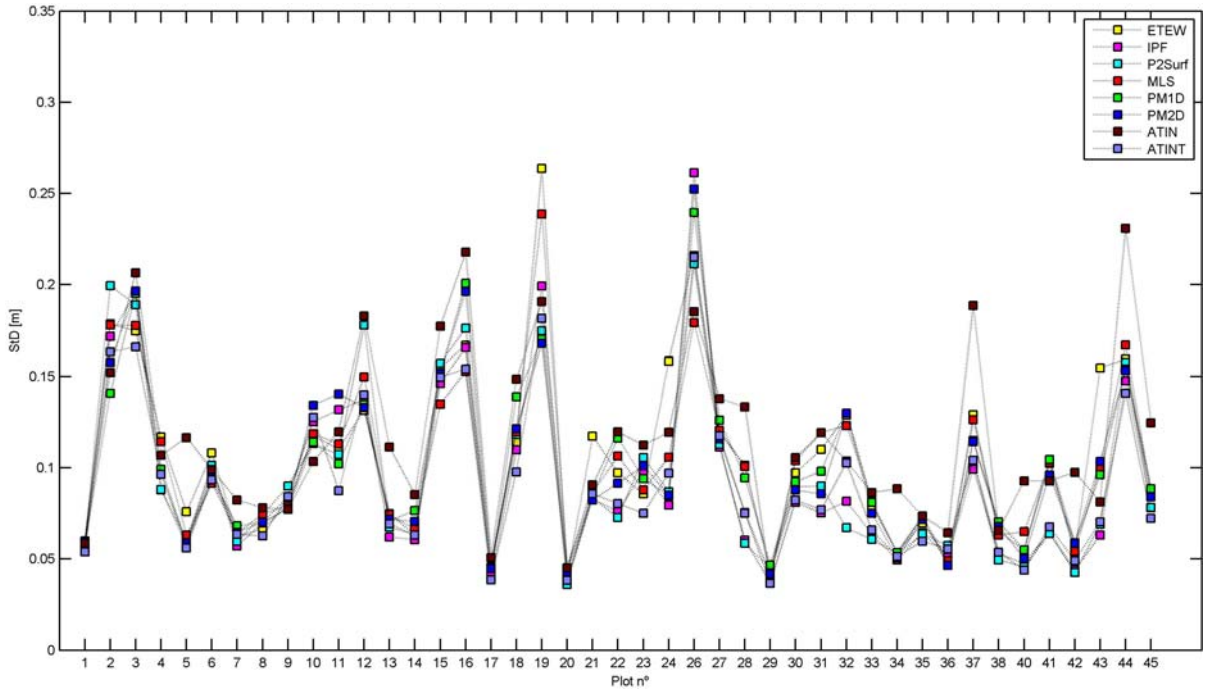


Figure 3: Values of the Standard deviation (STD) of residuals per plot for the eight filters.

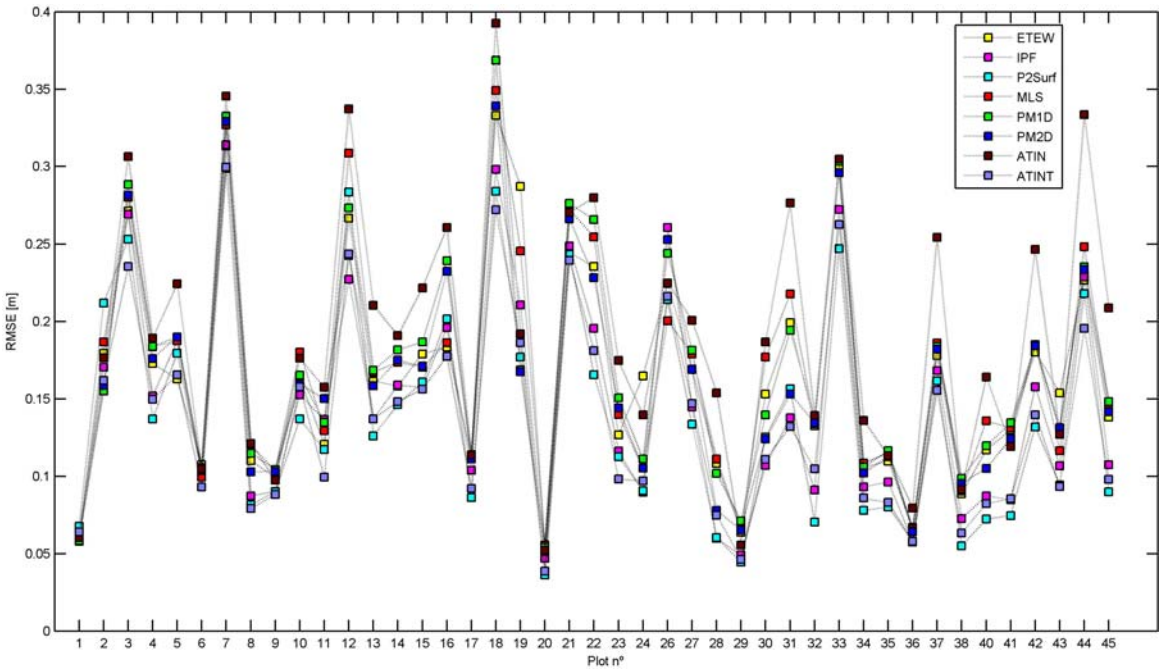


Figure 4: Values of the RMSE of residuals per plot for the eight filters.

Table 2 shows the same results for the eight filters when considering all the plots together, i.e., the 3 174 control points located within the 43 circular plots.

Statistical parametric tests of hypotheses were carried out to compare the mean and standard deviations of the residuals. By using a 5% level of significance the null hypothesis, i.e., the assumption that the mean values are equal was rejected (except for the mean of residuals obtained by using the P2Surf and ATINT filters). For the same level of significance, the tests indicate that the standard deviation values obtained with the filters P2Surf and ATINT are statistically equal and smaller than those obtained by using the other filters. These results show that both filters P2Surf and ATINT have similar performances, which are superior to those of the other filters. The ATIN filter, which is a different implementation of the Axelsson algorithm, has surprisingly the worst performance. In spite of these conclusions, the differences in the accuracy of the various DTM (maximum 6 cm) are not significant for work carried out in a forest environment.

Therefore, all the eight tested filters are suitable for the filtering of full-waveform LiDAR data collected in areas with great amount and high brushwood, chaotic eucalypt tree distribution and high tree density.

Table 2: Mean, standard deviation and RMSE values of residuals obtained by using the eight filters on LiDAR data within the 43 plots together.

	ETEW	IPF	P2Surf	MLS	PM1D	PM2D	ATIN	ATINT
Mean	0.10	0.09	0.08	0.12	0.12	0.11	0.14	0.08
STD	0.15	0.14	0.13	0.14	0.14	0.14	0.15	0.13
RMSE	0.18	0.16	0.16	0.18	0.18	0.18	0.21	0.15

Acknowledgments

The present study was funded by the Foundation for Science and Technology (FCT) of Portugal in the framework of the project PTDC/AGR-CFL/72380/2006 with co-funding by FEDER.

References

[1] Axelsson, P., 2000. DEM Generation from Laser Scanner Data Using Adaptive TIN Models. International Archive of Photogrammetry and Remote Sensing, Vol. XXXIII, Part B4, 110–17.

[2] Hodgson, M., and Bresnahan, P. 2004. Accuracy of airborne lidar-derived elevation: empirical assessment and error budget. PE& RS, vol. 70, no. 3, 331–40.

[3] Huising, E. and Gomes Pereira, L., 1998. Errors and accuracy estimates of laser data acquired by various laser scanning systems for topographic applications. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 53, no. 5, 245–61.

[4] Hyyppa, J., Hyyppa, H., Leckie, D., Gougeon, F., Yu, X., and Maltamo, M., 2008. Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. International Journal of Remote Sensing, vol. 29, no. 5, 1339–66.

[5] Reutebuch, S., McGaughey, R., Andersen, H.-E. and Carson, W., 2003. Accuracy of a high-resolution LiDAR terrain model under a conifer forest canopy. Canadian Journal of Remote Sensing, vol. 29, 527–35.

[6] Seo, S. and O Hara, C. 2008. Parametric Investigation of the Performance of LiDAR Filters Using Different Surface Contexts. PE& RS 74 (3), 343-62.

[7] Sithole, G. and Vosselman, G., 2004. Experimental comparison of filter algorithms for bare-earth extraction from airborne laser scanning point clouds. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 59, 85–101.

[8] Soininen, A., 1999 "TerraScan User's Guide", TerraSolid, 169.

[9] Vosselman, G., 2000. Slope Based Filtering of Laser Altimetry Data. International Archives of Photogrammetry and Remote Sensing, Vol. XXXIII, Part B3, Amsterdam 2000, 935–42.

[10] Zhang, K., Chen, S., Whitman, D., Shyu, M., Yan, J. and Zhang, C., 2003. A progressive morphological filter for removing non-ground measurements from airborne LIDAR data. IEEE Transactions on Geoscience and Remote Sensing, 41, 872-82.

[11] Zhang, K. and Whitman, D., 2005. Comparison of three algorithms for filtering airborne LIDAR data. PE& RS 71 (3), 313-24.

Evaluation of terrestrial LiDAR technology for the development of local tree volume equations

Demetrios Gatziolis[a], Sorin Popescu[b], Ryan Sheridan[b], and Nian-Wei Ku[b]

a USDA Forest Service, PNW Research Station, Resource Monitoring & Assessment, Portland, Oregon, dgatziolis@fs.fed.us

b Department of Ecosystem Science and Management, Texas A&M University

Abstract

Tree volume, a key parameter in all forest inventory systems, is typically calculated by using species-specific polynomial equations which are constructed by using measurements obtained from destructive tree sampling. Owing to cost, logistic, and sometimes policy constraints, typically only a small number of trees within the often large, multiregional distribution range of a species is destructively sampled. In areas characterized by diverse ecological, climatic, and forest structure conditions, such as the US Pacific Northwest region, this practice has lead to the development and use of volume equations whose prediction accuracy and precision at the local level is often unknown, even for common tree species. The detailed representations of tree stems offered by modern terrestrial LiDAR technologies are sometimes assumed to be an alternative, economical, non-destructive source of stem dimensionality information capable of supporting the development of local tree volume equations. We evaluate this assumption by comparing the merchantable volume of trees calculated from multiple diameter measurements obtained by experienced tree climbers at short intervals along the stem of standing trees in eastern Oregon to the volume estimated by processing corresponding terrestrial LiDAR data. Our methodology attempts to reconstruct the tree stem by exploring the mathematical morphology of terrestrial laser point clouds in voxel space and by utilizing elements of the graph theory. We found that if at least one stem segment near the upper end of the tree and a few other segments below are not obstructed when viewed from the laser scanning station, volume estimation errors no larger than 2% can be expected. For most trees, however, poor representation in the point cloud of the upper parts of the stem leads to an often substantial underestimation of volume. As expected, the accuracy of volume estimates improves with higher point densities.

Keywords: Terrestrial LiDAR, tree volume, voxel morphology

1. Introduction

Estimation of individual tree volume is a key component of forest inventory systems. Volume estimates are obtained via species- or species-group-specific equations that usually comprise two independent variables, tree diameter at breast height (DBH) and height, both characteristic, easy-to-measure descriptors of tree size. Volume equations come in a variety of forms including polynomial curves that approximate the tree diameter at regular intervals along the tree stem, or represent integrations of taper functions (Max & Burkhart, 1976). Pending on the specific portion of the woody tissue of a tree targeted, these equations can be used to predict the merchantable volume to a minimum stem diameter, the total volume of the main stem (rarely), or the volume with or without bark. Volume equations often form the basis for predicting tree biomass by using known specific gravity values for individual tree species.

The traditional approach used for developing volume equations involves destructive sampling. A set of trees believed to represent the targeted tree-species population is identified and the trees are subsequently felled. Diameter measurements made at regular, short intervals along the felled stems provide the data for developing taper functions and ultimately volume equations. The logistical complexity and cost associated with the traditional approach often hinders efforts aiming at the development of new or the refinement of existing volume

equations fit to local conditions. The next biggest impediment in the development of locally-unbiased tree volume equations in many regions is regulations that prohibit the harvesting of trees, or at least of trees over a certain size, and, hence, of obtaining a representative sample of the local tree population for a given species. In such circumstances, non-destructive-tree-sampling approaches could a useful, and likely the only viable, alternative for forest mensuration purposes, assuming of course that they meet accuracy and precision standards.

One such alternative comes in the form of terrestrial light detection and ranging (TLiDAR), a technology capable of generating abstract representations of objects illuminated with laser pulses from one or more near-ground locations. It has been increasingly used lately for forest applications including estimation of dendrometric parameters (Hopkinson *et al.*, 2004) and the assessment of tree-stem dimensionality (Thies *et al.*, 2004). TLiDAR data comprise a usually dense cloud of pulse returns or points from object surfaces precisely georeferenced in three dimensions. Processing of the point cloud could yield detailed representations of tree elements that range in size from the main stem to small twigs or leaves.

Pfeifer *et al.* (2004) managed to derive the structural architecture of trees scanned from multiple stations in the form of skeletons constructed by exploring the three-dimensional mathematical morphology in voxel space and they estimated stem thickness by fitting cylinders to skeleton sections. Many tree branches, especially those in the upper parts of the crowns were not reconstructed. Tree skeletons can also be derived by first organizing the point cloud into a geodesic graph (Verroust and Lazarus, 2000) which constitutes a three-dimensional network of points, each connected to a pre-selected number of closest neighbors. Given a reference point within the cloud, the minimum-distance path that connects each point to the reference can be computed. Connecting the centroids of points that belong to the same minimum-distance class produces the skeleton (Xu *et al.*, 2007). A variant of this approach designed to support automation was implemented by Cheng *et al.* (2007) who applied it to laser points pertaining to two small deciduous trees at leaf-off conditions collected from a single scanning station. For one of the two trees manual editing was necessary to obtain an accurate reconstruction of branch structure.

Bienert *et al.* (2007) used diameter profile fitting on tree stems the locus of which had been previously identified via segmentation of point density rasters and then applied a set of heuristics labeled ‘reliability factors’ to detect and correct diameter over- or underestimation.

More recently Lefsky and McHale (2008) estimated the stem, branch volume and canopy volume of 179 deciduous trees with complex structure scanned at leaf-off conditions in an urban setting. By examining the connectivity of voxels that contained at least one point, they first identified clusters of connected voxels. Processing of the clusters, application of heuristic rules, and manual editing yielded representations of the main stem and for the majority of larger branches, and volume was computed by cylinder fitting. The authors noted that computer memory limitations often enforced voxel resolution coarser than what it was considered optimal and that processing time requirements constrained recursive application of heuristic rules. As a surrogate of tree volume, they compared stem diameter measurements obtained with an optical dendrometer to corresponding diameters of fitted cylinders and reported high correlation between the two (R^2 between 96 and 98 percent).

The objective of this study was to evaluate the utility of TLiDAR data for assessing stem diameter and volume of trees represented by a range of sizes and belonging to two coniferous species. The evaluation process combined elements of the approaches mentioned above and it was based on *in-situ* measurements of stem diameter at regular intervals recorded by experienced tree climbers.

2. Methods

2.1. Field data

Nine fixed-radius (7.32 m) subplots in 4 Forest Inventory and Analysis (FIA) plots located in the Malheur National Forest of eastern Oregon, USA, dominated by coniferous tree species (*Ponderosa pine* and *Abies grandis*) were visited in August of 2009. Omnidirectional laser scanning was performed by using a Leica ScanStation II

instrument positioned on each subplot center. The specifications of the instrument are reported in Table 1. Nominal angular resolution of 0.1146°, or 2 cm at 10 m range, was used during the scanning. A selected subset of subplot trees, and five other trees outside the subplot limits but within the FIA hectare plot footprint (56.4 m radius) were scanned from additional stations with field of view just wide enough to include the entire tree and with finer angular pulse resolution. A total of 10 trees received high-density scans, and 6 of them were scanned from two different stations. Point data from independent scans targeting the same tree were co-registered automatically using proprietary software and reference targets placed appropriately within the scanning scenes. In a subsequent visit in early October 2009, the FIA database records from the latest regular inventory plot visit pertaining to each of the 9 subplots were updated using high precision field measurements that included DBH, height, and the location relative to the subplot center of all subplot trees. Ten of the 78 trees with DBH greater than 12.6 cm in the 9 subplots, and all other 10 trees scanned with high pulse densities were climbed by experience personnel who measured the over-bark diameter of the main tree stem at short, 1 to 1.5 m intervals. The measurements continued upwards the stem until the merchantable diameter threshold (10.2 cm) was reached. A distance tape was used to record the distance along the main stem at which each diameter measurement was obtained. Bark thickness was measured at every fifth diameter measurement. For trees growing on a slope, ground elevation differences between the base of the stem facing the scanning station and the uphill side were recorded, because the latter serves as reference for determining the elevation of the DBH measurements. All 20 climbed trees, a representative sample of the local distribution of tree sizes assessed by using data from the FIA databases, were free of structural deformities and in plain view from the scanning stations.

Table 6. Specifications of Leica ScanStation II instrument

Type	Dual axis compensated pulse laser
Pulse wavelength	532 nm (green)
frequency	< 50 kHz
footprint	6 mm at 50 m
range	300 m at 90% target albedo
Range precision	4 mm
Field of view	360° horizontal, 270° vertical
Scanning pattern	Top-down, bottom-up sequential
Intensity quantization	12-bit

2.2. Assessment of stem diameter and volume

The processes used to assess stem diameter and volume entailed determining the locus of individual tree stems in the point cloud, identifying the laser points belonging to each targeted tree, generating a tree representation in voxel space, processing the voxels to determine connectivity and extract the stem and branch architecture, and finally computing the dimensionality of the stems. All computations were performed using libraries ‘graph’ (Gentleman *et al*., 2010) and ‘RBGL’ (Carey *et* al., 2010) available in the R software (R Development Core Team, 2010) with external calls to custom scripts written in the C programming language.

For all omnidirectional scans a raster (25 cm cell resolution) ground surface was computed using the TiFFs software (Chen, 2007) and then used to convert point elevation to height values. Point subsets that included only returns no higher than 3 m from the ground were processed to compute point frequency (count) rasters of same

resolution and co-registered to the ground surface rasters. The approximate, near-ground, two-dimensional location of tree stems was then retrieved as the center of cell clusters in the count rasters with values much higher than those of surrounding raster cells. All laser points located horizontally within 1 m from the center of each cell cluster and with height value up to 3 m were subsequently identified. These groups of points represented the portion of each tree stem just above ground. For each point group, a single reference point in the scanner-facing direction of the stem and at DBH reference height was manually identified in point visualizations. In directional scans the reference point was selected directly from the point cloud. For each reference laser point, all other laser points horizontally within an arbitrary radius were selected, thereby forming point sets equal to the number of identified trees. The length of the radius was determined by inspecting visualizations of the corresponding point cloud, and was always large enough to include the tree stem in its entirety along with most of the branches, or all of them for free-standing trees (Figure 1a). A voxel representation of the point set was then generated for each tree, with each individual voxel labeled '1' if it contained one or more points, and '0' if empty. The resolution of the representation or, equivalently, the size of each voxel side, was determined by considering the angular resolution of the pulses and the distance between the scanner and the furthest voxel; it ranged between 8 and 15 cm. Note that excessively fine resolutions would produce a large number of filled voxels without any non-empty voxels as immediate neighbors, while with too coarse a resolution there would be too few empty voxels. Controlling the voxel size to ensure an adequate balance between filled and empty voxels was a critical prerequisite for retrieving the stem and branch architecture of the tree.

Each tree's voxel representation was processed with a variant of the connected components algorithm (Lumia *et al.*, 1983). By examining the 26 immediate neighbors of each voxel, the algorithm identified clusters of filled voxels where each member voxel of a cluster had at least one filled voxel as neighbor. Once all clusters had been identified, a unique identifier for each voxel, the coordinates of its center, and the identifiers of and Euclidean distance in three dimensions to all neighboring voxels were organized in files, one for each cluster. In the standard implementation of the algorithm, computer memory availability often constraints the size of the voxel representation that can be processed. In our variant, the algorithm operates in partitions of the representation and thus it can process very large number of voxels, such as those constructed from dense scans of large trees. Owing to the explicit topology present in each voxel representation, neighbor identification and distance computations are far simpler and less demanding on computing resources compared to working directly with the laser cloud points.

The information embedded in each of the cluster files mentioned above is in essence a graph G which can be expressed as G = (V, L), where V is a vector of voxel labels or nodes, and L is a list of minimum one to a maximum of 26 links or edges to the neighbors of each element in V. The distance between two neighboring nodes V_i and V_j can be retrieved from the cluster files and attached to the corresponding link L_{ij}. Since each cluster represented by G was identified by using the connected components algorithm, any V_j can be reached by any V_i by querying G to determined the sequence of links, or path, that has to be followed. Of the usually large number of paths that link two non-neighboring elements of G, of practical interest is the one that minimizes the distance or sum of weights of the visited links. In our application we computed the minimum distance path between the reference node V_r and all other nodes (Figure 1b) using Dijkstra's (1959) algorithm, calculated the number of shortest paths that pass through each node, and then we extracted the paths from V_r to the nodes that define the convex hull of G (Figure 1c). The collection of extracted paths resembles the stem and branch architecture of the tree. Tracking the nodes where two or rarely more paths merge starting from the reference node and progressing upwards the tree while considering the number of minimum paths crossing the nodes adjacent to those examined, leads to the identification and subsequent 'pruning' of the branches and to the subset $G_s = (V_s, L_s)$ of G that contains the nodes and links that represent the main stem of the tree.

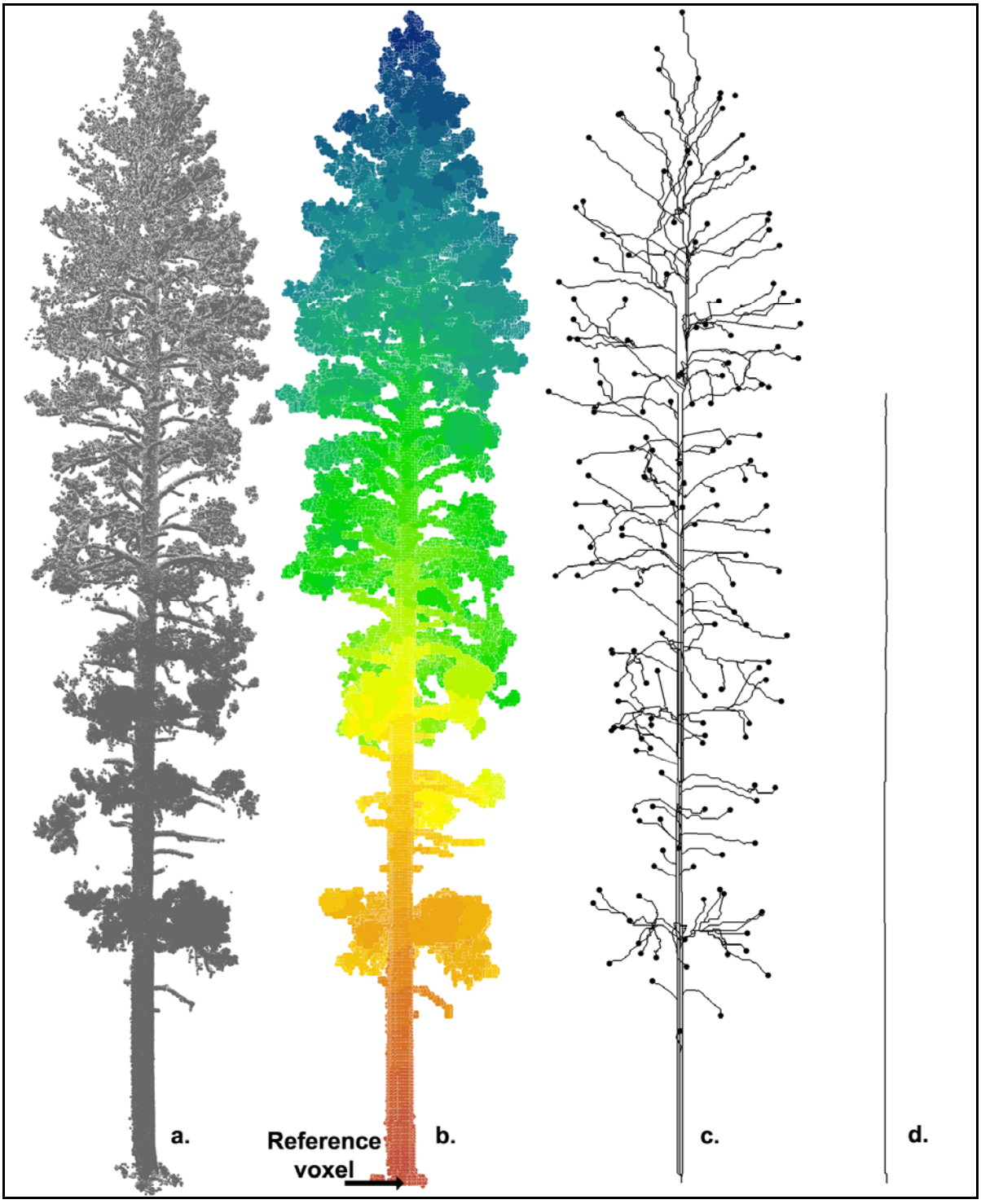


Figure 1. a. Subset of laser point cloud from dense directional scanning, b. Voxel representation colored by Dijkstra's distance from reference voxel at the base of the tree. Cooler colors indicate larger distances. c. Voxels defining the tree's convex hull denoted by dots and corresponding shortest paths to the reference voxel, and d. retrieved stem of the tree.

The sequence of nodes V_s and links L_s in G_s often exhibits localized meandering, especially in segments along the tree stem partially obstructed by branches and foliage situated between the stem and the scanning. Most of this meandering can be removed by fitting a spline through G_s (Figure 1d). In the last step of the process, all points in the immediate vicinity of any G_s element are identified and a sequence of short, often overlapping cylinders is fit to them so that the axis of each cylinder is in parallel to the corresponding spline section. The length and width of each cylinder is evaluated repetitively to minimize the chance that its dimensionality is influenced by points that actually belong to branches or foliage. For cylinders that pass the consistency test, the diameter and the location of the middle of the axis is recorded. Finally the volume of the stem is calculated as the sum of frustum of cones formed by the diameters calculated along the stem's length.

3. Results and discussion

For 28 of 78 subplot and 5 hectare plot trees, approximately one in three, occlusion by other trees or insufficient laser point density precluded assessment of DBH. All 28 trees were scanned with standard density from the subplot center. The relationship between measured and laser-estimated DBH is shown in Figure 2. As expected, the precision of estimates obtained by using high-density scans is higher compared to those obtained by using standard density. For trees scanned with standard density, precision appears to

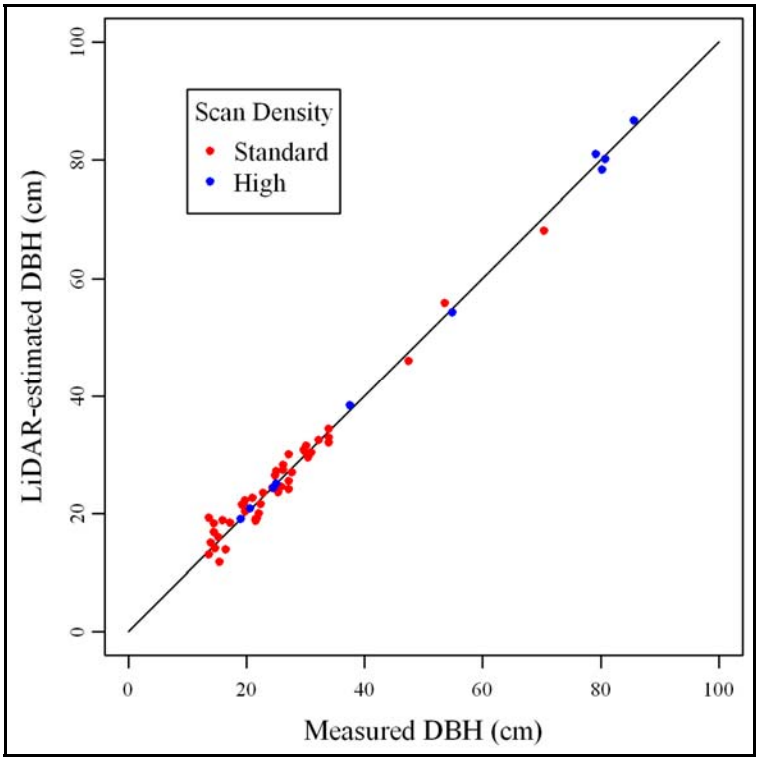


Figure 2. Relationship between measured and LiDAR estimated DBH.

decrease with diameter size. With the standard scan density used in this study, a stem at 10 m from the scanning station and with a diameter of 15 cm would be represented with a mere 7.43 points at a scan line across its axis. A planimetric assessment of diameter for this tree, similar to the one traditionally obtained by using dendrometers, would lead to a 7% underestimation of DBH. In theory, this negative bias is avoided by the cylinder fitting process, which capitalizes on the variability in the distance from the scanner to the cylinder surface to compute the radius of the cylinder and, hence, the diameter of the tree. For the tree in question, however, the expected standard deviation of distance to a horizontal cross section of its main stem at DBH elevation is only 1.65 cm which makes the computed cylinder radius particularly susceptible to noise in the measurement of range by the laser instrument or due to small objects near the surface of the stem. The effect of

these error sources is less pronounced for biggest trees; the expected standard deviation of distance to the stem for a 40 cm DBH tree is 4.39 cm and for an 80 cm tree is 5.59 cm, well above the noise level one would anticipate under normal conditions. On the other hand, departure from stem roundness may introduce error in the assessment of diameter for larger trees (Figure 2).

Stem reconstruction succeeded for all 20 climbed trees, but the accuracy of the reconstruction varied. With one exception, the volume of all trees scanned with standard density was underestimated substantially (mean -21.29%, std 10.25%) (Figure 4a-c). The volume of those scanned with high density

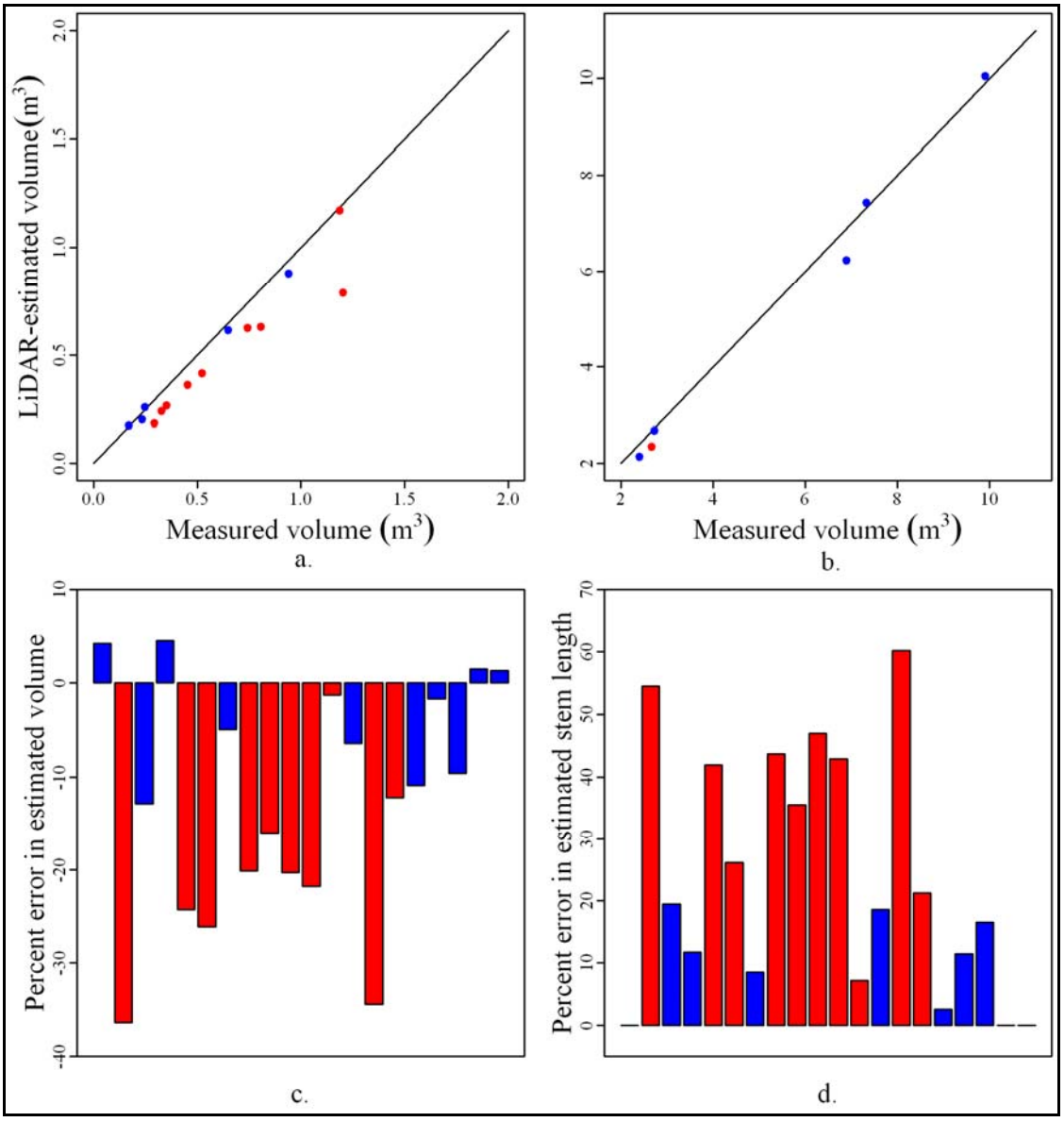


Figure 3. a. and b. Relationship between field measured and laser-estimated volume, c. magnitude of error in laser-derived estimates of merchantable stem volume of tree stem, and d. magnitude of error in estimates of merchantable stem length derived from laser-based reconstruction of individual tree stems. Trees scanned with standard pulse density are represented in red and those with high density in blue.

was also underestimated but the bias was smaller (mean -3.49%, std 6.37), and for 4 trees the bias was actually positive. For four trees the absolute volume error was below 2%. For all 14 trees with underestimation of merchantable volume greater than 4.5%, the stem reconstruction missed a portion of the upper part of the stem, and for 6 of those trees the missed portion was greater than 40% of the total merchantable stem length (Figure 3d). For several trees occluded portions of the stem were followed upwards the tree by portions that were successfully reconstructed (Figure 4). In those cases the volume of the missing part(s) could be retrieved by assuming a linear reduction in stem diameter between the upper diameter of the lower recovered stem section and the lower diameter of the upper section. Although this diameter interpolation improves the accuracy of the overall volume estimate it cannot be used to obtain reliable estimate of diameter that are needed for the development of volume equations.

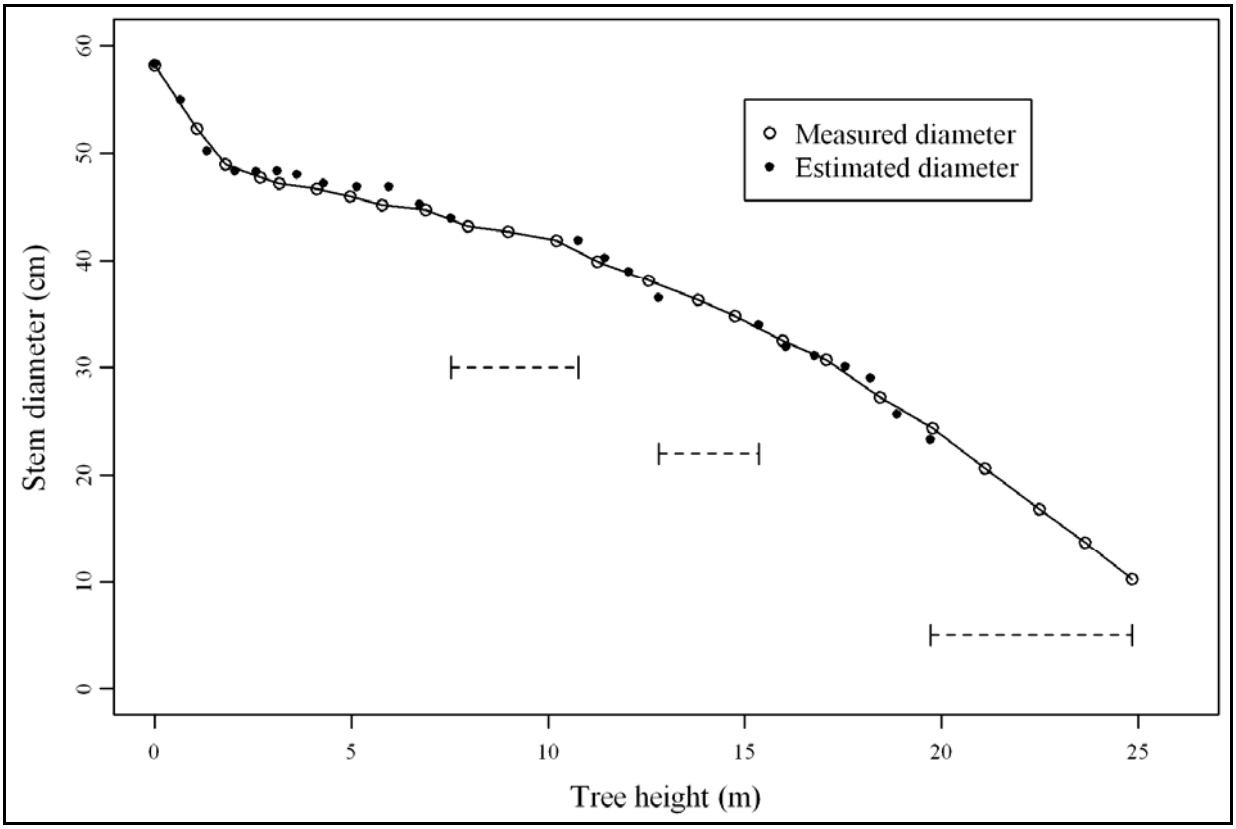


Figure 4. Laser-estimated and in-situ measured stem diameters at locations along the main stem of a ponderosa pine tree. Dash lines represent portions of the stem where reconstruction based on the laser point cloud failed. The continuous line linking measured diameters is for visualization purposes only.

These findings corroborate previous studies in which the upper parts of tree stems could not be reconstructed from terrestrial LiDAR data. When omnidirectional terrestrial LiDAR scanning is used to obtain a record of the plot or subplot vegetation structure, it is unlikely to expect that all trees in the scene will be reconstructed, except perhaps at leaf-off conditions. Where, however, only one tree is targeted for special purposes, including the development of localized volume equations, the operator of the laser instrument has a lot of flexibility in determining the location of the scanning stations, so as to maximize the chances for a successful, accurate, and precise reconstruction of the targeted stem. Experience from this study suggests that for a given pulse density at the tree level and tree species with predominantly horizontal branch structure, it is preferable to operate the

scanner from a longer distance and higher density setting than from closely underneath with lower density. This is because in the former setup the pulse has to travel shorter distance through foliage and branches.

4. Conclusion

We evaluated the potential of terrestrial LiDAR as an alternative approach to destructive tree sampling for assessing the dimensionality of tree stems and for the development of volume equations. We managed to estimate the merchantable volume with acceptable amount of error (< 2%) and retrieve the entire length of the merchantable stem only for 2 of the 20 trees we considered, although even for those two trees, a few stems diameters had to be interpolated. At the present state of algorithmic development, an unobstructed section at or near the top of the merchantable tree stem is a prerequisite for obtaining an accurate estimate of stem volume using terrestrial LiDAR data.

References

Bienert, A., S. Scheller, E. Keane, F. Mohan, and C. Nugent. 2007. Tree detection and diameter estimations by analysis of forest terrestrial laserscanner point clouds. In *Proc. ISPRS Commission III Workshop*, 36(3):50-55.

Carey, V., L. Long, and R. Gentleman. 2010. RBGL: An interface to the BOOST graph library. R package version 1.24.0.URL http://CRAN.R-project.org/package=RBGL.

Chen, Q. 2007. Airborne lidar data processing and information extraction. *Photogrammetric Engineering and Remote Sensing*, 73(2): 109-112.

Cheng , Z.L., X.P. Zhang, and B.Q. Chen. 2007. Simple reconstruction of tree branches from a single range image. *J. of Computer Science and Technology*, 22(6):846-858.

Dijkstra, E.W. 1959. A note on two problems in connection with graphs. *Numerische Mathematik* 1:269–271.

Gentleman, E.W., W. Huber, and S. Falcon. 2010. graph: A package to handle graph data structures. R package version 1.26.0. URL http://CRAN.R-project.org/package=graph.

Hopkinson, C., L. Chasmer, C. Young-Pow, and P. Treitz. 2004. Assessing forest metrics with a ground-based scanning lidar. *Canadian J. Forest Research*, 34:573-583.

Lefsky, M., and M. McHale. 2008. Volume estimates of trees with complex architecture from terrestrial laser scanning. *J. of Applied Remote Sensing*, 2:023521, 19 pp.

Lumia, R., L. Shapiro, and O. Zuniga. 1983. A new connected components algorithm for virtual memory computers. *Computer Vision Graphics and Image Processing*, 22:287-300.

Max, T.A., and H.E. Burkhart, H.E. 1976. Segmented polynomial regression applied to taper equations. *Forest Science*. 22:283-289.

Pfeifer, N., B. Gorte, and D. Winterhalder. 2004. Automatic reconstruction of single trees from terrestrial laser scanner data. *Int. Archives of Photogrammetry and Remote Sensing*, Vol. XXXV, B5, pp. 114-119.

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, URL http://www.R-project.org.

Thies, M., N. Pfeifer, D. Winterhalder, and B.G.H. Gorte. 2004. Three-dimensional reconstruction of stems for assessment of taper, sweep, and lean based on laser scanning of standing trees. *Scandinavian J. Forest Research*, 19(6):571-581.

Verroust, A., and F. Lazarus. 2000. Extracting skeletal curves from 3D scattered data. *Visual Computer*, 16:15-25.

Xu, H., N. Gossett, and B. Chen. 2007. Knowledge and heuristic-based modeling of laser-scanned trees. *ACM Transactions on Graphics*, 26(4).

Tree architecture and biomass assessment from terrestrial LiDAR measurements: a case study for some Beech trees (*Fagus sylvatica*)

MATHIEU DASSOT†‡, ADELIN BARBACCI†‡, AURELIE COLIN†, MERIEM FOURNIER‡ and THIERY CONSTANT†*

† INRA, UMR1092 LERFoB , F-54280 Champenoux, France

‡ AgroParisTech, UMR 1092 LERFoB, F-54000 Nancy, France

Running head: Terrestrial LiDAR for tree architecture and biomass assessment

Abstract

The assessment of the spatial distribution of the tree biomass by means of terrestrial LiDAR scanning in forest environment still remains a challenge. This paper aims at presenting a method coupling terrestrial LiDAR measurements (FARO LS 880 HE) to a point cloud processing software (InnovMetric Software Inc., PolyWorks) allowing for tree architecture and wood volume distribution modeling, at the tree scale, for two Beech trees (*Fagus sylvatica*). This method is compared to a detailed and labour-intensive way of tree biomass distribution assessment (tacheometer and manual measurements), considered here as the reference. Our results quantify the appropriateness of the T-LiDAR method to estimate wood volumes and to correctly represent the three-dimensional architecture of a tree and the spatial distribution of its woody biomass. However, some measurement parameters must be carefully considered and further refinements are still needed to automate the extraction of information from point clouds.

1. Introduction

The architecture acquired by a tree during growth is the result of its adaptation to environment, particularly in terms of competition for space, accession to light resources and posture control against gravity. Understanding the spatial distribution of the tree biomass is a key factor in assessing the functioning of forest ecosystems and estimating wood resources (timber wood and bioenergy resources). However, obtaining such information remains time-consuming and often destructive by traditional ways (field measurements). For a few years, airborne LiDAR scanning have been used to obtain non-destructive information about forest structure at the stand or the forest scale, but this method provides rather limited information at the tree scale. Nowadays, a complementary level can be achieved thanks to terrestrial LiDAR (T-LiDAR) scanning, allowing for the rapid, almost complete and accurate digitization of single trees or entire plots.

T-LiDAR scanning in forest environment to obtain the structural parameters of standing trees is a very active research area and appears to be useful for various fields of forest science such as forest inventories (Hopkinson *et al*. 2004, Wezyk *et al.* 2007), management (Murphy 2008) and ecology (Michel *et al*. 2008). Also, several works have been done in the topic of tree topology reconstruction by means of voxel-based processing (Gorte and Winterhalder 2004) or cylinder fitting (Thies *et al.* 2004). However, such works mainly focused on the reconstruction of the stem or the main branches of a tree. The assessment of the complete tree architecture and the three-dimensional (3D) distribution of the woody volumes for heavy branching trees by means of T-LiDAR scanning still remains a challenge.

This paper is part of a study on tree biomechanics which aims at assessing the effect of the unbalanced biomass of a tree on its internal wood quality, hence the effort devoted to model the tree architecture. In this paper, the objective is to present the first step of this study, i.e. to evaluate the capabilities of T-LiDAR scanning for tree architecture modeling and wood volume estimation by comparing it to a very detailed, labour-intensive and destructive way of tree architecture and biomass assessment.

2. Methods

2.1 Selected trees

Two beech trees (A and B) coming from a mixed stand dominated by sessile oak trees were selected in the state forest of Bride in North-East France (Table 1). The tree A was composed of 42 branches with insertion diameters ranging from 3 to 62 mm, and the tree B was composed of 19 branches with insertion diameters ranging from 26 to 78 mm (manual measurements). Both manual and T-LiDAR measurements were performed in leaf-off conditions, each method providing a three-dimensional model of the woody structure of the tree.

Table 1. Manual and T-LiDAR estimations of the main characteristics of the two trees.

	Tree A		Tree B	
	DBH (cm)	Tree height (m)	DBH (cm)	Tree height (m)
Manual	20.37	20.22	31.83	24.27
T-LiDAR	18.57	19.26	29.73	23.51

2.2 Manual description of the tree architecture and biomass

The first step of our study was to establish a reference 3D model for each tree and to manually estimate their woody volume, for the stem and for each branch (Figure 1). The method used was derived from the methodology already developed for other studies (Constant *et al*. 2003; Barbacci *et al*. 2009). This manual method allowing for the 3D reconstruction of trees resulted from the coupling of tacheometer (Leica TCR 307) measurements, i.e. 3D points, to manual measurements (i.e. diameter and length).

The tacheometer was used to digitize the 3D geometry of the standing tree, firstly for the trunk and the insertion segments of branches. Secondly, each branch was cut and placed in a branch-holder on the ground. The distal extremity of the distal quasi-linear segment belonging to a branch axis was digitized with the tacheometer. Then the segment was cut and its diameters measured as well as its length, giving access to its geometrical volume. Gradually, the whole branch was measured, and the data of its main segments processed to constitute a 3D branch mock-up which was virtually inserted on its corresponding starting segment from the trunk by a geometrical transformation. The stem was also processed in the same way. The result is a reference 3D tree mock-up composed by straight segments and corresponding to the spatial distribution of the main tree axes and their woody volumes which will be further mentioned as the geometrical estimation.

Furthermore, the weight of each main segment of branches (including the finest branching elements connected) was also recorded, allowing the weight of each branch to be determined. Additionally, small samples distributed in the whole tree were collected to assess a conversion factor between mass and volume from laboratory measurements. Their analysis led us to use the same uniform conversion factor for all the branches belonging to the same tree and another one for the stem. The estimation of the woody volume of the tree elements by this way is considered as the reference.

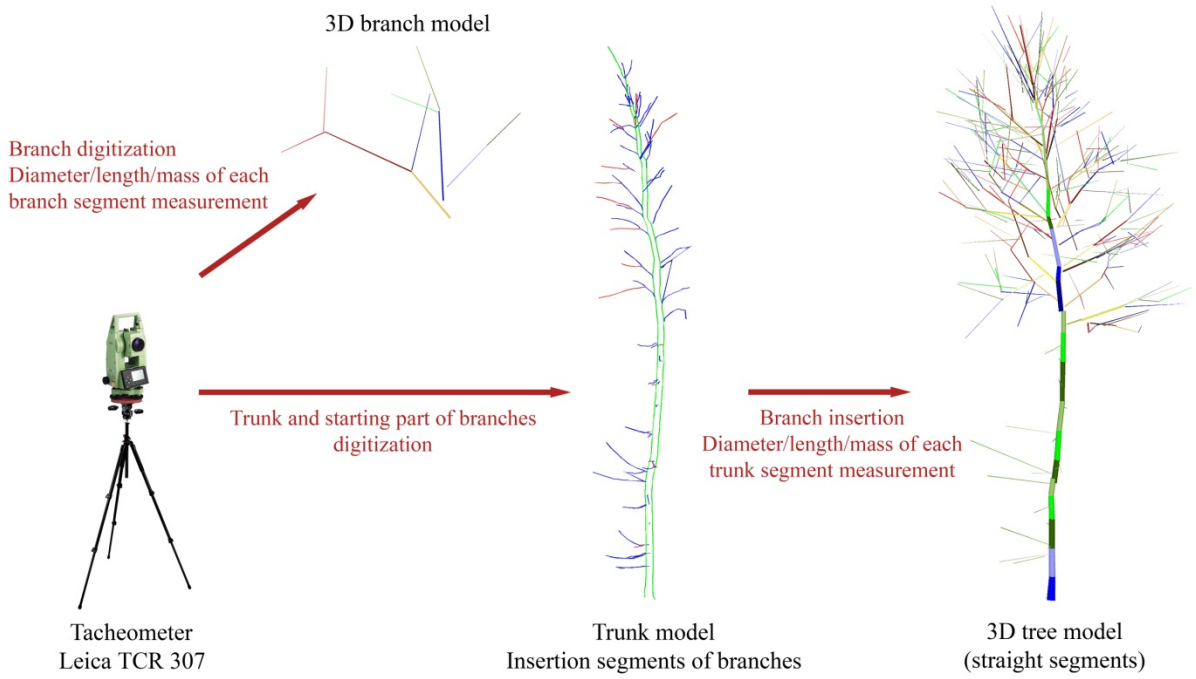


Figure 1. Manual assessment of the tree architecture (example of the tree A)

2.3 T-LiDAR scanning for tree architecture and biomass assessment

Before the previously described measurements, T-LiDAR scanning was performed with a phase-shift-based scanner, the FARO LS 880 HE. Several scans were made around each tree to reduce occlusions and ensure the digitization of the major part of the tree material. Three scans were made around the tree A (one scan every 120°) and two opposite scans were made for the tree B. The scans were performed with a distance from the trunk of about 6 m, with a scanning resolution of 1/2, i.e. a spatial resolution of about one point every 4 mm at 25 m. All the scans of the same tree were merged under the IM Align module of the PolyWorks software (point cloud processing, InnovMetric Software Inc.) by identifying at least three points common in both scans.

The characterization of the tree axes through geometrical reconstruction of point clouds was made thanks to the IM Inspect module of the PolyWorks software, which already proved its potential for forest inventory measurements (Hopkinson *et al*. 2004, Tansey *et al*. 2009). Several processing steps were defined to obtain the woody volume of each tree axis from its curved length and the variation of diameters along it. Firstly, polylines (a succession of points characterized by 3D coordinates) were fitted along both the stem and branches to obtain their 3D geometry and their curved length.

Two options were used for defining polylines depending on the size of the tree axis, (i) an automatic method allowing a polyline to be fitted as the neutral line of a tube-shaped axis (typically for the stem and the largest branches of the tree), and (ii) a manual method allowing a polyline to be anchored on point clouds (for the thin branches). Secondly, cylinder sequences were fitted on point clouds to provide the variation of diameters along polylines. If not measurable, a diameter of 3 mm was attributed to the extremity of the axis (approximately the size of a bud or a shoot). This geometrical reconstruction allowed to obtain a complex 3D mock-up of the tree (Figure 2). Then, polylines and cylinders characteristics were exported in a spreadsheet and diameters were attributed to their corresponding point of the polyline. Finally, volumes were estimated by applying a frustum of cone formula between two successive diameters, leading to the calculation of several volumes along a polyline which can be added to obtain a global volume for each axis, for each branch and for the entire tree.

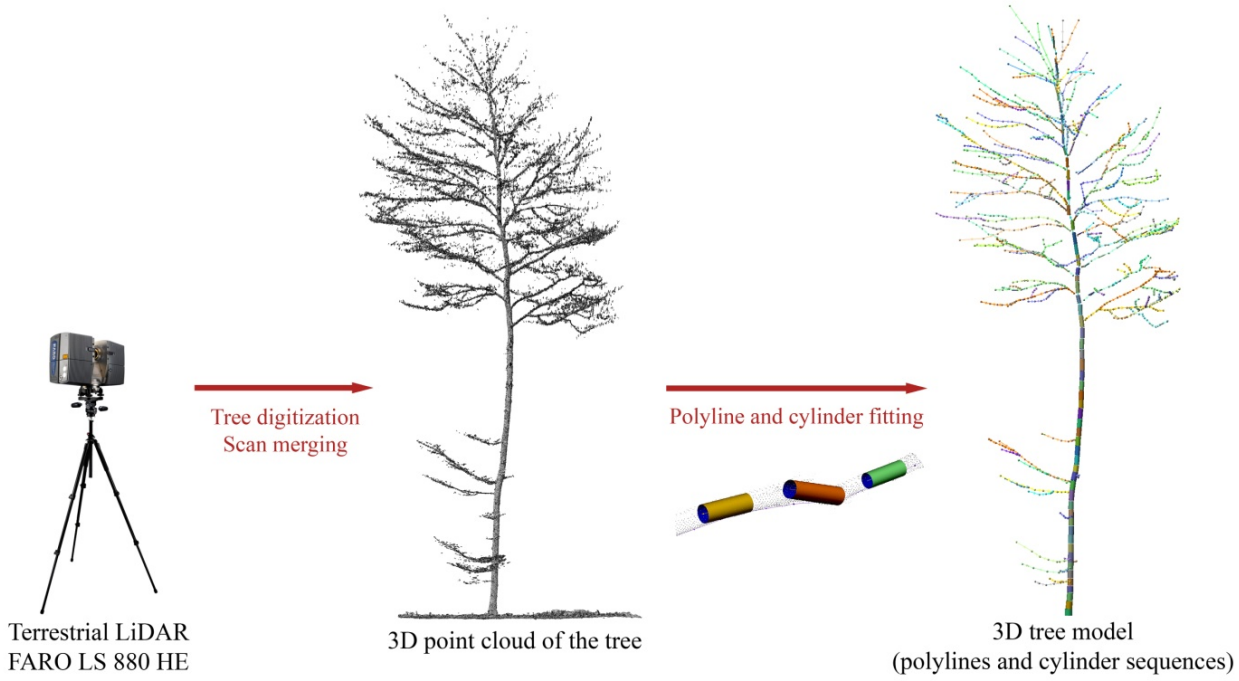


Figure 2. Assessment of the tree architecture using T-LiDAR (example of the tree A)

3. Results and discussion

3.1 Tree architecture

The visual comparison between reconstruction from point cloud and manual modeling of the tree architecture is very satisfactory, more especially for the tree A for which the complex tree structure appear to be quite complete, even for the thin branching elements. The comparison between both methods for the tree B reveals that a lot of the tree axes were missing, particularly for the thin branching elements and especially at the top of the tree (Figure 3). Keeping in mind that very thin shoots were not geometrically described, it seems that even in good scanning conditions like for the tree A the geometry of such woody elements cannot be reached without special conditions (higher number of scans leading to limited occlusions, higher resolution, short distance and no wind). Some differences between both models are also visible for the insertion angle of branches.

These differences are due to inaccuracies in determining the same insertion segment between those measured on the standing tree with the tacheometer and its simplified geometrical description made by linear segments on the ground. A promising result about branching was that, in both cases, no major difficulty was encountered to find the same number of branches of the order two with insertion diameters by both methods and whatever the completeness of T-LiDAR data (only some epicormic shoots of the tree A, representing insignificant volumes, could not be detected). These characteristics can be used as inputs to complete missing volumes through allometric relationships, or as criteria for other purposes linked to crown description or wood quality assessment.

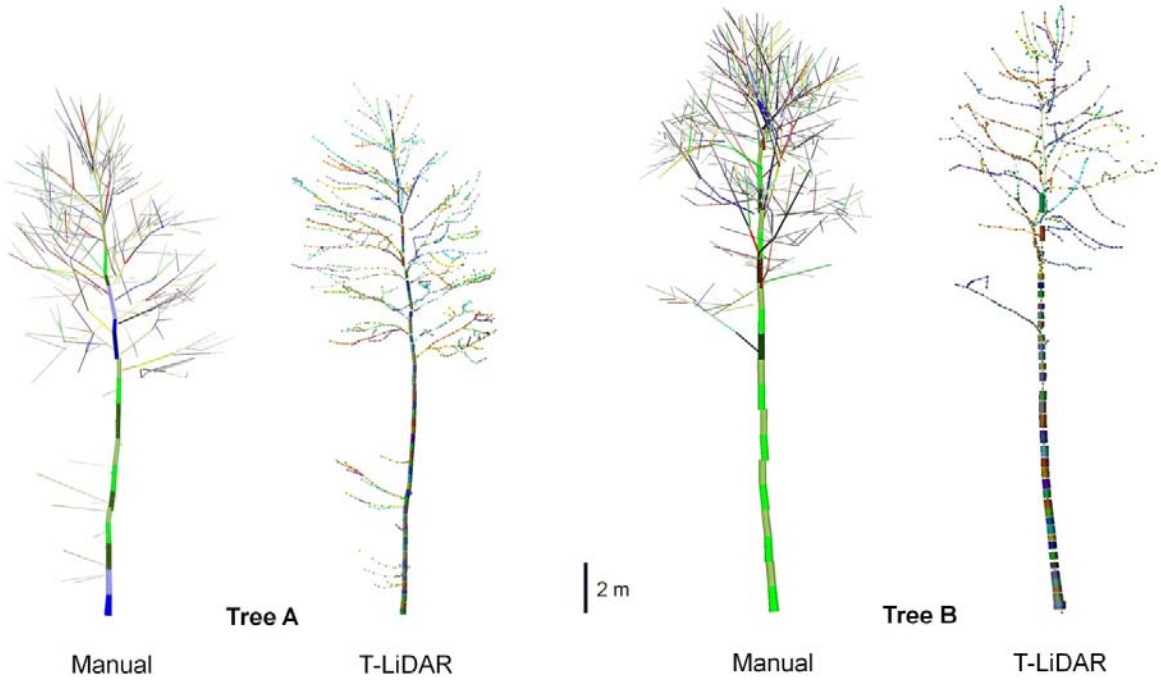


Figure 3. Comparison between the two ways of tree architecture assessment

3.2 Global volumes of trees

As presented in Table 2, the estimation of the wood volumes for a deciduous tree by means of T-LiDAR scanning gives rather good values and confirms its ability. Whether the stem volumes can be determined with a difference of less than 20% for a tree volume ranging from 0.3 to 1 m3, the branch volumes can differ depending on (i) the tree architecture itself (occlusions) and (ii) the tree environment and the weather circumstances governing the scanning conditions. Nevertheless, these results have to be weighted by the volume proportion of the branches which is very close to 20 % for both trees.

		Geometrical volume	Volume derived from mass	T-LiDAR volume
Tree A	Stem volume (m3)	0.316 (-10.8%)	0.298 (-5.4%)	0.282
	Volume of branches (m3)	0.054 (11.1%)	0.074 (-18.9%)	0.060
	Total tree volume (m3)	0.370 (-7.8%)	0.372 (-8.3%)	0.341

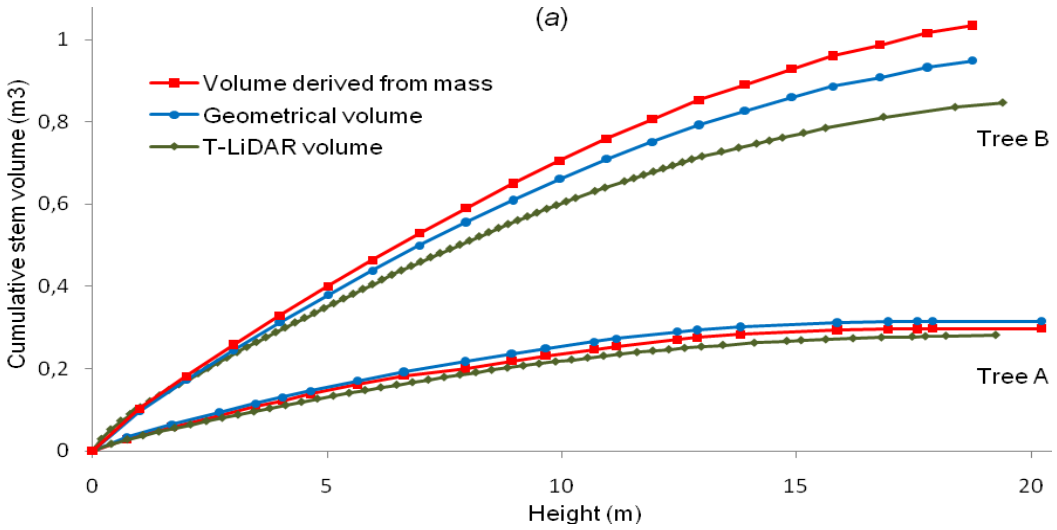
Tree B	Stem volume (m3)	0.949 (-10.8%)	1.034 (-18.1%)	0.847
	Volume of branches (m3)	0.211 (-40.7%)	0.219 (-42.8%)	0.125
	Total tree volume (m3)	1.160 (-16.2%)	1.253 (-22.5%)	0.972

Table 2. Comparison between the three ways of wood volume estimation. The error percentage for T-LiDAR estimation appears into brackets.

For the tree A, T-LiDAR data processing underestimates the total volume of the tree of 7.8% and 8.3% when compared to the geometrical volume and the volume derived from mass, respectively. The same comparison for the stem volume shows underestimations of 10.8% and 5.4%. The estimation of the total branch volume thanks to T-LiDAR shows an overestimation of 11.1% when compared to geometrical measurements, and an underestimation of 18.9% when compared to the volume derived from mass.

For the tree B, the T-LiDAR method underestimates the total volume of the tree of 16.2% and 22.5% when compared to the geometrical volume and the volume derived from mass, respectively. The T-LiDAR stem volume shows underestimations of 10.8% and 18.1%. The estimation of the total branch volume thanks to T-LiDAR scanning shows an underestimation of 40.7% when compared to geometrical measurements, and an underestimation of 42.8% when compared to the volume derived from mass. Once again, these bad results must be appreciated with respect to the fraction of the total volume concerned.

The cumulative wood volumes according to height for the stem and branches of the two trees are presented in Figure 4. For the tree A, having the smallest volume, the curves of the cumulative stem volume are very close to each other. For tree B, it seems to have clear trends especially after three meters height (Figure 4 (*a*)).



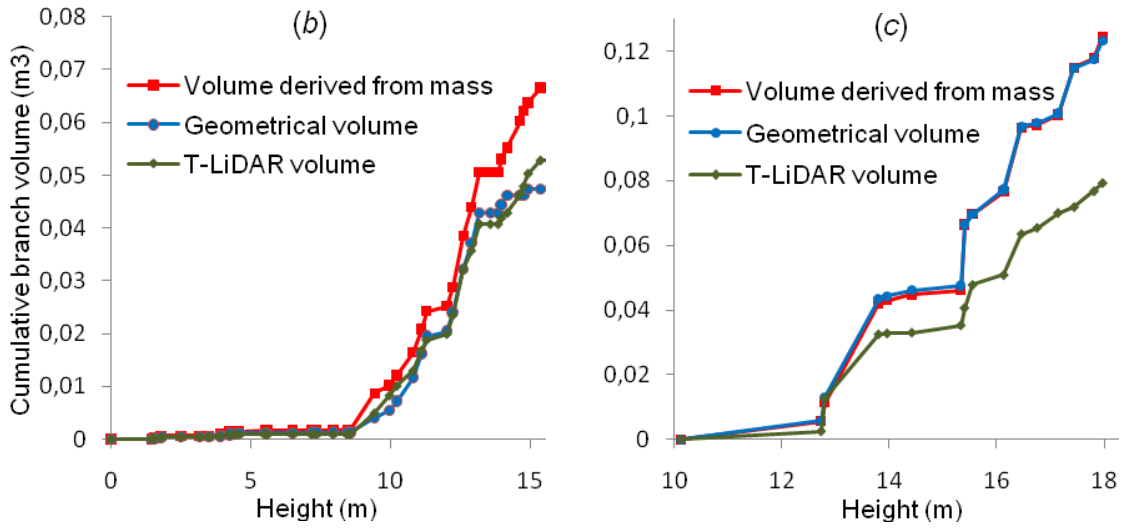


Figure 4. Cumulative wood volumes for the stems (a), for branches of the tree A (b) and for branches of the tree B (c).

Without being able to distinguish the factors, some explanations can be advanced. Firstly, our reference value can be slightly overestimated because of the conversion factor, which appears very close to one in any case. Indeed, this factor came from measurements performed on clearwood specimens and can lead to such a bias whether the conversion factor is overestimated. Secondly, geometrical volumes are estimated from perimeter measurement to obtain equivalent diameters, and if the cross-section was not circular, but more elliptical for instance, this induces a systematic bias overestimating the equivalent volume. Furthermore, the cross section of the stem probably has a trend to be less and less circular with height because of the presence of living or dead branches. This characteristic played also a role when obtaining T-LiDAR volumes because of the choice of a cylinder as primitive and the fitting procedure which can made diameters to be impacted by irregularities (noise or bark features). The use of contours (such as polygonal models) to compute their cross-section areas can be a way to test the validity of these assumptions in the future.

About the branches, the major part of the thin branching elements of the tree A was detected and successfully modelled, hence a logical overestimation when compared to the geometrical measurements (only concerning the main axes of the branches) and an underestimation when compared to the volume derived from mass. Despite the small diameters of its branches, branch volumes were quite accurately estimated for the tree A (Figure 4 (*b*)). Branches of the tree B globally had larger insertion diameters but many branches were not properly described due to occlusions and the quality of the 3D cloud built from only 2 scans, hence higher error levels when compared to the tree A (Figure 4 (*c*)). The effect of wind can have influenced the quality of the data but the scans were performed in wind conditions rather favorable.

3.3 Effect of branch diameter and height on volume estimation

The difference between the T-LiDAR estimation of the volume of branches and the volume derived from mass according to their height and their insertion diameter is presented in Figure 5 (the size of the discs being proportional to the insertion diameter manually measured). It appears that the modeling processes leads to both underestimations and overestimations of the volume of branches for the tree A (Figure 5(*a*)) and the tree B (Figure 5(*b*)). If no trend can be found for the impact of the height on volume estimation, underestimations are always found for branches having the largest insertion diameters. As expected, it also shows that the error in estimating the branch volume with T-LiDAR is higher for branches having a large insertion diameter. These

branches, often more complex and ramified, are not always well-described in point clouds unless the scan number increases and several thin branching elements cannot be detected or properly digitized.

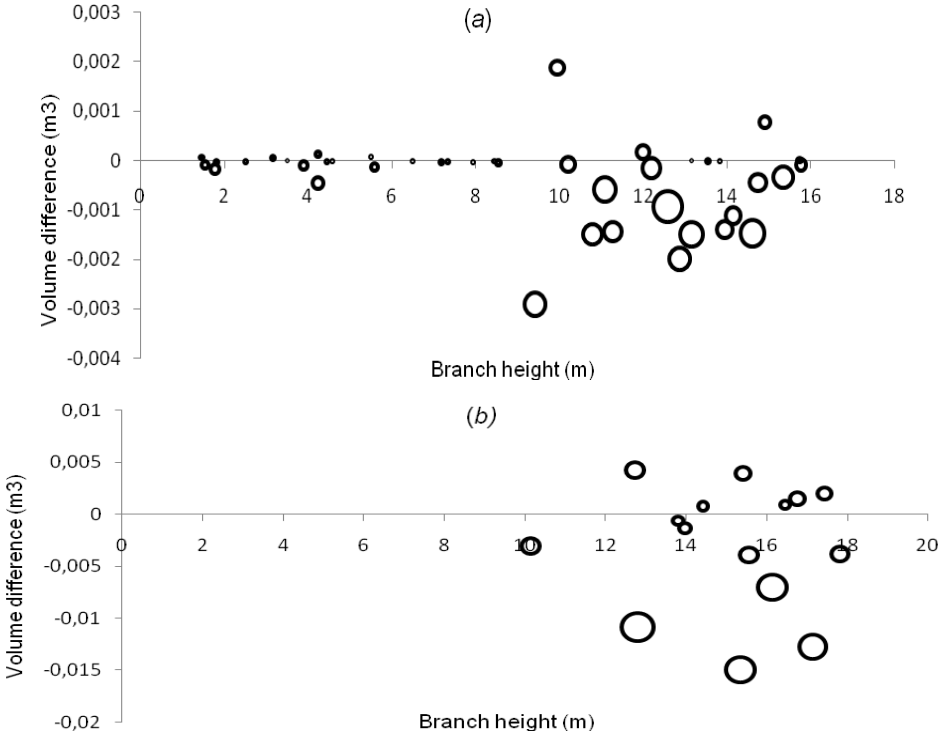


Figure 5. Difference between T-LiDAR and manual estimation of branch volume (derived from mass) according to branch height (x axis) and insertion diameter (proportional to the ring size), for the tree A (a) and the tree B (b). Volumes of the last four branches of the tree B were aggregated into one because no correspondences were established between T-LiDAR and manual measurements.

As presented in Figure 6, and contrary to the tree B for which branches never had an insertion diameter less than 2 cm, the tree A (Figure 6(*a*)) had several branches (epicormic shoots having a height between 1 m and 9 m) for which the insertion diameter was very small (less than 1 cm). For several of these branches, no diameter (and therefore no volume) could be found by point cloud processing. However, the correlation curve between manual and T-LiDAR measurements for insertion diameters shows a good value for the tree A ($R^2 = 0.82$). This correlation curve decreases to $R^2 = 0.4$ for the tree B. However, this result is largely influenced by a particular point which penalized the overall trend. In addition to diameter estimation, the other factor influencing the volume of branches could be the length of the different axes detected from T-LiDAR data, but this requires an analysis of the architecture of the branch which was not yet carried out.

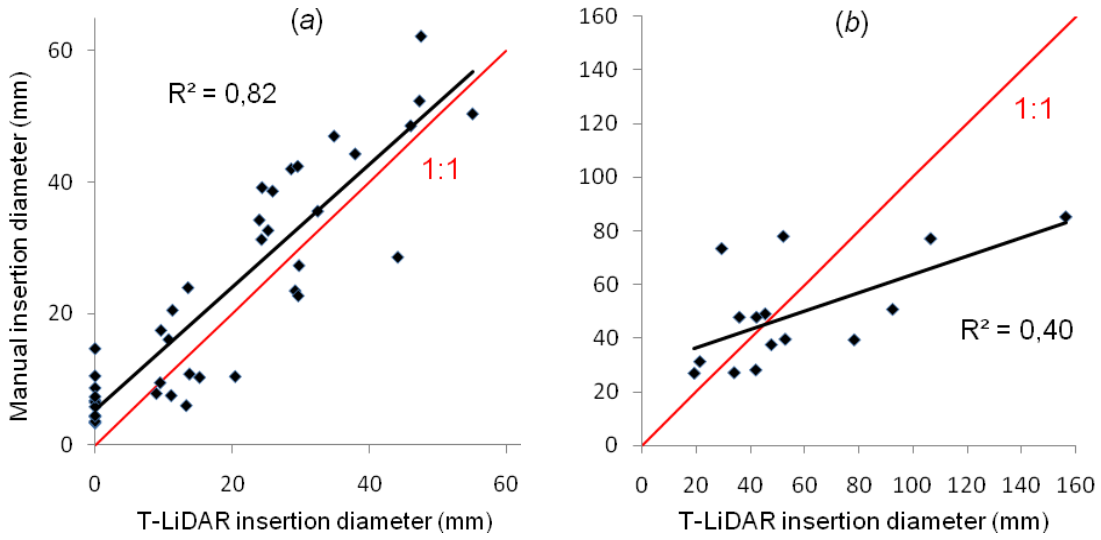


Figure 6. Correlation between T-LiDAR and manual measurements of the insertion diameter of branches, for the tree A (a) and the tree B (b).

3.4. Identification of measurement set-up key-factors

Our results show the appropriateness of the T-LiDAR method to correctly assess the architecture and the global wood volumes of a tree, but the quality of point clouds, and therefore the quality of polyline and cylinder fitting, may differ according to the measurement set-up used in the field. The results obtained for the tree A are encouraging, in contrast with the higher error levels obtained for the tree B.

The main problem of our study was the rate of occlusions in the upper part of trees caused by heavy branching. This phenomenon, which not only affects the trees presented here but all trees digitized by means of T-LiDAR scanners in forest environment, leads to the poor description of branches in the upper part of trees. This effect can be reduced by increasing the number of scans around trees. The comparison between the tree A (three scans) and the tree B (two scans) gives an example of the profit resulting. It shows that both polyline and cylinder fitting find benefits in increasing the number of scans, allowing for the detection of even more tree axes and their better description. For the tree B, contrary to the tree A, we cannot obtain the complete coverage of the trunk and numerous elements of branches were not detectable in the point cloud. A scan number of four (every 90°) seems to be very suitable to almost entirely digitize the structure of a tree.

Weather conditions must also be carefully considered. Particularly, wind is the most important factors reducing the quality of point clouds and makes the tree axes moving. It results in the poor detection of the finest branching elements, the poor description of the tree axes and makes more difficult cylinder fitting leading to aberrant diameter estimations.

For both trees, another issue lies in the fact that scans were merged under the PolyWorks software by means of points considered as identical between all the scans of a tree. This method led in merging biases of some millimeters, whereas using reference spheres in the field would have improved the merging quality (often less than 1.5 mm).

3.5 Working hardness and convenience

The detailed manual measurements presented in this study required the mobilization of four people during two week (one week per tree), whereas T-LiDAR scanning was performed by two people during half a day and point cloud processing carried out by only one person during two weeks (one week per tree, according to its complexity). Moreover, fitting polylines appears as a rapid method to assess both the 3D geometry and the

curving length of the tree axes in the same time and up to the finest diameter classes, with lesser handling than field measurements. Moreover, the non-destructive nature of T-LiDAR scanning is a significant advantage, particularly for periodic measurements studying the evolution of the geometry of a tree and its environment. Nevertheless, even if usable and relevant, our method remains laborious and future work is needed to automate point cloud processing, and therefore to facilitate the application of T-LiDAR techniques for forestry.

4. Conclusions

This study contributes to quantify some factors influencing the assessment of the tree architecture and biomass by means of T-LiDAR scanning. As demonstrated in this work, T-LiDAR scanners are able to correctly estimate the volume of the woody biomass of a deciduous standing tree in forest environment, but also to catch very detailed information about its woody architecture if sufficient time is devoted to perform enough scans, a full description requiring at least three scans. However, this demonstration would be stronger with a large number of trees. This future work is already in progress for about 100 trees belonging to different forest environments and species, with an adapted scanning protocol (FARO Photon 120 laser scanner, four scans per tree, resolution 1/4, merging with reference spheres). Efficient tools allowing for automated data processing providing crown architecture would also be welcome.

Acknowledgments

This work is supported by the French National Research Agency (ANR) through the EMERGE project (ANR BIOENERGIE 2008 BIOE-003) managed by Christine Deleuze, which aims at establishing reliable and generic distribution models of the tree biomass for several tree species, sizes and silviculture types. A part of this work is also funded by the French Forestry Office (ONF) through the contract Modelfor (2005-2010) and the French National Forest Inventory (IFN) through the contract 2008-CER-4148. Special thanks to Pierre Dalarun (Duwe-3d Company) for his help during the reconstruction processes.

References

BARBACCI, A., CONSTANT, T., NEPVEU, G., 2009, Theoretical and experimental study of a mechanical model describing the trunk behaviour of mature beech trees (*Fagus sylvatica* L.) under the static loading of the crown. *Trees*, **23**, pp. 1137-1147.

CONSTANT, T., MOTHE, F., BADIA, M.A., SAINT-ANDRÉ, L., 2003, How to relate the standing tree shape to internal characteristics : proposal of an experimental method applied to poplar trees. *Annals of Forest Science*, **60**, pp. 371-378.

GORTE, B., WINTERHALDER, D., 2004, Reconstruction of laser-scanned trees using filter operations in the 3D-raster domain. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, **36**, pp. 39-44. Available online at: http://www.isprs.org/proceedings/XXXVI/8-W2/GORTE.pdf (accessed 28 July 2010).

HOPKINSON, C., CHASMER, L., YOUNG-POW, C., TREITZ, P., 2004, Assessing forest metrics with a ground-based scanning lidar. *Canadian Journal of Forest Research*, **34**, pp. 573-583.

MICHEL, P., JENKINS, J., MASON, N., DICKINSON, K.J.M., JAMIESON, I.G., 2008, Assessing the ecological application of lasergrammetric techniques to measure fine-scale vegetation structure. *Ecological Informatics*, **3**, pp. 309-320.

MURPHY, G., 2008, Determining stand value and log product yields using terrestrial LIDAR and optimal bucking: a case study. *Journal of Forestry*, **106**, pp. 317-324.

TANSEY, K., SELMES, N., ANSTEE, A., TATE, N.J., DENNISS, A., 2009, Estimating tree and stand variables in a Corsican Pine woodland from terrestrial laser scanner data. *International Journal of Remote Sensing*, **30**, pp. 5195-5209.

THIES, M., PFEIFER, N., WINTERHALDER, D., GORTE, B.G.H., 2004, Three-dimensional reconstruction of stems for assessment of taper, sweep, and lean based on laser-scanning of standing trees. *Scandinavian Journal of Forest Research*, **19**, pp. 571–581.

WEZYK, P., KOZIOL, K., GLISTA, M., PIERZCHALSKI, M., 2007, Terrestrial laser scanning versus traditional forest inventory: First results from the Polish forests. Laser scanning 2007 and Silvilaser 2007. ISPRS Commission III WorkShop, Vol. 36, Part 3, Espoo, Finland, Sept. 12-14, 2007. Available online at http://www.isprs.org/proceedings/XXXVI/3-W52/final_papers/Wezyk_2007.pdf (accessed 28 July 2010).

Relating tree height variations to peat dome slope in Central Kalimantan, Indonesia using small-footprint airborne LiDAR data

*Hans-Dieter Viktor Boehm *†, Veraldo Liesenberg ‡ and Juergen Frank †*

viktorboehm@t-online.de

† Kalteng Consultants [www.kalteng.org], Kirchstockacher Weg 2, D-085635, Hoehenkirchen, Germany.

‡ Faculty of Geosciences, Geotechnique and Mining, TU Bergakademie Freiberg, Bernhard-von-Cotta-Str. 2, D-09599, Freiberg, Germany

Abstract

We investigated how measures derived from small footprint airborne Light Detection and Ranging (LiDAR) data can be used to evaluate the relationships between tree height (represented by the digital surface model, DSM) and peat swamp dome slope (digital terrain model, DTM). In August 2007 we mapped by helicopter different peat swamp forest (PSF) environments with Riegl LiDAR Technology LMS-Q560 in the Mawas area of the Ex-Mega Rice Project (EMRP) and in the Sabangau National Park (Central Kalimantan, Indonesia). In each LiDAR transect we used sample plots of 100x100m along the flown acquisition with a distance of 200m from each other. In each sample plot we calculated the peat dome slope and the tree height and we evaluate linear regression between both parameters. Our results showed that: a) the tree height increases by approximately 5m if the peat dome slope increases from 0.5 to 1.5m pro mille (m/km) from Sabangau River to the peat dome; and b) there is a relationship between both variables ($r^2 \geq 0.60$) that may be related with the permeability, interflow, water storage capability and nutrient availability in the domes. In a near future we intend to conduct regression analysis considering the field-measured stem volumes against together with LiDAR-derived tree height and Ortho-photos in the frame of the REDD knowledge.

1. Introduction

Natural tropical Peat swamp forests (PSF) are important for their rich biodiversity and because they represent important carbon pool (Page et al., 2002). However, PSF are decreasing due to conversion into farm land, by excessive draining, the use of shifting cultivation on a large scale, illegal logging and forest fire. Furthermore, managed land cover types, degraded ecosystems and large deforested areas have been also reverted to successional forest stages. This increases the interest for understanding in an ecological point of view and mapping such environments as they are recognized as an important source of carbon released in the atmosphere (Sorensen 1993, Page et al., 2002, Jaenicke et al., 2008, Ballhorn et al., 2009).

According to Hyde et al. (2007) airborne Light Detection and Ranging (LiDAR) data is nowadays the best single sensor to investigate biophysical parameters (e.g. three height and canopy diameter which are strongly correlated with above ground biomass and leaf area index). In peatland areas the great variety of ecosystems and is ecological rule is still not fully understood and the influence of different degrees of succession and the influence of selective logged areas (e.g. species composition, their structures and canopy properties) on global change issues also remain a big challenge. Consequently, to optimize the biophysical properties characterization, a better understanding of how LiDAR variations measurements could be useful for ecological studies in such critically endangered forests is still necessary

In August 2007 we mapped by helicopter different PSF locations with Riegl LiDAR Technology LMS-Q560 in Central Kalimantan, Indonesia. In this study, we evaluated the relationship between tree height and peat dome

slope in distinct relief conditions described in the next section. To demonstrate such relationship we extracted the tree height information (e.g. average and dominant) as well the slope from sample plots of 100x100m (i.e. divided in subplots of 50x50m) along the flown direction in a minimum distance of 200m.

Study Area description

Our study areas consist on four LiDAR transects located inside the Central Kalimantan province, Indonesia (Figure 1). One Landsat image acquired on February 15, 2003 shows the remaining forest in the region. In fact, peat swamp forests over the region were impacted with extensive logging activities in the 90th and by the implantation of the Ex-Mega Rice Project (EMRP). The failed EMRP with its 4000km channel system leads to severe peat damages with reasonable amount of carbon released to the atmosphere.

Two transects are located in Block E of the EMRP, locally known as Mawas. The Mawas area has a North-South Channel and the BOS Tuannan Station near river Kapuas which can be identified in the LiDAR data. The main forestry areas are characterized by altered primary forest patches, in which trees had been logged selectively until the end of the 1990s. Forest ground truth data were collected in this area in frame of the INDREX-II campaign (Hajnsek et al. 2009). The last transect is located in the Sabangau National Park along the CIMTROP (Centre for International co-operation in sustainable Management of TROpical Peatland) transect. This PSF was selective logged by the concession company Setia Alam Jaya using an extraction railway system up to 1997.

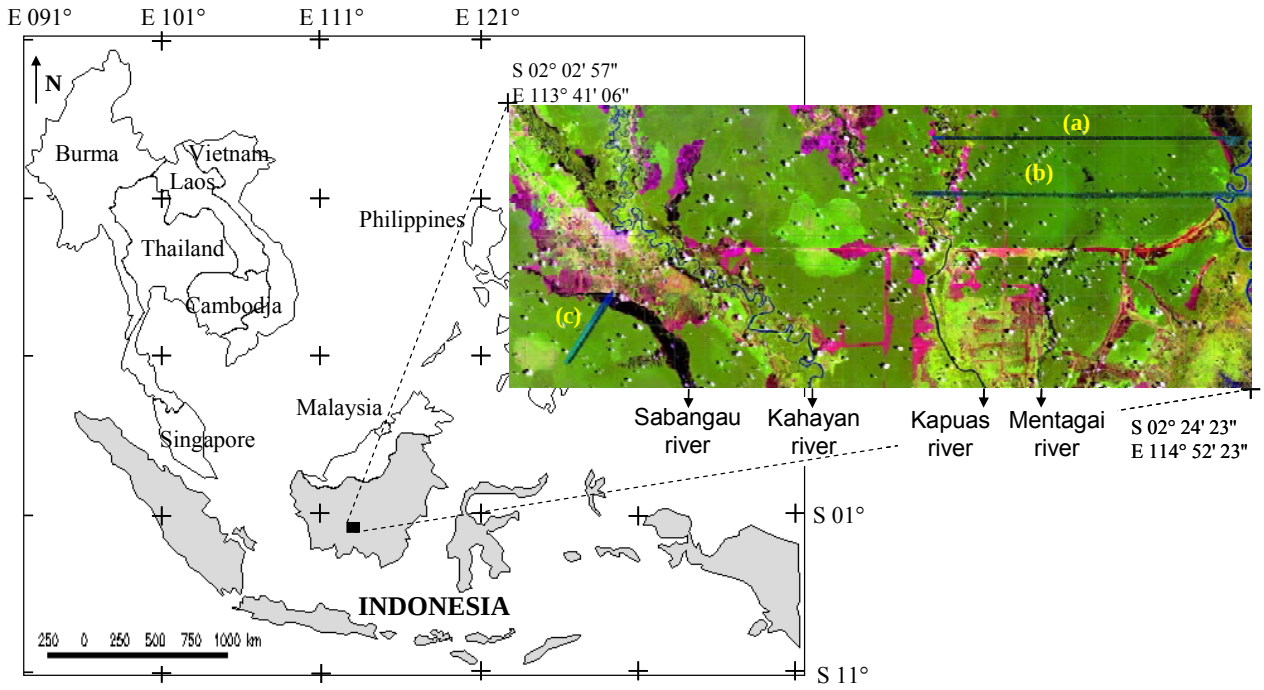


Figure 1. *Study area location with detail to a Landsat subset acquired on February 15, 2003. In this subset three letters indicate the location of three different LiDAR transects under investigation.*

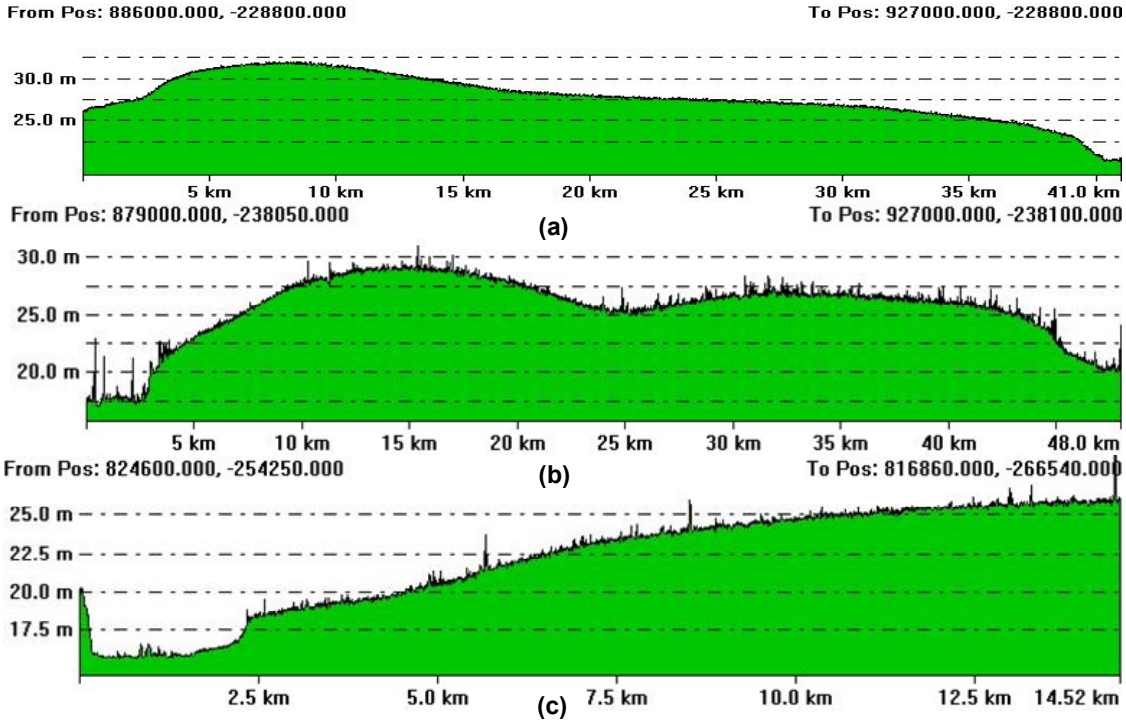


Figure 2. *LiDAR-DTM profiles for Mawas area in (a) km228.8 and (b) km238 and (c) Sabangau area. We selected for the cross-section a profile parallel to the visible logging transects. Transects were acquired with different lengths and have different scales.*

According to Hirano et al. (2007) the leaf area index of the peat swamp forest in the region in the dry monsoon is close to 6, typical for advanced and dense tropical rain forested areas. Areas of pasture, small agricultural fields, small villages and actual degraded forest resultant from selective logging can be also observed in the surroundings.

The three transects were selected based on the sample of three different relief conditions showed in Figure 2. In this figure we may noticed an asymmetrical peat dome on the left with a terrain peak close to 32m (Fig. 2a and Table 1). The second peat dome shows a double peat dome to the left with a peak close to 29m and with the upper part of Mentangai river (~25m) located in the middle. At the Kapuas river left the water level was 17.1m and on the right at Barito 20m on August 5, 2007 (Fig. 2b and Table 1). The third figure shows on the left side the Sabangau Catchment which up to 2.5m (18m) deep with the Sabangau river (15.5m) (Figure 2c and Table 1). The peat surface increases quickly with 0.17% up to 26m and further to the peat dome to 31m (results not shown in Fig. 2c).

Table 1*. Summary of the three LiDAR transects under study.*

Peat area transect	River level	Peat dome	Used trans. length	Max. slope (m / km)	Avg. avg. tree height	Avg. max. tree height
Mawas km228S (a)	17.8m	32m	41km	2.1m/km	12.8m	27.5m
Mawas km238S (b)	17.1m	29m	42km	1.3m/km	9.4m	26.3m
Sabangau km256S (c)	15.5m	26m	12km	1.7m/km	14.0m	29.4m

2. Material and Methods

2.1 LiDAR Data processing

The airborne LiDAR transects were acquired from August 5 to 7, 2007 in a helicopter campaign conducted by Kalteng Consultants and Milan Geoservice GmbH. We collected for the above described tracks approx. 4200ha of PSF with approx. 1.4 laser beams per square meter. The Riegl-airborne laser-scanner LMS-Q560 was mounted under the Bell 206 helicopter. Small footprint LiDAR data was collected for a flight altitude of approx. 500m with a scan angle of 60° with produced a swap-width of approx. 500m (Boehm et al. 2007, 2008). We used for this analysis the first and last pulse Laser echoes only, but the full-wave data are available for more detailed analyzes. The Laser scanner had a pulse rate of 66kHz resp. 100kHz with a beam divergence of 0.5mrad or a footprint of 0.25m. The ground backscattering in PSF through the canopy was responsible for 1 to 3% of the 0.5mrad Laser beams.

The DGPS reference station was located at the airport of Palangka Raya which has an elevation of 25.0m above sea level. The position and orientation of the LiDAR system on the helicopter was measured by an Inertial Navigation System (INS) and a GPS located on the tail boom with 256Hz. The Riegl LMS-Q560 airborne Laser scanner system itself allows height measurements with an accuracy of ±0.02m. After the correction of the attitude of the helicopter, the elevation accuracy of each Laser beam was ±0.15m with a root mean square error (RSM) of ±0.5m in both x- and y-direction. The processed laser beams were divided into ground surface and overground classes and converted in order to digital terrain model (DTM) and digital surface model (DSM), respectively, at a spatial resolution of 1m.

2.2 Data Analysis and Sampling Criteria

Observation data were compiled from sample plots of 100x100m collected in the flown acquisition. For validation purposes we divided each sample plot in four subplots of 50x50m for statistical analyzes. The spacing between each sample plot and/or subplot was 200m res. 500m in part of the Mawas area. We extract both DTM and DSM values for each measurement subplot and transect. In the slope determination we only account for the lowest values of the DTM in order to minimize the inclusion of the return signal coming from tree trunks and branches lying on top of the peat surface.

We calculated the slope by counting the difference between the DTM values of two samples and converted the altitude difference in a space of 200m and when applied to 500m into per mille (m per km). We proceed with the extraction of tree height for the DSM where we just account the signal coming from the dominant trees. Complementary we also calculated the average of all tree heights inside the sample and/or subplots. The number of plots varies according to each transect under study due to differences in length acquisition. The data from each transect was divided into training (70%) and validation (30%) datasets for statistical purposes described in the following section.

2.3 Statistical Modeling

The relationship between tree height and peat dome slope employed a linear regression analysis (i.e. yj=ax+b). The slope value for each sample plot was used as the predictor for tree height determination. The linear regression analysis was applied to the test dataset in order to perform the accuracy statistics. The accuracy statistics include the root mean square error (RMSE) (Eq. 1), Bias (Eq. 2) and their relative counterparts $RMSE_r$ (Eq. 3) and $Bias_r$ (Eq. 4) (Muukkone and Heiskanen, 2005).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - j_i\right)^2} \quad \text{Eq. 1}$$

$$Bias = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - j_i\right) \quad \text{Eq. 2}$$

$$RMSE_r = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - j_i)^2}}{y_m} \times 100 \quad \text{Eq. 3}$$

$$Bias_r = \left[\frac{1}{n}\sum_{i=1}^{n}(y_i - j_i)\right] y_m \times 100 \quad \text{Eq. 4}$$

where: j_i is the predicted value, y_i is the observed value, y_m is the mean of observed value and n is the number of plots in test dataset.

Complementary we also evaluated the analyses of the residuals (i.e. Observed value minus Predicted value) and we apply Cook's distance to identify outliers. Based on this distance we disregard up to 0.5% of the samples from the statistical analysis. In fact, some outliers were noticed in flat areas where no slope was observed. To examine if there were significant differences between the observed forest attribute (e.g. tree height) and those estimated from the linear regression, a non parametric test was applied. We expect the results might reject the hypothesis that the mean values of the observed and modeled forest properties differ significantly.

3. Results and Discussion

3.1 Mawas Peat swamp area at km 228

Initially we analyzed the asymmetrical peat dome located at km228.8 (South of the Equator) in the Mawas transect between Kapuas and Barito rivers (Figure 2a). We resampled the distance of our samples to 500m instead of 200m to facilitate graphic representation of the average tree height and peat dome slope (Figure 3). In this figure we found the highest tree height at km888.5 with 18.3m and the overall tree height average was 12.8m for 82 samples in the transect of 41km

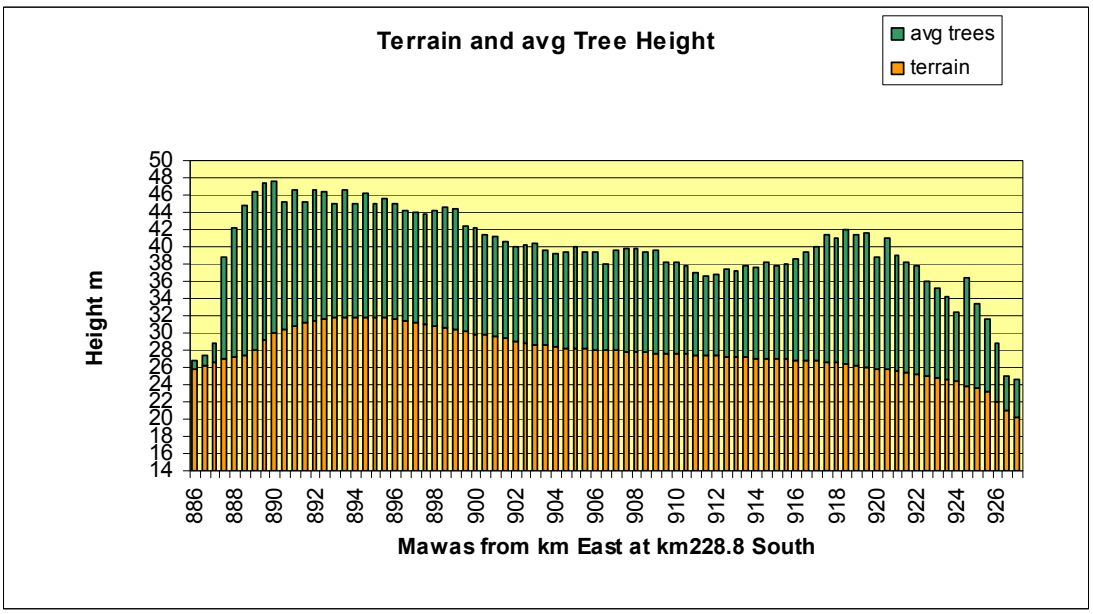


Figure 3. *Average tree height coupled with the terrain of the peat surface analyzed in 500m steeps over the Mawas transect of 41km, at km228.8 south. Sample size was 100x100m.*

Figure 4 shows the average tree height over the Mawas area. The highest average tree height at km888.5 was 18.3m (left side of the figure) and at km919 with 15.6m. This transect is relatively very flat and the tree height ranged mostly from 9 to 11m only. However, the average of the tree height over the 41km transect was close to 13m.

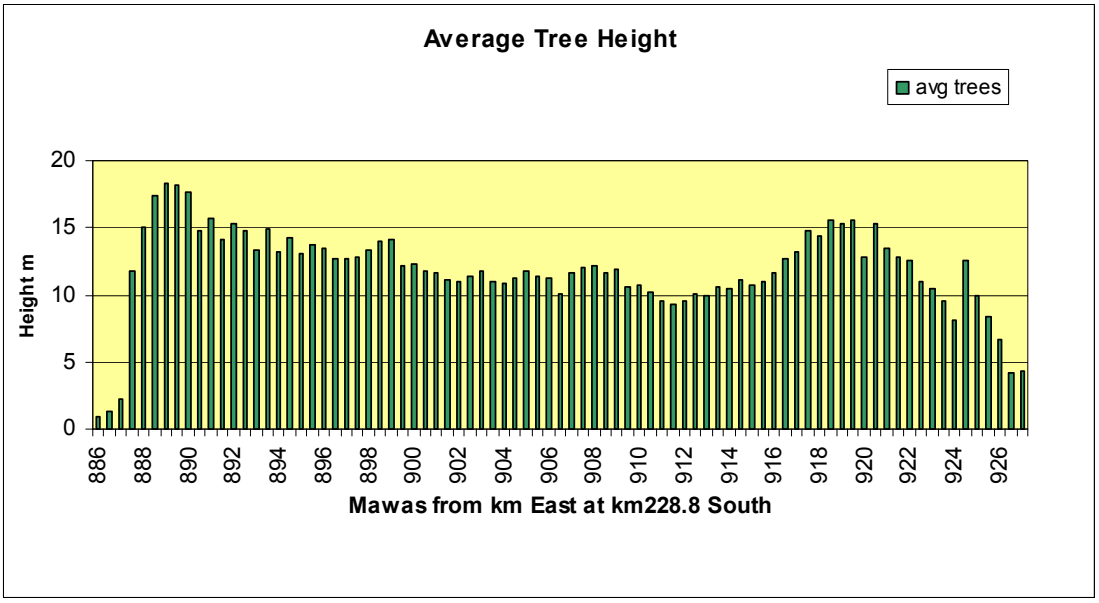


Figure 4. *Average tree height over the Mawas area of 41km, km228.8 South*

Figure 5 shows the relationship between tree height and peat dome slope over the Mawas area. Although height distributions of individual plots have not been fully investigated in our study, we noticed a stronger correlation with the maximum tree height (coefficient of determination of 0.78) with the slope than the average tree height (0.59). Maximum tree height is related to dominant height, a common parameter in forest inventories to asses the quality of forest stands over a certain region (Lefsky et al. 1999).

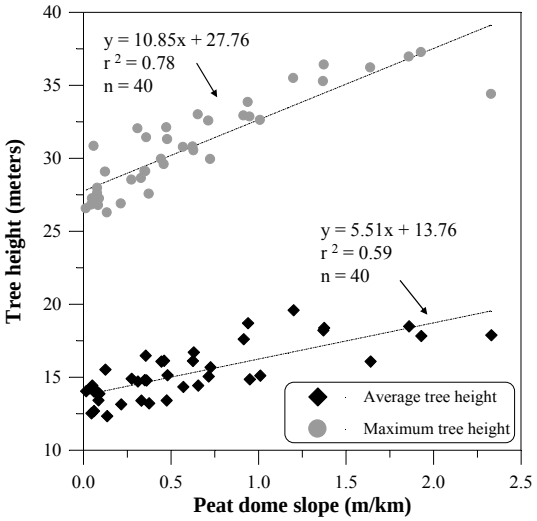


Figure 5. *Average and Maximum tree height as function of the peat slope for Mawas transect. The sample plots of 100x100m were analyzed every 200m from km887 to km895. The peat slope correlation between both tree heights are very good.*

3.2 Mawas Peat swamp forest area at km 238

The second peat dome cross-section was analyzed at km238 of the Mawas area. Figure 6 shows the average tree height coupled with the ground surface altitude. At km890.2 we identify the North-South channel of the Ex-Mega Rice Project. Forestry stands close to the channel had been logged in the past (indicated by the first arrow). Such interventions are better identified if we consider the average tree height information (Figure 7). The Mentangai river has his source at km904 with an altitude of 25m (indicated by arrow in the middle) and this river allows an easy access to this area with boots. At this Mawas cross-section the average tree height is not so high

probably due to the selective exploitation of the forest in the past. In this transect we noticed the highest of 13m (Figure 7). The average tree height over the 42km transect showed a value of 9.4m only. This indicates more logging activities compared to the Mawas transect illustrated in Figure 2a.

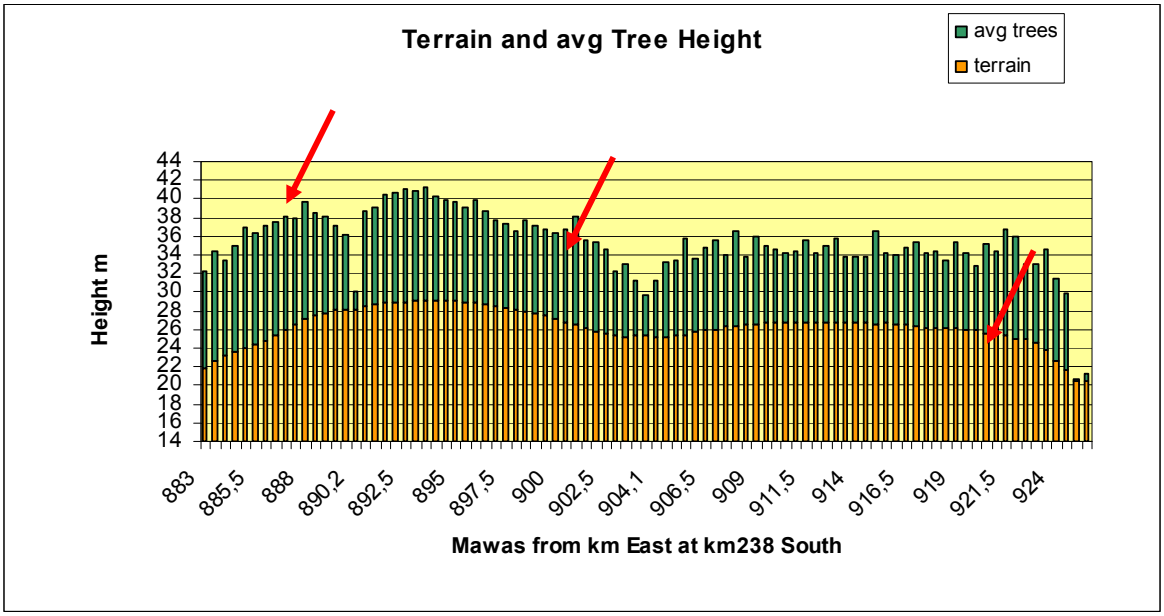


Figure 6. *Averaged tree height with peat surface for Mawas area at km238 using LiDAR-DSM and DTM-data.*

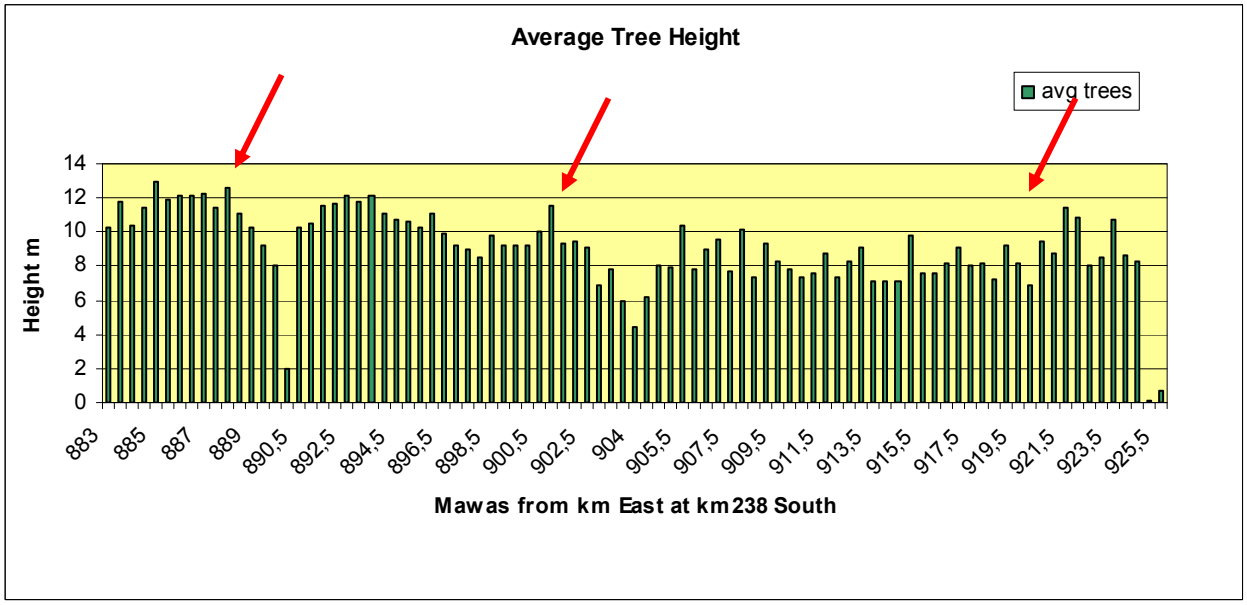


Figure 7. *Averaged tree height without the peat surface in Mawas area at km238.*

Due to the selective logging we did not found a strong correlation between tree height and peat dome slope. At km888 we found the highest averaged trees with 12.63m and a slope of more than 1.3m per km (Figure 8) that probably mask the relationship between both variables. At km890.2 the average tree height drops down, caused by the N-S channel and opened PSF. A similar result is shown on the right sight of the first dome to Mentangai river at km901 with 11.6m average tree height and slope down in order of 0.7m per km. At km911 on the second peat dome we found the lowest averaged tree height value with an average tree height of 7.5m. At km922 the average tree height increases again to 11.5m with the higher peat slope to river Barito.

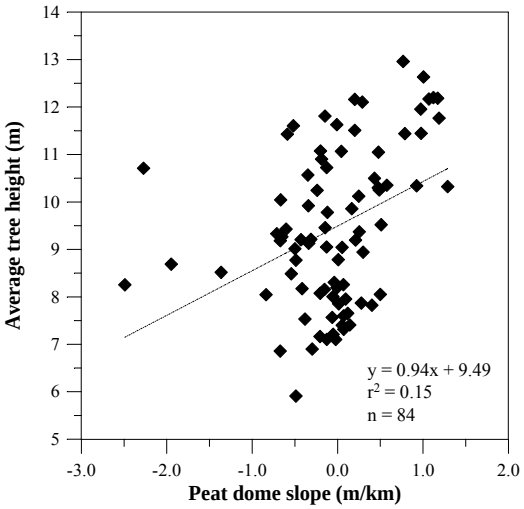


Figure 8. Average tree height as function of the peat dome slope for Mawas transect.

3.3 Sabangau National Park (CIMTROP peat swamp area)

The Sabangau CIMTROP transect in the National Park is another transect which was selectively logged in the past. In this transect, the Sabangau catchment in located on the left side of the profile with lowest ground elevation (15.5m) and in other side peat dome with 26m (Figure 9). The highest slope in this transect was up to 1.7m per km. For the selection of the sample plots we take into account only forest stands approximately 350 to 500m parallel to the former railway assuming less or moderate logging influence. Samples were reduced therefore to 50x50m and were analyzed every 200m

The average of the average tree height of the sample plots for this 12km transect was 14m, and was relatively higher that in Mawas area at km238 where we observed the value of 9.4m. At km258.4 South we found the highest averaged tree height with 17.1m whereas 12.63m in the Mawas area. Furthermore, the maximum tree height was also higher in this transect. We found the value of 34m in this transect whereas 29m for the Mawas area. Such results may be an indicative of logging severity in the past. As a result, the correlation between tree height and peat dome slope was better (Figure 11) than the Mawas area (Figure 8). In this transect the steepest peat surface was found at km259.2 with 0.30m-0.35m, that indicate a maximum slope of approximately 0.17%.

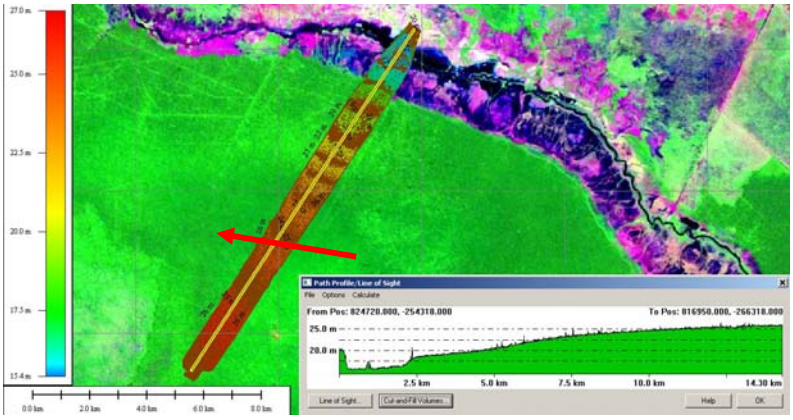


Figure 9. *LiDAR-DTM of CIMTROP transect in Sabangau National Park with 1m contour-lines and peat topography profile, Landsat image from 5.8.2007. The red arrow shows the highest slope of the transect with 0.17%.*

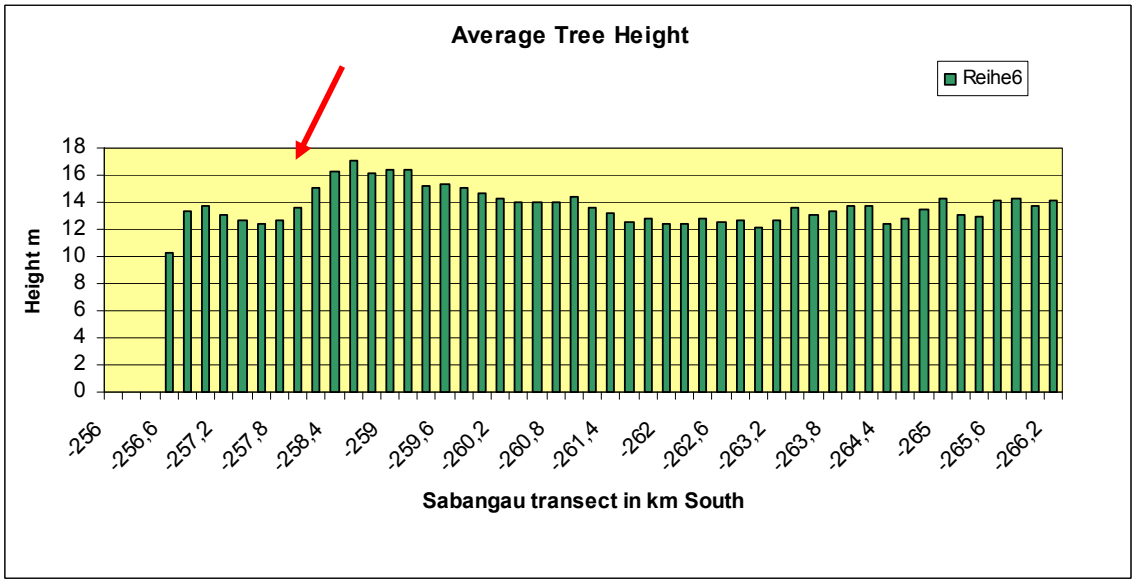


Figure 10. *Averaged tree height without the peat surface in Sabangau transect.*

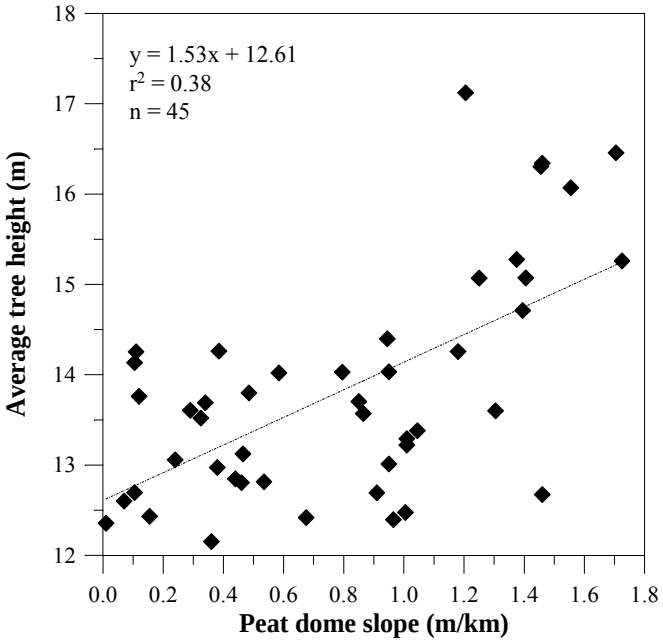


Figure 11. *Average tree height as a function of the Peat dome slope. The slope of the peat dome is up to 0.17%.*

3.4 Statistical approach for Mawas and Sabangau peat swamp areas

The maximum tree height of two less undisturbed peat swamp forest transects at Mawas and Sabangau transects were regressed against the predictor variable peat dome slope. Table 2 shows the main statistical results of the relationship between both variables using the validation dataset of both transects. The selected regression equations were obtained from the training dataset. The coefficient of determination (or explanatory power r^2) values of the maximum tree height were in order 0.51 and 0.65 for Mawas and Sabangau transects, respectively. These indicate that 51 and 65% of the variances in tree height were explained by the peat dome slope. The nature of this relationship may be noticed in Figure 12 for the entire dataset.

The tree height of each stand estimated from the peat dome slope (individual grid cell predictor) was compared to the ground-truth (Figure 13). Although the sample plots used to estimate the regression equation that related tree heights to the peat dome slope predictor variable were distributed independently throughout less disturbed peatland area, the mean difference between observed and predicted values for both transects were lower than 3m (Table 2 and Figure 14).

***Table 2**. Ordinary least square regression and RMSE calculated from the testing pixels.*

Mawas transect

Tree height	Test statistics						
	r^2	RMSE	RMSE r	Bias	Bias r	t	p-value
Maximum	0.51	1.93	7.00	0.08	0.31	3.27	<0.001

Sabangau transect

Tree height	Test statistics						
	r^2	RMSE	RMSE r	Bias	Bias r	t	p-value
Maximum	0.65	2.78	11.37	-0.10	-0.43	-2.46	<0.001

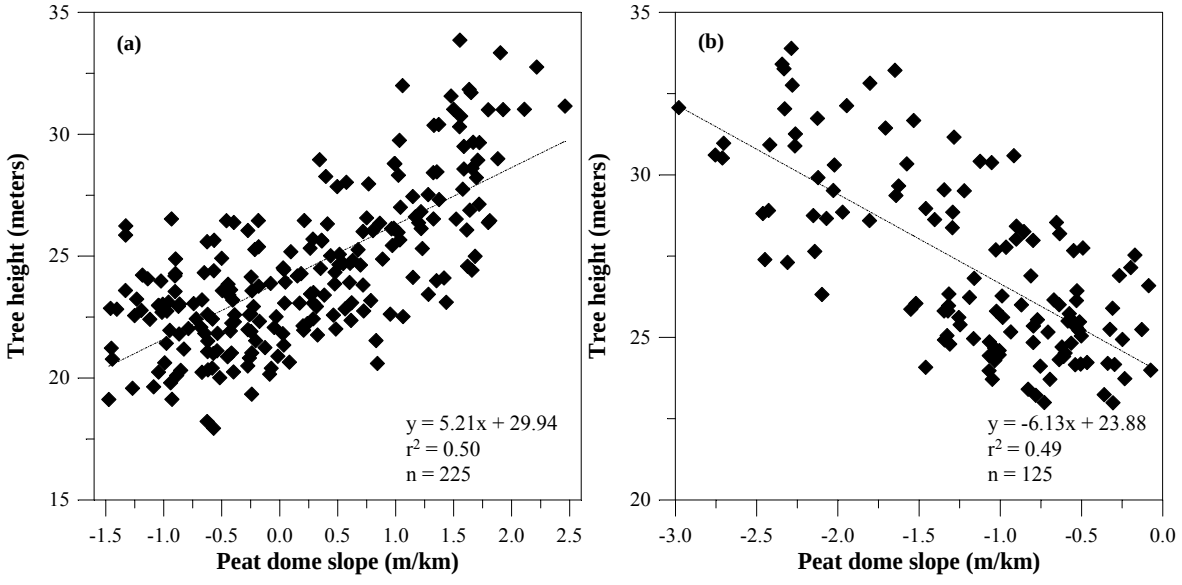


***Figure 12.** Relationship between maximum tree height against peat dome slope. The LiDAR attributes include complete dataset for (a) Mawas (Figure 2b) and (b) Sabangau (Figure 2c) transects. Negative and positive slopes values indicate in order descending and ascending relief.*

The estimated biases confirm such results showing low values for both sites and the non-parametric test showed the linear regression estimates and the observed data were significantly different (Table 2). Although a strong correlation between tree height and peat dome slope exist it seems highly possible that in both transects the trees have different hydrological demands according to the dome position (e.g. beginning, middle or top) implying different growth forms due to variations in permeability, interflow, water storage capability and nutrient avaibility. Information on tree height variation due to peat dome slope changes may be useful for further forest inventory assessment since biophysical properties (e.g. above ground biomass) may vary significantly according to the peat dome position.

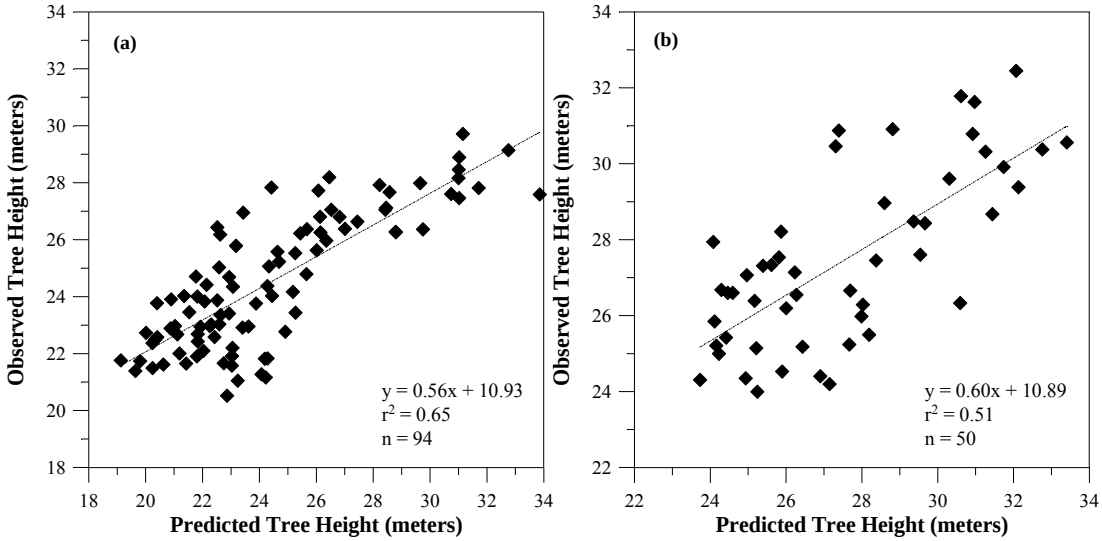


Figure 13. *Relationship between observed and predicted tree height. The LiDAR and statistical attributes include validation dataset for (a) Mawas (Figure 2b) and (b) Sabangau (Figure 2c) transects.*

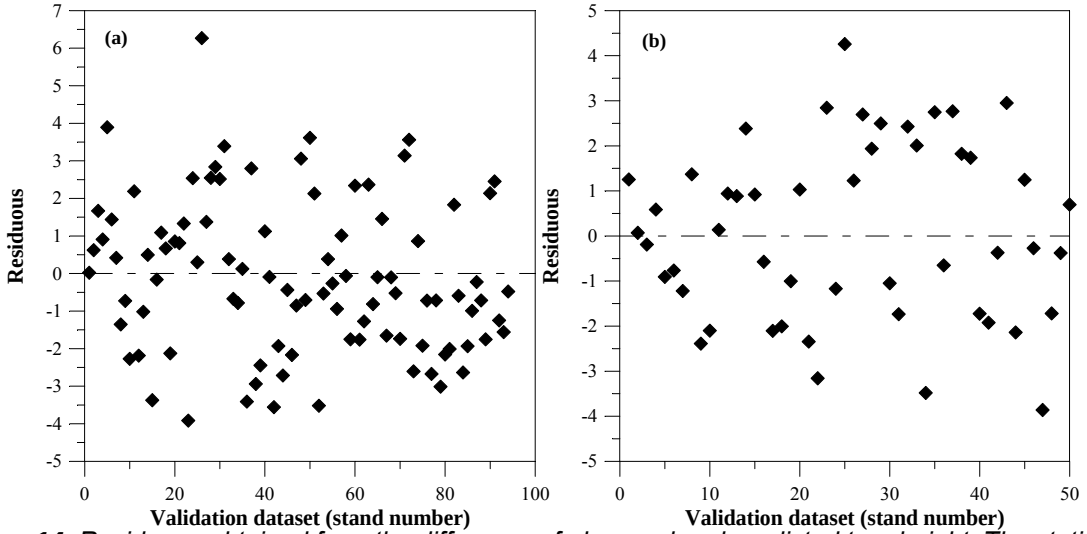


Figure 14. *Residuous obtained from the difference of observed and predicted tree height. The statistical attributes include validation dataset for (a) Mawas (Figure 2b) and (b) Sabangau (Figure 2c) transects. Positive and Negative values indicate in order over- and underestimation of the tree height.*

Final Remarks and implications for forest management

Regarding to the relationship between tree height and peat dome slope we observe a correlation in undisturbed forest areas, Fig.5. This relationship between both variables showed a coefficient of determination (or r^2) close or higher than 0.6. Such parameters reveals a strong correlation between both variables and may be related with the permeability, interflow, water storage capability and nutrient avaibility in the peat domes that needs to be investigated. In these transects, the intervention to the PSF was limitated and did not affect significantly the biophysical properties of the forest. In the linear regression analyses, the maximum tree height parameter showed to be a better predicted by peat dome slope than the average tree height. However, a lower correlation was found in transects where selective logging activities were common and take place in the past.

Information on tree height variation due to peat dome slope changes may be useful to assess the dependence of biophysical properties (e.g. above ground biomass) with peat dome slope in peatlands environments. Since peatlands act as a carbon sink, human interventions due to drainage practices for agriculture and selective logging of the peat swamp forest may have a stronger impact in the carbon release than relative flat areas.

Nonetheless, further research is still necessary in order to test the dependence of other biophysical parameters to peat dome slope and feature selection techniques for LiDAR data in different vegetation types in Indonesia as well as field work campaigns. In spite of the technique used for dependence assessment, interesting results will be probably achieved with the additional use of new LiDAR measurements over the area.

Acknowledgement

The funding for the LiDAR flights and data processing was done by Kalteng Consultants and partly by the Central Kalimantan Peatland Project (CKPP). We would to thank the operator Mr. Jan Giehler and technician Mr. Detlef Klante from Milan Geoservice GmbH and Mr. Mustafa Syafrudin for the good cooperation during the LiDAR flight campaign. Additionally we would to thank Dr. Marcel Silvius from Wetlands International and Prof. Dr. Hans Joosten from University of Greifswald for their delightful discussions on our research. The second author is supported by CNPq/Brazil.

References

BALLHORN, U, SIEGERT, F., MASON, M.C., LIMIN, S., 2009, Derivation of burn scar depths and estimation of carbon emissions with LIDAR in Indonesian peatlands. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, pp. 25-30.

BOEHM, H.-D.V., RIELEY, J.O., LIMIN, S., FRANK, J., SYAFRUDIN M., 2007, Successful Helicopter Flight Trials with Airborne Laser Scanning Technology to measure PSF height and Peat domes in Central Kalimantan. In *International Symposium, Workshop and Seminar on Tropical Peatland*, *"Carbon - Climate - Human Interactions - Carbon Pools, Fire, Mitigation, Restoration and Wise Use"*, 27 - 31 August, Yogyakarta, Indonesia.

BOEHM, H.-D.V., FRANK, J., 2008, Peat Dome Measurements in Tropical Peatlands of Central Kalimantan with a high-resolution Airborne Laser Scanner to achieve Digital Elevation Models. In *Proceedings of 13th International Peat Congress*, 8-13 June 2008, Tullamore, Ireland, Section 5: Tropical Peatlands.

HAJNSEK, I., KUGLER, F., SEUNG-KUK LEE, PAPATHANASSIOUS, K.P., 2009, Tropical Forest Parameter Estimation by Means of Pol-InSAR: The INDREX-II Campaign, *IEEE Transactions on Geoscience and Remote Sensing*, **47**, pp.481-493.

HIRANO, T., SEGAH, H., HARADA, T., LIMIN, S., JUNE, T., HIRATA, R., M. OSAKI, 2007, Carbon dioxide balance of a tropical peat swamp forest in Kalimantan, Indonesia, *Global Change Biology*, **13**, pp.412-425.

HYDE, P., NELSON, R., KIMES, D., LEVINE, E., 2007, Exploring LiDAR–RADAR Synergy - Predicting Aboveground Biomass in a Southwestern ponderosa pine forest using LiDAR, SAR, and InSAR, *Remote Sensing of Environment*, **106,** pp.28–38.

JAENICKE, J. RIELEY, J.O., MOTT, C., KIMMAN, P., SIEGERT, F., 2008, Determination of the amount of carbon stored in Indonesia peatlands, *Geoderma*, **147**, pp.151-158.

LEFSKY, M.A., COHEN, W.B., ACKER, S.A., PARKER, G.G., SPIES, T.A. AND HARDING, D., 1999, Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. *Remote Sensing of Environment*, **70**, pp.339–361.

MUUKKONEN, P., HEISKANEN, J., 2005, Estimating biomass for boreal forests using ASTER satellite data combined with standwise forest inventory data, *Remote Sensing of Environment,* **99**, 434-447.

PAGE, S., SIEGERT, F., RIELEY, J.O., BOEHM, H.-D.V., JAYA, A., LIMIN, S., 2002, The amount of carbon released from peat and forest fires in Indonesia during 1997. *Nature,* **420**, 61-65.

SORENSEN, K. W., 1993, Indonesian peat swamp forests and their role as a carbon sink, Chemosphere, *Chemosphere,* **27**, 1065-1082.

Derivation of forest structure using satellite, multifrequent radar and lidar data in Thuringian Forest, Germany

Claudia Hilbert[(1)], Christiane Schmullius[(1)], Manfred Zink[(2)]

(1)Department of Earth Observation, Friedrich-Schiller-University Jena, Grietgasse 6, 07743 Jena, Germany, claudia.hilbert@uni-jena.de, c.schmullius@uni-jena.de

(2)Microwaves and Radar Institute, German Aerospace Center (DLR), Oberpfaffenhofen, Germany, manfred.zink@dlr.de

Abstract

This paper presents first results of a study combining satellite, multifrequent radar and lidar data for characterising forest height and biomass. Biomass is an important structural parameter because it is the key to the carbon content. However, biomass is one of the large unknown components in the global carbon cycle. The synergy of pointwise lidar and multifrequent SAR data for forest mapping is promising. The study introduced here aims to combine TerraSAR-X, ALOS PALSAR and ICESat/GLAS data.

Some preliminary results for a test site in the Thuringian Forest are described. The Thuringian Forest is a low mountain range in eastern Germany and it mainly consists of temperate forests with Norway spruce, European beech and Scots pine. In the first part some preliminary results regarding the PALSAR data are shown. The second part focusses on the GLAS data: Two methods for filtering invalid GLAS shots are investigated. Moreover, different ICESat/GLAS waveforms parameters were calculated and compared to an airborne lidar based Digital Height Model (DHM) and a forest inventory data base.

1. Introduction

The global forests represent an important CO2 sink and forest degradation is one of the most relevant CO2 sources. Forest volume or biomass plays a critical role since it is directly related to the carbon content. About 50% of the dry plant matter is made of carbon. Therefore accurate and regularly repeated estimations of forest biomass are needed. However, existing data sets as ground based inventories cannot provide detailed and exact information about the global biomass and its temporal dynamics.

Remote sensing techniques have been shown to be a suitable complement to map the global forests and to derive biophysical parameters like above-ground biomass. Multispectral remote sensing data can be a good basis for continental biomass maps (Dong *et al.* 2003, Lu 2006). But characterising forest structure is difficult since the sunlight mainly gets reflected from the upper canopy. Optical data are particularly suitable to derive information about the vegetation condition and health.

Synthetic Aperture Radar (SAR) systems have a big potential to characterise forest structure because the radar backscatter is influenced by the geometric properties of a forest stand. However, the relationship between backscatter and biomass saturates at a certain biomass level. The exact saturation level depends on several factors, like wavelength, polarisation and weather conditions during the acquisition. Longer wavelengths penetrate deeper into the vegetation and allow in general better biomass estimations than short wavelengths. Moreover, cross-polarised data is more suitable than co-polarised data. The limit of L-Band, the longest wavelength available from space, is around 40-100t/ha (Luckmann *et al.* 1997, Imhoff *et al.* 1998). Other SAR

parameters and techniques, like interferometry and polarimetry, and multitemporal approaches can improve the biomass estimations. For example, Santoro *et al.* (2002) found a good relationship between the interferometric coherence and stem volume up to 350m^3/ha with an error of 10m^3/ha. The still existing saturation, radar effects like shadow and foreshortening and the dependence with the given environmental conditions during the acquisition complicate the usage of SAR data for a simple and robust biomass derivation. Moreover, any given wavelength interacts only with specific features (leaves, branches, trunks, etc.).

Light detection and ranging (Lidar) or laser remote sensing data are of special interest to derive directly the vertical dimension of a forest. The laser pulse gets reflected from the tree top, but it also penetrates the canopy. Depending on the system configuration, the sensor records the several distinct return echoes (discrete-return systems) or a continuous signal (waveform systems) from the upper part of the tree crown to the ground surface. Airborne discrete-return and also waveform lidar achieve high accuracies for tree height and biomass estimations also at high biomass levels (e.g. Næsset 1997, Means *et al.* 1999, Drake *et al.* 2002, Nelson *et al.* 2009). Discrete-return systems are already used operationally and support the common forest inventories especially in Scandinavian countries (Næsset *et al.* 2004). Since the launch of ICESat/GLAS in January 2003 also spaceborne lidar data are available. The waveform system was originally built for ice sheet measurements, but GLAS data show also high potential for forestry applications (Levsky *et al.* 2005, Levsky *et al.* 2007, Rosette *et al.* 2008, Chen 2010). However, the GLAS footprint size is on average 65 m in diameter and the distance between two footprints is about 170 m. Thus, GLAS offers a quite rough spatial resolution and the large footprint makes the waveform very sensitive to topography.

The synergy of different remote sensing methods offers a promising approach to overcome the limitations of the single data concepts. The study presented here combines spaceborne multifrequent SAR and lidar data to characterise forest structure. The overall goal of the study is to develop a validated and robust method to derive forest above-ground biomass. The basic idea is to use TerraSAR-X X-Band data to describe the horizontal structure and ALOS PALSAR L-Band data to identify the vertical structure. Complementary, ICESat/GLAS waveforms provide direct estimations of tree height for the footprint sites. A comprehensive understanding of the different signals will be acquired first. This comprises the comparison of the GLAS height measurements with the reference heights and the investigation of the relationship between SAR parameters (backscatter, interferometric coherence, texture) and biomass. In a second step, potential of fusing SAR and lidar data regarding its different spatial resolutions and data concepts will be analyzed. The final step will be the development of a model for retrieving biomass using a combination of GLAS and radar parameters. This paper will introduce some first results related to the GLAS waveform investigations and the PALSAR data analyses for a test site in the Thuringian Forest.

Study Area

The Thuringian Forest is located in the federal state Thuringia, central Germany, between 09°44' - 50°58'N and 11°38 -50°18'E (Fig. 1). It is a low mountain range extending about 50x110 km from northwest to southeast with ground elevations up to 900 m ASL and a hilly landscape. The main tree species are Norway spruce and European beech. In some parts Scots pine can be found. The forests are managed and silviculture is an important industry for the region. A strong storm called "Kyrill" occurred in the region in January 2007 and caused heavy damages. The size of the total area affected by the storm is about 11,000 ha and about 2.5 million m^3 of timber tumbled down.

The focus of this paper is a small site within the Thuringian Forest of about 3x15 km. The site, as from now called Schmiedefeld site because of the village in the north, is part of the biosphere reserve Vessertal. The area is characterized by a valley from north to south and steep slopes mostly facing to the east or west. In the northern part mainly spruce forests can be found whereas in the southern part also beech forests are common.

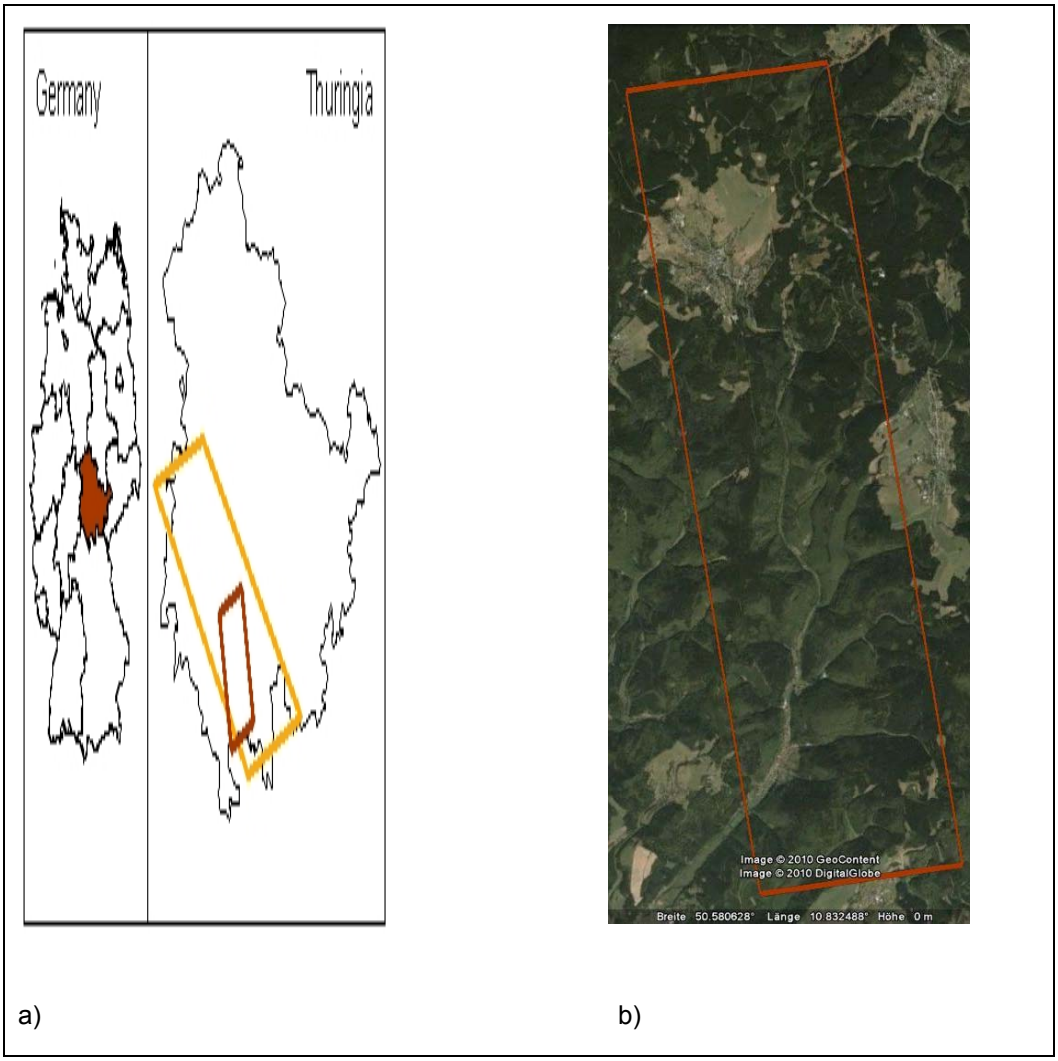


Figure 1. a) Overview of the geographic location of the study area. Big rectangle: Thuringian Forest, small rectangle: Schmiedefeld test site. b) Section of Schmiedefeld site. Image taken from GoogleEarth

Data

SAR data

ALOS PALSAR data for the years 2007-2009 in Fine Beam Single (FBS) and Fine Beam Dual (FBD) mode were obtained from the Japan Aerospace Exploration Agency (JAXA). 9 FBS scenes (HH) for the winter season between January and April and 12 FBD scenes (HH/HV) for the summer season between May and October are available for the Schmiedefeld site. All images were acquired with an incidence angle of 34.3°.

TerraSAR-X data were ordered from the German Aerospace Center (DLR). High Resolution Spotlight Mode from July-November 2009 is used. 11 acquisitions were done in dual polarisation HH/VV and 9 in single polarisation

HH. The incidence angles range from 22.8° - 47.5°. Furthermore, four Strip Map (SM) images in dual-polarization are available.

ICESat/GLAS

ICESat/GLAS data for all acquisition dates from February 2003 until March 2007, release 428, were ordered from the National Snow and Ice Data Center (NSIDC). The three different products GLA01 Global Altimetry, GLA05 Global Waveform-based range Corrections and GLA14 Global Land Surface Altimetry Data are used for this study.

An IDL based tool available from the NSIDC website was modified to process the binary GLAS into a simple ASCII format useable for later analyses. A total of 211 waveforms within forested areas were available for the Schmiedefeld site which had to be filtered for invalid data (see section 4.2). Since the exact footprint size changes between the different laser periods and even within one, the footprint size was adjusted for each record. The footprints were modelled as circles with the same area as the ellipse with the major axis and eccentricity of the transmitted pulse as given in the GLA14 dataset. The footprint sizes vary from 46.94 m – 63.56 m for the shots selected from the procedure described in section 4.2.

Ancillary data

The Thuringian Forest Agency provided a GIS data base with parameters of a common forest inventory. The data base includes information about main species, tree height, dbh, age, stocking volume, basal area, stocking degree, stocking layers and management type. The data are based on in-situ measurements from 1989-2010, i.e. the overall inventory is updated every ten years. The forest stands inside the Schmiedefeld site were measured between 2006 and 2010. The stem volume was calculated using allometric equations with a required accuracy of 20m^3/ha. To select some good stands suitable for later analyses, the data are validated using digital orthophotos with a resolution of 15 cm. Moreover, the following criteria were applied for selection: (1) date of the field measurement (2) damage of “Kyrill”-storm, (3) homogeneity, (4) canopy closure, (5) visual check with digital orthophotos, (6) area > 2 ha after edge erosion with a buffer of 25 m was applied.

Airborne Laser Scannig (ALS) data as last pulse ground and non-ground points are available from the Thuringian Land Surveying Office. The data were acquired during several flights between 2000-2005 with point densities up to 1 point/m^2. This induced to calculate a Digital Elevation Model (DEM) using the ground points and a Digital Surface Model (DSM) using the non-ground points with 5 m resolution each. The difference between both models resulted into a normalized DSM or here called Digital Height Model (DHM). The gridding of the lidar points was done using Surfer software package and the kriging interpolation method. Unfortunately, only last pulse data were available, but the heights of deciduous trees did not reveal any problems.

Finally, some meteorological data from the German Weather Service (DWD) are available. Parameters like temperature, precipitation, snow depth and wind speed as hourly and daily observations in the period 2006-2010 will help to analyse the SAR data.

2. Methods

Processing the SAR data

Both, PALSAR and TerraSAR-X images were acquired in SLC format. The processing of the SAR data included calibration, multi-looking, geocoding and topographic normalization and was done with the Gamma software

package. The ALS based DEM with 5 m resolution was used for the geocoding procedure. The images were processed to 25 m and 12.5 m for PALSAR FBD and FBS, respectively. 5 m and 10 m pixel size were chosen for TerraSAR-X HS data. Figure 2*a* shows an overview of the location of the PALSAR FBD frames. Frames 6 and 7 refer to t639/f1000 and t640/f1000 and provide six FBD images each for the Schmiedefeld site. Frame 6 and the image example for the Schmiedefeld site in Figure 2*b* will be further discussed in Section 5.1.

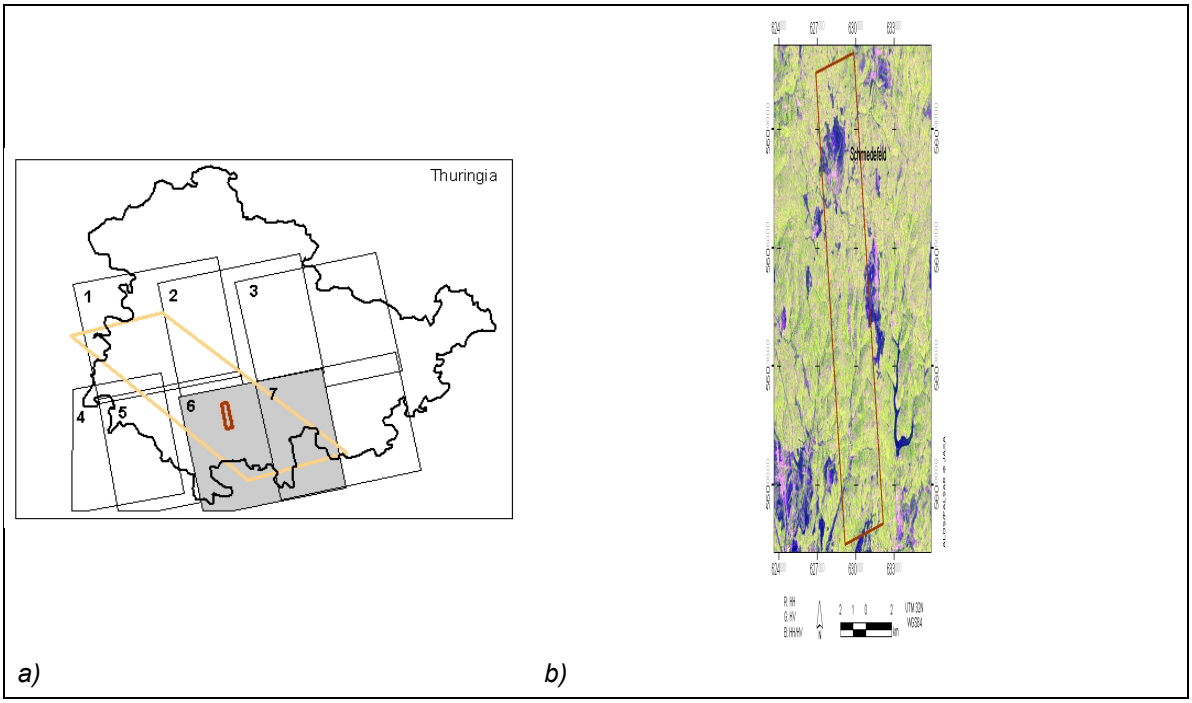


Figure 2: a) Location of PALSAR FBD frames and b) Section of frame no. 6 (t639/f1000), 7 September 2009, Schmiedefeld site – RGB composite of HH (Red), HV (Green) and HH/HV (Blue)

Filtering invalid GLAS data

Two methods were tested to remove outliers, i.e. to remove waveforms affected by the atmosphere or by waveform saturation. The first approach is based on the GLAS flags. Three criteria, namely waveform saturation, atmospheric influence and the existence of large off-nadir angle were checked by using the GL14 parameters Saturation Elevation Correction, the Full Resolution 1064 Quality Flag, the Elevation Usage Flag, the Range Correction Flag and the Attitude Flag 1.

The second method uses the ALS based DEM and an additional Signal Extent threshold. This procedure is based on the method proposed by Gonzàlez *et al.* (2010). The idea is to check if the surface elevation published in the GLA14 product and the surface elevation of the reference DEM are the same. If the difference between the two surface elevations exceeds a threshold of 75 m, the corresponding GLAS waveforms were removed as outliers. The Signal Extent threshold was added to get an absolute safety to remove all invalid shots. This was important because several waveforms with large Signal Extents which would produce unrealistic high height estimations were detected.

Calculating GLAS waveform metrics

The GLA14 data product was used to calculate several waveform metrics (Fig. 2). First, different height measures were extracted: Signal Extent as distance between Signal Begin and End, and a ground surface based height as the distance between the Ground Peak and Signal Begin. The Ground Peak usually refers to the Last Gaussian, but sometimes the maximum of the last two or five Gaussians represents the ground surface more accurately. The exact definition of the Ground Peak, however, revealed as challenging because of the complex waveform structures caused by the surface topography. Second, two parameters related to surface topography and canopy roughness were calculated, the Leading and Trailing Edge. They are defined as the distance between Signal Begin and End, respectively, and the position of the half maximum amplitude.

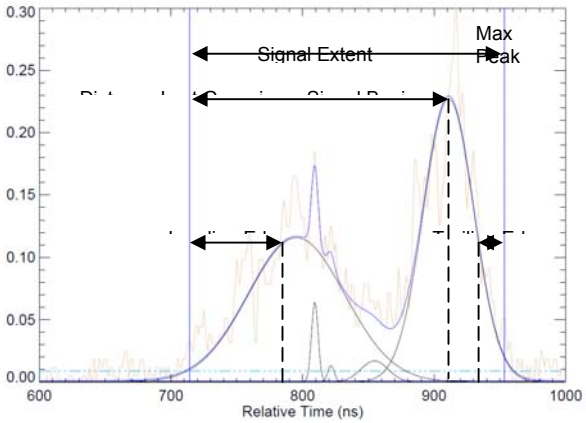


Figure 3. GLAS heights and relating waveform position. Red line: Return Pulse (V). Blue line: Alternate Modeled Waveform (V). Black line: Alternate Gaussian fits

3.

4. Results and discussion

PALSAR

Some first analyses for the six FBD images of track 639/frame 1000 were carried out. First, the data were investigated regarding the general properties of the backscatter and the temporal variation. Second, a first comparison to forest structure parameters were done.

Figure 4 demonstrates that the mean backscatter of beech is lower than of spruce (Mean for all six acquisitions: beech HH -9.01 dB, HV -14.03 dB vs. spruce HH -7.77 dB, HV -13.55 dB). The discrimination between the two species is, however, difficult because of the diversity of the backscatter (Mean Stdv for all six acquisitions: beech HH 1.04 dB, HV 1.22 vs. spruce HH 0.64, HV 0.64).

The temporal differences are quite small (<1 dB) but some trends can be detected. The mean backscatter for the dry acquisitions (see Tab. 1; 2 September 2007, 7 September 2009 and 23 October 2009) is in general lower than for the wet acquisitions (20 July 2008 and 23 July 2009). However, the mean backscatter of the image acquired on 18 October 2007 is similar to the dry acquisitions although a small amount of precipitation occurred at that day and the day before. This acquisition even shows the lowest mean backscatter for both beech and spruce. Another interesting effect can be identified for the beech stands: The image taken on 23 October 2009

with dry conditions is characterized by the highest maximum backscatter for HH and also HV (-3.9 dB and -9.4 dB, respectively). The maximum backscatter of the most humid acquisition on 23 July 2009 is around 2.5 dB lower. The investigation of the minimum backscatter does not show a similar effect and it is more consistent over time. Thus, the high dynamic range for the 23 October 2009 image of around 7 dB for HH and HV must has been caused by an object with very high backscatter. The other two dry acquisitions have a dynamic range of only 3-4 dB for HH and 4-5 dB for HV, the wet acquisitions 4-5 dB for HH and 5-7 dB for HV. The backscatter signatures for the 18 October 2007 image are more similar to the wet acquisitions as expected from the details listed in Table 1, but against the results for the spruce stands. The spruce backscatter properties demonstrate a lower dynamic range with 3 dB for HH and HV. It can be concluded that the environmental conditions obviously affect the backscatter. Some specific events probably occurred on 18 October 2007 and 23 October 2009, e.g. a short rain event directly before the acquisition or the beginning autumn season. The analysis of the other PALSAR images available for the area of he Thuringian Forest but not described in this paper will certainly enhance the understanding of the backscatter signatures which is the basis for the retrieval of biophysical parameters.

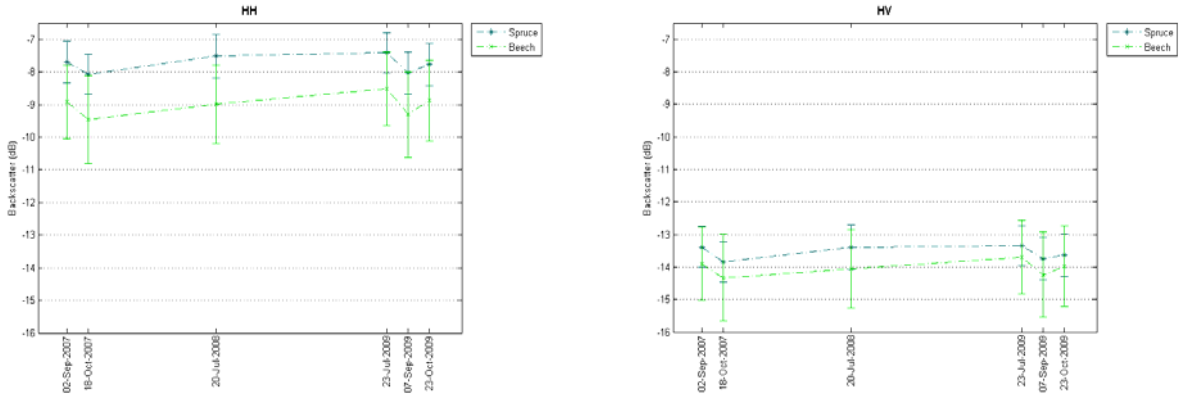


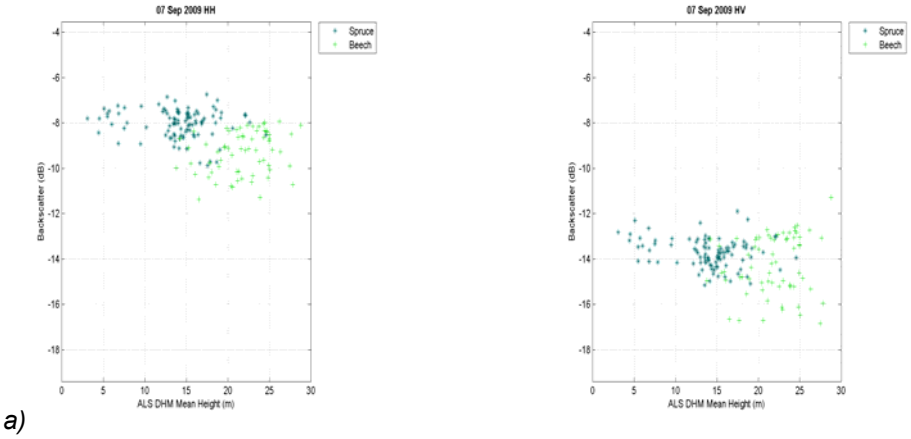
Figure 4. Mean backscatter and one standard deviation (vertical bar) for all FBD acquisitions of track 639/frame 1000

Table 1. Environmental conditions for all FBD acquisitions of track 639/frame 1000

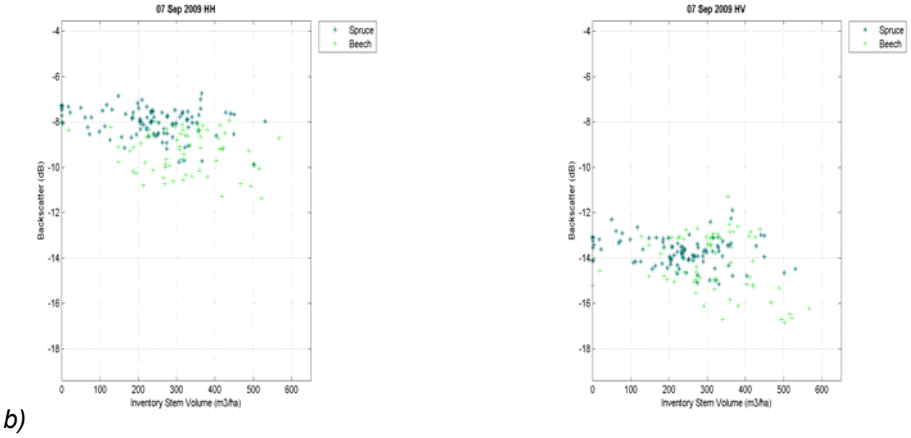
Acquisition date	Flight direction	Environmental conditions
02 Sep 2007	A	~13°C, dry, ~4 mm precipitation during the three days before
18 Oct 2007	A	~6°C, ~6 mm precipitation, ~5 mm precipitation during the day before
20 Jul 2008	A	~13°C, ~2 mm precipitation, ~18 mm precipitation during the five days before
23 Jul 2009	A	~19°C, ~15 mm precipitation, ~32 mm precipitation during the seven days before
07 Sep 2009	A	~13°C, dry, second dry day
23 Oct 2009	A	~7°C, dry, ~4 mm precipitation during the day before

Figure 5 presents a comparison between backscatter and ALS DHM Height and Inventory Stem Volume for the image acquired on 7 September 2009 as an representative example. The ALS DHM height was used as indirect

measure of biomass and volume, respectively. In future, the height will be also converted into volume using allometric equations. It can be seen that there is – against the assumption – only a low and even negative correlation. There are also significant differences between the two species as already indicated in Figure 4. The spruce backscatter spreads only by around 3 dB, whereas the range of the beech backscatter is around 4-6 dB. The stem volume range is also a bit higher for beech than for spruce (567 m^3/ha vs. 531 m^3/ha). In contrast, the beech heights range only by 15 m, but the spruce heights by 21 m. The structure of the beech stands is obviously more complex. Understorey and the complex phenology of a beech tree crown might be some reasons for the noticed backscatter diversity. These relations reveal that a species-specific analysis is critical to understand the backscatter behaviour and to develop a biomass model. The low correlation results are still not completely understood. Understorey vegetation like young trees or small shrubs might be one important factor influencing the backscatter but not considered in the inventory volume estimations or ALS DHM heights. Understorey grows particularly in stands where the sunlight can penetrate to the ground, i.e. rather in young or middle-aged stands than mature, dense stands. A three-day field survey in the Schmiedefeld site was conducted end of July 2010 and the analysis of the data is supposed to enhance the understanding of the relationship between backscatter and forest structure.



a)



b)

Figure 5. Comparison between backscatter and (a) ALS DHM Mean Height and (bB) Inventory Stem Volume for FBD image 7 September 2009, incidence angle 34.3°

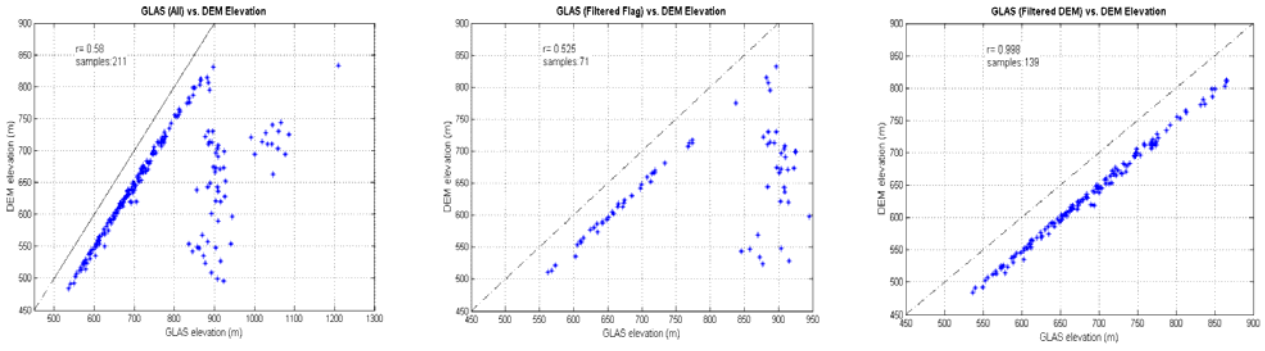
Filtering GLAS data

Table 2 summarises the results for the two filtering methods. The flag based approach detected 71 valid shots which are from three laser periods. Ten percent of the shots were removed as saturated and about three quarters as distorted by atmospheric influences. The data set did not contain any shots with a large off-nadir angle. The DEM based approach produced a complete different distribution: 139 shots from five laser periods remained. For example, the shots of laser period L3B were detected all as invalid when applying the flag based method. The DEM based approach identified only 14 outliers.

A comparison between the GLAS and reference surface elevation was done before and after filtering to check the working performance of the filtering methods (Fig. 6). It can be noticed that the data base without filtering contains several shots with significant higher surface elevation than for the reference DEM. This indicates that these shots are problematic since they would result into wrong waveform height metrics. The flag based filtering approach, however, does not detect all of them, but even some shots which have a valid surface elevation. The DEM based filtering removed all shots with an invalid surface elevation and the high correlation coefficient indicates that this filtering method works better. The GLAS elevations are on average 52.33 m (Stdv=5.19 m) higher because the GLA14 elevation referring to the waveform centroid was used here.

Table 2. Number of GLAS shots after filtering invalid data (flag based and DEM based method)

Laser Period	Date	All	Method Flags	Method DEM
L2B	Spring 2004	12	0	8
L2C	Summer 2004	2	0	0
L3A	Autumn 2004	47	34	0
L3B	Spring 2005	50	0	36
L3D	Autumn 2005	27	6	24
L3F	Summer 2006	15	0	13
L3H	Spring 2007	58	28	58
All		211	71	139



a) *b)* *c)*

Figure 6. GLAS versus ALS DEM surface elevations a) before filtering, b) filtering using flag based approach, and c) filtering using DEM based approach

ICESat/GLAS

The GLAS Signal Extent was compared to the maximum height derived from the ALS DHM and the inventory data base (Fig. 7 and 8). The maximum value of the reference heights was used because the Signal Extent refers to the top height. The correlation is moderate for both spruce and beech stands. Significant differences were observed for the specific acquisition seasons: The Signal Extent of the autumn data correlate well to the reference heights. The correlation for the spring data is moderate and for summer very low. The spring waveforms are from three different laser periods and the correlation results are indeed different for each of them: L2B (Spring 2004) r=0.35, L3B (Spring 2005) r=0.6 and L3H (Spring 2007) r=0.75. Thus it can be assumed, that not the different environmental conditions but rather the laser configurations lead to the different correlation results. The other height metrics based on the ground return peak did not reveal any significant improvements. The comparison with the ALS DHM performed better because this data set has a better spatial resolution than the inventory data base.

Since the test site is characterised by high topography and as indicated by the high Signal Extents shown in Figures 7 and 8, the correction for topographic influences is critical. The Leading and Trailing Edge are two key parameters (see Levsky *et al.* 2007). Both parameters correlate only fairly with the DEM Elevation Range (Fig. 9*a* and 9*c*) which is used as a terrain index similar to Levsky *et al.* (2005). The Trailing Edge Extent is generally lower than the real vertical elevation extent. The Leading Edge is more complex because it is influenced not only by the surface topography, but also by the canopy roughness. The correlation with the ALS DHM Height Range is higher than with the DEM Elevation Range for the summer data (Fig. 9*a* and 9*b*). The canopy roughness therefore plays an important role for the summer waveforms, i.e. during leaf-on conditions. The spring and autumn data seem to be influenced by surface and canopy roughness in a similar way. However, a separated analysis for spruce and beech stands showed that this effect is not species related. The investigation of the Leading and Trailing Edges reveal that the correction using these waveform metrics is not straight forward as assumed before: The slope effect cannot be removed by directly applying the Trailing Edge or Elevation Range. Rather a model including correction factor which is based on Leading and Trailing Edge has to be derived, as proposed by Levsky *et al.* 2007.

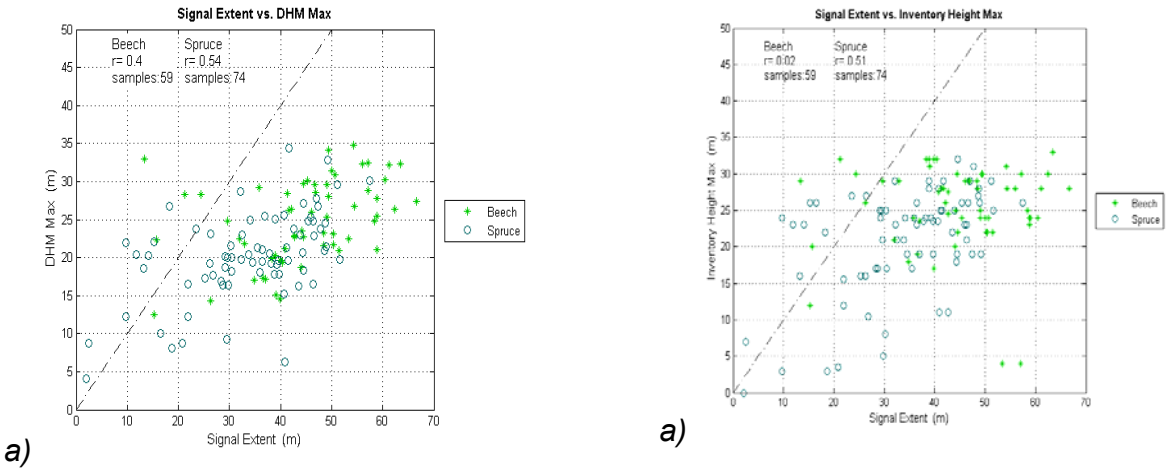


Figure 7. GLAS Signal Extent versus maximum ALS DHM height and maximum Inventory height (species specific)

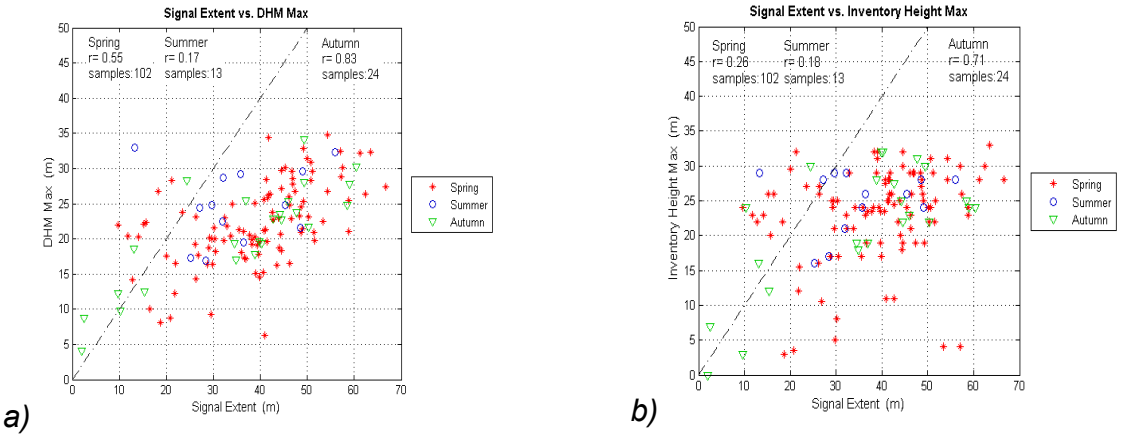


Figure 8. GLAS Signal Extent versus maximum ALS DHM height and maximum Inventory height (season specific)

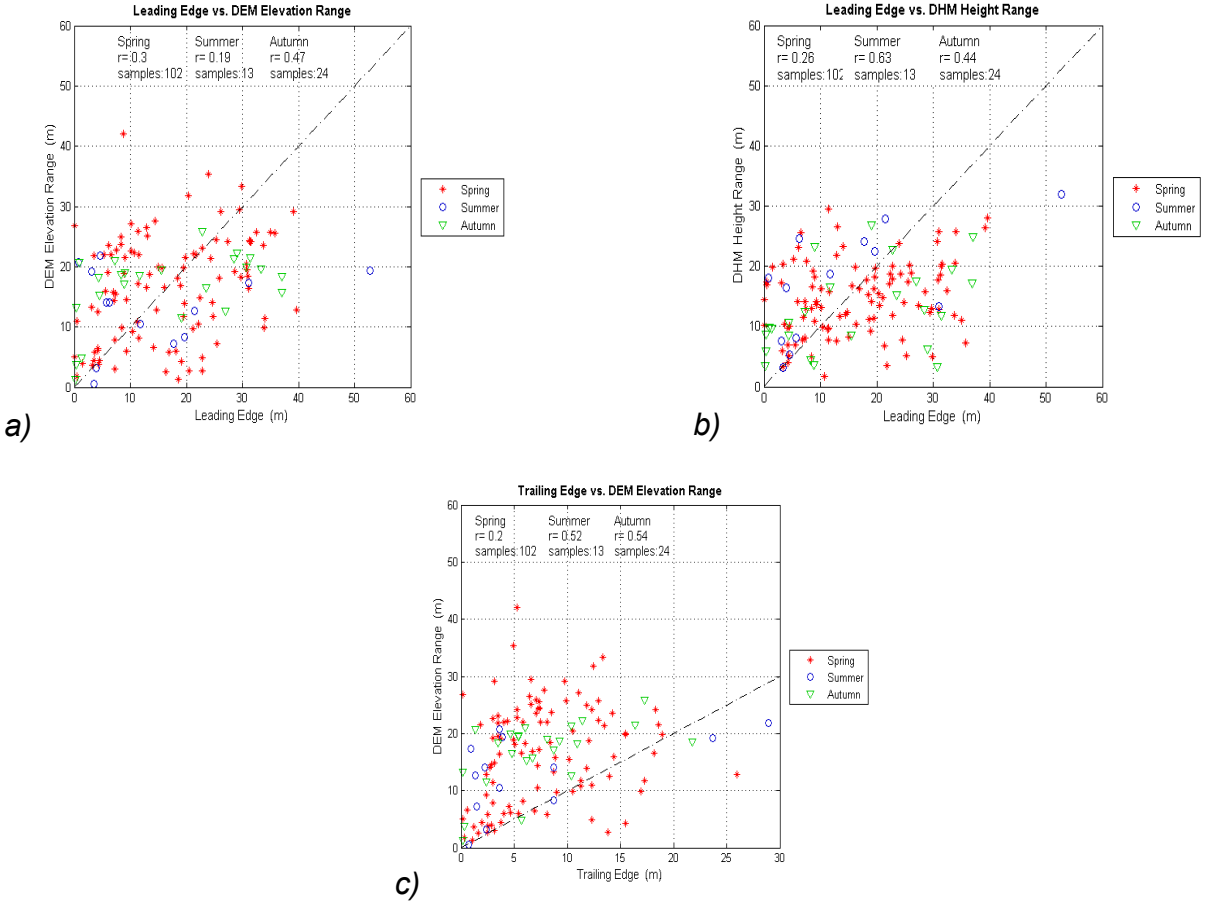


Figure 9. Leading and Trailing Edge: Leading Edge versus a) DEM Elevation Range and b) DHM Height Range, and c) Trailing Edge versus DEM Elevation Range

5. Conclusions

This paper presented some preliminary results of a study combining satellite SAR and lidar data to characterise forest stand structure. The overall goal is to develop a robust model to derive biomass. As a first step, the ICESat/GLAS and ALOS PALSAR data were compared to reference data of a forest inventory data base and an ALS based DHM to fully understand the remote sensing signals regarding the different forest properties.

Against the assumptions and previous findings of other researches (e.g. Luckmann *et al.* 1997, Watanabe *et al.* 2006), only a very low correlation between PALSAR backscatter and forest stem volume could be found. Further analyses are required and the inclusion of interferometric coherence estimations and the high resolution TerraSAR-X data are supposed to enhance the understanding and improve the relationship.

The first investigations of the GLAS data revealed the importance of filtering the data for outliers affected by atmosphere or waveform saturation before analysing. The waveforms show potential to estimate canopy height, but a season or even laser period specific analysis appears useful. The GLAS parameters, especially the Leading and Trailing Edge will be explored more in detail to develop a robust model to correct the Signal Extent for topographic effects and to derive the canopy height. Also some additional parameters like heights based on energy quantiles (e.g. Anderson *et al.* 2006, Sun *et al.* 2008) are planned to calculate. Another idea to improve the estimations is to stratify the data into slope and height classes.

The future working steps will focus on the improvement of the ongoing data analysis, i.e. the enhancement of the SAR and GLAS parameter set. An advanced topographic normalization is planned and the forest inventory data base will be continuously improved and updated. Furthermore, additional SAR images and GLAS footprints for the whole Thuringian Forest will be used to enhance the data basis. In this way it is expected to achieve a full understanding of the SAR and GLAS signals which will be then used to develop a model for biomass retrieval.

Acknowledgements

This research is funded by the German Aerospace Center (DLR). Moreover, the authors would like to thank the Thuringian Forest Agency, the Thuringian Land Surveying Office and the German Weather Service for providing their helpful data and their kind cooperation.

References

ANDERSON, J.E., MARTIN, M.E., SMITH, M.-L., DUBAYAH, R.O., HOFTON, A., HYDE, P., PETERSON, B.E., BLAIR, J.B. & KNOX, R.G. (2006): The use of waveform lidar to measure northern temperate mixed conifer and deciduous forest structure in New Hampshire. *Remote Sensing of the Environment*, 105, 248–261.

CHEN, Q. (2010): Retrieving vegetation height of forests and woodlands over mountainous areas in the Pacific Coast region using satellite laser altimetry. *Remote Sensing of the Environment*, 114, 1610-1627.

DONG, J., KAUFMANN, R.K., MYNENI, R.B., TUCKER, C.J., KAUPPI, P.E., LISKO, J., BUERMANN, W., ALEXEYEV, V. & HUGHES, M.K. (2003): Remote sensing estimates of boreal and temperate forest woody biomass: carbon pools, sources, and sinks. *Remote Sensing of the Environment,* 84, 393–410.

DRAKE, J.B., DUBAYAH, R.O., CLARK, D.B., KNOX, R.G., BLAIR, J.B., HOFTON, M.A., CHAZDON, R.L., WEISHAMPEL, J.F. & PRINCE, S.D. (2002): Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sensing of the Environment*, 79, 305–319.

GONZÀLEZ J.H., BACHMANN, M., SCHREIBER, R. & G. KRIEGER (2010): Definition of ICESat Selection Criteria for Their Use as Height References for TanDEM-X. *IEEE Transactions on Geoscience and Remote Sensing*, 48, 6, 2750-2757.

IMHOFF, M.L., CARSON, S. & JOHNSON, P. (1998): A Low-Frequency Radar Experiment for Measuring Vegetation Biomass. *IEEE Transactions on Geoscience and Remote Sensing,* 36, 6, 1988-1991.

LEVSKY, M.A., HARDING, D.J., KELLER, M., COHEN, W.B., CARABAJAL, C.C., DEL BOM ESPIRITOSANTO, F., HUNTER, M.O. & DE OLIVEIRA JR, R. (2005): Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical Research Letters,* 32, L22S02, doi:10.1029/2005GL023971.

LEVSKY, M.A., KELLER, K., PANG, Y., DE CAMARGO, P.B., HUNTER, M.O. (2007): Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms. *Journal of Applied Remote Sensing,* 1, 013537.

LU, D. (2006): The Potential and Challenge of Remote Sensing-Based Biomass Estimation. *International Journal of Remote Sensing*, 27, 5, 1297-1328.

LUCKMANN, A., BAKER, J., KUPLICH, T.M., YANASSE, C.C.F. & FRERY, A.C. (1997): A Study of the Relationship between Radar Backscatter and Regenerating Tropical Forest Biomass for Spaceborne SAR Instruments. *Remote Sensing of Environment*, 60, 1-13.

MEANS, J.E., ACKER, S.A., HARDING, D.J., BLAIR, J.B., LEVSKY, M.A., COHEREN, W.B., HARMON, M.E. & MCKEE, W.A. (1999): Use of Large-Footprint Scanning Airborne Lidar To Estimate Forest Stand Characteristics in the Western Cascades of Oregon. *Remote Sensing of the Environment*, 67, 298–308.

NÆSSET, E. (1997): Determination of mean tree height of forest stands using airborne laser scanner data. *ISPRS Journal of Photogrammetry & Remote Sensing*, 52, 49–56.

NELSON, R., RANSON, K.J., SUN, G., KIMES, D.S., KHARUK, V. & MONTESANO, P. (2009): Estimating Siberian timber volume using MODIS and ICESat/GLAS. *Remote Sensing of the Environment*, 113, 691-701.

ROSETTE, J., NORTH, P. & SUÁREZ, J. (2008): Vegetation height estimates for a mixed temperate forest using satellite laser altimetry. *International Journal of Remote Sensing*, 29, 5, 1475-1493.

SANTORO, M., ASKNE, J., SMITH, G. & FRANSSON, J.E.S. (2002): Stem volume retrieval in boreal forests from ERS-1/2 interferometry. *Remote Sensing of the Environment*, 81, 19-35.

SUN, G., RANSON, K.J., KIMES, S., BLAIR, J.B. & KOVACS, K. (2008): Forest vertical structure from GLAS: An evaluation using LVIS and SRTM data. *Remote Sensing of the Environment*, 112, 107–117.

WATANABE, M., SHIMADA, M., ROSENQVIST, A., TADONO, T., MATSUOKA, M., ROMSHOO, S.A., OHTA, K., FURUTA, R., NAKAMURA, K. & MORIYAMA, T. (2006): Forest Structure Dependency of the Relation Between L-Band σ0 and Biophysical Parameters*. IEEE Transactions on Geoscience and Remote Sensing,* 44, 11, 3154-3165.

Current Characteristics of Two Space Lidars:

ICESat-2 and DESDynI-Lidar

Ross Nelson[1], Amy Neuenschwander[2], K. Jon Ranson[1], and Bruce Cook[1]

1. Code 614.4, Biospheric Sciences Branch,NASA-Goddard Space Flight Center,
Greenbelt MD 20771 USA; Ross.F.Nelson@nasa.gov
2. Center for Space Research, University of Texas, Austin TX

The US plans to launch two space laser altimeters in the next decade – ICESat-2 and the DESDynI-Lidar. ICESat-2 (Ice, Cloud, and land Elevation Satellite), scheduled for launch in 2015, is primarily designed to monitor ice sheet height changes, sea ice freeboard, and, secondarily, to collect ranging data over much of the Earth's vegetated surfaces. ICESat-2 will be a photon-counting laser ranging system that requires not only new technology to collect the measurements but also new analysis techniques to extract information from the data. The DESDynI-Lidar (DESDynI-L, Deformation, Ecosystem Structure, and Dynamics of Ice-LiDAR), scheduled for launch in 2017, is primarily a vegetation mission designed to measure and monitor vegetation globally in conjunction with a second DESDynI satellite, the DESDynI-Radar (DESDynI-R), an L-band radar designed primarily, but not exclusively, to address the solid earth component of the overall mission. Below we describe the current configurations of both of these laser altimetry missions. It is noted, however, that both missions are in design phase and that the characteristics described in this paper may change. Due to budget constraints, some of these changes may be significant.

Though both three year missions have a vegetation science component, the scientific goals for ICESat-2 and DESDynI-L are significantly different. The vegetation science goal for ICESat-2 is the production of a global vegetation height surface with 3 m accuracy at a 1 km resolution. The ICESat-2 vegetation scientists expect to maintain this accuracy only within forest canopies of 75% crown closure or less under clear sky conditions. As crown closures approach 90%, identification of the ground signal in the photon cloud becomes problematic. ICESat-2 will support and complement the DESDynI-L forest structure measurements, but will do so only insofar as the primary ice objectives of the ICESat-2 mission are not compromised. Vegetation goals for DESDynI mission, which includes both the laser altimeter and the radar, aim to characterize the global carbon cycle through better measurements of vegetation height and canopy density, more accurate, precise estimates of aboveground biomass, and improved enumeration of vegetation disturbance and recovery.

The DESDynI laser altimeter will also make vegetation structural measurements to better characterize habitats to facilitate biodiversity assessments. With respect to biomass, DESDynI has as its primary goal the production of global estimates of woody biomass within ±20 t/ha (90% level of confidence, 1.6 σ) for biomass estimates in the range of 0-100 t/ha, within 20% (1.6 σ) of the estimate for biomass estimates in the range of 100-250 t/ha, and within 50 t/ha (1.6 σ) for biomass capacities exceeding 250 t/ha at a spatial resolution of 250 m globally at end of mission. In areas that support biomass consistently below 100 t/ha, that spatial resolution decreases from 250 m globally to 100 m since this biomass interval is within the range of sensitivity of the L-band radar. The laser altimeter will also provide transects of vegetation profiles over all biomes at a 25 m resolution, 30 m post-spacing, with a maximum 500 m across-track posting at end of mission with a 1 m vertical accuracy in conditions up to 98% canopy cover.

ICESat-2

The ICESat-2 space lidar system will employ photon-counting technology to acquire ranging measurements over ice sheets, sea ice, and terrestrial surfaces near-globally, i.e., up to 86º N and S. This technology is markedly

different from the Geosciences Laser Altimeter System (GLAS) aboard the first ICESat, which operated from January 2003 to October 2009 and which collected over 1.91 billion waveforms globally. GLAS employed a 40 Hz, single-beam, large footprint (50-72 m), near-infrared (1.064 um), waveform laser altimeter. In contrast, ICESat-2 will be transmitting a green pulse (0.532 um) with an expected 1 ns pulse width at 10 kHz. Although the vegetation community and many in the ice community would prefer pulse transmission in the near-infrared (1.064 um), the frequency doubling is necessitated by the fact that there are no space qualified near-infrared detectors capable of measuring photon returns with the rapidity needed to keep pace with target returns. A diffractive optic element will be used to split the green beam from a single laser onboard ICESat-2 into nine beams of varying brightness. These nine beams will exit the spacecraft arranged in a 3x3 square grid. The grid, when yawed 0.6º from the along-track vector of the spacecraft, will result in nine separate beams tracking across the Earth's ice sheets, sea ice, and terrestrial surfaces (Figure 1).

ICESat-2 will be collecting photon ranging information along 9 across-track beams arranged in 3 sets of three triplets. The triplets, or 3-beam clusters, will be spaced three kilometers apart, and within each triplet the 3 beams will be spaced 45-60 m apart. This arrangement was developed to permit calculation of both local and regional across-track slopes to help refine the ice sheet mass balance calculations and to increase sampling intensity globally. The nine beams will not be equally bright. The actual beam strengths are still under discussion; an initial scenario called for the outer triplets to be identical 25 uJ - 50 uJ – 25 uJ combinations and the center triplet to be a 50 uJ – 100 uJ – 50 uJ combination. Subsequent discussions have considered combinations where the three triplets distribute laser power identically, e.g. 25 uJ – 83 uJ – 25 uJ. The need for the unequal beam strengths is driven by the fact that a few of the science disciplines will be imaging both very bright and very dark targets that occur frequently, quickly, and abruptly, leading to the possibility of detector saturation problems. This is especially true with respect to measurement of sea ice extent and freeboard where open water leads (dark targets) are irregularly juxtaposed with sea ice (bright targets) and where measurements at these boundaries are critical to the calculation of sea ice thickness.

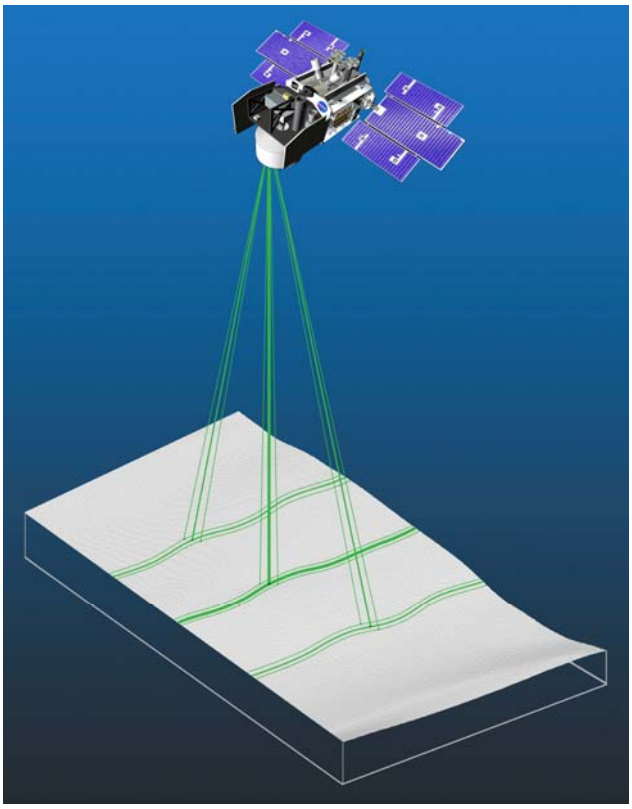
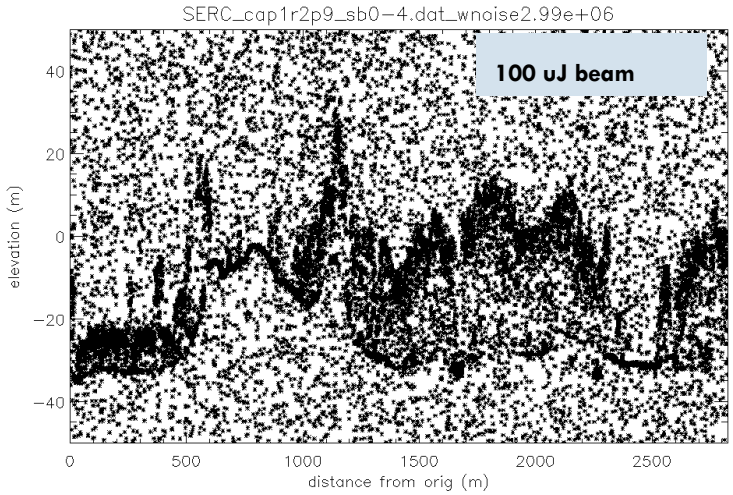


Figure 1. ICESat-2 green (0.532 um) 9-beam, photon-counting system.

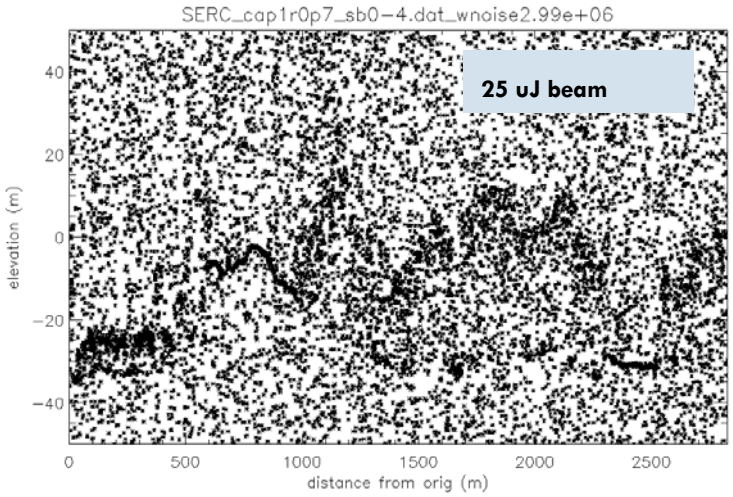
Within any one of the beams depicted in Figure 1, a 40 m diameter area will be illuminated on the ground and the returns from the central 10 m diameter area will be imaged to measure photon returns directly before, during, and after the expected pulse return. Within this vertical window, photon returns are time-stamped, enabling calculation of the range of a particular photon from the satellite. Green photons emitted by ICESat-2 laser, i.e., signal photons, cannot be differentiated from green photons returned from the atmosphere or scattered from adjacent targets, i.e., noise photons; all are collected by the satellite. Combined with the satellite velocity and laser pulse repetition frequency, the satellite moves along-track approximately 70 cm, the next pulse is emitted, and photons collected. Effectively, photons from 10 m diameter footprints are collected every 70 cm along-track, building up a collection of laser photons that have reflected preferentially from the target relative to the ambient noise level. The pointing knowledge associated with the location of each 10 m footprint is expected to be 6 m, 1 σ.

With such a system, the signal-to-noise ratios are very sensitive to the strength of the impinging laser beam, haze levels, and target reflectance properties. Figure 2 illustrates simulated ICESat-2 data acquired by an airborne photon-counting system over a ~25 m tall hardwood canopy in the mid-Atlantic US with crown closures in excess of 95%. Atmospheric noise has been added to simulate a daylight acquisition under mid-latitude, clear sky conditions. The same canopy trace is provided in both Figures 2A and 2B to illustrate expected differences in canopy returns between the strongest and weakest beams, i.e., the 100 uJ beam and a 25 uJ beam. The figure clearly illustrates the fact that the more signal photons that can be placed on target, the greater the likelihood of detecting both top-of-canopy and ground beneath the canopy.

The ICESat-2 satellite will be launched into a near-polar, 94ºinclination, 91 day repeat orbit; equatorial crossing distance between adjacent orbits ascending or descending orbits will be 29.6 km. In order to meet ecosystem scientific objectives, ICESat-2 will be off-pointed up to 1.6º during different repeat periods in the non-polar regions in order to fill in gaps between the sub-satellite ground tracks. By the end of year 2 of a 3 year mission, the maximum distance between adjacent across-track observations will be 1.2 km, barring cloud obscuration.



A.



B.

Figure 2. Simulated ICESat-2 data acquired by an airborne photon-counting system under clear-sky, daylight conditions with (A) a 100 uJ beam and (B) a 25 uJ beam.

DESDynI-Lidar

The DESDynI-Lidar, DESDynI-L, is the primary NASA vegetation mission, one designed to measure and monitor vegetation globally over a three year period in conjunction with a second satellite, the DESDynI-Radar. The radar will be mapping the entire globe and, in the context of vegetation, will be one of the primary tools used to detect the location and areal extent of forest change due to deforestation and fire. The DESDynI lidar measurements will be used as a global sampling tool to measure the structural characteristics of the world's forests in order to estimate global woody biomass and carbon.

The DESDynI three year mission will operate five individual lasers simultaneously over the Earth's terrestrial surfaces in order to acquire waveforms along 5 parallel profiles spaced 850 m apart, i.e., total across-track dimension, 3.4 km. Each near-infrared laser (1.064 um) will be transmitting 9-10 ns pulses at 240 hz, illuminating an area on the ground 25 m in diameter with an along-track post spacing of 30 m. By end of its 3 year mission, DESDynI-L will have collected, by off-pointing up to a maximum of 2.1º, enough profiles to fill in gaps so that the average across-track spacing between adjacent profiles will be 472 m, barring cloud obscuration.

Each waveform will be digitized to record the brightness of the waveform return in 1 ns vertical increments, i.e, 15 cm vertical bins. The pointing knowledge, i.e., the geolocation accuracy, of the centroid of a given DESDynI-L pulse is expected to be 12.5 m (3 σ).

Discussion

Both space lidars will be superior global sampling tools for vegetation assessment. DESDynI lidar scientific requirements call for forest height measurements with 1 m height accuracy (1 σ, top-of-canopy to ground) in areas with zero slope. With a 2.5 dB signal-to-noise ratio, the expectation is that DESDynI-L will be able to sense ground, i.e., return an identificable ground return, in forests growing on flat terrain where canopy densities are ≤98% canopy closure. ICESat-2 will be much more limited. During daylight acquisitions, we expect that no discernable ground return will be available in the photon data if canopy closures exceed 76-86%. In a hazy tropical environment, this lower limit may decreases to 35 – 50% for the 100 uJ beam during daylight acquisitions. At night, these percentages increase to 84-96%; the satellite is expected to spend approximately half it's mission life acquiring night data. In the design cases formulated by an ICESat-2 laser engineer, it is interesting to note that if ICESat-2 could be designed as a near-infrared, 1.064 um system, these canopy closure limits increase to 87-96%. With respect to vegetation measurements, there is an obvious penalty paid by making micropulse ranging measurements in the green.

As mentioned previously, these two satellite systems are currently in design phase and the designs for each are still being discussed. Airborne simulated data sets have been acquired and are being analyzed (1) to develop analysis techniques to extract forest structural information from photon-counting data - ICESat-2, (2) to develop fusion techniques for 25 m waveform lidar and coincident L-band radar - DESDynI, and (3) to determine, using simulated data, accuracy and precision limits associated with forest structural measurements, especially as regards the effects of off-angle pointing and the effects of slope on forest height measurements and biomass/carbon estimates.

Forest parameter assessment for Britain using satellite LiDAR

J.A. ROSETTE†‡, J.C. SUÁREZ‡, P.R.J. NORTH† and S. BATHGATE‡*

j.a.rosette@swansea.ac.uk

†Global Environmental Modelling and Earth Observation Research Group, Geography Department, Swansea University, Swansea, SA2 8PP, UK

‡Forest Research Agency of the Forestry Commission, Northern Research Station, Roslin, Midlothian, EH25 9SY, UK

Abstract

Through sampling of the Earth's surface, satellite LiDAR provides the opportunity of estimating vegetation parameters for large areas. An assessment is made of data from the Geoscience Laser Altimeter System (GLAS) aboard the Ice, Cloud and land Elevation Satellite (ICESat) for the estimation of forest parameters throughout Britain. Data used in this study were captured in February-March 2008 during the L3J laser operation.

The large footprint produced by GLAS is known to cause greater uncertainty in vegetation height estimations for sloped ground where signals returned from the ground surface and vegetation may be combined within the waveforms. An assessment is made of estimated ground elevation using Gaussian decomposition of the GLAS waveforms through comparison with mean ground elevation within GLAS footprint boundaries, calculated using values from a 10m resolution DTM (Ordnance Survey Land-Form Profile DTM product). Greater discrepancies are observed in more mountainous areas such as the Cairngorms in Scotland and Lake District in England, whilst differences are lower for example in the relatively flat region of East Anglia. However, despite this variation, ground elevations estimated from GLAS waveforms show 48% to be within 1m of the control DTM whilst 91.5% are within 5m.

The Forestry Commission in Britain maintains a comprehensive database of the forests that it manages on behalf of the public. For this research study, management and productivity information for these forests in 2007 are used as inputs to yield models in order to predict parameters for forest stands throughout Britain. The Forestry Commission yield models provide a good indication of expected vegetation parameters; however discrepancies can occur where management strategies or habitat conditions differ from those initially anticipated. For the L3J laser operation, 1090 GLAS footprints are positioned within stands recorded within the database. This has enabled vegetation height from yield model predictions and estimates using satellite LiDAR to be compared and offers the opportunity to reduce uncertainty within yield model predictions using observations from satellite LiDAR remote sensing to assess the yield class allocated to stands.

1. Introduction

The significance of the role of forests in mitigating climate change is underlined by international efforts such as reducing emissions from deforestation and degradation (REDD) and reporting on land use, land use change and forestry (FAO *et al.*, 2008; IPCC, 2003). Satellite LiDAR offers a means to quantify the distribution of vegetation and therefore to inform these needs and the potential for LiDAR measurements of vegetation on regional to global scales has been demonstrated by several authors (e.g. Helmer and Lefsky, 2006; Nelson *et al.*, 2008; Nelson *et al.*, 2009; Lefsky, 2010 in press). Issues concerning this are discussed by Dubayah *et al.*, 2010; Hurtt *et al.*, 2010; Nelson, 2010.

This paper presents an initial assessment of data available from the Geoscience Laser Altimeter System (GLAS) for vegetation parameter estimation on a sampling basis for Britain as a complement to conventional methods of forest monitoring and inventory.

1. ICESat/GLAS

This study uses satellite LiDAR data from NASA's Geoscience Laser Altimeter System (GLAS), a full waveform LiDAR profiler, aboard the Ice, Cloud and land Elevation Satellite (ICESat). ICESat was launched in January 2003 and the instrument has been operated for approximately month-long periods during February-March and October-November annually, plus during May-June earlier in the mission. 18 laser operation periods were undertaken until October 2009, when the final of the three lasers aboard the spacecraft ceased firing (GDFC, 2010). The satellite is the first of its kind to capture global LiDAR data for an extended period of time and therefore offers the opportunity to assess the potential of such sensors for vegetation analysis.

ICESat orbits at an altitude of 600km and has emitted pulses at 40Hz using a wavelength of 1064nm. GLAS produced footprints which are distanced at 172m intervals and whose diameters on the ground have varied between laser campaigns from a major axis of 148.6±9.8m to 51.2±1.7m. Further information about the mission and sensor are discussed by Schutz, 2002; Zwally *et al.*, 2002; Brenner *et al.*, 2003; Abshire *et al.*, 2005; Schutz *et al.*, 2005; NSIDC, 2010a.

2. Method

GLAS

Data from the GLAS L3J laser campaign were used in this study (data release 429). These were captured during February-March 2008. Mean footprint dimensions showed a major axis of 58.7±0.6m and an eccentricity of 0.52±0.04. Geolocation accuracy of footprints for this laser campaign has not yet been released.

Vegetation height was estimated from GLAS waveforms using GLA14 product (NSIDC, 2010b). This product provides a model fit to each waveform as the sum of up to six Gaussians.

The beginning of the waveform signal was assumed to be the uppermost intercepted canopy surface. The mean footprint ground elevation was estimated as the centroid of whichever of the lowest two Gaussian peaks has the greater amplitude. Maximum vegetation height within the footprint was calculated as the difference between these two elevations (Rosette et al., 2008).

Ordnance Survey Landform Profile product

The 10 metre resolution Land-Form Profile DTM from the Ordnance Survey was used to determine ground elevation for GLAS footprint locations. A buffer was created surrounding each footprint centre using the mean radius of the footprint major axis (29.35m for the L3J laser operation).

The mean elevation value within this area was calculated and this was compared with the estimated ground elevation identified within the GLAS waveforms.

For the terrain analysis, a filter was applied to exclude waveforms where the estimated ground surface differed from the local mean DTM elevation by more than 50m. These were assumed to be anomalous results, not corresponding to actual intercepted surfaces.

Forestry Commission Sub-Compartment Database and yield models

The Forestry Commission is the UK government department responsible for forestry. It is tasked with managing forest and non-forest resources (farmland, open mountain-tops, heathland, estuarine and riparian habitats) on behalf of the public. The majority of publicly-accessible woodland in Britain is managed by the Forestry Commission of Great Britain, a division of which, Forest Enterprise, maintains a sub-compartment database of

inventory data for the Forestry Commission forest management units throughout Britain. This database contains information relating to local habitat conditions (e.g. soil type, altitude, terrain) and management information (e.g. species, cultivation, planting year, initial spacing, thinning regime applied and yield class).

This information allows yield models to be applied which project a mean growth trend over time; this will vary between species and within species groups according to habitat conditions and management (Edwards and Christie, 1981). These models assume that tree density and stand composition at the time of planting remain constant and therefore do not take account of competitive suppression or dominance, nor of mortality due to natural or external factors.

The comprehensive sub-compartment database and the use of yield tables which are available for the majority of species common in the UK offer the opportunity for forest parameters to be estimated. For this study, software developed within Forest Research, and which incorporates yield models, was adapted to allow biophysical parameters to be calculated for stands managed by the Forestry Commission throughout Britain.

The sub-compartment database is updated annually and the 2007 database was selected for use in this study in order that stand status would most closely correspond with the growing season captured by the L3J laser campaign (February-March 2008). Yield models were applied to calculate top height within each stand. Due to selective thinning or retention of trees as wind breaks along paths for example, several species may be present within each stand. The calculated top height of the tallest species within each stand was selected and this was compared with the maximum canopy height estimated from the waveform analysis for corresponding footprints.

3. Results

Ground identification

Ground elevations estimated from 48% of waveforms were found to be within 1 metre of the mean calculated ground surface using the Ordnance Survey Land-Form Profile DTM. 91.5% of waveform-derived elevations were within 5m of the control DTM (Figure 1). The spatial distribution of differences requires further analysis, however findings suggest that greater discrepancies are observed in more mountainous areas such as the Cairngorms in Scotland and Lake District in England, whilst differences are lower for the relatively flat region of East Anglia.

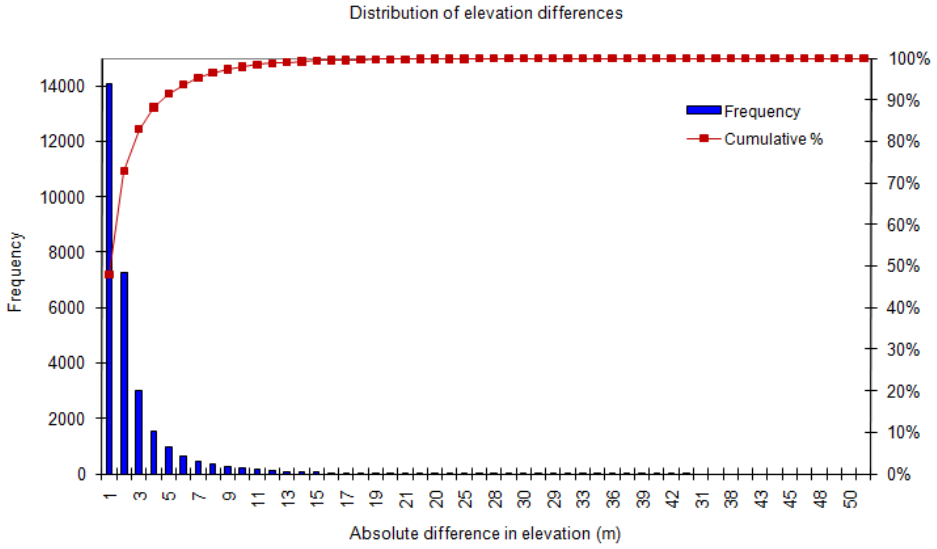


Figure 1. Distribution of differences between ground elevation identified within GLAS waveforms and mean elevation calculated within GLAS footprint boundaries from the Ordnance Survey Land-Form Profile DTM© (Crown Copyright/database right 2010).

Vegetation height estimation

An initial comparison of the tallest predicted top height within stands, calculated from yield models and using local conditions within the sub-compartment database, with maximum canopy height estimated using GLAS waveforms shows large variability and R^2 of just 0.37.Further investigation is needed to explore sources of uncertainty in these estimations compared with ground truth, however error is anticipated from both yield models and GLAS waveforms.

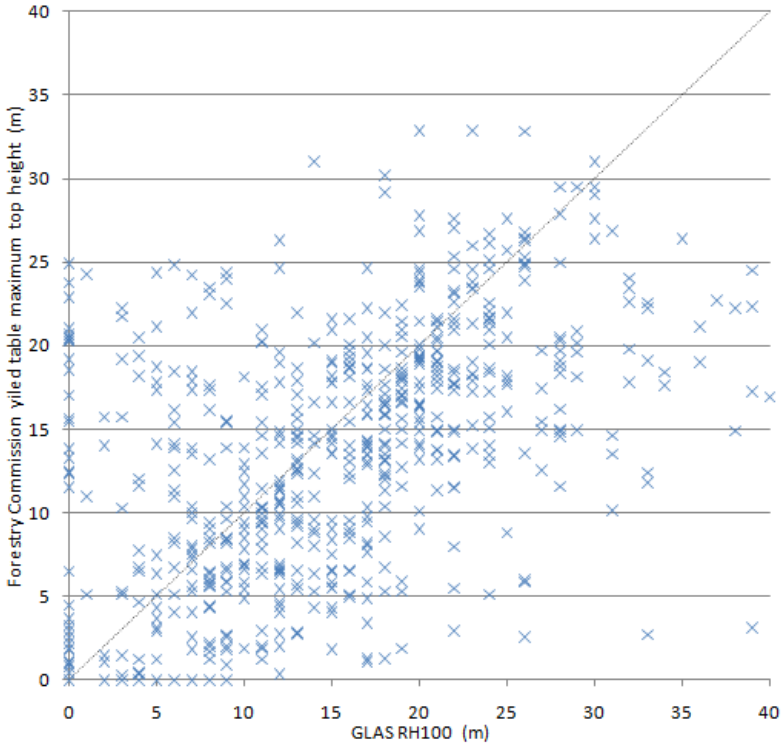


Figure 2. Relationship between top height of the tallest species within stands predicted by yield models and maximum canopy height estimated from coincident GLAS waveforms.

4. Discussion and Conclusion

Analysis of ground elevations estimated within GLAS waveforms have revealed close correspondence with mean elevations calculated within footprint boundaries from an independent 10m resolution DTM (Ordnance Survey Land-Form Profile product). Ground elevation identified within waveforms found 48% to be within 1m of the control DTM whilst 91.5% are within 5m. This suggests that for this laser campaign across Britain, beneath-canopy ground elevations within waveforms have been detected with sufficient accuracy to allow further applications.

Results indicate that greater discrepancies are seen among more mountainous regions of the country and this is to be expected since a ground surface within waveforms can be identified with less certainty on sloped, vegetated surfaces where terrain and vegetation can be at similar elevations. Please refer to Hyde *et al.*, 2005, for further discussions regarding waveform analysis for mountain environments.

A further source of error may be the control DTM itself. The DTM has a stated height accuracy of 2.5m or 5m, with lower accuracy of this product found in mountain and moorland areas (Digimap, 2010). This may contribute to the discrepancies found between GLAS estimates and the DTM at problematic locations.

A large variation is seen between yield model predictions and estimates of vegetation height from GLAS waveforms (R^2 = 0.37). Whilst top height is shorter than maximum vegetation height (being defined as the mean of trees with greatest diameter at breast height within 0.01ha sample plots (Edwards and Christie, 1981)) and error in estimated vegetation height from GLAS of several metres is expected (e.g. Lefsky *et al.*, 2005; Rosette *et al.*, 2008; Sun *et al.*, 2008), these do not fully explain the discrepancies seen.

Some caveats are noted for the country-wide yield model calculations carried out for this study. Where stands are newly-planted, height is calculated as zero whereas in reality planting stocking will typically be 20-60cm in height for stands less than a year old. Where stands are too young for available yield table data, linear interpolation is used between the first yield table age and zero, in order to predict top height. In fact growth is not linear and this may result in minor artefacts among heights of young stands. Finally, not all species' growth curves are currently incorporated within the yield models which are available electronically. In some cases different models are applied for species with similar growth responses whilst some species are excluded from analysis where yield models are not available.

Yield models provide a useful and well-established means of predicting vegetation parameters based on mean growth trends for given management practices and habitat conditions. However actual vegetation parameters may differ from those anticipated, particularly with increasing stand age at later stages of silvilculture (Forest_Research, 2009).

Due to the broad footprint size, combined signals from sloped terrain and vegetation will adversely affect the ability to identify the ground. Additionally, slope alone broadens the returned signal from the ground and reduces its amplitude (North *et al.*, 2010) meaning that the ground can be identified with less certainty. However ground has been found to be estimated to reasonable accuracy and consistency within GLAS waveforms (Sun *et al.*, 2008; Rosette *et al.*, 2010). Canopy height variability and its effect on the ability to identify the canopy top is likely to be an important contributory factor to errors from waveform-derived estimates (Lefsky *et al.*, 2007).

Data collected within UK national forest inventory field plots may offer an independent means of assessing both yield model and LiDAR waveform estimates of forest parameters. Uncertainty is known within sub-compartment database records and, if LiDAR data can offer observations of vegetation height within suitable levels of accuracy, these may be used together with records of stand age to review the yield class assigned to a stand and revise this if necessary.

The ICESat/GLAS satellite LiDAR mission has shown the potential capabilities of this technology for retrieval of biophysical parameters. Whilst the density of sampling and frequency of laser campaigns offer opportunities for repeat measurements which would not be feasible using conventional field measurements, a height error of several metres as currently found would produce a considerable discrepancy in carbon estimations (assuming a consistent bias or spatial patterns to errors). Future satellite LiDAR missions such as NASA's DESDynI (NASA, 2010) aim to significantly improve the density of LiDAR coverage and the accuracy of vegetation measurements.

Acknowledgements

We would like to thank staff at the Forest Research Northern Research Station for accommodating a visiting scientist, for permitting use of their data and models and for providing wide-ranging expertise in support of this work.

The use of the Ordnance Survey Land-Form Profile© product is also acknowledged; Crown Copyright/database right 2010, an Ordnance Survey/EDINA supplied service.

This research is funded by the UK Natural Environment Research Council (NERC) grant NE/F021437/1 and is a partnership project between Swansea University Geography Department and Forest Research Northern Research Station.

References

ABSHIRE, J., SUN, X., RIRIS, H., SIROTA, J., MCGARRY, J., PALM, S., YI, D. and LIIVA, P., 2005, Geoscience Laser Altimeter System (GLAS) on the ICESat Mission: On-orbit measurement performance. Geophysical Research Letters, 32: L21S02.

BRENNER, A., ZWALLY, H., BENTLEY, C., CSATHO', B., HARDING, D., HOFTON, M., MINSTER, J.-B., ROBERTS, L., SABA, J., THOMAS, R. and YI, D., 2003, Derivation of Range and Range Distributions from Laser Pulse Waveform Analysis for Surface Elevations, Roughness, Slope and Vegetation Heights; Algorithm Theoretical Basis Document Version 4.1. NASA Goddard Space Flight Center

DIGIMAP, 2010, Ordnance Survey Products. Available online at: http://edina.ac.uk/digimap/description/products/ (accessed June 2010).

DUBAYAH, R. O., SHELDON, S. L., CLARK, D. B., HOFTON, M. A., BLAIR, J. B., HURTT, G. C. and CHAZDON, R. L., 2010, Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica. Journal of Geophysical Research, 115

EDWARDS, P. N. and CHRISTIE, J. M., 1981, Yield Models for Forest Management; Booklet 48. The Forestry Commission

FAO, UNDP and UNEP, 2008, United Nations Collaborative Programme on Reducing Emissions from Deforestation in Developing Countries. [online] available from http://www.undp.org/mdtf/UN-REDD/docs/Annex-A-Framework-Document.pdf accessed May 2009

FOREST_RESEARCH, 2009, The CARBINE accounting model. URL: http://www.forestresearch.gov.uk/fr/INFD-633DXB Alice Holt Research Station, Surrey, UK (last date accessed May 2009)

GSFC, 2010, ICESat Home Page. URL: http://icesat.gsfc.nasa.gov/icesat Goddard Space Flight Center, Cryospheric Sciences Branch Code 614.1, Greenbelt, USA (last date accessed May 2010)

HELMER, E. H. and LEFSKY, M. A., 2006, Forest Canopy Heights in Amazon River Basin Forests as Estimated with the Geoscience Laser Altimeter System (GLAS). USDA Forest Service Proceedings, RMRS-P-42CD

HURTT, G. C., FISK, J., THOMAS, R. Q., DUBAYAH, R., MOORCROFT, P. R. and SHUGART, H. H., 2010, Linking models and data on vegetation structure. Journal of Geophysical Research, 115

HYDE, P., DUBAYAH, R., PETERSON, B., BLAIR, J. B., HOFTON, M., HUNSAKER, C., KNOX, R. and WALKER, W., 2005, Mapping forest structure for wildlife habitat analysis using waveform lidar: Validation of montane ecosystems. Remote Sensing of Environment, 96 (2005): 427-437.

IPCC, 2003, Good Practice Guidance for Land Use, Land Use Change and Forestry. Institute for Global Environmental Strategies (IGES) for the Intergovernmental Panel on Climate Change. [online] available from: http://www.ipcc-nggip.iges.or.jp/public/gpglulucf_contents.html accessed May 2009

LEFSKY, M., 2010 in press, A global forest canopy height map from the Moderate Resolution Imagine Spectroradiometer and the Geoscience Laser Altimeter System. Geophysical Research Letters, 37

LEFSKY, M., HARDING, D., KELLER, M., COHEN, W., CARABAJAL, C., DEL BOM ESPIRITO-SANTO, F., HUNTER, M., DE OLIVEIRA JR., R. and DE CAMARGO, P., 2005, Estimates of forest canopy height and aboveground biomass using ICESat. Geophysical Research Letters, 32 L22S02.

LEFSKY, M. A., KELLER, M., PANG, Y., DE CAMARGO, P. and HUNTER, M. O., 2007, Revised method for forest canopy height estimation from the Geoscience Laser Altimeter System waveforms. Journal of Applied Remote Sensing, 1: 1-18.

NASA, 2010, DESDynI: Deformation, Ecosystem Structure and Dynamics of Ice. http://desdyni.jpl.nasa.gov/ (accessed May 2010).

NELSON, R., 2010, Model Effects on GLAS-based Regional Estimates of Forest Biomass and Carbon. International Journal of Remote Sensing, 31 (5): 1359-1372.

NELSON, R., NÆSSET, E., GOBAKKEN, T., STÅHL, G. and GREGOIRE, T., 2008, Regional Forest Inventory using an Airborne Profiling LiDAR. Journal of Forest Planning, 13 (Special Issue 'Silvilaser'): 287-294.

NELSON, R., RANSON, K. J., SUN, G., KIMES, D. S., KHARUK, V. and MONTESANO, P., 2009, Estimating Siberian timber volume using MODIS and ICESat/GLAS. Remote Sensing of Environment, 113 (3): 691-701.

NORTH, P. R. J., ROSETTE, J. A. B., SUÁREZ, J. C. and LOS, S. O., 2010, A Monte Carlo radiative transfer model of satellite waveform lidar. International Journal of Remote Sensing, 31 (5): 1343-1358.

NSIDC, 2010a, ICESat Laser Operation Period Attributes. URL: http://nsidc.org/data/icesat/docs/glas_laser_ops_attrib.xls National Snow and Ice Data Center, Colorado, USA (last date accessed June 2010)

NSIDC, 2010b, ICESat/ GLAS Data at NSIDC. URL: http://nsidc.org/daac/icesat/ National Snow and Ice Data Center, Colorado, USA (last date accessed June 2010).

ROSETTE, J. A. B., NORTH, P. R. J. and SUÁREZ, J. C., 2008, Vegetation Height Estimates for a Mixed Temperate Forest using Satellite Laser Altimetry. International Journal of Remote Sensing, 29 (5): 1475-1493.

ROSETTE, J. A. B., NORTH, P. R. J., SUÁREZ, J. C. and LOS, S. O., 2010, Uncertainty within Satellite LiDAR Estimations of Vegetation and Topography. International Journal of Remote Sensing, 31 (5): 1325-1342.

SCHUTZ, B., ZWALLY, H., SHUMAN, C., HANCOCK, D. and DIMARZIO, J., 2005, Overview of the ICESat Mission. Geophysical Research Letters, 32: L21S01.

SCHUTZ, B. E., 2002, Laser Footprint Location (Geolocation) and Surface Profiles. NASA Goddard Space Flight Center

SUN, G., RANSON, K. J., MASEK, J., GUO, Z., PANG, Y., FU, A. and WANG, D., 2008, Estimation of Tree Height and Forest Biomass from GLAS Data. Journal of Forest Planning, 13 (Special Issue 'Silvilaser'): 157-164.

ZWALLY, H., SCHUTZ, B., ABDALATI, W., ABSHIRE, J., BENTLEY, C., BRENNER, A., BUFTON, J., DEZIO, J., HANCOCK, D., HARDING, D., HERRING, T., MINSTER, B., QUINN, K., PALM, S., SPINHIRNE, J. and THOMAS, R., 2002, ICESat's Laser Measurements of Polar Ice, Atmosphere, Ocean and Land. Journal of Geodynamics, 34: 405-445.

Assessment of branchiness in a Pinus pinaster plantation by terrestrial laser scanner data as a link between exterior and interior wood properties

HANS-JOACHIM KLEMMT†, THOMAS SEIFERT‡, STEFAN SEIFERT§, ANTON KUNNEKE‡, BRAND WESSELS‡ , HANS PRETZSCH§

hans-joachim.klemmt@lwf.bayern.de

†, Sachgebiet Waldbau, Bayerische Landesanstalt für Wald und Forstwirtschaft, 85354 Freising, D

‡, Department of Forest and Wood Science, Faculty of AgriSciences, Stellenbosch University, Stellenbosch, SA

§, Lehrstuhl für Waldwachstumskunde, Technische Universität München, 85354 Freising, D

Abstract

Forest inventories collect individual geometric tree parameters such as diameter at breast height (DBH), tree height, stem profiles etc. which are relevant especially for forest industries or forest owners. For the next processing steps after felling and removing trees from forests information on geometry and dimension of the logs as well as on wood quality related parameters is of interest especially for wood industries. In this respect there is a special interest for branchiness. This paper presents an algorithmic approach to determine the parameters branch whorl height and branch number for the most valuable parts of coniferous trees. The difference between branch whorl heights delivers valuable information on the volume and the length of possible logs without branches. In combination with the results for branch whorl heights this supports the decision where stems should be cut into logs for optimizing results. The presented algorithm was tested in a field campaign 2009 in a *Pinus pinaster* plantation in South Africa. The promising results show that terrestrial lidar technology is able to deliver information on forests stands which are of interest for forest industries as well as for wood industries. This link can be considered as an important extension of sustainable management oriented forest planning because it provides holistic information for different actors within the forestry-wood-chain.

IAPRS: international Archives of the photogrammetry, Remote Sensing and Spatial Information Sciences

1. Introduction

Tree branches are woody structural members connected to but not part of the central trunk of trees. Biologically they are of importance because they carry leaves, the above-ground plant organs specialized for photosynthesis. They also enable the transport of water and nutritions as well as of the results of photosynthesis between the stem and the peripherical organs of trees.

From wood utilization point of view branches are the most important wood defects. Especially for the utilization with high demands for stiffness or strength i.e. for construction solid woods, boards with many branches or with widespread branches are unwanted. Also knottiness is a major factor in value timber recovery in sawmills. Costs of production processes of high value timber increase linearly with an increasing number of branches and knots.

These stress ratio was the reason for many scientific projects in the past either on growth aspects of trees and branches or on the importance of branches for different wood utilization purposes. Representatives for the first group are e.g. Hapla et al. (2000), MacGuire et al. (1991), Mäkinen (1999), Meredieu et al. (2002), Moberg (1999), Persson et al., (1995), Polman et al., (1995), Sautter (1993), Stahl et al., 1995 or Witkowska (2008).

Examples for the second group are Agestam et al. (1998), Houllier et al. (1995), Jeffrey (1979) or Lippemaier (1977).

In the past the quantification of branchiness at standing mature trees was nearly impossible and branch measurements at felled trees was a time consuming work. Especially this fact lead to a third group of scientific works which concentrated on the modelling of branchiness based on sampled field data. Works which concentrated on this aspect are e.g. Colin et al. (1992), Kellomäkki et al. (1999), Mäkinen et al. (1999 a and b, 2001, 2003), Moberg (2001), Seifert (1999), Seifert (2003), Seifert and Pretzsch (2004) or Struck and Dohrenbusch (2000). Meanwhile new measurement technologies make it possible to achieve branchiness directly and more precise. The most important was the application of laser scanning technology for forest mensurational purposes. While airborne laser scanning is already established and state of the art for the achievement of areal forestry information terrestrial laserscanning is still under development for the application in forest practice (Roberts et al., 2007). The advantage of terrestrial laserscanning technology is that it is capable to measure single standing trees in details with high accuracy. In the last decade an increasing number of scientific studies have been done in this field. Hopkinson et al. (2002) evaluated a ground based lidar scanning system to assess its potential for tree-level forest mensuration data extraction. They found that all parameters (stem location, tree height, DBH, stem diameter and timber volume could be measured with good accuracy. Thies et al. (2004) developed a method and an algorithm for reconstructing the three-dimensional surface of stems based on terrestrial lidar data. They applied their methods for the reconstruction of a European beech tree (*Fagus sylvatica*) and a Wild Cherry tree (*Prunus avium*) and discussed suggestions for the further development. Henning and Radtke (2005) describe methods for indentifying trees in range images and co-registering range images from different vantage points. They also compared upper-stem diameters and branch heights derived from the range images to measurements made after the felling of a scanned collective of nine *Pinus taeda* trees. Results showed excellent agreement between the lidar-derived diameter estimates and caliper measurements for bole sections below the base of live crown. Less accurate estimates (<2 cm) were obtained for stem heights up to 13m. Results indicated the potential for accurate assessment of branch and whorl heights using ground-based scanning-lidar, with the greatest accuracy likely to be realized for branches near the base of the live crown and below it. Watt and Donoghue (2005) worked on the derivation of the most important forest inventory related parameters with TLS-technology. The results are conform to the results from Hopkinson et al. (2002). Bienert et al. (2007) presented an algorithmic approach to extract forest inventory related parameters automatically from three-dimensional TLS-point-cloud-data. The combination of already evaluated methods and its improvement lead to promising results towards a semi-automatically or automatically software system for the derivation of forest inventory parameters from this point clouds. Besides the volume results for grading purposes especially branchiness related information is of interest because this is the most important economically yield influencing factor. Particularly interested in this topic is plantation forestry with its main focus on economical optimized yield but it has also a big relevance for multifunctional forestry which has the aim to optimize wood production as well as several other functions. However, only a few attempts have been made to assess branchiness (Seifert and Seifert, 2006).

Within this study we concentrate on data from a Maritime pine plantation in South Africa. Maritime Pine (*Pinus pinaster*) is a pine originally native to the western Mediterranean region. The natural range ot the species extends from Portugal and Spain north to southern and western France, east to western Italy, and south to northern Morocco, with small outlying populations in Algeria and Malta (possibly introduced by man). It generally occurs at low to moderate altitudes, mostly from sea level to 600 m, but up to 2000 m in the south of its range in Morocco. It is widely planted for timber in its native area, being one of the most important trees in forestry in France, Spain and Portugal. It has become naturalised in parts of southern England, South Africa and Australia (Schütt et al., 1992).

The aim of this study is to gain preliminiary information on the possibility to assess branchiness by terrestrial laserscanning technology in plantations. It focuses on the partial aspects of the heights of branch whorls and the assessment of the number of branches at a whorl in the most valuable, lower parts of the stems. This study excludes the aspect of the possibility to measure branch size itself. Within this study an algorithm is developed to obtain the mentioned parameters automatically from three-dimensional laser scanning data. The accuracy of this algorithm is tested for data gained in a field campaign in a 30-year-old Maritime Pine plantation in the Western Cape Region (South Africa).

2. Material and Methods

Site description

The site where LIDAR measurements were made is compartment A33a of the La Motte MTO plantation. The area is situated about 3 kilometres west of the town of Franschoek in the Western Cape province of South Africa (Latitude 33° 54' 54" South, Longitude 19° 05' 20" East and Altitude 285m). The area has a Mediterranean climate with most rainfall in the cool winter months (MAP 950mm). Summers are normally dry and warm. The site is on a mid slope with steep to very steep terrain. The geology of the area is sandstone with shallow sandy to sandy loam dystrophic soils with a maximum depth less than one metre and a large percentage of the site with a rooting depth of about 50 cm. The limitation in depth is due to sandstone rock from which the soils were derived.

Description of the measurement device

For this study a Trimble FX phase shift terrestrial laser scanner was used. A technical description is given in table 1. This scanner was positioned four times on each edge of the stand. Data was processed after a co-registration of all settings. Laser scans were done in June 2009. Data was filtered by the "small areas"-filter form Trimble FX controller software to get rid of the noisy small areas especially in the crown and needle region. The application of this filter in combination with the maximum range (c.f. Tab. 1) of the used Trimble FX scanner lead to data sets with a high level of detail for stem axis and branches.

Table 1: Technical description of Trimble FX phase-shift terrestrial laser scanner

Performance	**System Specifications**	**Physical**
Range1,2. up to 31 m single pass; 38 m double pass Scanning speed . 175,000 points per second Typical scan time . 5 minutes (single pass) Standard deviation (range). 1 mm @ 15 m single pass, 0.3 mm @ 15 m double pass on 90% reflectivity Target acquisition . std dev. <1 mm @ 15 m Scan resolution. configurable: 1.5 mm @ 10 m; 4.2 mm @ 27 m Spot size. 2.3 mm @ 5 m; 8.6 mm @ 25 m	Scanner type. phase shift high speed scanner Laser wavelength. 690 nm (red) Laser type. continuous wave Laser power. 15mWt Laser class: (IEC EN60825-1). 3R Angular resolution. 0.002° Field of view. 360° x 270°	Dimensions. 466 mm L x 172 mm W x 245 mm H Weight .11 kg (24 lb) Power Supply • DC 18-24V Li-ion battery (4 hours) • AC 110–220V Power consumption. 30 W (55 W peak) Carry on case . 559 D x 355 W x 229 H mm (22 in D x 14 in W x 9 in H); weight 5.4 kg (12 lb) [Conforms to most airlines requirements for carry on baggage3] Transportation case. 927 D x 902 W x 508 H mm (36.5 in D x 35.5 in W x 20 in H in); weight 33.5 kg (74 lb) with scanner [Conforms to most airlines restrictions for checked baggage4] Environmental. calibrated Operating temperature. 5 °C to 45 °C, in non-condensing atmosphere

Stand characteristics

For evaluation purposes in the field diameter in breast height was measured as well as the height of branch whorls up to a height of 10 meters. The diameters in breast height of all standing trees were measured to the nearest diameter in August 2009, using a fibre diameter tape. Whorl heights were measured with a Mark IV height measurement instrument. The number of branches was counted from the ground after visual assessment. Within these study data for 31 trees were gained. Table 2 gives a short characteristic of the stand and a descriptive statistic on branchiness parameters of these data.

Table 2: Stand characteristics of the test site near Franschoek

Parameter	Per scanned area	Per ha
Size of scanning area (20m*20m)	400 m²	
Number of trees in scanned area	31	775
Mean diameter [cm]	29,1	
Mean height [m]	24,7	
Basal area [m²]		51,2
Volume [m³]		639, 4

Data preparation

Data preparation was done with RiScanPro Software, version 1.4.3. The data was imported, the segmentation of the three-dimensional point cloud in trees was done manually with this software. Data for each tree was exported in text file format. Further processing steps were done with the statistical computing language R (R Core Development Team, 2010).

For each manually separated tree data set, the data belonging to the ground or soil was eliminated automatically following a methodology presented by Teutsch (2007). This was done by subdividing the 3D-data for each tree according to their x and y-coordinates within four sectors of the same size. For each sector the point with the minimum z-value was found. With the plain3d of the scatterplot3d package of R a three-dimensional linear regression model was fitted to these points. Assuming that all datasets below the resulting regression plane are not part of the tree they were eliminated.

Description of the algorithmic approach for automated determination of branchiness

a) Virtual pruning

Within the next processing step termed virtual pruning datapoints belonging to branches were eliminated from the point cloud around the stem. Therefore for a 25cm slice at the base of the tree as well as for top of the tree circles were fitted on the data. For the center of the circle at the base and at the top of the tree a cylinder was calculated with the doubled radius for the base circle, resulting in a frustrum of a cone. In the next step all data of the point cloud were eliminated which lied outside the frustrum. Figure 1 illustrates this processing step schematically.

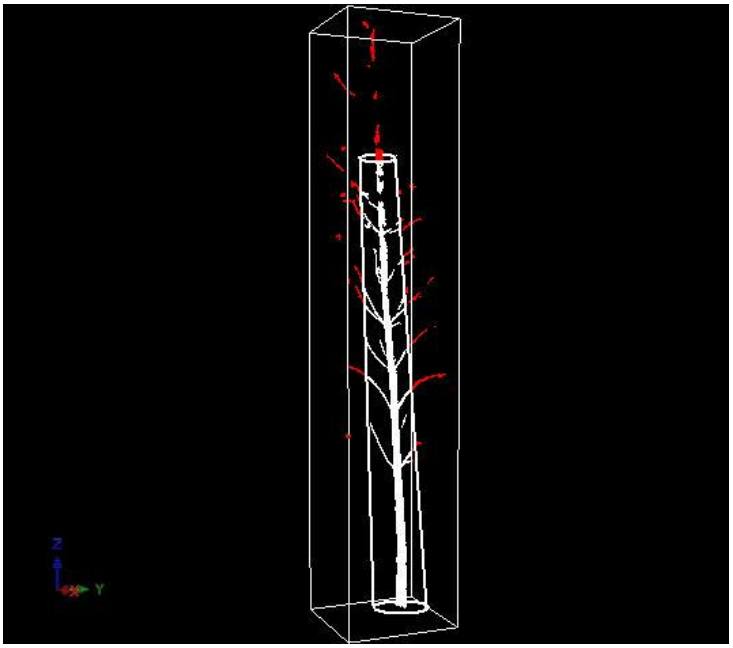


Figure 1: Schematical Illustration of the virtual pruning processing step for tree no. 37 of Franschoek Plantation. Red marked data points outside the frustrum of the cone have been removed for further processing steps.

b) Circle fitting and testing its goodness of fit

Within the next step we assumed that it is possible to detect branch whorls by the results of circle fitting to stem axis. For parts of stem axis without branches circle fitting should work very well while the goodness of fit within branch whole sections should be worse due to additional points of the branches on the laserscan. Therefore after the virtual pruning step circles where fitted on the resulting tree data, slicing the stem in vertical distances of 5cm. On these slices the least square circle fitting routine of the circular package of R was applied. Figure 2 shows the goodness of fit (expressed by QFactor) exemplarily for one tree out of the data set. A threshold value of 0.1 (determined based on a variation for all trees and a comparison of the results with field observations) was set for all trees. This treshold seperates good and bad circle fits. Following the values along the stem axis from the ground to the top it was assumed that first time the QFactor value was higher than the set threshold value the height was reached where a branch whorl is located. (It was also assumed that below a height of 1.5m no branch whorl exists).

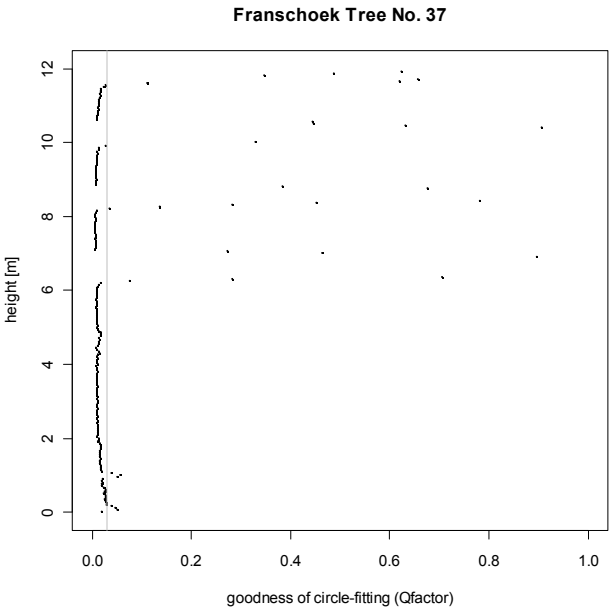


Figure 2: Goodness of fitting values for tree no. 37

c) Density based clustering for determination of the number of branches at a whorl

At a distance of 25 cm over the determined height of the branch whorl a 10 cm slice was cut out of the data sets for each tree. These slices contain reflection data from the stem axis as well as from branches. On these data a density based cluster algorithm was applied (DBScan from the fpc-package in R; Ester et al., 1996) which tries to build separate clusters for each branch as well as for the stem axis. Figure 3 shows the data for such a slice and the results of application of the DBScan algorithm. The resulting number of branches was calculated by the following term:

Number of branches per whorl = number of clusters – 1

This subtraction takes into account that one cluster for the stem axis was generated.

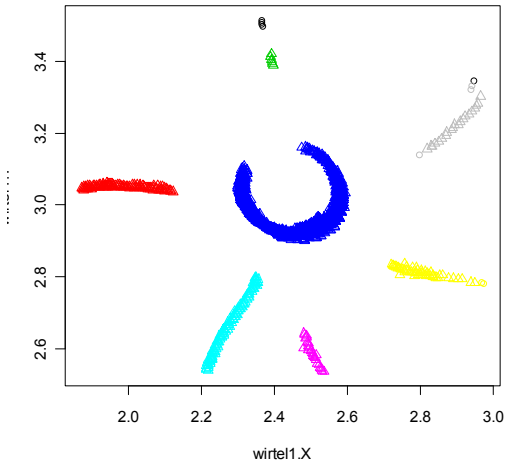


Figure 3: 10 cm slice at a distance of 25 cm above a calculated branch whorl height. Branches and stem axis are drawn in different symbols as a result of the application of the DBScan algorithm.

3. Results

Dbh-measurement

Figure 4 shows a comparison between the measured dbh (diameter in breast height) in field and the automatically calculated dbh. The relationship is linear and positive (ad. R^2=0.96). The mean difference between field measured dbh and automatically derived dbh is -0.45cm which lead in consequence to an underestimation for terrestrial laser scanner data. To test the fitted values linear RMA regression (package smart) was applied on the data. The estimate of the slope (b) = 1.010335 with confidence limits {0.937799; 1.088482} indicates that the hypothesis (H0<>1) can be rejected as the value b is contained in the interval of 95% confidence limits. The hypothesis that the elevation differs significantly from 0 can be rejected as the value of a lies in the interval of 95% confidence limits.

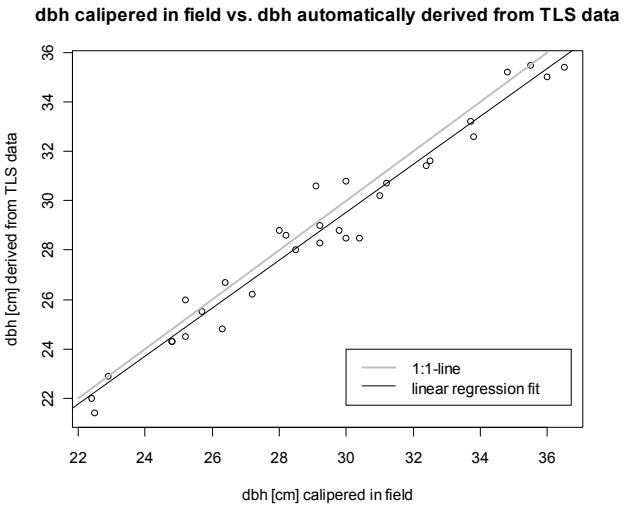


Figure 4: Comparison of field measurement vs. automatically scanner derived dbh-values for Pinus pinaster trees in Franschoek study area.

Branch whorl heights

In figure 5 (a to c) a comparison for the automatically derived branch whorl heights and the heights measured in field is shown.

For the lowest, **first whorl** (fig. 5a) above the ground the mean value for the manual measurement was 5.33 m while automatically determined value value was 5.35m. The relationship between both shown in figure 5 a is also linear and positive with an adjusted R^2-value of 0.92 for RMA regression. Also the estimates for the slope and the elevation indicate that neither slope differs from 1 nor elevation differs from 0.

For the **second whorl** (fig. 5b) the mean value for the manual measurement is 6.98 m while the automatically derived value is 7.08m. Here also the relationship is linear and positive (ad. R^2=0.88). There also the estimates for the slope and the elevation indicate that neither slope differs from 1 nor elevation differs from 0.

For the **third whorl** (fig. 5c) the mean value for the manual measurement is 8.55 m while the automatically derived value is 8.97m. Here also the relationship is linear and positive although variance increased (ad. R^2=0.67). Also the estimates for the slope and the elevation indicate that neither slope differs from 1 nor elevation differ from 0.

Number of branches per whorl

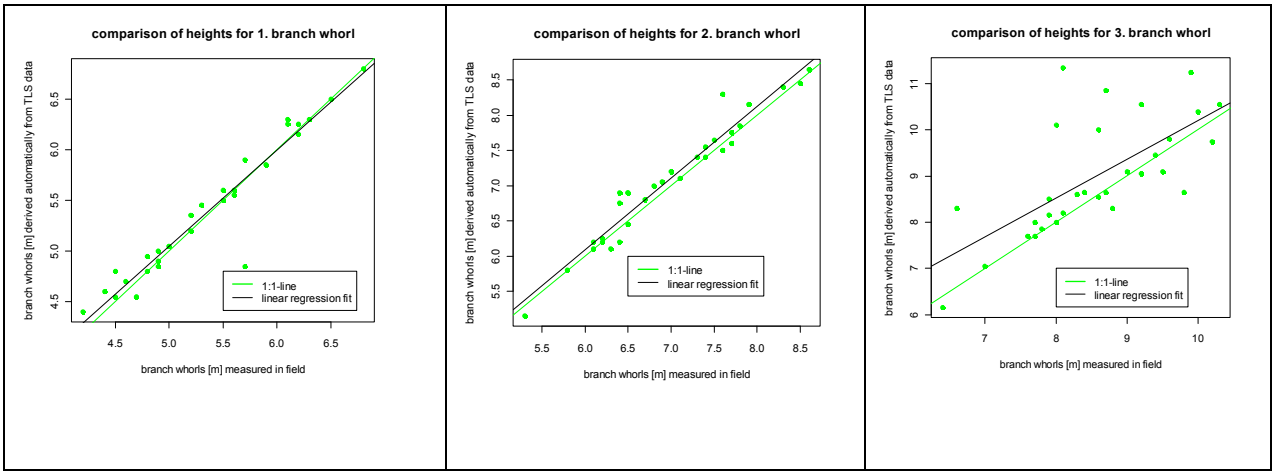


Figure 5 (a-c from left to right): Comparison of field measured branch whorl heights and automatically derived heights from terrestrial laser data for the first 3 whorls above the ground within the most valuable parts of the Pinus pinaster trees in the Franschoek study area.

To ensure that branch numbers got compared for the same whorls a manual verification and in some cases new assignment of comparable branch whorls took place. In figure 6 the deviation between the automatically computed number of branches per whorl and the counted number from the ground is shown. For the first whorl above the ground the developed algorithm found in 13 cases an identical number of branches and in 18 cases a lower number. For the second whorl in one case one branch more than counted from the ground was detected by the automated system, in twelve cases it was identical and in 18 cases there was a negative deviation. For the third whorl there was also in one case a positive deviation by one, an identical number in 16 cases and a negative deviation in 13 cases.

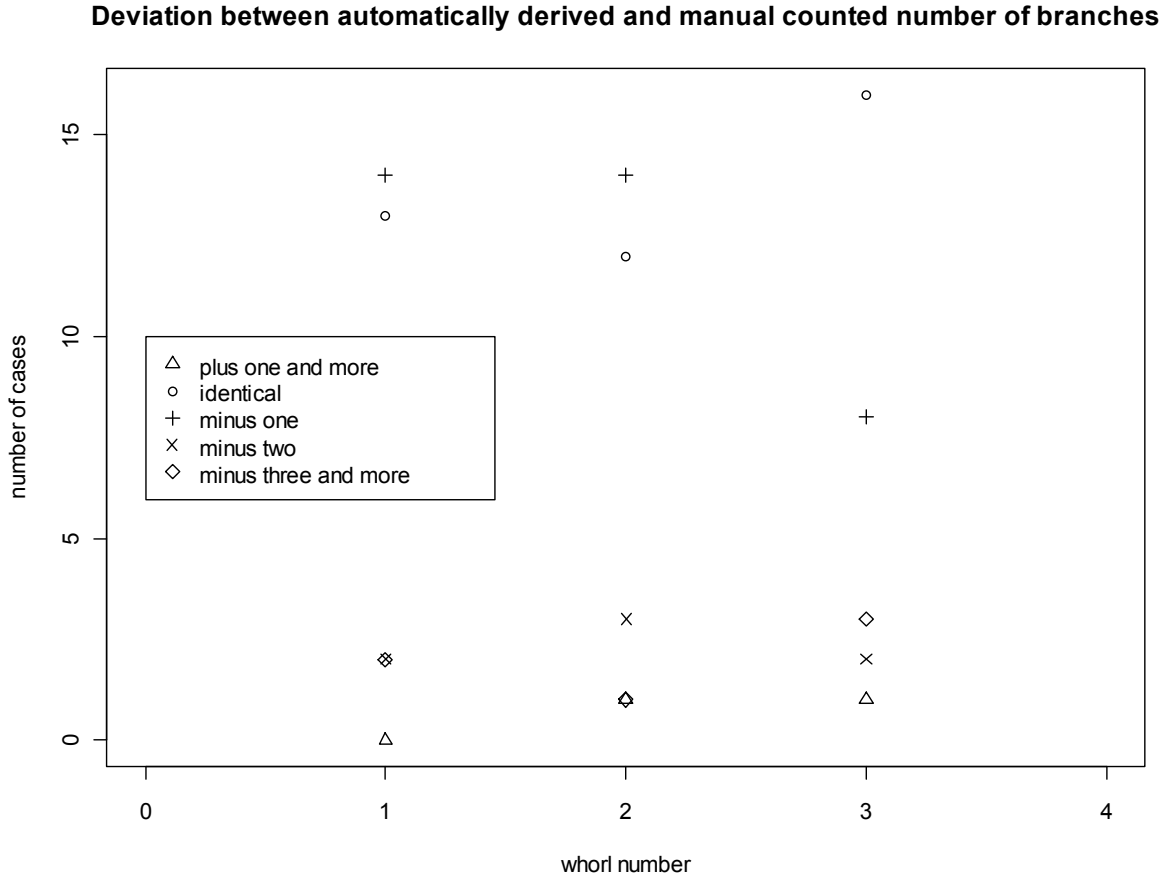


Figure 6: Deviation between the automatically derived number of branches at the trees in Franschoek Pinus pinaster plantation and the counted number from the ground. "Minus one" indicates for example that the number of branches detected by the developed algorithm was lower by one than the number of branches at a whorl counted from the ground.

4. Discussion

The cross-section or the diameter in breast height can be modeled by different geometric shapes, such as circles (Henning et al. 2004, Bienert et al., 2006), B-Splines (Pfeifer and Winterhalder, 2004), polygons (Wezyk, 2007) or cylinders (Pfeifer et al., 2004, Hopkinson et al., 2004, Thies et al., 2004). Within this study we applied circle fitting statistics because tests have proven that this methodology is more robust especially for sliced data with no complete arcs. The comparison between traditional dbh measurement in field and terrestrial laser scanning derived values has shown a high level of accordance which supports the results of Hopkinson et al. (2004) or Wezyk et al. (2007). This shows that it is possible to obtain reliable dbh values automatically from terrestrial laser scan data for this simple stand situations (homogenous stand structures, no undergrowth layer, regular stem shape a.s.o.).

In the past only a few scientific papers dealt with the derivation of branchiness related topics from terrestrial laser scanning data. Although Henning and Radtke (2006) did not provide a rigorous comparison of branch measurements with lidar derived estimates they concludes that the possibility is given that ground based lidar

could be used to determine branch heights at or below the base of the live crown as well as parameters like branch diameters, azimuth, inclination or possible length. The results of these study support this opinion. For plantation trees in South Africa good results were found for the branchiness parameters “branch whorl height” and “number of branches per whorl”. For the first considered parameter the level of accordance between field measurements and the estimations from terrestrial laser scanner data decreased with an increasing distance of the whorl above the ground. One reason for this finding could be that there have been problems in the definition of the “height of a branch whorl” for the traditional measurements. A closer look on the laser derived data has shown that in some cases within the field campaign whorls were recorded which have not been whorls but only single branches within the internodial sections of the stem. The developed algorithm can be readjusted to detect each irregularity within the stem axis so that for this parameter no further development should be necessary in the future. For the second parameter “number of branches” within this study a tendency towards an underestimation for the automatically derived number estimation was found. The reason for this might be the sub-optimum positioning of the scanner at the edges of the *Pinus pinaster* stand. Further studies should work on an optimization of the positioning of terrestrial laser scanning devices in forest stands to receive optimum results.

Within this work deliberately the diameters of branches got not computed e.g. following a methodology presented by Pfeifer and Winterhalder (2007) because no validation data was available. Further works should try to derive information on this third branchiness related parameter as well. Also further development could try to compare automatically TLS based branch diameters with results of well calibrated models for this tree species. It was obvious that the lower branch whorls could be detected with the presented method. Branch whorls and branch diameters higher up the crown were usually not detectible because of shading effects of the stem and lower crown parts in Pinus pinaster. Methods to determine with mathematical models based on lidar measured crown shape information have been proposed here (Seifert and Seifert, 2006).

5. Conclusions

The results of this study have shown that it is possible to achieve information on the branchiness related parameters “whorl height”, which is relevant for the yield of branchless bole sections and “branch number per whorl” for simple stand situations in plantations more or less automatically from terrestrial laser scanning data. The presented algorithmic approach should be extended by a method for optimization of the positions of the terrestrial laser scanner to each other. Further on the virtual pruning step should be advanced for irregular stem forms.

The application of the presented methodology on data of tree stands, which are more structured vertically and horizontally, and the comparision with traditional field measurements should give indices for further improvement.

Nevertheless the developed algorithm is considered to be a first practicable approach to deduce wood property features from external high detail measurement of standing trees with a modern technical device. Further applications could try to integrate this branchiness information from terrestrial laser scanner data in sawing simulation programs (Murphy, 2009; Seifert, 2010) which would enable a more precise determination of the economical optimized time of harvesting of trees in the future.

Acknowledgement

The authors thank Sean Dane (Optron) for the field campaign in Franschoek with the Trimble FX Laserscanner and for the data provision and Daniel Alexander Müller and Andrea Oumeddah-Seitz for data preparation. We also thank the Bavarian State Forest Service and the European Union (Interreg IIIa) for funding preceding fundamental research related with this topic.

References

Agestam, E., Ekö, P-M., Johannson, U., 1998: Timber quality and volume growth in naturally regenerated and planted Scots pine stands in S.W. Sweden. Studia Forestalia Suecia, No. 204, 17 p.

Bienert, A., Scheller, S. Keane, E., Mohan, F., Nugent, C., 2007: Tree detection and diameter estimations by analysis of forest terrestrial laser scanning point clouds. Proceedings of the ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007, p. 50-68

Chasmer, L., Hopkinson, C., Treitz, P., 2006: Investigating laser pulse penetration through a conifer canopy by integrating airborne and terrestrial lidar. Can. J. Remote Sensing, Vol. 32, No. 2, pp. 116

Colin, F., Houllier, F., 1992: Branchiness of Norway Spruce in northeastern France: prediciting the main crown charactersitics from usual tree measurements. Ann. Sci For. (1992), 49, 511-538

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996: A Density-Based Algorithm

for Discovering Clusters in Large Spatial Databases with Noise. Institute for Computer Science, University of Munich. Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96).

Hapla, F., Oliver-Villanueva, J.-F., Gonzalez-Molina, J.-M., 2000: Effect of silvicultural management on wood quality and timber utilisation of Cedrus atlantica in the European Mediterranian area. Holz- als Roh- und Werkstoff 58 (2000), p. 1-8.

Henning, J.G., Radkte, P., 2006: Detailed stem measurement of standing trees from ground-based scanning lidar. Forest Science 52(1). 2006, p. 67-80.

Hopkinson, C., Chasmer, L., Young-Pow, C., Treitz, P., 2004: Assessing forest metrics with ground-based scanning lidar. Can. J,. For. Res. 34: 573-583

Houllier, F, Leban, J-F., Colin, F., 1995: Linking growth modelling to timber quality assessment for Norway Spruce. Forest Ecology and Management 74 (1995), p. 91-102

Ikonen, V.-P., Kellomäkki, S., Peltola, H., 2003: Linking tree stem properties of Scots pine (Pinus sylvestris L.) to sawn timber properties through simulated sawing. Forest Ecology and Management 174 (2003), p. 251-263.

Jeffrey, W.G., 1979: The relationship between quality and quantity in spruce timber production. Forestry, Vol 52, No. 1, p. 47-68 [6]

Kapp, G.B., Lujan, R., 1997: Species and site selection for timber production on farm boundaries in the humid Atlantic lowlands of Costa Rica and Panama. Agroforestry Systems 35. p. 139-154.

Kellomäkki, S., Ikonen, V.-P., Peltola, H., Kolström, T., 1999: Modelling the structural growth of Scots pine with implications for wood quality. Ecological Modelling 122 (1999), 117-134

Kiraly, G., Brolly, G., 2007: Tree height estimation methods for terrestrial laser scanning in a forest reserve. ISPRS workshop on Laser Scanning 2007 and SilviLaser 2007 (proceedings), Espoo, Finland, p. 211-215

Laasasenaho, J., Kouivuniemi, J., 1990: Dependence of some stand characteristics on stand density. Tree physiology, 7, p. 183-187

Lacointe, A., 1999: Carbon allocation along tree organs: A review of basic processes and representation in funcitonal-structural tree models. Ann. For. Sci. 57 (2000), p. 521-533.

Lippemaier, P., 1977: Beziehungen zwischen Schnittholzqualität und Rundholzqualität bei Fichte. Forstwiss. Cbl. 96 (1977), S. 162-169 [4]

MacGuire, D., Kershaw, J.A. Jr., Hann, D. W., 1991: Predicting the effects of silvicultural regime on branch size and crown wood core in Douglas-fir. Forest Science, Vol. 37, No. 5, p. 1409-1428

Mäkelä, A., Mäkinen, H., 2003: Generating 3D-sawlogs with a process-based growth-model. Forest Ecology and Management 184 (2003), p. 337-354

Mäkinen, H., Colin, F., 1998: Prediciting branch angle and branch diameter of Scots pine from usual tree measurements and stand structural information. Can. J. For. Res. 28, p. 1686-1696.

Mäkinen, H., 1999: Effects of stand density on radial growth of branches of Scots pine in southern and Central Finland. Can J. For. Res. 29: p. 1216-1224

Mäkinen, H., Colin, F., 1999: Predicting the number, death, and self-pruning of branches in Scots pine. Can. J. For. Res., 29, p. 1225-1236 (1999)

Mäkinen, H., Ojansuu, R., Niemistö, P., 2001: Prediciting external branch characterstics of planted silver birch (Betula pendula R.) on the basis of Routine stand and tree measurements. Forest Science 49 (2) 2003, p. 301-317

Mäkinen, H., Song, T, 2002: Evaluating models for branch characteristics of Scots pine in Finland. Forest Ecology and Management 158 (2002), p. 25-39.

Mäkinen, H., Mäkelä, A., 2003: Prediciting basal area of scots pine branches. Forest Ecology and management, 179 (2003), p. 351-362

Mäkinen, H., Ojansuu, R., Sairanen, P., Yli-Kojola, H., 2003: Prediciting branch characteristics of Norway Spruce (Picea abies L. KARST) from simple stand and tree measurements. Forestry Vol. 76, No. 5, 2003, p. 525-546

Meredieu, C., Caraglio, Y., 2002: External indicators of living branches with missing rings within a tree crown of Corsian pine. Forestry Vol. 75 (2002), p. 569-578

Middleton, G.R., Carter, R.E., Munro, B.D., Mackkay, J.F.G., 1989: Losses in timber values associated with distorted growth in immature Douglas-fir. Forestry FRDA Report, 50, 27 p.

Miller, W.E., Kessler, K.J. Jr., Ohman, J.H., Eschle, J.T., 1978: Timber quality of Northern Hardwood regrowth in the Lake States, Forest Sci, Vol. 24, No. 2, 1978, pp. 247-258

Moberg, L., 1999: Variation in knot size of Pinus sylvestris in two initial spacings trials. Silva Fennica 33(2), p. 131-145

Moberg, L., 2001: Models of internal knot properties for Picea abies. Forest Ecology and Management 147 (2001), p. 123-138

Murphy, G., 2009: Determining stand value and log product yields using terrestrial lidar and optimal bucking: a case study, Journal of Forestry, p. 317-324.

Perrson, B., Perrson, A., Stahl, E., Karlmats, U., 1995: Wood quality of Pinus sylvestris at various spacings. Forest Ecology and Management 76 (1995). p. 127-138

Pinto, I., Pereira, H., Usenius, A., 2003: Analysis of log shape and internal knots in twenty Maritime pine (Pinus pinaster Ait.) stems based on visual scanning and computer aided reconstruction. Ann. For. Sci, 60 (2003), p. 137-144

Polman, J.E., Militz, H., 1996: Wood quality of Douglas-fir from three stands in the Netherlands. Ann. Sci. For. 53 (1996), p. 1127-1136

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, URL http://www.R-project.org

Raulier, F., Ung, C.-H., Begin, J., 1998: Analytical estimation of branchwood volume in sugar maple, linked to branchiness. Trees, 1998 (12), p. 395-405

Roberts, JW., Tesfamichael, S., Gebreslasie, M., van Aardt, J., Ahmed, FB, 2007: Forest structural assesment using remote sensing technologies: an overview of the current state of the art. Southern Hermisphere Forestry Journal, 69 (3): 183-203.

Sautter, H.U., 1993: Astigkeitsverhältnisse und Schnittholzqualität in einem weitständig begründeten Fichtenbestand. AFZ 17/ 193, p. 875-878

Schütt, P., Schuck, J., Stimm, B., 1992: Lexikon der Forstbotanik. Ecomed. 581 S.

Seifert, T., 1999: Modelling wood quality of Norway Spruce (Picea abies) depending on silvicultural treatment. Proceedings of the third workshop “Connection between Silviculture and Wood Quality through modeling approaches and simulation software” of the IUFRO working party S5.01.04, La Londe-Les-Maures, France, p. 534-540.

Seifert, T., Pretzsch, H., 2004: Modelling growth and quality of Norway spruce with the growth simulator SILVA. Proceddings of the 4th workshop “Connection between Forest resources and wood quality: Modelling approaches and simulation software”, IUFRO working party S5.01.04, Harrison Hotsprings, Brsitish Columbia, Canada, p. 562-574.

Seifert, S., Seifert, T., 2006: Bodengestüztes Laserscanning zur Erfassung der Nadelbiomasse bei Fichte. Tagungsband der Jahrestagung der Sektion Ertragskunde im Deutschen Verabnd Forstlicher Forschungsanstalten, S. 86-97

Seifert, T, Klemmt, H.-J., Seifert, S., Kunneke, A., Wessels, C.B., 2010: Integrating terrestrial laser scanning based inventory with sawing simulation. In: Proceedings (extended abstracts): Precision Forestry Symposium, March 2010, Stellenbosch, South Africa.p. 12-13

Stahl, E., Karlmats, U., 1995: Yield, wood properties and timber harvest at establishment of seed-tree and shelterwood regeneration systems. Studia Forestalia Suecica, 197 (1995), 15 p.

Struck, Dohrenbusch, 2000: Development of a new branchiness index ASIX – A simple tool to describe branchiness in young decidous forest stands. Ann. For. Sci 57 (2000), 811-818

Teutsch, Ch., 2007: Model-based analysis and evaluation of point sets from optical 3D-Laser Scans. Ph.D. Thesis. Univ. Magdeburg. 153 p.

Thies, M., Pfeifer, N., Winterhalder, D., Gorte B., 2004: Three-Dimensional Reconstruction of stems for Assessment of Taper, Sweep and Lean based on laser scanning data of standing trees. Scand. J. For. Res. 19: 571-581.

Wezyk, P., Koziol, K., Glista, M., Pierzchalski, M., 2007: Terrestrial laser scanning versus

traditional forest inventory first results from the polish forests. In: ”Proceedings of ISPRS

Workshop on laser scanning 2007 and SilviLaser.” Espoo, Finland. 12-14 September, 2007. 424-430.

Witkowska, I., 2008: Spruce quality in mixed stands. M.Sc. Thesis, Swedish University of Agricultural Sciences., No. 104, Souther Swedish Forest Research Centre, 51 p.

Towards an inventory of quality attributes of individual trees using airborne laser scanning and multispectral data

*Johannes Breidenbach *†, Erik Næsset†, Vegard Lien†, Terje Gobakken†, Svein Solberg‡*

†Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences, 1430 Ås, Norway

‡ Norwegian Forest and Landscape Institute, 1430 Ås, Norway

Abstract

The semi-individual tree crown approach (semi-ITC) was used to predict crown base heights (CBH) on the level of single crown segments based on airborne laser scanning (ALS) derived metrics. The root-mean-squared-differences (RMSD) on the segment level were smallest for spruce. However, they were larger than the standard deviation of the measured CBH for pine and birch. The RMSD values were also larger compared to other studies. This can in part be explained by the fact that the semi-ITC approach incorporates errors of the segmentation algorithm. As a consequence, all instead of only correctly identified trees were considered in modeling which results in more realistic RMSD values. After aggregating the individual segment predictions to the plot level, the RMSD values were smaller than the standard deviations of the field measurements and comparable to other studies. The relative RMSD values for birch, spruce, pine and all species were 51.61, 35.22, 49.28, and 13.89%, respectively.

Keywords: Forest inventory, timber quality, individual tree crown segmentation, nonparametric regression

1. Introduction

Over the last few years, many studies have focused on extracting individual tree properties from airborne laser scanning (ALS) data. While paying attention mainly to the tree segmentation algorithms, it has been common practice to just consider “correctly identified” segments in the subsequent analysis of tree properties to be predicted. Thus, segmentation errors inherent in most segmentation algorithms have to a large extent been neglected. The small number of studies attempting to use information from individual tree crown segmentation algorithms (ITC) for forest inventories have generally resulted in biased estimates after aggregation to larger units such as stands. Although forest inventories based on ITC have been offered as a commercial product for some time, it is just recently that statistically sound approaches allowing unbiased predictions using ITC have been described (Breidenbach et al. 2010; Lindberg et al. 2010).

In addition to variables such as timber volume and tree species, information on timber quality is essential for an efficient forest management. Consequently, crown base heights (CBH) are, due to their close relation with the branch distribution along the stem, measured in several forest inventory programs. Various studies using either the area-based (Andersen et al. 2005; Maltamo et al. 2010) or the ITC approach (Næsset and Økland 2002; Popescu and Zhao 2008; Maltamo et al. 2009) have shown the prospects of ALS for predicting the crown base height.

In this study, we apply the ITC approach proposed by Breidenbach et al. (2010) and analyze its ability to predict the crown base height. Because the approach accepts none, one, or several trees to be within a given segment, it is called “semi-ITC”. As a consequence, variance induced by errors in the segmentation of tree crowns is considered inherently which allows for a realistic computation of error statistics.

2. Material and methods

2.1 Data

The study area is located in the municipality of Aurskog-Høland, in southeast Norway. The managed forest under consideration is dominated by Norway spruce (*Picea abies*) and Scotch pine (*Pinus sylvestris*), which together have a proportion (based on timber volume) of 91%. The most common deciduous tree species is birch (*Betula* spp.) with a proportion of 6%.

Field work was carried out in the dormant season of 2007-2008. Species, tree position and diameter at breast height (DBH) were measured for all trees with a DBH >5 cm on circular plots of size 1000 m^2 (36 plots) and 500 m^2 (4 plots). On each plot, crown base height (height to the whorl with living branches in 3 of 4 quadrants) was measured on a sample of 10 and 5 trees of the dominant and co-dominant species, respectively. The CBH of the unmeasured trees were predicted based on species-specific models based on DBH and plot basal area developed from the measured trees. The stem volume of each tree was computed by means of volume equations based on diameter and height as predictor variables. A total of 4002 trees were tallied. A measurement of the CBH was available for 680 trees.

The ALS data were acquired on 12 June 2006 from an altitude of approximately 800 m above ground. The Optech ALTM3100 scanner was operated with a scanning angle of +/- 5° and a beam divergence of 0.3 mrad. The scanner setup resulted in a point density of approximately 7.4 m^{-2}. In addition to elevation measurements, intensity values were recorded for every echo.

Aerial images were acquired on 29 June 2005 with a Vexcel UltraCam D digital camera from 3100 m above ground. The sensor resolution was 7500x11500 pixels in the panchromatic and 2400x3680 pixels in the multispectral bands. The near infrared (NIR), red (R), and green (G) bands were available for analysis.

2.2 Methods

The semi-ITC approach can roughly be subdivided into four steps: i) data preparation, ii) delineation of crown segments and computation of explanatory variables, iii) data analysis and statistical modeling, and iv) imputation of nearest neighbors and upscaling to larger areas such as plots or stands. See Breidenbach et al. (2010) for further details.

The data preparation, i.e. step i), includes derivation of digital elevation models, normalization of ALS elevations to derive heights above ground, and fusion of ALS and optical data. Once the basic data were available, crown segments were delineated (Solberg et al. 2006) from the canopy height model as part of step ii). A total of 2283 segments had their centroides within a sample plot. The segments were then characterized by their areas as well as metrics (Næsset 2002) derived from the ALS height. Mean values per segment were derived for each segment from the ALS intensity and multispectral data. Subsequently, the segments were intersected with the trees measured on the ground to combine remote sensing and field data. For segments containing more than one field measured tree, single tree volumes were summed up by species and the mean crown base height was determined.

The same remote sensing derived metrics as selected by Breidenbach et al. (2010) comprised the predictor variables that were used to model the proximity between crown segments in step iii). The predictor variables used were metrics such as mean and maximum height, segment area and canopy density measures. Using mean CBH as well as total and tree species-specific timber volumes per segment as responses, a most similar neighborhood (MSN) (Moeur and Stage 1995) model was fitted. This statistical model basically allows determining the similarity of crown segments in terms of their covariates. Consequently, response variables from nearest neighboring segments with observed ground truth values (references) were used for predictions of segments without ground truth values (targets). At this point, it seems worth noting that segments without measured trees belonged to the references to avoid overestimates caused by oversegmentation errors.

In step iv), a cross validation procedure was applied in which all segments from one sample plot were left out at a time. The segments with cross-validated predictions were then joined with the other segments that were not

used for modeling but had their centroid within a sample plot. Subsequently, the predicted response variables of the segments within a sample plot were aggregated on plot level to derive tree species-specific CBH. The root-mean-squared-difference (RMSD) as the square root of the difference between observed and predicted values was used to assess the goodness-of-fit of the models. See Breidenbach et al. (2010) for further details on the data and methods.

3. Results and discussion

The comparison of predicted and observed CBH at the segment level revealed that a considerable amount of variation remained unexplained (Table 7, Figure 4 A). The CBH prediction with ALS was best for spruce. The RMSD values of birch and pine were even larger than the standard deviation. This indicates that a null-model (i.e., the mean measured CBH) is better suited for a prediction of CHB for these tree species on the segment level than the model based on ALS derived variables. A less accurate prediction of CBH for deciduous trees based on ALS metrics was also observed by Popescu and Zhao (2008). However, their prediction of CBH for American pine species was very accurate.

The reported RMSD values are larger compared to other studies carried out on the single tree level (cf. Popescu and Zhao (2008) and references therein). This may in part be explained by the fact that in our study segmentation errors were considered in the prediction instead of reporting predictions of the correctly identified trees only. The correctly identified trees are in tendency the large trees. Consequently, considering only them for analyses may result in overoptimistic results.

Upscaling the segment predictions to the plot level resulted in all RMSD values of the CBH being smaller than the standard deviation of the field data (Table 8, Figure 4 B). This positive effect of aggregating predictions of segments to the plot level was also observed for other response variables (Breidenbach et al. 2010). While tree species-specific results are often not reported, the RMSD of CBH for all species is slightly better than was found in other studies (see Maltamo et al. (2010) and references therein). Nonetheless, it should be noted that since the response variable (CBH) was not measured but predicted for many of the trees based on DBH and plot basal area, the actual variance was most likely somewhat underestimated.

While the results are promising, considering other predictor variables and tree species-specific CBH as response variables during modeling may further improve the prediction accuracy.

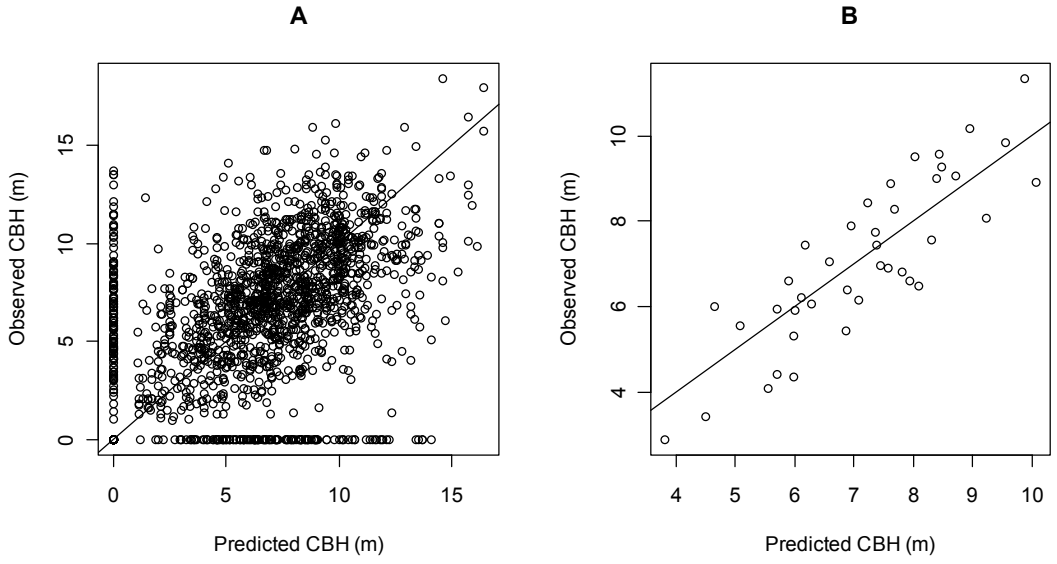


Figure 4: Observed vs. predicted crown base height (CBH) on the segment level (A) and aggregated to plot level (B).

Table 7: Comparison of observed and predicted crown base heights at the segment level.

	Mean of field measurements (m)*	Standard deviation of field measure-ments (m)*	Coefficient of variation of field measure-ments (%)	RMSD (m)	RMSD (%)
Birch	0.86	2.61	302.28	3.32	384.89
Spruce	3.07	3.91	127.21	3.33	108.43
Pine	3.52	4.68	132.77	4.80	136.22
All species	6.31	3.81	60.45	3.75	59.41

* Empty segments were considered as zero CBH.

Table 8: Comparison of observed and predicted crown base heights aggregated to plot level.

	Mean of field measurements (m)	Standard deviation of field measure-ments (m)	Coefficient of variation of field measure-ments (%)	RMSD (m)	RMSD (%)
Birch	6.46	3.96	61.29	3.33	51.61
Spruce	4.64	2.94	63.26	1.63	35.22
Pine	8.74	4.40	50.30	4.31	49.28
All species	7.10	1.92	26.99	0.99	13.89

References

Andersen, H. E., McGaughey, R. J. and Reutebuch, S. E. (2005). Estimating forest canopy fuel parameters using LIDAR data. Remote Sensing of Environment, 94(4), 441-449.

Breidenbach, J., Næsset, E., Lien, V., Gobakken, T. and Solberg, S. (2010). Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. Remote Sensing of Environment, 114(4), 911-924.

Lindberg, E., Holmgren, J., Olofsson, K., Olsson, H. and Wallerman, J. (2010). Estimation of tree lists from airborne laser scanning by combining single-tree and area-based methods. International Journal of Remote Sensing, 31(5), 1175-1192.

Maltamo, M., Bollandsås, O., Vauhkonen, J., Breidenbach, J., Gobakken, T. and Næsset, E. (2010). Comparing different methods for prediction of mean crown height in Norway spruce stands using airborne laser scanner data. Forestry, (Published online).

Maltamo, M., Peuhkurinen, J., Malinen, J., Vauhkonen, J., Packalén, P. and Tokola, T. (2009). Predicting Tree Attributes and Quality Characteristics of Scots Pine Using Airborne Laser Scanning Data. Silva Fennica, 43(3), 507-521.

Moeur, M. and Stage, A. R. (1995). Most similar neighbor: An improved sampling inference procedure for natural resource planning. Forest Science, 41(2), 337-359.

Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sensing of Environment, 80(1), 88- 99.

Næsset, E. and Økland, T. (2002). Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. Remote Sensing of Environment, 79(1), 105-115.

Popescu, S. and Zhao, K. (2008). A voxel-based lidar method for estimating crown base height for deciduous and pine trees. Remote Sensing of Environment, 112(3), 767-781.

Solberg, S., Næsset, E. and Bollandsås, O. M. (2006). Single tree segmentation using airborne laser scanner data in a structurally heterogeneous spruce forest. Photogrammetric engineering and remote sensing, 72(12), 1369-1378.

Consequences of stand structure on timber quality

J.C. SUAREZ†, B.A.GARDINER†, M. DI LUCA‡, J. GOUDIE‡ and K. POLSSON‡*

juan.suarez@forestry.gsi.gov.uk

† Forest Research the Agency of the Forestry Commission, Northern Research Station, Roslin, Midlothian, EH25 9SY, UK

‡Ministry of Forests, Victoria, B.C. Canada

Abstract

Sitka spruce (Picea sitchensis (Bong.) Carr.) represents 49.2% of all conifer species planted in Britain. Initial site conditions and management practices have profound impacts on the final quality of home-grown timber. To measure the impact of different silviculture strategies, a timber quality model for Sitka spruce is being developed at Forest Research. The Timber Quality model for Conifers (ConTQ) focuses on compression wood, stem straightness, wood density, spiral grain angle, branchiness and knot density.

ConTQ was tested with airborne LiDAR data taken in 19 50m x 50 m plots in two study areas in Britain, managed at different thinning intensities. Canopy Height Models generated from LiDAR were analysed in Definiens professional, where a ruleset for delineating individual tree canopies was constructed. This procedure creates tree lists containing parameters such as tree height, dbh, canopy length, canopy width and the spatial location of each individual. Those lists were entered into the Canadian TASS model for running realistic scenarios of growth over 20 years.

The analysis of the time-series scenarios generated from TASS using baseline cartography from LiDAR shows a power law relationship between Dominance status, defined by a ratio between active and potential foliar volume for each tree, and two timber quality parameters: stem straightness and wood density. These relationships are dynamic and change in time as a result of tree growth and management practices.

The results show volume increment associated to age and Site Index (SI) to be inversely correlated to Wood Density. On the contrary, Stem Straightness is directly correlated to volume increment and its effect is lowered by wind exposure. Spacing has a detrimental effect on straightness and wood density. The use of Dominance helps to understand the spatial variability of the timber quality parameters being analysed and its likely progress in time, especially after thinning operations or the impact of windthrow.

1. Introduction

Forest cover in Britain extends over 2.8 million hectares, representing about 12% of the total land area of the country (Forestry Commission, 2007). Most of the plantation effort was undertaken in the uplands between 1950 and 1990, characterised by a rapid process of afforestation in areas of difficult terrain, aided by improvements in cultivation techniques and machinery.

The majority of the species being planted were conifers. Although, Sitka spruce (Picea sitchensis (Bong.) Carr.) alone covered 730,000 hectares or 49.2% of all conifer species in 2007. Three quarters of those plantations happened in the uplands of Scotland and Northern England (Forestry Commission, 2007).

As most plantations from the 1960s are coming into production now, some concern has grown about the characteristics and properties of the timber being harvested. According to Macdonald and Hubert (2002), the general trend is for increasing knot sizes, reduced wood densities, increased juvenal core size and higher percentages of compression wood. These problems are associated to a progressive widening of initial spacing over the years and wind exposure.

Therefore, the primary aims of this work will be to determine the capability of Airborne Laser Scanning (ALS) for mapping the allometric characteristics and the spatial location of individual trees growing on a commercial stand. Then, a distance-dependent model (TASS) will be used to determine the dynamic components that determine tree growth based on competition between individuals and the ultimate establishment of a Dominance hierarchy. In this process, each Dominance type will be defined by its effective Foliar Volume (FV). This factor will be responsible for producing photosynthates in enough supply to sustain annual bole increment and configure the allometric characteristics of the stems. Finally, the Timber Quality model for Conifers (ConTQ) will be used to estimate the changes in stem properties, such as Wood Density and Stem Straightness, derived from the process of consolidation of the Dominance hierarchy in forest stands.

A pending question relates to the capability of trees to adapt their physical stem properties in time in response to disturbances or man-made interventions. Therefore, TASS will be used to simulate dynamic responses of the stand community to structural changes during a 20-year simulation period. By growing trees in time and monitoring their changes in Dominance status, it will be possible to detect processes of adaptation in the quality parameters of the individual stems.

Material

Study areas

Two monitoring areas were selected in Aberfoyle (SW Scotland) at 56º10' N, 4º22' W and Kielder Forest (North of England) at 55º 14` N 2º 35` W. Mensuration data were taken for 12 thinned plots at various levels of intensity in Aberfoyle and 5 unthinned plots in Kielder Forest. All plots were planted with Sitka spruce.

Plot dimensions were 50 m by 50 m. The work was undertaken between 2002 and 2004, one month before and after the ALS surveys took place to avoid data decorrelation. Measurements consisted of diameter at breast height (DBH at 1.3 m) and the position of each tree (located with a Total Station and DGPS). Height and Canopy dimensions were taken only for a selection of trees with a Sonic Vertex III (Haglöf, Sweden).

ALS data.

ALS was acquired in September 2002 (Aberfoyle) over an area of 17.5 Km^2. Kielder was surveyed in 2004, covering an area of 20 Km^2. Both surveys were undertaken by the English Environment Agency using an Optech ALTM2033 scanner. Sampling density varied from 3-4 to 8-10 returns per m^2 (Table 1).

Table 9. Characteristics of the ALS surveys in the study areas

Parameter	Aberfoyle 2002	Kielder 2004
Sensor	Optech ALTM2033	Optech ALTM2033
Date	19 September 2002	25 April 2004
Laser pulse frequency	33,000 Hz	33,000 Hz
Flying altitude	1000 m	1000 m
Beam divergence	0.3 mrad	0.3 mrad
Scanning angle	20 degrees	10 degrees
Sampling intensity	3-4 returns per m^2	8-10 returns per m^2
Position accuracy	X,Y < 40 cm	X,Y < 40 cm
Elevation accuracy	Z < 9- 15 cm	Z < 9- 15 cm

Data format	First, last returns and intensity (8-bit)	First, last returns and intensity (8-bit)

The cloud points generated by ALS were processed in Terrascan 007.008 Software (Terrasolid Ltd, Finland) for filtering ground hits.

Once the ground hits were selected, they were exported to Surfer 8.0 (Golden Software, USA) to be gridded into a Digital Terrain Model (DTM) using a kriging interpolation method with a linear model without anisotropy at a grid resolution of 0.5 m. First return points were also gridded using this method to create a Digital Surface Model (DSM).

Finally, the subtraction of DSM and DTM produced a normalised Canopy Height Model (CHM) for each stand.

Tree analysis

A new canopy delineation algorithm was designed to identify the location and properties of each individual tree. This method used the CHM generated in the previous steps as an input data for a routine written for Definiens Developer 7.0 (Definiens Imaging, Germany). The characteristics of this canopy delineation algorithm can be found in (Suárez et al., 2009).

The algorithm located tree tops and delineated individual crowns inside each plot. Three plots were randomly selected in each study area for training allometric functions. The selection of trees was done manually by linking detected individuals with measured trees in the field. Each detected tree canopy contained information about maximum pixel height (i.e. the height of the tree) and canopy area. These two parameters were used to construct a ALS-based model showing the true Height and the DBH for each individual tree. The models were run in SAS (SAS Institute Inc. USA) to look at all possible combinations and to select those models statistically more significant. Table 2 shows the models for Height and DBH in the two study areas.

Table 2. Height and DBH recovery models from ALS. H is the LiDAR Height and A is the Crown Area.

Year	model	R^2	Equation	P-value Intercept	P-value first param.	P-value second param.
Aberfoyle	Height	0.96	$5.359+0.653*H+0.008*H^2$	0.06	0.009	0.148
n=80	DBH	0.88	11.728+1.709*H+0.458*A	1.8E-06	1.17E-26	8.21E-17
Kielder	Height	0.99	0.462+1.009*H	0.1943	<0.0001	N/A
n=39	DBH	0.79	-3.384+1.306*H+0.208*A	0.315	<0.0001	0.0034

The results provided a level of detection of individual trees of 85% in Aberfoyle and 91% in Kielder. Height was accurately predicted in both areas (c.100%), whereas DBH was overpredicted by 10%.

2. Methods

The ConTQ model

ConTQ is a model developed at Forest Research used to simulate wood properties in forest stands (Gardiner et al, 2005). It focuses mainly on Sitka spruce, as this species remains the main crop in the country. The model takes yield class, mean DBH, Top Height and thinning regime as inputs. The use of yield models allows ConTQ to simulate scenarios of possible wood property changes that may be introduced by site conditions and management regime (i.e. thinning intensity) up to the age of the simulation.

The tree height and diameter estimates derived from the yield models are used as inputs into a taper function that specifies an annual average stem profile for the mean tree in the stand. A tree density profile is then derived based on calculated ring widths from the taper model and ring number (from the pith) for all locations along the tree stem. The diameters along the stem profile are calculated using the Achim et al., (2006) taper model as a function of height above ground, tree height and age:

$$d = \theta_5 DBH\left(1 - z/Ht\right)^{(\theta_1+\theta_2(1-z/Ht))}.\left(1+\theta_3 e^{-\theta_4(z/Ht)}\right) \text{ (cm)} \quad (1)$$

where:

$$\theta_1 = 0.9448 - 0.00246\left(\frac{Ht}{dbh}\right)$$

$$\theta_2 = -0.356 - 0.00838\,Age$$

$$\theta_3 = 0.3422$$

$$\theta_4 = 3.0361 + 0.1892\,Age$$

$$\theta_5 = 0.8478$$

d = stem diameter over bark at height z above the ground (cm)

z = height along the tree stem above the ground (m)

Ht = total tree Height (m)

The Specific Gravity (SG) model is used for estimating wood density at 0% moisture content. The model has been derived from the analysis of discs taken at different test areas in Scotland using a CT-scanner. The general formulation of the model is a function of Ring Number (RN) and Ring Width (RW):

$$SG = 0.4322 \cdot (1 + 0.5829 \cdot e^{-(RN/5.3193)}) \cdot (1 + 0.0020 \cdot RN - 0.0472 \cdot RW) \quad (2)$$

The SG values are then multiplied by 1,000 to generate wood density values in kg m^{-3}.

The Stem Straightness (SS) model is empirically derived as a function of DBH and DAMS. DAMS is a scoring system and it is used within the ForestGALES model to measure local windiness. The DAMS scores have been calculated for the whole country in 50 m x 50 m grid intervals (Bell et al., 1995). The formulation of Stem Straightness in ConTQ is:

$$SS = 0.855 - 0.0398 \cdot DAMS + 0.1147 \cdot DBH \quad (3)$$

Running the ConTQ model with tree lists generated by TASS

The tree lists generated by ALS were entered into the TASS model (Mitchell, 1975) in order to run a 20 year growth simulation (Figure 1). The scenarios produced by the model were used to estimate annual increments on Height, DBH, Canopy Length, Canopy Width, Foliar Volume (FV) and Maximum Foliar Volume (FV_{max}). Dominance status was calculated for each tree as $-ln(FV/FV_{max})$ for each year.

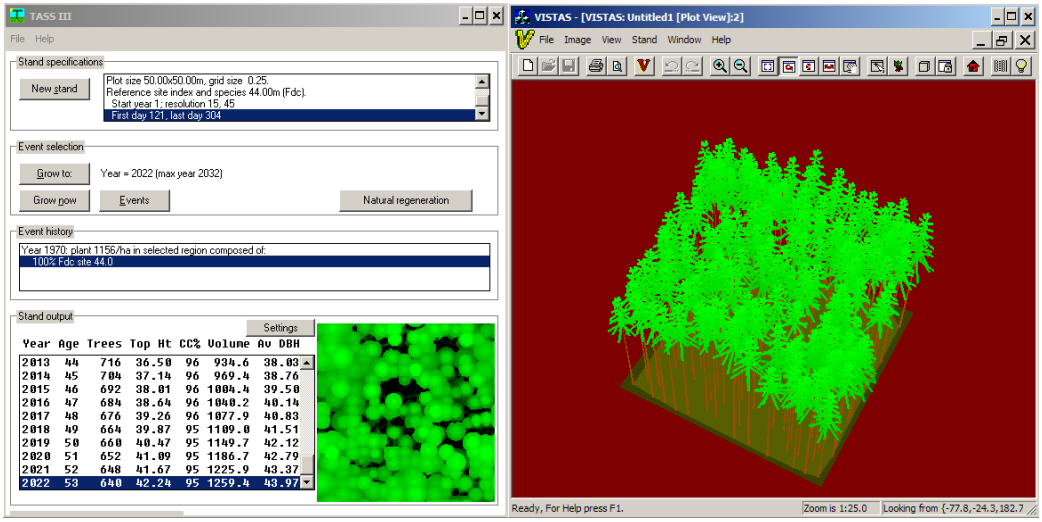


Figure 1. Growth simulations with TASS using tree lists generated by LiDAR.

The tree lists generated from TASS were input into the ConTQ model, containing information about each individual tree: Tree ID (integer), Eastings and Northings (m), Tree Height (m), DBH (m), Yield Class (m^3 $year^{-1}$ ha^{-1}), Canopy Length (m), Canopy Width (m) and Canopy Area (m^2).

The Wood Density and Stem Straightness values were then calculated for each tree in a plot at the year of reference (2002 in Aberfoyle and 2004 in Kielder Forest) and at 5-year intervals afterwards.

Top Height and Mean DBH were used in ConTQ to grow each individual tree backwards in time. Top Height and mean DBH at the year of reference were used to create the average tree dimensions in each plot. Then, the individual tree values were divided by the mean tree. The ratio of each individual tree in relation to a mean tree allowed the program to calculate the annual ring increments from the planting year using the taper model described in Eq. 1. Ring Number (RN) and Ring Width (RW) were derived from Age and individual DBH and Height value increments each year.

SG was calculated for each ring using Eq. 2. As Moisture Content (MC) was 0%, the SG had to be corrected for Mass and Volume increments at 12% MC (average moisture content in commercial timber products). This was undertaken using Simpson (1993):

$$SG_B = \frac{SG_M}{\left(1+\frac{MC}{100}\right)} \Bigg/ \left(1+0.265\cdot\left(\frac{30-MC}{30}\right)\cdot\frac{SG_M}{\left(1+\frac{MC}{100}\right)}\right) \quad (4)$$

Where, SG_B = Basic Specific Gravity (at MC 12%),

SG_M = Specific Gravity of wood being measured,

MC = Moisture Contents (in this case 12%)

The SG_B values were converted into wood density in kg m^{-3} by multiplying it by 1,000. The estimated Wood Density was presented in the model as the average value for each ring. Figure 2 and Table 3 depict an example of the estimation of Wood Density along the stems of different Dominance types.

Dominant trees were characterised by rapid annual volume increments. Average density for the entire stem was low compared to other classes but standard deviation was larger. The main source of variability was the juvenal core (depicted in red in Fig. 2 close to the pith at radius = 0).

Suppressed and meta-suppressed were characterised by a slower Volume increment than dominants. This created narrow growth rings, where early wood was not well developed and late wood became proportionally more important. This situation produced higher Wood Density values. Table 3 displays larger average density values for the suppressed and meta-suppressed in comparison with the rest of the trees. In addition, the juvenal core seemed to be smaller in the lower ranks.

Stem Straightness values were negatively correlated to windiness and positively to DBH (Eq. 3). Trees growing in the most exposed sites developed poorer stem forms. Likewise, smaller trees develop poorer forms than other with larger diameters. The connection between DBH and Stem Straightness in Sitka spruce suggested two possible mechanisms: leader losses and competition for light.

Table 3. Example of Wood Density values estimated for five Dominance classes in Aberfoyle Plot 1 (kg m^{-3}).

Dominance	$-\text{Ln}(FV/FV_{max})$	mean	sdev	max	min
Dominant	0.87	444.11	136.63	781.78	316.40
Co-Dominant	1.51	446.91	101.87	781.78	358.45
Subdominant	2.49	468.61	64.76	781.78	402.22
Suppressed	3.50	474.41	62.08	781.78	403.54
Meta-Suppressed	10.44	521.87	57.52	781.78	458.58

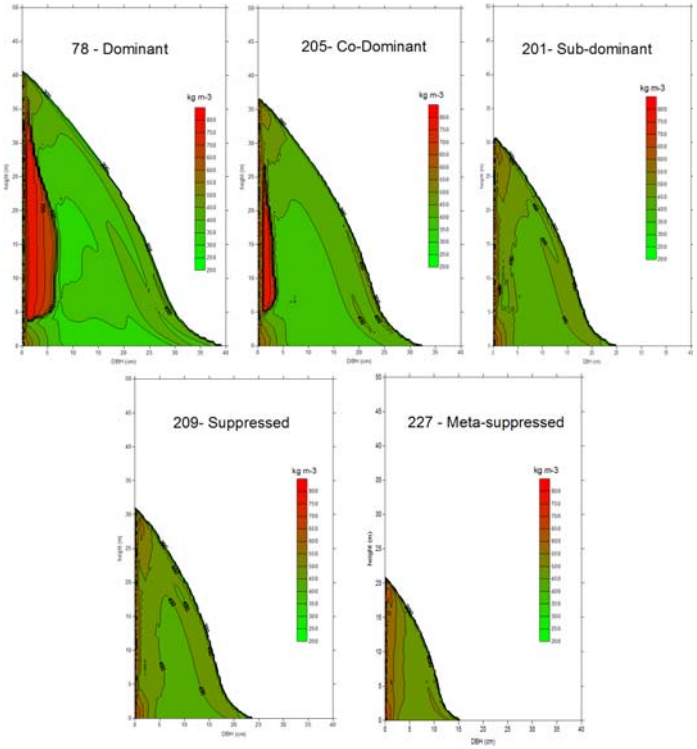


Figure 2. Example of Wood Density at different parts of the stem estimated by ConTQ for five different Dominance classes (-Ln(FV/FVmax)). The X-axis represents DBH (cm) and the Y-axis is Height (m).

3. Results

Aberfoyle

Wood Density and Stem Straightness were calculated for each tree in the 12 plots in Aberfoyle for the simulation period after the reference date. The results were analysed at five yearly intervals (year 5, 10, 15 and 20). Power Law models were fitted to relate Dominance status and timber quality properties (Table 4 and 5).

Table 4. Power Law models of Wood Density against Dominance. The Power Law equations are shown in the left columns from year 5 until year 20 (end of the simulation). The right columns show the coefficients of determination (R^2) for each model.

Density	YR5	YR10	YR15	YR20	R^2	YR5	YR10	YR15	YR20
Plot1	$380.30x^{0.106}$	$391.19x^{0.0954}$	$396.92x^{0.0928}$	$403.26x^{0.0937}$	Plot1	0.62	0.69	0.68	0.75
Plot2	$399.49x^{0.1643}$	$410.76x^{0.1074}$	$415.59x^{0.085}$	$419.56x^{0.0793}$	Plot2	0.42	0.54	0.52	0.54
Plot3	$384.74x^{0.1288}$	$396.26x^{0.1103}$	$403.08x^{0.1018}$	$409.41x^{0.1027}$	Plot3	0.50	0.61	0.65	0.67
Plot4	$392.89x^{0.0886}$	$402.60x^{0.0806}$	$408.41x^{0.0788}$	$413.00x^{0.0838}$	Plot4	0.60	0.71	0.71	0.73
Plot5	$406.99x^{0.1133}$	$413.84x^{0.1022}$	$418.53x^{0.0936}$	$421.87x^{0.0941}$	Plot5	0.63	0.72	0.74	0.73
Plot6	$408.62x^{0.1387}$	$418.79x^{0.0975}$	$423.59x^{0.0884}$	$428.50x^{0.0782}$	Plot6	0.56	0.60	0.62	0.62
Plot7	$387.30x^{0.093}$	$392.76x^{0.0858}$	$399.00x^{0.085}$	$403.43x^{0.088}$	Plot7	0.68	0.76	0.74	0.76
Plot8	$376.75x^{0.0744}$	$381.19x^{0.087}$	$385.78x^{0.0938}$	$392.90x^{0.0927}$	Plot8	0.68	0.80	0.77	0.76
Plot9	$388.92x^{0.0839}$	$393.03x^{0.0853}$	$399.27x^{0.0835}$	$404.28x^{0.084}$	Plot9	0.58	0.69	0.75	0.76
Plot10	$389.42x^{0.1117}$	$399.14x^{0.1014}$	$405.56x^{0.0956}$	$410.45x^{0.0962}$	Plot10	0.72	0.77	0.75	0.75
Plot11	$410.69x^{0.0791}$	$416.51x^{0.0743}$	$420.40x^{0.0768}$	$422.82x^{0.0827}$	Plot11	0.39	0.55	0.58	0.69
Plot12	$407.87x^{0.0957}$	$416.70x^{0.0785}$	$420.47x^{0.0758}$	$426.07x^{0.0735}$	Plot12	0.48	0.51	0.50	0.53

Table 5. Power Law models of Stem Straightness classes against Dominance. The Power Law equations are shown in the left columns from year 5 until year 20 (end of the simulation). The right columns show the coefficients of determination (R^2) for each model.

Straightness	YR5	YR10	YR15	YR20	R^2	YR5	YR10	YR15	YR20
Plot1	$6.28x^{-0.3642}$	$6.64x^{-0.321}$	$7.03x^{-0.2826}$	$7.23x^{-0.2414}$	Plot1	0.65	0.72	0.75	0.69
Plot2	$5.03x^{-0.5022}$	$5.33x^{-0.3513}$	$5.90x^{-0.2884}$	$6.41x^{-0.2589}$	Plot2	0.40	0.55	0.54	0.57
Plot3	$5.16x^{-0.2886}$	$5.69x^{-0.2964}$	$6.33x^{-0.3154}$	$6.80x^{-0.3019}$	Plot3	0.50	0.64	0.72	0.69
Plot4	$5.69x^{-0.315}$	$6.06x^{-0.2839}$	$6.61x^{-0.2932}$	$6.94x^{-0.2665}$	Plot4	0.60	0.71	0.75	0.72
Plot5	$4.99x^{-0.3336}$	$5.50x^{-0.3258}$	$6.11x^{-0.3321}$	$6.64x^{-0.3349}$	Plot5	0.65	0.75	0.77	0.75
Plot6	$5.06x^{-0.5201}$	$5.35x^{-0.3846}$	$5.92x^{-0.3475}$	$6.45x^{-0.3277}$	Plot6	0.60	0.63	0.67	0.68
Plot7	$5.75x^{-0.3369}$	$6.18x^{-0.2954}$	$6.75x^{-0.2813}$	$7.16x^{-0.2753}$	Plot7	0.67	0.76	0.74	0.75
Plot8	$6.66x^{-0.3199}$	$7.03x^{-0.2909}$	$7.41x^{-0.2701}$	$7.51x^{-0.2218}$	Plot8	0.80	0.81	0.75	0.67
Plot9	$5.59x^{-0.3375}$	$6.07x^{-0.3137}$	$6.61x^{-0.2928}$	$7.02x^{-0.2775}$	Plot9	0.62	0.70	0.76	0.77
Plot10	$5.38x^{-0.342}$	$5.84x^{-0.3325}$	$6.43x^{-0.3337}$	$6.93x^{-0.3379}$	Plot10	0.71	0.77	0.75	0.76
Plot11	$4.64x^{-0.2564}$	$5.14x^{-0.2587}$	$5.75x^{-0.2776}$	$6.33x^{-0.2991}$	Plot11	0.40	0.58	0.63	0.70
Plot12	$5.11x^{-0.328}$	$5.48x^{-0.2782}$	$6.06x^{-0.2872}$	$6.55x^{-0.2769}$	Plot12	0.50	0.52	0.54	0.53

The intercepts (x = 1) in the Power Law equations in Table 4 showed a progressive increase of wood density across the simulation period. Values ranged between 380 and 410 at year 5 and between 392 and 426 at year 20. The value of the intercepts was inversely related to site conditions. So, plots with higher Site Index (SI) and therefore higher volume increment presented lower Wood Density values (Fig. 3).

The effect of age was also evident on the intercept values. As stands grew older, Wood Density increased, which shifted the intercept values up (Fig. 3). In terms of forest management, the consequences of these results meant that longer retentions, in the absence of re-spacing interventions or natural disturbances, should produce a general improvement of the density of the timber being grown. In this particular case, the general outcome, if allowing these plots to grow beyond 50 years old, would be an increment in Wood Density above the average in Britain, with considerable gains in the stiffness and the strength of the lumber material.

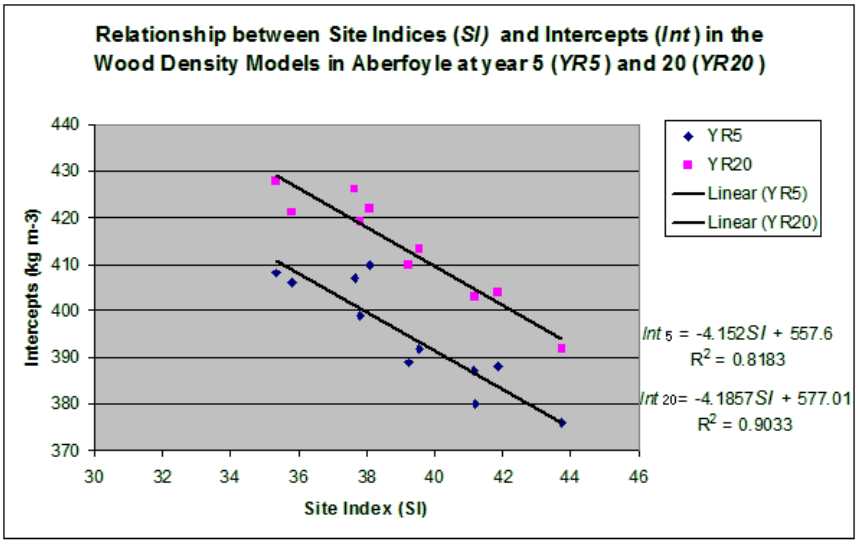


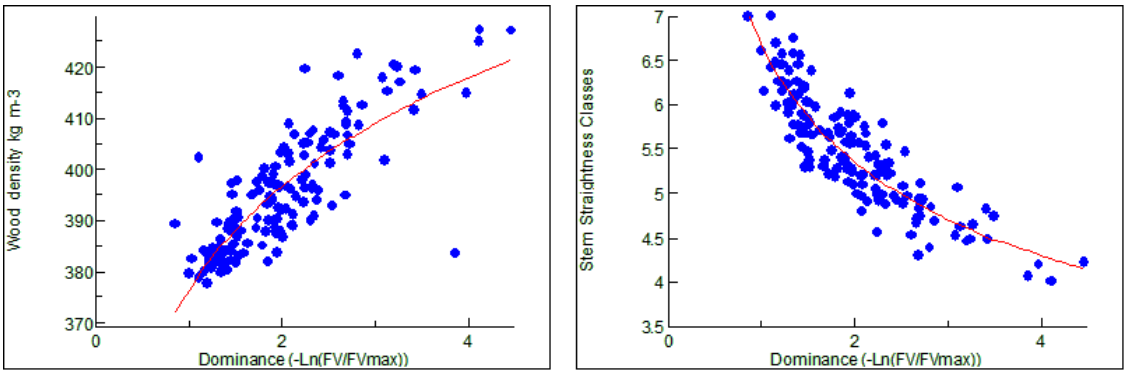
Figure 3. Relationship between the intercepts (Int_{AGE}) in the Wood Density Power Law models related to Site Index (SI). The two lines show a linear relationship between these two factors at age 5 (YR5) and 20 (YR20).

The exponent values were all positive, showing an unequal response between Dominance types and Wood Density (Fig. 4). All plots, apart from Plot 2 and 6, showed similar exponents during the simulation periods (values between 0.08 and 0.10).

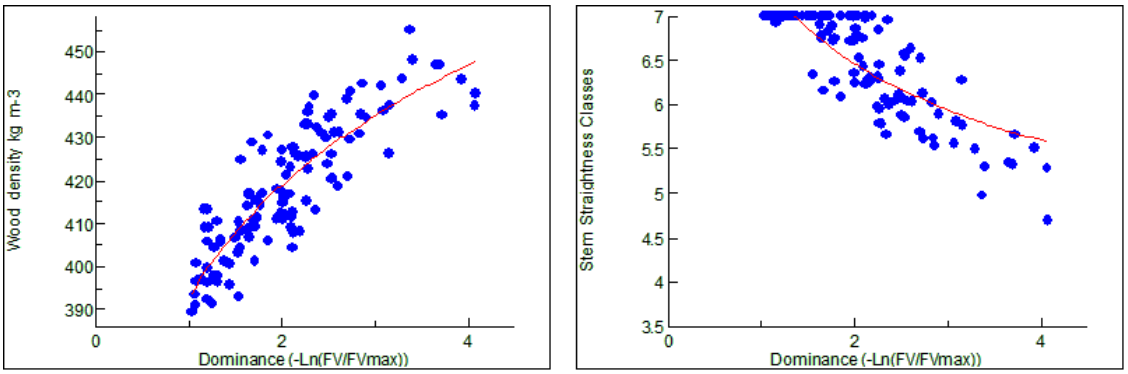
In terms of Stem Straightness, the intercept values in the plots ranged between 4 and 7 (maximum value in the Straightness scoring system). According to Eq. 3, DBH increment improved the straightness scores of the logs. So, the longer the stand kept growing, the larger the number of logs found in the higher Straightness classes. This effect was enhanced by the process of self-thinning due to competition, which produced the progressive elimination of the lower Dominance classes, normally associated to poorer stem forms. The effect of DBH increase was compensated in Eq. 3 by the use of DAMS as a measure of windiness. The windiest places normally presented a higher proportion of bent stems as noticed by Macdonald et al. (2009).

The coefficients of determination presented different patterns in the two timber quality parameters being analysed. In the estimations of Wood Density, R^2 increased from year 5 to year 10 in all plots and then became stable afterwards with little variation. This was mainly attributed to a process reconstruction of the Dominance structure after thinning. The coefficients proved that there was a clear relationship between Dominance types and Wood Density, which improved in time as the Dominance structure was becoming more stable.

The estimation of Stem Straightness presented a different process in terms of R^2. Between year 5 and 10 the coefficients experience an increase. However, once they reached a maximum level they started to decrease again. The reason for this apparent decorrelation is illustrated in the example in Plot 8 (Fig. 4). Between year 5 and 20, the R^2 dropped from 0.80 to 0.67. As the scoring system used for stem straightness only envisages classes from 1 to 7, as soon as any tree reached this level the value could not go any further.



Year 5



Year 20

Figure 4. Example of Wood Density (left column) and Stem Straightness (right column) estimations with ConTQ in Plot 8 across the simulation period. The X-axis depict the Dominance types being from 0 the most dominant to 4 for suppressed trees. The Power Law fits are depicted as a red line.

Kielder

Wood density and Stem Straightness were calculated for the five unthinned plots in Kielder Forest. The procedures followed the same methods used in Aberfoyle.

The results in Kielder followed the same patterns already described in Aberfoyle with slight differences (Tables 6 and 7). The intercepts in the Power Law models for Wood Density were substantially higher than in Aberfoyle. Site indices were lower in this area, which affected the trends for annual Volume Increment below the ones observed in Aberfoyle. Therefore, the slower growing potential in Kielder created the conditions for higher Wood Density values (Fig. 5). The average requirements for passing the lowest grading test (C16), according to EU-Standards, are 370 kg m^{-3}, with an absolute minimum value of 310 kg m-3 (CEN, 2003). According to the results in Fig. 6.13, all the plots in Kielder passed the C24 (optimal) test, which requires an average of 420 kg m^{-3} with an absolute minimum of 350 kg m^{-3}. In contrast, only half of the plots in Aberfoyle approached the C24 grading at year 20 and none of them at year 5. Therefore, the effect of SI on Wood Density properties seemed very important.

Table 6. Power Law models of Wood Density against Dominance in Kielder. The Power Law equations are shown in the left columns from year 5 until year 20 (end of the simulation). The right columns show the coefficients of determination (R^2) for each model.

Density	YR5	YR10	YR15	YR20	R^2	YR5	YR10	YR15	YR20
Plot8	$435.04x^{0.082}$	$451.44x^{0.057}$	$453.13x^{0.065}$	$457.95x^{0.063}$	Plot8	0.64	0.70	0.65	0.64
Plot9	$435.89x^{0.079}$	$449.15x^{0.060}$	$452.04x^{0.063}$	$456.99x^{0.061}$	Plot9	0.54	0.68	0.63	0.65
Plot10	$435.94x^{0.079}$	$445.61x^{0.065}$	$448.49x^{0.066}$	$450.71x^{0.070}$	Plot10	0.64	0.65	0.69	0.73
Plot11	$427.03x^{0.081}$	$439.69x^{0.068}$	$447.65x^{0.064}$	$454.61x^{0.057}$	Plot11	0.67	0.66	0.65	0.59
Plot12	$416.78x^{0.082}$	$429.53x^{0.069}$	$438.23x^{0.063}$	$444.88x^{0.061}$	Plot12	0.73	0.71	0.70	0.73

Table 7. Power Law models of Stem Straightness against Dominance in Kielder. The Power Law equations are shown in the left columns from year 5 until year 20 (end of the simulation). The right columns show the coefficients of determination (R2) for each model.

Straight.	YR5	YR10	YR15	YR20	R^2	YR5	YR10	YR15	YR20
Plot8	$4.489x^{-0.312}$	$4.479x^{-0.254}$	$4.918x^{-0.299}$	$5.251x^{-0.312}$	Plot8	0.62	0.69	0.63	0.63
Plot9	$4.432x^{-0.323}$	$4.512x^{-0.274}$	$4.887x^{-0.292}$	$5.211x^{-0.297}$	Plot9	0.52	0.66	0.62	0.63
Plot10	$4.019x^{-0.286}$	$4.173x^{-0.248}$	$4.592x^{-0.276}$	$5.072x^{-0.312}$	Plot10	0.62	0.62	0.67	0.72
Plot11	$4.633x^{-0.298}$	$4.721x^{-0.273}$	$4.987x^{-0.282}$	$5.272x^{-0.286}$	Plot11	0.66	0.65	0.66	0.64
Plot12	$5.141x^{-0.329}$	$5.198x^{-0.288}$	$5.433x^{-0.280}$	$5.765x^{-0.293}$	Plot12	0.73	0.71	0.69	0.72

Age also acted as a factor contributing to the progressive increase in Wood Density. In general, the plots in Kielder increased 15-35 kg m-3 between year 5 and 20. The effect of spacing on density increments was more important in this study area than in Aberfoyle. As these plots remained unthinned and with average spacings around 2 m, this created the conditions for higher Wood Density increments through constraints on canopy expansion due to close competition. On the contrary Aberfoyle, with average spacings of 5 m and similar planting age, presented smaller Wood Density increments in comparison over the same simulation period (between 10 and 20 kg m-3).

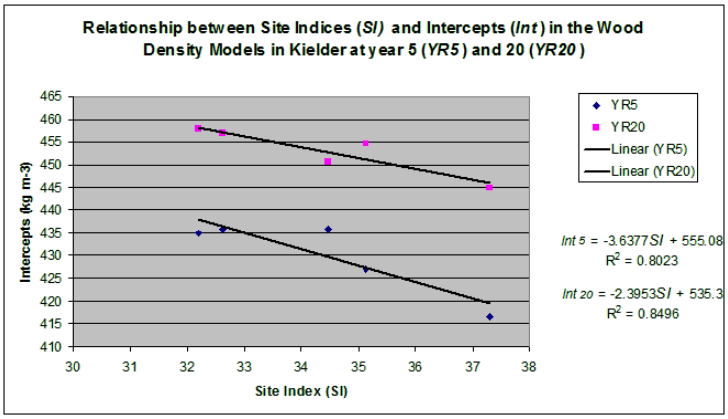


Figure 5. Relationship between the intercepts (Int) in the Power Law models in Table 6 and Site Index (SI). The two lines show a linear relationship between these two factors at age 5 (YR5) and 20 (YR20).

The exponents in the Power Law models were slightly lower than in Aberfoyle. This meant that the Dominance structure was fully established in those plots from the beginning of the simulations and remained unaltered afterwards. As a result, the Wood Density differences between the Dominance groups were slightly smaller in Kielder Forest. In Aberfoyle, the thinning practices and localised windthrow created the conditions for canopy expansion that mainly benefited the dominant trees. The end result was an increase of the differences between the largest and the smallest trees in those plots.

Stem Straightness remained similar in all plots apart from plot 9 and 10 that experienced an increase in their R^2 values towards the end of the simulation period. The intercepts showed a trend towards a progressive increase in Straightness similar to that observed in Aberfoyle. However, the increase was slower due to the smaller DBH increments. The exponents were becoming slightly steeper towards the end of the simulations as the large trees were starting to expand crowns, colonising the space left by the meta-suppressed that were progressively eliminated by competition. Like in Aberfoyle, Stem Straightness kept improving with age, especially in the dominant and co-dominant groups. However, they remained far from the class 7 threshold. This was due to higher windiness scores in this area compared to Aberfoyle.

4. Discussion and Conclusion

LiDAR analysis was used to generate detailed inventory data about the characteristics and location of individual trees in a group of plots in two monitoring areas. These data were combined with a distance-dependent individual tree growth model (TASS) to run simulations in time about the changing characteristics of each tree as a result of initial vigour (as measured initially by LiDAR) and competition with neighbours. The tree lists generated with TASS at each time intervals were input into a timber quality model (ConTQ) developed at Forest Research that is able to make estimations of Wood Density and Stem Straightness.

The predictions of these two timber quality parameters made by ConTQ were linked to Dominance status as defined by the TASS model. The result was a collection of Power Law models that helped to explain the influence of the Dominance characteristics of each individual on the progressive development of the physical characteristics of the stems.

The conclusions of this work were that Wood Density was mainly affected by SI. Age and management practices acted as modifiers of the main trends. Dominance helped to explain the spread of the Wood Density values within a stand. In terms of Stem Straightness, the influencing factors were DBH and Windiness, as measured by the DAMS scoring system. Age also contributed to a progressive increment of the straightness scores by constant Volume increment and the elimination of small trees in the lower Dominance ranks by competition.

The operation of ConTQ with tree lists can provide foresters with a much more flexible tool for testing the consequences derived from the implementation of different silviculture strategies on a forest stand. The model simulations showed that the level of response to increased spacing was followed differently by each individual in the forest community according to its position in the hierarchy. Dominance status, as defined by TASS, summarises the capability of each tree to respond to new growing conditions and the long-term consequences of creating those conditions in the first place. Therefore, it can be assumed that, in the near future, the combination of tree models and remote sensing techniques can create a new paradigm in the use of tools for forest management.

Acknowledgements

We would like to thank staff at the Ministry of Forest in Victoria (Canada) for accommodating a visiting scientist, for permitting use of the TASS model and for providing wide-ranging expertise in support of this work.

This research is funded by Forest Research.

References

Achim, A., Gardiner, B.A., Leban, J.M. and Daquitaine, R. (2006). 'Predicting the branching properties of Sitka spruce grown in Great Britain'. New Zealand Journal of Forestry Science, 36 (2/3): 246-264.

Bell, P. D. , Quine, C. P. & Wright, J. A. (1995). The use of digital terrain models to calculate the windiness scores for the windthrow hazard classification. Scottish Forestry, 49: 217 - 225.

CEN. (2003). 'Structural timber - Strength classes. EN338:2003'. European Committee for Standardisation, Brussels. 14 p.

Forestry Commission. (2007). 'Forestry statistics 2007 – a compendium of statistics about woodland, forestry and primary wood processing in the United Kingdom'. 150 p. Forestry Commission. Edinburgh.

Gardiner, B. A., Achim, A., Leban, J.-M. and Bathgate, S. (2005). 'Predicting the timber properties and performance of Sitka spruce through the use of simulation software'. In Proceedings of 5th IUFRO Workshop on Connection Between Forest Resources and Wood Quality. G. Nepveu (ed). Waiheke Island, New Zealand, 20 – 27 November 2005.

Macdonald, E. and Hubert, J. (2002). 'A review of the effects of silviculture on timber quality of Sitka spruce'. Forestry, 75: 107-138.

Macdonald, E., Gardiner, B. and Mason, W. L. (2009). 'The effects of transformation of even-aged stands to continuous cover forestry on conifer log quality and wood properties in the UK'. Forestry, 23:1-22.

Mitchell, K.J. (1975). Dynamics and Simulated Yield of Douglas fir. Forest Science. Vol 21 no 4 pp 1-40.

Suárez, J.C, di Lucca, M., Goudie, J., Polsson, K., Xenadis, G., Gardiner, B.A. and Perks M. (2009). 'An individual canopy delineation algorithm based on Object-Oriented segmentation and classification'. Proceedings of Silvilaser 2009, 9th International Conference on LiDAR Applications in Forest Assessment & Inventory, October 2009, University of Texas.

3D segmentation of forest structure using an adaptive mean shift based procedure

ANTÓNIO FERRAZ †‡§, FRÉDÉRIC BRETAR †¶, STÉPHANE JACQUEMOUD ‡, GIL GONÇALVES §, LUISA PEREIRA |

antonio.ferraz@ign.fr

† Institut Géographique National, Laboratoire MATIS, Saint Mandé, France

‡ Institut de Physique du Globe de Paris, Géophysique spatiale et planétaire, Paris, France

§ Instituto de Engenharia de Sistemas e Computadores de Coimbra, Portugal

| Escola Superior de Tecnologia e Gestao de Agueda, Universidade de Aveiro, Portugal

Abstract

Plant communities display a vertical structure based on the size and growth pattern of the dominant species. To a large extent, this pattern, called vertical stratification, depends on the climatic zone. Vertical structure analysis consists in detecting the number of layers and their limits within a forest stand. So far, there is a lack of robust approaches applied to airborne laser scanning (ALS) data that properly segment the different strata of forests having complex structures. In this study, we propose a procedure to characterize vertical forest stratification based on the mean shift (MS) algorithm. The MS is a non-linear filter that searches for local density maxima (modes). It is a non-parametric and unsupervised approach, which only requires a single criterion, the kernel bandwidth. Since the forest point cloud is a multi-modal distribution, the MS is used to find the modes which are supposed to be the barycenters of vegetation features. Once achieved, the modes are grouped together according to height range and the corresponding ALS points are assigned to each vegetation strata. Due to their complex pattern, using a single scale over the whole space is not recommended for the analysis of such environments. On this basis, the modes are computed using a variable kernel bandwidth according to the forest pattern. To depict such a pattern, we propose a new technique that segments the main forest layers at the plot level: overstory, understory, and surface vegetation. The procedure has been carried out on 45 plots of a Portuguese forest mainly composed of eucalyptus (*Eucalyptus globulus*) and pine (*Pinus pinaster*) trees that can be strongly populated by understory and surface vegetation.

1. Introduction

Forest and woodland display simple single-storied structures or more complex multi-storied structures. Dominant and codominant trees form the overstory layer. Subdominant trees still have access to light but do not occupy the upper canopy. The understory layer generally grows in the shadow underneath the uppermost layer. It is made of suppressed trees, juvenile trees, and tall shrubs. The surface layer is immediately above the ground. Small shrubs and herbaceous plants characterize it. All these layers may have different density, thickness, and water content. Vertical stratification consists in detecting the presence or not of the different layers within a forest stand and in estimating their limits. It is known to play a crucial role in the distribution of fuels, and consequently, in fire behaviour, habitat quality, microclimatic conditions, carbon storage, etc.

Many authors showed the potential of multi-echo airborne laser scanning (ALS) data to compute digital terrain models (DTM) over vegetated areas, or to extract forest variables (Hyyppa et al., 2008). However, most studies focus on the canopy layer, which does not fully characterize heterogeneous forests such as Mediterranean ecosystems. Only a few are interested in the vertical segmentation of forest structures. Zimble et al. (2003) and Maltamo et al. (2005) characterize forest plots as single-storied or multi-storied by

analysing the height distribution of vegetation hits. Riano et al. (2003) approach vertical forest stratification by applying clustering techniques directly to the 3D raw ALS data. As pointed out by these authors, the cluster algorithm works well when there are only two structures, but it fails in the presence of more complex structures. Morsdorf et al. (2010) discriminate vegetation strata over treated plots applying a supervised cluster analysis on a two-dimensional feature space. The latter is spanned by the height and the intensity of the ALS points. This study takes advantage of the fact that the vegetation layers are unispecies. Some authors used intensity values to discriminate trees (Holmgren and Persson, 2004; Kim et al., 2009). They assume that the ALS points corresponding to trees of the same species and age class have similar values. However, the layers seldom display such a property. Moreover, due to calibration problems (unknown instrument specifications) and uncertain interpretation of the digital numbers obtained by the instruments (fuzziness caused by the vegetation itself), using echoes intensity for ALS vegetation studies remains a challenge.

Among the ALS studies that focus on canopy layers, emphasis is often placed on tree height and/or crown width. Very few studies try to assess the vertical component of the topmost layer (Holmgren and Persson, 2004; Barilotti et al., 2008; Popescu and Zhao, 2008). They usually deal with the crown base height by analysing the vertical profile of the laser hits within a single crown with different techniques. Usually, individual crowns are delineated from the canopy height model (CHM) by image segmentation. The main disadvantage of such a method is that the dominated and smaller trees are invisible in the CHM. Thus, mapping the overstory vertical component at larger scales is highly dependent on plot homogeneity. Other authors try to figure out the nature of the bi-storied or multi-storied overstory by means of tree-based approaches. Reitberger et al. (2009) apply a sophisticated method based on normalized cut segmentation to depict dominated trees. Despite good performance, it is highly site-dependent since several empirical parameters are involved.

In this work, we propose a new mean shift (MS) based procedure to segment forest. Since its reformulation by Comaniciu and Meer (2002), the MS algorithm has been mainly applied to image segmentation. The processing of unstructured ALS point clouds using it was first proposed by Melzer (2008) to extract power lines. It is a non-linear filter that searchs for local density maxima (modes) and, unlike other methodologies, that requires no prior geometrical or statistical knowledge. Since the point cloud is a multi-modal distribution, the MS is used to find the ALS modes that are supposed to be the vegetation features barycenters. A global mode corresponds to a forest layer, such as the overstory, while a local mode points out a tree crown or a shrub. The only parameter used by the MS technique precisely relies on this concept of scale. The area inside which the modes (the kernel bandwidth) are calculated is of crucial importance. In the following, we first study the impact of the kernel bandwidth on the data set, and then propose a procedure that permits forest stratification that takes into account scaling effects.

2. Material and methods

2.1 Study area

The study area is located in north-west Portugal (40°36' N, 8°25' W), nearby the city of Águeda (figure 1(a)). The selected area covers 900 ha at an altitude ranging from 70 m to 220 m, with gentle to steep slopes (figure 1(b)). The landscape is predominantly composed of woodlands dominated by eucalyptus (Eucalyptus globulus) with some stands of pine (Pinus pinaster). One also finds shrublands as well as agricultural fields, and a few built-up areas. The eucalyptus stands correspond to regular and irregular plantations, the management of which is mainly done by short rotations of about 10-12 years to feed the pulp and paper industry. In spite of a limited extension, the study area is composed of various eucalyptus stands in terms of age and growth (plantation, coppice or mixed). The forest stands can be also populated by understory and surface vegetation. In such cases, these strata are mainly composed of juvenile trees (eucalyptus, pine, acacia, and oak), gorse (Ulex europaeus), tree heath (Erica arborea), carqueja (Pterospartum tridentatum L. Willk), fern, and herbaceous plants.

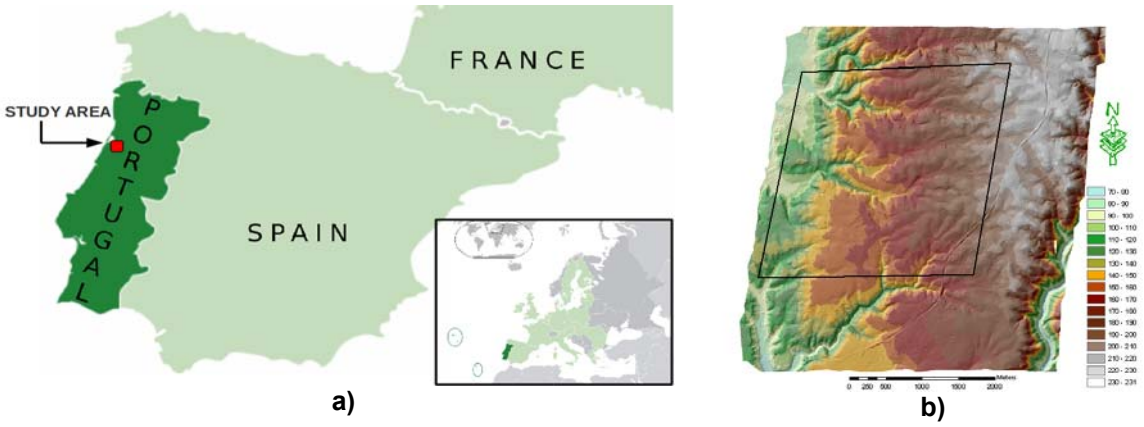


Figure 1. Study area: (a) localization and (b) delimitation over the digital terrain model

2.2 Field data collection

An extensive forest inventory was performed on plots defined by a systematic sampling method. 45 plots were selected in a regular grid made of 325 m x 325 m cells, superimposed on a forest land-cover map produced by aerial photointerpretation. 43 out of 45 plots correspond to forest, 43 mainly composed of eucalyptus and 2 of pine. The geometrical center of each cell defines the center of the forest plots. The sampling plots are circular and composed of two concentric circles. An outer circle (400 m^2), hereafter called Plot, and an inner circle (200 m^2), called Subplot. The plot delineation was performed using a measuring tape. The characterization of vegetation complies with the official Portuguese forest inventory manual published by the Direcçao Nacional de Florestas (DGF, 1999). Detailed information about this field experiment can be found in Pereira and Gonçalves (2010).

2.3 Airborne laser scanning data

The ALS data were acquired in July 14, 2008 using the RIEGL LMS-Q650 laser scanner in a full-waveform mode*, in the framework of a research project funded by the Portuguese Foundation for Science and Technology (FCT)*. The acquisition parameters are listed in table 1. The manufacturer delivered the point cloud after processing of the full-waveform data. To calculate the effective height of the objects, the point cloud is normalized by calculating a digital terrain model (DTM). The ground points are first classified using the TerraScan software (Soininen, 2010). The DTM is then defined by taking the lowest values within a neighborhood of 0.3 m. The missing pixels are obtained using a Delaunay triangulation. Note that the points classified as ground are not removed from the dataset. In the following, we consider all points as vegetation hits.

Table 1.ALS acquisition parameters

ALS sensor	RIEGL LMS-Q5600
Wavelength	1064 nm
Scan angle	45°
Pulse rate	150 kHz
Effective measurement rate	75 kHz
Beam divergence	0.5 mrad
Ground speed	46.26 m/s
Flying height above the terrain	640 m
Swath	479 m
Sidelap	70%
Single run density	3.3 pt/m^2
Expected final density	9.9 pt/m^2
Distance between lines	150m
Spot diameter	30cm

2.4 Adaptive mean shift

The mean shift (MS) is a non-linear filter based on the Parzen window kernel that looks for local density maxima (modes) in a set of data samples (Comaniciu and Meer, 2002). The solution of the algorithm iteratively converges towards the local maximum:

$$\boldsymbol{X}^{t+1} \leftarrow \boldsymbol{X}^{t} + m_{h,G}\left(\boldsymbol{X}^{t}\right) \tag{1}$$

,where t denotes the number of iterations, h the kernel bandwidth, and $m_{h,G}(\boldsymbol{X})$ is the so-called mean shift vector defined by:

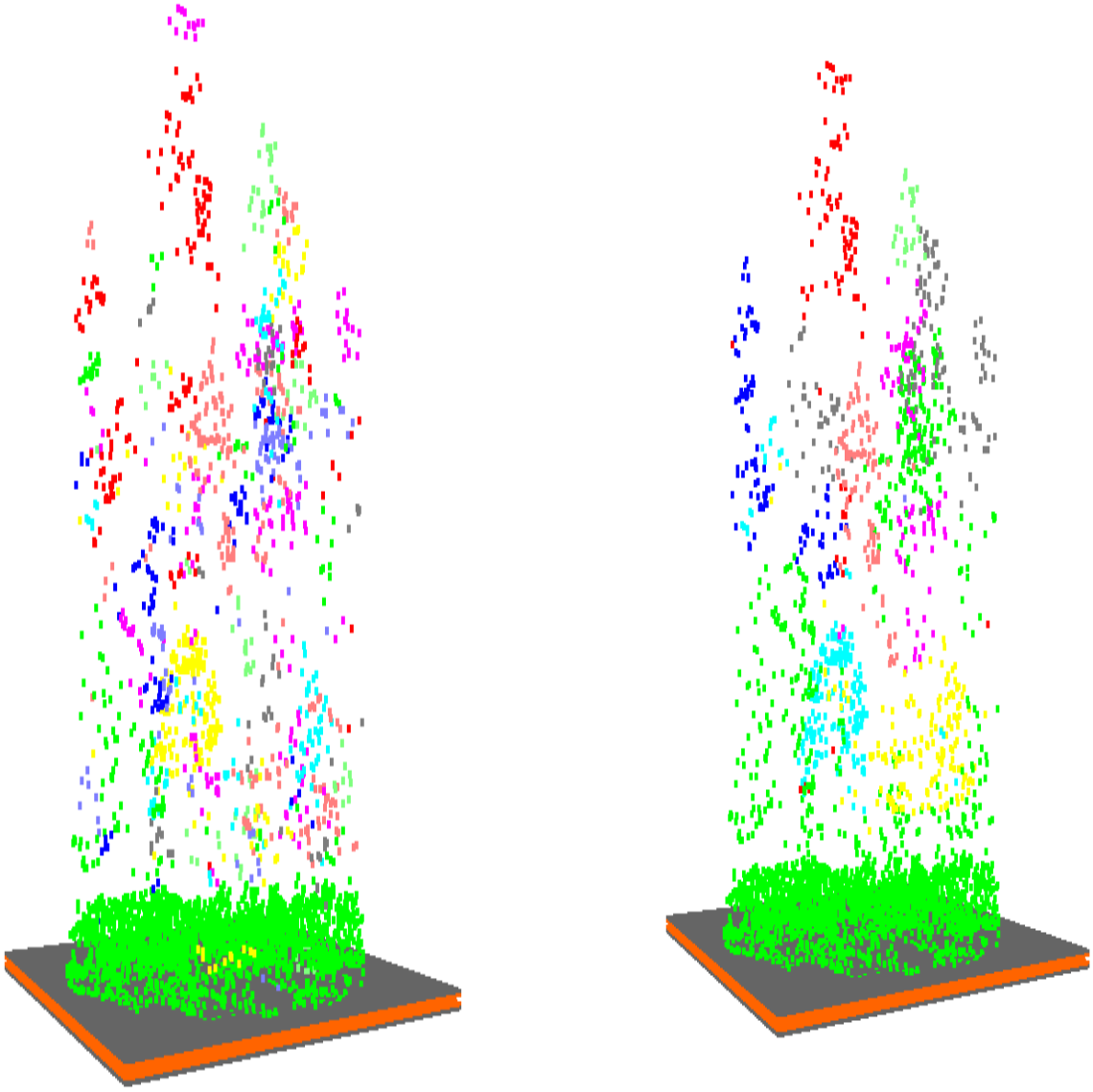
$$m_{h,G}\left(\boldsymbol{X}\right) = \frac{\sum_{i=1}^{n} \boldsymbol{X}_i \cdot g\left(\left\|\frac{\boldsymbol{X}-\boldsymbol{X}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{\boldsymbol{X}-\boldsymbol{X}_i}{h}\right\|^2\right)} - \boldsymbol{X} \tag{2}$$

$\boldsymbol{X}_{i=1,\ldots,n}$ are the n ALS data points in a three-dimensional space and $g(\boldsymbol{X})$ is the derivative of the Epanechnikov kernel $k(\boldsymbol{X})$:

$$k\left(\boldsymbol{X}\right) = \begin{cases} c_{k,d}\left(1-\left\|\boldsymbol{X}\right\|^2\right) & \text{if inside the unit sphere} \\ 0 & \textbf{if outside} \end{cases} \tag{3}$$

Comaniciu and Meer (2002) proved that the MS algorithm converges on a stationary point. It can easily be extended to a distance-based segmentation technique, grouping together all the modes that are closer than a certain distance $r \in$. The MS segments are then retrieved by aggregation of the basins of attraction of the corresponding convergence points. In the following, we set $2r = h$ for all experiments. The choice of the kernel bandwidth h is critical because it strongly impacts on the results. A small kernel width leads to several distinct modes (small basins of attraction, more and smaller objects), while a large kernel width aggregates small structures into larger ones (small number of modes with large basins of attraction). The determination of an optimal value of h is actually a major challenge for an efficient MS segmentation. As far as the vertical component is concerned, the forest layer depth increases with height. Typically, scrubby vegetation strata are thinner than mature tree layers. The optimal value of h that allows a proper segmentation of the shrubs may fragment a tall tree into several segments (lower branches, top foliage, etc.). Figure 2(*a*) results from a segmentation with $h = 1$ m. While the surface vegetation has a coherent shape, the higher features are oversegmented. Increasing h improves the overstory segmentation, but it may cause merging between close small vegetation features.

For instance, if $h = 4$ m (figure 2(*b*)) the denser surface vegetation will attract the sparse understory, causing an undersegmentation of the scene. Thus, using a single scale over the entire space is not recommended for the analysis of forest environments. As far as the MS technique is concerned, several statistical approaches deal with the scale selection (Comaniciu, 2003, Huang and Zhang, 2008; Bo et al., 2009). Regarding the task dependent applications the kernel bandwidth can be provided by the user (Comaniciu and Meer, 2002). Here it is calculated as a function of the height range: overstory, understory, and surface vegetation.



a) b)

Figure 2. Mean shift segments over Subplot 30 using (a) $2r = h = 1$ m *and (b)* $2r = h = 4$ m.

A height frequency histogram usually shows that forest plots are characterized by two dense zones: the overstory and the ground. A third zone may arise when dense understory is present. In this respect, we apply the MS segmentation technique to compute the basins of attraction of the two denser zones by calculating the two main modes of the ALS point cloud. Since we study the vertical profile at the plot level, the influence of the horizontal coordinates is removed in the computation. Thus, for planimetric distances, the kernel bandwidth is set to the plot diameter. The vertical extension of the kernel is defined as the value that forces the ALS points to converge towards two modes only. To calculate such a bandwidth, we set the initial kernel bandwidth to 1 m and increment it by 1 m at each iteration until the MS algorithm accomplishes only two modes. The borderline between the basins of attraction defines the overstory height threshold.

We consider that forest plots in which the 95% height percentile is lower than 5 m are composed of a maximum of two layers. However, in plots with taller trees, a third layer may coexist. In such a case, the understory height threshold is set to 1 m. This assumption and its consequences are discussed below. Some results are presented in figure 3 using height frequency histograms.

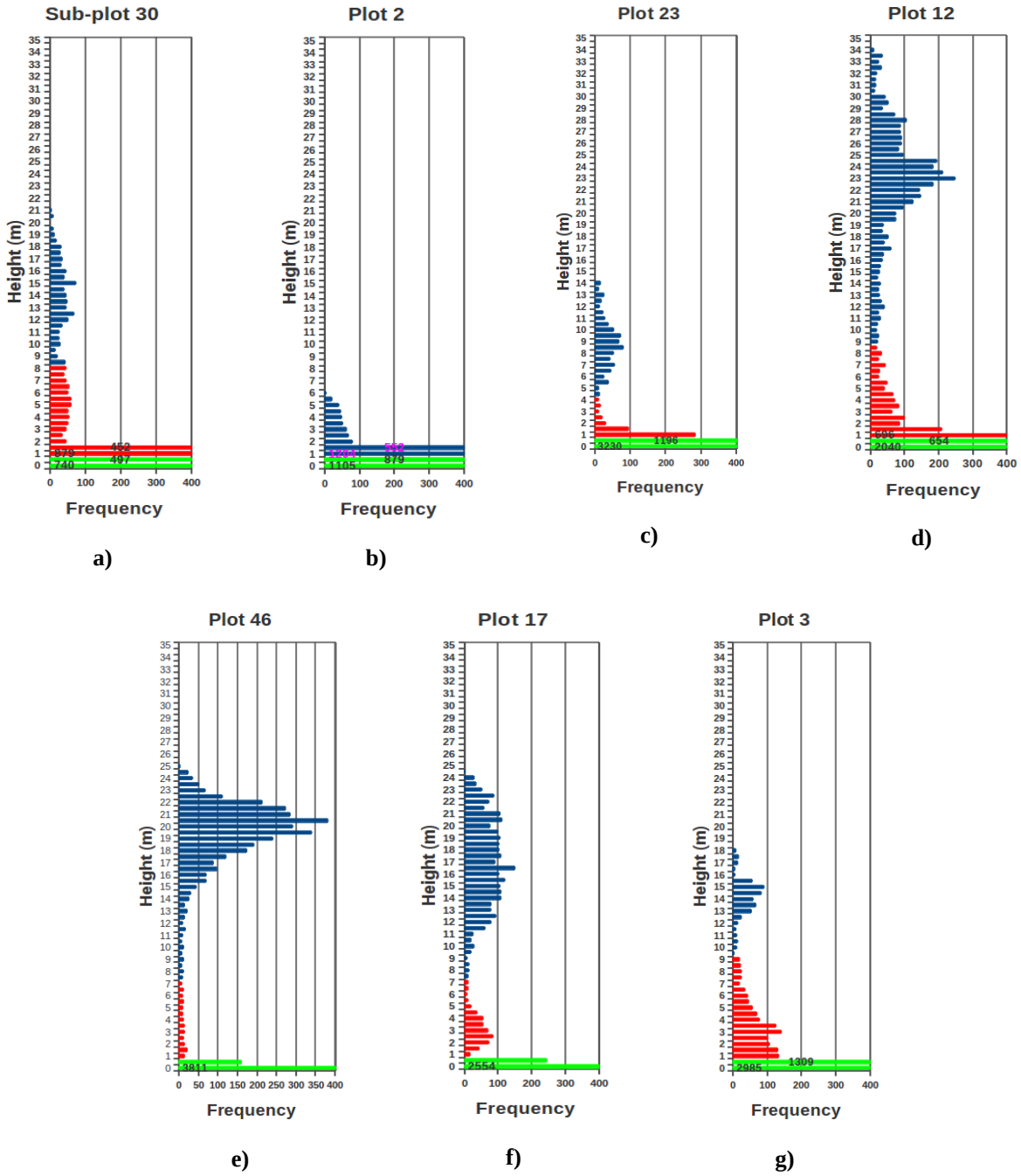


Figure 3. Forest stratification of seven plots. The results are illustrated by a height frequency histogram. (a) Subplot 30; (b) Plot 2; (c) Plot 23; (d) Plot 12; (e) Plot 46; (f) Plot 17; (g) Plot 3. The different colors define the strata delimitations. The histograms are limited to 400 points; for bins over that value, the real frequency value is given.

The initial solution of the segmentation at the plot level is then refined at a thinner scale to get a better discrimination of forest strata. The MS is reapplied with a modulation of the kernel bandwidth. Because vegetation features, e.g. the eucalyptus crowns, are not spherical we adopt two different metrics for planimetric and height distances. The horizontal component is kept constant to 1 m while the vertical component of the bandwidth is empirically set to 0.66 times the overstory amplitude (difference between the

highest and the lowest bin of the topmost distribution) and to 0.5 times the understory amplitude. As for the surface vegetation, the bandwidth is set to 1 m in all directions.

2.5 Adaptive mean shift procedure

The procedure to detect strata works iteratively from the bottom to the top of the forest structure, by adapting the kernel bandwidth within the defined height ranges (figure 3). First, the 5% height percentile, $w_{l=1}$, of the data points $\boldsymbol{X}_i = (x_i, y_i, z_i)$ is calculated. The adaptive kernel bandwidth, h, is set to a value ranging within the heights defined in section 2.4. Therefore, the modes $\boldsymbol{X}_i = (x_i, y_i, z_i)$ for each ALS point, $\boldsymbol{X}_i$, are computed using the corresponding value of h. All modes that are closer than r are grouped together creating MS segments, $\boldsymbol{C}_{p\in\square}$. The forest layer, $\boldsymbol{F}_{l=1}$ (figure 4(*b*)), is a set of $\boldsymbol{X}_i$ for which the corresponding MS segments are closer than $s_{l=1}$ from $w_{l=1}$. The ALS assigned points are not taken into account to further calculations. This step improves the segmentation by removing the influence of the lower layers, which are usually denser

PROCEDURE

REPEAT

1: $\forall z_i \in \boldsymbol{X}_i, \quad w_l = P_{0.05}$

2: $\forall \boldsymbol{X}_i, \quad \boldsymbol{X}_i^* = \lim_{t\to\infty} \boldsymbol{X}_i^t$ → equation (1)

3: $\forall \boldsymbol{X}_i, \quad \boldsymbol{C}_p = \left\{ \boldsymbol{X}_i \mid \forall u = 1,...n \quad \left\| \boldsymbol{X}_i - \boldsymbol{X}_u \right\|^2 \le r^2 \right\}_{p\in\square}$

4: $\forall p, \quad \boldsymbol{F}_l = \left\{ \boldsymbol{X}_i \mid \exists \boldsymbol{X}_i \in \boldsymbol{C}_p \quad \left\| z_i - w_l \right\|^2 \le s_l^2 \right\}_{l\in\square}$

5: $\forall \boldsymbol{X}_i, \quad \boldsymbol{X}_i = \left\{ \boldsymbol{X}_i \mid \boldsymbol{X}_i \notin \boldsymbol{F}_l \right\}$ → removal of the assigned points

6: $\boldsymbol{X}^*_i = \varnothing$ → reinitialize modes

UNTIL $n = 0$

than the higher. When two regions with different densities are adjacent, the points of the sparser region are likely to be shifted towards the denser one (figure 2(*b*)). In the second iteration the 5% height percentile of the remaining points, $w_{l=2}$, is calculated in order to define the new value of h (figure 4(*c*)) . The procedure keeps on until all points are assigned to one layer.

Note that the value of s_l defines the resolution of the forest stratification, i.e., the number of strata. Since we want three strata, we set s_l to the amplitude of layer l. However, it can take lower values to thinner analyses, for example to depict bi-storied overstories.

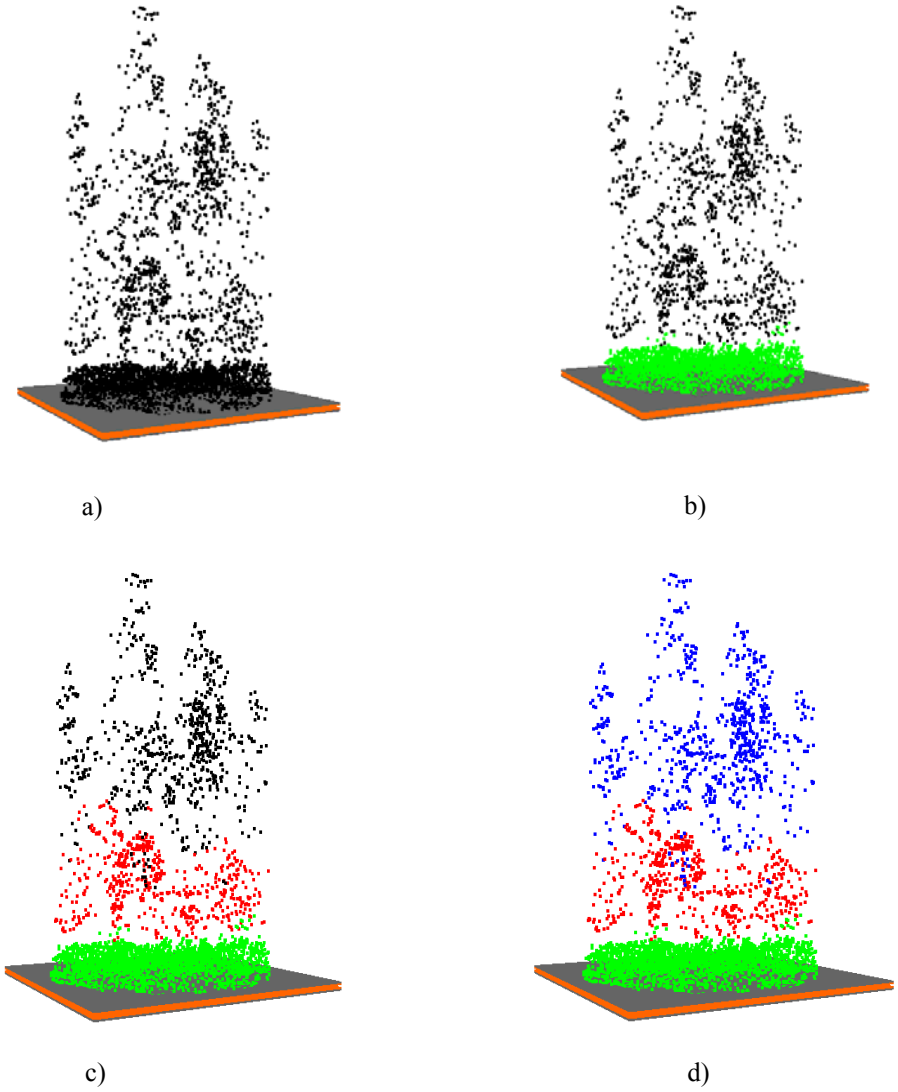


Figure 4. Procedure iterations on Subplot 30 (200 m2). (a) Original ALS points (black). (b) First iteration, points assigned to the first layer (green) with $w_1 = 0$ m*, thus* $2r = h = 1$ m *and* $s = 1$ m*. (c) Second iteration, points assigned to the second layer (red) with* $w_2 = 2.38$ m*, thus* $2r = h = 3.75$ m *and* $s = 7.5$ m*. (d) Third iteration, points assigned to the third layer (blue) with* $w_3 = 8.04$ *m, thus* $2r = h = 8.3$ m *and* $s = 12.5$ m*.*

3. Results and discussion

The reliability and robustness of the procedure has been assessed on all defined circular plots (400 m^2). Figure 5 shows the results over six plots with various forest structures: single layer plots (figure 5(*d*)); multi layer plots (figure 5(*c*), figure 4), plots with single overstory stands (figure 5(*e*)), plots composed of different stands (figure 5(*f*)), juvenile plots (figure 5(*a*)) and adult stands (figure 5(*d*)). Plot 2, Subplot 30 and Plot 12 (respectively, figure 5(*a*) and figure 4 and figure 5(*c*)) are characterized by a denser surface vegetation (mean height of about 1.3 m, 1.5 m and 1 m, respectively) while Plot 23 (figure 5(*b*)) is populated by moderate surface vegetation (mean height of about 0.40 m) with some species 1 m tall.

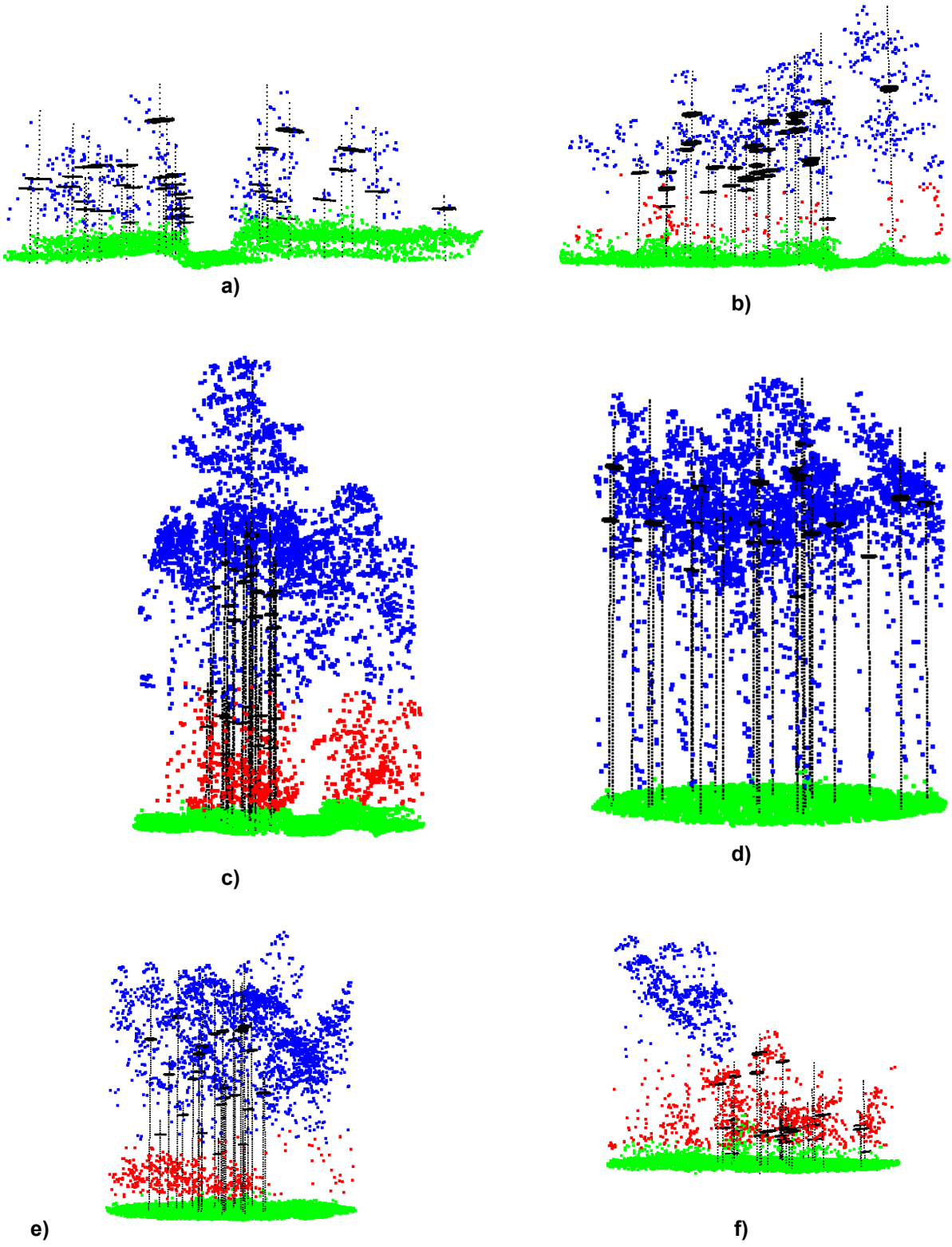


Figure 5. Results of the MS-based procedure over five eucalyptus plots and one pine plot. The surveyed tree metrics are also displayed. (a) Plot 2; (b) Plot 23; (c) Plot 12; (d) Plot 46; (e) Plot 17; (f) Plot 3. The plots have, respectively, mean age of 3, 6, 13, 60, 10 and 3 years.

Figures 4 and 5 show that the original points progressively converge by means of the MS algorithm, producing coherent segments at least along the vertical component. It can be visually assessed that are well assigned to layers. Only two strata are retrieved both in figures 5(*a*) and 5(*d*). In Section 2.4, only two segments were defined for juvenile plots (figure 3(*a*)) while three segments were expected for adult stands. However, figure 5(*d*) shows that only two strata have mushroomed from the procedure. Due to the lack of understory (this plot only has trunk reflections beneath the overstory), the 5% height percentile at the second iteration equals 9.25 m. Figure 3(*e*) shows that the kernel bandwidth at the second iteration is directly set to one of the overstory stratum. All ALS hits lying on the tree stem were then assigned to the overstory. Thus, by applying our approach, the number of layers retrieved is inherent to the forest pattern. Moreover, it provides a fuzzy stratification of the forest where different layers can interpenetrate. It is a more realistic result than a simple stratification only based on height thresholds (figures 4 and 5). Although the surface vegetation height threshold was initially set to 1 m, the procedure is able to figure out the shape of this layer. In figures 4 and 5(*a*), points up to 2 m height are classified as surface vegetation. A similar analysis can be done for the upper strata. Moreover, the global framework can provide multi layer segmentation. For instance, by setting s_l to half the forest layer amplitude, the forest plot can be stratified with a thinner resolution (single-storied and bi-storied). The post-processing of each layer derived from a coarser analysis is certainly a field to investigate.

4. Conclusion

In this paper, we explore the MS potential to segment vegetation features. Additionally, a unsupervised and non parametric approach is proposed to extract forest layers. It is carried out in three dimensions providing genuine 3D segments. Thus, the procedure is able to compute large areas at the thinner resolution. The only parameter that needs to be defined (the kernel bandwidth) is here adapted as a function of the forest pattern. To analyse such pattern, we introduce a technique that stratifies forest at the plot level by computing the basins of attraction of the two denser zones. Since the MS applies to a joint space (spatial and attributes), additional attributes of the ALS points (intensity, number of echoes, etc.) may be introduced in the calculation of the MS vector. Additionally, the analysis of different kernel functions and metrics may help to better fit the tree crowns aiming at a thinner forest stratification but also at a better fit the tree crowns in single tree extraction. In the near future, we pretend to validate our procedure using forest inventory data acquired in the frame of the FCT project.

References

Barilotti, A., Sepic, F. and Abramo, E., 2008, Automatic detection of dominated vegetation under canopy using Airborne Laser Scanning data. In Proceedings of the International Conference Silvilaser 2008, 17-19 September 2008, Edinburgh, UK, pp. 134-143.

Bo, S., Ding, L., Li, H., Di, F. and Zhu, C., 2009, Mean Shift-based clustering analysis of multispectral remote sensing imagery. International Jounal of Remote Sensing, 30, pp. 817-827.

Comaniciu, D. and Meer P., 2002, Mean shift: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24, pp. 603-619.

Comaniciu, D., 2003, An Algorithm for Data-Driven Bandwidth Selection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25, pp. 281-288.

DGF, 1999, Manual de instruções para o trabalho de campo do inventario florestal nacional. Direcção Geral das Florestas, Lisboa.

Holmgren, J., and Persson, A., 2004, Identifying species of individual trees using airborne laser scanner. Remote Sensing of Environment, 90, pp. 415-423.

Huang, X. and Zhang, L., 2008, An adaptive mean-shift analyses approach for object extraction and classification from urban hyperspectral imagery. IEEE transactions on Geoscience and Remote Sensing, 46, pp. 4173-4185.

Hyyppa, J., Hyyppa, H., Leckie, D., Gougeon, F., Yu X. and Maltamo, M., 2008, Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. International Journal of Remote Sensing, 29, pp. 1339-1366.

Kim, S., McGaughey R. J., Andersen, H.-E., Schreuder. G., 2009, Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. Remote Sensising of Environment, 113, pp.1575-1586.

Maltamo, M., Packalén, P., Yu, X., Eerikainen, K., Hyyppa J. and Pitkanen, J., 2005, Identifying and quantifying structural characteristics of heterogeneous boreal forest using laser scanner data. Forest Ecology and Management, 216, pp. 41-50.

Melzer, T., 2008, Non-parametric segmentation of als point clouds using mean shift. Journal of Applied Geodesy, 13, pp. 159-170.

Morsdorf, F., Mårell, A., Koetz, B., Cassagne, N., Pimont, F., Rigolot, E., Allgöwer, B., 2010, Discrimination of vegetation strata in a multi-layered Mediterranean forest ecosystem using height and intensity information derived from airborne laser scanning. Remote Sensing of Environment, 114, pp. 1403-1415.

Pereira, L. and Gonçalves G., 2010, Accuracy of a DTM derived from full-waveform laser scanning data under unstructured eucalypt forest: A case study. In Proceedings of The XXIV FIG International Congress, 11-16 April 2010, Sidney, Australia, pp. .

Popescu, S. and Zhao, K., 2008, A voxel-based lidar method for estimating crown base height for deciduous and pine trees. Remote Sensing of Environment, 112, pp. 767-781.

Reitberger, J., Schnorr Cl., Krzystek, P. and Stilla U., 2009, 3D Segmentation of single trees exploiting full waveform LIDAR data. ISPRS Journal of Photogrammetry and Remote Sensing, 64, pp. 561- 574.

Riano, D., Meier, E., Allgower B., Chuvieco E. and Ustin, S., 2003, Modeling airborne laser scanning data for the spatial generation of critical forest parameters in fire behaviour modeling. Remote Sensing of Environment, 86, pp. 177-186.

Soininen, A., 2010, TerraScan User's guide. Available online at: http://www.terrasolid.fi/en/users_guide/terrascan_users_guide (Accessed 24/07/2010).

Zimble, D., Evans, D., Carlson, G., Parker, R., Grado, S. and Gerard, P., 2003, Characterizing vertical forest structures using small-footprint airborne LiDAR. Remote Sensing of Environement, 87, pp. 171-182.

Quantifying the height and density of understory trees and saplings using multi-return airborne lidar

*MARIE-CLAUDE JUTRAS PERREAULT and BENOÎT ST ONGE**

st-onge.benoit@uqam.ca

Department of Geography, University of Quebec at Montreal, P.O 8888, box Centre Ville, Montréal, QC, Canada, H3C 3P8

Abstract

While lidar surveys are now being used to acquire data on forest canopy attributes, managers are still looking for tools to map the understory. Making the proper choice of sylvicultural treatments does indeed require information on the lower strata under canopy trees. Using an Optech ALTM3100, we have acquired a lidar coverage with up to four returns per pulse over a southern boreal forest landscape in Canada. In seven plots (576 m^2), each divided in 16 quadrats, we have counted the number of saplings, and estimated their average height per quadrat. Subcanopy trees (suppressed or intermediate) were all precisely positioned and measured. The lidar point cloud was binned by 1 m vertical intervals and the return counts per bin were standardized (to account for occlusion by higher trees). For saplings, a height error of 1 m or less between field data and lidar-based predictions was observed in half of the quadrats. Individual subcanopy trees were studied by extracting the returns between 3 to 16 m located inside a 3D space defined by the 4 crown radii measured in the field, the tree 3D position and the inclination of each field measured tree. The standardized return counts were calculated per 1 m vertical bins within that volume for each tree. The largest inflexion of the count per height-bin curve was automatically detected. Over 306 trees, the inflection height differed from field height by at most 2 m in 50% of cases. Finally, the correlation between the number of subcanopy trees and saplings per hectare and the number of detected local maxima was 0.48 for the saplings and 0.75 for the subcanopy trees. We conclude that the lidar point clouds contain rich information on the understory, even when the canopy surface is closed.

1. Introduction

The old growth boreal forests of North America exhibit complex vertical and horizontal structures due to multiple events of fire, windthrow, insect outbreaks, and natural mortality. Despite the relative low number of canopy tree species (22), their overall diversity is augmented by several low strata vegetation layers such as mosses, herbs and shrubs (DeGrandpré *et al*. 2003). In addition, numerous slow growing subcanopy trees may be present. The characterization of these subcanopy plants can greatly enhance our understanding of biodiversity and forest dynamics and reduce the uncertainty of biomass estimates (Welder *et al*. 1996, George and Bazzaz 2003). Moreover sustainable sylvicultural practices can be enhanced based on better data on saplings and trees of intermediate heights.

Optical remote sensing is limited to describing the canopy surface trees. Promising microwave approaches for mapping 3D forest structure (e.g. PolinSAR, as per Williams and Jenkins 2009) and related techniques (e.g. Polarization Coherence Tomography) are maturing but many of these use airborne scanning lidar as a reference (Tebaldini 2008) to disambiguate predicted vegetation profiles. Lidar sensors themselves have recently evolved into multiple return or full waveform systems. These innovations open up the possibility of better detecting subcanopy vegetation. The first such studies concerned the vertical layering of stands (i.e. identifying single *vs* multiple layered stands) based on tree height variance (Zimble *et al*. 2003) or by thresholding height percentile values (Maltamo *et al*. 2005). By combining percentile data and cluster analysis, Riaño and collaborators (2003) extracted the height of the subcanopy trees in addition to other structural attributes. Other studies focused on the detection of subcanopy shrub layers. Martinuzzi and collaborators (2009) employed a selection algorithm (varSelRF, from the *R* statistical package) to identify the most important variables among several forest height and topographic metrics for predicting the presence of subcanopy shrubs. They obtained an accuracy of 83% in detecting the presence-absence of shrubs using the proportion of grounds returns (height = 0 m), the proportion of vegetation returns between 1 and 2.5 m in

height and the percent slope times the cosine of aspect transformation. Hill and Broughton (2009) have combined a lidar dataset acquired in leaf-off canopy / leaf-on subcanopy conditions to a full leaf-on dataset to map the presence-absence of a subcanopy. Combining the first returns of the leaf-on survey with the last returns of the leaf-off dataset, they detected the presence-absence of understory trees and shrubs in 77% of the cases. Moreover, to predict density and the Lorey's mean height of understory trees, Maltamo and collaborators (2005) have used regression models. They obtained a R^2 of 0.87 (sd of 0.312) for the prediction of the logarithmic number of trees and a R^2 of 0.76 (sd of 0.130) for the prediction of the logarithmic Lorey's mean height. Finally, Lee and collaborators (2007) have attempted to characterize the overall vertical structure of forest stands, and recently, Morsdorf and collaborators (2010) have used intensity and height to discriminate three distinct strata with clusters.

The goal of our study is to develop methods to assess the height and density of understory vegetation using multi-return lidar. Specifically, our objectives are i) to determine the height of subcanopy trees and saplings from unstandardized and standardized lidar hit counts, and ii), to estimate subcanopy tree and sapling density by applying a local maxima method to voxelized and thresholded (subcanopy) lidar canopy height models.

Study region and materials

The study site falls within the Training and Research Forest of Lake Duparquet (TRFLD, 79° 22' W, 48° 30' N), in the Province of Quebec, Canada. It is characterized by small hills with elevations comprised between 227 m and 335 m. The mixed vegetation is composed of common boreal species, and dominated by balsam firs (Abies balsamea L. [Mill.]), paper birch (Betula papyfifera [Marsh.]), trembling aspen (Populus tremuloides [Michx]), jack pine (Pinus banksiana [Lamb.]) and white spruce (Picea glauca [Moench] Voss.). Most stands are mature or over-mature and reach heights of 25-30 m.

The lidar data was acquired on July 12th 2007 using an Optech ALTM3100. Strip overlap was sufficient to avoid data gaps. The first returns density was approximately 3.2 hits m^{-2} (single density, i.e. outside strip overlaps), 7.5 hits m^{-2} within strip overlap, and 13.4 hits m^{-2} for all returns. The last returns were classified as ground and non-ground using Terrasolid's Terrascan. The ground-classified last returns were used to build the digital terrain model (DTM). Table 1 presents the key survey and lidar instrument parameters.

Table 1. Specification of the lidar data acquisition

Specification	2007
Lidar	ALTM3100
Power (µJ)	110
Flight altitude (m AGL)	600
Divergence (mrad)	0.3
Footprint size at nadir (cm)	18
Pulse frequency (Hz)	50 000
Max. scan angle (degrees)	10
First return density (hits m^{-2})	3.2

Field data was acquired in July and August 2007. Seven square plots of 576 m^2, composed of 16 quadrats, were located and sampled. The plots were chosen for their representativeness of particular stands and to maximize variability between them, with regards to dominant species and understory density. Within each plot, every tree with a minimum diameter at breast height (DBH) of 5 cm was geolocated with a total station linked to a differential GPS (DGPS) to obtain sub-metre absolute accuracy and centimetre relative accuracy. The height of these individual trees was measured using a Vertex or a height pole and the radii of the crowns were recorded along 4 orthogonal azimuths. The trees were classified as dominant, codominant, intermediate or suppressed, and then a percentage of defoliation was noted. For the leaning trees, the degree of inclination and the leaning azimuth was taken. Saplings (here defined as trees higher than 1 m but with a DBH lower than 5 cm) were counted and classified as softwood or hardwood for each quadrat. The mean height of the saplings in a quadrat was noted when their height was relatively uniform, which was the case for 27 quadrats.

Figure 1 clearly shows a lidar point cloud example in which subcanopy vegetation is visible.

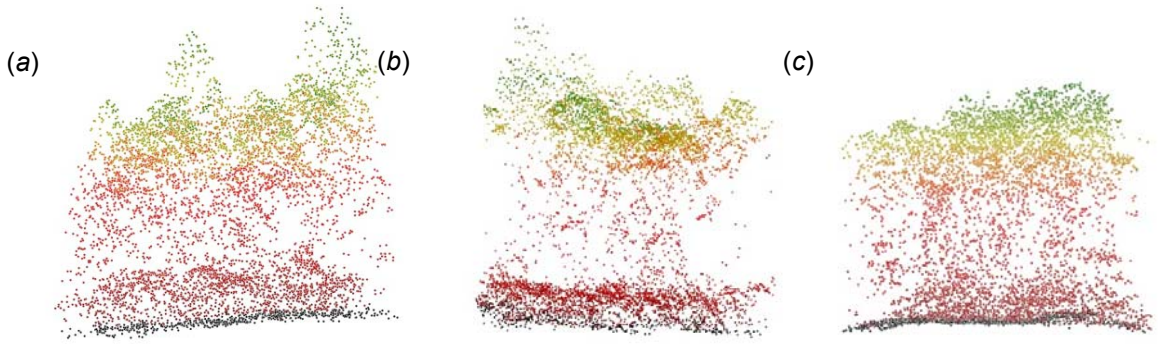


Figure 1. Vertical distribution of the point clouds for three different plots. (a) Canopy cover: white spruce, trembling aspen and paper birch; understory: mountain maple (Acer spicatum [Lam.]), balsam fir and paper birch. (b) Canopy cover: trembling aspen; understory: balsam fir, white spruce and mountain maple. (c) Canopy cover: jack pine; understory: white spruce, balsam fir and paper birch.

2. Methods

2.1 Lidar data preparation and binning

The analyses presented in this paper are essentially based on lidar return counts per 1 m height bin. The total number of return per bin was calculated per plot (576 m^2), per quarter-plot (144 m^2), and per plot quadrat (36 m^2). This allows the assessment of scale effects on the results. A visibility class was manually attributed to each quadrat and each individual subcanopy trees to document the fact that low vegetation was either directly visible from the sky or was occluded by taller trees.

To compensate for the progressive attenuation of laser pulses as they travel though the canopy and are partly or entirely reflected before reaching subcanopy levels, we have developed a standardization method similar to the one proposed by Riaño and collaborators (2003) (equation 1). Due to lack of space in this paper, we cannot show the mathematical reasoning but only the resulting equation used to correct the number of returns:

$$N_c = N_r / N_i \quad (1)$$

where N_r is the number of observed lidar returns of any type (first, second, etc.) within the 1 m vertical bin to be standardized, N_i is the number of incident lidar pulses (not returns) that travelled within the bin (whether they ended in this bin or below), and N_c is the corrected (standardized) number of returns within the bin. Note that N_i corresponds to the total number of incident laser pulses on the canopy surface minus the number of pulses that were completely stopped above the bin (whose last return happened above the bin).

Figure 2 shows a comparison of the unstandardized and standardized vertical distributions of returns. Moreover, the point cloud was cleaned to help certain analyses, such as the detection of local maxima under the canopy. First, the number of returns was counted per 1 m^3 voxel. A first filtered version was created by removing returns within voxels having a count of one, and a second version by removing returns within voxels having a count of two or less. Figure 3 shows the effect of point cloud cleaning.

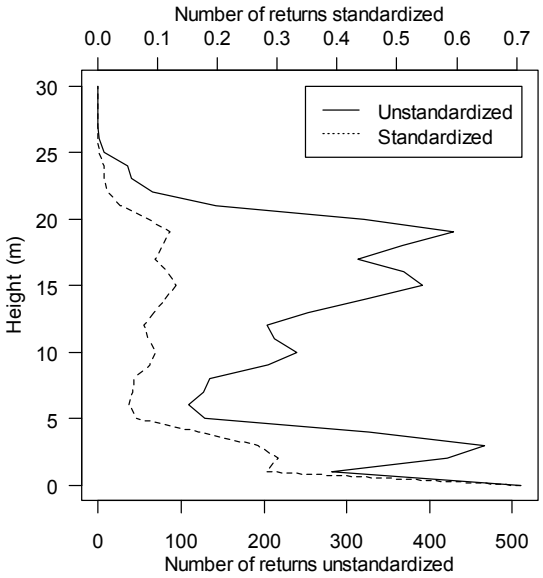


Figure 2. Comparison between standardized and unstandardized

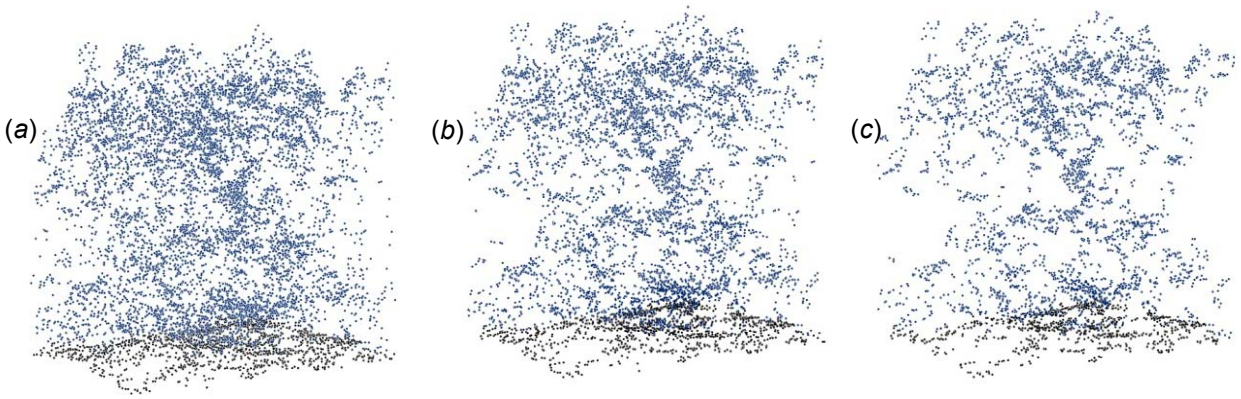


Figure 3. Suppression of the isolated returns in a plot. (a) Complete point cloud. (b) Returns within voxels having more than one return. (c) Returns within voxels having more than two returns.

2.2 Height of saplings and subcanopy trees

For saplings, the Laplacian of the curve formed by the count per 1 m vertical bin was computed. Than a 3 m wide moving window was used to detect the strongest local inflexion (figure 4). The first inflexion detected higher than 1 m was taken as the average height of the saplings. In the field, the average height of saplings was measured only when a clear homogeneous layer could be seen, yielding 27 quadrat observations. The difference between the lidar-predicted and field height was less than a meter in 52% for the unstandardized point cloud (< 2 m in 85% of case), and 33% for the standardized version (< 2 m in 74% of case).The height of subcanopy trees was first estimated without automatic detection and delineation. For each tree, its XY position and the field measured radii in four directions were used to draw the crown perimeter and extrude a column that ran from the ground to the top of the canopy. The lidar returns within each column corresponding to the 306 single trees were extracted. For leaning trees, the column was centred on the tree apex, not on the stem XY position. The Laplacian of the curve formed by the return count per 1 m vertical bin within each column was computed. The maximum Laplacian score located between the arbitrary levels of 3 and 16 m was taken as the predicted height of subcanopy trees. In the case where more than one height corresponded to the highest Laplacian score, the lower one was retained. When compared to field-measuredheight, 39% of lidar-predicted heights differed by less than 1 m, and 58% by less than 2 m, when using the unstandardized point cloud. The standardized version did not improve the results. Suppressed trees were better detected than intermediate trees (43% of the former trees differed by less than 1 m compared to 31% in the latter case, see table 2). Suppressed hardwood and softwood trees height was estimated with the same accuracy, however, the height of intermediate softwood trees (< 1 m in 35% of case) was better estimated than of similar hardwood trees (< 1 m in 21% of case). Moreover, 46% of trees classified as well visible from above had a height difference of less than a meter, compared to 32% of occluded trees. The height of leaning trees (33% < 1 m) was not evaluated as well as that of vertical trees (41% < 1 m)

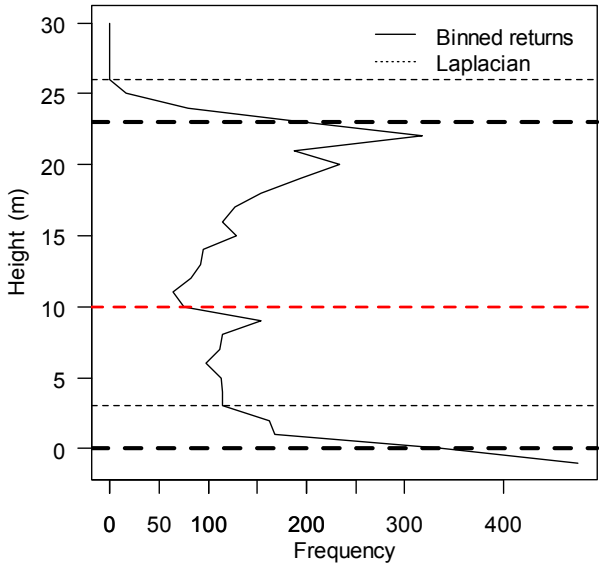


Figure 4. Canopy height profile with the local maximum inflexion detected with the Laplacian filter. The highest large black dashed lines correspond to the top of the canopy and the lowest one to the ground. The light dashed lines are intermediate inflexions and the red one corresponds to the height of subcanopy trees.

Table 2. Number and percentage of trees with an error less than 1 or 2 meters for: (a) suppressed or intermediate softwood (SW) or hardwood (HW) trees, (b) trees visible or not through the point clouds, (c) vertical or leaning trees.

(*a*)

		Suppressed			**Intermediate**			**TOTAL**		
	<=	2m	1m	total	2m	1m	total	2m	1m	total
SW	nb	70	49	114	51	31	88	121	80	202
	%	61	43	100	58	35	100	60	40	100
HW	nb	44	31	71	12	7	33	56	38	104
	%	62	44	100	36	21	100	54	37	100
TOT.	nb	114	80	185	63	38	121	177	118	306
	%	62	43	100	52	31	100	**58**	39	100

(*b*)

	High visibility			**Low visibility**		
<=	2m	1m	total	2m	1m	total
nb	52	40	87	57	34	106
%	60	46	100	54	32	100

(*c*)

	No inclination			**Inclination**		
<=	2m	1m	total	2m	1m	total
nb	133	92	227	44	26	79
%	59	41	100	56	33	100

2.3 Density of saplings and subcanopy trees

The density of saplings and subcanopy trees was estimated using a local maxima approach. Returns below certain height thresholds, i.e. within the vertical strata of subcanopy trees and saplings, were extracted. Isolated returns (within voxels having a single return) were removed. A subcanopy height model (SCHM) was created using the Inverse Distance Weighted interpolation algorithm of ArcGIS (v 9.3). The local maxima within circular windows whose radii were fixed or varied according to the height of the central pixel were identified using an in-house C++ application.

To estimate the density of saplings, the SCHM was created using the unfiltered returns within 1 and 4 m and the local maxima detected using a window of a fixed radius of 0.3 m. The number of detected local maxima per quadrat was compared to the in situ sapling stem count. The correlation between these two counts was relatively high for hardwood saplings (0.47) but much less for softwood saplings (0.11) (table 3). However, the RMS errors (in number of stems per quadrat) were quite similar (respectively of 12.21 and 14.21 trees for hardwoods and softwoods respectively).

To estimate the density of subcanopy trees, we have varied the height threshold used to extract the SCHM returns from plot to plot as a single overall threshold could not be applied due to the variability of the vertical structure of plots. Table 4 shows the correspondence between a thresholded point cloud and thresholded trees for one of the plot. First, the curve of the number of lidar returns per 1 m bin was computed at plot level and the first inflexion of this curve, which normally corresponds to the top of the sapling layer, was detected. The lower threshold was set to 1 m above this first inflexion. The high threshold was set to the second highest inflexion, which normally corresponds to the top of the subcanopy tree stratum. The points between the low and the high threshold were extracted and filtered (isolated returns were eliminated using voxel counts). The SCHM for subcanopy tree detection were then computed on the unfiltered returns, on the returns filtered based on voxels having more than one return, and those having more than two returns. Here, local maxima were detected using a variable radius window according to:

$$radius = 1.0 + 0.01 * H^{1.6} \quad (2)$$

where H is the height of the SCHM center pixel (equation 2). Due to the rather low tree counts per quadrat, results are presented only per quarter plot. The overall correlation between detected local maxima and in situ

subcanopy tree stem counts using voxels with more than one return is 0.75 (R^2 = 0.56) with a RMSE of 5 trees (table 5, figure 5). Hardwood trees were not detected as well as softwood trees. Figure 6 shows the location of detected local maxima compared to the true position of trees in the field.

Table 3. Correlation and RMSE between the per-quadrat number of local maxima and in situ sapling stem counts.

	MIN	MAX	MEAN	SD	R	RMS
LOCAL MAX	6	36	17.68	6.79	-	-
HARDWOOD SAPL.	0	60	17.52	13.89	0.47	12.21
SOFTWOOD SAPL.	0	26	6.13	5.61	0.11	14.21
SAPL.	3	68	23.66	14.72	**0.48**	14.15
SAPL. AND OPPR. TREES	5	68	25.48	14.18	0.49	14.57

Table 4. In grey are the inflexions (infl.) detected on the binned data filtered by a Laplacian (both lidar returns and trees) with a 3 m wide moving window. The uncolored rectangle is the inflexion corresponding to the ground. It's important to note that some high trees' heights are missing in the field data. That explains the low number of trees in the high strata.

	LIDAR		**FIELD**	
H (m)	Nb ret.	Infl.	Nb trees	Infl.
0	356	87	0	0
1	137	256	0	0
2	100	66	0	-8
3	71	43	8	-11
4	57	4	11	1
5	67	-60	7	4
6	117	-7	7	0
7	74	67	7	0
8	50	-18	7	-2
9	92	-30	9	5
10	80	29	2	5
11	63	-23	4	0
12	103	-111	2	4
13	174	-145	0	1
14	248	-90	1	0
15	264	34	0	-1
16	214	127	2	0
17	137	179	0	2
18	35	133	0	-1
19	4	33	1	0
20	2	2	0	1
21	2	2	0	0

Table 5. Correlation coefficients and RMSE between the local maximum and the number of subcanopy trees, subcanopy softwood and hardwood trees. Thresholds were applied to the point clouds and the correlations were calculated on all returns, on the returns filtered with voxels having more than one return, and with voxels having more than two returns.

	ALL		> 1		> 2	
	R	RMS	R	RMS	R	RMS
SUBCANOPY TREES	0.70	5.34	**0.75**	4.54	0.56	6.92
SOFTWOOD	0.74	7.54	0.69	5.14	0.56	4.70
HARDWOOD	0.14	12.21	0.31	8.90	0.16	6.51

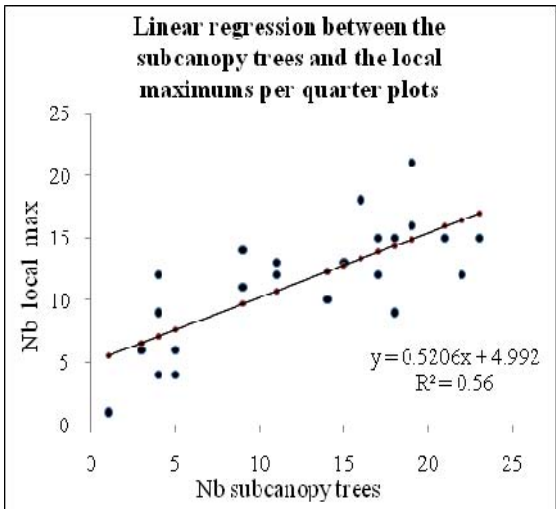


Figure 5. Linear regression

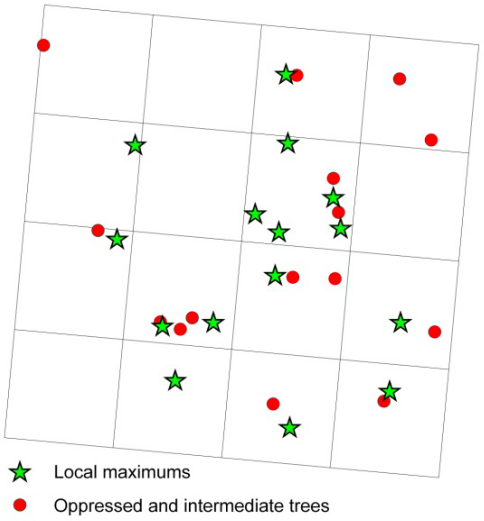


Figure 6. Real position of the understory trees compared with those detected by the local maximum.

3. Discussion and conclusion

Results show that lidar has potential for evaluating subcanopy vegetation structure. It allows the estimation of saplings and tree density below the canopy surface, as well as their individual or stratum height. The detection of sapling yielded better results than that of subcanopy trees. This may be due to the fact that saplings tend to form relatively dense and uniform layers that create a clear peak in the curve showing the number of returns per bin. Taller softwoods were in general better detected than equally high hardwoods while in lower strata, hardwoods were better characterized. It might be related to the higher proportion of softwood compare to hardwood in the superior strata (67%) or to their more compact shape. Contrary to expectations, data standardization to account for signal attenuation as it travels through the canopy did not improve results.

The estimation of subcanopy tree density based on the extraction of local maxima on filtered data yielded good results. The density of saplings was not as accurate, probably because these tend to be interspersed

and do not form typical dense convex shapes. Here again, results varied depending on strata and broad species type, with softwood trees being better assessed in higher saplings layers, and hardwood trees in lower layers. As expected, tree visibility has an impact on the results. Moreover, the success of tree and sapling characterization varies from plot to plot, specifically in the case of two plots with significantly higher levels of error.

To improve accuracy, the filter strength was adapted for each plot according to first return density. This did not however lead to a marked improvement of results but on the contrary has decreased it. This situation may be cause by a variable degree of penetration through the canopy or particular characteristics of the vertical structure.

Future amelioration of performance could come from finding an optimal binning height different than 1 m (e.g. 0.25, 0.50 m). Combining the detection of the inflexions in the curve of binned counts to the actual count value could lead to improved results as a strong inflexion close to a very low count could be a strong indication of a subcanopy layer boundary.

To improve the estimation of sapling and tree density, more sophisticated filtering approaches are being considered. In addition to suppressing isolated returns, it may be useful to also delete small isolated groups of returns. What is more, the thresholding method for extracting the part of the point cloud to be submitted to analysis could be improved by using height percentile, instead of absolute height, as a criterion. Local maxima could in addition be extracted for more than one stratum in order to detect more trees. Alternatively, an algorithm for detecting point clusters could also help detecting trees on all subcanopy layers.

Much remains to be done to achieve reliable characterization subcanopy vegetation results. However, we have demonstrated that there exists a strong subcanopy signal in the lidar point cloud that correlates to the height and density of saplings and trees. Improvements in algorithms and usage of ancillary data such as canopy surface species types, age and origin of the stands, etc. could possibly lead to results of great importance in characterizing the overall biomass of stands, the habitat diversity, and optimal sylvicultural treatments.

References

DE GRANDPRÉ, L., BERGERON, Y., NGUYEN, T., BOUDREAULT, C. and GRONDIN, P.,

2003, Composition and Dynamics of the Understory Vegetation in the Boreal Forests of Quebec. In *The Herbaceous Layer of Forests of Eastern North America*, F.S. Gilliam and M.R. Roberts (Ed.), pp. 238-264 (New York: Oxford University Press).

GEORGE, L.O and BAZZAZ, F.A., 2003, The Herbaceous Layer as Filter Determining Spatial

Pattern in Forest Tree Regeneration. In *The Herbaceous layer in Forests of Eastern North America*, F.S. and M.R. Roberts (Ed.), pp. 265-282 (New York: Oxford University Press).

HILL, R.A. and BROUGHTON, R.K., 2009, Mapping the understorey of deciduous woodland from

leaf-on and leaf-off airborne LiDAR data: A case study in lowland Britain. *ISPRS Journal of Photogrammetry & Remote Sensing*, **64**, pp. 223-233.

LEE, A., LUCAS, R. and BRACK, C., 2007, Quantifying Vertical Forest Stand Structure Using

Small Footprint Lidar to Assess Potential Stand Dynamics. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, **36** (8/W2), pp. 213-217.

MALTAMO, M., PACKALÉN, P., YU, X., EERIKÄINEN, K., HYYPPÄ, J. and PITKÄNEN, J.,

2005, Identifying and quantifying structural characteristics of heterogeneous boreal forests using laser scanner data. *Forest Ecology and Management*, **216**, pp. 41-50.

MARTINUZZI, S., VIERLING, L.A., GOULD, W.A., FALKOWSKI, M.J., EVANS, J.S.,

HUDAK, A.T. and VIERLING, K.T., 2009, Mapping snags and understory shrubs for a LiDAR-based assessment of wildlife habitat suitability. *Remote Sensing of Environment*, **113**, pp. 2533-2546.

MORSDORF, F., MÅRELL, A., KOETZ, B., CASSAGNE, N., PIMONT, F., RIGOLOT, E.,

ALLGÖWER, B., 2010, Discrimination of vegetation strata in a multi-layered Mediterranean forest ecosystem using height and intensity information derived from airborne laser scanning. *Remote Sensing of Environment*, **114**, pp. 1403-1415.

RIAÑO, D., MEIER, E., ALLGÖWER, B., CHUVIECO, E. and USTIN, S.L., 2003, Modeling

airborne laser scanning data for the spatial generation of critical forest parameters in fire behaviour modeling. *Remote Sensing of Environment*, **86**, pp. 177-186.

TEBALDINI, S., 2008, Forest SAR tomography: A covariance matching approach. In IEEE Radar

Conference, 26-30 May 2008, Rome, Italy, pp. 1-6.

WELDER, M., GEYER, R., HEINDL, B., HAHN, S. and TENHUNEN, J.D., 1996, Leaf-level Gas

Exchange and Scaling-up of Forest Understory Carbon Fixation Rates with a “Patch-Scale” Canopy Model. *Theoretical and Applied Climatology*, **53**, pp. 145-156.

WILLIAMS, M.L. and JENKINS, L.G., 2009, GeoSAR and DBInSAR: Combining “X” with “P”

for Tropical Forest “C”.*Photogrammetric Engineering & Remote Sensing*, **75**, 7, pp. 738-743.

ZIMBLE, D.A., EVANS, D.L., CARLSON, G.C., PARKER, R.C., GRADO, S.C. and GERARD,

P.D., 2003, Characterizing vertical forest structure using small-footprint airborne lidar. *Remote Sensing of Environment*, **87**, pp. 171-182.

Derivation of 3D landscape metrics from airborne laser scanning data

Werner Mücke†, Markus Hollaus† and Martin Prinz‡*

wm@ipf.tuwien.ac.at

†Institute of Photogrammetry and Remote Sensing, Vienna University of Technology, Gußhausstraße 27-29, 1040 Vienna

‡Department of Conservation Biology, Vegetation Ecology and Landscape Ecology, University of Vienna, Rennweg 14, 1030 Vienna

Abstract

One of the key-research topics in landscape ecology is the analysis and characterization of landscape pattern and structure. A description of these two features is commonly achieved through derivation of various metrics (e.g. Contagion, Dominance or Fractal Dimension) for the assessment of landscape connectivity, fragmentation and patch shape complexity.

Up to now, only very few analyses in landscape ecology have been carried out on the basis of airborne laser scanning (ALS) data, while the majority is based on either aerial or satellite based imagery supported by conventional field survey. However, airborne and space borne images exhibit a critical drawback in comparison to laser measurements. Images can not display information from below the canopy surface, as the measurement method is not able to penetrate it. Consequently, the derivation of landscape metrics from such data is merely 2D. It can not account for the vertical structure of vegetation, a key element in forestry and the assessment of structural diversity and, as such, landscape ecology.

The laser pulses, on the other hand, are able to penetrate through little gaps in the canopy surface and can provide information on the vertical and horizontal distribution of vegetation. The aim of this study is to make use of the 3D information and penetration capability of ALS for the derivation of novel landscape metrics. The presented approach exploits the collected information about the vegetation layer structure in order to describe not only if two landscape patches are connected, but how this connection is composed in terms of vertical structure of the plants building the patches. It therefore integrates knowledge of under storey or herbaceous vegetation into the shape metrics. Additionally, 3D shape metrics that relate the surface of a patch or corridor canopy surface to its enfolded volume are introduced. In this way, information about the three-dimensional interconnection of adjacent landscape patches is obtained.

1. Introduction

TransEcoNet, which stands for transnational ecological networks, is a project funded by the European Union regional development fund with the aim of elaborating strategies for development and management of ecological networks in Central Europe (TransEcoNet 2010). The background of this project is founded on the fact that protected areas, like national parks or conservation areas, are often isolated regions within a less or unprotected matrix, like intensively agriculturally used areas, traffic corridors or human settlements. Therefore, *TransEcoNet* aims at the analysis and assessment of landscapes regarding existing ecological processes, patterns and structures. Areas of significant ecological interest, so-called hot spots, are explored by traditional methods of landscape ecology concerning the state of the landscape functionality. Although *TransEcoNet* is a research project, its study areas comprise extended regions. Extensive manual field work, supported by remotely sensed imagery is carried out in order to map land use patterns, landscape connectivity, fragmentation, patch shape complexity and biodiversity indicators. A description of these features is commonly achieved through derivation of various metrics, such as Contagion, Dominance or Fractal Dimension. As a project partner, the Institute of Photogrammetry and Remote Sensing (TU Wien) is responsible for research on the use of airborne laser scanning (ALS) data for the derivation of biodiversity relevant quantities in order to support the ecological assessment.

Up to now, only very few analyses in landscape ecology have been carried out on the basis of ALS data, while the majority is based on either aerial or satellite based imagery supported by field survey. However, airborne and space borne images exhibit a critical drawback in comparison to laser measurements. Images cannot display information from below the canopy surface, as the measurement method is not able to penetrate it. Consequently, the derivation of metrics for quantitative and qualitative description of the landscape surface is merely 2D. It cannot account for the vertical structure of vegetation, a key element in forestry and the assessment of biological diversity, functionality and, as such, landscape ecology (Maier and Hollaus, 2008). ALS on the other hand, as it is an active measurement system, is able to penetrate through little gaps in the canopy surface. It can therefore provide information not only on the horizontal, but also on the vertical distribution of the vegetation.

Research papers have shown the need and potential of the integration of the third dimension into landscape structure indices (Jenness 2004, Höchstetter 2009). And while the usage of penetration rates of laser scanning data is common practice in forestry related studies (Naesset 2002, Korpela et al. 2009, Hollaus et al. 2008), their applications in landscape ecology are very rare. With this study we tend to make use of the 3D information and penetration capability of ALS for the derivation of novel landscape metrics. The presented approach exploits the collected information about the vegetation layer structure in order to describe not only if two landscape patches are connected, but how this connection is composed in terms of vertical structure of the plants building the patches. It therefore integrates knowledge of different vegetation layers into the shape metrics.

Study area and data set

The study area is located in the southern part of the Leithagebirge, a low mountain range on the borders of the federal states Lower Austria and Burgenland in Austria. The test site features a complex of semi natural and agricultural landscape composed of forest, vineyards and fields. Furthermore, patches of elongated as well as compact shape with different levels of vegetation height or under storey are present (figure 1a).

1.

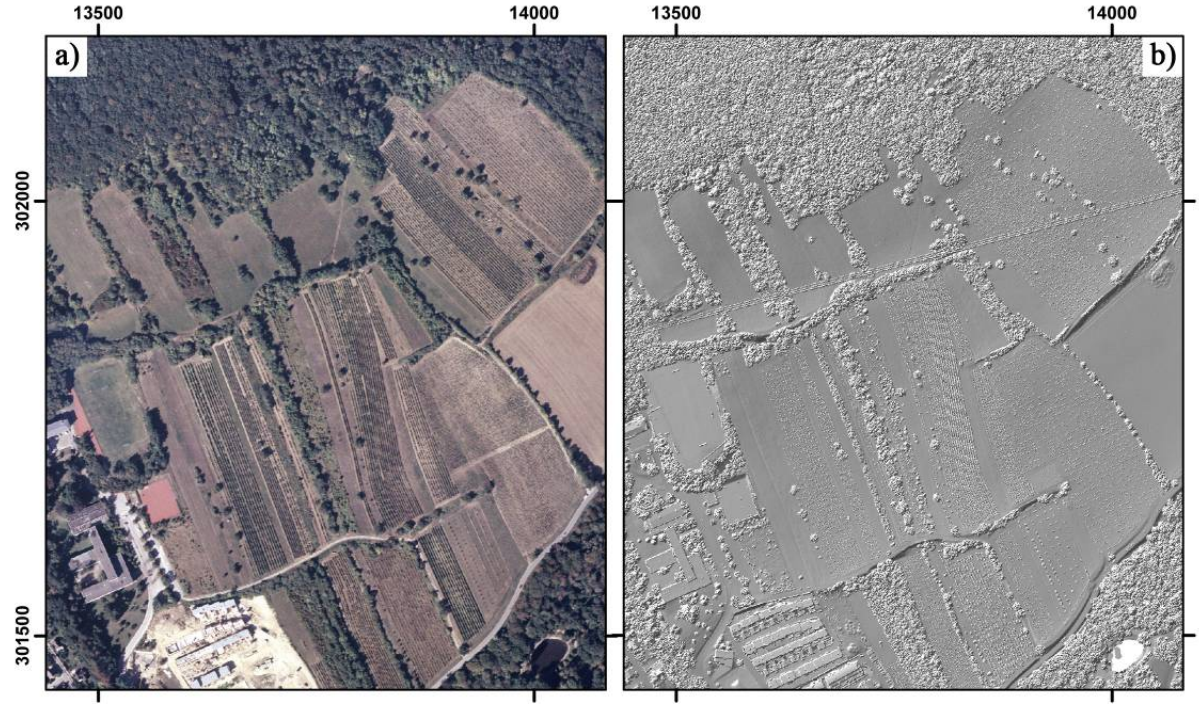


Figure 1: (a) True colour orthophoto of the study area. (b) Landscape dependent digital surface model (DSM).

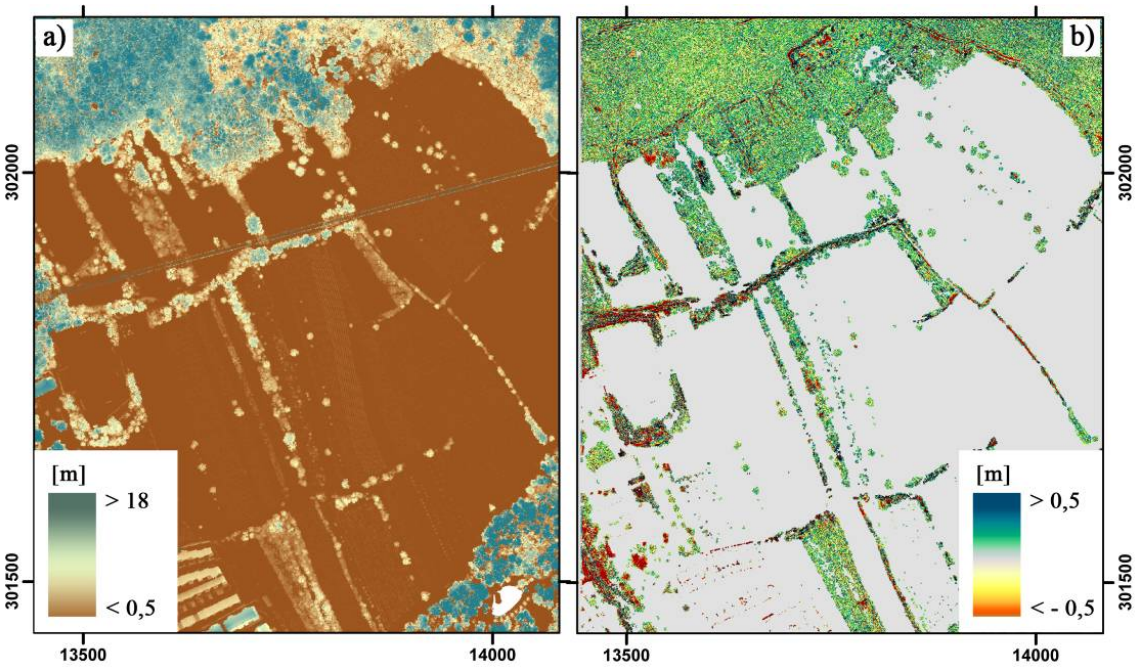


Figure 2: (a) Normalised DSM (nDSM), the colour coding represents the normalised heights. (d) difference model of nDSM and DSMdiff.

The utilised laser data were acquired in march 2007 under leaf-off conditions. The employed sensor was a RIEGL LMS-Q560 full-waveform laser scanner carried by a fixed wing aircraft. Due to a rather large overlap of the single ALS strips, a very high point density of 18 returns per m² on average was achieved. Based on the first and single returns a landscape dependent digital surface model (*DSM,* figure 1b) was calculated as described in Hollaus et al. (2010). The concept is based on the derivation and fusion of two different types of surface models dependent on the surface roughness. One model is calculated from the highest laser returns within a computing cell (DSM_{max}), the other by moving least squares interpolation (DSM_{mls}).

The latter tends to alter the corresponding object heights due to smoothing effects, which is an undesired consequence in vegetated areas. Consequently, the heights are underestimated which can lead to differences in the range of several decimetres. Therefore, the method employs the DSM_{max} only in vegetated areas and the non-vegetated areas are represented by the DSM_{mls}. A digital terrain model (*DTM*) was calculated from the last returns in the test site by hierarchical robust filtering (Kraus and Pfeifer 1998). The normalised DSM (*nDSM*) was derived as the difference of *DSM* and *DTM* (figure 2c). Subsequently, two height models, one based on the highest and one based on the lowest point in a grid cell are computed. The difference model of the two, the so-called differential *DSM* (DSM_{diff}), is expected to depict the vertical extents of the branches, needles and leaves. Compared to using the standard *nDSM*, the deviation is up to 0.5 m in height, especially in areas with less or no under storey (figure 2d). All derived height models have a spatial resolution of 0.5 m.

2. Method

2.1 Derivation of a vegetation mask

To improve performance of the methods over large areas and to reduce computing times all of the following calculations are limited to vegetated areas only. This is achieved by the application of a vegetation mask, which is derived in a multi step approach by morphological image processing based on the *nDSM* and a measure of local transparency, the so-called echo ratio (*ER*) (Höfle et al. 2009). The first step is to discriminate between elevated and non-elevated objects within the *nDSM* by simple height thresholding. As

even the lowest vegetation is interesting for our study, the height threshold is defined with 10 cm. As the result still includes artificial objects like buildings, the *ER* is employed in the second step as a distinct parameter for further separation. It basically describes whether a laser shot is able to penetrate a surface or not (figure 3a). Building areas, compared to vegetation, usually feature no penetration at all and can therefore be reliably separated with the *ER.* In the third step, the individual results are combined (figure 3b) and morphological opening with a kernel size of 3x3 pixels is applied to eliminate isolated pixels (figure 3c). The final raster layer is a binary map representing the classified vegetation pixels as 1 and everything else as 0. For further processing, the connected vegetation pixels are vectorised as shown in figure 3d.

2.2 Segmentation of vegetated areas

The aim of the segmentation step is to extract homogenous features like shrubbs, single tree crowns or sub-tree crowns for larger distributed single tree crowns (e.g. large deciduous trees). The derived segments are subsequently used as a reference unit for the calculation of structure parameters using the original 3D point cloud. Based on the *nDSM* an edge-based segmentation procedure is applied. This segmentation approach has already been tested in densely forested areas (Höfle et al., 2008, Hollaus et al. 2009) and in densely built-up urban areas (Höfle and Hollaus 2010). The main idea of the segmentation is to delineate convex objects (i.e. tree crowns) in the *nDSM* by finding concave edges between the convex objects.

The constraints for the concave edge detection are a minimum curvature in the direction perpendicular to the direction of maximum curvature and an *nDSM* threshold of again 10 cm, as we want to include also the lowest vegetation. The *nDSM* is masked with the prior computed vegetation mask and a window of 7x7 *nDSM* raster cells is employed for calculation of the curvatures. The derived potential edge areas are skeletonised to extract the final edge map, which corresponds to the segment boundaries. Finally, a connected component labelling and vectorisation of the connected region boundaries are applied to derive the segment polygons (figure 4b).

2.3 Penetration index

Diversity of structure leads to diversity of species, because different ecological niches are created (Tews et al. 2004). Not only structural diversity on the surface, but particulary in the third dimension leads to a further diversification of habitats. Especially birds and insects are species groups who benefit the most of diverse habitats (Burel F. 1992). Structural diversity usually results from any kind of disturbance (Roxburgh et al. 2004). This can either be of natural cause, like fires, windfall or damages caused by game animals, or by human activity, which is mostly agricultural or due to forest maintainance. As landscape elements are permanently changing systems, they become of high relevance for nature conservation if the changes, e.g. in structural complexity, happen to fast. In combination with a lack of stepping stones and corridors they can lead to local extinction of species.

The distribution of the laser echoes in the vegetation allows us to draw conclusions on its structural complexity. Due to the fact that the ALS data were acquired during leaf-off conditions, the penetration of the vegetation by the laser pulses is assumed to be very good. The lack of returns below the upper most branches is therefore considered to stem from the absence of under storey, rather than from occlusion by high trees. A so-called penetration index is calculated as a measure of penetrability and geometric structure.

The 3D point cloud is reduced to the vegetated areas by intersection with the vegetation mask. Further, it is divided into a terrain point cloud and a vegetation point cloud using an *nDSM* height threshold of 10 cm. Subsequently, the points are assigned to height levels, which were derived as percentage of the maximum occurring point height within a grid cell of 0.5 m. The defined levels are 0 to 33% (*L1*), 34 to 66% (*L2*) and 67 to 100% (*L3*). Then the ratio of the total number of points within a grid cell and the number of points within the same grid cell and respective height level is computed. The derived segmentation of the vegetated areas is used to accumulate the raster values to the tree crown segments as spatially better represented objects. Each segment is assigned the according values from the resulting three raster maps. Mean values for each

segment are derived and a decision tree based strategy is used to classify the segments, the results of which represents the penetration index map (figure 5).

2.4 Vegetation surface - volume ratio

As the majority of geometric analysis in landscape ecology is based on airborne or space borne imagery, conventional landscape shape metrics derived from these data sources are merely 2D. In chapter 1 the demand for a description of the three-dimensional interconnection of adjacent patches was argued.

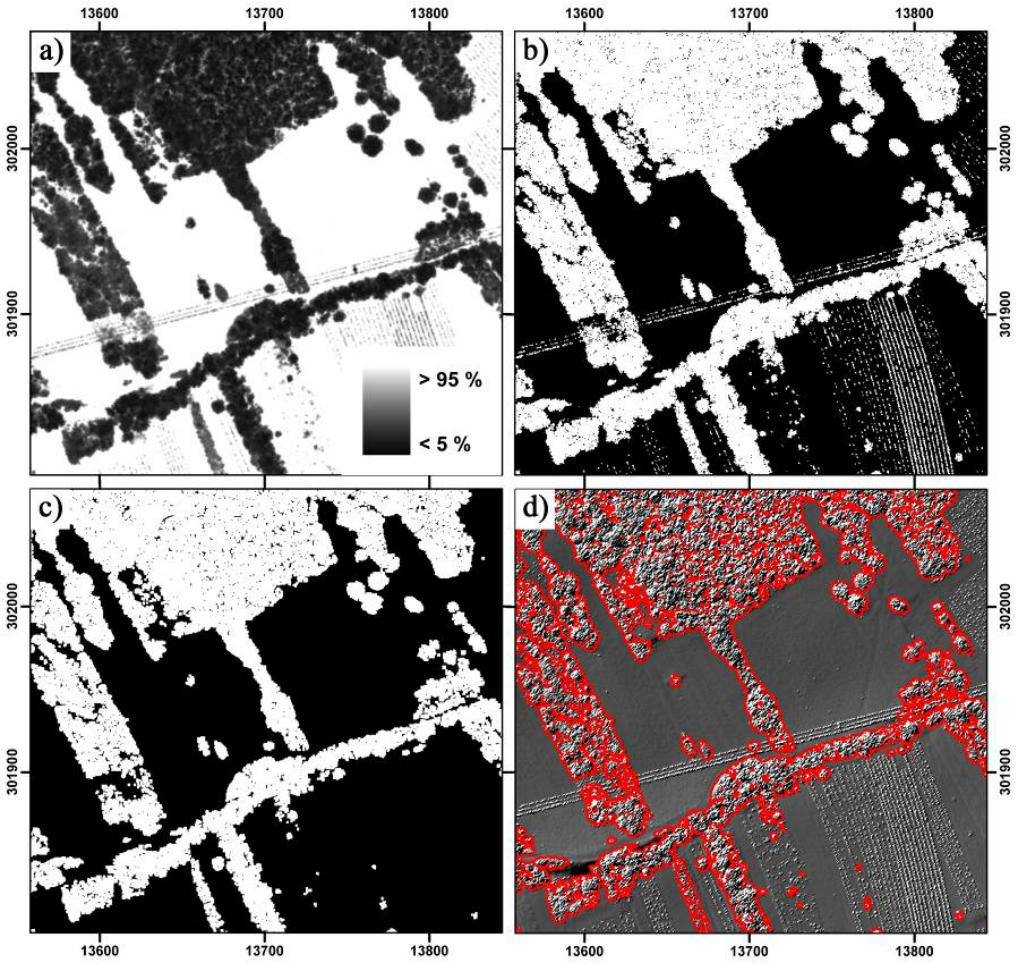


Figure 3: (a) Echo ratio (ER) in percent; darker areas indicate penetrable surfaces. (b) Binary map where nDSM > 0.10 m combined with ER > 85%; white areas represent potential vegetation regions. (c) Vegetation regions after morphological opening with 3x3 kernel. (d) Final vectorised vegetation mask. The resolution of the raster images is 0.5 m.

The main concept of the described approach is the relation of a patch's surface, defined as the area of its enveloping canopy, to the volume enclosed by it. So if a vegetation object is shaped like a sphere, meaning it is very compact, it will show a very small surface compared to its enclosed volume. This will be the case e.g. for dense hedgerows or forested areas with a sufficient presence of under storey. On the contrary, if a vegetation object is more branched, its surface will get larger compared to the volume it encloses. This results from large trees with high crown height variability and only few or even no under storey.

The surface is computed as the sum of all visible lateral faces and the top face of a cell column in the raster domain. The workflow is implemented in GRASS GIS (GRASS GIS 2010). First, the DSM_{diff} is filtered by four

different 3x3 kernels, each calculating the height difference between the center pixel and each of the four neighbouring pixels by subtraction of the neighbouring from the center pixel. Secondly, the values of the four difference raster layers are added if the difference is positive, indicating that it is a cell on the ascending part of the surface. Multiplying the respective sum of the differences with the raster resolution the area of all lateral faces of the cell columns is obtained, which is subsequently added to twice the cell area to generate the final surface. The computation of the volume is achieved by multiplication of the respective value of the DSM_{diff} and the cell size. Finally, the ratio of the surface and the volume are computed for each raster cell and assigned to the vegetation segments in the same manner as described in section 3.3.

3. Results and discussion

Vegetation mask

Reducing the area of actual computation to only the vegetated areas increased performance of the applied methods significantly and therefore efficient analysis of the reduced original point cloud was possible. As the vegetation mask was derived from raster layers of 0.5 m, it is very accurate, as can be seen in figure 3d.

However, in the case of very dense plant cover the *ER* becomes very similar for vegetation and artificial (impenetrable) objects. Under these circumstances discrimination could not be achieved, which lead to undesired holes in the vegetation mask that could only be corrected manually.

But from experience from previous studies (Höfle et al. 2008, Höfle et al. 2009, Höfle and Hollaus 2010) it can be stated that this does not occur very frequently. Still, if the results appear to be insufficiently because of this matter, of course the integration of additional information (e.g. building outlines from the cadastre) might be useful.

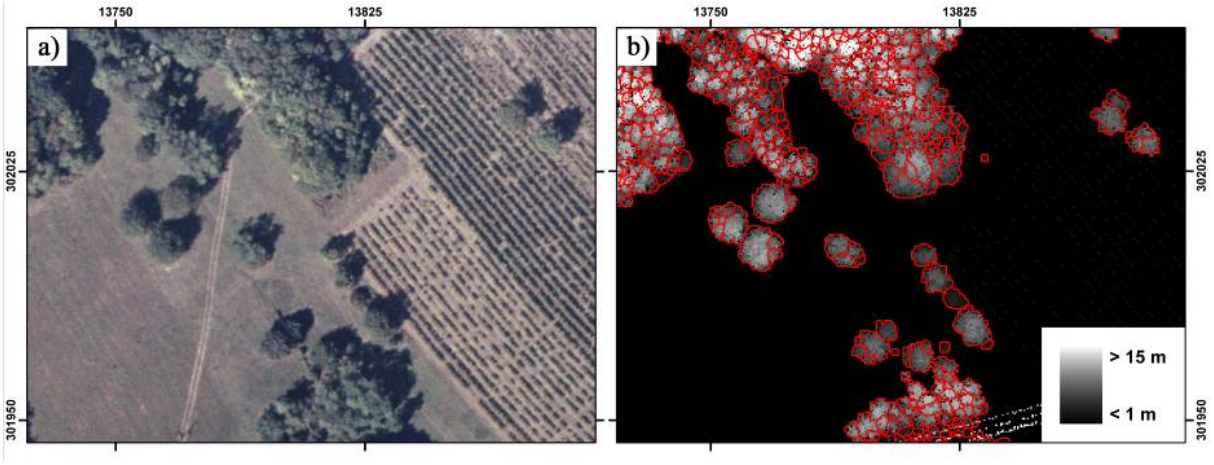


Figure 4: (a) True colour orthophoto. (b) nDSM overlaid with the result of the vegetation segmentation.

Segmentation of vegetated areas

The delineation of single trees or tree crowns in dense deciduous forests is generally difficult. As the applied segmentation algorithm detects convex objects separated by concave areas, it works very well for single trees with clearly distinct crowns, as can be seen in figure 4b. But especially older or larger deciduous trees often develop large crowns with multiple maxima which results in multiple convex areas and are therefore represented by more than one segment. A further limitation occurs in very dense young deciduous forest, characterised by a smooth canopy surface. Because of the less distinct crown shapes, the resulting segments often include multiple trees.

Penetration index

The resulting penetration index map for the study area can be seen in figure 5. For evaluation purposes, three profiles of the 3D point cloud, which are ment to display the structural diversity of a particular area, were created. In the chosen study area four dominant types of vegetation structure as described in section 3.3 could be identified: *L1 + L3* > 80% (red), *L2 + L3* > 80% (light green), *L3* > 80% (dark green) and equally distributed structure (yellow). Below the profiles the corresponding lines from the penetration index map are given. They demonstrate that the classification result corresponds very well with the actual structure of the forest. Deviations could be observed in areas with high local variations. Due to the comparably small cell size of 0.5 m these local variations are in general very well represented in the raster layer. However, this leads to a rather noisy impression in these areas and therefore the assignment to the segments was carried out. Hence, an averaging takes place which can no longer represent these inner segment variations.

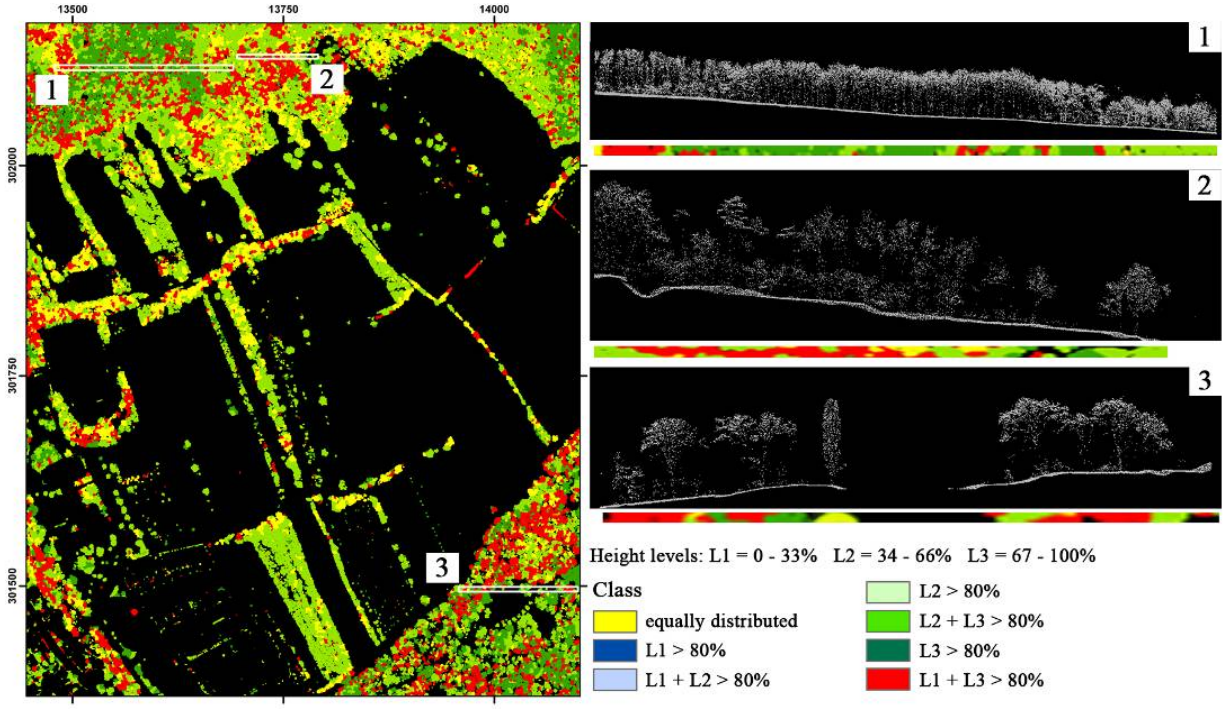


Figure 5: left: Penetration index map. Right: Profile views of the ALS point cloud. The location of the profiles can be found in the left image. Below the profiles the corresponding lines from the penetration index map are given.

Surface - Volume ratio

The vegetation surface and volume ratio can be seen as a proxy for the compactness of a particular landscape element. Changing compactness along a geometric element implies a change in structure and consequently permeability. This permeability is of significance for certain species, e.g. highly adapted birds, whose requirements do not allow structural changes within their habitats. In figure 6b the computed vegetation surface to volume ratio is shown. A high voltage powerline runs right through the study area crossing several vegetation corridors. It is clearly visible in the ratio image that the character of the vegetation structure is changing significantly below the powerline. For evaluation of the results, visual examination of the 3D point cloud had to be used, because of the lack of an adequate ground truth measurement method for the proposed surface to volume ratio. A profile view is given in figure 6c and it can be seen that the changing of the corridor vegetation character, as indicated by the ratio, is supported by the 3D point cloud. In this case the power line acts as a natural barrier, which is a disturbance in this particular habitat or corridor and can lead to a decrease of migration and, subsequently, inbreed and extinction.

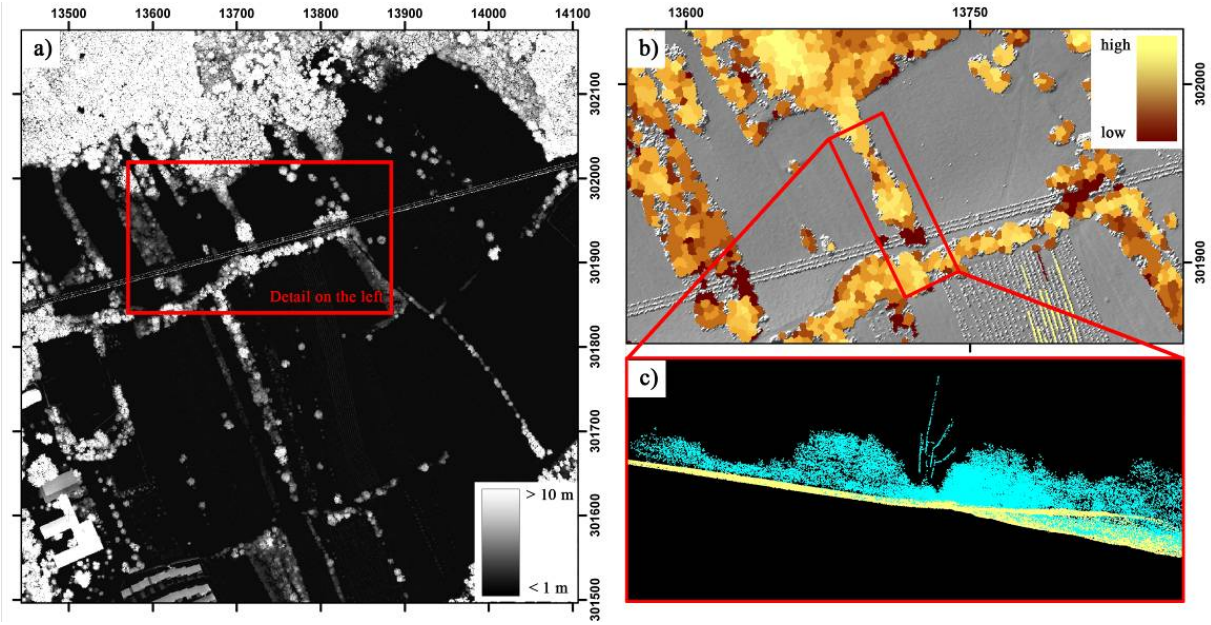


Figure 6: (a) color coded nDSM (b) surface to volume ratio (c) profile of point cloud showing the differing character of vegetation in particular area.

As an alternative method for the calculation of canopy surface from digital height models Jenness (2004) gave a recommendation based on triangulation. He stated that, compared to simple raster based area calculations, his method was more accurate. However, the resolution of the employed raster models was 90 m, so by far less than what is achievable with modern small footprint ALS systems, as it was used in this study. We do not expect the difference between raster and triangulation based surface calculation to be significant with a grid cell size of 0.5 m.

4. Conclusions

The penetration of the canopy surface by the laser pulses provides knowledge which is of great significance. It could be shown that 3D shape metrics derived from data acquired with ALS are a suitable indicator for the evaluation of structural diversity in vegetated areas. The hereby produced maps highlight vegetated areas with less interconnection and can help with the identification of natural barriers and can be a valuable addition to the textural information of orthophotos in traditional field work.

Acknowledgements

This study was funded by the TransEcoNet project implemented through the CENTRAL EUROPE Program co-financed by the ERDF.

References

BUREL, F., 1992, Effect of landscape structure and dynamics on species diversity in hedgerow networks. Landscape Ecology, 6:161–174

GOETZ, S.J., STEINBERG, D., BETTS, M.G., HOLMES, R.T., DORAN, P.J., DUBAYAH, R. AND HOFTON, M., 2010, Lidar remote sensing variables predict breeding habitat of a Neotropical migrant bird. Ecology. 91:1569-1576.

GRASS GIS, 2010, Available online at http://grass.itc.it/ (accessed 07/2010).

HOLLAUS, M., MÜCKE, W., HÖFLE, B., DORIGO, W., PFEIFER, N., WAGNER, W., BAUERHANSL, C. AND REGNER, B., 2009, Tree species classification based on full-waveform airborne laser scanning data";Talk: Silvilaser 2009, College Station, Texas, USA; 10-14-2009 - 10-16-2009; in: "Proceedings of SILVILASER 2009", ISBN: 978-1-61623-997-8; 9 pages.

HOLLAUS, M., MANDLBURGER, G., PFEIFER, N. AND MÜCKE, W., 2010, Land cover dependent derivation of digital surface models from airborne laser scanning data. International Archives of Photogrammetry, Remote Sensing and the Spatial Information Sciences. PCV 2010, Paris, France. Vol. 39 (3), 6.

HÖCHSTETTER, S., 2009. Enhanced methods for analysing landscape structure: Landscape metrics for characterising three-dimensional patterns and ecological gradients. Rhombos Verlag. ISBN: 3941216139.

HÖFLE, B., HOLLAUS, M., LEHNER, H., PFEIFER, N. AND WAGNER, W., 2008, Area-based parameterization of forest structure using full-waveform airborne laser scanning data";Talk: Silvilaser 2008, Edinburgh, Scotland; 09-17-2008 - 09-19-2008; in: "Proceedings of SilviLaser 2008 8th international conference on LiDAR applications in forest assessment and inventory", (200, ISBN: 978-0-85538-774-7; 9 pages.

HÖFLE, B., MÜCKE, W., DUTTER, M., RUTZINGER, M. AND DORNINGER, P., 2009. Detection of building regions using airborne lidar: A new combination of raster and point cloud based GIS methods. GI_Forum 2009 – International Conference on Applied Geoinformatics, Salzburg.

HÖFLE, B. AND HOLLAUS, M., 2010, Urban vegetation detection using high density full-waveform airborne lidar data – combination of object-based image and cloud analysis. In: W. Wagner and B. Székely (Editors), ISPRS TC VII Symposium - 100 Years ISPRS, Vienna, Austria, pp. 281-286.

JENNESS, J.S., 2004, Calculating landscape surface area from digital elevation models. Wildlife Society Bulletin, Vol. 32, No. 3, pp. 829-839.

KRAUS, K. AND PFEIFER, N., 1998, Determination of terrain models in wooded areas with airborne laser scanner data, ISPRS Journal of Photogrammetry and Remote Sensing, Volume 53, Issue 4, Pages 193-203, ISSN 0924-2716, DOI: 10.1016/S0924-2716(98)00009-4.

KORPELA, I., KOSKINEN, M., VASANDER, H., HOLOPAINEN, M. AND MINKKINEN, K., 2009, Airborne small-footprint discrete-return LiDAR data in the assessment of boreal mire surface patterns, vegetation, and habitats, Forest Ecology and Management, Volume 258, Issue 7, Pages 1549-1566, ISSN 0378-1127, DOI: 10.1016/j.foreco.2009.07.007.

MAIER, B. AND HOLLAUS, M., 2008, Waldstrukturerfassung mittles Laserscanning im Schutzwald, Die kleine Waldzeitung (invited), 3 (2008), 9 - 11.

NAESSET, E., 2002, Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data, Remote Sensing of Environment, Volume 80, Issue 1, Pages 88-99, ISSN 0034-4257, DOI: 10.1016/S0034-4257(01)00290-5.

ROXBURGH, S. H., SHEA, K. AND WILSON, J.B., 2004, The intermediate disturbance hypothesis: patch dynamics and mechanisms of species coexistence. Ecology 85:359–371.

TEWS, J., BROSE, U., GRIMM, V., TIELBÖRGER, K., WICHMANN, M.C., SCHWAGER, M. & JELTSCH, F. 2004, Animal species diversity driven by habitat heterogeneity/diversity: the importance of keystone structures. Journal of Biogeography 31: 79–92.

TRANSECONET, 2010. Available online at: http://www.transeconet.eu/ (accessed 07/2010).

comparing canopy penetration of repeated als ACQUISITIONS

SVEIN SOLBERG[*]*, ERIK NÆSSET*[§] *and HOLGER LANGE*[*]

svein.solberg@skogoglandskap.no

*Norwegian Forest and Landscape Institute, P.O. Box 115, N-1431 Ås, Norway

§Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences, N-1432 Ås, Norway

Abstract

The airborne laser scanning (ALS) penetration rate, i.e. the ratio of ground echoes to total echoes, is a proxy for gap fraction. Hence, ALS has a potential for monitoring forest properties that are related to gap fraction, such as leaf area index, canopy cover and disturbance. Furthermore, two gap types may be distinguished: While a pulse that only produces a ground echo most likely hit a large, between-tree gap, a pulse that produces a ground echo as the last of several returns most likely hit smaller, within-canopy gaps. This may be utilized to distinguish between disturbance types such as defoliation and tree removal. However, the ALS penetration rate needs to be calibrated with gap fraction measurements in the field, because it is influenced by technical properties of the acquisition. The aim of this study was to quantify the magnitude of this influence, by comparing repeated acquisitions with different technical specifications. We had at hand 12 ALS acquisitions which could be combined into six pairs, from four spruce and pine dominated forests in Norway. We established 20x20 m grids, and for each grid cell we extracted three penetration variables: first echo penetration, last-of-many echo penetration, and total (i.e., first and last echo). We log-transformed the penetration variables (P_1 and P_2) from two laser acquisitions, and fitted the no-intercept, linear model $\log(P_1) = \beta \log(P_2)$, applying total least squares regression analysis. In a majority of the cases, the penetration variables were very similar, i.e. they deviated by <10%. For the first echo penetration the slopes varied from 0.87 to 1.07 and the R^2 values ranged between 0.91 and 0.99. For the last-of-many echo penetration, there was generally weaker correspondence with slopes varying from 0.78 to 1.02, and R^2 values ranging from 0.60 to 0.94. Finally, for the total penetration there was again stronger agreement with slopes in the range 0.83-1.03 and R^2 values from 0.88 to 0.99. In conclusion, it seems that the penetration ability of different ALS scans in many cases are very similar, and further research may reveal ranges of standardized settings for which field inventory can be redundant.

1. Introduction

The near-vertical pulses from airborne laser scanning (ALS) penetrate into the forest canopy layer, and the magnitude of this penetration is closely related to canopy properties such as gap fraction and canopy density. The penetration also affects any other metric derived from ALS over a forest, such as the vertical height distribution of echoes which is used to estimate tree height and standing volume.

In an ALS based forest health, or disturbance, monitoring, the strong and linear relationship between ALS penetration and effective leaf area index (LAI_e) can be utilized where LAI_e serves as an operational variable for defoliation (Solberg et al. 2006, Richardson et al. 2009):

$$LAI_e = \beta \ln(P^{-1}) \quad , \qquad (1)$$

where β is a parameter to be calibrated for each acquisition, and P is an ALS penetration variable. Here, the estimated calibration factor β can be split into two factors, i.e. $\beta = \beta_1 \cdot \beta_2$. The former, β_1, varies with physical properties of the trees, mainly foliage angle distribution, and will vary mainly in space and remain fairly constant within a tree species and hence also over time for any

given place. The latter, β_2, varies with the ability of the laser pulses to penetrate the canopy, i.e. the technical properties of the acquisition, which in a monitoring program may vary over time along with changes in sensor, flight altitude, pulse repetition frequency (PRF) and other factors. Hence, in a monitoring scheme with multi-temporal data sets it is in particular the variation in β_2 that needs to be accounted for through the calibration procedure. State-of-the-art methods for calibration utilize point-based, hemispherical measurements of light transmission through the canopy using either the LAI-2000 Plant Canopy Analyzer or hemispherical photography (Solberg et al. 2009). One problem of these methods is that the random errors are about the same magnitude as the typical variation we have experienced in β_2. In addition to technical problems such as light scattering in the canopy, random errors are significant because only a moderate number of measurements can be performed: they are time-consuming and sometimes difficult e.g. due to inappropriate weather conditions. Thus, it is typically difficult to determine whether an observed change in β_2 for a set of repeated acquisitions represents a real change or simply results from erroneous field measurements.

It appears that the understanding of how technical acquisition properties affect ALS penetration is limited, and that the penetration rate and related forest metrics vary relatively little across different acquisition settings (mostly <10%). Hopkinson (2007) suggested that canopy penetration increases with decreasing "peak pulse power concentration", defined as peak pulse power divided by footprint area. The apparent explanation was that with decreasing peak pulse power concentration larger foliage backscatter areas would be needed in order to raise the backscatter energy above a threshold required for pulse detection. This would mean that if all other properties were kept constant, canopy penetration would increase with increasing footprint size or decreasing peak pulse power. This is apparently contradictory to Lovell et al. (2003) and Morsdorf et al. (2006) who argued that ALS pulses often have too large footprints to be able to penetrate small canopy gaps, i.e. increasing the footprint size would reduce canopy penetration. Effects of pulse repetition frequency, flying altitude, footprint size and scan angle on various tree metrics have been found, these effects have generally been moderate or small (Nilsson 1994, Næsset 2004; 2009, Holmgren et al. 2003, Yu et al. 2004, Chasmer et al. 2006, Goodwin et al. 2006). However, a general limitation of the approach in these studies is the use of field inventory data: While the acquisitions themselves produce extremely large data sets which should give very small mean random errors only a small fraction of these ALS data is actually used in the analyses, those from the location of the field samples. Another problem is that sometimes it is not explicitly stated in the literature what type of ALS penetration it is referred to.

The idea with this study was to compare ALS penetration rates directly by overlaying pairs of ALS data sets gathered from repeated acquisitions. In this way, we have a novel approach as compared to existing studies. Rather than involving field measurements and the correspondingly small fractions of ALS data, we utilize entire ALS data sets. In this way systematic differences in penetration rate can be accurately quantified. In addition, we studied three alternative penetration variables separately.

2. Materials and methods

ALS data were obtained from four study areas in conifer forest areas in Norway dominated by Norway spruce (*Picea abies*) and Scots pine (*Pinus silvestris*). The study areas were covered by repeated ALS acquisitions, where the repetitions had either different sensors, different flight altitudes or different pulse repetition frequencies. From these acquisitions we generated six pairs of ALS data.

We overlaid grids over the areas, and derived ALS penetration rates for each grid cell. Preliminary analyses demonstrated that 20x20 m grid resolution was appropriate, i.e. being little influenced by random errors which tend to increase with decreasing grid cell size. Hence, we used this spatial resolution throughout the study.

For each grid cell the penetration rate was derived by pulse counting, after the echoes were classified as either canopy echo or ground echo. Echoes having a height of more than 1 m above the ground were assigned as 'canopy echo', while echoes below that threshold were assigned as 'ground echo'. We studied three alternative penetration rates: (1) P_F = the fraction of pulses that produced a ground echo as the first echo; (2) P_L = the fraction of pulses that produced a ground echo as the last of two or more echoes; and (3)

P_{tot} ('total penetration') = P_F + 0.5P_L. Here the value 0.5 is arbitrarily set, and represent the idea that a pulse that produce a canopy-first echo and then a ground-last echo has hit a canopy part with a 50% gap fraction.

The data sets provided, however, did not contain information on individual pulses, but only individual echoes. Thus, we could not establish the relationship between subsequent echoes for individual pulses. The number of pulses was therefore estimated, and we used the following formulas:

$$P_F = \frac{N_{gS} + N_{gF}}{N} \quad \text{and} \quad P_L = \frac{N_{gL} - N_{gF}}{N} \quad . \qquad (2)$$

Here N_{gS} denotes the number of single echoes at the ground, N_{gF} the first-of-many echoes at the ground (extremely few), N_{gL} the number of last-of-many echoes at the ground, and finally *N* the total number of pulses. Some of the acquisitions were accomplished with the Optech ALTM 1210 and 1233 sensors which have separate recorders for first and last echoes. In these cases N_{gS} was always zero, and N_{gF} was the number of first echoes at the ground. The number of pulses was estimated as $N = N_S + 0.5(N_F + N_L)$, where N_S was the number of single echoes (not relevant for ALTM 1210 and 1233), N_F was the number of first or first-of-many echoes, and N_L was the number of last or last-of-many echoes. This formula should work for both types of sensors, and should also be robust against problems when the number of first and last echoes was not identical.

To model the relationship between the penetration rates of repeated ALS acquisitions, we consider that any penetration variable should be 0 for a theoretical, opaque canopy, and 1 for bare ground (e.g. a clear-cut), respectively. Hence, if two laser acquisitions produce different penetration rates, then the relationship between them should be curvilinear and has to include the two extremal points (0,0) and (1,1), and interpolate between those. An intuitive model respecting this constraint is:

$$P_1 = P_2^{\gamma} \qquad (3)$$

where P_1 and P_2 are the penetration rates of two ALS acquisitions to be compared, and the parameter γ represents their relationship and corresponds to the β_2 parameter discussed in the introduction. We log-transformed this into the linear, no-intercept model

$$\ln(P_1) = \gamma \cdot \ln(P_2) \qquad (4)$$

which we fitted with total least squares analysis (TLS). Rather than minimizing the squared vertical distances between the data points and the fitted line as in ordinary least squares (OLS), the TLS minimizes the squared slant distances between the data points along a line perpendicular to the fitted line. The fitting was accomplished by using an iterative procedure with the Newton-Raphson optimization in the PROC NLP procedure of the SAS software package (Anon. 2009).

3. Results

In the six comparisons the slope parameter varied from 0.78 to 1.07 (Table 1). The random errors were apparently small and the coefficient of determination (R^2) varied from 0.62 to 0.99, the majority being above 0.9. The penetration rate of first echoes had the strongest agreement, where 0.87 was the value for γ that deviated the most from 1.0. The R^2 values here were 0.91 or higher. The penetration rate of the last-of-many echoes showed the weakest agreement, containing parameter estimates varying from 0.78 to 1.01 and R^2 values down to 0.60. It should be noted also that the three last comparisons had a time difference, and although the vegetation changes are likely to be small they cannot be completely excluded.

Table 1. Results of no-intercept total least squares regression analyses of Eq. 4. The parameter estimate of γ is given, as well as the coefficient of determination (R^2, parenthesized) for the $P_1 = P_2^{\gamma}$ version of the model. 20x20 m grids were used. The acquisitions differed with regard to sensor (Optech ALTM), altitude and pulse repetition frequency (PRF)

Data pair	Sensor	Altitude, m a.g.l	PRF, kHz	P_F	P_L	P_{tot}	Time difference
1	-	-	100 vs 50	1.07 (0.99)	0.78 (0.90)	0.91 (0.99)	No
2	-	2000 vs 1100	-	1.02 (0.99)	1.00 (0.92)	0.98 (0.98)	No
3	-	530 vs 840	-	0.99 (0.91)	1.01 (0.62)	0.99 (0.97)	No
4	1233 vs 3100	-	33 vs 50	0.87 (0.95)	0.86 (0.76)	0.83 (0.94)	9 months
5	1233 vs 3100	600 vs 750	33 vs 100	0.98 (0.93)	0.89 (0.60)	0.93 (0.88)	22 months
6	-	1000 vs 1500	94 vs 78	1.01 (0.98)	1.02 (0.94)	1.03 (0.98)	2 months

In the third case there was a problem with the estimation of the number of pulses. Along one edge of the scanned area, there were considerable differences between the number of first and last echoes. The edge corresponded to maximum scan angle, many last echoes ended up outside the area of interest, and these last echoes had apparently been discarded in the preprocessing by the operator. For example, some grid cells contained five times more first echoes than last echoes. We recalculated the penetration variables of this study area using alternative ways of estimating the number of pulses, and obtained somewhat different results: For first echoes γ=1.00 (R2=0.99); for last echoes γ=1.00 (R^2=0.72); and for total penetration γ=1.01 (R^2=0.78).

4. Discussion

The agreement in penetration rate for different acquisitions is influenced by both systematic and random factors. In several cases the systematic difference between penetration rate of different acquisitions was negligible, and in a majority of the cases (13 out of 18) it was less than 10%. In general, the differences were largest for last echo penetration. To some extent the differences could be explained as artefacts, i.e. when pulses were coming at an angle against vertical along the border of the study areas the estimation of the number of pulses was very uncertain.

The systematic differences observed here correspond to the differences in the β parameter in the effective leaf area index model, i.e. $LAI_e = \beta \cdot \ln(P^{-1})$. Hence, this study demonstrates that the β parameter in most cases would vary within ±10%. If we assume that the range of technical acquisition settings in the present study is representative for those commonly used in operational ALS campaigns today, then calibration with field measurements would need to be quite accurate in order to not erroneously introduce systematic errors. For some applications, it might be that field calibration are not mandatory comparing the results of the acquisitions directly, as long as the acquisitions are carried out within a given range of settings. However, more studies are needed before that can be recommended, and in particular the effects of some proprietary producer settings in echo detection algorithms are unknown and should be better understood.

The study clearly shows that it is necessary to specify what type of penetration variable (first echo, last-of-many echoes or total penetration) is used when discussing the effects of technical acquisition properties. The results for these different penetration variables are often different. There was a general tendency that the penetration rate of the first echoes was oppositely related to the penetration rate of the last-of-many echoes. If there is a high penetration rate of first echoes, it means that many of the pulses create a ground echo only, leaving few multiple-echo pulses and hence few pulses that could produce a last-of-many echo penetration. However, this means that if we use total penetration rate in an application, then these two oppositely related properties will tend to cancel each other out. This is seen in the comparisons of the total penetration rates, where the parameter estimates were generally close (0.83-1.03).

Acknowledgements

We acknowledge the Research Council of Norway for funding parts of this study through the REMFOR project.

References

ANON. 2009. Proc NLP, SAS/OR(R) 9.2 User's Guide: Mathematical Programming, 2009. http://support.sas.com/documentation/

CHASMER, L., HOPKINSON, C. AND TREITZ, P., 2006. Investigating laser pulse penetration through a conifer canopy by integrating airborne and terrestrial lidar. *Canadian Journal of Remote Sensing* 32(2), pp. 116–125.

GOODWIN, N. R., COOPS, N. C., & CULVENOR, D. S. (2006). Assessment of forest structure with airborne LiDAR and the effects of platform altitude. *Remote Sensing of Environment*, 103, 140-152.

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

HOPKINSON, C. 2007 The influence of flying altitude, beam divergence, and pulse repetition frequency on laser pulse return intensity and canopy frequency distribution, *Canadian Journal of Remote Sensing* 33 (2007), pp. 312–324.

LOVELL JL, JUPP DLB, CULVENOR DS, COOPS NC (2003) Using airborne and ground-based ranging LiDAR to measure forest canopy structure in Australian forests. *Canadian Journal of Remote Sensing* 29: 607-622.

MORSDORF, F., FREY, O., MEIER, E., ITTEN, K. AND ALLGÖWER, B., 2006. Assessment on the influence of flying height and scan angle on biophysical vegetation products derived from airborne laser scanning. In: 3d Remote Sensing in Forestry, 14-15. Feb. 2006, Vienna, Austria.

NÆSSET, E., 2004: Effects of different flying altitudes on biophysical stand properties estimated from canopy height and density measured with a small-footprint airborne scanning laser. *Remote Sensing of Environment*. 91:243-255.

NÆSSET, E. 2009. Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. Remote Sensing of Environment, 113: 148-159.

NILSON, T. (1994). Estimation of tree heights and stand volume using airborne Lidar systems. Report 57. Swedish University of Agricultural Sciences.

RICHARDSON, J.J., MOSKAL, L.M., & KIM, S.-H. (2009). Modeling approaches to estimate effective leaf area index from aerial discrete-return LIDAR. *Agricultural and Forest Meteorology*, 149, 1152–1160.

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.

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.

YU, X., HYYPPÄ, J., KAARTINEN, H. AND MALTAMO, M., 2004. Automatic detection of harvested trees and determination of forest growth using airborne laser scanning. *Remote Sensing of Environment* 90, pp. 451–462.

Scale-specific effects of disturbance and environment on vegetation patterns of a boreal landscape

Running title: Scale-specific pattern of vegetation

UDAYALAKSHMI VEPAKOMMA*, MARIE- JOSÉE FORTIN

udayalakshmi.vepakomma@utoronto.ca

Department of Ecology & Evolutionary Biology, University of Toronto,

25 Harbord Street Toronto, Canada.

Abstract

An array of ecological factors like available resources, disturbance regimes at varying spatio-temporal scales and environmental conditions, influence forest succession and alter the resulting landscape patterns. Disturbance regimes create spatio-temporal patterns of tree mortality and thus a mosaic of forest ages. Environmental factors such as elevation, slope, aspect, soils, and vegetation edges may increase spatial variability in disturbance regimes, which may then translate into spatial patterns of forest vegetation. Insights into such ecological processes and their interactions are vital in setting conservation targets, assessing current managed systems, or in making ecologically informed management decisions. Thus the main objectives of our study are (i) to investigate the relationships between vegetation patterns of a boreal landscape and its disturbance history, and (ii) to test whether these relationships contrast with environmental heterogeneity. We also would like to examine whether the vegetation responses to the interplay of disturbances and environment vary as a function of the spatial scale of observation. We used recent lidar and imagery of parts of North American mixedwood boreal forest to develop accurate generalisations across multiple scales across the landscape, spatial patterns of disturbances, environmental heterogeneity, structure and composition of vegetation. The forest responses at multiple resolutions were investigated using wavelet analysis.

1. Introduction

Ecological processes such as disturbances occurring over a hierarchy of spatial and temporal scales structure the terrestrial ecosystems. In boreal forests, disturbance regimes create spatio-temporal patterns of tree mortality and are considered as a dominant force structuring a complex mosaic of forest ages (Pickett and White 1985). Disturbances increase available resources, alter micro-climate and create new establishments. Environmental factors such as elevation, slope, aspect, soils, and vegetation edges may increase spatial variability in disturbance regimes, which may then translate into spatial patterns of forest vegetation. In turn, variability in vegetation structure affects the spread of disturbances and influences the flow of energy, matter and composition within and between systems (Forman and Gordon 1986). Resulting heterogeneity affects also many other aspects of ecological dynamics such as foraging behaviour and wildlife habitat. However, a question of interest for forest dynamics that still remains is how these multiple factors combine to drive succession and produce vegetation patterns at various scales.

Ecological theories have proposed role of environment (Whittaker 1953) or disturbance as dominant force (Pickett and White 1985) in structuring forest communities. Studies also have showed changes in composition as a function of topographic gradient (del Moral and Watson 1978). However, very few studies compared multiple ecological factors, which suggest that the dominant factors controlling vegetation patterns vary with the extent of study area and manner in which the study is characterised. Yet, ecological relationships between pattern and process may occur along a continuum of scales. Although a particular scale of interest would be informative when investigating a system, there would be no single correct scale of analysis (Levin 1992). Insights into how process, pattern and their relationship change with scale of observation are vital in setting conservation targets, assessing current managed systems, or in making ecologically informed management decisions.

Understanding heterogeneity across scale can only be achieved through identification of the dominant scales at which processes are expressed and the association of processes with variables describing landscape structure. This requires substantial amount of empirical data collected at multiple scales. Due to limitations in the available tools and methods, examinations of pattern-process relationships are often conducted at coarse scale resolutions. Remote sensing is a potential alternative to field based measurements across multiple spatial scales (e.g., tree, plot to landscape or region) and multiple time intervals. Lidar is a helpful tool that can provide required level of details to the scale in question from very fine, of a tree, to broad scale, at a landscape, on forests, that it is now increasingly drawing the attention of ecologists and natural resource managers (Lim et al. 2003, Lefsky et al., 2004). To perform a multiscale analysis with lidar data, wavelets can be used. Wavelets are useful scalable method of spatial pattern analysis for raster data that can be used to summarize pattern at multiple scales (Keitt and Urban 2005). Wavelets can be used to decompose data into its scale-specific components and associate portions of total signal variance with those scales (Csillag and Kabos 2002).

Thus the main objectives of our study are (i) to investigate the relationships between vegetation patterns of a boreal landscape and its disturbance history, and (ii) to test whether these relationships contrast with environmental heterogeneity using lidar data in combination with ancillary data. We also examine whether the vegetation responses to the interplay of disturbances and environment vary as a function of the spatial scale of observation. Our study area is a relatively broad extent, nearly 80 km^2, of North American mixedwood boreal forest (Quebec, Canada). To aid in developing accurate generalisations across multiple scales across the landscape, spatial patterns of disturbances, environmental heterogeneity, structure and composition of vegetation were created at a fine grain using recent lidar data and high resolution imagery. The forest responses at multiple resolutions were investigated using wavelet analysis.

2. Methods

2.1 Site description

Located in the Boreal Shield Ecozone (approx. 48°30'N, 79°22'W; Quebec, Canada), the Lake Duparquet Research and Teaching Forest (LDTRF) extending in approximately 80 km^2 area is a forest mosaic of old-growth and matured stands, originated from fires occurred during 1760 – 1944, and a combination of insect (spruce-budworm), management intervention (clear or partial cuts) and other small-scale disturbances like canopy gaps (Figure 1). Largely composed of mixedwood

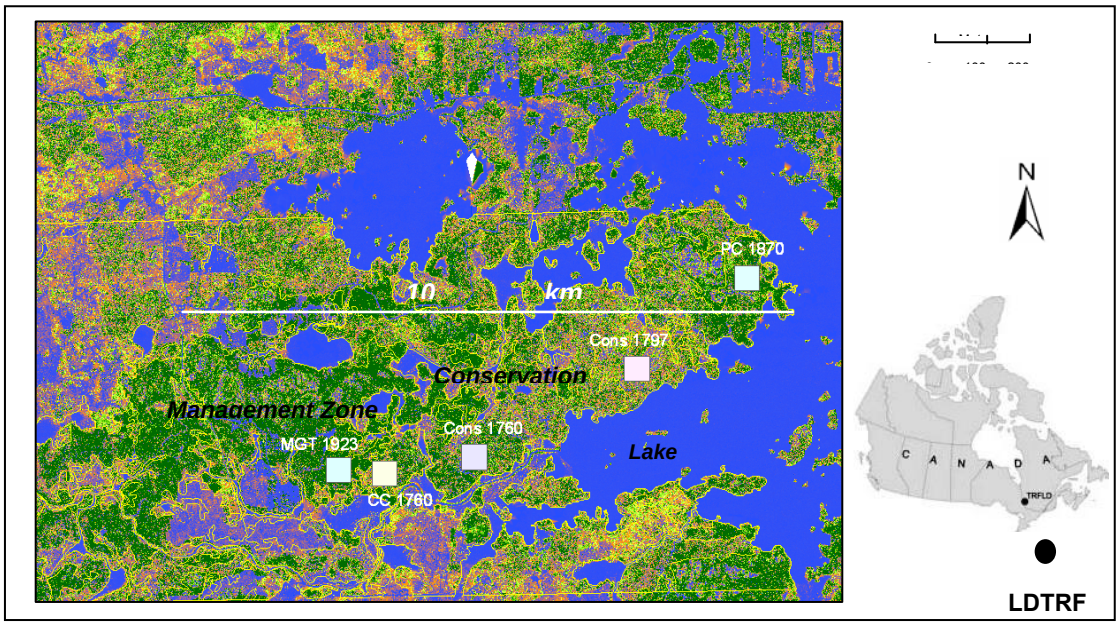


Figure 1. Location of the Lake Duparquet Teaching and Research Forest (LDTRF), Quebec, and the 10 km. transect and study windows representing stands originated from different disturbances seen overlaid on the lidar Canopy Height Model (CHM) where purple to dark green indicate lowest to the highest canopy height.

stands of height lower than 30 m, these forests have most common boreal tree species: jack pine, white and black spruce, balsam fir, eastern white cedar, white birch, trembling aspen, balsam poplar and eastern larch. The area is characterized by small hills with elevations varying between 227 m and 386 m. The climate is subpolar, subhumid, and continental with an average annual temperature of 0.8°C and annual precipitation of about 857 mm. The frost-free period lasts for nearly 64 days, while the length of the growing season averages 160 days (Environment Canada 1993).

About 25% of this 80 km^2 research forest is designated as conservation zone, while the rest is devoted to developing, applying and demonstrating various management interventions. In this study we considered a 10 m × 10 km transect across the LDRTF, and five sample windows of 512 m × 512 m each representing stands with different disturbance history (*viz*., (i) 1760 fire (243 yr Time Since Last Disturbance, TSLD stand) (ii) 1797 fire (206 yr TSLD) and (iii) 1870 fire with partial cuts within the Conservation Zone (133 yr TSLD) (iv) 1760 fire with an intervention of clear cut in 1982 (20 yr TSLD) (v) 1923 fire with clear cut in 1992 (10 yr TSLD).

2.2 Data

Vegetation structure, species groups and topographic variables

The lidar survey was carried out from 14 to 16 August 2003 using Optech's ALTM2050 lidar flown at 1,000 m AGL, such that the first and last returns were recorded for each pulse, with a 50% overlap between adjacent swaths. Table 1 presents the key survey and lidar instrument parameters. The data were registered to high-end carrier phase differential GPS ground profiles and the inter-swath geometrical fit was improved using the TerraMatch algorithm (Terrasolid Ltd.,

Helsinki). The last returns were classified as ground or non-ground using the Terrascan algorithm (Terrasolid) followed by manual verification and editing. The Digital Surface Model (DSM) was created by taking the highest lidar first return within each1 m × 1 m grid cell and supplementing the missing values with interpolated vegetation heights obtained using the IDW (inverse distance-weighted) algorithm. Similarly, Digital Terrain Model (DTM) was created by using the lowest ground return within a 1 m × 1 m grid and filling the empty cells with the IDW algorithm (Figures 2a, 2b). A high correspondence (r^2 of 0.86 and RMSE of 1.85 m) was found between lidar derived canopy height and 77 field measured trees. Surface slope (as a percentage) and aspect (in degrees) were created using Spatial Analyst. ArcGIS v9.2 (Figures 2c, 2d).

Details of tree species groupings were obtained from the integrated forest ecosystem maps of the third decennial forestry inventory at 1:20000 produced by the Quebec`s natural resources Department (MRNFPQ, 2003). These were based on aerial photo interpretation and several ground based inventories conducted during 1991-2003 (Figure 2g).

Table 1. Lidar data acquisition parameters

Lidar survey configuration	Details
Date of acquisition	14-16 August
ALTM sensor model	2050
Flight altitude (m AGL)	1000
Pulse power (µJ)	200
Divergence (mrad)	0.2
Pulse frequency (Hz)	50000
Max. scan angle (°)	15
Footprint size (cm)	20
First return density (hits / m^2)	2.5
Last return density (hits / m^2)	0.2

Environmental variables – surface temperature and moisture

Plant moisture is a biophysical parameter that is directly associated with vegetation stress and biomass reduction. Land surface temperature is a good indicator of vegetation health, moisture and thermal conditions. These two environmental variables were mapped using radiometrically corrected and orthorectified Landsat ETM image of 8 July 2005 (Figures 2e, 2f). The surface moisture was generated by calculating the Wetness component of the Tasseled Cap transformation (TCT) for the Landsat ETM image using the coefficients developed by Huang et al (2002). The wetness component of the TCT being a contrast between shortwave-infrared and visible/ near-infrared reflectance, determines the amount of moisture held by the vegetation or soil and also can be an indicator of vegetation density (Kimes et al., 1986).

The spectral radiances of the 60m resolution ETM thermal band (Markham and Barker, 1986) were converted into satellite brightness temperature using the following relationship that is similar to the Planck equation with two free parameters (Schott and Volchok, 1985; Wukelic et al., 1989):

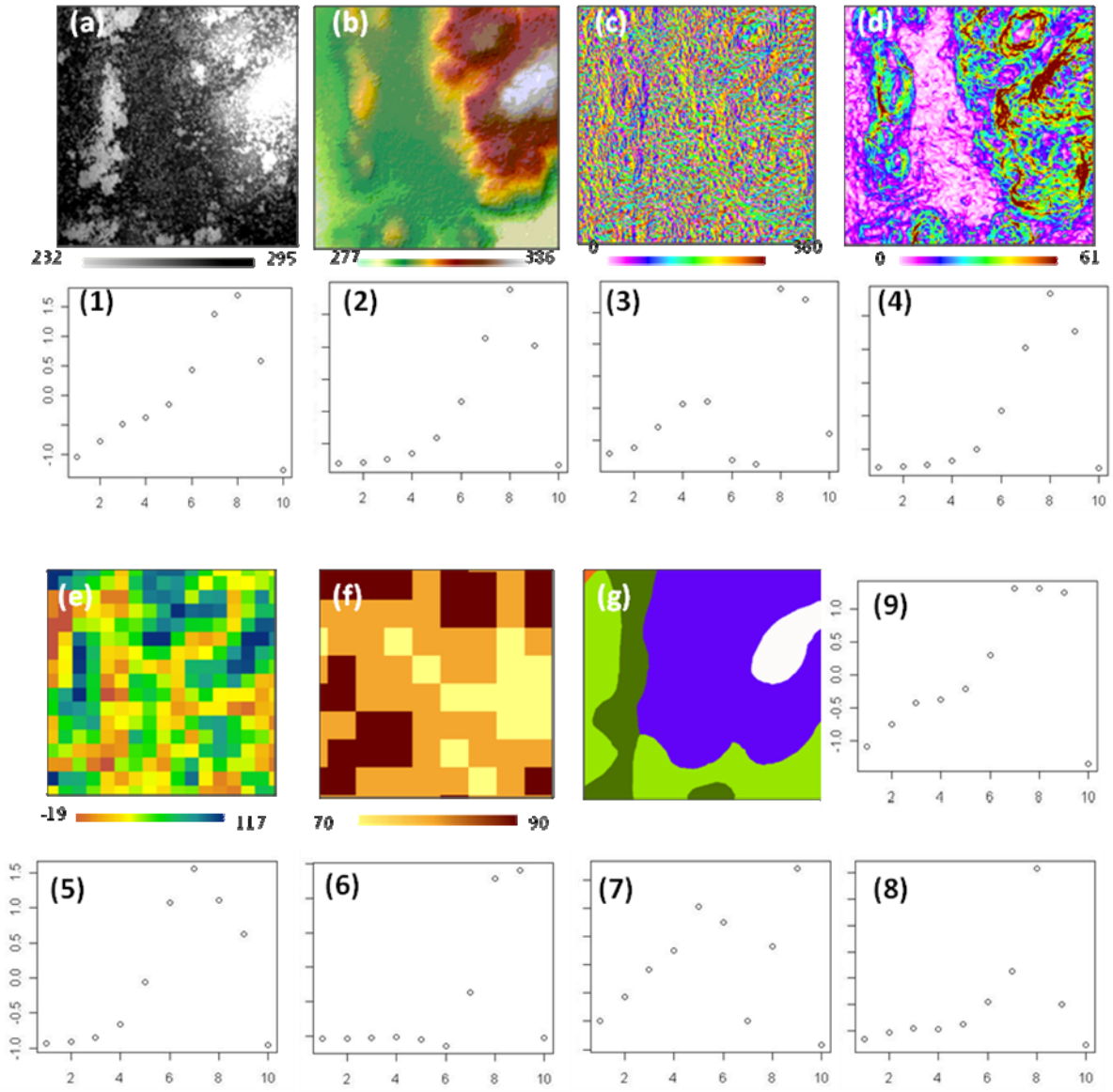


Figure 2. Vegetation patterns in relation to the topographic and environment factors – a sample window of the 1760 fire stand within the Conservation Zone. (a) DSM (in meters ASL) (b) Ground elevation (in meters

ASL) (c) Aspect (in degrees) (d) Percent Slope (e) Wetness index (f) Surface Temperature (in F) (g) Species groups (Dark green: Grey Pine with a mix of balsam fir; light green: balsam fir with paper birch; blue: Black spruce; white: Spruce mix with paper birch). (1) is the plot of the wavelet variance across scale (scalogram) of DSM and (2)-(6) are the plots of covariance of DSM against the independent variables, while (7) to (9) are the scalograms of the vegetation type stratified DSM. In plots (1) to (9) x-axis represents the scale of the wavelet while the y-axis variance of the wavelet coefficients.

$$T_B = \frac{K_2}{\ln\left(\frac{K_1}{L} + 1\right)} \quad (1)$$

where L the blackbody radiance for a temperature, T_B in Kelvin, integrated over the ETM thermal band, and K_1 and K_2 are two free parameters with the values of K_1= 60.776 mWcm-2sr-1μm-1, K_2=1260.56 K.

Table 2. Distribution (% to the total area of the sample window) of species groups in the stands.

Species groups	CS1760	CS1797	CP1870	CT1760	M1923
Balsam fir	-	16.1	-	-	-
Balsam poplar (>75%), balsam fir (<25%)	15.2	-	-	-	-
Balsam fir (<75%), paper birch (<25%)	25.9	27.9	3.6	58.2	-
Balsam fir (>75%), deciduous (<25%)	-	-	30.3	-	-
Jack pine (>50%), balsam fir and birch (<50%)	-	-	-	2.1	-
Paper birch (>75%), balsam fir (<25%)	52.5	4.2	-	-	3
Spruce	0.3	51.8	-	-	-
Spruce (>75%), paper birch (<25%)	6.1	-	-	-	-
Paper birch (>75%), jack pine (<25%)	-	-	53.4	-	-
Balsam poplar	-	-	12.7	-	-
Balsam poplar (>75%), jack pine (<25%)	-	-	-	-	45.6
Paper birch (>75%), spruce (<25%)	-	-	-	-	0.1
Paper birch (>50%), jack pine and spruce (<50%)	-	-	-	9.1	-
Jack pine	-	-	-	-	51.3
Balsam fir (<75%), spruce (<25%)	-	-	-	30.5	-

Disturbance history

We used the stand initiation maps created by Dansereau and Bergeron (1993) based on dendrochronology to identify the different times of origin since fire (TSF) for each stand. These ranged from 1760 to 1944. History of management interventions as provided in the information system of the research forest was used.

2.1 Wavelet analyses

Wavelet analysis is a method of quantitatively partitioning the variance into scales from a time series data , spatial transects (1-D) or images (2-D) concurrently that help in identifying patterns different scales or frequencies (Daubechesis 1992). It involves successive passes of a wavelet template (wavelet transform, WT) of

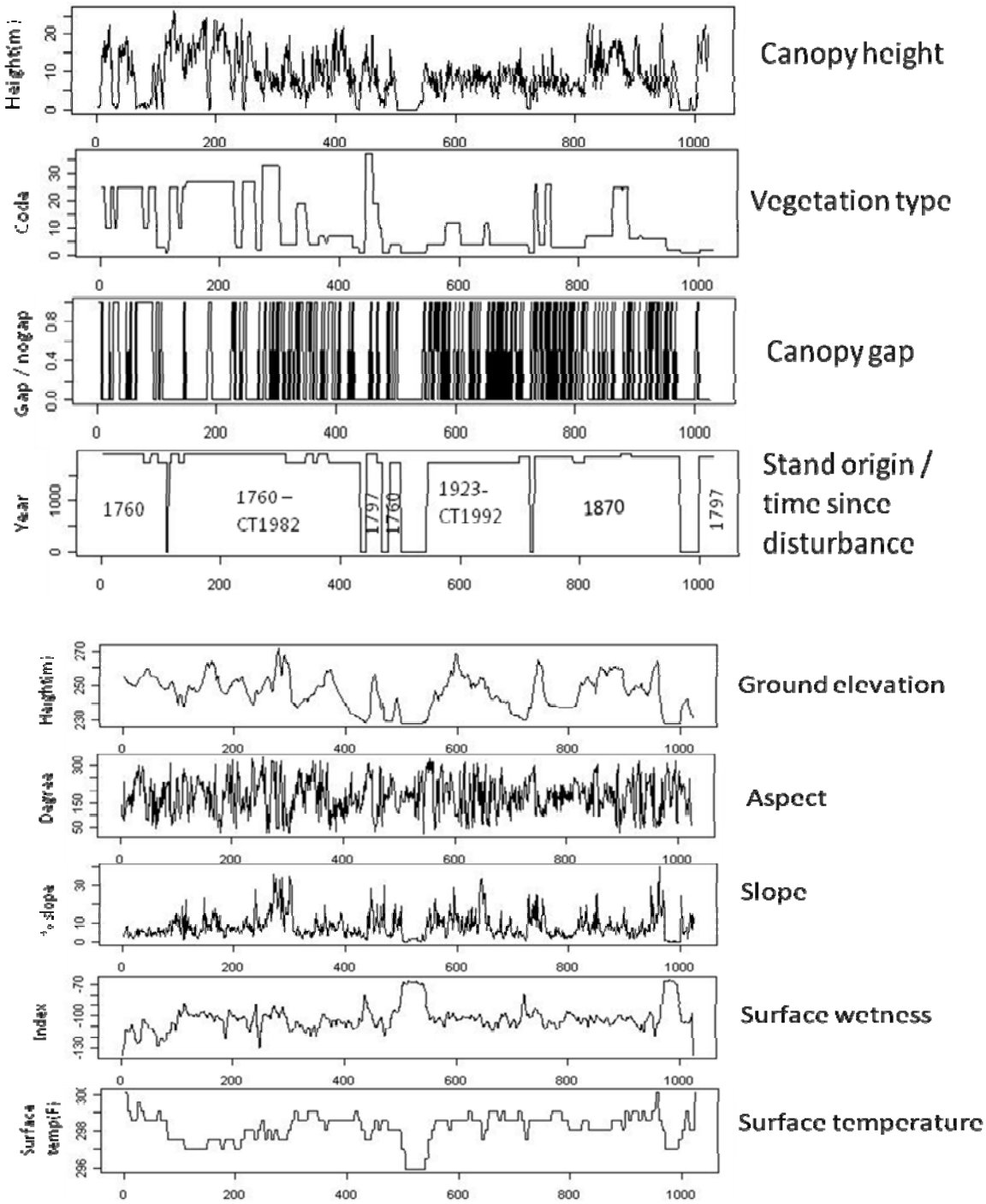


Figure 3. High variance and complexity in high resolution data like that of surfaces generated using lidar data seen in this example along a 10 km (10 m wide) transect across the research forest.

increasing size over a data series and assesses degree of similarity of pattern and the shape of the template at each size and location (Bradshaw and Spies, 1992). Wavelet coefficients are the values that describe the fit or correlation between the wavelet at a given scale) and the data array (at a particular location). To determine scale-specific information independent of location, we computed wavelet variance at each level of decomposition. Wavelet variance is the sum of the squared wavelet coefficients weighted by the number of coefficients at a given scale (Percival 1995). This 'multi-resolution analysis' (Mallat 1999) is one of the key attributes of wavelet analysis. A given decomposition level represents the information extracted from a particular scale. In recent years, WT have been found useful in the study of multi-scale and non-stationary processes over finite spatial and temporal domains. Applications of wavelet analysis have been found as effective tools not only in signal and image processing (compression, filtering and fusion), there have been several ecological applications using time series, 1-D transect data and 2-D images. 2D wavelets on remotely sensed data were applied to analyse grassland productivity (Csillag and Kabos 2002), tree crown feature identification (Falkowski et al. 2006, Strand et al. 2006), vegetation dynamics using NDVI (Martinez and Gilbert, 2009) and insect defoliation (James et al, 2010).

In this study, we first use 1-D transect to understand patterns of variability across the variables, and illustrate the utility of using a very high resolution remote sensing data for multi-scale analyses (Figure 3). This also guided us to remove noise at finer scales using wavelet filtering (Figure 4). We then used a two-dimensional maximum overlap discrete wavelet transform (MODWT; Percival and Walden 2000) with a simple 'Haar' basis to identify significant scales of spatial pattern in vegetation structure, in correspondence with environmental and topographical variables. Wavelet template sizes that increase as powers of two (e.g., 2^2, 2^3, 2^4, etc.) in terms of pixels were applied on the lidar DSM, ground elevation, slope, aspect, surface temperature and wetness index. For the purposes of this analysis, the term "scale" refers to the different sizes of wavelet template (here, level 1 equals to a resolution of 2 m, level 2 to 4 m and so on) , and the term "level" to refer to the separate sets of coefficients produced through variance decomposition. A given decomposition level represents the information extracted from a particular scale to identify (e.g., Figures 2.1). We stopped the transform when the scale was equal to the sample window considered i.e. at 2^9. Levels that exhibit high wavelet variance can be considered of interest (Bradshaw and Spies 1992). Extending this idea to wavelet covariance (Figures 2.1-2.6), peaks in wave-covariance between a dependent and an independent variable occurs when the two variables exhibit a shared pattern (Keitt and Urban, 2005).

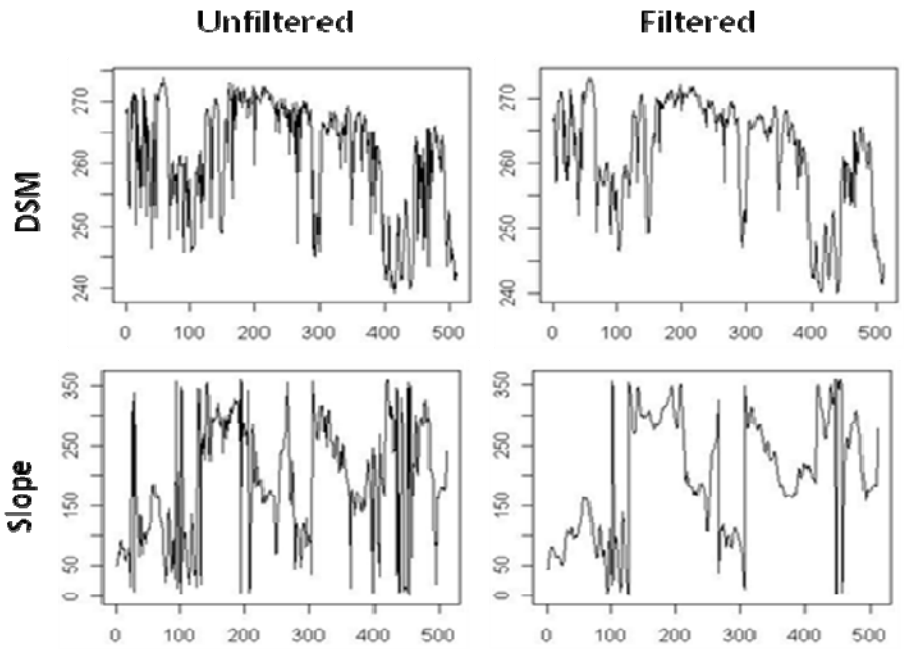


Figure 4. Wavelet denoising – an example of smoothing highly varying in high resolution data.

Our computations are based on software package Waveslim implemented in R (Whitcher 2005). Regression was then applied in the wavelet domain to fit a multiple linear regression model to the coefficients at each level of the transform with DSM coefficients as the response and topographic and environmental factor coefficients as explanatory variables at each level. We applied a stepwise model selection with minimisation of Akaike's information criterion to arrive at an optimal model at each level.

3. Results

Vegetation and disturbance history

Although the maximum ground elevation, surface temperature and moisture levels are highly varying across the chronosequence of stands, the mean of these variables was consistent irrespective of the stand age, 133 yr TSLD being an exception (Table -3). Species groups of stands in the conservation zone are mixedwoods in the 233 yr and 133 yr TSLD, but with a predominance of softwood species in 203 yr TSLD (Table-2). The managed stands are either predominantly softwood (20 yr TSLD) or predominantly hardwood (10 yr TSLD). As expected, mean canopy height decreased with the stand age.

Table 3. Summary statistics of the sample windows of stands with different disturbance history.

DSM

Stand	Statistic	DSM(m)	CH(m)	ELV(m)	SLP(%)	ASP(°)	WT	ST(F)
CS1760	*Min*	232.6	0.0	231.9	0.0	0.0	-117.1	77.2
	Max	295.8	29.5	276.5	61.0	360.0	-90.3	79.0
	Mean	256.6	9.2	247.4	7.2	182.1	-105.1	77.9
	Std.	11.4	6.8	9.4	6.4	103.1	4.5	0.4
CS1797	*Min*	237.7	0.0	238.0	0.0	0.0	-132.3	78.5
	Max	343.3	24.9	33.6	73.9	360.0	-93.2	86.2
	Mean	291.1	5.3	285.8	17.0	184.1	-104.4	80.4
	Std.	22.9	4.1	24.2	13.5	102.8	7.3	1.4
CP1870	*Min*	232.3	0.0	231.6	0.0	0.0	-117.1	76.1
	Max	294.2	27.6	286.7	64.1	360.0	-97.7	80.0
	Mean	264.1	11.7	252.4	9.2	179.6	-106.1	78.1
	Std.	12.5	6.4	11.1	7.4	103.1	3.5	0.4
CT1760	*Min*	230.8	0.0	230.8	0.0	0.0	-151.1	76.1
	Max	284.1	30.7	266.3	59.9	360.0	-95.6	84.4
	Mean	247.6	8.0	239.6	5.2	181.5	-114.1	78.2
	Std.	8.9	5.4	*6.2*	5.4	104.5	12.6	1.0
M1923	*Min*	239.6	0.0	239.6	0.0	0.0	-146.0	77.0
	Max	291.8	30.7	270.0	56.0	360.0	-102.0	81.7
	Mean	270.2	17.1	253.1	5.9	181.5	-114.6	78.3
	Std.	8.3	6.6	5.3	4.5	104.4	8.1	0.6

Surface height of the canopy; CH – Canopy height = DSM-Ground Elevation; ELV – Ground Elevation; SLP – Slope; ASP – Aspect; WT – Wetness Index; ST- Surface Temperature;

Patterns across the transect in 1-D

Each of the chosen variables shows different patterns of variability even as a simple line transect of coarser 10 m resolution (Figure 3). Canopy height is a mixture of complex fine scale pattern with respect to vegetation type, while intermediate to broad scale with disturbance history. Response of canopy height to other variables is not clearly evident. Patterns of elevation seem intermediate while surface wetness and temperature are broad scale. By appropriately enhancing or reducing certain coefficients of the wavelet decomposition, the signal reconstructed through inversion algorithms can reveal certain desirable features or suppress undesirable ones i.e. noise. Thus, we used various options of denoising algorithms available in Waveslim to reduce the noise and selected the one that retained essentially the same shape of the underlying surface in the variables extracted through the lidar data (e.g., Figure 4). Wavelet based denoising has been applied earlier to remove noise from SAR images (e.g., Dai et al., 2004) and more recently to filter terrain data (e.g., Tate et al., 2005).

Wavelet analyses

Plots of the wavelet variance of all the variables show that local pattern of wavelet variance differed between stands originated from different disturbances (Figure 2). Except aspect, large amounts of variance at coarser spatial scales and little fine scale variability of all variables is seen across all stands. Wavelet variance of the DSM peaked at scale 8 but with varying rates of increase and dissimilar patterns at finer scales across stands. The highest amplitude of variance of DSM occurred in 1870 stand and the lowest in the 1923 stand. However, the pattern of wavelet variance of the DSM was completely different at sub-stand scales (Figure 2) with two scales (3 and 8) dominating the pattern of vegetation in the younger managed stands.

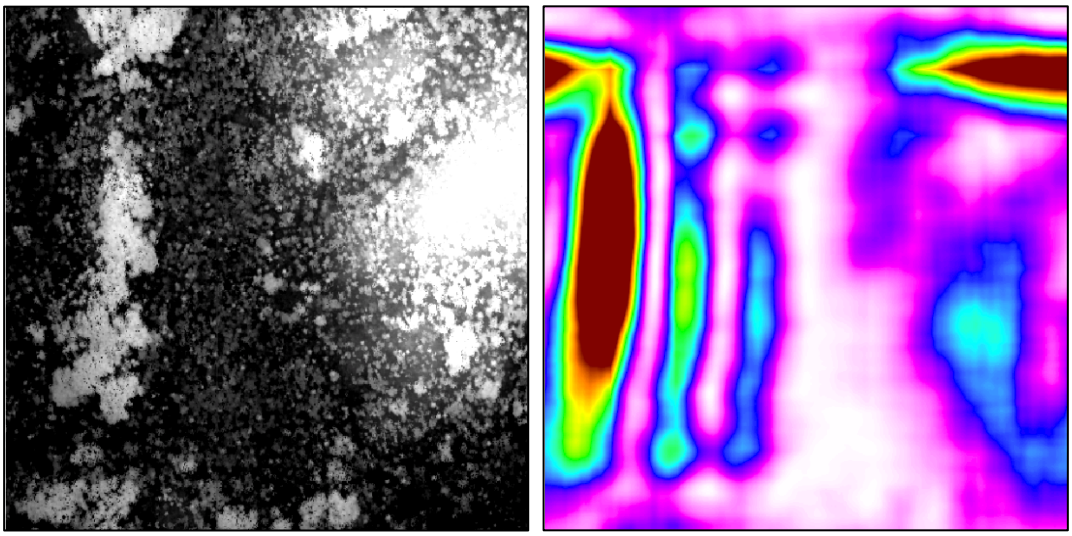


Figure 5. Local wavelet variance image of DSM at scale 8 of 1760 stand in the conservation zone. White lines indicate high (bold line), medium (dashed line) and low (dotted line) variance zones.

Influence of environmental or topographical factors

Typical of noisy multi-scale patterns, up to scale 4 all variables are uncorrelated with DSM and show different patterns at coarser scales in the wavelet correlations

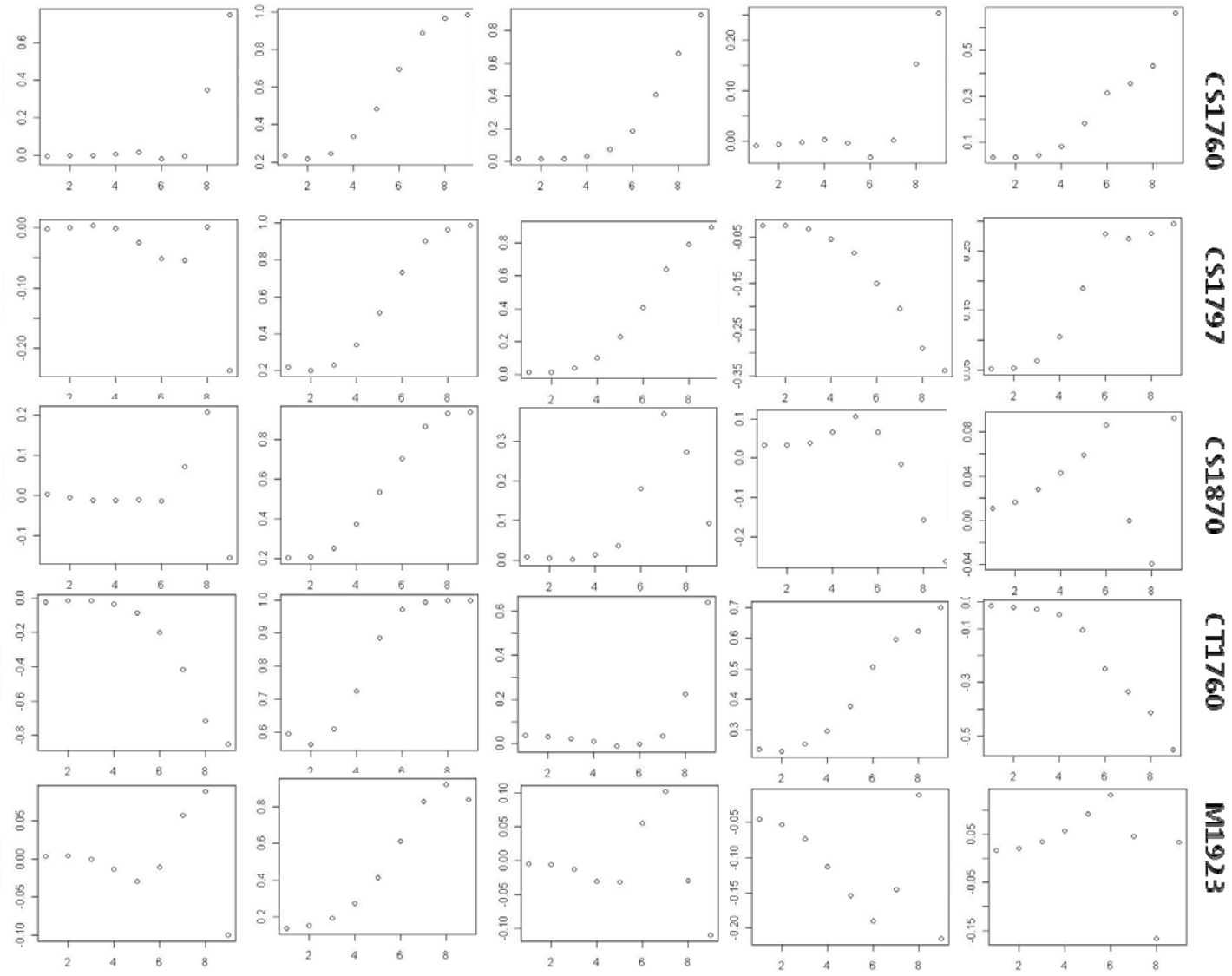
Table 4. Regression models of vegetation in different disturbance stands

STAND	LEVEL	INTERCEPT	ASP	ELEV	SLOPE	STEMP	MOIST	F+	AIC	R2
CS1760	0	358.549	0.0004	0.9752	-0.0258	-1.1259*	0.0669*	115500	9972890	0.68
	1	360.8415	0.0006	0.9754	-0.0267	-1.1332*	0.0686*	121100	9476561	0.7
	2	363.814	0.001	0.9754	-0.0286	0.9754*	0.0712*	136700	8352314	0.72
	3	373.4388	0.0019	0.9751	-0.0303	-1.1736*	0.0752*	170700	6618256	0.77
	4	406.707	-	0.9745	-0.0288	-1.2815*	0.0807*	284700	4848916	0.81
	5	507.8946	-0.0117	0.9697	-	-1.6125*	0.0737*	408900	3192902	0.86
	6	687.2518	-0.0523*	0.9601	0.0558	-2.1895*	0.0526	539700	1667639	0.91
	7	1030.1016	-0.1072*	0.9544	0.0852	-3.3051*	0.0396	1243000	492032.5	0.96
	8	1196.4219	-0.149*	0.9522	0.0897	-3.8395*	0.0271	2734000	88075.41	0.98
CS1797	6	19.1472	0.0362*	0.964	-0.0695*	-	0.0841*	23940000	283717.2	0.99
	7	71.5689	0.0756*	0.9828	-0.0884*	-0.2258	0.0563	48570000	79963.52	0.99
	8	25.5074	0.0186	0.9771	-0.1075*	-	0.1419*	189700000	10964.48	0.99
CP1870	6	1841.4772	0.0063	18980	-0.1639	-6.3621*	-0.5276*	898000	1637303	0.94
	7	2209.246	-0.0196	1.0859	-0.2441*	-7.6661*	-0.7149*	1747000	668326.3	0.97
	8	2650.6805	-0.0208	1.0989	-0.4001*	-9.1769*	-0.7911*	3126000	161856.3	0.98
CT1760	6	233.1369	-0.0718*	1.06	0.2477*	-0.7853	-0.06	1199000	451350.7	0.96
	7	264.0127	-0.1572*	1.0298	0.3344*	-0.8141	-0.0612	262138	118845.5	0.98
	8	28.0858	-0.2026*	1.0667	0.2779*	-	0.0068	13200000	13333.02	0.99
M1923	6	289.4665	0.0313	0.931	-0.0742	-0.0373	-0.8851	246700,0	987170.9	0.82
	7	492.4843	0.1352*	0.9249	-0.1823*	-1.6435	-0.0968	519200	294192.8	0.91
	8	-483.3944	0.1614*	1.0635	-0.0887	1.6314	0.2764*	1041000	54109.31	0.95

Table 5. Regression models for the sub-stand units.

STAND	LEVEL	INTERCEPT	ASP	ELEV	SLOPE	STEMP	MOIST	F+	AIC	R^2
CS1760	1	-5803.7215	0.0072	0.6305*	0.4925*	19.7945*	-	73.68	16524.1	0.25
	2	-5117.6175	-	0.5388*	0.9077*	17.5685*	-	333.3	12309.5	0.53
	3	-3578.8066	-	0.296	2.0586*	12.5943*	-	1031	6301.803	0.78
	4	-2823.5015	0.0678	0.5733*	1.6191*	9.7979	-	876.8	5051.98	0.8
	5	-9995.8335	0.1326	1.3001*	-	33.2338	-	24460	531.0025	0.99
	6	-66175.4838	0.3667*	0.5876*	-	223.0564	2.747*	5614	341.2745	0.96
	7	4059.385	-	1.0223*	-	-13.766	-0.4811	66060	16.8829	0.99
	8	893.4482	-	1.1809*	-	-3.1109*	-	61180	0.9198	0.99

Numbers in italics and '*' indicate the most significantly and other important contributing factor (based on the t-statistic) respectively towards the pattern of vegetation. '+' indicates that the F statistic is significant at p-level 0.


CS1760
CS1797
CS1870
CT1760
M1923

across the stands (Figure 6). With an exception in 1797 stand, elevation is highly positively correlated with DSM in the coarser scales. High positive correlation of slope in 1760 – conservation and 1760 clear cut stand, while high negative correlation of aspect in 1797 stand is seen (Table 4). Further exploring scale-specific patterns through wavelet regression, we observe that irrespective of the disturbance history, topographic factors, specifically elevation, greatly influence the patterns of vegetation across all scales. Though not significantly higher, surface temperature in older stands, and soil moisture in the younger stands also seem to influence vegetation. However, elevation, slope and surface temperature at finer scales are contributing equally while elevation alone explains the patterns in vegetation when considering sub-stand units (Table 5).

4. Discussion

There is a need for fine to landscape scale assessments of vegetation patterns to improve our understanding of the effect of ecological processes such as disturbances and environmental factors to aid in setting conservation targets, assessing current managed systems, or in making ecologically informed management decisions. Towards this, it is essential to capture the spatial and temporal variability over areas that are large enough to represent the ecosystem and its natural and managed disturbance regimes. Conventional empirical data, that generally is coarse at landscape scale, may not be able to capture variability at multiple scales. Although extensively used in image processing (Ulfarsson et al., 2003; Sakamoto et al., 2005), remote sensing and wavelets, are also shown to be invaluable tools in such multi- scale analyses. For example, for ecological feature extraction i.e. plant location and size (Falkowski et al., 2006, Strand et al., 2008), regional scale vegetation change using NDVI (Martinez and Gilbert, 2009) and insect defoliation (James et al, 2010). However, identification and quantification of very fine to broad scale-specific ecologically relevant patterns across multiple spatial scales has not been made.

In this study using lidar, we investigate whether the relationships between vegetation patterns of a boreal landscape and disturbance history as well as environmental heterogeneity can be detected using wavelet analysis and correlations at multiple scales. However, as the spatial resolution of lidar is so high, the data were first filtered, using wavelet, to be able to match the spatial resolution scales at which the disturbance events occur. Then using the filtered lidar data, we present the potential application of very high resolution canopy surface models generated using lidar to highlight and quantify heterogeneity in vegetation patterns, effect and significant contribution of disturbances and environmental factors towards the variability at multiple scales.

Boreal forest dynamics are driven by disturbances (Bergeron, 2000). In this study, the results confirm that variability in vegetation pattern across multiple scales is affected by disturbance history. Although weak at very fine scales, these patterns are further influenced by topographic factors, specifically elevation over scale 4 (i.e. from a spatial resolution of 16 m). Our study shows that the variability was better explained by different environmental factors and that at different scales.

It is therefore important to start planning forest studies using a multi-scale framework as it is possible to do by combining lidar data with wavelet analysis. The wavelet filtering used in this research can also be applied in several other forest ecosystems such as temperate forest as well as mixed boreal forest to determine the appropriate scale to investigate the relationship between environmental factors and forest spatial structure.

References

Bergeron, Y. 2000. Species and stand dynamics in the mixed woods of Québec's southern boreal forest. Ecology, 81, p. 1500–1516.

Bradshaw, R. H. W. and T. Spies. 1992. Characterizing canopy gap structure in forests using wavelet analysis. Journal of Ecology 80:205-215.

Bradshaw, G. A. and M.-J. Fortin. 2000. Landscape heterogeneity effects on scaling and monitoring large areas using remote sensing data. Geographic Information Science 6 61–68.

Cazelles, B., M. Chavez, D. Berteaux, F. Menard, J. O. Vik, S. Jenouvrier and N. C. Stenseth. 2008. Wavelet analysis of ecological time series. Oecologia 156:287-304.

Csillag, F. and S. Kabos. 2002. Wavelets, boundaries, and the spatial analysis of landscape pattern. Ecoscience 9:177-190.

Dai, M., Peng, C., Chan A.K., Loguinov, D. 2004 Bayesian wavelet shrinkage with edge detection for SAR image despeckling. IEEE Transactions in Geoscience and Remote Sensing 42(8): 1642-1648.

Dale, M. R. T. and M. Mah. 1998. The use of wavelets for spatial pattern analysis in ecology. Journal of Vegetation Science 9:805-814.

Danserau, P. & Bergeron, Y. (1993) Fire history in the southern part of the boreal forest of northwestern Quebec. Canadian Journal of Forest Research, 23, 25-32.

Daubechies, I. 1992. Ten lectures on wavelets. Society for industrial and applied mathematics, Philadelphia, PA.

del Moral, R., and A. F. Watson. 1978. Gradient structure of forest vegetation in the Central Washington Cascades. Vegetatio 38:29–48.

Falkowski, M. J., A. M. S. Smith, A. T. Hudak, P. E. Gessler, L. A. Vierling and N. L. Crookston. 2006. Automated estimation of individual conifer tree height and crown diameter via two-dimensional spatial wavelet analysis of lidar data. Canadian Journal of Remote Sensing 32:153-161.

Forman, R.T.T. and Godron, M. 1986. Landscape Ecology. JohnWiley and Sons, New York.

Huang, C., Wylie, B., Yang, L., Homer, C., & Zylstra, G. (2002). Derivation of a tasselled cap transformation based on Landsat 7 at-satellite reflectance. International Journal of Remote Sensing, 23, 1741– 1748.

James P.M.A., Fleming, R.A., Fortin, M.J. 2010. Spatial significance testing using wavelets and null models: Spruce budworm defoliation as a case study. Landscape Ecology 25: 873-887.

Keitt, T. H. and D. L. Urban. 2005. Scale-specific inference using wavelets. Ecology 86:2497-2504.

Kimes, D. S. et al 1986, Directional reflectance distributions of a hardwood and pine forest canopy, IEEE Trans Geoscience Remote Sens Vol. GE-24, 281-293.

Lim, K., P. Treitz, K. Baldwin, I. Morrison, and J. Green. 2003. Lidar remote sensing of biophysical properties of tolerant northern hardwood forests. Canadian Journal of Remote Sensing, 29, p. 658−678.

Lefsky, M.A., Cohen, W.B., Parker, G.G. & Harding, D.J. (2002) LIDAR remote sensing for ecosystem studies. BioScience, 52, 9–30.

Levin, S. A. 1992. The problem of pattern and scale in ecology. Ecology 73: 1943–1967.

Mallat, S. 1999. A Wavelet Tour of Signal Processing. Academic Press.

Markham, B. L. and J. L. Barker. 1986. Landsat MSS and TM post-calibration dynamic rangers, exoatmospheric reflectance and at-satellite temperatures. EOSAT Landsat Tech. Notes (Aug.): 3-8.

Martínez, B., Gilabert, M.A. 2009. Vegetation dynamics from NDVI time series analysis using the wavelet transform, Remote Sensing of Environment 113 1823–1842

MRNFPQ (Ministère des Ressources naturelles, de la Faune et des Parcs Québec). 2003. Zones de Végétation et Domaines Bioclimatiques du Québec. Gouvernement du Québec, Ministère des Ressources naturelles, de la Faune et des Parcs, Direction des Inventaires Forestieres, Code de diffusion, 2003-3043.

Percival, D. B. and A. T. Walden (2000) Wavelet Methods for Time Series Analysis, Cambridge University Press.

Percival, D. B. and A. T. Walden. 2000. Wavelet Methods for Time Series Analysis. Cambridge University Press, New York.

Percival, D. P. 1995. On estimation of the wavelet variance. Biometrika 82:619-631.

Pickett, S.T.A. and White, P.S. (eds.). 1985. The Ecology of Natural Disturbance and Patch Dynamics. Academic Press, New York

Schott, J. R. and W. J. Volchok. 1985. Thematic Mapper thermal infrared calibration. Photogrammetric Engineering and Remote Sensing, Vol. 51, No. 9, 1351-1357.

Strand, E.K., Vierling, L.A., Smith, A.M.S. and Bunting, S.C. 2008. Net changes in aboveground woody carbon stock in western juniper woodlands, 1946–1998. J. Of Geophysical Research, 113: G01013.

Tate.J.N., Brundson, C., Charlton, M., Fotheringham, S., Jarvis, C.H. 2005. Smoothing / filtering lidar digital surface models. Experiments with loess regression and discrete wavelets. Journal of Geographical Systems 7: 273-290

Ulfrasson, M.O., Benediktsson, J.A. and Sveinsson, J.R.2003, Data fusion and feature extraction in the wavelet domain. International Journal of Remote Sensing, 24, pp. 3933–3945.

Whitcher B (2005) Waveslim: basic wavelet routines for one-, twoand three-dimensional signal processing. R package version 1.5 http://www.image.ucar.edu/staff/whitcher/book/

Prediction of stem attributes by combining airborne laser scanning and measurements from harvesting machinery

JOHAN HOLMGREN†, ANDREAS BARTH‡, HENRIK LARSSON†, and HÅKAN OLSSON†*

Johan.Holmgren@srh.slu.se

† Swedish University of Agricultural Sciences, Department of Forest Resource Management, SE-90183, Umeå, Sweden

‡ The Forest Research Institute of Sweden, SE-75183, Uppsala, Sweden

Abstract

A method for prediction of stem attributes by combining harvester measurements with Airborne Laser Scanning (ALS) data including an algorithm for automatic segmentation of tree crowns based on tree crown models was validated. Each harvested tree was linked to the nearest tree crown segment derived from the ALS data. For each segment several variables were derived from ALS data which were used for imputation of crown segments. The relative sub-stand-level cross-validated RMSE of stem volume, mean tree height, mean stem diameter, and stem density (stems per ha) estimates were 11%, 7%, 9%, and 18%, respectively. The imputation of stem data collected by harvesters could in the future be used for bucking simulations of not yet harvested forest stands in order to predict product recovery.

1. Introduction

In Swedish forestry it is important to be able to control the timely flow of round-wood from the logging operations to the saw- and pulp-mills. The mills have to produce products in current demand. This is a serious management problem since the quality of data traditionally supplied by pre-harvest inventories provides a too poor basis for sufficient estimation of the round-wood recovery. The common systems focus mainly on estimation of stand average properties, for example stem volume per hectare, but these data are neither accurate enough nor detailed enough for control of round-wood deliveries.

Detailed information of round-wood recovery can be recorded by harvesting machinery performing final felling. This information is normally only utilized to measure logging production and provide input data for controlling transportation to the mills, however, these data could potentially be used to estimate the round-wood recovery of remaining stands with similar forest characteristics. This could be possible if there is a data source that can link stem data from harvested areas to non-harvested trees with similar characteristics. Airborne laser scanning (ALS) can be used for detection and estimation of tree attributes for individual trees (*e.g.* Hyyppä and Inkinen 1999; Hyyppä *et al.* 2001, Persson *et al.* 2002, Koch *et al.* 2006, Solberg *et al.* 2006). These tree crown delineation algorithms provide estimates of position and characteristics of individual trees, such as stem diameter, height, and crown width. By equipping harvesters with high precision positioning systems, tree positions could be measured during harvesting, making it possible to automatically link tree stem data with tree crown polygons derived from ALS data. A future operational system could continuously receive stem data sent from operating harvesters and these data could be used to continuously update a reference database enabling improved prediction of stem attributes for not yet harvested forest stands. The imputation of stem data files could then be used for bucking simulations in order to predict product recovery. The objective of this study was to evaluate a method where tree stems are imputed based on tree crown segmentation in ALS data.

2. Material and Methods

2.1 Study area

The study was located at a test site in northern Sweden (64º 6' N, 19º 11' E) where 20 forest stands, planned to be harvested, were selected. The forest stands were dominated by Scots Pine or Norway Spruce. Four field plots within each forest stand were systematically placed on a grid with 50 m inter-node distance.

2.2 Field data

The field plot inventory was performed from June to August 2008. For all trees on a 10 m radius field plot, the stems were callipered 1.3 m above ground level (DBH) if the DBH was at least 40 mm, tree species were recorded, and stem position were measured relative to the field plot centre using ultrasonic triangulation with three transponders. For all measured trees on the plot, an identification number was painted on the tree stem. The plot centre position was measured using DGPS, which is expected to produce sub-meter accuracy under optimal conditions. The spatial patterns of tree positions and tree sizes from field data and tree detection with ALS data were used as input for the algorithm by Olofsson *et al.* (2008) for correction of DGPS measured plot centre and tree positions.

2.3 Harvester data

Trees on the plots were harvested during winter 2009. A special module of a harvester GIS program was developed for registration of trees on the plots used in this research project. The position of the harvester and the field plots were visible on a digital map and if a tree was harvested within a field plot, the operator was asked to enter the tree identification number painted on the tree stem. The stem data were recorded according to the StanForD standard (Anon 2008a) and the tree identification number was saved in an associated stem-file. The stem diameter was measured by the harvester continuously along the processed part of the tree stem and registered with 10 cm intervals. The program for registration of tree stems did not operate in an optimal way for the specific harvester used in this case, which produced linking errors. The linking was manually verified and because of errors seven plots were excluded from further analysis: these were all plots in one stand, two plots in another stand, and one plot in a third stand.

2.4 Airborne laser scanner data

The laser data were acquired on the 3rd and 5th of August 2008 using the TopEye MKII system S/N 425 operated onboard a helicopter from an altitude of 500 m and 250 m above ground level for main strips and cross strips, respectively. Because of overlapping strips the measurement density varied between plots but the average density, including both first and last returns, was 15 laser returns per m^2.

2.5 Tree crown segmentation

The segmentation procedure consisted of two phases: (1) a training phase where known tree locations on sample plots were used in order to predict optimal parameter settings as a function of variables that can be derived from ALS data, and (2), a prediction phase in order to produce tree crown segments for the entire area. The segmentation algorithm used six steps, described in the following sections.

Create a height model. A Digital Surface Model (DSM) was created by assigning each raster cell the highest z-value of laser returns located within the cell. A Digital Elevation Model (DEM) was created and nDSM was calculated as the difference between DSM and DEM. A canopy height model (CHM) was created to describe the height of tree crowns at each location. A raster cell value of the CHM was set to the nDSM value if the cell value was above a height threshold (2 m), otherwise the value was set to zero. Some raster cells with a zero value could be inside a tree crown due to the lack of laser returns at that location or because the laser beam penetrated through gaps in the canopy. In order to set a value to these raster cells with no data, these areas need to be distinguished from areas outside of the tree crown, *i.e.* the crown area needs to be defined. First, a binary crown area raster was derived from the nDSM, in which a raster cell was set to the value one if the corresponding nDSM raster cell had a value greater than a height threshold (2 m);

otherwise the raster cell was left as a zero value. The crown area raster was then updated by morphological closing with a structure element of 3×3 cells. The next step was to update the CHM based on the crown area raster in order to set non-zero values of the CHM at crown locations, *i.e.* value of one in the crown area raster. Neighbour cells with values greater than zero within the smallest window needed to cover at least one value greater than zero were used to update the CHM raster cell. If only one non-zero value was found within the window, then that value was assigned; otherwise, the mean value of non-zero values was assigned.

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

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

Different models were tested for a raster cell if the CHM had a height value greater than zero: parameter *a* was set to the value of the CHM at the model origin, and different radii were tested, *b* = 0.5, 0.7, r_{max}, where r_{max} is the maximum expected radius set at a proportion of model height (parameter *a*). In order to obtain higher correlation values some laser returns were removed in the same way as done by Solberg *et al.* (2006): laser returns with other laser returns within a radius of 0.3 m which had a higher z-value were removed. This pre-processing increased the correlation because returns inside tree crowns were discarded making the correlation with a solid geometric model higher. The correlation surface was smoothed by filtering the surface three times with a 3×3 Gaussian kernel.

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

Merging segments. The aim of the preceding steps was to create segments that are small enough to include small trees but large enough to make it possible to test models, *i.e.* include more than a minimum number of returns. The next step was to merge segments with the aim of removing segments that only covered part of a tree. For each segment, models (*i.e.* Equation 1) were used to decide if the segment should be merged or not to a neighbour segment. A neighbour segment was defined as a segment having an edge in common with the tested segment. The model origin was placed at the raster cell with the highest value of the correlation surface (segment centre) within the tested segment and a test value was calculated using only laser data within that segment. The model was also placed at the centre of a neighbour segment and a test value was calculated using only laser data within the tested segment. If the value at the tested segment centre yielded a higher test value than any where the model origin was at a neighbour segment centre, the segment was not merged; otherwise, the segment was merged with the neighbour segment for which the highest test value was calculated. The test value used to decide whether to merge two segments was the weighted correlation between h and z-values (weight = distance above ground level) multiplied by a penalty factor, *p*, calculated with Equation 2, where r_p is predicted and r_o is the observed radius-height ratio for a tree crown.

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

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

Prediction of radius-height ratio. In order to predict the best radius-height ratio (*i.e.* the r_p value) to be used for the segmentation, empirical data were used from field plots with known tree locations. Several segmentations, each with one expected radius-height ratio, were tested and the results were evaluated by counting number of trees within each of the segments for the segments that were located inside a field plot. For each segment laser variables were also derived. For this study five classes (percentile 20, 40... 100) were defined using the maximum height of the segment, but other variables, such as variables related to tree species, could also be used. The radius-height ratio with the highest proportion of one field tree within a segment was judged as the best radius-height ratio for a class. For this study a cross validation procedure was used by excluding field data for a specific square kilometre when producing the training data used for segmentation of that square kilometre.

Segmentation of the area. After prediction of the radius-height ratio segmentation could be performed for the entire laser scanned area with different predicted radius-height ratios for different classes derived from laser data. Tree crown variables, *e.g.*, percentiles, were finally derived for the segments.

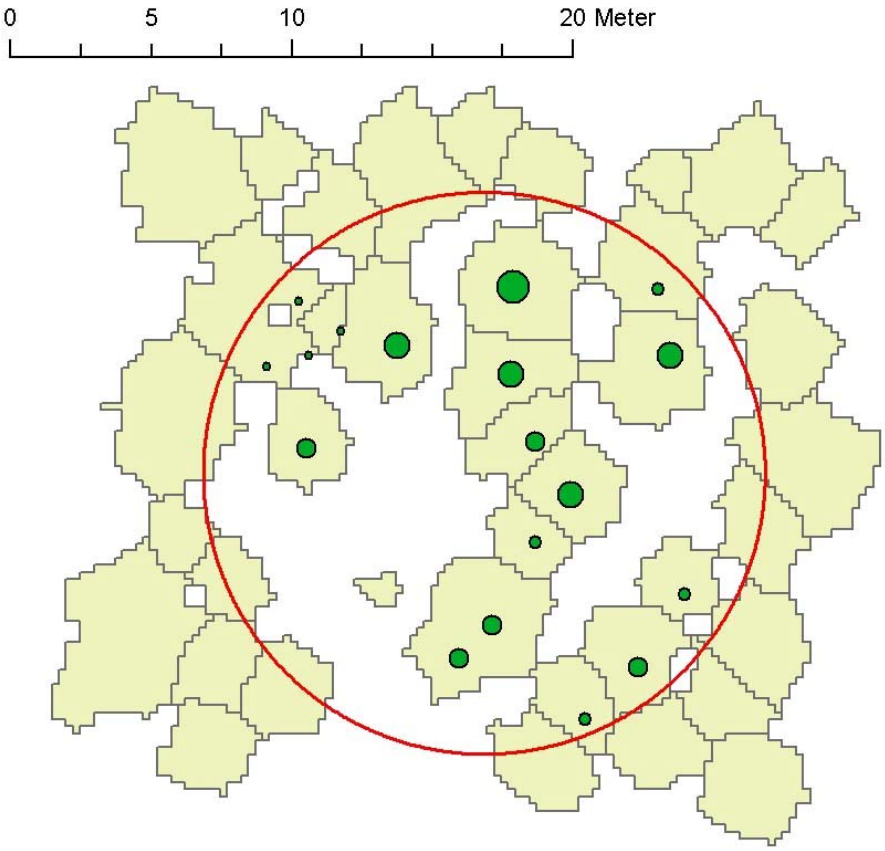


Figure 1. Tree crown segments and harvester measured trees (small circles, size proportional to stem diameter), within a field plot (large circle).

2.6 Imputation

Tree stems were linked to a laser derived crown polygon if they were inside the polygon. If a tree stem was outside any crown polygon the stem was linked to the nearest tree crown polygon (Figure 1). The MSN method (Moeur and Stage 1995) was used for imputation of tree crowns using the yaImpute package as part of the R-project (Anon 2008b). Only segments fully contained within a field plot were used for training but all segments with predicted tree top location within a field plot were used for prediction. The sum of stem

volume, DBH, stem diameter at last cut and tree height, from zero, one, or several tree stems that had been linked to a segment were used as the dependent variables and variables extracted from ALS data for a segment were used as the independent variables for the imputation. The variables extracted from ALS data were: laser height percentiles, crown width, crown length, mean and standard deviation of height of laser returns, mean intensity, number of returns in the crown, and proportion of first returns.

2.7 Validation

Predictions of stem lists were validated on a sub-stand level. A sub-stand was defined as the total area of the two to four field plots that were located on a regular grid with 50 m inter-node distance within the same forest stand. The RMSE and Bias were calculated using Equations 3 and 4, where y_i is the predicted variable for stand i and N is number of stands.

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

$$Bias = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\widehat{y_i} - y_i)} \quad (4)$$

The error index (*EI*) was calculated with a fixed number of intervals (50 mm intervals) according to Reynolds *et al.* (1988), where p_j is the predicted and o_j is the observed number of trees in class j and n is the total number of trees:

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

3. Results

For the segments fully contained within a plot 62% contained one tree, 15% contained two trees, 5% contained three trees, 2% contained four or more trees, and 16% contained no harvester processed tree stems. If only one of the trees within a segment would have been linked 72% of the number of trees linked to a segment fully contained within a plot would have been classified as detected trees. The proportion of not linked segments was 14% of the number of trees linked to a segment fully contained within a plot. The imputed and harvester measured sum of stem volume were compared at segment level (Figure 2). When aggregated at sub-stand level the RMSE values were between 7 and 18% for all validated variables when all segments within the same sub-stand were excluded from training data (Figure 3, Table 1). The RMSE values did not differ much for two tested cross validation procedures: (A) excluding only segments from the same plot, and (B) excluding segments from the same forest stand from training data. The bias was lower for validation procedure A as compared to procedure B for estimation of all variables except for stem number estimates. For stem diameter distribution estimations at sub-stand level, the average EI values were 0.36 and 0.38 for validation procedures A and B, respectively.

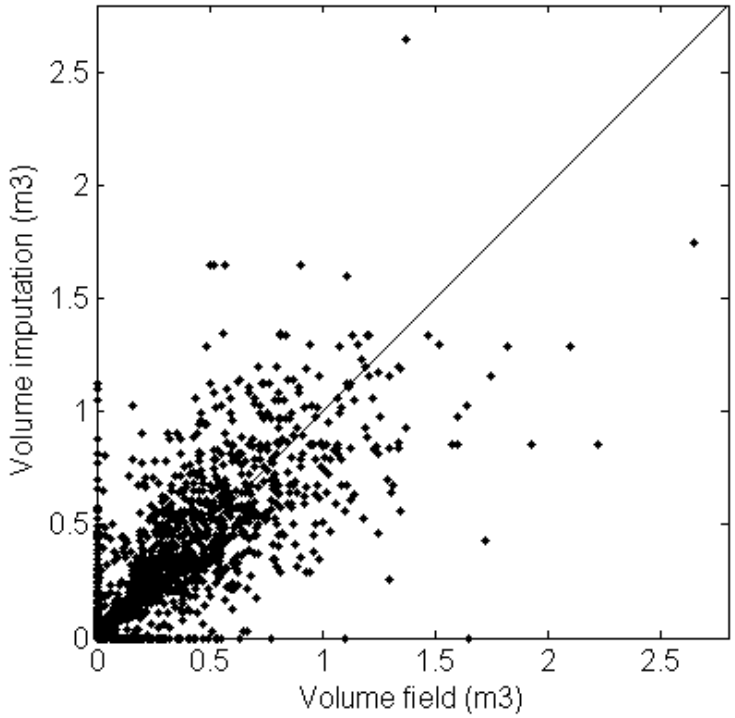


Figure 2. Stem volume per segment (m^3) from imputation plotted against harvester measured stem volume; validation by excluding segments from same sub-stand from training data.

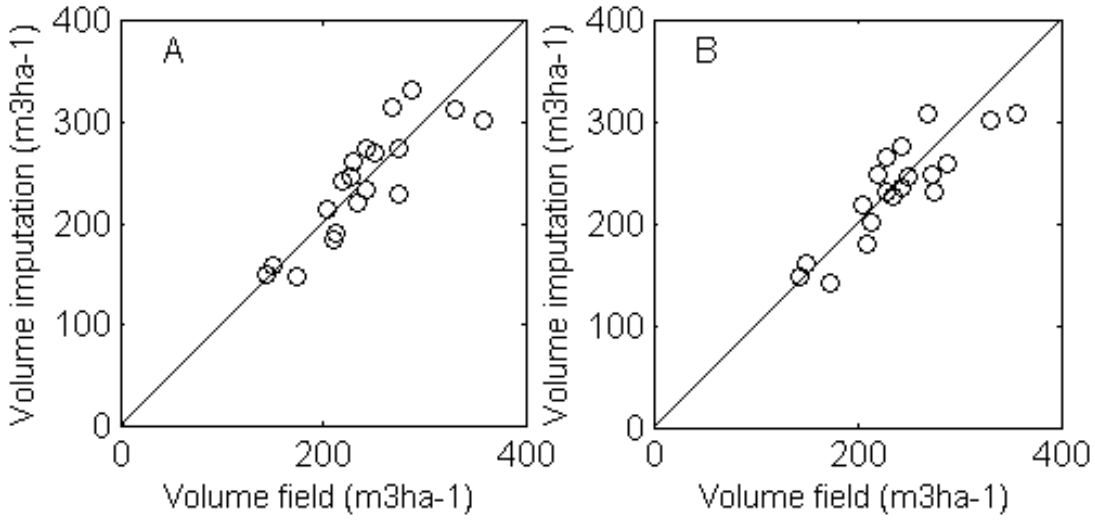


Figure 3. Imputation of stem volume at sub-stand level plotted against harvester measured stem volume (m^3ha^{-1}), for (A) excluding segment on the same plot, and (B) excluding segments from the same sub-stand from training data.

Table 1. Imputation results validated at sub-stand level with (A) excluding segment from the same plot, and (B) excluding segments from the same sub-stand from training data.

		RMSE	Rel. RMSE (%)	Bias	Rel. Bias (%)
Stem volume ($m3ha^{-1}$)	A	28.1	11.7	0.76	0.32
	B	27.2	11.4	-5.66	-2.37
Mean tree height (m)	A	0.85	5.95	-0.17	-1.20
	B	1.07	7.44	-0.26	-1.81
Mean stem diameter (mm)	A	17.2	8.30	-4.32	-2.08
	B	19.0	9.13	-5.58	-2.69
Stem number (ha^{-1})	A	121	16.2	24.7	3.32
	B	131	17.6	16.6	2.23

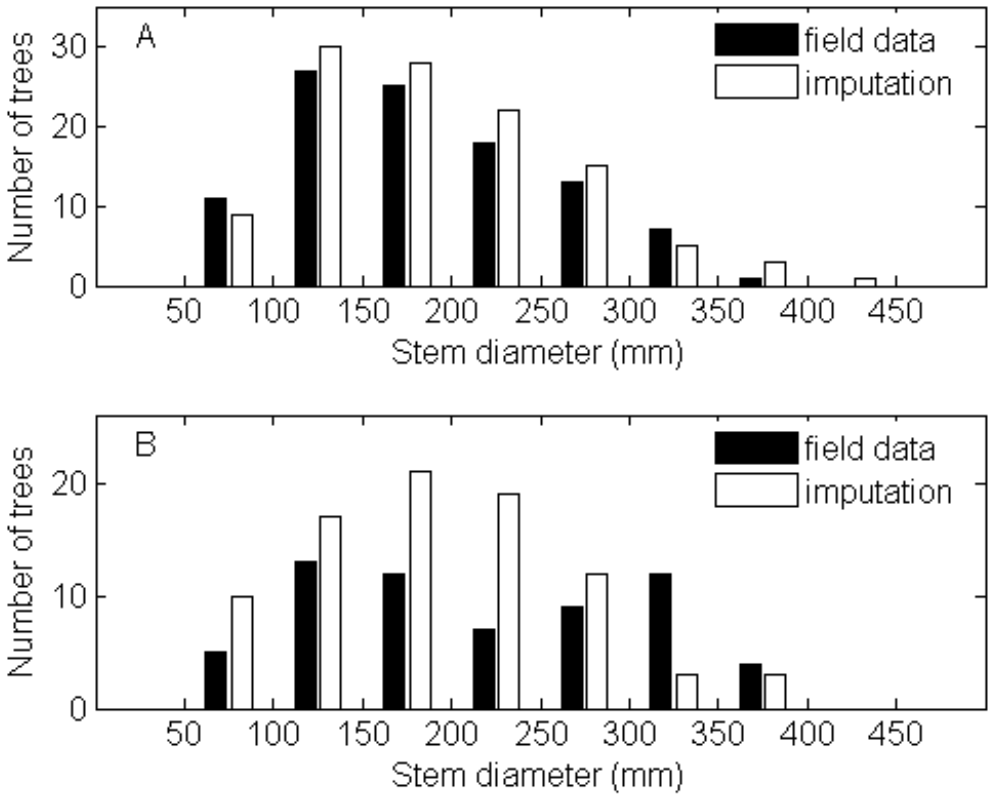


Figure 4. Number of stems in 50 mm stem diameter (DBH) interval from field data and imputed trees; (A) the sub-stand with lowest EI and (B) the sub-stand with highest EI; for both histograms was validation done by excluding segment from the same sub-stand from training data.

4. Discussion

The aim of this research is to develop an automatic method where crown polygons delineated from ALS data are used for imputation of tree stems measured by harvesting machinery. In an operational context, a reference database with tree stems linked to ALS data could be continuously updated.

Segments with no linked trees could be caused by false tree crown segments, linking errors, but also the lower size limit for a tree to be harvested. The linking errors could be caused by relative positioning errors of field trees, leaning trees, or problems to automatically correct the position of some field plots using field plot matching. The problem with either none or several trees linked to a segment was solved by using all crown segments for the imputation and using the sum of stem attributes (*e.g.* total stem volume of all trees linked to a segment) as reference variables. The tree crowns with none linked trees were usually estimated as segments with no or a low volume (*i.e.* usually less than 0.5 m^3). The estimates were aggregated to sub-stand level which produced estimates with low bias for stem volume, mean tree height, mean stem diameter, and stem density. The imputation results could possibly be improved by deriving more variables from ALS data, for example, those related to tree species. The results obtained from imputation of the sum of stem volume for none, one, or several trees linked to ALS derived segments are similar to results obtained by Breidenbach *et al*. (2010).

It is possible that MSN methods produce low accuracy estimates if a small training dataset is used because then the probability is high that there are no similar data points in the reference database. The most realistic validation of this study was to use cross validation where all segments from the same sub-stand were excluded as training data (procedure B). However, in this study, cross validation by excluding only segments from the same plot was also tested (procedure A), simulating a situation with a large amount of reference data as well as having training data similar to the stand to be predicted. The problem with this cross validation procedure is that unrealistic estimates could be produced if the within stand variation is low and the between stand variation of the data is high. In such a case the segments would be imputed from the same forest stand because of differences between stands. However, the RMSE values were similar for the two validation procedures. The RMSE values where similar to those obtained in other studies in Scandinavian forests with validation performed at stand level but were low compared to other studies where validations were performed at plot level (Næsset *et al*. 2004).

One advantage of the imputation of tree stems, compared with using for example regression analysis for prediction of stem attributes, is that imputed tree stems can be used for bucking simulations in order to predict product recovery. In the future the potential of the imputation of tree stems will be evaluated using bucking simulations. The remote sensing task is to find out if not only tree stems with similar sizes, as was evaluated in this study, can be imputed, but also if quality parameters can be predicted, such as tree species and number of branches.

Acknowledgements

This research was financed by the Swedish Research Council Formas, as part of the WW-IRIS project within the ERA-Net Wood Wisdom project, and by the Heureka research programme. We would like to thank Heather Reese for comments on the manuscript.

References

ANON, 2008a, The Standard for Forest Data (StanForD). Available online at: www.skogforsk.se (accessed 20 July 2010).

ANON, 2008b, The R Foundation for Statistical Computing (R-project). Available online at www.r-project.org (accessed 20 July 2010).

BREIDENBACH, J., NÆSSET, E., LIEN, V., GOBAKKEN, T., SOLBERG, S., 2010, Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. *Remote Sensing of Environment*, **114**, pp. 911-924.

HYYPPÄ, J., and INKINEN, M., 1999, Detection and estimating attributes for single trees using laser scanner. *The Photogrammetric Journal of Finland*, **16**, pp. 27-42.

HYYPPÄ, J., KELLE, O., LEHIKOINEN, M. and INKINEN, M., 2001, A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. *IEEE Transactions on Geoscience and Remote Sensing*, **39**, pp. 969–975.

KOCH, B., HEYDER, U. & WEINACKER, H., 2006, Detection of individual tree crowns in airborne lidar data. *Photogrammetric Engineering & Remote Sensing*, **72**, pp. 357–363.

MOEUR, M., and STAGE, A.R., 1995, Most Similar Neighbor - An improved sampling inference procedure for natural-resource planning. *Forest Science*, **41**, pp. 337-359.

NÆSSET, E., GOBAKKEN, T., HOLMGREN, J., HYYPPÄ H., HYYPPÄ, J., MALTAMO, M., NILSSON, M., OLSSON, H., PERSSON, Å., and, SÖDERMAN, U., 2004, Laser scanning of forest resources: the Nordic experience. *Scandinavian Journal of Forest Research*, **19**, pp. 482- 499.

OLOFSSON, K., LINDBERG, E., and HOLMGREN, J., 2008, A method for linking field-surveyed and aerial-detected single trees using cross correlation of position images and the optimization of weighted tree list graphs. In Proceedings of SilviLaser 2008: 8th international conference on LiDAR applications in forest assessment and inventory. Hill R.A., Rosette, J. and Suárez, J. (Eds). 17-19 September 2008, Edinburgh, UK., pp. 95-104.

PERSSON, Å., HOLMGREN, J. and SÖDERMAN, U., 2002, Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering & Remote Sensing*, **68**, pp. 925–932.

POLLOCK, R., 1996, The automatic recognition of individual trees in aerial images of forests based on a synthetic tree crown image model. PhD thesis, University of British Columbia, Vancouver, Canada.

REYNOLDS, M.R., BURK, T.E., and HUANG, W.-C., 1988, Goodness-of-fit tests and model selection procedures for diameter distribution models. *Forest Science*, **34**, pp. 373–399.

SOLBERG, S., NÆSSET, E. and BOLLANDSÅS, O. M., 2006, Single tree segmentation using airborne laser scanner data in a heterogeneous spruce forest. *Photogrammetric Engineering & Remote Sensing*, **72**, pp. 1369–1378.

Estimation of stem attributes using a combination of terrestrial and airborne laser scanning

EVA LINDBERG†, JOHAN HOLMGREN†, KENNETH OLOFSSON†, and HÅKAN OLSSON†*

Eva.Lindberg@srh.slu.se

† Swedish University of Agricultural Sciences, Department of Forest Resource Management, SE-90183 Umeå

Abstract

A new method to combine terrestrial laser scanning (TLS) with airborne laser scanning (ALS) has been evaluated for estimation of stem diameters (DBH) and stem volume at single tree level. The aim is to measure tree stems in field plots with TLS to use as training data for wall-to-wall estimation of forest variables based on ALS data. DBH and tree positions were estimated from TLS data in six field plots. Trees estimated from TLS data and tree heights and positions estimated from ALS data were co-registered to automatically link TLS and ALS measured trees. DBH estimated from TLS data and tree height measured in field for a sub-sample of trees were used to train regression models based on ALS derived tree crown segments. At tree level, the root mean square error (RMSE) was 46.6 mm (15.6%) for DBH, 9.8 dm (3.8%) for tree height, and 200.6 dm^3 (34.4%) for stem volume.

1. Introduction

Modern remote sensing techniques such as airborne laser scanning (ALS) have made it possible to estimate attributes of individual trees, e.g. height, diameter at breast height (DBH) and stem volume, if a sub-sample of trees with known positions is available as training data (Persson *et al.* 2002). To co-register tree crown segments from ALS data with the corresponding field trees, accurate tree positions measured in field are required. These measurements are usually done manually. Manual field surveys could be automated by introducing terrestrial laser scanning (TLS) (Thies *et al.* 2004). TLS provides highly accurate three dimensional data consisting of distance measurements from the scanner to surrounding surfaces.

The combination of data from TLS and ALS provides a possibility to implement a forest inventory system with minimal need for manual measurements. The combination of TLS and ALS has been used to measure canopy structure (Lovell *et al.* 2003) but we have not found any other studies yet where it has been used to estimate information about tree stems for forest management planning. Combination of remote sensing data at single tree level with field data requires co-registration of the data sources. Since GPS positions measured below a canopy are inaccurate, the positions of the field data must be adjusted (Olofsson *et al.* 2008). Co-registration of remote sensing data and data from TLS must also take the shaded zones in the TLS data into account.

The objective of this work is to create a method to use TLS data as training data to estimate attributes of tree stems from ALS data. As a first step, we have validated a new method to estimate DBH from TLS data. We have furthermore developed a method to automatically co-register the trees stems from the TLS data with the corresponding tree crown segments from the ALS data. Finally, the accuracies of the ALS estimates trained with TLS data have been compared with the corresponding ALS estimates trained with manual field measurements.

2. Material

Study area

The study area is located in the southwest of Sweden (lat. 58° N, long. 13° E). The dominating tree species are Norway spruce (*Picea Abies*), birch (*Betula spp*), and Scots pine (*Pinus Sylvestris*).

Field data

Six rectangular 80 × 80 m^2 field plots were allocated (table 1) during the autumn 2006. Within the field plots, all trees were measured using a calliper and the tree species were recorded. For a sub-sample of trees, the height was also measured. The positions of reference points near the field plots were measured using a RTK GPS with cm accuracy. The positions of all trees were measured relative to the reference points using a total station. Stem volume was calculated for the sub-sample of trees where the height was measured in field using functions by Brandel (1990). For deciduous trees, the function for birch was used. Linear regression models were fitted for the sub-sample of trees where the height was measured to estimate the stem volume of all trees for pine, spruce and deciduous trees.

Table 1. Summary of field plot data.

Plot	Number of trees	Species composition of basal area (%)			DBH (mm)		
		Spruce	Pine	Deciduous†	10^{th} percentile	Mean	90^{th} percentile
1	369	87	10	3	230	322	411
2	424	99	0	1	222	297	386
3	357	79	0	21	150	290	422
4	410	99	0	1	173	237	302
5	384	98	0	2	130	233	315
6	332	98	0	2	180	322	452
Total	2276	93	2	5	178	268	395

†5 % of the basal area originated from deciduous trees of which almost all were birch.

Airborne Laser Scanning, ALS, data

The acquisition of ALS data was performed on April 24, 2007 using a TopEye MKII ALS system with wavelength1024 nm carried by a helicopter. The flying altitude was 130 m above ground and average point density was 30 m^{-2}. First and last returns were saved for each laser pulse. Laser returns were classified as ground or non ground using a progressive Triangular Irregular Network (TIN) densification method (Axelsson 2000) in the TerraScan software (Soininen 2004) and the ground returns were used to derive a Digital Elevation Model (DEM).

Terrestrial Laser Scanning, TLS, data

The acquisition of TLS data was performed in September 2006 using an Optech ILRIS-3D with wavelength 1535 nm. The scan density was 0.27 × 10^{-3} rad horizontally and vertically. The laser scanner has a 40º×40º field of view. The complete 360º scenes were collected by scanning one scene at a time and rotating the scanner horizontally. The TLS reflections were rectified to create a complete point cloud for every scanner position using the software Polyworks. At least three common points were selected in each pair of scenes. The software rectified the scenes against each other, another scene was added and new common points were selected. Finally all scenes were rectified against each other at the same time and the TLS reflections were geo-referenced.

3. Methods

The Hough transform was used to find rough estimates of the positions and diameters of the tree stems from the TLS data. The estimates were used as initial values to fit circles along the tree stems to estimate the DBH. The result was validated against the field data from the manual field measurements.

The ALS data were segmented into tree crowns and features based on the ALS data were extracted for each segment. The segments were co-registered with the tree stems estimated from the TLS data and regression models were created with DBH, tree height and stem volume of the estimated tree stems as dependent variables and features from the ALS segments as independent variables.

Finally, the regression models were used to estimate DBH, tree height and stem volume based on ALS features. The result was co-registered with and validated against the field data from the manual field measurements. The creation of the regression models and the validation was done by using leave one out cross validation for one field plot at a time.

3.1 Processing of TLS data

Localisation of tree stems. To obtain positions of the tree stems, the TLS reflections were projected onto a raster image. Only data points within a small height span of 1 m, centred at breast height, were used. Tree stems in this image look like half circles facing the position of the sensor. Foliage can be visible as noise in the image (figure 1). Assuming that the tree stems look like arcs in the stem profile image, the Hough transform (Gonzalez and Wintz 1987) was used to find probable positions of the trees for four diameter classes, 20, 30, 40 and 50 cm. This gave several possible solutions. To find the optimal position for each tree, the result with the highest value in the Hough transform was chosen.

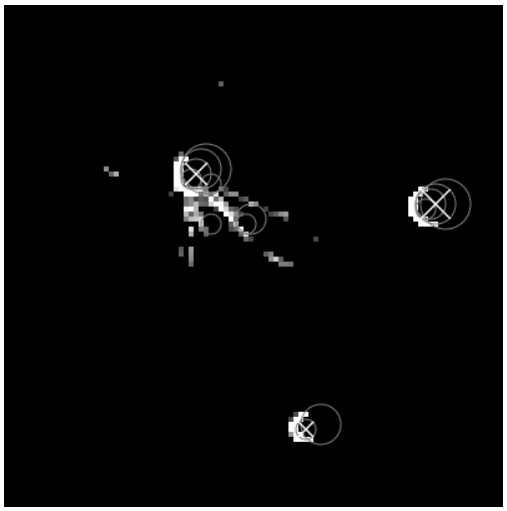


Figure 1. Raster of projected TLS data as white and gray pixels. Gray circles are the probable tree position candidates of different diameter classes.The final tree positions chosen by the algorithm are marked by a white X.

Estimation of diameters. Diameters along each stem were estimated by fitting circles in two steps. In the first step, the resulting position and diameter from the Hough transform were used to identify TLS reflections and circles were fitted to the TLS reflections by minimizing the sum of squared errors. In the second step, the positions and diameters of the fitted circles were used to identify a new set of TLS reflections with less noise than the first set and new circles were fitted to those TLS reflections. Finally, the DBH of the stem was estimated by estimating a linear function for the diameter at different heights and using that function to estimate the diameter at 1.3 m above ground.

To reduce the number of tree stems for which the estimation has a great risk of being inaccurate, the final list of tree stems included only tree stems ≤ 45 meter from the scanner (II in the result section) and with a set of TLS reflections with width ≥ 0.5 $\times$ r and ≤1.5 $\times$ r where r=the estimated diameter/2 (III in the result section) (figure 2).

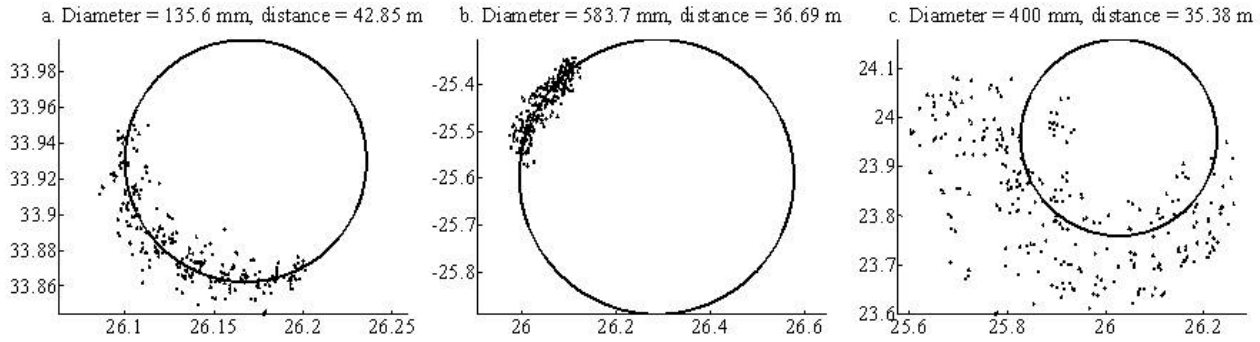


Figure 2. Examples of TLS reflections and fitted circles. a. The TLS reflections have a width ≥ 0.5 × r and ≤1.5 × r. b. The TLS reflections have a width < 0.5 × r. c. The TLS reflections have a width >1.5 × r.

3.2 Processing of ALS data

Delineation of tree crowns. Tree crowns were automatically segmented based on geometric tree crown models. A canopy height model (CHM) was derived with a raster cell value set to the maximum height above ground for laser returns within the raster cell. A crown area raster was created using morphological operations. CHM raster cells were updated using interpolation if they were at crown area locations but with no data or if the height value was below a threshold. A correlation surface (CS) was created using geometric models, i.e. generalized ellipsoids of revolution (GER). For all raster cells with CHM values above a threshold, correlations between laser heights and GER models were calculated and the raster value was set to the maximum correlation found at the specific location. The CS was smoothed using a Gaussian kernel. A starting point, a seed, was placed at each raster cell with a CHM value above a threshold and with a positive CS value. For each seed, the current location was updated by changing the position to the neighbour cell with the highest value of the smoothed CS. This was repeated until the position could not be updated because a local maximum of the smoothed CS was reached. The seeds with the final location at the same local maximum defined a segment. The next step was to merge segments. For each segment, the GER models were used to decide if the segment should be merged to a neighbour segment (Holmgren and Wallerman 2006, Holmgren *et al.* 2010).

3.3 Combination of ALS and TLS data

Co-registration of sources. TLS data from an area with trees have shaded zones where the data is obscured from the scanner. This can introduce errors when co-registering with aerial data. To avoid this problem, shaded zones were masked out in the field plots depending on where tree stems were found in the TLS data. To co-register with aerial data, the position-image-method by Olofsson, Lindberg et al. (2008) was further developed and adapted so that the shaded parts of the data were excluded. The tree linking algorithm was the same as in (Olofsson *et al.* 2008).

Estimation of forest variables. DBH, tree height and stem volume were estimated with regression models from features extracted from the ALS segments. The parameters of the regression models were estimated by using DBH estimated from the TLS data as training data and as a comparison by using values from the manual field measurements as training data. The result was two different sets of estimated parameters for each regression model. For the TLS data, the height of the closest tree where the height was measured in field was used in the regression model

For the trees from the manual field measurements, the regression model for the stem volume included only spruce trees. For the tree stems estimated from the TLS data, two different regression models were used; a separate model including only spruce trees and a weighted model including all trees where the volume of a

tree was calculated as a weighted mean of the stem volume of a pine, spruce and birch with the same DBH and height. The weights were the proportion of stem volume of each tree species in the training data.

3.4 Validation

The final ALS based estimates were validated using leave one out cross-validation. One field plot at a time was excluded from the dataset and the rest of the dataset was used as training data to estimate forest variables for the excluded field plot. The root mean square error (RMSE) and bias were calculated for the estimated DBH, tree height and stem volume at tree level.

The tree height was validated only for trees where the height was measured in field. The stem volume was validated only for spruce trees since two of the models were based only on spruce trees.

4. Results

Table 2 shows the RMSE and bias of DBH estimated from TLS data.

Table 2. RMSE and bias of DBH estimated from TLS data and validated against manual field measurements. I. All tree stems, II. Tree stems ≤ 45 meters from the scanner, III. Tree stems with a set of TLS reflections with width ≥ 0.5 × r and ≤1.5 × r, IV. Restrictions II + III.

	Number of stems measured from TLS data†	Number of stems with manually measured tree	Mean DBH (mm)	RMSE		Bias	
				(mm)	%	(mm)	%
I	1793	1032	290.7	38.0	13.1	1.6	0.5
II	1321	984	289.6	36.3	12.5	1.8	0.6
III	1708	1005	292.2	34.4	11.9	-0.2	-0.1
IV	1270	958	291.1	32.3	11.1	-0.1	0.0

†Including tree stems outside the field plots.

The mean DBH of the trees estimated from the TLS data that could be linked to tree crown segments from the ALS data was 300.3 mm, the RMSE was 34.0 mm and the bias was 0.8 mm. Tables 3-5 show the RMSE and bias of the DBH, tree height and stem volume estimated from ALS data.

Table 3. DBH estimated from ALS data.

Training data	RMSE DBH		Bias DBH	
	(mm)	%	(mm)	%
Trees estimated from TLS data	46.6	15.6	0.2	0.1
Trees from manual field measurements	45.5	15.2	0.1	0.0

Table 4. Tree height estimated from ALS data. The training data included only trees where the height was measured in field.

Training data	RMSE tree height		Bias tree height	
	(dm)	%	(dm)	%
Trees estimated from TLS data	9.8	3.8	0.4	0.1
Trees from manual field measurements	9.6	3.7	-0.3	-0.1

Table 5. Stem volume estimated from ALS data. The stem volume was validated only for spruce trees (see section Validation).

Training data	RMSE stem volume		Bias stem volume	
	(dm^3)	%	(dm^3)	%
Trees estimated from TLS data, separate model for spruce trees	200.6	34.4	34.1	5.9
Trees estimated from TLS data, weighted model for all tree species	199.0	34.1	21.0	3.6
Trees from manual field measurements, separate model for spruce trees	203.3	34.9	1.3	0.2

5. Discussion

The accuracy of the DBH estimated from TLS data was almost independent of the distance from the scanner which may be related to the small vertical field of view of the scanner. Estimates close to the scanner would probably be more accurate if TLS data from a scanner with a larger field of view were used. Additionally, noise reduction is an essential part of the analysis. However, some tree stems will always be obscured by other trees closer to the scanner. To solve this it is useful to estimate the accuracy of the estimated DBH directly from the TLS reflections. Since the estimates are used only as training data for the ALS data, it is then possible to select only the tree stems where the accuracy is likely to be high.

Several previous validations of estimates of DBH from TLS data have been based on small samples with around 10-20 trees (Thies and Spiecker 2004, Watt and Donoghue 2005, Henning and Radtke 2006, Bienert *et al.* 2007) but two larger samples have also been used (Hopkinson *et al.* 2004, Maas *et al.* 2008). The accuracy achieved in this study was higher than in one study based on time of flight TLS data collected in pine forest (Hopkinson *et al.* 2004) and lower than in one study based on continuous wave TLS data collected in mixed forest (Maas *et al.* 2008). Both studies were based on TLS data collected at multiple positions which was not the case in this study.

The combination of TLS data with remote sensing data requires co-registration of the data sources. One source of errors is erroneous linking of trees between the two data sources. The erroneous linking leads to errors in the training data and also has the effect that the accuracy appears lower than it actually is since the estimates are validated against the wrong trees.

The estimates of DBH and stem volume with trees estimated from TLS data as training data had a positive bias because the subset of trees that could be linked to tree crown segments from ALS data had a positive bias. The reason might be that the segmentation of tree crowns from ALS data identifies trees with large crowns and dense branches that cause an overestimation of DBH from TLS data. The small positive bias of the DBH was enlarged since the stem volume is a function of DBH to the power of 1.8-2.2 (Brandel 1990).

The bias was smaller for the estimates from a weighted model for all tree species, possibly because the estimation of DBH was less accurate for spruce than for the other tree species.

Previous studies have achieved slightly higher accuracy for estimation of DBH, tree height and stem volume from ALS data (Persson *et al.* 2002, Solberg *et al.* 2006, Vauhkonen *et al.* 2008, Vauhkonen *et al.* 2010). In this study, the RMSE of DBH was 45.5 mm with manual field measurements as training data and 46.6 mm with TLS data as training data. The corresponding numbers for tree height were 9.6 dm versus 9.8 dm and for stem volume 203.3 dm3 versus 200.6 dm3 (only spruce) and 199.0 dm3 (weighted model for all tree species). The reason for the lower accuracy with TLS data as training data is that the tree stems estimated from TLS data are not completely accurate. For the tree heights, the lower accuracy may also be related to the lower number of trees where the height was measured in field.

6. Conclusions

This study has presented a new method to combine ALS and TLS data for forest inventory. The method uses tree stems derived from TLS data as training data for tree crown segments from ALS data. The only manual measurements that are needed are the heights of a sub-sample of trees in each plot. This reduces the need for manual field measurements which means that a field inventory could be done more efficiently and with smaller risk of errors due to the human factor. The time spent on measuring the DBH could be used for collecting other data in the field plots. The accuracy of the estimates from ALS data with TLS data as training data was almost as high as with manual field measurements as training data. This means that the method has the potential to become a part of an automated forest inventory system when terrestrial laser scanning technology develops.

Acknowledgements

We would like to acknowledge Håkan Larsson at the Swedish Defence Research Agency who has collected, rectified and geo-referenced the TLS data. We would also like to acknowledge Blom Sweden who collected and pre-processed the ALS data. This project was financed by the Swedish Research Council Formas.

References

AXELSSON, P., 2000, DEM generation from laser scanner data using adaptive TIN models. International Archives of Photogrammetry and Remote Sensing, 33, pp. 111-118.

BIENERT, A., SCHELLER, S., KEANE, E., MOHAN, F., and NUGENT, C., 2007, Tree detection and diameter estimations by analysis of forest terrestrial laserscanner point clouds: ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007 Espoo, pp. 50–55.

BRANDEL, G., 1990, Volymfunktioner för enskilda träd: tall, gran och björk [Volume functions for individual trees: Scots pine (Pinus sylvestris), Norway spruce (Picea abies) and birch (Betula pendula& Betula pubescens)]. Swed. Univ. Agric. Sci., Dept. For. Yield Res, 26.

GONZALEZ, R., and WINTZ, P., 1987, Digital image processing. 1987: Reading, MA: Addison-Wesley), pp. 425-427.

HENNING, J. G., and RADTKE, P. J., 2006, Detailed stem measurements of standing trees from ground-based scanning lidar. Forest Science, 52, pp. 67-80.

HOLMGREN, J., BARTH, A., LARSSON, H., and OLSSON, H., 2010, Prediction of stem attributes by combining airborne laser scanning and measurements from harvesting machinery: Silvilaser 2010, the 10th International Conference on LiDAR Applications for Assessing Forest Ecosystems.

HOLMGREN, J., and WALLERMAN, J., 2006, Estimation of tree size distributions by combining vertical and horizontal distribution of laser measurements with extraction of individual trees: Workshop on 3D Remote Sensing in Forestry, pp. 168–173.

HOPKINSON, C., CHASMER, L., YOUNG-POW, C., and TREITZ, P., 2004, Assessing forest metrics with a ground-based scanning lidar. Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere, 34, pp. 573-583.

LOVELL, J. L., JUPP, D. L. B., CULVENOR, D. S., and COOPS, N. C., 2003, Using airborne and ground-based ranging lidar to measure canopy structure in Australian forests. Canadian Journal of Remote Sensing, 29, pp. 607-622.

MAAS, H. G., BIENERT, A., SCHELLER, S., and KEANE, E., 2008, Automatic forest inventory parameter determination from terrestrial laser scanner data. International Journal of Remote Sensing, 29, pp. 1579-1593.

OLOFSSON, K., LINDBERG, E., and HOLMGREN, J., 2008, A method for linking field-surveyed and aerial-detected single trees using cross correlation of position images and the optimization of weighted tree list graphs: SilviLaser 2008, 8th international conference on LiDAR applications in forest assessment and inventory., pp. 95-104.

PERSSON, A., HOLMGREN, J., and SODERMAN, U., 2002, Detecting and measuring individual trees using an airborne laser scanner. Photogrammetric Engineering and Remote Sensing, 68, pp. 925-932.

SOININEN, A., 2004, Terra Scan for MicroStation, user's guide. Terrasolid Ltd., Jyvaskyla, Finland, 132.

SOLBERG, S., NAESSET, E., and BOLLANDSAS, O. M., 2006, Single tree segmentation using airborne laser scanner data in a structurally heterogeneous spruce forest. Photogrammetric Engineering and Remote Sensing, 72, pp. 1369-1378.

THIES, M., PFEIFER, N., WINTERHALDER, D., and GORTE, B. G. H., 2004, Three-dimensional reconstruction of stems for assessment of taper, sweep and lean based on laser scanning of standing trees. Scandinavian Journal of Forest Research, 19, pp. 571-581.

THIES, M., and SPIECKER, H., 2004, Evaluation and future prospects of terrestrial laser scanning for standardized forest inventories. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36, pp. 192-197.

WATT, P. J., and DONOGHUE, D. N. M., 2005, Measuring forest structure with terrestrial laser scanning. International Journal of Remote Sensing, 26, pp. 1437-1446.

VAUHKONEN, J., KORPELA, I., MALTAMO, M., and TOKOLA, T., 2010, Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. Remote Sensing of Environment, 114, pp. 1263-1276.

VAUHKONEN, J., TOKOLA, T., MALTAMO, M., and 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, pp. 441-459.

Fourier transform of waveform Lidar for species recognition - data requirements

Nicholas R. Vaughn1 and L. Monika Moskal1

1School of Forest Resources, University of Washington, Box 352100, Seattle, WA 98195

Abstract:

Waveform Lidar information is typically analyzed only after decomposing waveforms into a sum of Gaussian peaks. Under the assumption that some important information may be lost in the decomposition, an attempt was made to transform the waveform into the spectral domain using a fast Fourier transform. This approach was successful at distinguishing three deciduous species with 75 % accuracy (kappa=0.62), using a classification tree approach.

The data set density used in this work was about 10 light pulses per square metre (lppm) near nadir at ground level. This allows for an analysis of data density effects on the ability of the classification method to correctly identify a given species. The data were reduced, by removing waveforms at uniform intervals, into subsets containing approximately 80, 60, 40, and 20 % of the original density. This resulted in densities of approximately 8, 6, 4 and 2 lppm.

Surprisingly, not all reductions of data were found to decrease the ability of this method to correctly identify tree species. In fact the 80 % density showed marginal improvement over the full density. The 60, 40 and 20 % densities decreased classification accuracy by 10 to 20 %. The results indicate that pulse density has only slight, yet sometimes unpredictable effect on the classification accuracy outcome.

1. Introduction

Airborne Lidar data has been shown to provide estimates of stand characteristics, such as height and canopy cover, that have very high precision (Næsset et al., 2004; Andersen et al., 2006). Being an active sensor, not dependent upon light conditions, Lidar data is often preferred over data from various aerial and space-borne sensors. However, species recognition is one area where Lidar has not yet excelled. For this reason it is still common for species information to be obtained from two-dimensional hyperspectral imagery. However, as Lidar instruments improve we should see corresponding improvements in the ability to classify species using Lidar data alone.

The most obvious differences between many tree species are those involving color and physical structure. Though Lidar typically works with only one “color”, a frequency in the near-infrared range of the spectrum, a few examples exist of some success using Lidar intensity data alone to discriminate tree species (Ørka et al., 2007; Kim et al., 2009). Others have worked with variables objectifying tree shape and structure in various ways (Liang et al., 2007; Brandtberg et al., 2003; Brandtberg, 2007). Encouraging results have also been achieved with the combination of intensity and structure variables (Vauhkonen et al., 2009; Holmgren and Persson, 2004; Kim, 2007; Ørka et al., 2009).

More recently, a slightly different form of Lidar, called waveform or fullwave Lidar, has become more readily available (Mallet and Bretar, 2009). This instrument digitises segments of the return signal at a very high sample rate, resulting in data that resemble a wave with peaks and troughs. Several authors have investigated the abilities of waveform Lidar to distinguish species characteristics. At first, this data was used to detect peaks that may be missed by a on-board peak detector system, which resulted in denser discrete point datasets. Reitberger et al. (2006) showed that such data was useful for species detection. To obtain these peak locations, waveforms are typically decomposed into a series of Gaussian or similar forms

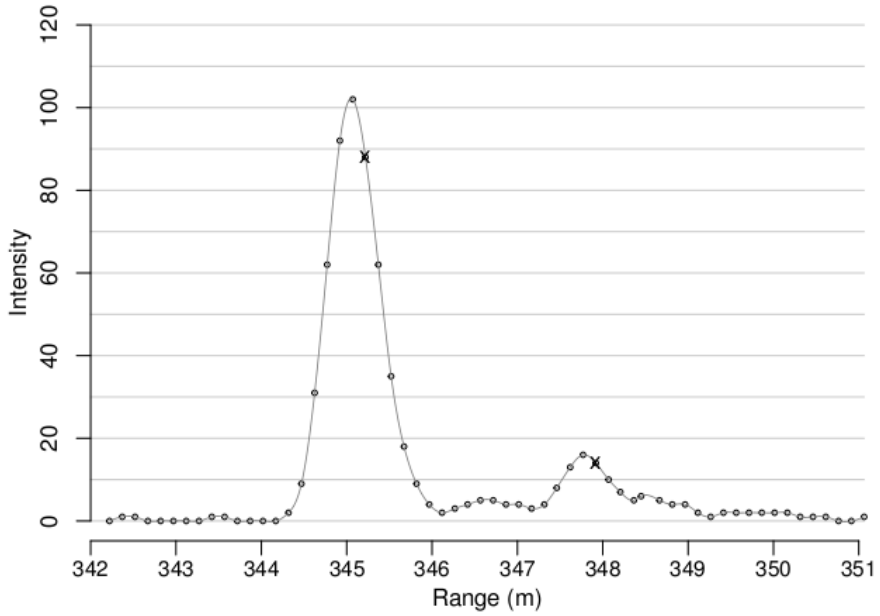
(Chauve et al., 2007). Some works suggest that storing shape parameters from individual peaks may lead to further improvements in classification (Wagner et al., 2008; Litkey et al., 2007; Hollaus et al., 2009).

As useful as individual peak modeling has been, information is always lost when simplifying data using any model with fewer parameters than data points. The pulse width of output signal on most Lidar systems in in the order of 0.5 to 2 m, and this signal has a strong smoothing effect on the shape of the returned signal. While deconvolving the data may reduce much of this smoothing, fitting the data to waveforms will filter out higher frequency patterns. Some of these patterns may help distinguish one species from another. We put this theory to test by transforming individual waveforms using a discrete Fourier transform (Vaughn et al., 2010). This linear transform rebuild the signal as a composition of sine waves, allowing one to analyze the importance of differing frequencies in the observed data. Results showed that wavelengths as low as 0.36 m were important for classification.

Spatial scale of raster data often has large influence on classification results, and optimal scales differ by data type and approach (Treitz and Howarth 2000, Marceau et al. 1994, Woodcock and Strahler 1987). Also depending on the application, discrete Lidar data density may also highly influence results. Liu et al. (2007) found that point density affects DEM accuracy. One advantage of this particular approach is that it should be less dependent upon data density, because the simple means used as discriminating variables for each tree should be stable with only relatively few waveforms hitting the tree. In this paper we apply the same transformation to datasets with reduced numbers of waveforms compared to the original dataset to test if this is indeed the case.

2. Methods

Figure 1: Difference between waveform and discrete Lidar data. The 60 waveform samples are shown as circles and a spline fit to these data appears as a solid gray line. A peak detector might detect two peaks at about 345 and 348 m and return the intensity value when the peak is detected as shown with exes. Due to inherent limitations, real time peak detection algorithms usually produce a slight lag in peak location.



The study site

The Washington Park Arboretum in Seattle, Washington is operated by the University of Washington Center for Urban Horticulture. The Arboretum, which is approximately 93 ha in size, is planted with more than 10,000 cataloged woody plant specimens representing numerous genera. In addition much of the Arboretum

contains several species native to Western Washington State, such as Douglas-fir (Pseudotsuga menziesii (Mirbel) Franco), western redcedar (Thuja plicata Donn ex D. Don), bigleaf maple (Acer macrophyllum Pursh), black cottonwood (Populus balsamifera L. ssp. trichocarpa (Torr. & A. Gray ex Hook.) Brayshaw), and red alder (Alnus rubra Bong.). The non-native trees in the arboretum are planted in groups by genus. Native species are also dispersed throughout arboretum and can be found clustered into their own groups or sparsely mixed within the non-native trees. While many of the planted trees are open-grown, overlapping crowns are typical within the native tree groups. However, tree densities are rarely as high as one would expect in natural stands in the vicinity.

Data processing

We applied waveform data provided by Terrapoint USA, who flew a RIEGL LMS-Q560 airborne laser scanner with waveform signal digitization, over the Washington Park Arboretum on August 8th, 2007. This instrument was set to digitise waveforms at a sample interval of about 1 ns, or 15 cm in linear distance. Scan angle ranged from -30 to 30 degrees, and the pulse frequency was set at 133,000 Hz, resulting in a pulse density of about 10 pulses per square metre (lppm) near nadir at ground level. For comparison, this would yield about 20 points per square metre in a comparable first and last return discrete point dataset. One example waveform is displayed in figure 1. The range from the instrument is shown on the horizontal axis, while the unitless intensity value is shown on the vertical axis. Two points are marked with exes indicate data points that might be returned by a traditional on board peak detector. A single 4.5 km looped pass in the North-South direction for the length of the Arboretum, lasting about 6 min provided nearly 49 million waveforms. Each waveform contained a minimum of 60 consecutive samples, which covers a linear distance of about 9 m. In many cases, depending on target height, the number of samples was 120 or even 180 for a given waveform, but this is not necessarily consecutive data. Because of the discrepancy in the number of waveforms, only the first 60 samples were kept from each waveform. This should cover 9 m of the path of the waveform starting from the surface of the target. As a result, in trees taller than about 9 m, ground strikes will not be recorded within the retained waveform data.

Within the same Arboretum, Kim et al. (2009) geolocated and measured characteristics of the trees in 18 field plots within the Arboretum. The locations of these plots are shown in figure 2. The field plots were installed systematically so that at least one plot is measured in each genus group of interest. Within each plot about ten to twenty example trees were identified and measured during the summer of 2005. Each of the trees measured in the field plots has been mapped into UTM coordinates using a angle and distance from one of three known points within the plots. These points were located with survey-grade GPS units and these data were later differentially corrected for optimal accuracy.

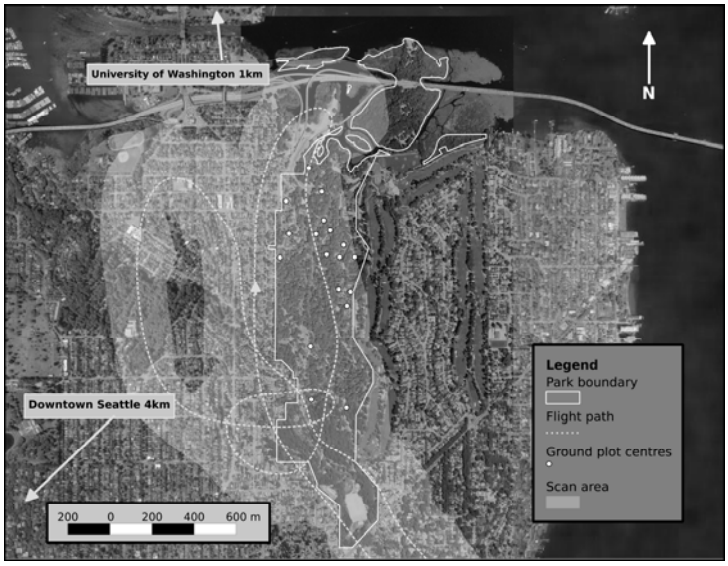


Figure 2: Map of the Washington Park Arboretum with flight path and ground plot locations.

In order to associate waveform data to individual trees on the ground, the tree crowns had to be delineated in mapping coordinates. We therefore used a discrete point dataset, built from the waveform dataset, to create a raster image containing a digital canopy height model. In this elevation model the highest return elevation, above the DEM, within each grid cell was stored. We used the method described by Hyyppä et al. (2001a) to obtain an initial set of polygons representing the crown outlines of individual trees in the

Arboretum. This method works in an iterative manner, such that at each step neighboring pixels are added to clusters surrounding local maxima of a low-pass filtered canopy height model. Under such an algorithm, groups of trees are often mistaken for single trees (Hyyppä et al., 2001b), though in the Arboretum this should be less of a problem. The resulting polygon for each tree in the training data set was visually inspected and, if necessary, improved upon by hand. All waveforms were identified that contained data at any height above ground within the outline of each tree.

Fourier Transform

In the analysis of time series, several tools are available to look for non-random patterns within the data. One such tool is the discrete Fourier transform, which allows one to look at the data in the “frequency domain”. In doing so, we may discover frequencies of strong influence within the time series. This transform converts the original data into a set of coefficients representing the individual influence of sine waves from a known set of frequencies. Large coefficients are associated with heavy influence and imply that a higher amount of periodicity at the given interval is detected in the data. Fast versions of the transformation exist under the name Fast Fourier Transform (FFT), and have a relatively low upper limit on computing time (Singleton, 1979).

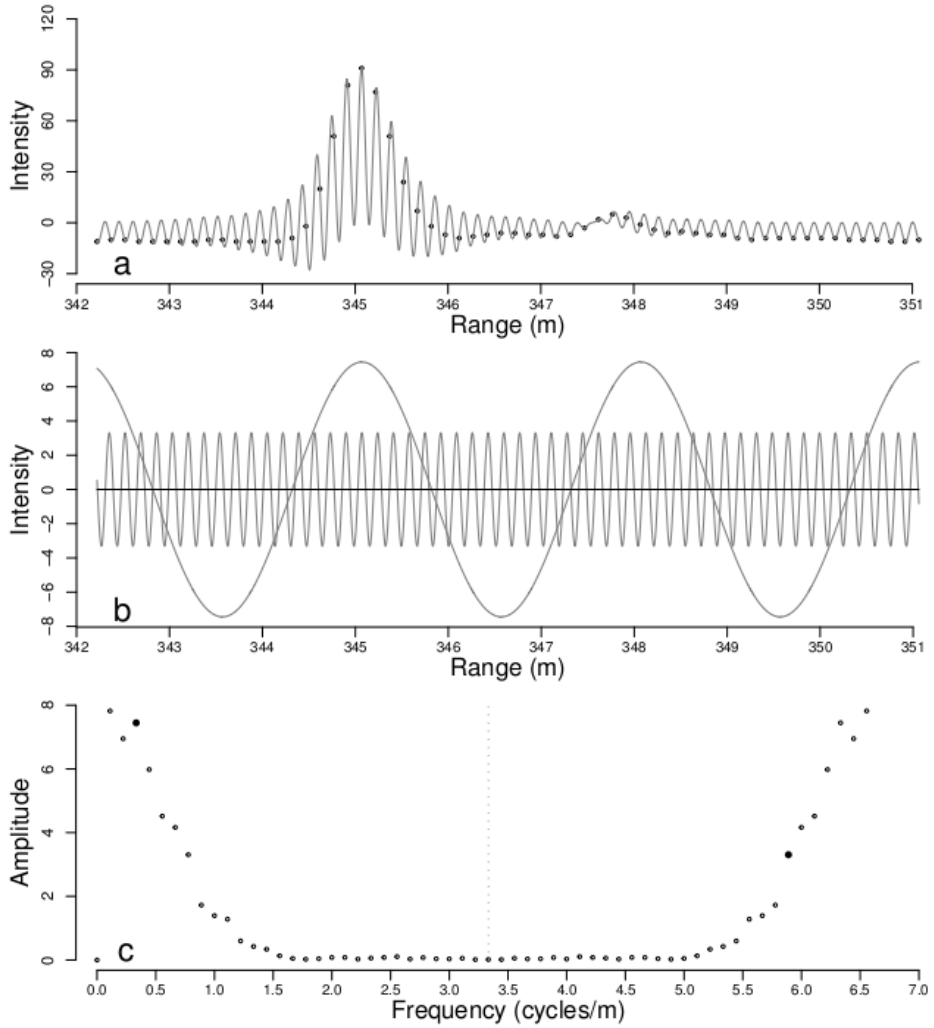


Figure 3: Example of spectral decomposition of one waveform. In panel a the composite wave returned by the transformation is plotted along with the original mean-centered intensity values. As the transformation disregards range information, the wave composition has been re-translated back into the range scale. In panel b, two examples of the 60 component waves are drawn. In panel c, the amplitudes of all 60 component waves are plotted against their frequency.

Figure 3 shows the transform of an example waveform taken from the training data. Typically, the mean is subtracted from each sample value. In figure 3(a), the mean centered waveform appears along with the complex waveform fit by the FFT algorithm. A result of modeling with the same number of variables as data is that all of the sample points all fall exactly on the composite wave. Two example component waves are depicted in figure 3(b). The amplitude of each wave represents the contribution of that particular frequency to the composite wave shown in figure 3(a). How the influences of these two example frequencies compare to the rest is depicted in figure 3(c). The first frequency is 0, and represents an intercept term. The rest of the amplitudes are symmetrical as per a restriction of the FFT algorithm. The fft function in the R programming language (R Development Core Team, 2009) was used to compute the FFT. In figure 3(c), the amplitudes of the component waves from figure 3(b) are shown with solid dots.

The frequencies of the component waves used by the FFT algorithm are determined entirely by the number of sample points. In order to ensure that the same frequencies are used by the transformations of two time series, one series must contain the same number, or a power of 2 multiple of the number, of samples in the other series. As mentioned above, 60 samples were kept from each waveform so that this condition could be met. However, this 60-sample waveform, stretching about 9 m will cross the boundary of some tree crown outlines. This means that some waveforms will contain data for parts of neighboring trees. We decided that no action would be done to correct for this due to the added difficulties this would create.

As waveforms in this study were restricted to have 60 samples each, the number of amplitude values returned by the FFT algorithm is then also 60. The influence of frequencies above a given level, known as the Nyquist frequency, cannot accurately be measured. As a result, with 60 samples per waveform, we can consider the amplitudes of only the first 30 frequencies to be useful. Not coincidently, this is the point at which the amplitudes start to mirror those of lower frequencies in figure 3(c). This mirroring is a restriction used by the algorithm to keep the number of variables, from being greater than the number of sample points.

For each tree, the averages (across all waveforms hitting a tree) for each of these 30 useful amplitudes were stored as variables named with a leading “M” followed by the frequency identification number (M1 through M30). Additionally, the standard deviations of each frequency were recorded as variables V1 through V30. Additionally, the average intensity value was kept for each waveform. This easily computed value represents the total amount of light reflected from each pulse. The average and standard deviation of these values for each tree were recorded as MI and VMI. In total, for each tree in the training data set, there are 62 variables that will be considered for use in classification as described in the next section.

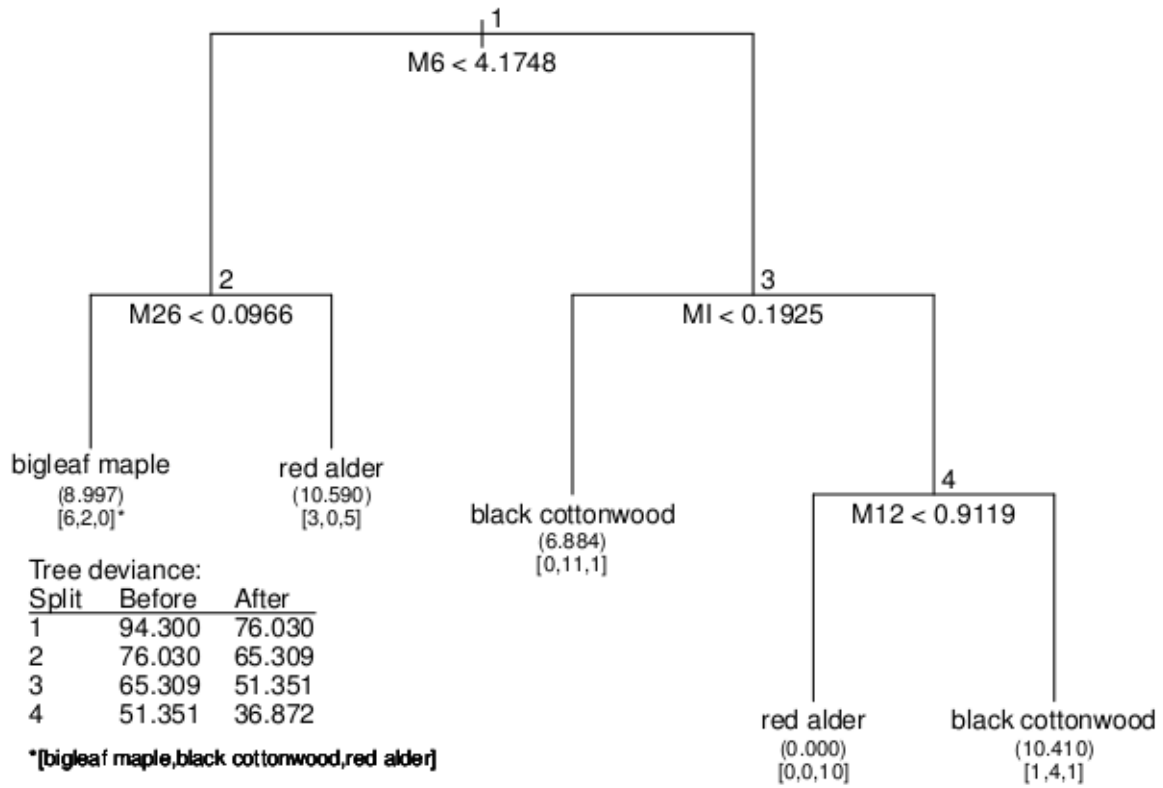
The FFT algorithm assumes an equal time period between samples, however in some cases the range values within each pulse data are not equally spaced. To greatly simplify analysis, these facts were ignored, as a violation of this assumption is not too concerning in this case. The displacement of an occasional sample point should have little impact on the results. In most cases the difference in intensity between two neighboring samples is very small. We are also ignoring that our series is not of a periodic origin, as are our sinusoidal basis functions. It does little harm to pretend that our series repeats itself in both directions ad infinitum.

Classification

To apply the FFT information to classify tree species, we used a classification tree approach (Breiman et al., 1984, page 18). Classification tree algorithms recursively split the data into two parts based on a value of the most locally powerful predictor variable. The R library tree contains a function of the same name for modeling with these regression and classification trees (Venables and Ripley, 2002, page 266). Given a class variable as a response and a list of potential predictor variables, the function will compute a tree of “appropriate” size. Here appropriate is determined by an internal algorithm. Splits are added to the tree's branches sequentially, until a very large tree is produced. At each split, the variable that most reduces the tree deviance under a

multinomial model is chosen. Cross-validation is used to determine the optimal tree size, as too large of a tree will overfit the data, and too small a tree will perform poorly.

Figure 4: The result of a classification tree fit to the full-density training data. Listed above each split is the variable and level used to split the remaining trees into two subgroups. The reduction in tree deviance resulting from each split is shown in the bottom left. Leaf deviances are shown in parenthesis, actual class memberships in square brackets. The variables M6, M12, and M26 correspond to the mean influence of the wavelengths 1.80, 0.82 and 0.36 m, respectively.



We attempted to correctly classify all trees belonging to three native hardwood species: red alder, black cottonwood, and bigleaf maple. These species represent common hardwood species that grow naturally in the arboretum, and therefore are represented well in the field data. Figure 4 shows the classification tree obtained by fitting the entire training dataset. Each fork represents the optimal separation of the remaining trees into two sub-groups. The variable and split value used is shown atop the fork. Each end node is labeled with the species most represented in the group of trees that have not been eliminated when traversing the tree from the root. Below each leaf the deviance within that leaf and actual class membership are presented. The total tree deviance is the sum of the individual leaf deviances, and the reductions in tree deviance as each split is added are shown in a table in the bottom left of the figure. One branch of this tree was “pruned” because both leaves predicted the same species.

With limited training data available, the trained model could not be applied to a separate dataset for validation. Therefore, a cross-validation technique was used to estimate the actual predictive power of this technique on new data. The species of each tree was determined from a tree trained from all the other trees in the training data. The numbers involved in such a process makes refinement of each tree unpractical, and thus each tree was built from the built-in defaults of the tree function. In a non-academic application, the tree building could be better optimised and this may result in slight improvements in the classification accuracy.

Data reductions

To test the technique at different levels of data density, the original dataset was systematically reduced. Within the waveform data for each tree in the training data, we divided the waveforms into five groups of waveforms representing every fifth waveform. Each group started from one of the first five waveforms. For example the first group was comprised of the first waveform, sixth waveform, eleventh waveform, and so on. Thus the reduced datasets were created by sequentially removing these groups starting from the fifth group and down to the second group. The resulting new datasets then contained 80, 60, 40, and 20 % of the original waveforms. Data densities represented by these datasets were approximately 8, 6, 4, and 2 lppm, respectively. For each of these reduced datasets and the original dataset the classification process described above was performed, and results were recorded in order to be put into table form for subsequent comparison.

3. Results and Discussion

Table 1 shows the classification results for the full-density dataset. The overall classification accuracy, or the portion of correctly classified trees, was 75 %. The associated kappa value was 0.615. For individual species, 70 % of maples, 82 % of cottonwoods and 71 % of alders were correctly classified. While differences in leaf reflectance likely play a part in the results, we believe that differences in tree structure lead to stronger classification.

As presented in a previous paper (Vaughn et al., 2010), the wavelengths most often chosen by the classification tree algorithm as partitioning variables are 1.80, 0.82 and 0.36 m. These wavelengths likely correspond differences between species across different components of a tree. For example the leaves of bigleaf maple are typically about 15-30 cm in width, while the leaves of red alder and black cottonwood are much smaller. This difference might be expected to show up in shorter wavelengths, and this is the case in the example tree shown in figure 4. The variable M26 helps distinguish red alder from bigleaf maple. The two longer wavelengths may represent differences in branch to branch distance and leaf retention rates between the species. Red alder has fairly thin leaves allowing more visible light through, and may be able to retain more leaves further into the canopy than the other species.

The full-density results were surprising as we had expected the red alders and cottonwood trees to be more easily separated from the bigleaf maples. This is because the maples represented in the dataset are nearly all open-grown and the leaves of bigleaf maple are a much larger target than those of other trees. The larger leaves should provide at least a more consistent first peak height. The cottonwoods and red alders are growing in closer proximity and appear to have similar growth forms to the naked eye under this condition. However, judging from an near-infrared raster image created from the discrete point data, red alder and black cottonwood also have very similar reflectance of the near-infrared wavelength used by the RIEGL LMS-Q560. This implies that structural differences between the two species contributed highly to the classification. As branch to branch distances may play a role, we wonder if tree growth rate would affect the results. More needs to be done to figure out what features of the tree are contributing most to the classification abilities of this technique.

Table 1 Results of the classification when each tree species is predicted from a classification model incorporating all other trees using the full-density dataset.

	Predicted			
Species	bigleaf maple	cottonwood	red alder	Producer accuracy (%)
bigleaf maple	7	2	1	70.0
cottonwood	2	14	1	82.4
red alder	1	4	12	70.6
User accuracy (%)	70.0	70.0	85.7	75.0†
†This number represents overall accuracy.				

Table 2 shows how the technique responded to different densities of waveforms. Amazingly, the 80% density resulted in improved performance. Additionally, the 20% density provided greater classification ability than

both the 60% and the 40% densities. The kappa values behaved in a similar manner, as expected when the same trees are used for all densities. These results suggest that the height, and perhaps the speed, at which a waveform Lidar mission is flown should not greatly affect the results of this classification technique as much as the would affect techniques that may rely on high point cloud density.

Table 2: The classification results under systematically reduced datasets representing 100, 80, 60, 40 and 20 % of the original density of waveforms.

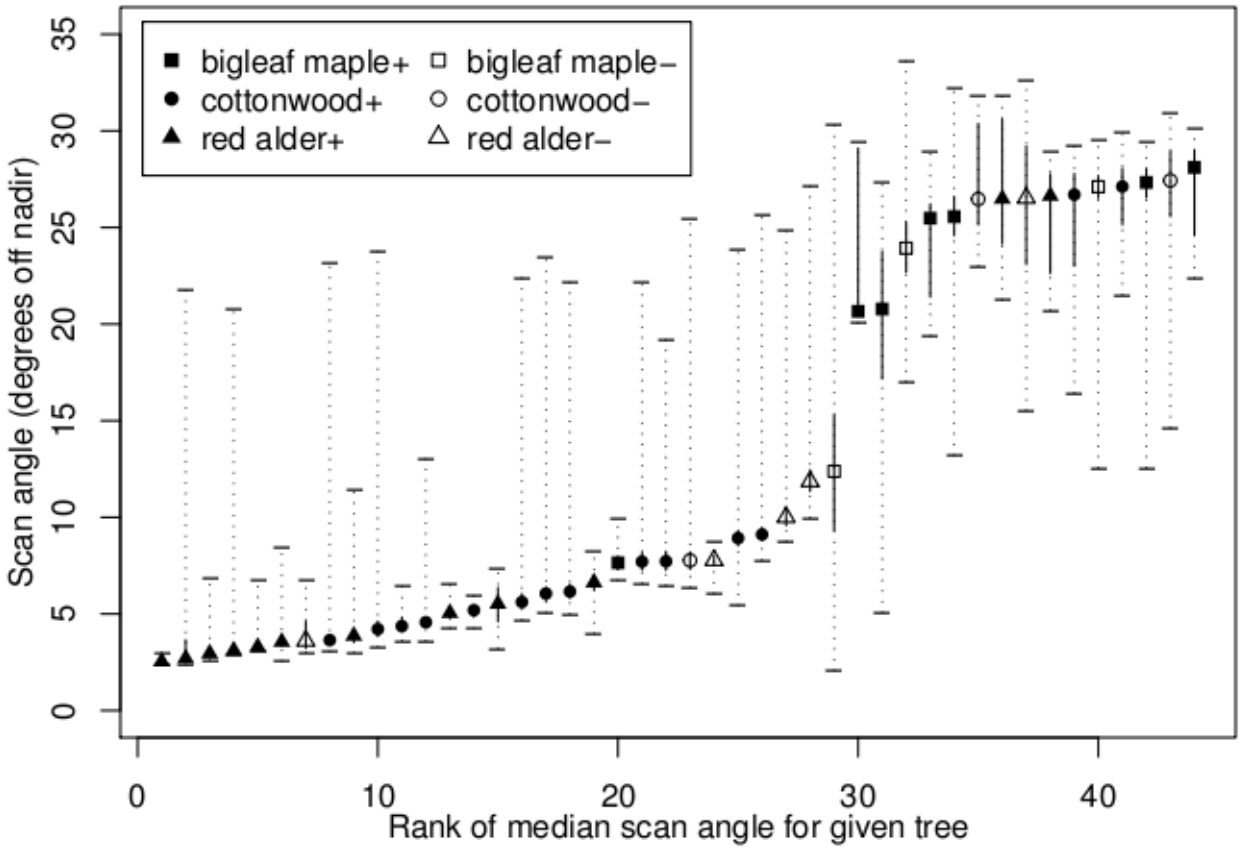
Density	Species	Producer accuracy†	User accuracy	Kappa
(%)		(%)	(%)	
100	all	75.0		0.62
	bigleaf maple	70.0	70.0	
	cottonwood	82.4	70.0	
	red alder	70.6	85.7	
80	all	81.8		0.72
	bigleaf maple	90.0	75.0	
	cottonwood	82.4	73.7	
	red alder	76.5	100.0	
60	all	63.6		0.43
	bigleaf maple	50.0	62.5	
	cottonwood	64.7	52.4	
	red alder	70.6	80.0	
40	all	54.6		0.29
	bigleaf maple	30.0	37.5	
	cottonwood	76.5	56.5	
	red alder	47.0s	61.5	
20	all	65.9		0.48
	bigleaf maple	60.0	54.6	
	cottonwood	70.6	66.7	
	red alder	64.7	73.3	
†For all species this number represents overall accuracy.				

We believe that the demonstrated robustness of the technique to data density is due mainly to the fact that the stronger classification variables are actually sample means. The sample mean is a very efficient estimator of a population mean, and large numbers of samples are not needed to get a fairly good estimate. Therefore, whether the mean is calculated from thousands of waveforms or simply hundreds, the sample mean should be very close. However, such a reduction of data should not be expected to have no effect as the sample mean can be affected by extreme outliers in the data. If, when collecting a waveform Lidar dataset, one ends up with a higher proportion of unusual waveforms, the results would suffer accordingly.

There are several technical difficulties that were overlooked in order to more directly test the efficacy of the FFT for species detection. Finding methods to address any of these difficulties will likely improve upon the

results presented here. First, we did nothing to account for the differences in scan angle between trees and species. Because tree structure differs horizontally from vertically, scan angle likely plays a part in the dominant wavelengths that are seen in the FFT of the waveform data. The dataset does not provide enough coverage to test this technique for a standard range of scan angles for all trees. Figure 5 shows the full range of scan angle for each tree in the training data. In this figure correctly classified trees are indicated by filled-in symbols. There is a clear discrepancy in the number of correct classifications in the trees with generally higher scan angles on the right side of the figure.

Figure 5: The range of scan angles found for each tree in the dataset. Dotted lines represent the full range of scan angle, while solid lines represent the 25th to the 75th percentiles. Trees correctly classified in the full dataset are represented with filled-in symbols at the tree's median scan angle.



A second simplification was the reduction of all waveforms to exactly 60 samples due to requirements of the FFT. There were two cases where this might severely affect results. The first case is the loss of data because more than 60 samples were available in a given waveform. About 62, 32, and 5 % of waveforms contained 60, 120, and 180 samples respectively. Another case is when waveforms cross crown outlines, such that only a portion of the samples contained in the waveform pertain tot he given tree. In this case some of the waveform data for a tree is actually describing other trees of unknown species. These simplifications may have a drastic effect. However, a standardization of the number of samples in a waveform is by far the easiest way to ensure that the component frequencies modeled by the FFT are the same across all waveforms.

4. Conclusion

The technique described above, despite some simplifications, shows much promise. One important feature is that, despite significant reduction in data density, the technique did not respond with large decreases in effectiveness. This is due largely to the sample mean being a highly efficient estimator of population mean. As such, any modification of this technique that incorporate other statistics may not scale as well as we have

seen here. However, the number of samples in a waveform should be the same despite the details of the Lidar acquisition, and methods that rely only on the density within a waveform should see similar results. Standardization of scan angles along with future increases in crown segmentation accuracy should not affect this feature of the method, and will likely lead only to improvements in accuracy.

Acknowledgments

The authors wish to thank Terrapoint USA for providing the dataset used in this study, as well as the University of Washington Precision Forestry Cooperative, the University of Washington Remote Sensing and Geospatial Analysis Laboratory for their support. Also Bob McGaughey and Steve Reutebuch at the USDA Forest Service Pacific Northwest Research Station provided their experience and their assistance in working with the raw and processed data.

Bibliography

ANDERSEN, H., REUTEBUCH, S. AND MCGAUGHEY, R., 2006, A rigorous assessment of tree height measurements obtained using airborne lidar and conventional field methods. Canadian Journal of Remote Sensing, 32, pp. 355-366.

BRANDTBERG, T., 2007, Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar. ISPRS Journal of Photogrammetry and Remote Sensing, 61, pp. 325-340.

BRANDTBERG, T., WARNER, T.A., LANDENBERGER, R.E. AND MCGRAW, J.B., 2003, Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in North America. Remote Sensing of Environment, 85, pp. 290-303.

BREIMAN, L., FRIEDMAN, J.H., OLSHEN, R.A. AND STONE, C.J., 1984, Classification and Regression Trees, The Wadsworth statistics/probability series (Belmont, California: Wadsworth, Inc).

CHAUVE, A., MALLET, C., BRETAR, F., DURRIEU, S., PIERROT-DESEILLIGNY, M. AND PUECH, W., 2007, Processing full-waveform lidar data: modelling raw signals. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36, pp. 102-107.

HOLLAUS, M., MÜCKE, W., HÖFLE, B., DORIGO, W., PFEIFER, N., WAGNER, W., BAUERHANSL, C. AND REGNER, B., 2009, Tree species classification based on full-waveform airborne laser scanning data. In Proceedings of the Silvilaser 2009 Proceedings, October, College Station, TX.

HOLMGREN, J. AND PERSSON, Å., 2004, Identifying species of individual trees using airborne laser scanner. Remote Sensing of Environment, 90, pp. 415-423.

HYYPPÄ, J., KELLE, O., LEHIKOINEN, M. AND INKINEN, M., 2001a, A segmentation-based method to retrieve stem volume estimates from 3-d tree height models produced by laser scanners. IEEE Transactions on Geoscience and Remote Sensing, 39, pp. 969-975.

HYYPPÄ, J., SCHARDT, M., HAGGRÉN, H., KOCH, B., LOHR, U., PAANANEN, R., SCHERRER, H., LUUKKONEN, H., ZIEGLER, M., HYYPPÄ, H., PYYSALO, U., FRIEDLÄNDER, H., UUTTERA, J., WAGNER, S., INKINEN, M., WIMMER, A., KUKKO, A., AHOKAS, E. AND KARJALAINEN, M., 2001b, HIGH-SCAN: The first European-wide attempt to derive single-tree information from laserscanner data. The Photogrammetric Journal of Finland, 17, pp. 58-68.

KIM, S., 2007, Individual tree species identification using LIDAR- derived crown structures and intensity data. PhD thesis, College of Forest Resources, University of Washington, Seattle, WA.

KIM, S., MCGAUGHEY, R., ANDERSEN, H. AND SCHREUDER, G., 2009, Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. Remote Sensing of Environment, 13, pp. 1575-1586.

LIANG, X., HYYPPÄ, J. AND MATIKAINEN, L., 2007, Deciduous-coniferous tree classification using difference between first and last pulse laser signatures. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36, pp. 253-257.

LITKEY, P., RÖNNHOLM, P., LUMME, J. AND LIANG, X., 2007, Waveform features for tree identification. In Proceedings of the Proceedings of the ISPRS Workshop: Laser Scanning 2007 and SilviLaser 2007, September, pp. 258-263.

LIU, X., ZHANG, Z., PETERSON, J. AND CHANDRA, S., 2007, The effect of LiDAR data density on DEM Accuracy. In Proceedings of the Proceedings of the MODSIM07 International Congress on Modelling and Simulation, pp. 1363-1369.

MALLET, C. AND BRETAR, F., 2009, Full-waveform topographic lidar: State-of-the-art. ISPRS Journal of Photogrammetry and Remote Sensing, 64, pp. 1-16.

MARCEAU, D., HOWARTH, P. AND GRATTON, D., 1994, Remote sensing and the measurement of geographical entities in a forested environment. 1. The scale and spatial aggregation problem. Remote Sensing of Environment, 49, pp. 93-104.

NÆSSET, E., GOBAKKEN, T., HOLMGREN, J., HYYPPÄ, H., HYYPPÄ, J., MALTAMO, M., NILSSON, M., OLSSON, H., PERSSON, Å. AND SÖDERMAN, U., 2004, Laser scanning of forest resources: the Nordic experience. Scandinavian Journal of Forest Research, 19, pp. 482-499.

ØRKA, H., NÆSSET, E. AND BOLLANDSÅS, O.M., 2007, Utilizing airborne laser intensity for tree species classification. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36, pp. 300-304.

ØRKA, H., NÆSSET, E. AND BOLLANDSÅS, O., 2009, Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. Remote Sensing of Environment, 113, pp. 1163-1174.

R DEVELOPMENT CORE TEAM, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria 2009.

REITBERGER, J., KRZYSTEK, P. AND STILLA, U., 2006, Analysis of full waveform LiDAR data for tree species classification. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 36, pp. 228-233.

SINGLETON, R.C., 1979., Mixed Radix Fast Fourier Transforms. In Programs for Digital Signal Processing I.D.S.P. Committee (Ed.) (IEEE Press).

TREITZ, P. AND HOWARTH, P., 2000, High Spatial Resolution Remote Sensing Data for Forest Ecosystem Classification:: An Examination of Spatial Scale. Remote sensing of environment, 72, pp. 268-289.

VAUGHN, N., MOSKAL, L.M. AND TURNBLOM, E.C., 2010, Fourier transformation of waveform Lidar for species recognition. Remote Sensing Letters In press.

VAUHKONEN, J., TOKOLA, T., PACKALEN, P. AND MALTAMO, M., 2009, Identification of Scandinavian Commercial Species of Individual Trees from Airborne Laser Scanning Data Using Alpha Shape Metrics. Forest Science, 55, pp. 37-47.

VENABLES, W.N. AND RIPLEY, B.D., 2002, Modern applied statistics with S (Springer Verlag).

WAGNER, W., HOLLAUS, M., BRIESE, C. AND DUCIC, V., 2008, 3D vegetation mapping using small-footprint full-waveform airborne laser scanners. International Journal of Remote Sensing, 29, pp. 1433-1452.

WOODCOCK, C. AND STRAHLER, A., 1987, The factor of scale in remote sensing. Remote Sensing of Environment, 21, pp. 311-332.

Estimation of stem volume by using 3D tree segments derived from full waveform LiDAR data

JOSEF REITBERGER†, MARCO HEURICH‡ and PETER KRZYSTEK†*

josef.reitberger@hm.edu

†University of Applied Sciences München, Department of Geoinformatics, 80333 Munich, Germany

‡Bavarian Forest National Park, Department of Research, 94481 Grafenau, Germany

Abstract

Recently, full waveform LiDAR could be successfully applied to segment single trees in 3D even in the lower forest layers and to classify deciduous and coniferous trees at a high accuracy level of 95%. This paper highlights the stem volume estimation of single trees which are represented by 3D tree segments as a result of a novel segmentation approach using the normalized cut algorithm. In this segmentation the forest area is subdivided into a voxel space and a bipartite graph which is formed by the voxels and similarity measures between the voxels. This procedure has several advantages compared to segmentations using only on the canopy height model (CHM). Most important is the capability to identify dominated trees without local maxima in the CHM and small trees in the lower forest layers. The paper focuses on the interesting question whether this new 3D segmentation improves the estimation of single tree parameters like the stem volume. The segmented trees are represented by alpha shapes which envelop the laser points of the trees and provide a measure for the crown volume. Several features (i.e. crown volume, tree and crown height and maximum projected crown area) are derived from the segments. From the segment features and a reference stem volume derived from field measurements (i.e. diameter at breast height, tree height) the parameters of a multiple linear regression model are estimated using least squares. The experiments were conducted in the Bavarian Forest National Park using full waveform data captured with the Riegl LMS-Q560 scanner at a point density of 25 points/m^2 both in leaf on and leaf off condition. In the case of coniferous trees the study shows a high coefficient of determination (R^2) of about 0.90 both for the conventional watershed segmentation and the new normalized cut segmentation. In the case of deciduous trees R^2 increases from 0.66 to 0.79 in leaf off condition and from 0.67 to 0.74 in leaf on condition for the normalized cut segmentation.

1. Introduction

The development of new approaches to forest inventory utilising remote sensing data has been an important research issue in the past. Beside area based methods, techniques for single tree extraction from LIDAR data have been investigated for mapping forests at the tree level and for identifying important parameters, such as tree height, crown size, crown base height, tree species, and stem volume. Novel methods for single tree detection tackle conceptually the segmentation problem with a 3D approach instead of using only the CHM (Wang et al., 2008). In combination with full waveform data Reitberger et al. (2009) could successfully demonstrate that the detection rate of single trees could be significantly improved in overall terms. Interestingly, the improvement was most changing in the lower forest layers with 20%. Thus, it seems that the fusion of 3D techniques and full waveform data push the single tree approach to a new level of accuracy. Consequently, other parameters like the estimation of the stem volume might be enhanced using the 3D volume of segmented trees. Recently, stem volume estimation was investigated by several groups. Hyyppä et al. (2005) combine ALS first/last pulse data with CIR orthophoto and get an overall accuracy of 25 to 30% in terms of random errors. For smaller tree groups the random error amounts to 34 to 40%. Heurich (2006) uses in his experiments first/last pulse data acquired by the TopoSys system at a nominal point density of 10 points/m^2. Based on a 2D watershed segmentation the multiple regression ends up with an RSME error of 27% for coniferous trees and 35% for deciduous trees. Finally, Vauhkonen et al. (2010) report on RSME errors for stem volume of 31% when imputing single tree attributes using tree height, intensity and alpha shape metrics. It is common for all reports mentioned here, that the imprecision of the field data affects the

results significantly. Furthermore, the study of Maltamo et al. (2007) shows that timber volume can only be estimated with about 30% RMSE accuracy if derived as a function of the tree height and the crown diameter.

The objective of this paper is (i) to review a new method that segments single trees in 3D with the normalized cut segmentation, (ii) to show how the tree volume is approximated by alpha shapes, (iii) the method to correlate field data with variables derived from the tree crown and the 3D segments, and (iv) to explain experimental results about the tree species classification and the stem volume estimation.

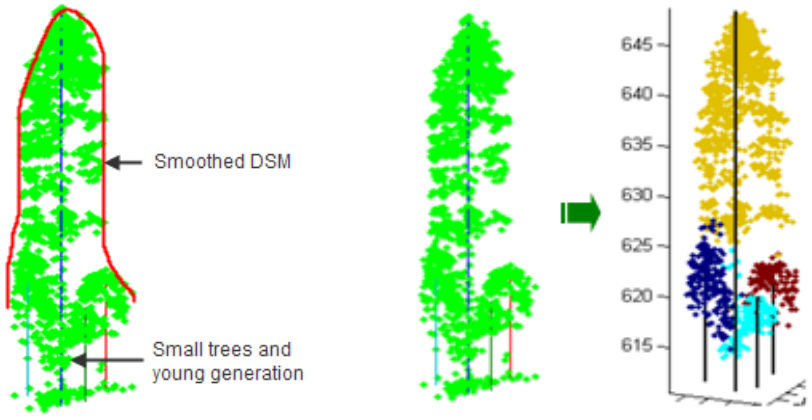
The paper is divided into seven sections. Section 2 briefly illustrates the overall concept beginning from the full waveform data. In section 3 the single tree segmentation and the tree species classification are highlighted. Section 4 shows the stem volume derivation from segment features using a multiple regression. Section 5 illustrates the results which have been obtained from full waveform data acquired in May 2006 and 2007 by the Riegl LMS-Q560 scanner in the Bavarian Forest National Park. Finally, these results are discussed with conclusions in sections 6 and 7.

2. Concept

Our concept for stem volume estimation comprises the following steps. At first, the waveforms are decomposed by a robust adjustment scheme to obtain 3D points and their attributes (Reitberger et al., 2008a). Then, individual trees are delineated by a watershed segmentation of the CHM. Furthermore, a novel 3D segmentation method which bases upon the normalized cut algorithm is applied. The next step is a tree species classification into deciduous and coniferous trees using characteristic features of the individual segments. Finally, the regression coefficients between the stem volume of reference data and characteristics of the segments are calculated. In the case of watershed segments the projected crown area, the tree height and the tree species are used as characteristics. For the normalized cut segments the crown volume and the height of the crown are introduced as additional characteristics.

Single tree segmentation and tree species classification

The watershed segments of the trees are found by applying the algorithm of Vincent and Soille (1991) to the CHM, which is previously derived from the 3D coordinates of the laser reflections (Reitberger et al. 2008a). Thereby an appropriate smoothing of the CHM is necessary to reduce the effects of over- and under-segmentation.
However, all the CHM based segmentation methods suffer from the drawback that dominated trees in the intermediate and lower tree height level cannot be recognized since they are invisible in the CHM (Figure 1a).



a) b)

Figure 1. Comparison of different segmentation methods in the case of smaller trees in the understorey. a) Watershed segmentation and b) Normalized cut segmentation

The key idea of the new 3D segmentation technique (Reitberger et al., 2009) is to rigorously apply a normalized cut segmentation (Shi and Malik, 2000) to a voxel representation of the forest area. This makes is possible to detect also smaller trees which are not represented by local maxima in the CHM (Figure 1b). This segmentation uses the positions (x_i, y_i, z_i) of the laser reflections and optionally the pulse width W_i and the intensity I_i of the waveform decomposition. Additionally, stem positions provided either by a special stem detection method (Reitberger et al., 2007) or by a watershed segmentation of the CHM can be used as prior knowledge.

The normalized cut segmentation applied in the voxel structure of a ROI is based on a graph *G*. The two disjoint segments A and B of the graph are found by maximizing the similarity of the segment members and minimizing the similarity between the segments *A* and *B* (Figure 2a) solving the cost function

$$NCut(A,B) = \frac{Cut(A,B)}{Assoc(A,V)} + \frac{Cut(A,B)}{Assoc(B,V)} \quad (1)$$

with $Cut(A,B) = \sum_{i \varepsilon A, j \varepsilon B} w_{ij}$ as the total sum of weights between the segments A and B and $Assoc(A,V) = \sum_{i \varepsilon A, j \varepsilon V} w_{ij}$ as the sum of the weights of all edges ending in the segment A. The weights w_{ij} specify the similarity between the voxels and are a function of the LiDAR point distribution and features calculated from W_i and I_i. A minimum solution for (1) is found by means of a corresponding generalized eigenvalue problem (Reitberger et al., 2009). It turned out that the spatial distribution of the LiDAR points mainly influences the weighting function. The features derived from the LiDAR points attributes from W_i and I_i only support in second instance the segmentation result. Furthermore, we found that prior knowledge about the position of the single tree significantly backs the segmentation. Thus, the local maxima of a CHM as a result of a watershed segmentation or stem positions – provided by a special stem detection technique – improve the weighting function and, hence, the segmentation result. Note that the approach is not dependent from full waveform LiDAR data. It can also successfully be applied to conventional LiDAR data just providing 3D point coordinates. The figure 2b shows complex situations where the normalized cut segmentation works excellent and a conventional watershed segmentation fails.

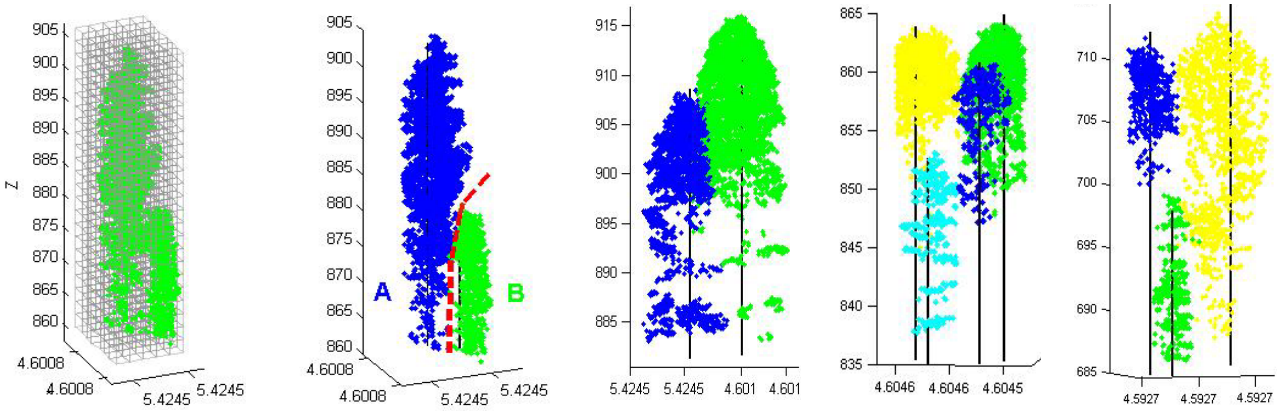


Figure 2. Single tree segmentation using normalized cut. a) Subdivision of ROI into a voxel structure and division of voxels into two tree segments and b) Segmentation results with the reference trees as black lines

For the tree species classification we introduce four groups of saliencies $S_t = \{S_g, S_I, S_W, S_n\}$ which are calculated from the N_t LiDAR points $\boldsymbol{X}_i^T = (x_i, y_i, z_i, W_i, I_i)$ $(i = 1,...,N_t)$ in the segments. Table 1 summarizes the definition of the most significant saliencies (Reitberger et al., 2008a).

Table 1. Definition of saliencies ("Sal.") used in classification

Sal.	Definition	Sal.	Definition
S_g^2	Mean horizontal distances of layer points to tree trunk	S_W	Mean pulse width of single and first reflections in the entire tree segment
S_I^2	Mean value of calibrated intensity in entire tree segment	S_n^2	Relation of the number of single reflections to the number of multiple reflections

Both an unsupervised and a supervised classification are performed in order to identify coniferous and deciduous tree species. The Expectation-Maximization algorithm (Heijden et al., 2004) turned out as suitable for an unsupervised clustering into the main tree species. The supervised classification was a maximum likelihood classification using an appropriate number of reference trees as a training subset (Reitberger et al., 2008b).

2.2 Derivation of stem volume

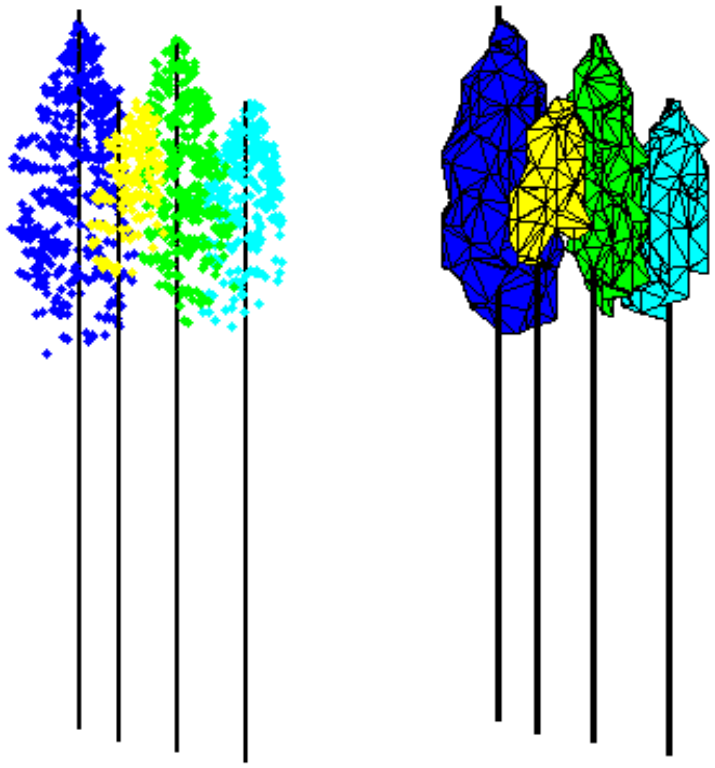
Since the stem volume cannot be directly measured from ALS data an indirect calculation is conducted according to equation (2) using measurable segment features (Table 2).

$$V_{stem} = b_0 + b_1 H_{tree} + b_2 A_{crown} + b_3 V_{crown} + b_4 H_{crown} + b_5 H_{tree}^2 + b_6 A_{crown}^2 + b_7 V_{crown}^2 + b_8 H_{crown}^2 \quad (2)$$

The coefficients $b_0 \ldots b_8$ in (2) are estimated in a least squares adjustment whereby V_{stem} is the known stem volume of reference trees. In the case of the normalized cut segments the crown points are separated from possible existing stem reflections at first in order to calculate the parameters V_{crown} and H_{crown}. According to Reitberger et al. (2007) these points are all the laser hits above the crown base height h_{base}. Finally, the 3D Alpha-Shape triangulation (Figure 3) was performed by making use of the Computational Geometry Algorithms Library (CGAL).

Table 2. Segment features used for derivation of stem volume

Segment feature	Definition	Watershed segments	Ncut segments
H_{tree}	Height of tree	Difference between highest laser reflection within segment and interpolated DTM height	
A_{crown}	Area of tree crown	Area of watershed polygon	Maximum 2D convex hull of laser points in segment
V_{crown}	Volume of tree crown	-	Crown surface by using 3D Alpha-Shapes
H_{crown}	Height of tree crown	-	Difference between highest and lowest crown point



a) b)

Figure 3. Example for crown surface triangulation by using 3D Alpha-Shapes. a) Laser points of four normalized cut segments and b) the corresponding crown triangulation

3. Experiments

Experiments were conducted in the Bavarian Forest National Park (49° 3' 19" N, 13° 12' 9" E), which is located in South-Eastern Germany along the border to the Czech Republic. There are four major test sites of size between 591 ha and 954 ha containing sub alpine spruce forest, mixed mountain forest and alluvial spruce forest as the three major forest types.

3.1 Field data

18 sample plots with an area size between 1000 m^2 and 3600 m^2 were selected in two test sites E and C. Reference data for all trees with DBH larger than 10 cm have been collected for 688 Norway spruces (*Picea abies*), 812 European beeches (*Fagus sylvatica*), 70 fir trees (*Abies alba*), 71 Sycamore maples (*Acer pseudoplatanus*), 21 Norway maples (*Acer platanoides*) and 2 lime trees (*Tilia Europaea*). Tree parameters like the DBH, total tree height, stem position and tree species were measured and determined by GPS, tacheometry and the 'Vertex III' system. The stem volume of the reference trees was calculated from DBH, tree height and species specific parameters according to Kennel (1973). Furthermore, the trees are subdivided into 3 layers with respect to the top height h_{top} of the plot, where h_{top} is defined as the average height of the 100 highest trees per ha (Heurich, 2006). The lower layer contains all trees below 50% of h_{top}, the intermediate layer refers to all trees between 50% and 80% of h_{top}, and finally, the upper layer contains the rest of the trees. Naturally, the reference data were updated for the individual flying dates of the LiDAR data. Table 3 summarizes the characteristics of the individual sample plots.

Table 3. Characteristics of sample plots

Plot name	21	22	55	56	57	58	59	60	64	65	74	81	91	92	93	94	95	96
Age [a]	160	160	240	170	100	85	40	110	100	100	85	70	110	110	110	110	110	110
Size [ha]	0.20	0.20	0.15	0.23	0.10	0.10	0.10	0.10	0.12	0.12	0.30	0.30	0.36	0.25	0.28	0.29	0.25	0.30
Altitude [m]	860	885	610	640	765	710	810	890	835	875	720	690	764	767	766	768	750	781
N/ha	500	540	830	340	450	440	2150	380	430	810	700	610	260	170	240	250	240	200
N lower layer	37	19	77	31	0	10	76	8	13	53	11	29	31	13	7	15	6	30
N interm. layer	14	60	21	19	4	4	85	22	4	26	33	59	11	3	2	4	0	3
N upper layer	48	29	20	27	41	30	54	27	35	35	165	96	54	27	59	54	53	26
Deciduous [%]	66	79	5	10	0	14	1	100	87	96	29	100	75	100	66	97	10	86

3.2 LiDAR data

Table 4. Characteristics of ALS campaigns

Time of flight	May '06	May '07
Data set	I	II
Foliage	Leaf-off	Leaf-on
Scanner	Riegl LMS-Q560	Riegl LMS-Q560
Pts/m^2	25	25
HAAT [m]	400	400
Footprint [cm]	20	20

LIDAR data of several ALS campaigns are available for the test sites. Full waveform data have been collected by Milan Flug GmbH with the Riegl LMS-Q560 scanner. Table 4 contains details about the point density, leaf-on and leaf-off conditions during the flights and the footprint size. The term point density is referring to the nominal value influenced by the pulse repetition frequency (PRF), flying height, flying speed and strip overlap.

3.3 Segmentation and tree species classification

The watershed segmentation and the normalized cut segmentation were applied to all the plots and data sets in a batch procedure without any manual interaction (Tables 5 and 6). The tree positions from the segmentation are compared with reference trees if (i) the distance to the reference tree is smaller than 60% of the mean tree distance of the plot and (ii) the height difference between h_{tree} and the height of the reference tree is smaller than 15% of h_{top}. If a reference tree is assigned to more than one tree position, the tree position with the minimum distance to the reference tree is selected. Reference trees that are linked to one tree position are so-called 'detected trees' and reference trees without any link to a tree position are treated as 'non-detected' trees.

Table 5. Results of different segmentation methods with data set I (leaf off)

Method	Detected trees per height layer [%]				False pos. [%]
	lower	intermediate	upper	total	
Watershed	5	21	77	48	4
Normalized Cut	21	38	87	60	9

Table 6. Results of different segmentation methods with data set II (leaf on)

Method	Detected trees per height layer [%]				False pos. [%]
	lower	intermediate	upper	total	
Watershed	5	20	79	48	4
Normalized Cut	17	32	86	58	10

If we focus on data set I, we can highlight how the normalized cut segmentation is superior to the watershed segmentation. The watershed segmentation performs in this case rather poor in the lower forest layer and detects almost no trees. In total, it ends up at an overall detection rate of 48%. Instead, the detection rate of the normalized cut segmentation is by 16% better in the lower and intermediate layer. Even in the upper layer the improvement is 10%, whereas the overall detection rate increases by 12%. Apparently, the waveform decomposition detects not only the reflections of the dominant trees but also of the dominated small trees. The 3D segmentation takes full advantage of this and yields clearly better segmentation results.

When we compare data set I (leaf-off) and data set II (leaf-on) we can also address the question whether the foliage condition affects the detection rate. As expected, the detection rate is in leaf-on situation by approximately 4% worse in the lower and intermediate layer for the normalized cut segmentation. But in the upper layer the results are almost equal, leading to an overall loss of accuracy of 2% in leaf-on situation.

We apply a supervised classification between deciduous and coniferous trees to the watershed and normalized cut segments, where one fifth of the trees were randomly selected as a training data set by keeping the proportion between the tree species. The classification method was applied 20 times in order to minimize the impact of the selection procedure on the results. Thus the numbers in table 7 refer to averaged classification values.

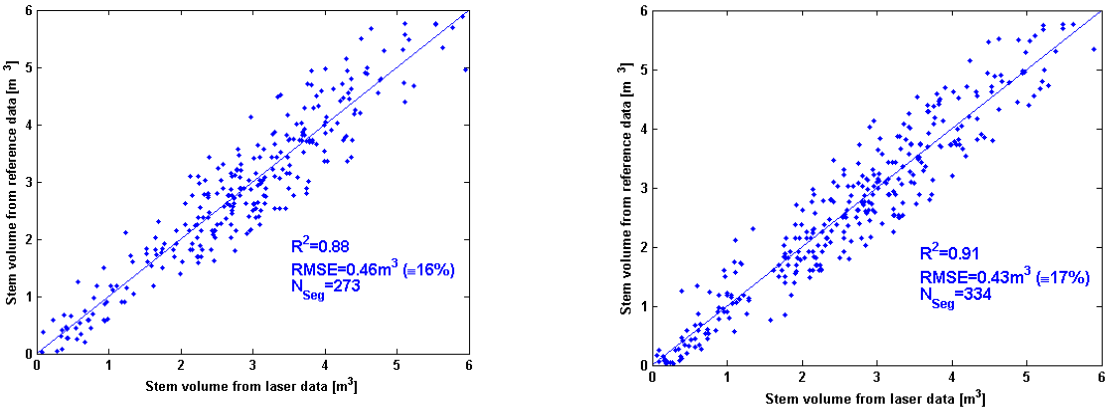
Table 7. Results of supervised ML classification between deciduous and coniferous tree species for segments of data set I (leaf off) and data set II (leaf on)

Saliency	Overall accuracy (%) for data sets I and II (supervised classification)			
	I		II	
	Watershed	Normalized Cut	Watershed	Normalized Cut
S_g^2	83	78	85	82
S_I^2	81	81	93	94
S_W	82	79	53	51
S_n^2	94	93	65	63
S_g^2, S_I^2, S_W, S_n^2	96	94	95	95

The results show that the classification works best with a combination of all saliencies. Furthermore, the classification results are almost the same for the watershed and normalized cut segments. Apparently, the applied saliencies are highly characteristic both for the dominant trees in the upper layer and for the dominated trees in the lower and intermediate layer. Moreover, except for the tree shape related saliency S_g^2 it is not relevant whether one or more trees of the same tree species are in one segment. Most interestingly, the intensity related saliency S_I^2 turns out as the best feature in the leaf-on case (data sets II). As expected, this feature is worse in the leaf-off case. Instead, the saliency S_n^2 dedicated to penetration behavior of the waveforms influences the classification results significantly in the leaf-off case (data set I). The saliency S_W dedicated to the pulse width of the reflections works in general better in the leaf-off case. Finally, the saliency S_g^2 representing the tree geometry has an almost constant impact on the classification in leaf-on and leaf-off situations. As far as the foliage condition is concerned, data sets I and II show equal classification accuracies, where the significance of the individual saliencies is different.

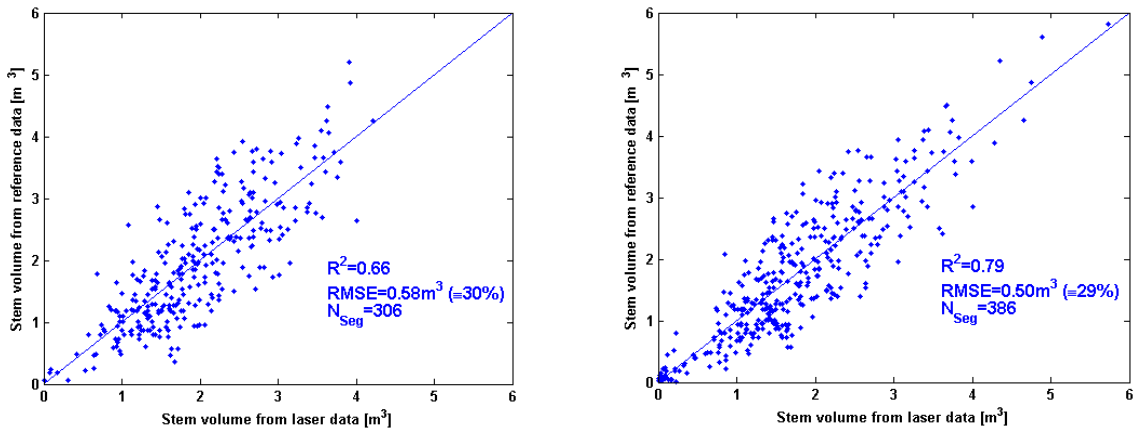
3.4 Derivation of stem volume

The results of the stem volume derivation for the different date sets, tree species and segmentation methods are illustrated in figures 4, 5, 6 and 7.



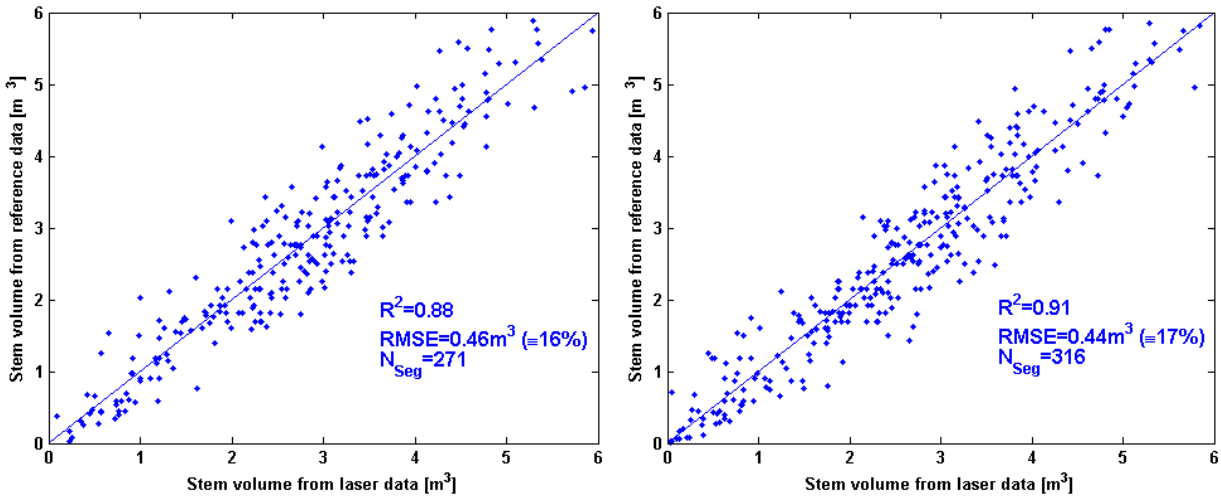
a) b)

Figure 4. Stem volume regression for coniferous trees using data set I (leaf off). a) Watershed segments and b) Normalized cut segments



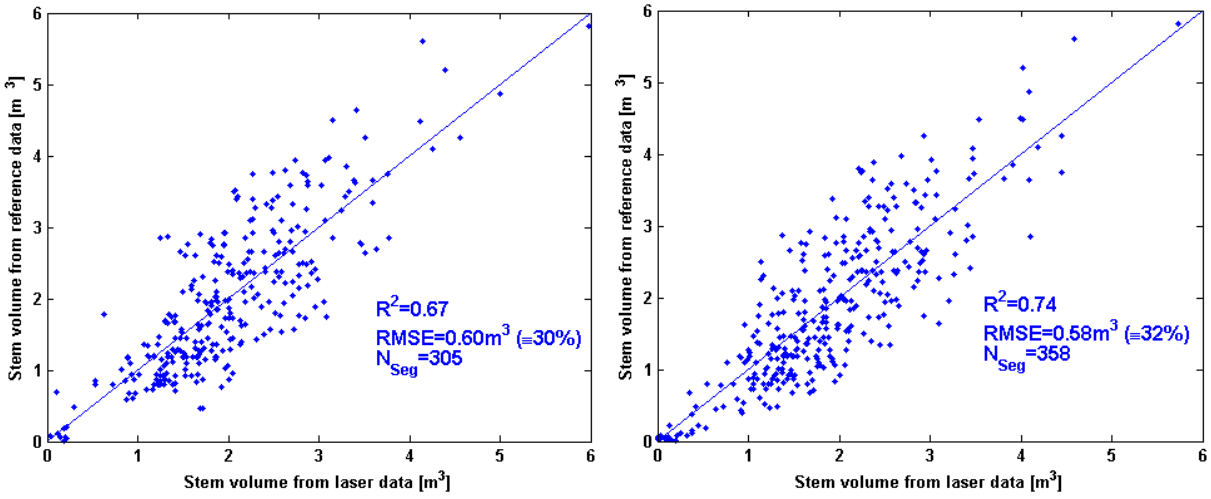
a) b)

Figure 5. Stem volume regression for deciduous trees using data set I (leaf off). a) Watershed segments and b) Normalized cut segments



a) b)

Figure 6. Stem volume regression for coniferous trees using data set II (leaf on). a) Watershed segments and b) Normalized cut segments



a) b)

Figure 7. Stem volume regression for deciduous trees using data set II (leaf on). a) Watershed segments and b) Normalized cut segments

A comparison of the different tree species shows that the stem volume of coniferous trees can be estimated more accurately than the stem volume of deciduous trees. This finding is in accordance with the results of Heurich (2006).

If we focus on the segmentation method a higher coefficient of determination R^2 in the case of the normalized cut method becomes apparent. This increase is marginal for coniferous trees, but very obvious for deciduous trees, especially in the leaf off case. The RMSE [m^3] of the normalized cut segmentation gets smaller than for the watershed segmentation. Thus, the difference between calculated and measured stem volume is smaller for the normalized cut segmentation in absolute terms. However, the relative RMSE [%] is practically the same for both segmentation methods, since the smaller RMSE of the normalized cut segmentation is compensated by smaller mean value of stem volume.

It is also remarkable that due to the higher detection rate of the normalized cut segmentation the stem volume is determined for more trees. Especially for data set I, which represents the leaf-off case, the number of small deciduous trees is considerably higher. Interestingly, we found that the calculated regression coefficients are suitable both for these small trees and for the tall trees.

Finally, the results are virtually the same for both data sets. The most significant difference is reflected in the 3D segmentation of deciduous trees. Here, the estimated stem volume is more accurate in the leaf-off case.

4. Discussion

The experiments clearly prove that the new segmentation working in 3D can be viewed as a breakthrough for single tree segmentation. If this method is combined with full waveform LIDAR data the detection rate of single trees is improved by more than 20% in the lower forest layers. We got a slightly higher detection rate in leaf-off situation because of the higher penetration rate of laser hits. Furthermore, the segmentation of deciduous trees is more difficult with ALS data acquired in summer since the crowns of neighboring trees merge considerably and, hence, most of the laser hits are reflected at the leaves of the outside section of the trees. Contrary, if ALS data result from a winter campaign more reflections occur on the thicker tree parts (e.g. stem, branches) which leads to a better characterization of the tree structure. Thus, the leaf-off situation seems to be the more appropriate flying time to segment trees in 3D, at least for mixed mountain forests that are scanned with a high point density.

When viewing at the results of the classification between coniferous and deciduous trees we attained an overall accuracy of approximately 95% for both data sets and segmentation methods. These excellent

results are provided mainly by the waveform specific saliencies S_n^2 in leaf off and S_l^2 in leaf on condition. Note, that a high classification result is mandatory for the stem volume estimation, since the regression models must be calculated separately for deciduous and coniferous trees.

The results of the stem volume estimation are very good for the coniferous trees. However, the stem volume of deciduous trees is approximated by the factor 2 worse. One reason for this is the imprecision of the reference stem volume for deciduous trees influenced mainly by the irregular shape of their stems. Deciduous stems are often oval instead of round, not cylindrical, and sometimes the stem branches at a height of only few meters. In comparison, the coniferous stems range from the ground to the crown and are structured more regularly. Thus, the manually measured parameters DBH and tree height lead to a more accurate reference stem volume for coniferous trees. Another reason for the superior stem volume estimation of coniferous trees is their consistent relation between crown dimension and stem volume. The crown basal area of spruces increases continuously with the tree age and, furthermore, spruces react slowly to exemptions from neighboring trees. Contrary, beeches respond rapidly to changing lighting conditions, which leads to a modified relation between crown dimension and stem volume.

All in all, the results are better than the findings of Heurich (2006), who used solely first/last pulse data in the same test areas and a 2D watershed segmentation. Thus, it seems that the full waveform data push this accuracy gain for both segmentation methods. It is remarkable that the relative accuracy (i.e. RMSE [%]) of the stem volume is independent from the segmentation method for both tree species. However, we have to keep in mind that the normalized cut segmentation detects significantly more trees correctly. Thus, the stem volume is provided for more trees.

5. Conclusions and outlook

Since the stem volume estimation is strongly influenced by factors like field measurements and allometric assumptions, more accurate reference data are necessary to evaluate the advantage of the 3D segmentation. Thus, measurements with a terrestrial laser scanner will lead to a more precise stem volume of the reference trees and to a verification of the volume of the 3D segments. Moreover, we need more insight how the features in (2), which are calculated from the 3D segments, are influenced by segmentation errors like the wrong assignment of voxels to a neighboring tree.

References

Heijden, F. van der, Duin, R.P.W., Ridder, D. de and Tax, D.M.J., 2004, *Classification, parameter estimation and state estimation – An engineering approach using MATLAB*. John Wiley & Sons Ltd, The Attrium, southern Gate, Chichester, West Sussex PO19 8SQ, England

Heurich, M., 2006, Evaluierung und Entwicklung von Methoden zur automatisierten Erfassung von Waldstrukturen aus Daten flugzeuggetragener Fernerkundungssensoren. *Forstlicher Forschungsbericht München*. Nr. 202, ISBN 3-933506-33-6. http://meadiatum2.ub.tum.de/. (Accessed December 16, 2008).

Hyypää, J., Mielonen, T., Hyypää, H., Maltamo M., Yu X., Honkavara E., Kaartinen, H., 2005. Using individual tree crown approaches for forest volume extraction with aerial images and laser point clouds. *ISPRS WG III/3 V/3 Workshop "Laser scanning 2005",* Enschede, the Netherlands, September 12 – 14.

Kennel, E., 1973, Bayerische Waldinventur 1970/71. *Forstliche Forschungsberichte München.* Nr. 11, 136 S.

Maltamo, M., Packalén, P., Peuhkurinen, J., Suvanto, A., Pesonen, A., Hyyppä, J., 2007. Experiences and Possibilities of ALS Based Forest Inventory in Finland. *Proceedings of the ISPRS Workshop Laser Scanning 2007 and SilviLaser 2007,*Volume XXXVI, PART 3/W52, 12 – 14th September 2007, Espoo, pp. 270 – 279.

Reitberger, J., Krzystek, P. and Stilla, U., 2007, Combined tree segmentation and stem detection using full waveform LiDAR data. *Proceedings of the ISPRS Workshop Laser Scanning 2007 and SilviLaser 2007*,Volume XXXVI, PART 3/W52, 12 – 14th September 2007, Espoo, 332 – 337.

Reitberger, J., Krzystek, P. and Stilla, U., 2008a, Analysis of full waveform LiDAR data for the classification of deciduous and coniferous trees. *International Journal of Remote Sensing*, Vol. 29, No. 5, 1407 – 1431.

Reitberger, J., Krzystek, P., Stilla, U., 2008b, 3D segmentation and classification of single trees with full waveform LiDAR data. *Proceedings of SilviLaser 2008, 8th international conference on LiDAR applications in forest assessment and inventory*, Edinburgh, UK, 216 – 226.

Reitberger, J., Schnörr, C., Krzystek, P., and Stilla, U., 2009, 3D segmentation of single trees exploiting full waveform LiDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 64, No. 6, 561 – 574.

Shi, J., Malik, J., 2000, Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, 888 – 905.

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

VINCENT, L. and SOILLE, P., 1991, Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, Vol. 13, No. 6, 583 – 598.

Wang, Y., Weinacker, H., Koch, B. and Sterenczak, K., 2008, LIDAR Point Cloud Based Fully Automatic 3D Single Tree Modelling in Forest and Evaluations of The Procedure. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 37 (Part B6b), 45 – 51.

Single tree detection using full waveform laser scanner data

Sandeep Gupta†, Holger Weinacker† and Barbara Koch†*

sandeep.gupta@felis.uni-freiburg.de

†Dept. of Remote Sensing and Landscape Information Systems (FeLis), Faculty of Forestry, Albert-Ludwigs University, Tennenbacher str. 4, 79106 Freiburg, i.Br., Germany

Abstract

In the presented work, single tree detection and shape reconstruction was carried out using full waveform airborne laser scanning (ALS) data. The medium density data was acquired from the mixed forest area in the Rhine valley near Karlsruhe, Germany. Study was conducted in the rectangular plots, each of size 2500 m^2, containing 12 tree species of varying age and height. A modified partitioning algorithm was applied to extract the clusters of single tree. Extracted tree clusters were reconstructed to represent their approximate shape in the form of 3-D convex polytope. A standard validation procedure was implemented for accuracy assessment of the extracted individual clusters of each tree species with reference data. An overall 52.7% of tree species were correctly detected in such a complex forest condition. The algorithm worked comparatively well for extracting the evergreen conifers than the deciduous trees despite their lower distribution in the test plots. The presence of tree species of different age groups in different layers with high canopy density was the main factor that caused an average performance of the algorithm in accurately detecting the single tree. The outcome and problem analysis shows the scope of further refinement of the algorithm for the better estimate of single tree in such a forest condition.

1. Introduction

Airborne laser scanning, one of the active optical remote sensing technologies, provides data that make it possible to detect and isolate the individual trees (Hyyppä *et al.* 2001, Persson *et al.* 2002). Measurements of tree attributes are of critical importance in the derivation of estimates of standing timber volume, and biomass (Lim *et al.* 2003). The availability of resources and rapid development in technologies has paved the path for harnessing the potential of airborne laser scanner (ALS) data at its fullest for the tree-level inventory. Several studies has been carried out in the past on the application of airborne LIght Detection And Ranging (LIDAR) data for vegetation related information retrieval using different methods (Nilsson 1996, Hyyppä and Inkinen 1999, Persson *et al.* 2002, Hyyppä *et al.* 2006, Wang *et al.* 2008, Vauhkonen *et al.* 2009). The forest information from LIDAR data are being retrieved with two main approaches – normalized digital surface model (nDSM)-based and the individual tree detection-based (Maltamo *et al.* 2006). The main drawback of nDSM based approach is that the trees and young regenerations in the intermediate and lower forest layers are seldom detected (Reitberger *et al.* 2008a). The single tree crown contour delineation at different height quintiles with a hierarchical morphological approach has been presented by Wang *et al.* (2008). A three-dimensional (3-D) segmentation technique based on the normalized cut segmentation method using variable full waveform LIDAR point density showed an improvement in the tree detection rate (Reitberger *et al.* 2008a). The false detection also increases with the increase in the detection rate showing the substantial decline in the reliability of the approach (Pitkänen *et al.* 2004, Reitberger *et al.* 2008a). With the use of the local maxima, several studies have been conducted in the past for the tree top detection, majority in the image domain, with some degree of adaptations (Persson *et al.* 2002, Popescu *et al.* 2002, Pitkänen *et al.* 2004, Koch *et al.* 2006, Tiede *et al.* 2008).

Among the vector-based methods, clustering is one of the promising approaches for single tree detection. Several mechanisms exist for clustering, among which, the *k*-means is the most popular iterative partitioning based approach. In this, the ALS data of the study area is being partitioned into a group of clusters using a distance criterion (Jain *et al.* 1999). The research work on the single tree detection with ALS data using the different clustering mechanisms has been reported in the past (Morsdorf *et al.* 2003, Morsdorf *et al.* 2004, Cici *et al.* 2008, Doo-Ahn *et al.* 2008, Reitberger *et al.* 2008b). The smoothened DSM derived local maxima as seed points in the *k*-means method with the first and last pulse high density ALS data to delineate the

single tree in the Swiss National Park was used (Morsdorf *et al.* 2003). In contrast to the approach used in the presented work, instead of scaling-down the z-coordinates, Morsdorf *et al.* (2003) scaled it up by a factor of three to accommodate the aspect ratio of pine tree crowns, which ranged from 3-6. It is noted that the *k*-means works well when a data set has "compact" or "isolated" clusters (Mao and Jain 1996). Therefore, it is advantageous to scale down the height value of the normalized raw points as well as seed points in the z dimension. This helps in minimizing the squared error function, which is the ultimate objective of the *k*-means method (Gupta *et al.* 2010). Additionally, the present study differs from Morsdorf *et al.* (2003) in two ways: first, the smoothing was avoided on the nDSM and second, the unwanted local maxima points were deleted using a search algorithm based threshold distance. The shape of the detected point clouds of clusters of tree crowns can be geometrically reconstructed and shown in different forms such as convex hull (Morsdorf *et al.* 2003, Ko *et al.* 2009).

The objective of the presented work is to examine the potential of supervised *k*-means approach to detect single trees using the full waveform ALS data in the selected plots of temperate forest with multi-tier mixed tree species distribution.

2. Materials and Methodology

2.1. Study area and field data characteristics

General characteristics of the Study area

Investigations were carried out in the forest area in the Rhine valley near Karlsruhe, Germany. The Gauss Krüger coordinates (in meters) for the upper left corner are 3456375, 5433820 and the lower right corner 3458025, 5432980. The test area contains seven rectangular field plots, each of size 50m*50m. The fields are characterised by multi-tier deciduous and coniferous tree species with dense canopy. All the fields, except field 3, are dominated by deciduous tree species (table 1).

Table 1. Tree type Distribution.

Field Id	Conifer (%)	Deciduous (%)
1	9.4	90.6
2	4.1	95.9
3	61.1	38.9
4	22.7	77.3
5	4.0	96.0
6	25.5	74.5
7	26.7	73.3

The tree type, tree species name, their common name with species symbols used in the presented work is shown in table 2.

Table 2. Tree Species distribution.

Tree species	Common name	Symbol	Tree type	%
Pinus sylvestris	Scots pine	Kie	Conifer	13.9
Picea abies	Norway spruce	Fi	-do	2.2
Pseudotsuga menziesii	Douglas fir	Dgl	-do	1.6
Carpinus betulus	Hornbeam	HBu	Deciduous	9.5
Prunus avium	Cherry	Kir	-do	23.7
Quercus petraea	Oak	Ei	-do	24.3
Quercus rubra	Red Oak	REi	-do	0.9
Fagus sylvatica	European Beech	Bu	-do	21.1
Robinia pseudoacacia	Black Locust	Rob	-do	0.6
Tilia cordata	Linden or Basswood	Li	-do	0.9
Betula pendula	Silver Birch	Bi	-do	0.3
Acer pseudoplatanus	Sycamore Maple	Bah	-do	0.9

Field measurements

The state forest administration (FVA) of the Federal State of Baden-Württemberg, Germany, provided the Forest inventory data from summer 2006. All trees above 7cm diameter at breast height (DBH) were measured. A Vertex® instrument was used to measure the two top heights of the main tree and one top height of the dominated tree. The arithmetic mean of the height measurements of all the dominating trees in each field was calculated as an average top height for the respective plot (Straub *et al.* 2009). The heights of the remaining trees were estimated using the stand height curves with the DBH as an input variable (Korn-Allan *et al.* 2004). Additional fieldwork was conducted for the plot establishment and coordinates measurement.

2.2. LIDAR data and Orthophoto characteristics

The full waveform laser scanner data was acquired in August 2007 by TopoSys GmbH using the "Harrier 56/G4" Riegl LMS-Q56O scanner mounted on a helicopter and flown at a height of 450m above ground level. To achieve a high point density (16 points m^{-2}), the study area was flown twice during day and night with more than 50% of overlapping. The DTM, DSM and nDSM were calculated in TreesVis (Weinacker *et al.* 2004) environment. The raw full waveform LIDAR points were normalized using the DTM to ensure the absolute height of the object and to eliminate the influence of the terrain. The normalized LIDAR data was used in the clustering. Optical data in four spectral channels were collected by TopoSys GmbH with the aid of an RGB/NIR line scanner in July 2008, at 700m above ground level at a rate of 682 pixels per line. The important flight and technical parameters of the laser scanner and RGB/NIR line scanner are described by Straub *et al.* (2009).

2.3. Single tree detection

All the data processing work was carried out in a 100m by 100m area covering each of the seven fields. Subsequently, for the accuracy assessment, the data were clipped to the plot size. This was done to avoid the omission of some of the reference tree top points that were intersected at the edge and/or were present outside the rectangular field.

Local maxima detection and processing

Local maxima points were extracted from the nDSM above 5m height having a gray value larger than the gray value of all its 8 neighbors. The nDSM was not smoothened in this study. In the next step, the local maxima points that were too close to each other were excluded based on a threshold distance (d_{thres}) with a search algorithm. The filtered local maxima as external seed points were finally used in the *k*-means algorithm. The d_{thres}, varied depending on the forest conditions. The value of d_{thres} for younger trees with single and narrow crown at the top or conifers was found between 2-4m. While, the d_{thres} for trees with relatively older ones with wider crown with more intermittent peaks at the top or broadleaved trees was found between 4-6m.

Supervised k-means

Jain *et al.* (1999) have described the simple *k*-means algorithm. The simple *k*-means algorithm was supervised in this study to use the nDSM-derived local maxima as external seed points for process initialization. This reduces the time and machine cost of using trial and error approach with several repetitions and cross-checking each time visually with reference data in order to select an appropriate *k* clusters for each plot. The output was improved by reducing the height value of the data points and that of the external seed points by an empirically found value (i.e. half) before the initialization. The supervised *k*-means ran over the field data for generation of single tree clusters.

2.4. 3-D Reconstruction of Individual tree clusters

The shape of the tree crown clusters, commonly present in the convex form, was reconstructed using the QHull approach (Barber et al. 1996). QHull is a general dimension code for computing convex hulls using

Quickhull algorithm (Berg et al. 1997). Those rare clusters whose polytopes could not be formed were discarded.

2.5. Validation procedure

The detected tree tops of single tree clusters in each plot have been validated with the corresponding field measured tree tops. Different approaches to validate the detected tree tops has been presented (e.g. Persson et al. 2002, Morsdorf et al. 2003, Heurich et al. 2004, Koch et al. 2006, Wang et al. 2008, Koch et al. 2009, Reitberger et al. 2009). In this work, the five validation classes for accuracy assessment have been adopted.

- Exact - only one detected tree top with respect to the nearby reference tree top. The 3-D Euclidean distance (ED) between the reference tree top and the detected tree top point (3-D ED) is below 3m.
- Nearly Exact - one detected tree top nearly at the same height level nearby a reference tree top and 3-D ED is 3-5m.
- Split - more than one neighboured detected treetops up to the 3-D ED of 5m.
- Missing - no detected tree top nearby the reference tree top up to the 3-D ED of 5m. It also includes the reference points for which there is no detected tree tops within the radius of 3m created from each reference point.
- Extra - detected tree tops within the field boundary for which there is no reference tree top point up to the 3-D ED of 5m. In addition, it includes those detected points within the field boundary for which there is no reference point within the radius of 3m created from each reference point.

The ‘Extra’ and ‘Missing’ point is considered as ‘commission error’ and ‘omission error’, respectively. When expressing the accuracy in percent, two types of accuracy classes were used (Congalton 1991), the “user’s accuracy” and the “producer’s accuracy”. The clusters obtained were evaluated against the field inventory data. The validation performed was based on the spatial relationship between the field measured and the detected tree tops with reference to the above mentioned five classes.

3. Results and Discussion

After running of the supervised *k*-means algorithm, the 3-D cluster points of single trees were extracted above 5m height in all the field plots. This was done to avoid the effect of low ground vegetation and other smaller objects during the clustering process. From the 3-D points of single tree clusters, tree top points were extracted. The tree height is calculated as the maximum height of the LIDAR points belonging to a specific cluster. After accuracy assessment, the result has been presented in table 3.

Table 3. Single tree accuracy.

Field ID	Producer accuracy (%)	User accuracy (%)	False Detection (%)
1	65.6	42.0	58
2	51.0	35.7	64.3
3	61.1	50.0	50
4	63.6	46.7	53.3
5	39.6	41.2	58.8
6	72.3	50.7	49.3
7	36.7	22.9	77.1

It is evident from table 3 that the producer’s and user’s accuracies are varying greatly in the deciduous dominated field plots. The high accuracy in field 6 is attributed to the presence of fewer tree species nearly at the same height level where the algorithm worked well. The lower accuracy in the field 5 is mainly due to the mixed distribution multi-layered Oak and European beech, while in the field 7 it is due to the presence of

mixed tree species of dense canopy and varying age groups. In both the cases, the algorithm did not worked well. The false detection is high in fields 2 and 7. It was assumed that the presence of higher proportions of cherry trees of low height and small crown diameter in both the plots might have caused the increase in the false detection. It is obvious that the performance of the algorithm was found average. The multi-tier, mixed tree species with varying age groups and dense canopy closure might have played a vital role in lower accuracies and increase in false detection.

Table 4 shows the species-wise distribution of the reference tree top and detected tree tops in the 'Exact' and 'Nearly Exact' validation classes with their height range in each field. The field columns were left blank when there are no tree species detected in the 'Exact' and 'Nearly Exact' validation classes. The height range shown in table 4 of the detected tree tops of the respective tree species is of the trees falling under the two validation classes only.

Table 4. Height distribution of the reference and detected tree tops in the 'Exact' and 'Nearly Exact' validation classes in each field.

Field Id	Tree species	Rct	RHrange [m]	DHrange [m]	E	E (%)	NE	NE (%)	E+NE	E+NE (%)
1	Hornbeam	11	08-21	9.2-19.9	7	63.6	2	18.2	9	81.8
	Scots pine	3	20-24	22.1-24.3	3	100	0	-	3	100
	Cherry	18	08-10	7.7-12.3	7	38.9	2	11.1	9	50.0
2	Scots pine	2	22-27	23.5-24.3	0	-	2	100	2	100
	Cherry	45	08-16	10.2-18.9	13	28.9	8	17.8	21	46.7
	Oak	2	23-25	23.5-26.5	2	100	0	-	2	100
3	Scots pine	16	21-31	17.3-28.7	10	62.5	3	18.8	13	81.3
	Red Oak	3	26	24.9	1	33.3	0	-	1	33.3
	European Beech	9	15-18	17.2-20.2	2	22.2	1	11.1	3	33.3
	Douglas fir	5	16-23	17.2-26.0	2	40.0	1	20.0	3	60.0
	Black Locust	2	20-23	21.8-22.0	2	100	0	-	2	100
	Norway spruce	1	8	-	0	-	0	-	0	-
4	Hornbeam	15	13-19	14.1-18.9	7	46.7	2	13.3	9	60.0
	Scots pine	5	23-35	26.1 - 32.8	3	60.0	1	20.0	4	80.0
	European Beech	2	21	21	1	50.0	0	-	1	50.0
5	Oak	74	11-19	12.1-18.3	9	12.2	24	32.4	33	44.6
	Scots pine	4	17-22	16.7-22.5	0	-	4	100	4	100
	European Beech	19	13	14.8	0	-	1	5.3	1	5.3
	Linden	3	14	15.9	0	-	1	33.3	1	33.3
	Silver Birch	1	17	17.1	1	100	0	-	1	100
6	European Beech	34	13-23	15.1-26.0	10	29.4	14	41.2	24	70.6
	Scots pine	12	24-30	20.3-29.3	3	25.0	7	58.3	10	83.3
	Sycamore Maple	1	20	-	0	-	0	-	0	-
7	European Beech	3	26-28	26.3-28.5	2	66.7	1	33.3	3	100
	Scots pine	2	29-32	30.5-31.1	0	-	2	100	2	100
	Norway spruce	6	14-32	13.4-30.5	0	-	3	50.0	3	50.0
	Sycamore Maple	2	9	10.9	0	-	1	50.0	1	50.0
	Cherry	12	5-9	-	0	-	0	-	0	-
	Hornbeam	4	20-24.2	20.8-24.2	0	0.0	2	50.0	2	50.0
	Oak	1	26	-	0	-	0	-	0	-

where, Rct = reference tree top count, RHrange [m] = height range of the reference tree in meters, DHrange [m] = height range of the detected tree in meters, E = 'Exact' detected tree top count, E (%) = 'Exact' detected tree tops in percent, NE = 'Nearly Exact' detected tree top count, NE (%) = 'Nearly Exact' detected tree tops in percent, E+NE = 'Exact' + 'Nearly Exact' detected tree tops together and E+NE (%) = 'Exact' + 'Nearly Exact' detected tree tops together in percentage.

One Norway spruce of height 8m in field 3, one sycamore maple of height 20m in field 6 and, twelve cherry tree of height range 5-9m and one oak tree of height 26m in field 7 were not *et al* detected. The cherry trees present in the fields 1 and 2 constitute roughly 56% and 92% part of all the tree species in their respective fields. Nearly 49% of the cherry trees ranging from 8-16m in height in the fields 1 and 2 were detected. The d_{thres} for filtering out the external seed points were kept relatively low in the fields 1 and 2, 4m and 3.5m, respectively, as compare to 4.5m in the field 7. In the field 7, cherry trees were 40% of the total tree composition and were of low height (5-9m). These were lying below the with dense canopy covered upper-tier trees. It is likely that the presence of dense canopy cover may have resulted in insufficient backscattered laser signals from the under storey trees. Another possibility is the poor detection of local maxima from the interpolated nDSM. The mixed tree species composition with dense canopy coverage, inappropriate d_{thres}, and low extraction of external seed point could be the possible reasons for null estimation of cherry trees in the field 7.

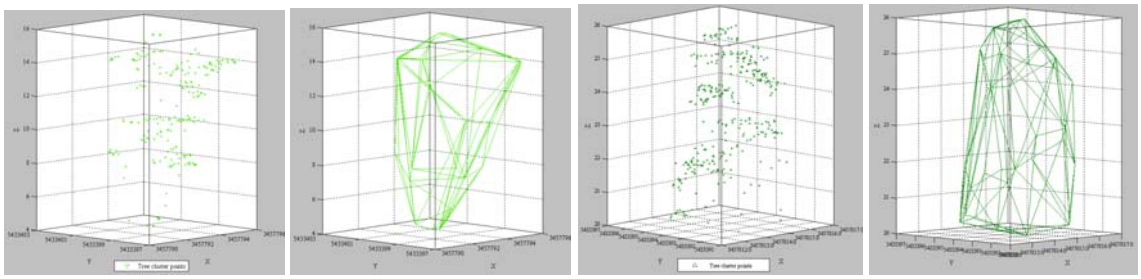
In the field 5, oak is a dominant (73%) tree species with comparatively low maturity level. Roughly, 45% of the oak trees were detected (table 4) after running the supervised *k*-means algorithm (d_{thres} = 2m). In the field 5, the oak trees were in different age groups with varying height (11-19m). The trees detected were mainly mature. It was noticed from the reference data that nearly 47% of the oak tree in field 5 were 11-14m tall and rest were between 14-19m in height. The oak trees detected below the 14m and above 14m height were around 12% and 32%, respectively (table 4). Thus, higher is the detection probability of a tree when it is in the upper canopy layer. By above examples, it is obvious that the tree age and height, tree species, layer of the trees and canopy density beside the terrain plays an important role in single tree remote sensing. It is noticeable from table 4 that Scots pines, despite their low presence (13.9%), were found in all the 7 fields and more than 80% of the Scots pines were detected by applying the algorithm. This was probably due to their presence in the upper canopy layer.

Most of the Hornbeam trees present in the three fields (1, 4 and 7) were attained maturity. Hornbeam is most dominant tree species in field 4 with 60% detection rate and is second most dominant in the field 1 with nearly 82% detection rate. Field 7 has the 50% of detection rate for Hornbeam. Around two-third of the hornbeam tree was successfully detected in all the three fields. The European beech is present in five fields (3, 4, 5, 6 and 7). Most of the European beeches are at mature stage. Field 6 is dominated by European beech with 72% distribution over the area. The detection rate in field 6 is 70.6%. The most poor detection rate of European beech (5.3%) was found in field 5 where nearly 84.2% trees were under 'Missing' category and 10.5% trees were under 'Split' category. The reason could be attributed to the fact that almost three-forth of the European beech are younger and lying beneath the layer of oak and Scots pine. The dense canopy and inter-locking branches of some of the oak and Scots pine trees might have contributed to very low detection rate of European beech in field 5.

An overall 52.7% of tree tops together for 'Exact' and 'Nearly Exact' validation classes were found by applying the supervised *k*-means. During the research investigation, it was found that the d_{thres} setting becomes more difficult with increasing forest complexity. For example, the plot containing mainly trees with wider canopies requires higher d_{thres} because local maxima from smaller peaks will most likely represent only branches, hence needs to be eliminated. Whereas, local maxima from a peak in a plot containing younger trees will most likely to be a treetop, hence requires comparatively smaller d_{thres}.

The triangulated facets of the convex hulls were computed from the vertices using QHull algorithm. In all the cases, the facets were successfully formed even when there was a gap in clusters containing LIDAR points obtained after running the algorithm. The shape of the 3-D convex polytope formed from each tree clusters represents the approximate shape of the respective individual tree. Two examples of sample clusters of

European beech and Scots pine (Figures 1a and 1c) and their respective 3-D convex polytopes (Figures 1b and 1d) have been presented. The convex hull can be further used in the area and volume estimation.



(a) (b) (c) (d)

Figure 1. The sub-figures showing the 3-D ALS clusters of a) European beech and c) Scots pine, and the corresponding b) and d) 3-D convex polytopes.

4. Conclusions

The clusters of single trees above 5m height were extracted by applying the supervised *k*-means algorithm on the full waveform ALS data for 7 test plots in the forest area near Karlsruhe, Germany. Traditional *k*-means generates arbitrarily bad grouping of objects due to random seed selection procedure and repeated run of the algorithm after cluster analysis to meet the fitness criteria (Jain *et al.* 1999). The supervised *k*-means yields comparatively fair results by partitioning the LIDAR data after seeding is done externally and once the height value of the LIDAR points are scaled down to half before initialization of the process so as to bring those LIDAR points closer height-wise (Gupta *et al.* 2010). The terrestrial data of the field plots with tree position and other related attributes helped in validating the result at species level (tables 4 and 5). Majority of the trees species are deciduous (82.3%) with Oak, cherry, hornbeam and European Beech. Among the evergreen conifers, 78.6% trees were Scots pine.

The validation procedure was devised and five validation classes were defined for accuracy assessment. Out of the five validation classes, two were used for goodness criteria in the Euclidean 3-D space. The mixture of multi-storey tree species of varying height quintile (table 4) with dense canopy cover poses a big challenge in detecting the tree species by the algorithm in the presented study. In all the 7 field plots, more than 80% of the Scots pines were detected by applying the algorithm.

The average producer's and user's accuracies among all the field plots are 55.7% and 41.3%, respectively (table 3). The variations in the producer's and user's accuracies are roughly between 40-65% and 36-42%, respectively, for the deciduous dominated fields (1, 2 and 5). In case of mixed species fields (3, 4, 6 and 7), the producer's and user's accuracies are varying nearly between 37-72% and 23-51%, respectively (table 3). The variation in the height range and average height for 'Exact' validation class was 0-2.5m and 0-1.4m, respectively, and that for the 'Near Exact' validation class was 0.3-4.3m and 0.3-4.2m, respectively.

The result showed that the algorithm performance was better for the upper tier trees which are relatively mature and older as compared to the relatively younger tree species lying under-storey. In case of under-storied tree species, sufficient number of external seed points could not be generated using local maxima method, mainly because of interference from the first-tier trees and high tree density. The result in the tested field plots could not be improved even by varying the d_{thres}. However, it was assumed that more robust parameter setting and smoothing of nDSM before local maxima calculation could have reduced the false detection.

In general, the value of d_{thres} for younger trees with narrow crown at the top is lower as compare to the relatively older trees with wider crown and more intermittent peaks at the top. It was found that the scaling factor has substantial impact in the data partitioning. The relatively better clustering was due to the empirically found scaling factor for the z-coordinates, which was kept as half for all the plots.

The significance of pre-knowledge about the forest conditions from the field-inventoried data cannot be ruled out in setting the variables in the algorithm. The quantity and quality of the result mainly depend on the flight parameters, forest condition, LIDAR point density, d_{thres} and external seed points. The point density of extracted cluster has a direct effect on the clustering and shape of 3-D convex polytope.

In the future, the single tree detection and modeling task will focus on the utilization of ALS data and other associated parameters viz. intensity and pulse width in different forest conditions based on the specific rule sets. Separate study cannot be ruled out by using leaf-on and leaf-off datasets for better discrimination of tree species, particularly in complex forest sites.

Acknowledgements

The author acknowledges the Landesgraduiertenförderungsgesetz (LGFG) Commission, Baden-Württemberg, Germany for providing financial support during the research work within the framework of doctoral scholarship program. The author would like to thank the Forest Research Institute of Baden-Württemberg (FVA) for providing the reference data for the Karlsruhe study area. The author also acknowledges his gratitude to his fellow colleagues Christoph and Johannes who have further conducted the field work in Karlsruhe test sites for getting more forest parameters.

References

BARBER, C.B., DOBKIN, D.P. and HUHDANPAA, H.T., 1996, The Quickhull algorithm for convex hulls. ACM Trans. on Mathematical Software, 22(4), pp. 469-483.

BERG, M.D., KREVELD, M.V., OVERMARS, M. and SCHWARZKOPF, O., 1997, Computational Geometry - Algorithms and Applications, ISBN 3-540-61270-X, pp. 233-244 (Heidelberg: Springer-Verlag).

CICI, A., KEVIN, T., NICHOLAS, J. T., SARAH, S. and JÖRG, K., 2008, Extraction of vegetation for topographic mapping from full-waveform airborne laser scanning data. In Proceedings of SilviLaser 2008, 17-19 September 2008, Edinburgh, U.K., pp. 343-353.

CONGALTON, R. G., 1991, Considerations and techniques for assessing the accuracy of remotely sensed data. Remote Sensing of Environment, 37, pp. 35-46.

DOO-AHN, K., WOO-KYUN, L. and HYUN-KOOK, C., 2008, Estimation of effective plant area index using LiDAR data in forest of South Korea. In Proceedings of SilviLaser 2008, 17-19 September 2008, Edinburgh, U.K., pp. 237-246.

Gupta, S., WEINACKER, H. and KOCH, B., 2010, Comparative analysis of clustering-based approaches for 3-d single tree detection using airborne fullwave lidar data. Remote Sensing, 2(4), pp. 968-989.

HEURICH, M., PERSSON, Å., HOLMGREN, J., and KENNEL, E., 2004, Detecting and measuring individual trees with laser scanning in mixed mountain forest of central europe using an algorithm developed for swedish boreal forest conditions. In International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36 (part 8/W2), Freiburg, Germany, pp. 307-312.

HYYPPÄ, J. and INKINEN, M., 1999, Detecting and estimating attributes for single trees using laser scanner. The Photogrammetric Journal of Finland, 16, pp. 27-42.

HYYPPÄ, J., KELLE, O., LEHIKOINEN, M. and INKINEN, M., 2001, A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. IEEE Transactions on Geosceince and Remote Sensing, 39(5), pp. 969-975.

HYYPPÄ, J., Yu, X., HYYPPÄ, H. and MALTAMO, M., 2006, Methods of airborne laser scanning for forest information extraction. In EARSeL-Proceedings of International Workshop on 3D Remote Sensing in Forestry, 14-15 February 2006, Vienna, Austria, pp. 63-78.

JAIN, A.K., MURTY, M.N. and FLYNN, P.J., 1999, Data Clustering: A Review. ACM Computing Surveys, 31(3), pp. 264-323.

KO, C., SOHN G. and REMMEL, T. K., 2009, Classification for Deciduous and Coniferous Trees Using Airborne LiDAR and Internal Structure Reconstructions. In Proceedings of SilviLaser 2009, 14-16 October 2009, Texas, USA, pp. 36-45.

KOCH, B., HEYDER, U. and WEINACKER, H., 2006, Detection of individual tree crowns in airborne lidar data. Photogrammetric Engineering & Remote Sensing, 72(4), pp. 357-363.

KOCH, B., HEINZEL, J., GUPTA, S., WANG, Y. and WEINACKER, H., 2009, Extraction of forest parameter from laser data. In Proceeding of the DGPF 2009 Conference, March 24-26 2009, Jena, Germany, pp. 345-354.

KORN-ALLAN, E., V.D. GOLTZ, H., BLUST, M. and NOTHDURPT, A., 2004, Verfahrenshandbuch Betrieb sinventur. - Version 1.1., Landesforstverwaltung Baden-Württemberg.

LIM, K., TREITZ, P., WULDER, M. and ST-ONGE, B., 2003, LiDAR remote sensing of forest structure. Progress in Physical Geography, 27, pp. 88–106.

MALTAMO, M., EERIKÄINEN, K., PACKALÉN, P. and HYYPPÄ, J., 2006, Estimation of stem volume using laser scanning-based canopy height metrics. Forestry, 79(2), pp. 217-229.

MAO, J. and JAIN, A.K., 1996, A self-organizing network for hyperellipsoidal clustering (HEC). IEEE Trans. Neural Netw., 7, pp. 16–29.

MORSDORF, F., MEIER, E., ALLGÖWER, B. and NÜESCH, D., 2003, Clustering in airborne laser scanning raw data for segmentation of single trees. In International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 34 (part 3/W13), Dresden, Germany, pp. 27-33.

MORSDORF, F., MEIER, E., KÖTZ, B. and ITTEN, K.I., 2004, Lidar based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. Remote Sensing of Environment, 92, pp. 353-362.

NILSSON, M., 1996, Estimation of tree heights and stand volume using an airborne lidar system. Remote Sensing of Environment, 56(1), pp. 1-7.

PERSSON, Å., HOLMGREN, J. and SÖDERMAN, U., 2002, Detecting and measuring individual trees using an airborne laser scanner. Photogrammetric Engineering & Remote Sensing, 68, pp. 925−932.

PITKÄNEN, J., MALTAMO, M., HYYPPÄ, J. and WEI YU, X., 2004, Adaptive methods for individual tree detection on airborne laser based canopy height model. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36 (pasrt 8/W2), Freiburg, Germany, pp. 187-191.

POPESCU, S.C., WYNNE, R.H. and NELSON, R.F., 2002, Estimating plot-level tree heights with lidar: local filtering with a canopy height based variable window size. Computers and Electronics in Agriculture, 37, pp. 71-95.

REITBERGER, J., KRZYSTEK, P. and STILLA, U., 2008a, 3D segmentation and classification of single trees with full waveform LIDAR data. In Proceedings of SilviLaser 2008, 17-19 September 2008, Edinburgh, U.K., pp. 216-226.

REITBERGER, J., KRZYSTEK, P. and STILLA, U., 2008b, Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees. International Journal of Remote Sensing, 29(5), pp. 1407-1431.

REITBERGER, J., SCHNÖRR, Cl., KRZYSTEK, P. and STILLA, U., 2009, 3D segmentation of single trees exploiting full waveform LIDAR data. ISPRS Journal of Photogrammetry and Remote Sensing, 64(6), pp. 561-574.

STRAUB, C., DEED, M., WEINACKER, H. and KOCH, B., 2009, Using Airborne Laser Scanner Data and CIR Orthophotos to Estimate the Stem Volume of Forest Stands. Photogrammetrie - Fernerkundung - Geoinformation, 2009(3), pp. 277-287.

TIEDE, D., LANG, S. and HOFFMANN, C., 2008, Domain-specific class modelling for one-level representation of single trees. In Object-Based Image Analysis - Spatial concepts for knowledge-driven remote sensing applications, T. Blaschke, S. Lang and G. Hay (Eds.), pp. 133-151 (New York: Springer).

VAUHKONEN, J., TOKOLA, T., PACKALÉN, P. and MALTAMO, M. 2009. Identification of Scandinavian commercial species of individual trees from airborne laser scanning data using alpha shape metrics. Forest Science 55(1), pp. 37-47.

WANG, Y., WEINACKER, H., KOCH, B. and STEREŃCZAK, K., 2008, Lidar point cloud based fully automatic 3d single tree modelling in forest and evaluations of the procedure. In International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 37 (part B6b), Beijing, China, pp. 45–52.

WEINACKER, H., KOCH, B., HEYDER, U. and WEIANCKER, R., 2004, Development of filtering, segmentation and modelling modules for lidar and multispectral data as a fundament of an automatic forest inventory system. In International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36 (part 8/W2), Freiburg, Germany, pp. 50-55.

A comparison of support vector and linear classification of tree species

Johannes Heinzel§, Olaf Ronneberger‡ and Barbara Koch§*

johannes.heinzel@felis.uni-freiburg.de

§University of Freiburg, Department of Remote Sensing and Landscape Information Systems, 79106 Freiburg, Germany

‡University of Freiburg, Department of Computer Science and Centre for Biological Signalling Studies (BIOSS), 79110 Freiburg, Germany

Abstract

Data from full-waveform airborne LiDAR becomes more and more available. At the same time high dimensional feature spaces are increasingly used for remote sensing related classifications. To handle those data sets advanced classification techniques are often regarded as producing more accurate results compared to less complex classification methods. Within this paper an empirical comparison of classification results from linear discriminant analysis (LDA) and support vector machines (SVM) is carried out. For this we used an innovative and comprehensive set of reflection specific full-waveform LiDAR features. Altogether 231 features were available, which are compositions of the geometric reflection properties, their statistical distribution, the position within the laser beam as well as the spatial position.

The LDA was conducted with a stepwise inclusion of variables to achieve a ranking of their importance. SVMs are tested with different configurations concerning the multiclass classification method and the kernel. The parameters of the SVM are optimized using a grid search algorithm. A generally approved method to achieve a nonlinear variable ranking from the SVM is not available until now. Therefore, both classifiers are additionally executed with the most important variables from the stepwise LDA and are compared to a set with all variables. Verification was done by a 24-fold cross validation with leaving out complete independent sample plots.

Using the complete feature set SVM outperforms LDA for the four main species pine, spruce, oak and beech (79.22%) and for the deciduous-coniferous (94.43%) classification. When the variables are restricted to only the top ranked variables the results for the LDA increase and come very close to the SVM accuracies. SVM accuracies remain constant using LDA selected features. As a surprising result, LDA with an optimized feature set even outperforms SVM in the case of six species(61.81%). This indicates a need to develop specific nonlinear variable ranking mechanisms for SVMs and to test their influence on SVM classification results.

1. Introduction

The automated classification of tree species from airborne surveyed data is a major scientific and practical aim of forestry related remote sensing applications. In general there are two basic principles by which the solution of the classification can be influenced. This is first the selection of the features and second, the choice of the classifier.

Concerning the first option early attempts used mainly colour infrared (CIR) images (Pinz and Bischof 1990, Meyer *et al.* 1996) whose capability were often more or less limited to the distinction of coniferous and deciduous trees when working with a precision on a single tree scale (Holmgren and Persson 2004). Most recent studies try to exploit the potential of light detection and ranging (LiDAR) data while either using the spatial distribution pattern of the point cloud (Brandtberg *et al.* 2003, Brandtberg 2007, Tokola *et al.* 2008) or, still less common, by inspecting the geometric properties of the single reflections (Litkey *et al.* 2007, Hollaus *et al.* 2009, Heinzel and Koch 2010).

The second option, which concerns the classifier, has also experienced major technical improvements. When LiDAR data is the last decades' innovation on the feature and data side, then support vector machines (SVMs) seem to develop as one of the most innovative methods on the classifiers' side. Their introduction to remote sensing is still in its infancy and most of the studies are from a period within the last five years (Pal and Mather 2005, Waske and Benediktsson 2007, Ørka *et al.* 2009b). While some authors indicate their advantage over other classifiers in remote sensing applications, there are still numerous studies relying on more basic but well proven methods like linear discriminant analysis (LDA) (Lucas *et al.* 2008, Kim *et al.* 2009, Ørka *et al.* 2009a).

This study presents a direct empirical comparison of the results from LDA and SVM classifiers using the same dataset for tree species classification. Furthermore, it considers different configurations of the SVM, concerning the kernel and the multiclass method. It also explains a complete SVM classification procedure which comprises the optimization of parameters and the *t*-fold cross validation. The dataset consists of features which are compositions of the geometric properties of LiDAR reflections, different values for their statistical distribution within a grid cell and their location within the laser beam as well as in space. Altogether, 231 compositions represent the feature space for both classifiers.

In comparison to several other studies which aim to distinguish coniferous from deciduous or a maximum of up to three species (Holmgren *et al.* 2008, Ørka *et al.* 2009a) we present results for different numbers of classes, which are referred to as classification depths. The discrimination of up to six species was tested under the conditions of a mixed temperate forest. Furthermore, we describe a grid based classification where no information about the delineation of the tree crown is known. The classification of automatically delineated crowns, e.g. Holmgren and Persson (2004) and Reitberger *et al.* (2008), bears the risk that failures in single tree detection increase the failure in species classification.

Data

Study area

The study area is located in south-western Germany close to the city of Karlsruhe as shown in Figure 1. It comprises a rectangular area of about 10 km² and has an even terrain. The forest stands are characterized by a high diversity concerning the deciduous-coniferous composition, the tree age, the layering and the density of the stands. The main species in a typical temperate forest of this type are pine (*Pinus sylvestris*), spruce (*Picea abies*), oak (*Quercus petraea*) and beech (*Fagus sylvatica*). Additionally, there are a considerable number of the secondary species hornbeam (*Carpinus betulus*) and cherry (*Prunus avium*) within our study area.

LiDAR data

LiDAR data was captured during a flight campaign in the summer of 2007 with a Harrier 56 system by TopoSys® GmbH. The recorded full-waveforms were preprocessed with the software RiAnalyze 560© from Riegl LMS GmbH to extract singular reflections. In addition to the pure three dimensional coordinates we extorted the reflection specific amplitude, the width and the intensity of the signal. The intensity is defined as the total power received over time by the reflected signal. Furthermore, we got the position of each reflection and the total number of reflections within the corresponding beam. Due to a special flight planning we achieve a point density of at least 16 points/m².

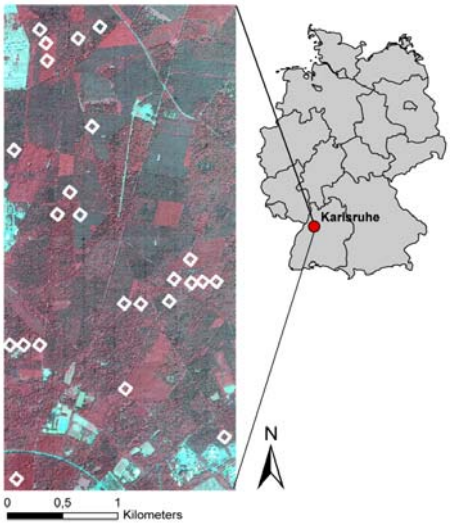


Figure 5. Location and overview of the study area. The image shows a CIR aerial photograph with sample plots indicated as white squares.

Reference data

Tree species specific reference data was collected during multiple phases in 2008, 2009 and 2010. Within 24 square plots of 30 m side length the position of the crown centre was measured for each individual tree and recorded together with the referring tree species. Additionally the crown class was estimated from which only dominant, co-dominant and intermediate were selected due to their visibility from above. According to Hildebrandt (1996) these three upper classes make up to 97% of the whole timber volume. Training data was manually selected as explained in Heinzel and Koch (2010). The tree top positions were allocated to single crowns and delineated by following their outline on a normalized digital surface model (nDSM) and colour infrared (CIR) images. Afterwards these polygons were intersected with the feature grids to receive sample cells of known tree species. Altogether the plots were selected to enclose an equal number of samples for each species.

Feature Extraction

From the LiDAR point data a comprehensive set of composed features was derived. The features were originally computed for a slightly different dataset to exploit the information of secondary parameters from the waveform geometry and is comprehensively explained in Heinzel and Koch (2010). The point cloud data is projected onto a grid with a cell size of 1 m. Each grid layer represents one feature which consists of four components. These are first the primary geometric information such as the width, the amplitude and the intensity of the reflection as well as the total number of reflections within a laser beam. Second, a measure for the statistical distribution of the reflection specific values within a grid cell and third and fourth, the position of the reflection within the ray and the position in space respectively. Altogether 231 variables were available.

2. Methods

Within this compare

ative study two major classification methods are used. These are first the LDA and second the SVM classification.

Linear discriminant analysis

LDA is a widely-used classification method in remote sensing image classification. In this study we use the software package SPSS to conduct all LDA classifications. Due to the well known principles and functionality of LDA we only give a brief summary of the settings and framework we use.

LDA aims to find new variables that allow an optimized linear separation of the groups. The general form of the according discriminate function Y is a linear combination of the feature variable X_i with $i = 1,...,d$:

$$Y = a_0 + a_1 X_1 + a_2 X_2 + ... + a_d X_d \quad (1)$$

The discriminate coefficients a_i are determined by maximizing the ratio of the explained sum of squares and the unexplained sum of squares. To receive a ranking of the original 231 variables concerning their importance for discrimination we conduct a forward stepwise LDA. The variable which decreases Wilks' lambda most is iteratively included in the analysis with each step. We use an F-value of 3.84 as threshold for inclusion and 2.71 for exclusion according to Pollar *et al.* (2007).

As a summary of the classification principle it can be said that LDA tries to find the average object of each group and allocates new objects to the group with the most similar mean value. Further LDA aims on the reduction of variables used for classification and is limited to the linear discrimination of groups.

Support vector machines

SVMs use a different classification principle. They try to find those training objects (vectors) of a group which are most similar to the other second group. These objects are then called support vectors because they form the decision boundary for classification. The new object is then compared to each of these support vectors. A

SVM is primarily a two class separation tool. To handle multiple classes there are two major methods. These are the separation of the multiclass problem into all possible two class pairs (one versus one) or the comparison of each class against the rest (one versus rest). The theory of support vector machines is explained in several scientific papers and tutorials (Burges 1998, Bennett and Campbell 2000). A separate description of the SVM theory would go beyond the scope of this paper. Instead we give an explanation of the framework in which the SVM classification should be used to receive optimal results. To conduct fast SVM computations we use the LIBSVMTL programming library (Ronneberger 2004).

The general form of the decision function of the SVM is given in equation (2).

$$f(\boldsymbol{x}) = sign(\sum_{i=1}^{l} y_i \alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i) + b) \quad (2)$$

Here $\boldsymbol{x}$ is the vector to be classified, $\boldsymbol{x}_i$ is the i^{th} support vector, y_i is the known class of $\boldsymbol{x}_i$, α_i is the lagrangian multiplier and b is the bias. $K(\boldsymbol{x},\boldsymbol{x}_i)$ is a kernel function that allows a nonlinear classification while remaining in the input feature space. One of the most universal kernels is the radial basis function (RBF) as shown in equation (3).

$$K(\boldsymbol{x}, \boldsymbol{x}_i) = \exp\left(-\gamma \| \boldsymbol{x} - \boldsymbol{x}_i \|^2\right) \quad (3)$$

$$\text{s.t.} \qquad \gamma > 0$$

A complete SVM classification procedure can be subdivided in three hierarchical levels (Figure 6). To receive meaningful results it is important to follow this scheme. The upper level comprises the gridsearch method that allows an optimization of the kernel parameter γ and the cost factor C which is used for estimating α_i (Tso and Mather 2009). This is important since both variables differ for each classification problem and therefore no standard values exist. Starting at an initial value they are exponentially increased for a defined number of times. In case of a linear SVM only C has to be optimized. The mid level contains an accuracy assessment for each variable pair determined by the gridsearch. We use a 24-fold cross-validation referring to the 24 sample plots on the test site. With this we receive a mean accuracy for each of the parameter settings. The lower level conducts the actual multiclass SVM classification. When using the one versus one method altogether $k(k - 1) /2$ SVMs are required for k classes.

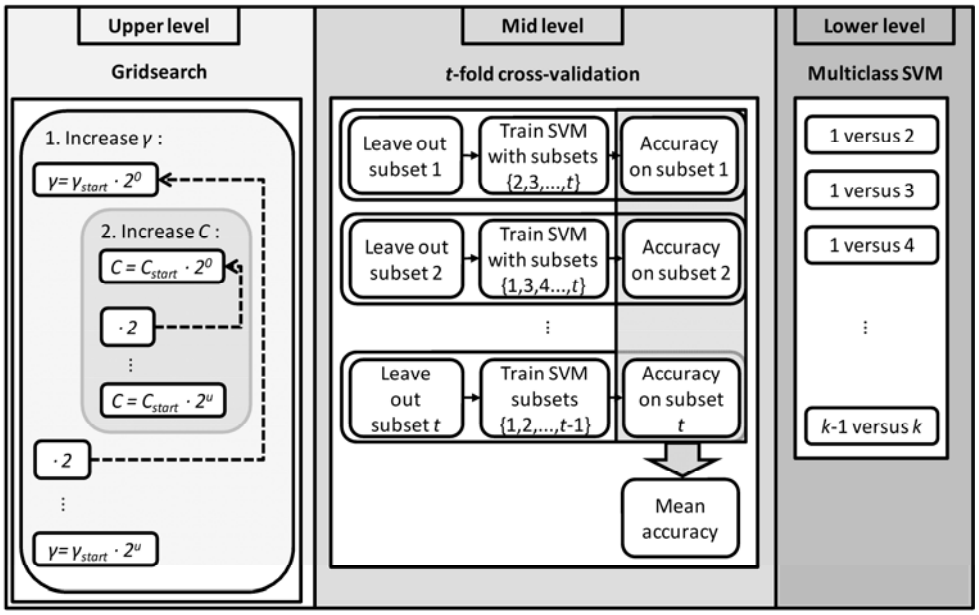


Figure 6. The hierarchical levels which are required to conduct an optimal SVM classification.

After plotting the accuracies into the referring parameter grid the optimal parameterization can be easily determined as the global maximum of the grid values. These values are then used to train the final SVM model.

We tested SVM classifications after the above described method with different configurations. These include the kernel, where we used both RBF and linear SVM and the multiclass methods one versus one as well as one versus rest. Supplementary to the usage of all feature we conducted all SVM classifications with the same ranked features as used for the LDA. This way of feature ranking can be described as a filter ranking which means, that the ranking is done by a method separated from and prior to the actual classifier (Kohavi and John 1997).

3. Results

The classification results of different SVM configurations are shown in Table 1. The columns distinguish the use of an RBF kernel and a linear SVM as well as the two different multiclass methods. The rows show the overall accuracies for three classification depths. The RBF kernel with a one versus one (ovo) method achieves the best results for all classification depths. We therefore chose this SVM setting for further comparison with the LDA.

Table 10. Overall accuracy (%) for SVM classification with different configurations

	One versus one		One versus rest	
	RBF	Linear	RBF	Linear
Six species	56.10	55.90	53.84	47.67
Main Species	79.22	78.33	77.40	73.51
Con/Dec	94.43	93.90	94.43	93.90

Table 2 gives an overview of these comparative classification results. The columns distinguish LDA and SVM accuracies while they are further subdivided for the use of all 231 variables of the feature set and those variables selected by the stepwise discriminant analysis. The number of the selected variables by the stepwise LDA is 64 for the six species, 44 for the main species and 48 for deciduous-coniferous classification. It can be seen that the SVM outperforms LDA for the main species and the deciduous-coniferous when using all features. LDA is better in the case of six species with selected features. A further difference between both classifiers is that LDA accuracies increase when limiting the feature space to the stepwise included variables. On the other hand SVM accuracy results nearly stay constant.

Table 11. LDA and SVM classification accuracies (%) in comparison for rank selected and non-ranked features as well as for different number of classes.

	LDA		SVM (RBF/ovo)	
	All features	Top features	All features	Top features
Six species	57.89	61.81	56.10	55.26
Main Species	75.07	79.59	79.22	79.83
Con/Dec	90.18	93.10	94.43	94.73

Additionally we observed the classification accuracies when the top ten variables were included in a stepwise procedure. Therefore, we determined the variables which occur most often on each of the upper ten ranks during the 24-fold stepwise discriminant analysis. It can be seen from Figure 3 that the most relevant improvement of accuracy occurs with the first three variables. This is the case for both classifiers and all classification depths.

The three most important variables are very similar for all classification depths. In the first place it is the arithmetic mean of the intensity within a grid cell from laser beams with only one reflection, in the second place the mean of the signal width using all targets within the beam and in the third place the mean of the total number of targets within a beam. In all cases the considered reflections were limited to the upper 5 m of the canopy. Deciduous-coniferous classification has the same variables within the top three ranking but uses the number of targets in place one and the intensity variable in place three.

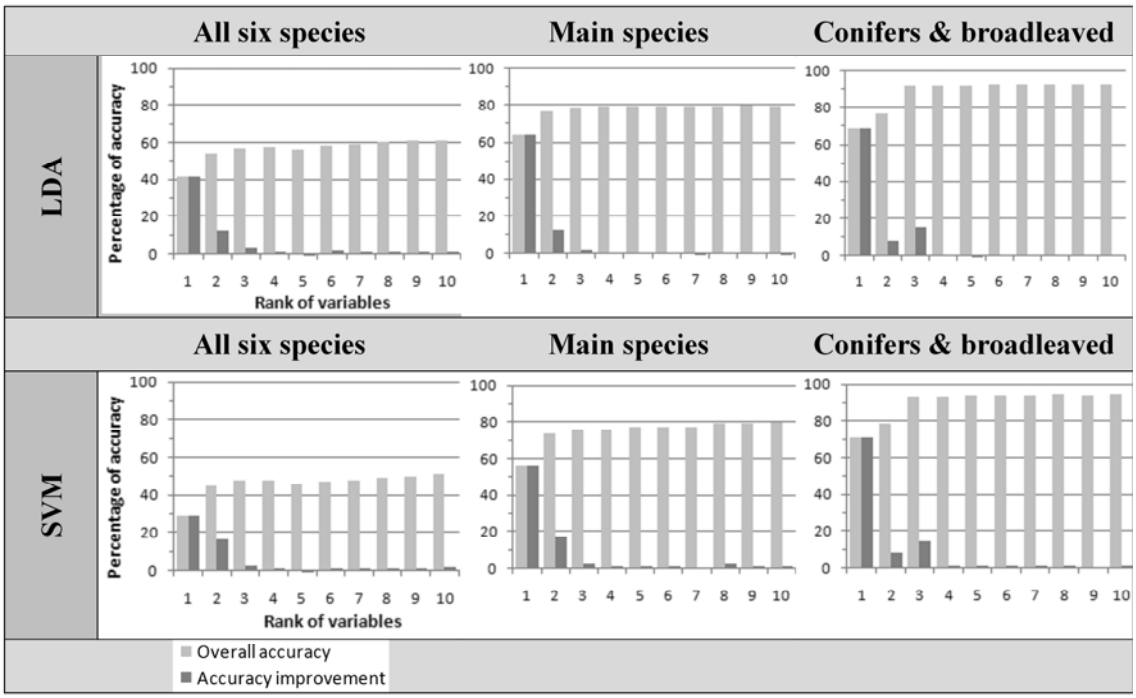


Figure 7. Classification accuracies for the averaged top ten variables. The accuracy is given for each number of topmost ranked variables together with the improvement in comparison to the next lower number.

4. Discussion

The initial test of different SVM configurations reveals the advantage of the RBF kernel in comparison to a linear SVM and the superiority of the one versus one multi class method. These results confirm the study of Pal and Mather (2005) who classified a multi- and hyperspectral remote sensing dataset and came to the same conclusions. We have also pointed out the gridsearch method which is obligatory to receive optimal SVM results. This step seems often to be neglected in application oriented studies which use SVMs as a black box tool.

The comparative classification results show that in most cases using the complete feature set SVM outperforms LDA classification more than 4% in overall accuracy. After limiting the feature set to the selected features from stepwise LDA the accuracies for the main species and deciduous-coniferous trees come very close to the SVM results (see Table 2). In the case of six species we have the unexpected phenomenon that LDA even overtops SVM. However, it should be considered that SVM results may underlie small variations due to the step size in the gridsearch algorithm. Nevertheless, step sizes should not be too small to avoid

overfitting of the model. The increase in accuracy after removing noise producing variables refers to the general character of LDA. All three classification depths clearly show this behaviour. In contrast SVM results stay more or less constant which points to a higher stability of SVM in high dimensional feature spaces (Chen and Ho 2008).

Focusing the iterative inclusion of the ten uppermost ranked features shows a very similar development of the results for both classifiers. Within all classification depths the major increase in accuracy is restricted to the top two or three variables. The inclusion of further variables only results in minor changes of accuracy. Since SVM classification uses the same rank selected features derived by stepwise LDA, it can be assumed that the ranking is not optimal for the SVM. Therefore we see the necessity for future work in developing an SVM specific ranking and testing the classifier with a customized feature selection.

Altogether the often indicated superiority of non-linear SVMs in comparison to more simple classifiers (Gokcen and Peng 2002, Chen and Ho 2008, Ørka *et al.* 2009b) is only partly confirmed with this study when using the complete feature set. With an LDA optimized set of variables we achieve surprisingly good accuracies for this classifier. Even if our restriction to a single example does not allow any generalization, it indicates that the differences in performance depend on the extent the classifiers fit to the problem at hand (Gokcen and Peng 2002). It can be concluded that even if one classifier outclasses another in theory, as shown for SVM and LDA by Gokcen and Peng (2002), in the individual case the superiority need not to be mandatory.

Acknowledgements

This work was financed by the Deutsche Forschungsgemeinschaft (DFG). We would also like to thank TopoSys® GmbH for good cooperation and providing the high quality LiDAR data.

References

BENNETT, K. P., and CAMPBELL, C., 2000, Support vector machines: hype or hallelujah? *ACM SIGKDD Explorations Newsletter*, **2**, pp. 1-13.

BRANDTBERG, T., 2007, Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar. *Isprs Journal of Photogrammetry and Remote Sensing*, **61**, pp. 325-340.

BRANDTBERG, T., WARNER, T. A., LANDENBERGER, R. E., and MCGRAW, J. B., 2003, Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in North America. *Remote Sensing of Environment*, **85**, pp. 290-303.

BURGES, C. J. C., 1998, A tutorial on Support Vector Machines for pattern recognition. *Data Mining and Knowledge Discovery*, **2**, pp. 121-167.

CHEN, C. H., and HO, P. G. P., 2008, Statistical pattern recognition in remote sensing. *Pattern Recognition*, **41**, pp. 2731-2741.

GOKCEN, I., and PENG, J., 2002, Comparing linear discriminant analysis and support vector machines. *Lecture Notes in Computer Science*, **2457**, pp. 104-113.

HEINZEL, J., and KOCH, B., 2010, Exploring full-waveform LiDAR parameters for tree species classification. *International Journal of Applied Earth Observation and Geoinformation* (submitted).

HILDEBRANDT, G., 1996, *Fernerkundung und Luftbildmessung* (Heidelberg: Wichmann).

HOLLAUS, M., MÜCKE, W., HÖFLE, B., DORIGO, W., PFEIFER, N., WAGNER, W., BAUERHANSL, C., and REGNER, B., 2009, Tree species classification based on full-waveform airborne laser scanning data. In *Silvilaser 2009*, 14-16 October 2009, College Station, USA, pp. 54-62.

HOLMGREN, J., and PERSSON, A., 2004, Identifying species of individual trees using airborne laser scanner. *Remote Sensing of Environment*, **90**, pp. 415-423.

HOLMGREN, J., PERSSON, A., and SODERMAN, U., 2008, Species identification of individual trees by combining high resolution LIDAR data with multi-spectral images. *International Journal of Remote Sensing*, **29**, pp. 1537-1552.

KIM, S., MCGAUGHEY, R. J., ANDERSEN, H. E., and SCHREUDER, G., 2009, Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. *Remote Sensing of Environment*, **113**, pp. 1575-1586.

KOHAVI, R., and JOHN, G. H., 1997, Wrappers for feature subset selection. *Artificial Intelligence*, **97**, pp. 273-324.

LITKEY, P., RÖNNHOLM, P., LUMME, J., and LIANG, X., 2007, Waveform features for tree identification. In P. Rönnholm, H. Hyyppä, and J. Hyyppä (Eds.), International Archives of Photogrammetry, Remote Sensing and Spatial Information, Vol. 36 Part 3/W52, *Proceedings of the ISPRS Workshop 'Laser Scanning 2007 and SilviLaser 2007'*, 12-14 September 2007 (Espoo, Finland, pp. 258-263.

LUCAS, R., BUNTING, P., PATERSON, M., and CHISHOLM, L., 2008, Classification of Australian forest communities using aerial photography, CASI and HyMap data. *Remote Sensing of Environment*, **112**, pp. 2088-2103.

MEYER, P., STAENZ, K., and ITTEN, K. I., 1996, Semi-automated procedures for tree species identification in high spatial resolution data from digitized colour infrared-aerial photography. *Isprs Journal of Photogrammetry and Remote Sensing*, **51**, pp. 5-16.

ØRKA, H. O., NAESSET, E., and BOLLANDSAS, O. M., 2009a, Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. *Remote Sensing of Environment*, **113**, pp. 1163-1174.

ØRKA, H. O., NÆSSET, E., and BOLLANDSÅS, O. M., 2009b, Comparing classification strategies for tree species recognition using airborne laser scanner data. In *Silvilaser 2009*, 14-16 October 2009, College Station, USA, pp. 46-53.

PAL, M., and MATHER, P. M., 2005, Support vector machines for classification in remote sensing. *International Journal of Remote Sensing*, **26**, pp. 1007-1011.

PINZ, A. J., and BISCHOF, H., 1990, Constructing a neural network for the interpretation of species of trees in aerial photographs. In *10th International Conference on Pattern Recognition* 16-21 June 1990, Atlantic Citys, USA (Omaha, USA: IEEE Computer Society), pp. 755-757.

POLLAR, M., JAROENSUTASINEE, M., and JAROENSUTASINEE, K., 2007, Morphometric analysis of tor tambroides by stepwise discriminant and neural network analysis. *Engineering and Technology*, **33**, pp. 16-20.

REITBERGER, J., KRZYSTEK, P., and STILLA, U., 2008, Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees. *International Journal of Remote Sensing*, **29**, pp. 1407-1431.

RONNEBERGER, O. 2004. LIBSVMTI - a Support Vector Machine Template Library. Available online at: http://lmb.informatik.uni-freiburg.de/lmbsoft/libsvmtl/index.en.html (accessed 16/10/2010).

TOKOLA, T., VAUHKONEN, J., LEPPÄNEN, V., PUSA, T., MEHTÄTALO, L., and PITKÄNEN, J., 2008, Applied 3D texture features in ALS based tree species segmentation. In G. J. Hay, T. Blaschke, and D. Marceau (Eds.), International Archives of Photogrammetry, Remote Sensing and Spatial Information, Vol. 38 Part 4/C1, *GEOBIA 2008*, 5-8 August 2008, Calgary, Canada (Calgary, Canada: International Society for Photogrammetry and Remote Sensing).

TSO, B., and MATHER, P. M., 2009, *Classification Methods for remotely sensed data* (Boca Raton, London, New York: CRC Press).

WASKE, B., and BENEDIKTSSON, J. A., 2007, Fusion of support vector machines for classification of multisensor data. *Ieee Transactions on Geoscience and Remote Sensing*, **45**, pp. 3858-3866.

USE OF LIDAR FOR OPERATIONAL FOREST INVENTORY IN THE PACIFIC NORTHWEST, USA: CURRENT STATUS AND FUTURE DIRECTIONS

Hans-Erik Andersen, USDA Forest Service Pacific Northwest Research Station, Anchorage, Alaska, USA

As government land management agencies and private companies in the Pacific Northwest of the United States (Washington, Oregon and southern Alaska) begin to consider the use of lidar in operational forest inventory systems, a number of important issues have emerged that can have a significant impact on the success of these projects. In this presentation, we discuss several of these emerging questions related to the use of lidar in operational forest inventories and discuss how they are being addressed in practice, including 1) Do we have a standardized set of data acquisition specifications? Numerous mission specifications can influence the utility of lidar data for inventory applications, including time of year, data density, among others. The utility of lidar as a monitoring tool could depend upon standardization of specifications to allow for detection of meaningful changes. Acquisition of highly-accurate plot coordinates has also become more important. 2) Do we have the tools to view, analyze, and process large raw lidar datasets? Lidar data is generally difficult to manipulate in standard GIS environments. Tools are becoming more widely available, but they vary in terms of functionality and cost. 3) Do we have the models to relate lidar metrics to inventory attributes of interest? The efficiency of lidar forest inventories will directly depend on how well the models describe the relationship between lidar-derived information and inventory parameters. 4) Do we have the statistical framework to develop parameter estimates (and the variances associated with these estimates) over large areas? The use of lidar in operational forest inventories can require the development of new sampling designs, drawing from both model-based and model-assisted theory. 5) How can lidar be most useful to forest managers and planners in the short term and long term? Lidar will be used at the regional, strategic, and tactical levels of management, but the required products will differ depending on the management objectives. Lidar scientists need to prioritize short-term management needs in order to maximize impact.

ALS based estimation of plot volume and site index in a eucalyptus plantation with a nonlinear mixed effect model that accounts for the clone effect

*PETTERI PACKALÉN *†, LAURI MEHTÄTALO† and MATTI MALTAMO†*

† Faculty of Science and Forestry, University of Eastern Finland, P.O. Box 111, FI-80101 Joensuu, Finland.

Abstract

Most ALS studies have been carried out in semi-natural forests but some research has also been carried out in plantations. Results indicate that methods similar to those which are used in semi-natural forest are also usable in plantation forestry. The aims of this study are to investigate (1) how accurately the plot volume (V) may be estimated by ALS data in eucalyptus plantations, and (2) how to estimate the site index (SI) directly by combining ALS data and stand age. The study was conducted in a pulpwood plantation growing *Eucalyptus urograndis* in Bahia state, Brazil. The data included field measurements from 55 stands and 28 different clones; all the trees in a stand belong to the same clone. The V and SI were estimated by means of nonlinear mixed effect modelling in order to take into account the stand within clone hierarchy of the data. The obtained accuracies are quite good if compared to those obtained in semi-natural forests. The RMSE was 8.2% for V and 2.7% for SI when clone effect was used in prediction. It is worth noting that in eucalyptus plantations the clone is always known and therefore its use does not restrict the applicability of this method. If none of the sample plots is located within a particular clone, only the fixed part of the model may be used. In the estimation of SI also the stand age was taken from the stand register data. This does not restrict the applicability of the method either, since the planting date is always known and trees are of the same age. Therefore the SI may be predicted in a wall-to-wall manner across the whole plantation. Precision forestry applied in plantations differs many ways from the forestry practiced in semi-natural environment. ALS based forest inventory has a lot of potential in pulpwood plantations when the unique features of plantation forestry are taken into account.

1. Introduction

Although most ALS (Airborne Laser Scanning) studies have been done in semi-natural forests, some research has also been carried out in plantations (e.g. McCombs et al. 2003, Wack et al. 2003, Roberts et al. 2005, Donoghue et al. 2007, Rombouts et al. 2008, Hopkinson et al. 2008, Tesfamichael 2009, Zonete et al. 2010). The focus in these studies has been the accuracy of ALS based stand attribute estimates including leaf area and growth. Results indicate that methods similar to those used in semi-natural forests also seem to be appropriate for plantation forestry. So far, most studies have dealt with conifer plantations, an exception being the articles by Wack et al. (2003), Zonete et al. (2010), and the PhD thesis by Tesfamichael (2009, cf. journal publications) in which eucalyptus was considered. Both area based method (Næsset 1997) and individual tree detection (Hyyppä and Inkinen 1999) have been used studies in plantations.

Plantation forestry differs in many ways from the forestry practiced in semi-natural environments. There are several kinds of forest plantations. The aim may be to provide wildlife-habitat, biological diversity, or other services in addition to wood production: in such cases multi-species forest plantations are favoured (FAO 2002). The type of plantation considered here, however, is a mono-species eucalyptus plantation which is targeted to the pulpwood production only. In eucalyptus plantations, trees are planted in rows using a fixed stand density. There is no natural regeneration, thus trees within a stand are of the same age and single storied. Typically, all trees in a stand belong to the same clone; the aim is to select the most productive clone to a particular stand. Productivity is quantified by the site index (SI), which is usually based on the development of a dominant height over the rotation period.

Plantation forestry also occupies a special situation regarding remote sensing based forest assessment. Although the stand level information about age and tree species is often available also for semi-natural forests, this information is especially accurate in even age plantations where trees are of the same age and also the clone is known in certain cases. This makes it more feasible to combine existing stand register data

and remote sensing data and to carry out estimation in a wall-to-wall manner. Accurate age at tree level also enables the estimation of attributes which are normally not feasibly estimated by remote sensing, such as the variation of site index (SI) within a stand.

The aims of this study are to investigate (1) how accurately the plot volume (V) may be estimated by ALS data in Eucalyptus plantation, and (2) how to estimate SI directly by combining ALS data and age taken from the stand register. In order to utilize efficiently all available information, the estimation was carried out with a nonlinear mixed effect model that accounts for the clone effect.

Material

Study area and field data

The study area is a pulpwood plantation growing eucalyptus in Bahia state, Brazil. The plantation is owned by Veracel and the pulp mill itself is sited within the test area. The cultivated *Eucalyptus* species is *E. urograndis*, which is a hybrid between *E. grandis* and *E. urophylla*. This eucalyptus hybrid is highly productive and is one of the main species currently used for pulp production in Brazil (Siverio et al. 2007).

A network of 195 circular sample plots with a radius of 13 meters was established and measured in August–September 2008. Sample plots were placed in 55 forest stands with three or four plots per stand. Satellite positioning was used to determine the position of each plot center using a real-time differential correction signal from the OmniSTAR satellite (http://www.omnistar.com). The trees were growing in rows and the tree spacing was fixed, giving a density of 833 stems per hectare. All trees were recorded in the field for diameter at breast height (*d*) and quality, and every seventh tree on each plot was measured for height (*h*). Näslund's (1937) *h*-*d* curve was fitted by stands and used to predict heights for trees without height measurement. Stem volumes were calculated as a function of *d* and *h* using a clone and age class specific model constructed in-house in Veracel.

Plot volume (V) was calculated by aggregating from tree to plot level, and dominant height (HD) was calculated as the mean height of the 100 thickest trees at breast height per hectare. Stand age (T) was taken from the plantation database in which it was registered with the precision of one month. There were 28 different clones in the 55 stands from which the field data were collected. All trees in a stand belonged to the same clone.

The site index (SI) was predicted using the following form of the Chapman–Richards Equation (e.g., Clutter et al. 1983):

$$SI = HD\left(\frac{1-e^{-\beta_1 t_{reference}}}{1-e^{-\beta_1 t_{current}}}\right)^{\beta_2}, \quad (1)$$

where $t_{reference}$ is the reference age of 7 years, HD is the current dominant height, $t_{current}$ is the current age, and β_1 (0.3341) and β_2 (1.1442) are known model parameters in Veracel.

Airborne laser scanning data

ALS data were collected on August 16, 2008 using an Optech ALTM 3100 laser scanning system. The test site was measured from an altitude of approximately 1200 m above ground level using a field of view of 30 degrees. Pulse repetition frequency was set to 50 000 pulses per second, which resulted a nominal sampling density of about 1.5 measurements per square meter. The footprint was about 35 cm at ground level.

A digital terrain model (DTM) was generated from the ALS data. First, laser points were classified as ground and non-ground points using the method reported by Axelsson (2000). Then a raster DTM with a 1-meter pixel size was interpolated using ground points and an inverse distance weighting algorithm (Lloyd and Atkinson 2002). Finally, the raster DTM was subtracted from the ellipsoidal heights of laser points in order to scale the ALS data to the above-ground level (AGL).

2. Methods

2.1 Explanatory variables

The laser scanner used captures a maximum of four range measurements for each submitted pulse. These echo categories are 'first of many', 'last of many', 'only' and 'intermediate'. After preliminary tests it was decided that only the echo categories 'first of many', and 'only' would be used in this study, since the exclusion of 'last of many' and 'intermediate' echoes did not significantly decrease the accuracy of the SI and V estimates. This set contains all of the first – or surface – echoes since an 'only' echo may also be considered as a first echo.

Numerous height and density metrics were calculated from the combined set of 'first of many' and 'only' echoes. The principle of the area based method was used here. The first step was to calculate height distributions for each sample plot using the heights of the AGL data. All the laser hits were considered, also those lying on the ground. Height quantiles for 5, 10, 20,…, 80, 90, 95% (*h5,…, h95*) were computed, and the corresponding densities (*p5,…, p95*) were calculated for the respective quantiles. Height quantiles were calculated by summing the heights at AGL. For instance, the metric h50 is the height at which 50% of the cumulative height has accumulated and p50 is the number of laser hits below h50 divided by all the laser hits on the plot. In addition, the mean (*hmean*) and standard deviation (*hstd*) of heights at AGL were calculated by plots. These metrics form a set of candidate explanatory variables used for modeling SI and V.

2.2 Modeling of site index and plot volume

The same form of the Chapman-Richards Equation which is used to predict the SI in this plantation is used as the starting point in the modeling of SI (see Eq. 1). However, instead of modeling the HD separately and inserting the resulting predictions into Eq. 1, the SI was modeled directly. This was done by replacing the HD of Eq. 1 with a linear dominant height model that uses ALS based explanatory variables, and fitting the equation as a nonlinear model into the modeling data. We also incorporated nested random effects into the model to account for the effects of the hierarchical data (sample plot within stand within clone) on the model residuals. To begin with, however, the form of the HD model was chosen by using linear regression. At this stage, both manual insertion and deletion of explanatory variables and a stepwise selection based on the Akaike information criterion were used.

Preliminary analyses for the model of V showed a linear relationship between ln(V) and the predictors. That is why a nonlinear model of the exponential form was used to model the total volume. This led to a model that has the observed linear relationship and is unbiased for V. This model was also fitted as a mixed effects model to account for data hierarchy.

The nonlinear mixed effect models were fitted by using the *nlme* routine (Lindstrom and Bates 1990, Pinheiro and Bates 2000) in the R environment (R Development Core Team 2009) by using the method of maximum likelihood.

2.3 Accuracy assessment

Accuracy assessment was carried out at three different levels: for the fixed part of the mixed effect model, by using the predicted clone effect, and by using the predicted stand and clone effects. The accuracy of estimates was evaluated in terms of relative root mean squared error and bias at the plot level:

$$\text{RMSE-\%} = \frac{\sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}}}{y_{mean}} \times 100 \qquad (2)$$

$$\text{BIAS-\%} = \frac{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)}{n}}{y_{mean}} \times 100 \qquad (3)$$

where n is the number of plots, y_i is the observed value for plot i, $\hat{y}_i$ is the predicted value for plot I, and y_{mean} is the observed mean of the variable in question.

3. Results

Nonlinear mixed effect models

The model for the SI is as follows:

$$SI_{cki} = \left(\frac{1}{\beta_3 + \beta_4 h90_{cki} + \beta_5 h90_{cki}^2 + \beta_6 \sqrt{hmean_{cki}}}\right)\left(\frac{1-e^{-\beta_2 reference}}{1-e^{-\beta_2 current,ck}}\right)^{\beta_1} + b_c + b_{ck} + \varepsilon_{cki} \quad (4)$$

where SI_{cki} is the SI for clone *c*, stand *k*, and plot *i*; b_c is a random clone effect; b_{ck} is a random stand effect (stands are nested within clones); and ε_{cki} is the residual for clone *c*, stand *k*, and plot *i*. β_1 and β_2 are the known parameters of the SI model (the same as in Eq. 1). The random effects and residuals are assumed to be normally distributed with mean zero and constant variance.

The model for the plot volume is as follows:

$$V_{cki} = e^{\left(\beta_7 + \beta_8 \log(h10) + \beta_9 \log(h50) - \beta_{10}\sqrt{p90} + b_c + b_{ck}\right)} + \varepsilon_{cki} \quad (5)$$

where V_{cki} is V for clone *c*, stand *k*, and plot *I*; b_c is a random clone effect; b_{ck} is a random stand effect (stands are nested within clones); and ε_{cki} is the residual for clone *c*, stand *k*, and plot *i*. The random effects and residuals are assumed to be normally distributed with mean zero and constant variance. Preliminary analyses showed increasing residual variance with respect to the prediction, so the residual was assumed to be normally distributed with mean zero and variance $var(\varepsilon_{cki})=\sigma^2 h50^{2\delta}$, where σ and δ are the parameters of the variance function. The parameter estimates for both models are listed in Table 1.

Table 1. Parameter estimates for the fixed independent variables and estimated variances for the random effects at the clone, forest stand, and plot level.

Equation 4			Equation 5		
Coefficient	Estimate	SE	Coefficient	Estimate	SE
β_3	0.0977596	0.0028945	β_7	11.415382	2.298940
β_4	-0.0034367	0.0001932	β_8	0.142945	0.077417
β_5	0.0000432	0.0000031	β_9	1.460686	0.095138
β_6	-0.0008237	0.0003894	β_{10}	-1.133237	0.236214
Random parameters			Random parameters		
$Var(b_c)$	0.478231^2		$Var(b_c)$	0.064235^2	
$Var(b_{ck})$	0.726614^2		$Var(b_{ck})$	0.065542^2	
$Var(\varepsilon_{cki})$	0.659612^2		$Var(\varepsilon_{cki})$	$0.015319^2 h50^{4.236}$	

Accuracies

The accuracy and bias of the estimates of SI and V are presented in Table 2. Level denotes the level of grouping used for obtaining the predictions. Level 'Fixed' means that only the fixed part of the model (Eq. 4 or 5) is used, level 'Clone' means that the predicted clone effect is used in prediction, and 'Stand' means that also the predicted random effect for the stand within clone is used in the prediction.

The estimates of V were already rather accurate without random effects (Table 2), however, the inclusion of random clone effects improved the accuracy from 11.82 to 8.22, and the further inclusion of random stand effects decreased the RMSE-% to 4.90. There was some bias in the estimates of V when only the fixed part of the model was used, without random components.

The accuracy of estimation of SI was 3.18% when only the fixed part of the model was used in the prediction. The inclusion of random effects improved the accuracy notably but not quite as much as in the case of V. The SI estimates were virtually unbiased at all levels.

Table 2. The RMSE and bias of the estimates of site index (SI) and volume (V) at different levels of grouping.

	SI		V	
Level	RMSE-%	BIAS-%	RMSE-%	BIAS-%
Fixed	3.18	0.41	11.82	2.20
Clone	2.66	0.10	8.22	0.41
Stand	1.68	0.02	4.90	-0.07

4. Discussion

Precision forestry applied to plantations differs in many ways from the forestry practiced in semi-natural environments. ALS based forest inventory and assessment has a lot of potential in pulpwood plantations when the unique features of plantation forestry are taken into account.

In this study, SI was modeled directly in one stage. Another option is a two-stage approach, where the HD is first modeled and then the resulting prediction is used in the known SI curve. Direct modeling approach is justified since the only use of the predicted HD is the prediction of SI. In addition, the model was fitted in a nonlinear form, where the SI was treated as the independent variable as such, without transformations. These two methodological choices led to unbiased prediction of SI, which could not have been guaranteed by using the two-stage method and/or a linearized form of the applied HD model. Also, the applied nonlinear model for plot volume led to unbiased prediction of volume. With the selected model form, the non-homogeneous variance was satisfactorily homogenized by a variance function of the power form.

Using the clone and stand effects in prediction improved the accuracy of the prediction considerably. This improvement resulted from using all the plot-specific observations of the response (V or SI) for the stand or clone under consideration in prediction. For clones that do not have measurements available, this level of accuracy is not possible, because the prediction can be done only at the 'Fixed' level. For clones with measurements from different stands, the prediction can be done at the 'Clone' level. Prediction at the 'Stand' level can be done only if sample plot measurements from that particular stand are available. The practical use of the model arises from the possibility to use the clone effect in prediction for stands with no measured sample plots. In those cases, the prediction could be close to the reported accuracy at the clone level.

The drawback of using predictors based on ALS data is that the models are local. The properties of point cloud produced by ALS is dependent on the scanning configuration, e.g., what is the pulse repetition frequency or flying altitude; also, the responses of different sensors differ (Naesset 2009). Therefore, such models as presented here must be made by plantations but in practical terms this is not a major restriction. If several sensors or different scanning configurations are used to collect ALS data from a plantation, the

sensor effect may be taken into account as a random effect in a similar manner as were the clone and stand considered here. However, this definitely decreases the accuracy.

The clone level explained a considerable amount of the between-stand variation for both models, and including the clone effect in the prediction improves the accuracy considerably. This observation leads to suggesting the placing of sample plots over the whole inventory area so that some measurements are available for all the clones. Then the prediction could be carried on a grid to the whole plantation at the clone level. Final inventory results may be obtained by aggregation to stand level, or cell level results may be used as such. For instance, SI estimated in a wall-to-wall manner provides new possibities to carry out precision forestry.

References

CLUTTER, J.L., FORTSON, J.C., PIENAAR, L.V., BRISTER, G.H. and BAILEY, R.L., 1983, *Timber management: a quantitative approach*. (New York: John Wiley & Sons).

DONOGHUE, D.N.M., WATT, P.J., COX, N.J. and WILSON, J., 2007, Remote sensing of species mixtures in conifer plantations using Lidar height and intensity data. *Remote Sensing of Environment*, **110**, pp. 509–522.

FAO, 2002, *Forest plantation productivity. Report based on the work of W.J. Libby and C. Palmberg-Lerche.* In *Forest Plantation Thematic Papers, Working Paper 3.* Forest Resources Development Service, Forest Resources Division. FAO, Rome.

HOPKINSON, C., CHASMER, L. and HALL, R.J., 2008, The uncertainty in conifer plantation growth prediction from multi-temporal lidar datasets. *Remote Sensing of Environment*, **112**, pp. 1168–1180.

HYYPPÄ, J. and INKINEN, M., 1999, Detecting and estimating attributes for single trees using laser scanner. *The Photogrammetric Journal of Finland*, **16**, pp. 27–42.

LINDSTROM, M.J. and BATES, D.M., 1990, Nonlinear mixed effects models for repeated measures data. *Biometrics*, **46**, pp. 673–687.

MCCOMBS, J.W., ROBERTS, S.D. and EVANS, D.L., 2003, Influence of fusing lidar and multispectral imagery on remotely sensed estimates of stand density and mean tree height in a managed loblolly pine plantation. *Forest Science*, **49**, pp. 457–466.

NÆSSET, E., 1997, Estimating timber volume of forest stands using airborne laser scanner data. *Remote Sensing of Environment*, **51**, pp. 246–253.

NÆSSET, E. 2009, Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. *Remote Sensing of Environment*, **113**, pp. 148–159.

PINHEIRO, J.C. and BATES, D.M. 2000, *Mixed-effects models in S and S-PLUS*, (New York: Springer).

R Development Core Team, 2009, *R: A language and environment for statistical computing*. (Vienna: R Foundation for Statistical Computing). URL http://www.R-project.org.

ROBERTS, S.D., DEAN, T.J., EVANS, D.L., MCCOMBS, J.W., HARRINGTON, R.L. and GLASS, P.A., 2005, Estimating individual tree leaf area in loblolly pine plantations using LIDAR-derived measurements of height and crown dimensions. *Forest Ecology and Management*, **213**, pp. 54–70.

ROMBOUTS, J., FERGUSON, I.S., and LEECH, J.W., 2008, Variability of LiDAR volume prediction models for productivity assessment of radiata pine plantations in South Australia. In *Proc. of SilviLaser 2008, 8th international conference on LiDAR applications in forest assessment and inventory*, R. Hill, J. Rosette, and J. Suárez (eds). Edinburgh, pp. 39–49.

SIVERIO, F.O., BARBOSA, L.C.A., MALTHA, C.R.A., SILVESTRE, A.J.D., PILO-VELOSO, D. and GOMIDE, J.L., 2007, Characterization of lipophilic wood extractives from clones of *Eucalyptus urograndis* cultivated in Brazil. *BioResources*, **2**, pp. 157–168.

TESFAMICHAEL, S.G., 2009, Assessment of structural attributes of even-aged Eucalyptus grandis forest plantations using small-footprint discrete return lidar data. PhD thesis, University of KwaZulu-Natal, South Africa.

WACK, R., SCHARDT, M., LOHR, U., BARRUCHO, L. and OLIVEIRA, T., 2003, Forest inventory for Eucalyptus plantations based on airborne laser scanner data. In *Proc. of the International Society for Photogrammetry and Remote Sensing Symposium, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **34**, 3/W13. The Netherlands, pp. 40–46.

A model-based approach for estimating the height distribution of eucalyptus plantations using low-density ALS data

LAURI MEHTÄTALO†, ANNI VIROLAINEN†, JUKKA TUOMELA† and JUKKA NYBLOM*

lauri.mehtatalo@uef.fi

† Faculty of Science and Forestry, University of Eastern Finland, P.O. Box 111, FI-80101 Joensuu, Finland.

‡Department of Mathematics and Statistics, University of Jyväskylä, P.O. Box 35 (MaD), 40014 University of Jyväskylä, Finland

Abstract

Under some simplifying assumptions, observations collected by airborne laser scanners can be regarded as observed heights of the canopy surface of the stand. The canopy surface is a surface generated by individual tree crowns. Assuming that individual tree crowns have a solid surface at the top side, and that the stand is viewed directly from above, the height of canopy surface at a given point is defined as the maximum over the tree-specific surfaces at that point. Our earlier studies have shown how the probability of having the canopy surface above a given height depends on (i) the stand density, (ii) the distribution of tree heights, (iii) the shape of individual tree crowns, and (iv) the spatial pattern of tree locations (either random or strictly regular). Based on this result, we have proposed a method based on maximum likelihood for estimating the stand density and height distribution using the observations on the canopy surface. This paper applies the model to the situation of the area-based approach, estimating the mean and dominant height of eucalyptus plantations (E. urograndis in Bahia state, Brazil) with known stand density and rectangular pattern of tree locations. First, a set of 18 training sample plots with known tree heights are used to estimate the parameters specifying the shape of individual tree crowns. These estimates are then used to estimate the distribution of tree heights for another set of 18 evaluation sample plots. The results indicated poorer performance when compared to the widely used empirical area-based approach. The reason was most likely an unrealistic model for individual tree shape. Possible improvements will be studied in the future.

1. Introduction

The area-based approach to the forest inventory using airborne laser scanning (ALS) is rapidly developing to an operational tool for forest inventory (e.g. Naesset et al 2004). The idea of the approach is to combine information measured on ground sample plots with wall-to-wall information collected by low-pulse density (about one pulse per m^2) discrete return airborne laser scanners. A regression approach, either parametric or non-parametric, is used in generalizing the relationship of ALS data and ground measurements from sample plot locations to other locations. The predictors of the regression model are different characteristics of the laser hits, such as quantiles of the laser data, or proportion of laser hits below a given threshold height. We call this approach as the *empirical area-based approach* in contrast to the *model-based approach* proposed in this paper

The ALS data collection procedure provides data that includes direct measurements of the canopy surface. This is the main difference to other remote sensing methods, such as to the use of satellite images and aerial photographs, and explains the good performance of the method. However, the relationship between the laser data and tree characteristics is not very well known, and selection of the predictors to the models of the empirical area-based approach is based mostly on intuition and empirical findings, not on well-established theory on how laser observations are generated by individual tree crowns.

From a theoretical point of view, there are two main factors that have an effect into the ALS observations: (i) the vertical profile of the forest canopy, and (ii) the physical properties of the laser measurements. A good theoretical model would integrate both of these into a single model. A starting point for such a model was proposed in Mehtätalo and Nyblom (2009). Their model expressed the vertical profile of laser observations for a stand as a function of stand density, and the parameters of the height distribution of trees and the shape of individual tree crowns. The model was used to estimate the stand density and tree height distribution in simulated forest stands. However, the model included several simplifications, such as the assumption of random tree locations within the stand. Mehtätalo and Nyblom (2010) developed the model into a systematic, square grid spatial pattern, and evaluated the model both using simulated data and by using pre-processed empirical data. However, both above-mentioned studies assumed that the canopy surface is solid, the shape of a single tree as a function of tree height is known in advance, and the laser pulses are taken directly from above at an infinitesimal point. In general, the model satisfactorily considered the process that generates the vertical profile of the forest canopy, but did not tackle the physical properties of the laser measurements.

The aim of this study was to apply the model of Mehtätalo and Nyblom (2010) into the situation of the area-based forest inventory, and to test it with empirical data. The idea is to first estimate the parameters specifying individual tree crowns using laser data from plots with known tree heights. The information on the crown shape is then used with another set of plots to estimate the distribution of heights. The stand density was not estimated, because the rectangular pattern of tree locations was known in advance.

Material

The study material includes sample plots from the pulpwood plantation growing eucalyptus in Bahia state, Brazil. On each sample plot, all trees were callipered for diameter (*d*). Every seventh tree was measured for height (*h*). Näslund's *h*-*d* curve was fitted by stands and used to predict heights for trees without height measurement. Predicted heights from this model are hereafter regarded as the true heights, and the prediction error in them is ignored.

ALS data were collected on August 16, 2008 using an Optech ALTM 3100 laser scanning system. The flight altitude was 1200 m above ground level using a field of view of 30 degrees. The pulse density at the ground was about 1.5 measurements per square meter and the footprint diameter was about 35 cm. However, to reduce the computational burden, we used only every 6^{th} pulse of the original data. After this, we had on average 122 pulses for each of the 530 m^2 sample plot (0.23 pulses per m^2). The observations below 5 meters were taken as ground hits and treated as zeroes in the analysis.

The plots of this study are a subsample of the original data described in Packalen et al (2010). A total of 18 pairs of sample plots were selected systematically according to the canopy height properties so that they would represent the whole range of mean canopy heights in the plantation. The plots of each pair were then randomly assigned to the training and evaluation datasets. The same set of sample plots was used in Vauhkonen et al. (2010).

2. Methods

2.1. The model for canopy height

Let $Z(v)$ be the height of canopy surface at a fixed point v, i,.e., the vertical distance between the ground level and the canopy surface. The canopy surface at a given point can be thought as the maximum over the tree-specific surfaces at that point.

Mehtätalo and Nyblom (2009, 2010) showed that the probability of having $Z(v)$ below the height z above ground is the probability that none of the neighbor tree crowns at height z extends to point v,

$$P(Z(v) \leq z) = \prod_{i \in \mathcal{N}} F\{w(z, \|v - u_i\| | \boldsymbol{\theta}) | \boldsymbol{\xi}\}, \quad (1)$$

where

$\mathcal{N}$ is the set of neighboring trees of point *v* (see below)

$F(h|\boldsymbol{\xi})$ is the cumulative distribution function of tree height (e.g., the Weibull cdf),

$w(z,r|\boldsymbol{\theta})$ is a function that gives the total height for a tree that has crown radius *r* at height *z* above ground,

$\|v - u_i\|$ is the distance between the fixed point *v* and the location u_i of tree *i*

θ includes the parameters of the height distribution and

ξ includes the parameter that specify the shape of individual tree crown as the function of tree height

Equation (1) assumes independent and identically distributed tree heights within the stand. Furthermore, it assumes that tree crowns are solid so that the observations of *Z*(*v*) do not penetrate into tree canopies. The horizontal cross section of a tree crown is assumed to be circular or at least the orientation is assumed to be uniform between 0 and 360 degrees so that trees are circular on average. The set of neighboring trees $\mathcal{N}$ includes the trees that are located so close to the point *v* that their crowns (practically) have a nonzero probability to extend to the point *v*. In this study, we used a neighborhood of 16 trees (4 trees from 4 rows).

When laser observations are used, no information on the tree locations u_i with respect to *v* is available. However, it is natural to assume that the observations are placed uniformly over the stand area. Then it is justified to estimate $P(Z \leq z)$ as the mean of $P(Z(v) \leq z)$ over the stand area. However, if trees are planted into rows with distance *l* between rows, and having distance *m* between trees of a row, then the stand consists of *N* cells of size *m* by *l*, which all have equal mean of $P(Z(v) \leq z)$ (we forget the cells that are at the stand edge). Then it is enough to take the mean of $P(Z(v) \leq z)$ over only one cell. This mean is the integral of the right hand side of the equation 1 over the cell, divided by the cell area

$$G(z|\boldsymbol{\theta},\boldsymbol{\xi}) = P(Z \leq z) = \frac{1}{lm} \int_0^{lm} \prod_{i \in \mathcal{N}} F\{w(z, \|v - u_i\| | \boldsymbol{\theta}) | \boldsymbol{\xi}\} \, dv_1 \, dv_2. \quad (2)$$

By the definition of the cumulative distribution function (cdf, Casella and Berger 2001), this is the cdf of the laser observations, if the observations are taken at infinitesimal points directly from above and the other previously specified assumptions on the stand structure are met. Function (2) depends on two parameters: $\boldsymbol{\xi}$, which specifies the distribution function of the tree heights, and **θ,** which specifies the shape of an individual tree crown. In contrast to Mehtätalo and Nyblom (2010) the stand density is assumed to be known. If it were unknown, then the cell area, the bounds of the integral, and tree locations u_i would be functions of stand density.

2.2 Estimation

The probability density function (pdf) of the non-zero laser observations is obtained by differentiating (2) with respect to *z* as

$$g(z|\boldsymbol{\theta},\xi) = G'(z|\boldsymbol{\theta},\xi) \text{ for } z>0. \quad (3)$$

In addition, the probability to have canopy height of zero is $P(Z \leq 0) = G(0|\boldsymbol{\theta},\xi)$.

Assuming that the laser observations are an i.i.d. sample from the distribution specified by (2), the log likelihood is the sum of logarithmic densities over the observations. In addition, the likelihood has a term for the ground hits (see Mehtätalo and Nyblom 2009, 2010),

$$\ell(\boldsymbol{\theta},\xi) = \sum_{j=1}^{M} I(z_j > 0) \ln g(z_j|\boldsymbol{\theta},\xi) + M_0 \ln G(0|\boldsymbol{\theta},\xi), \quad (4)$$

,where M_0 is the number of hits ground hits. The assumption of independence is likely violated because there are several observations per tree in the dataset. Nevertheless, we keep calling $\ell(\boldsymbol{\theta},\xi)$ the likelihood.

Estimation of parameter ξ or $\boldsymbol{\theta}$ is based on maximizing the likelihood $\ell(\boldsymbol{\theta},\xi)$ with respect to the parameter of interest. The variance-covariance matrix of the estimation errors can be approximated by the inverse of the negative Hessian matrix at the solution (e.g. Casella and Berger, 2002). However, the variances are expected to be underestimates due to the lack of independence among the observations (Mehtätalo and Nyblom 2009).

2.3 The assumed functions for crown shape and height distribution

The height distribution of trees was assumed to be of the Weibull form

$$F(h|\xi) = 1 - \exp\left\{-\left(\frac{h}{\beta}\right)^{\alpha}\right\}, \quad (5)$$

where $\xi = [\alpha \quad \beta]'$.

The height of a tree that has radius *r* at height *z* above ground was assumed to be of the form

$$w(z,r|\boldsymbol{\theta}) = \frac{-ry_0 - \frac{b^2}{a^2}z^*x_0 + \frac{b}{a}\sqrt{b^2z^{*2} + a^2r^2 - (x_0r - y_0z^*)^2}}{b^2 - \frac{b^2}{a^2}x_0^2 - y_0^2} \quad (6)$$

where

$$a = \sqrt{b^2\frac{(1-x_0)^2}{b^2-y_0^2}} \text{ and } z^* = \max\left(z, \frac{x_0r}{y_0+b}\right), \; a,b>0,\; x_0<1,\; y_0\leq 0.$$

This function is based on assuming the crown shape of a tree of height *H* to be an ellipsoid that is centered at (x_0H, y_0H) and has the half axes *aH* and *bH* (see Figure 1); this function has been found to fit well to old-

growth Norway spruce and Scots pine trees in Finland (unpublished). However, requiring that the ellipsoid passes through the tree top gives a condition, which was used to eliminate parameter *a*. The function *w* above results from solving this equation of ellipse for tree height *H*. The function *w* has three parameters: the relative height of the maximum radius (x_0), the relative maximum width of the crown (y_0+*b*) and a parameter controlling the shape of the crown (*b*).

To be specific, the ellipsoid shape is assumed only for the part above the maximum crown width. Below that height the crown is assumed to be cylinder with the diameter equal to the maximum crown width. This assumption is realistic because this is how a solid crown looks like when seen from above.

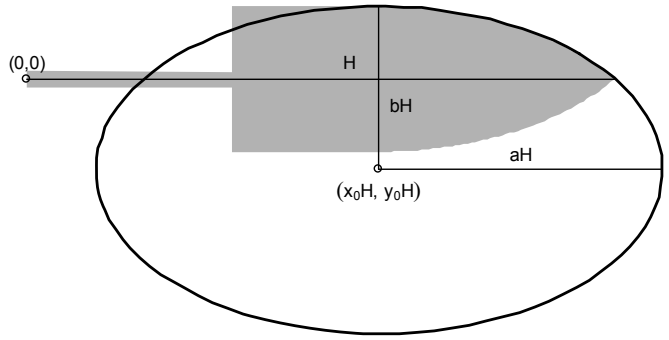


Figure 1. *Graphical representation of the applied function of crown shape. The shaded object demonstrates a fallen tree of height H.*

2.4 The estimation procedure

The estimation procedure had the following five steps.

1. The Weibull distribution was fitted to the known heights to get the estimate $\hat{\xi}$ for each of the 18 training plots.
2. The estimates of ξ from step 1 were used in the likelihood (3), which was maximized with respect to the crown shape parameter vector $\boldsymbol{\theta}$. This resulted in estimates of vector $\boldsymbol{\theta} = [x_0, y_0, b]^t$ for each training plot.
3. The plot-specific estimates of the crown shape parameters were explored with respect to the mean height of the canopy hits, and a regression was fitted to explain the observed trend.
4. The fitted regressions were applied to the plots of the evaluation data to predict the parameters $\boldsymbol{\theta}$.
5. The predictions of $\boldsymbol{\theta}$ from step 4 were used in the likelihood (3), which was maximized with respect to the Weibull parameters $\xi = [\alpha, \beta]^t$. This resulted in an estimated height distribution for each plot of the evaluation data.

In a pairwise fitting approach, the steps 3 and 4 above were omitted. Instead, the estimated crown shape of the pair of the plot under consideration was used in step 5.

For comparison, also the empirical area-based method was implemented for the dataset. In that implementation, the model of Packalen et al (2010) was fitted for both the dominant and mean heights in the training dataset of this study.

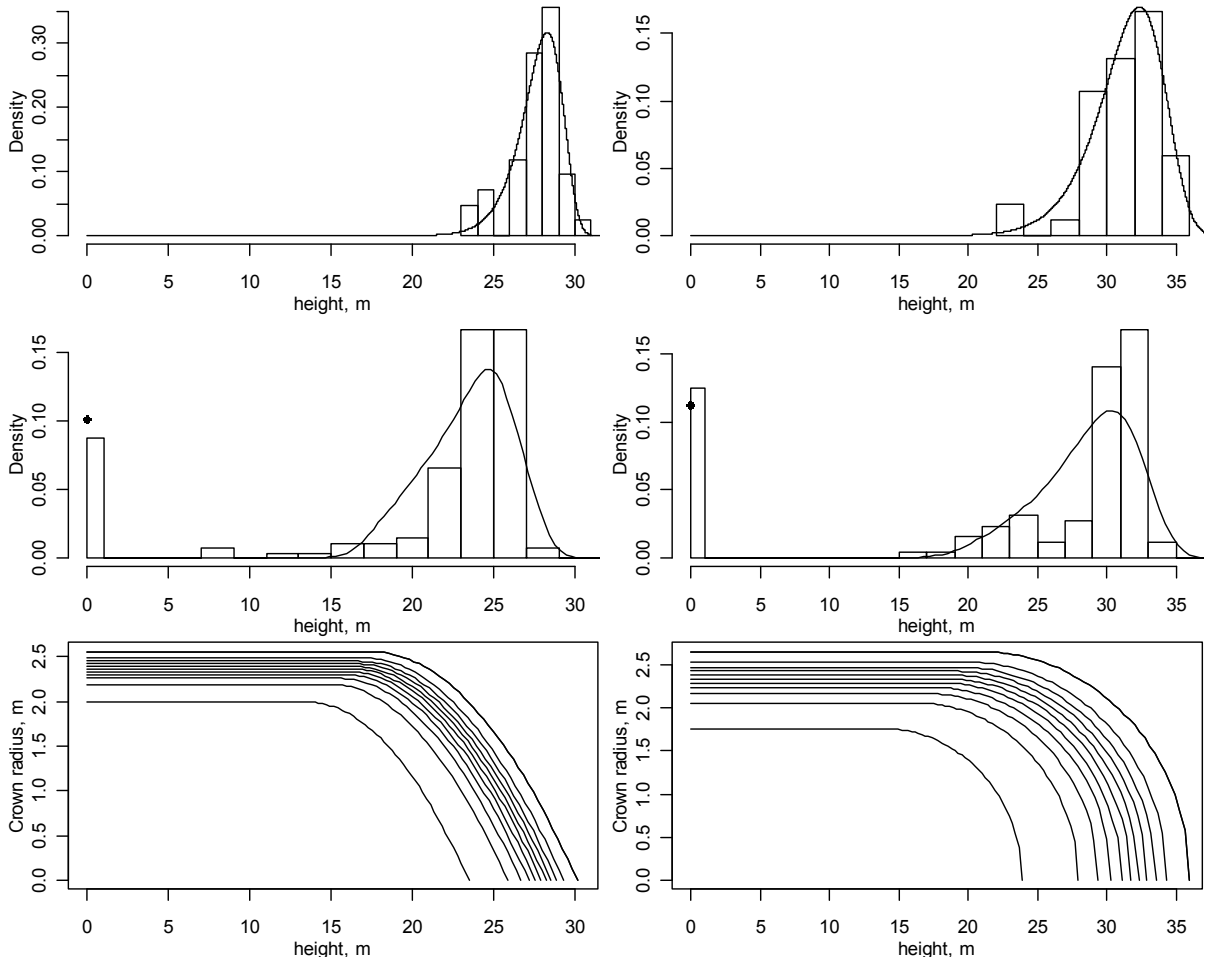


Figure 2. Examples of the fit into the modeling data on two plots. The upper graphs show the true tree heights (histogram), and the fitted Weibull density (line). The middle graphs show observed distribution of Z (histogram), and the fitted density $g(z|\theta,\xi)$ (line) and $G(0|\theta,\xi)$ (point) of step 2 of the estimation procedure. The lowest graphs demonstrate the estimated crown profiles for trees with heights at the 0.01th , 0.1st ,0.2nd , …, 0.9th , and 0.99th quantiles of the fitted Weibull height distribution.

2.5. Evaluating the methods

To evaluate the estimated height distributions, estimated and true mean and dominant heights (mean height of 100 tallest trees per ha) were computed for the evaluation plots. The true values were based on the true trees of the plot.

The estimated values for the model-based approach were based on numerical evaluation of the expected value of the height distribution of all and dominant trees. 95 % confidence intervals for these estimates were computed by using a Monte Carlo approach. In that approach, the approximate asymptotic estimation errors

of the plot-specific Weibull parameter estimates were used to generate 1000 realizations of the Weibull parameters for each evaluation plot. The mean and dominant heights were numerically evaluated for each realization. The interval between the 25th smallest and 25th largest values of the 1000 H and H_{dom} values was then taken as the 95% confidence limit for the plot.

The predictions of the empirical area-based approach were obtained by direct application of the fitted regression models into the evaluation data. Confidence intervals could not be computed for this approach.

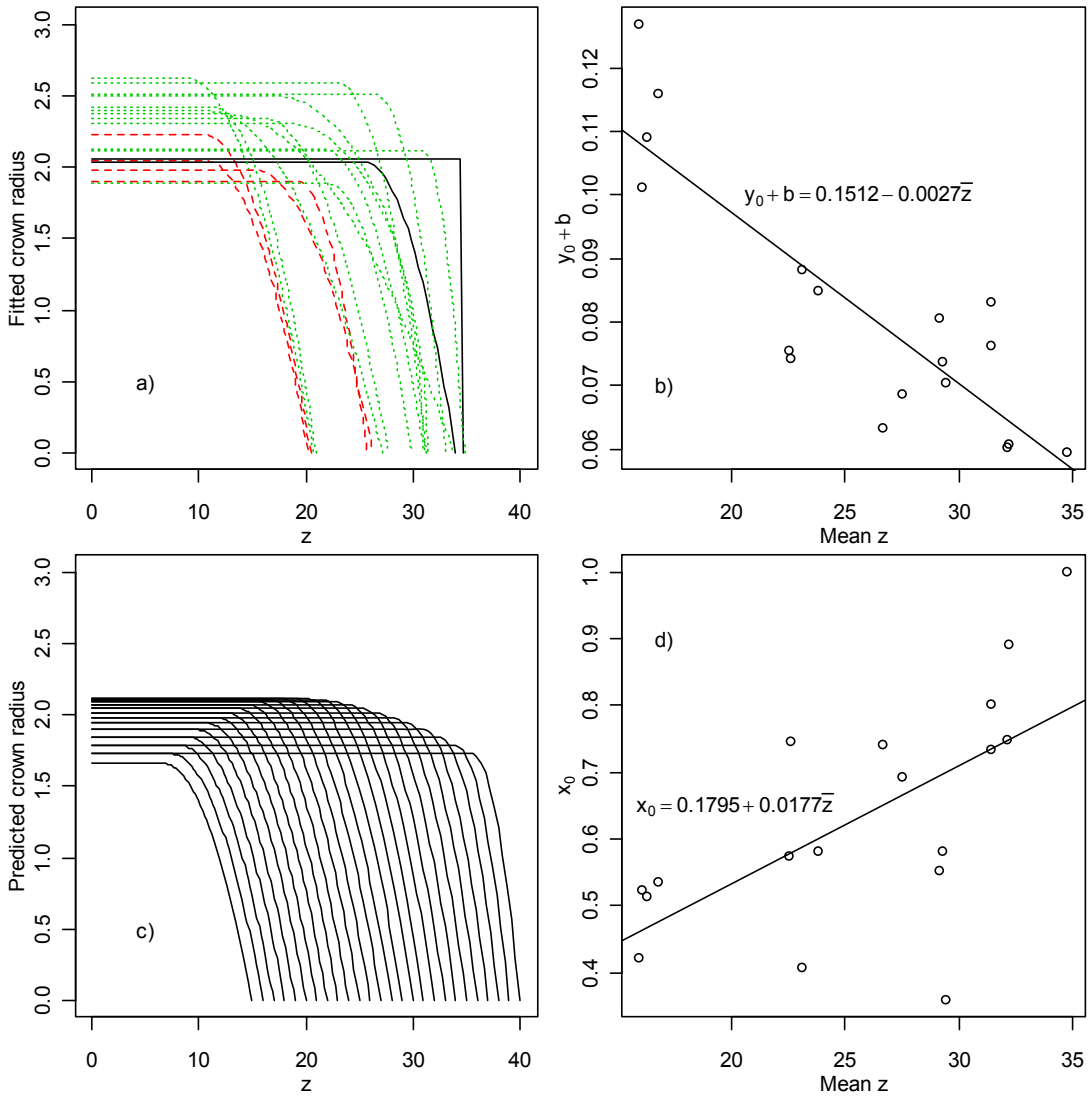


Figure 3. Demonstration for the modeling of individual tree crown shape. Graph (a) shows the estimated crown profiles for a tree of a typical height on each plot of the training dataset. Graphs (b) and (d) show the estimated values for the parameters of a crown shape as a function of average nonzero z- values of the plot, and trendlines fitted to these data. Graph (c) demonstrates the predicted crown profiles when average nonzero z- value varies from 15 to 40 meters.

3. Results

Figure 2 shows examples of the fitted Weibull diameter distribution on two sample plots of the training dataset. The Weibull function fits quite well to the tree heights, but the resulting fit of function (3) is not that good. The bottom graphs show that the estimated crown shape may be very different on different plots. This can be seen also in Figure 3(a). Figures 3(b) and 3(d) show that the relative crown width and length are smaller on plots of tall trees than on plots of short trees. Regression models explaining these trends are shown in Figures 3(b) and 3(d). Predictions of these models in Figure 3 (c) show that the absolute shape of

tree crowns does not much vary among plots of different mean canopy height, but the crown is at a higher level on plots of larger trees.

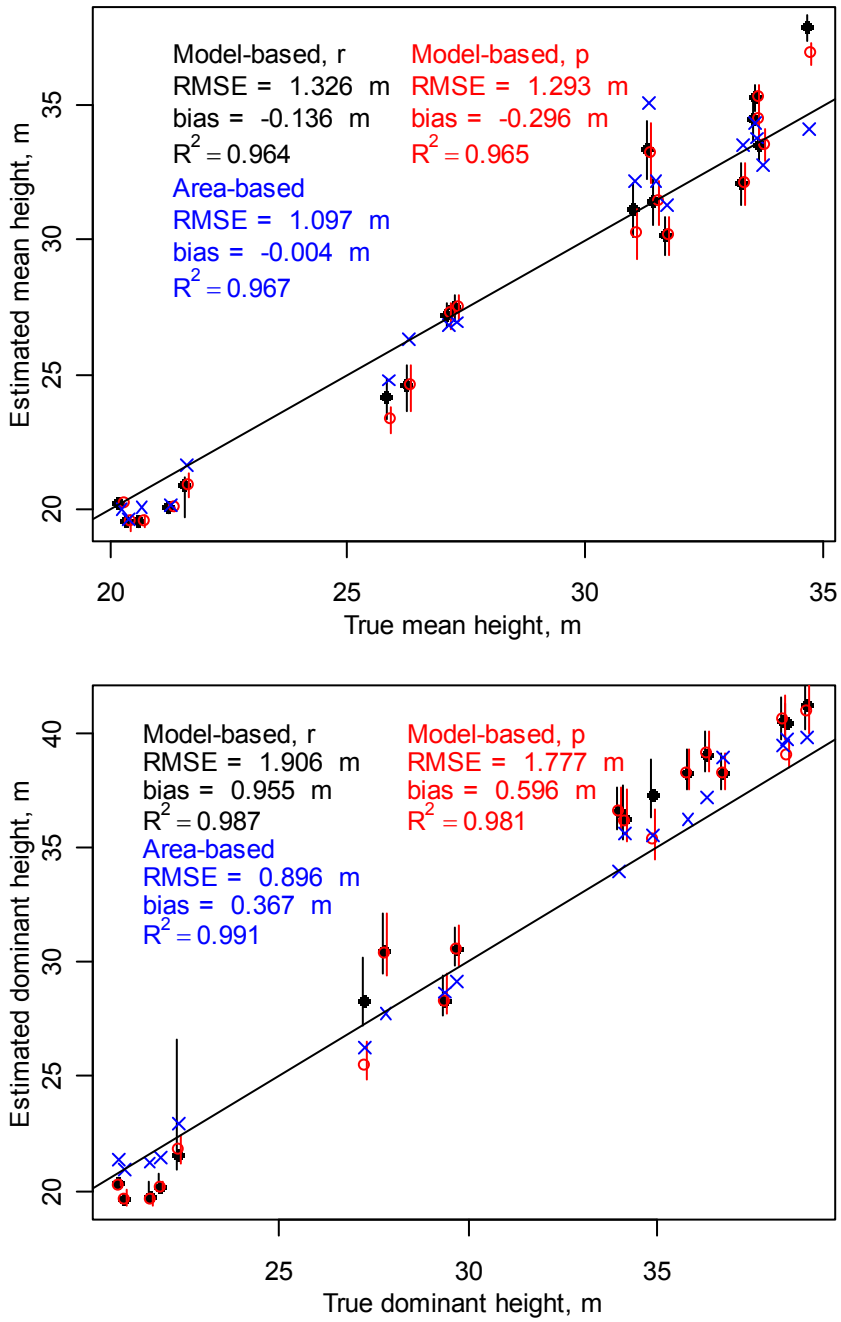


Figure 4. The estimated mean and dominant height of the validation plots on the values based on ground measurements and their confidence intervals. The predictions are based on the model-based method with predicted crown shape (black solid), the model-based method with crown shape of the corresponding pair of the training dataset (red open), and on the empirical area-based approach (blue cross).

Figure 4 shows the estimated mean heights (top) and dominant heights (bottom) for the evaluation plots using the three alternative methods. The RMSE of mean height was 1.33 m for the model based regression approach, and 1.29 m for the model-based pairwise approach. These values were clearly higher than the RMSE of the empirical area-based approach. Similar, but even stronger differences were observed in the case of dominant height. The mean and dominant heights were underestimated for plots of small trees. For

the plots of large trees, no noticeable bias can be observed for the mean height, but the dominant height was overestimated. The dominant height had quite strong positive bias.

Figure 5 shows examples of the fits from steps 4 and 5 for typical plots of small (left), medium sized (middle) and large trees (right). Function (3) does not fit very well to any of the plots. Especially, the fitted proportion of ground hits is much higher than the observed value. The estimated distribution of tree heights seems, however, good for the plot of middle-sized trees. For the plot of large trees, the height distribution has a heavy right tail. For the plot of small trees, the maximum predicted height is equal to the minimum of the measured heights, because all laser observations on this plot (middle left graph) were below the minimum measured tree height (bottom left graph). These observations largely explain our earlier notes on Figure 4.

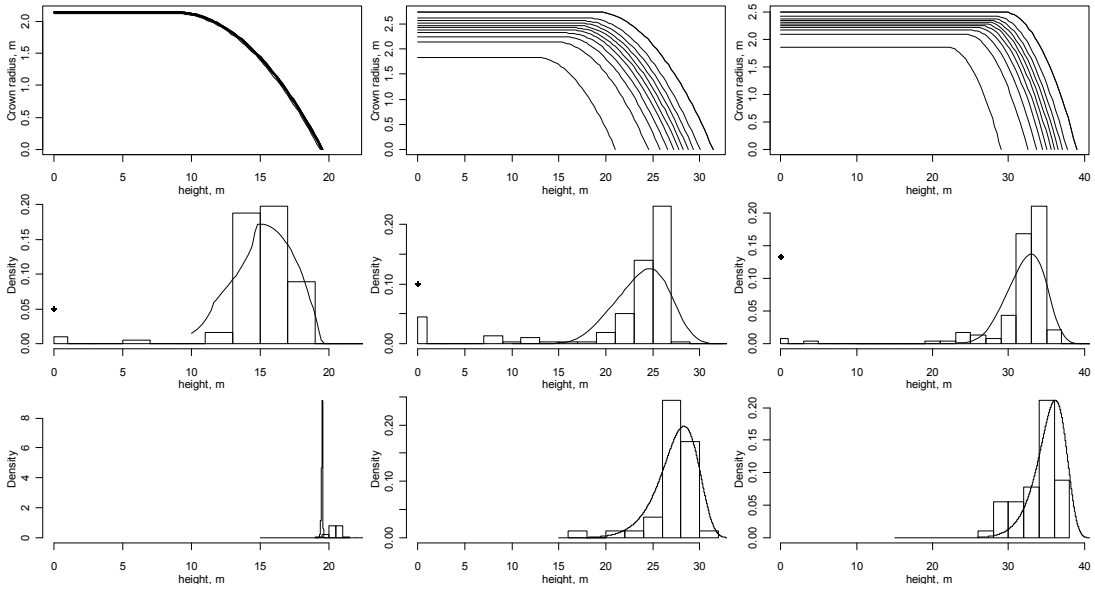


Figure 5. Example fits of step 5 of the estimation procedure for typical stands having small (left), middle-sized (middle), and large (right) trees. The upper graph shows the predicted crown shape for different tree heights, the middle graphs laser observations (histogram) and the fitted density $g(z|\theta,\xi)$ *(line) and* $G(0|\theta,\xi)$ *(point), and the bottom graphs true tree heights (histogram) and the estimated Weibull height distribution (line).*

4. Discussion

This paper developed and evaluated a new, model-based approach for the ALS forest inventory. The approach is based on a model that expresses the distribution of canopy height for a given stand density, height distribution, spatial pattern of tree locations, and individual tree crown shape. Detailed description of the model can be found in Mehtätalo and Nyblom (2009, 2010). In this study, the model was generalized for a rectangular spatial pattern. A special feature of the applied situation was that the stand density was known in advance. In addition, this was the first trial to use the model to estimate the parameters of individual tree crown shape.

The results were not as good as one could have expected. We believe that the main reason is an unrealistic assumed model for the individual tree crown. Especially, we assumed that the tree crown has a solid top surface. However, the laser return seems practically always to return from the inner parts of the crown, because the returning energy needs to exceed some limit before it is interpreted as a return (Gatziolis et al 2010). This is most likely the reason for the observed discrepancy between the true tree heights and laser observations in the left graph of figure 5. This effect should be included in our model, for example, by introducing an additional parameter for the penetration into the model individual tree crown.

Another problem was the assumption of similar relative shape among the trees. This assumption was found unrealistic. We relaxed this assumption by allowing different relative shape for different plots. However, a better approach could be a model that allows variation in relative shape also within a plot. In addition, alternative functions could be tested for the crown shape (e.g., Rautiainen et al, 2007).

The proposed method is theoretically justified because the parameters have interpretations arising from the stand structure and tree crown properties. In addition, the applied method of maximum likelihood has a strong theoretical basis, providing justified means to construct confidence intervals for the estimates, and to make likelihood-based tests between different models. However, these intervals may be too narrow because of the lack of independence among observations. In addition, they are valid only if the assumed model is correct. In this study, this was evidently not true, leading to too narrow confidence intervals for the mean and dominant heights.

The definition of the distribution function (Eq. 2) includes an integral over a two-dimensional plane. This function is differentiated with respect to *z* and the derivative evaluated repeatedly at each observation to get the likelihood. For asymptotic inference, the likelihood needs still to be differentiated twice with respect to the parameters. This computational burden currently restricts the applicability of the method. This is also the main reason for that we did not search for the best possible models for tree canopies in this study. Instead, we just went through the chain of computations to demonstrate how this estimation could be done. Comparison and evaluation of the model with different assumed models is left to the future. In addition, possible approximations are searched for to decrease the computational requirements of the method.

References

CASELLA, G. and BERGER, R.L. 2002. Statistical inference. 2ndedition. (Pacific Grove: Duxbury).

gatziolis, D., fried j.s., and monleon, v.s. 2010. Challenges to estimating tree height via LiDAR in closed-canopy forests: a parable from Western Oregon. Forest Science, 56(2), pp. 139–155.

mehtäTALO, I AND NYBLOM, J. 2009. Estimating forest attributes using observations of canopy height: a model-based approach. Forest Science, 55(5), pp. 411–422.

mehtäTALO, I AND NYBLOM, J. 2010. A model-based approach for ALS inventory: application for square grid spatial pattern. Submitted MS.

Næsset, E., GOBAKKEN, T., HOLMGREN, J, HYYPPÄ, H:, HYYPPÄ, J., MALTAMO, M, NILSSON M., OLSSON, H., PERSSON, Å. and SÖDERMAN, U. 2004. Laser scanning of forest resources: the Nordic experience. Scandinavian Journal of Forest Research, 19, pp. 482–499.

PACKALÉN, P., MEHTÄTALO, L. and MALTAMO, M., 2010. ALS based estimation of plot volume and site index in a eucalyptus plantation with a nonlinear mixed effect model that accounts for the clone effect. Proceedings of Silvilaser 2010, the 10th International Conference on LiDAR Applications for Assessing Forest Ecosystems. September 14th - 17th, 2010, Freiburg, Germany.

Rautianen, m., mõttus, m., stenberg, p. and ervasti, s. 2007. Crown envelope shape measurements and models. Silva Fennica, 42(1), pp. 19–33.

VAUHKONEN, J., ENE, L, GUPTA, S., HEINZEL, J., HOLMGREN, J., PITKÄNEN, J., SOLBERG, S., WANG, Y., WEINACKER, H., HAUGLIN, K.M., LIEN, V., PACKALÉN, P. GOBAKKEN, T., KOCH, B., NÆSSET, E., TOKOLA, T. and MALTAMO, M., 2010 Comparative testing of single-tree detection algorithms. Proceedings of Silvilaser 2010, the 10th International Conference on LiDAR Applications for Assessing Forest Ecosystems. September 14th - 17th, 2010, Freiburg, Germany.

An evaluation of the field sampling design of the first operational LiDAR based site quality survey of radiata pine plantations in South Australia.

JAN Rombouts†‡, IAN Ferguson†, jERRY Leech§ & dARIUS cULVENOR¤*

rombouts.jan@forestrysa.com.au

†Department of Forest and Ecosystem Science, University of Melbourne, Vic 3010, Australia;

‡ForestrySA, PO Box 162 Mt Gambier, SA 5290, Australia;

§Forestry Systems, Mt Gambier, Australia;

¤CSIRO, Sustainable Ecosystems, Clayton, Vic, Australia

Abstract

The first operational LiDAR based site quality survey of radiata pine plantations was conducted in South Australia in 2009. All nine and ten year old plantations, 28 sites in all, covering 9,365 hectare scattered across an area of 10,000 km^2, were assessed. Area-based methods were applied and field data were collected to calibrate new prediction models for stand volume (i.e. the criterion of site quality). Decisions with regard to the field sampling design were to some extent guided by prior research, but were conservative so as to ensure robustness of the results. This study evaluated two key aspects of the sampling design – sampling intensity and sample selection method. It also posed the question whether the decision not to re-use pre-existing data and models was justified in light of the survey results. Firstly the precision of prediction models fitted to three independent datasets of different size were compared. Secondly a simulation experiment was conducted to test the sensitivity of model precision to a range of sample sizes and sample selection methods (random, stratified random and systematic sampling). This experiment was run with and without inclusion of pre-existing data. Results indicated that the sampling intensity applied in the survey could have been significantly reduced without significant loss of model accuracy. Models were only sensitive to the sample selection method when samples were small. The new prediction models proved to be significantly different from any pre-existing models. However, simulation results suggested that when new data are scarce multi-campaign models calibrated using new and pre-existing data may offer increased prediction accuracy. Given the high cost of field data collection these findings have practical importance for future surveys.

Keywords: site quality, radiata pine, LiDAR, model calibration, sampling design

1. Introduction

For the past sixty years site quality information has underpinned many aspects of South Australian radiata pine plantation management including yield regulation (growth modelling and inventory), silviculture (thinning and fertiliser regimes) and monitoring of long-term land use sustainability (Lewis *et al.*, 1976, O'Hehir and Nambiar, 2010).

The South Australian site quality system applies total volume production to a small-end underbark diameter of 10 cm at reference age (9 or 10 years) as the criterion of site productivity. Site quality assessment is in essence a process of mapping the spatial variation of stand volume within forest compartments. Trials commenced in 2002 demonstrated the feasibility of LiDAR based site quality assessment (Rombouts, 2006, Rombouts *et al.*, 2008, Rombouts *et al.*, 2010). In late 2007 two South Australian radiata pine growers ForestrySA (FSA) and Green Triangle Forest Products (GTFP) decided to adopt LiDAR based site quality assessment operationally and scheduled a first LiDAR site quality survey for 2009.

The trials conducted up to that point had demonstrated the effectiveness of area-based methods for site quality assessment. Good results had been obtained using LiDAR data of low density (0.3 pulses m^{-2}) captured at relatively high flying altitude (2800 m). These findings were instrumental in selecting operational

LiDAR data capture parameters for the survey. The trials also showed that LiDAR prediction models were sensitive to “campaign” effects, where the factor campaign referred to the complex of LiDAR data capture parameters, LiDAR instruments, service providers, as well as transient environmental factors (weather). This prompted the decision to ignore any pre-existing models and to collect the necessary field data for calibration of new models.

The objective of field data collection was to provide adequate data to enable fitting of model(s) representative of the true relationships between LiDAR and forest variables across the survey area. Several studies had evaluated the effect of field sample size on the accuracy of predictions (Juntilla *et al.*, 2008, Gobakken and Næsset, 2008, Maltamo *et al.*, 2009) but the results of studies in mixed boreal forest could not be directly applied to even-aged radiata pine plantations. Earlier trials had produced some evidence (albeit inconclusive) that LiDAR prediction models were site specific. To err on the side of caution it was decided that each site had to be represented in the sample. It was conservatively estimated that 6-10 field plots per site would yield sufficient data to fit robust site specific models. Inspired by (Schreuder *et al.*, 1990) it was also decided to apply a stratified random sampling approach, whereby stratification was based on the distribution of the LiDAR predictor variable across the site.

The aim of this paper was to review these key sampling design decisions in light of the survey outcomes. The focus thereby was on the properties of the LiDAR stand volume prediction models calibrated using the new LiDAR and field data. Were the prediction models sensitive to sample size and/or sample selection method? Was site variability indeed significant? Could any productive use have been made of pre-existing data?

2. Methods

2.1 Sites

Figure 1 shows the geographic distribution of the 1999, 2000 and 2001 plantations included in the survey. Sampling of the 2001 plantations was deliberately deferred by one year so that measurement could take place at reference age nine. The evaluation in this paper therefore only pertains to the 1999-2000 plantations (20 FSA sites and 8 GTFP sites). The sites cover a wide range of growth conditions (rainfall and soil type) present in the so-called Green Triangle of SE South Australia and SW Victoria. Figure 1 also shows the nine trial sites established in 2002, 2006 and 2007. See Rombouts *et al*. (2010) for more specifics.

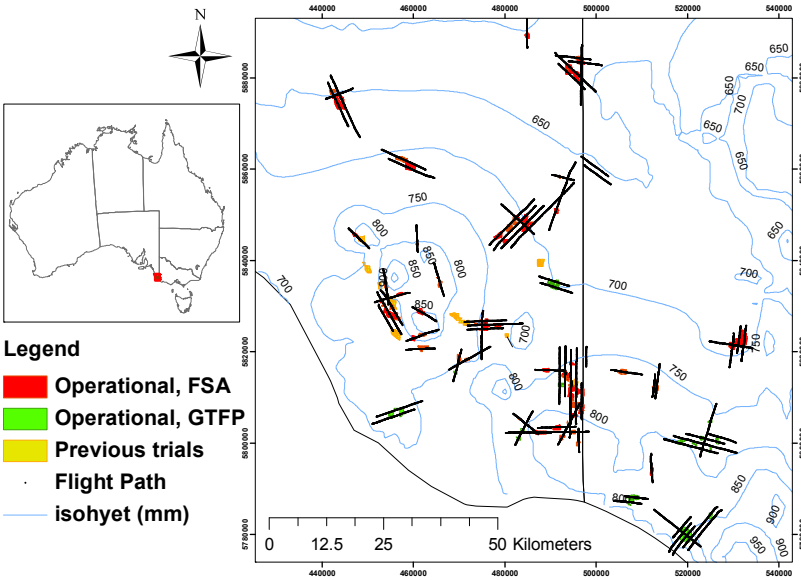


Figure 8: Operational and trial sites. Dotted lines represent flightpaths (operational survey only).

2.2 LiDAR data

Table 1 shows LiDAR system and data capture parameters of operational and trial campaigns. Figure 1 shows flight paths (operational survey only). Campaigns 2007LD and 2007HD covered the same five sites but were flown at different altitude. The LiDAR data sets were used as provided by the supplier, including the classification of LiDAR returns into ground or non-ground points using TerraScan and proprietary systems. The ground points provided the basis for the construction of a Digital Terrain Model (DTM) using an ESRI

ArcGIS™ implementation of Delaunay triangulation. The height of LiDAR points above ground level was calculated as the difference between a point's z value and the z value of its projection on the DTM.

Candidate LiDAR predictor variables were derived from the height distribution of LiDAR returns observed in either field plots or grid cells created by overlaying a 25x20 m rectangular grid over the areas under assessment. Only first return LiDAR data were used (i.e. the first recorded echo of a pulse). Data located outside areas mapped as effective plantation were removed. Computed LiDAR variables were:

- mean height ($\overline{h}$), mean quadratic height ($\overline{h}_q$), standard deviation of heights (s_h), maximum height ($h_{\max}$), average of the maximum height in each plot quadrant ($h_{\max 4}$)
- percentile heights of the ordered cumulative height distribution (h_0, h_{10},…,h_{90}), proportion of returns in height frequency distribution classes (d_0, d_{10},…, d_{90}).
- proportion of ground returns (p_g), point density as first returns m^{-2} (n_f), scanning angle approximated as scanning angle to plot centroid (α_{scan}).

Table 12: LiDAR system details and flight parameters of operational and trial campaigns

	Operational	Trials			
	2009	2002	2006	2007 LD	2007HD
System	ALTM Gemini	ALTM 3025	ALS 50	ALTM 3100	ALTM 3100
Service provider	SP1	SP1	SP2	SP1	SP1
Date	20/12/2008	7/07/2002	9/04/2006	20/08/2007	20/07/2007
Flying altitude (m asl)	2,500	1,100	1,040	2,800	1,100
Pulse Frequency Rate (Hz)	50,000	25,000	73,000	33,000	33,000
Max scanning angle (dgrs)	15.0	15.0	13.5	16.0	12.5
Beam divergence (mrad)	0.25	0.20	0.22	0.20	0.20
Footprint diam. (cm)	50	22	23	56	22
Mean pulse density (m^{-2}) [range]	0.4 [0.2-1.5]	0.8 [0.5-2.1]	2.7 [1.2-9.5]	0.3 [0.2-0.4]	2.6 [2.3-3.2]
Returns per pulse	1-4	1st/last	1-4	1-4	1-4

2.3 Field data

Table 2 provides details of field data collected during the operational survey and earlier trials. Operational FSA and GTFP field plot locations were selected independently at each site after stratification based on the LiDAR predictor variable $\overline{h}_q$. Stratification involved overlaying a 25x20m grid over the areas; computing $\overline{h}_q$ in the 25x20m cells; ranking of cells from lowest to highest $\overline{h}_q$ and subdivision of the ranked set into n groups with approximately equal numbers of cells, with n equal to the number of plots to be established at the site. One cell was randomly selected from each group, providing the plot locations. Cells located on the plantation boundary were rejected because they often had low point counts. The 31 FSA-validation plots were randomly selected. The trial plots were purposively selected to cover soil types, age classes and other features of interest (see Rombouts *et al.* (2010)).

Table 2: Field data collected during operational survey and trials

	Operational			Trials		
	FSA	FSA-validation	GTFP	2002	2006	2007
nr. plots	142	31	50	28	79	62
nr. sites	20	20	8	1	3	5
age	9-10	9-10	9-10	10	7-11	9-10
V_{10} - mean	150.9	147.8	130.5	159.2	167.4	140.2
V_{10} - range	36.7-328.1	47.2-294.2	4.5-301.8	35.2-242.8	35.9-303.3	34.4-304.1

All field plots were rectangular (25x20m). Plots were measured either by FSA or GTFP field crews. All tree stems were measured for diameter at breast height (1.3m). Plot Predominant Height (PDH) was estimated based on the heights of the largest tree in each of the four plot quadrants. Volume to 10 cm underbark small-end diameter (V_{10}) was predicted using a stand volume equation with plot basal area, stocking and PDH as predictor variables. For more detail on plot measurement procedures refer to Rombouts *et al.* (2010).

2.4 Comparison of prediction models fitted to data sets of different size

Earlier work had demonstrated the effectiveness of single variable, linear models, fitted using Ordinary Least Squares (OLS) or mixed effect maximum likelihood (ME) regression techniques as predictors of stand volume:

OLS: $$V_{10,ij} = \beta_0 + \beta_1 \overline{h}_{q,ij} + \varepsilon_{ij} \quad (1)$$

ME: $$V_{10,ij} = \beta_0 + b_{0,j} + (\beta_1 + b_{1,j})\overline{h}_{q,ij} + \varepsilon_{ij} \quad (2)$$

where $b_{0i} \sim N(0, \sigma_0^2)$ $b_{1i} \sim N(0, \sigma_1^2)$ $\varepsilon_{ij} \sim N(0, \sigma^2)$

with $V_{10,ij}$ and $\overline{h}_{q,ij}$ volume and mean quadratic height of plot i at site j; β_0 and β_1 fixed effect parameters to be estimated; $b_{0,j}$ and $b_{1,j}$ random effects for site j to be estimated.

Models were fitted to three independent datasets of decreasing size using equations (1) and (2). These data sets were: FSA dataset with 142 plots, the GTFP dataset with 50 plots and the FSA-validation dataset with 31 plots (see Table 2). After confirming that $\overline{h}_q$ was the preferred predictor variable - alternative (combinations of) predictor variables were tested - the three models were compared graphically and statistically. Each of the three models were applied to each of the three datasets to compare bias and RMSEs of predictions. In the case where the model was applied to its own calibration dataset bias and RMSE were computed using a leave-one-out cross validation approach.

2.5 Simulation of alternative sample sizes and sample selection schemes

A resampling experiment was conducted to simulate the effect on prediction models of alternative sample sizes and sample selection schemes. The objectives of this analysis were similar to those of a study by Maltamo *et al.* (2009) which examined different sampling strategies for field data collection in the context of airborne laser scanning forest inventory. However, the sample selection schemes, the modelling and resampling approach applied were substantially different.

The resampling dataset was created by pooling the FSA, GTFP and FSA-validation data sets (223 plots). The resampling approach was as follows:

for each of *i* iterations

- randomly draw a validation plot from the data set
- draw a sample of size *n* from the remaining data (222 plots) using one of five sample selection methods
  - fit OLS and ME models (equations 1 and 2) to the data
  - calculate prediction error based on validation plot: $e_{sample,i} = V_{10,i} - \hat{V}_{10,i}$

Five sample selection methods were simulated:

- simple random sampling (**RS**)
- systematic sampling (**SS**)
- stratified random sampling, with stratification based on covariate $\overline{h}_q$ (**SRHq**)
- stratified random sampling, with stratification based on site, with each site contributing the same number of plots (**SRSite**)
- stratified random sampling with stratification based on site, with each site contributing the same number of plots but with concentration of the sample in 10 (out of 28) randomly selected sites (**SR10Sites**)

The range of sample sizes tested was between 5 and 84 and was constrained by the size of the simulation dataset and sample selection scheme.

Systematic sampling was implemented by:

- ranking the 222 plots based on the predictor variable $\overline{h}_q$
- randomly selecting a "seed" plot
- selecting every k^{th} plot, starting from the seed plot and working up and down the ranking (with k=222/sample size).

Strata under the SRHq sampling scheme were formed by ranking the 223 plots based on the predictor variable $\overline{h}_q$ and creating 5 $\overline{h}_q$ intervals comprising equal numbers of plots (three strata had 43 plots and the other two strata had 44).

The number of iterations was set at 30,000 after confirming that this ensured stable results for the various combinations of sample selection scheme and sample size. Average bias and Root Mean Square Error (RMSE) were computed based on the 30,000 observed prediction errors $e_{sample,i}$

2.6 Augmenting the calibration dataset with pre-existing data

Earlier work had shown that campaign effects could be effectively accounted for using models with campaign dummy variables. Confirmation was sought that this was still the case for the dataset created by merging the 2002, 2006, 2007LD, 2007HD and 2009 data sets. A model was fitted with following structure (for conciseness only the OLS model is shown) :

$$\text{OLS:} \qquad V_{10,ij} = \beta_0 + (\beta_1 + \beta_2 C_{2006} + \beta_3 C_{2007LD} + \beta_4 C_{2007HD} + \beta_5 C_{2009})\overline{h}_{q,ij} + \varepsilon_{ij} \qquad (3)$$

with $V_{10,ij}$ and $\overline{h}_{q,ij}$ as defined above; C_{2006}, C_{2007LD}, C_{2007HD} and C_{2009} campaign dummy variables; $\beta_0,\ldots,\beta_5$ fixed effect parameters to be estimated and the residuals ε_{ij} normally distributed with mean zero.

A second simulation experiment was conducted to test the merit of adding pre-existing data to the calibration dataset while using a model with structure such as equation (3). To this end the method described in section 2.5 was changed in two ways: (i) to each sample drawn the data of campaigns 2002, 2006 and 2007HD (but not 2007LD) were added and (ii) instead of equations (1) and (2), equation (3), or its mixed-effect equivalent, were used.

3. Results

OLS and ME models were fitted to three independent data sets of decreasing size. The preferred predictor variable was confirmed to be $\overline{h}_q$. Additional predictor variables or variable transformations did not materially improve model fit. Dummy variables analysis and F-tests showed that the slopes and intercepts of the three models were not significantly different.

Table 3: Comparing predictive accuracy (bias and RMSE) of three models fitted to three independent data sets. The values in italics are based on leave-one-out cross-validation (see 2.4).

Model applied		Data set the model was applied to					
		FSA		FSA validation		GTFP	
		%bias	%RMSE	%bias	%RMSE	%bias	%RMSE
FSA	OLS	*0.0*	*10.9*	-1.3	11.2	0.4	13.5
(n=142)	ME	*0.1*	*10.7*	0.0	10.1	-1.7	13.5
FSA validation	OLS	1.2	10.9	*0.0*	*11.6*	2.7	14.3
(n=31)	ME	1.5	11.0	*0.4*	*12.2*	3.1	14.6
GTFP	OLS	0.3	10.8	-1.0	11.6	*0.1*	*14.3*
(n=50)	ME	0.3	10.8	-1.0	11.6	*0.2*	*14.9*

Table 3 displays bias and RMSE of predictions calculated when models were applied to each of the three datasets. Models calibrated using smaller plot samples did not consistently produce poorer predictions. Taking the FSA dataset as the most significant example: the RMSEs recorded for the FSA, FSA-validation and GTFP models were 10.9%, 10.9% and 10.8% respectively. Bias typically increased as models were applied outside their calibration dataset, but not to unacceptable levels. Mixed effect models were not consistently more accurate indicating that site effects were weak or non-existent.

Figure 2 shows the result of the first simulation experiment using OLS models. RMSE was expressed as a percentage of average plot volume (145.9 m^3 ha^{-1}). For sample sizes above 30-40 the %RMSE was hardly influenced by the sample selection method. For sample sizes below 10 random sampling stood out as the least effective sampling method. Reducing the sample size from 222 to 30 plots changed the %RMSE from 11.5 to 12%.

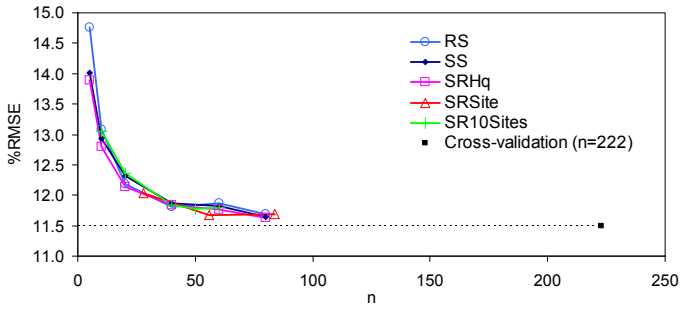


Figure 9: relative RMSE of stand volume prediction as function of sample size and selection method. Each data point represents the average of 30,000 iterations. Applied model: equation (1). The dotted line indicates the RMSE for full sample size=222 (i.e. equivalent to leave-one-out cross validation).

Bias ranged between -0.1 and 0.5% and seemed uncorrelated with sample size and selection scheme. Even the highest value was of no material concern for practical application. Results for mixed-effect models were mostly consistent with OLS models but RMSE were comparatively higher for smallest sample sizes. For conciseness these results were not shown.

Figure 3 shows equation (3) fitted to the pooled dataset (operational + trial data). Each campaign dummy variable was significantly different from zero and from every other dummy variable. The difference between the parameters for C_{2009} and C_{2007LD} was nearly 10 times the parameter standard error notwithstanding the similarity of the 2009 and 2007LD LiDAR data point density and flying height (see Table 1). The difference was even greater than between the C_{2009} and C_{2007HD} parameters. This strongly suggested that flying height and point density differences were not the main causes of the observed campaign effects. Interestingly the dummy variables increased in value each successive campaign.

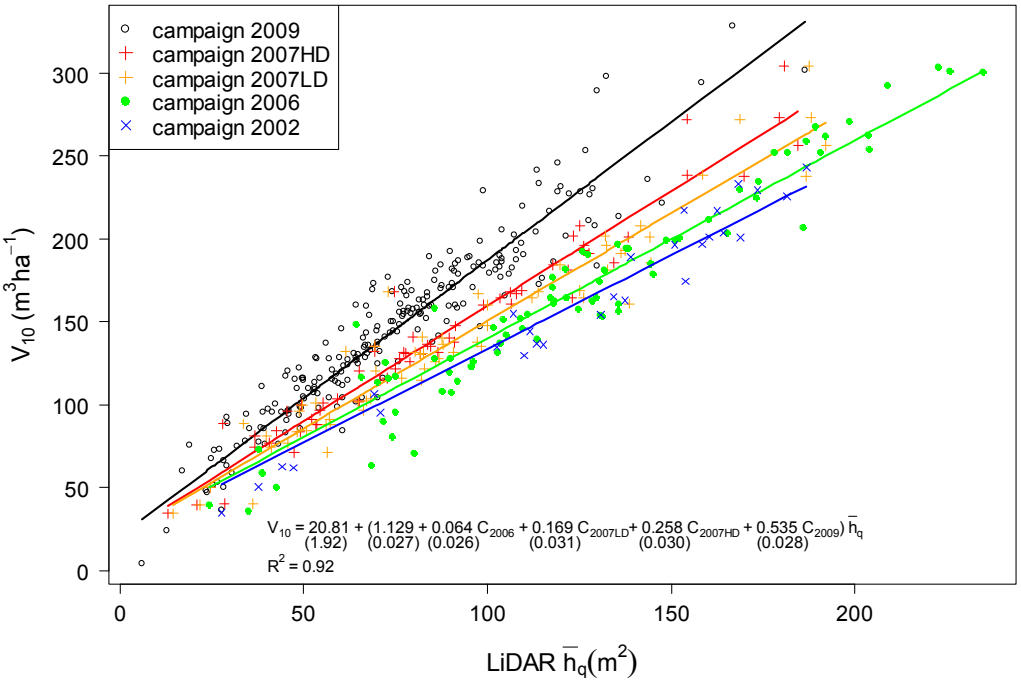


Figure 10: Model with campaign dummy variables fitted to the complete dataset (= operational + trial data)

Figure 4 shows the results of the second simulation experiment whereby pre-existing calibration data were added to the samples drawn from the 2009 dataset. Consistent with Figure 2, RMSEs were expressed relative to the average volume of the 2009 plots. Comparison of Figures 2 and 4 suggests that pooling of pre-existing data had little effect when sample sizes were sufficiently large (>40). However as sample size dropped below 40 increasingly large gains in precision were observed, most noticeably for random sampling. This may be explained by the fact that equation (3) constrains all campaign lines through an intercept estimated using both new and pre-existing data. This prevents the very large prediction errors that may occur when equation (1) and (2) models are fitted to samples with a very narrow range of the predictor variable $\overline{h}_q$ or with high leverage outliers. Such samples have a higher probability of being drawn when sample selection is random and when samples are small.

Bias was not correlated to sample size or selection method and ranged between -0.4 and -1.0%. Mixed-effect modelling results were marginally better than corresponding OLS results (results were not shown because of space constraints).

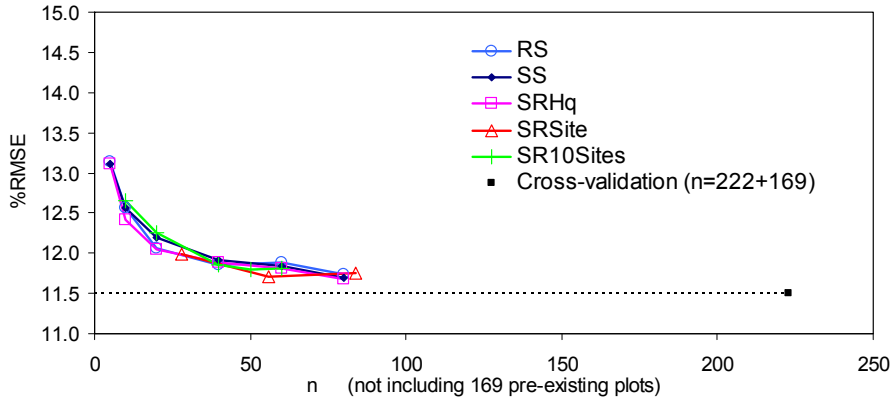


Figure 11: relative RMSE of stand volume prediction as function of sample size and selection method. Each data point represents the average of 30,000 iterations. 169 pre-existing plots were added to each drawn sample. Applied model: equation (3). The dotted line indicates the RMSE for sample size=222.

4. Discussion

Importantly the first operational LiDAR based site quality survey did not present major inconsistencies with earlier trial work. Simple linear models with a single LiDAR predictor variable ($\overline{h}_q$) produced RMSE in the order of 11% of mean stand volume, which was comparable to earlier trial results. The prediction model fitted to new LiDAR and field data proved to be very different from earlier prediction models justifying the decision to develop new models rather than re-use existing ones. Volume predictions were 8-18% higher for the same $\overline{h}_q$ than those of the most similar earlier model (Figure 3). By comparison, in a study by Næsset (2008) volume prediction bias reached up to 10.7% when prediction models were applied to LiDAR data captured using different sensors and data capture parameters.

Predictions made using models fitted to three independent datasets with 142, 50 and 31 plots respectively were of comparable accuracy and simulation experiments showed that model accuracy gains were very modest between sample sizes 30 and 222. It appears therefore that the prescribed numbers of plots (6-10 per site, 142+50 in total) could be significantly reduced without material loss in model accuracy.

The comparison of OLS and mixed effect model predictions indicated that site effects in models were too weak to be of practical significance. The simulation experiments failed to demonstrate that stratified random sampling with stratification based on site was superior to other sample selection methods (in particular stratified random sampling with stratification based on $\overline{h}_q$). Based on these results, the survey prescription to represent every site in the sample could be relaxed and travel cost, for example, could in future be taken into consideration when selecting sites for data collection. Similarly, calibration data could be pooled across regions or forest estates to take advantage of economies of scale.

Stratified random sampling, with stratification based on $\overline{h}_q$ (SRHq), consistently performed as well or better than the other sampling schemes simulated in the experiment. This mirrors the results of the simulation study by Maltamo *et al.* (2009) who found that stratification based on LiDAR variables was most effective for stand volume prediction.

The observed campaign effects are likely to be caused by a complex of interacting factors (LiDAR instruments, LiDAR data capture parameters and data processing algorithms, transient environmental effects – see Næsset (2008) and Rombouts *et al.* (2010) for background) but the available data were poorly suited for a detailed analysis of the relative contribution of each of those factors to the overall effect.

The study did confirm however that it was possible to model campaign effects quite effectively using simple campaign dummy variables. The second simulation experiment indicated that pooling of pre-existing data

and the use of multi-campaign models increased the precision of predictions, but only if the sample of new data was small (n<40). It is likely that better techniques could be developed to leverage pre-existing data than the dummy variable models used in this study (i.e. Bayesian approaches, mixed-effect models with campaign random-effects). This would appear to be a promising area of research.

5. Conclusion

The first operational survey confirmed the accuracy and feasibility of LiDAR based site quality assessment. The results showed that:

(1) Sampling intensity applied in the survey could have been significantly reduced without significant loss of model accuracy.

(2) Models were only sensitive to the sample selection method when samples were small.

(3) New prediction models were significantly different from any pre-existing models justifying the decision to develop new models.

(4) Multi-campaign models calibrated using new and pre-existing data may offer increased prediction accuracy, when new data are scarce.

This detailed review of the field sampling design identified considerable scope for optimisation of sampling intensity and sample selection prescriptions. Revised prescriptions will be applied in future surveys.

Acknowledgements

The considerable financial and logistic support of ForestrySA in conducting this study is gratefully acknowledged. The authors also thank Green Triangle Forest Products for contributing data to this study.

References

GOBAKKEN, T. & NÆSSET, E. (2008) Assessing effects of laser point density, ground sampling intensity, and field sample plot size on biophysical stand properties derived from airborne laser scanner data. Canadian Journal of Forest Research, 38, 1095-1109.

JUNTILLA, V., MALTAMO, M. & KAURANNE, T. (2008) Sparse Bayesian estimation of forest stand characteristics from Airborne Laser Scanning. Forest Science, 54, 543-552.

LEWIS, N. B., KEEVES, A. & LEECH, J. W. (1976) Yield Regulation in South Australian Pinus radiata plantations. Bulletin No 23, Woods and Forests Department, South Australia, pp 174.

MALTAMO, M., BOLLANDSAS, O. M., NÆSSET, E., GOBAKKEN, T. & PACKALEN, P. (2009) Different field sampling strategies for field training plots in ALS inventory. Proceedings of SilviLaser 2009, October 14-16 - College Station, Texas USA.

NÆSSET, E. (2008) Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. Remote Sensing of Environment, doi:10.1016/j.rse.2008.09.001.

O'HEHIR, J. F. & NAMBIAR, E. K. S. (2010) Productivity of three successive rotations of P. radiata plantations in South Australia over a century. Forest Ecology and Management, 259, 1857-1869.

ROMBOUTS, J. H. (2006) Exploring the potential of airborne LiDAR for site quality assessment of radiata pine plantations in South Australia: initial results. Paper presented to Research Working Group 2, Forest Measurement and Information Systems, Biennial meeting, 21-24 November 2006, Woodend, Australia.

ROMBOUTS, J. H., FERGUSON, I. S. & LEECH, J. W. (2008) Variability of LiDAR volume prediction models for productivity assessment of radiata pine plantations in South Australia. Conference Proceedings SilviLaser 2008, Sept. 17-19, 2008 - Edinburgh, UK.

ROMBOUTS, J. H., FERGUSON, I. S. & LEECH, J. W. (2010) Campaign and Site effects in LiDAR prediction models for Site Quality assessment of radiata pine plantations in South Australia. International Journal of Remote Sensing, 31:5, 1155-1173.

SCHREUDER, H. T., LI, H. G. & WOOD, G. B. (1990) Model-dependent and design-dependent sampling procedures - a simulation study. U.S. For. Serv. Rocky Mt. For. range Exp. Stn. Res Pap. RM-291.

Using delta values of multi-temporal first-return small footprint airborne laser scanner data to predict change of tree biomass in mountain spruce forests.

OLE MARTIN BOLLANDSÅS† and ERIK NÆSSET†*

olebo@umb.no

†Dept. of Ecology and Natural Resource Management, Norwegian University of Life Sciences, Box 5003, N-1432 Ås, Norway

Abstract

The mountain forests are likely to undergo a pronounced growth in biomass stocks because of increased mean temperatures and changed precipitation as a result of a changing climate. The biomass increase is both attributed to an upward advance of the current tree line and also an increased growth of the current forests. In order to monitor changes in such forests, which often are found in remote areas and are structurally diverse, methods based on remote sensing can be more efficient compared to traditional field based inventories. Traditional field based methods can be very costly both because of high transportation costs to get the field crew on site, but also because of the high number of field plots necessary to capture the biomass variations. This study explored the capability of multi-temporal airborne laser scanner data to estimate the change of total aboveground tree biomass in the mountain forest. More specifically we used the observed difference between first return height percentiles and density variables obtained from airborne laser scanner data on two points in time (2005=T1 and 2008=T2) to model the biomass change observed in field during the same time period by means of an ordinary least squares model. Our data covered four growth seasons and the mean change of biomass during this period was 17.5 t/ha with an initial mean biomass of 135.1 t/ha. Results showed that the differences between laser height percentiles at T1 and T2 were most correlated to the observed biomass change. The model fit as expressed by R^2 was 0.87, and the model root mean square error was 4.4 t/ha which corresponds to 25 % of the observed change in biomass.

1. Introduction

Monitoring of changes in forests is important to get relevant data on biomass changes needed for reporting on emissions and removals of CO_2. Especially for those countries that have ratified the Kyoto protocol, it is of importance in order to meet the demands for reporting on land use, land use change and forestry (LULUCF) activities (Höhne *et al*. 2007). These activities are defined as afforestation, reforestation, deforestation, forest management, revegetation, crop land management, and grazing land management.

Detecting changes in biomass for large areas for reporting purposes requires efficient inventory systems that can be both accurate and give data for sufficiently short time intervals. There are several possible methods that can be applied that can be categorized as either field based inventories or remote sensing. Field based inventories are most often based on field plot sample surveys with systematic or random design for the area in question. Remote sensing is performed from air- or spaceborne platforms and may be combined with ground observations for calibration.

Remote sensing techniques have the advantage compared to strictly field based methods, that they cover large areas with use of relatively little time resources. The accuracy and efficiency will of course vary between methods. Another advantage of remote sensing techniques is that they can be repeated with short time intervals. For a country for example, it can take several years to complete a full field based inventory, whereas an air- or spaceborne sensor can collect data from the same area in much less time. Moreover, the percent of the area sampled with remote sensing techniques is often significantly larger than for field based methods. Furthermore, for certain areas with limited access, field based methods can be very costly because

of the extra resources needed for transportation of field crews. In mountain areas far from developed infrastructure for example, the use of remote sensing is highly relevant. In such areas, traditional forest inventories have also been limited because these areas comprise little merchantable timber and they have been of marginal interest from an economical point of view. From a carbon reporting perspective, however, these areas are very interesting because it is in this region the most pronounced changes due to climate change are likely to take place (Kupfer and Cairns 1996, Holtmeier and Broll 2005).

Currently, airborne laser scanning (ALS) is one of the most efficient remote sensing techniques to produce precise estimates of biophysical properties of trees and forests (e.g. Næsset 2007, Næsset and Nelson 2007). During the last decade, a few studies have explored if ALS also can be utilized for detection of change in forests (Yu *et al*. 2003, 2004, 2005, 2006, 2008, St-Onge and Vepakomma 2004, Gobakken and Næsset 2004, Næsset and Gobakken 2005, Solberg *et al*. 2006, Hopkinson *et al*. 2008). All these studies report good results regarding the capability of multi-temporal ALS datasets to be utilized for change detection and estimation of properties such as tree height, timber volume and biomass in forests.

Since few studies of biomass change have been carried out in mountain forests, we wanted to test if multi-temporal ALS data sets could be utilized to estimate changes also in this type of forest, which is characterized with fairly low growth rates. We used data from a growth period of four seasons from a mountain forest in southeastern Norway.

2. Materials and methods

2.1. Study area

The study area is an experimental site that was established in 1930 for regeneration studies. The site is situated in the southeast of Norway at approximately 61º 44´ N, 10º 35´ E, 700-1150 m a.s.l.

2.2. Field data

The field data were collected on plots which were quadratic and rectangular in shape. Each plot was visited in the spring of 2005 and in the fall of 2008 which means that the data spans four full growth seasons. The sizes of the plots were variable and ranged between 96.2 and 1134.1 m^2 with an average of 375.1 m^2. The number of plots was 27. For each of these plots, diameter at breast height (DBH) and species were registered for every tree together with heights of sample trees. Biomass of all trees was calculated using single tree equations. Since there are no consistent set of biomass equations available for Norway, we used the allometric equations calibrated on data for entire Sweden by Marklund (1988), which have DBH and tree height as independent variables. Biomass was scaled to per hectare values for all plots. Change in biomass, which is the response variable in this study, was calculated as the difference between per hectare field observed biomass in 2008 and the corresponding value of 2005. A summary of the field data is found in Table 1. For all plots, the changes in biomass were positive in the four year period. The mean change was 17.5 t/ha with minimum and maximum values of 3.0 t/ha and 53.5 t/ha, respectively.

Table 1. Summary of data

Variable	Mean	SD	Min	Max
Proportion of spruce	94.7	4.7	85.7	100.0
Proportion of pine	0.9	2.8	0.0	13.9
Proportion of hardwood species	4.4	4.4	0.0	13.6
Plot area (m^2)	375.1	325.5	96.2	1134.1
Biomass in 2005 (t/ha)	135.1	57.2	12.2	295.0
Biomass in 2008 (t/ha)	152.6	66.2	17.9	348.5
Biomass change between 2008 and 2005 (t/ha)	17.5	11.1	3.0	53.5

2.3. Laser data

Laser data were collected to correspond in time to the two field campaigns. The two laser data acquisitions were carried out the 26th of June 2005 the 28th of August 2008, respectively. A fixed-wing Piper PA31 Navajo aircraft was used in both acquisitions. In the 2005 acquisition an Optech ALTM-3100 laser scanner was used, operated from an altitude of approximately 600 m a.g.l. In the 2008 acquisition, an Optech ALTM-GEMINI was used and flying altitude was 600 m a.g.l. The point density on the ground for the ground plots used in this study was approximately 3.4 and 4.7 m^{-2} in 2005 and 2008, respectively.

An initial processing of the data was accomplished by the contractor (Blom Geomatics, Norway). Planimetric coordinates (*x* and *y*) and ellipsoidal height values were computed for all echoes. Both flight campaigns were flown as four single strips with no overlap, but two perpendicular strips were flown to correct for systematical errors between strips. Triangular Irregular Networks (TIN) based on the ground echoes were established by means of the densification algorithm (Axelsson 2000) of the TerraScan software (Anon. 2005) for both acquisitions. TIN models were created from the planimetric coordinates and corresponding heights of the laser echoes classified as ground points. The heights above the ground surface were calculated for all echoes by subtracting the respective TIN heights from the height values of all echoes recorded.

The height distribution created from the first echoes was used to calculate plot-level percentiles for 0, 10,…, 90, and 100% of the heights (h_0, h_{10},…,h_{90}, h_{100}). We labeled these variables “height variables”. Furthermore, cumulative proportional canopy densities (d_0, d_1,…,d_8, d_9) were calculated for 10 uniform fractions across the height range of the laser height distribution. We labeled these “density variables”. We also calculated mean height of laser echoes (*mean*) and the corresponding coefficient of variation (*cv*). The height distribution contained only those laser points which were classified as above-ground echoes, using a threshold value of 2 m. The h_{10} variable denotes the height at which the accumulation of laser echoes in the vegetation is 10% for the first echoes. The d_1 variable denotes the cumulative proportion of laser echoes that is accumulated from the top of the canopy at 10% of the maximum laser measured canopy height, relative to the total number of echoes. For further details see Næsset (2004b). In addition, differences between all corresponding laser variables obtained in the data from the two different acquisitions (2008 and 2005) were calculated. We labeled these variables “delta values”. For example, δp_{80} is the difference in height between the 80th height percentile from the first echoes from the 2008 data and the corresponding percentile from the 2005 data. These were used as the potential explanatory variables in this study. We also used square transformations of these variables.

2.4. Plotting mean differences of laser variables

Different variables derived from laser data collected over forest are affected differently by development such as growth, mortality or harvest. This is because the distribution of biomass changes both vertically and horizontally as a consequence of these processes. To illustrate this and for better conceptual understanding, we made graphical plots of the mean change for the height and density variables over all plot.

2.5. Estimation of biomass change

In the current study, the change of biomass (t/ha) was modeled directly by means of the delta values described above. An ordinary least square model was fitted using the REG procedure of SAS (SAS Institute Inc. 1990), and all delta values were candidate variables to explain the variation in the ton per hectare change of biomass (*δbiomass*) over the four year period. The model appears in equation 1:

$$\delta biomass = \beta_0 + \sum_{k=1}^{p} \beta_k x_k + e$$

where β_0 is the constant term, β_k is the parameter estimate for explanatory variable *k* (x_k), *p* is the total number of variables in the model, and *e* is the error term. The final explanatory variables were selected among 23 potential laser variables and square transformations of these variables so that the total potential number of variables was 46. The variable selection was based on an initial stepwise selection where inclusion and removal of variables was based on a significance level of 0.05. A further removal of variables was then manually carried out based on evaluation of the collinearity among the statistically selected variables. The collinearity assessment was based on the condition number (largest eigenvalue over the

smallest eigenvalue). Condition numbers <100 indicates no serious problems due to collinearity (Montgomery *et al*. 2001, p339). Furthermore, the accuracy of the change estimate was assessed by leave one out cross validation. By means of the PRESS option of the REG procedure, cross validated differences were acquired and by *t*-tests it was assessed if they were significantly different from zero.

In addition to analyzing the differences between predicted and observed biomass change at full range of the data, we also split the data material in two parts according to fixed intervals of observed biomass change, to assess if the predictions where sound for plots with little as well as large observed changes. The limit between small and large changes was set at 12 t/ha.

3. Results

Figures 1 and 2 are graphical representations of the mean change for the different height and density variables, respectively. Figure 1 shows that the most pronounced changes were found for the high height percentiles. The magnitude of the change had a decreasing trend towards the low percentiles. For the density variables, Figure 2 shows the opposite trend. The delta values of low density were greater than those dependent on the biomass conditions only in the upper parts of the canopy.

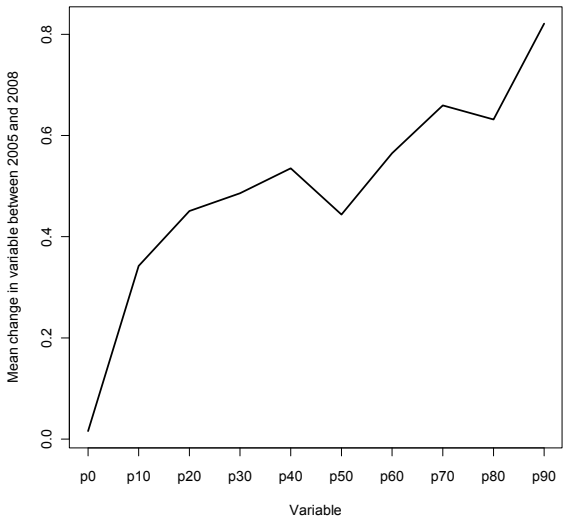


Figure 1. Mean changes for different height variables between 2005 and 2008.

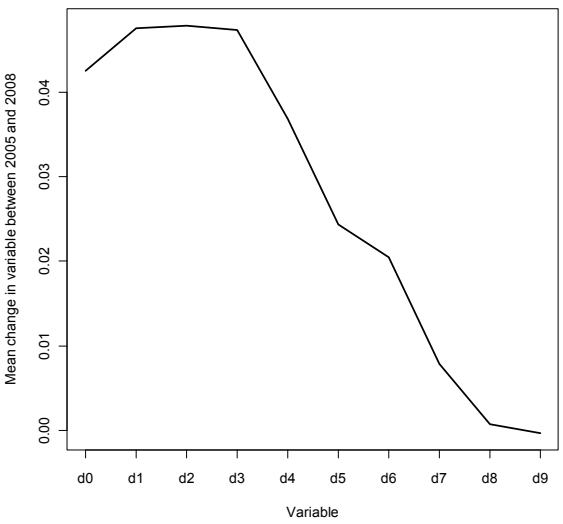


Figure 2. Mean changes for different density variables between 2005 and 2008.

The results from the modeling of *δbiomass* appear in table 2. The table shows that four different variables where selected under the selection criteria used in this study. It appeared that δp_0, δp_{70} were selected together with $\delta p_{10}{}^2$ and $\delta mean^2$. The model fit as represented by R^2 was 0.87 and the root mean square error estimate (RMSE) was 4.4 t/ha. The condition number (CN) was 4.4.

Table 2. Selected variables, model fit, RMSE, and condition number (CN) for models for direct prediction of biomass change.

Variables	R^2	RMSE	CN
δp_0, δp_{70}, $\delta mean^2$, $\delta p_{10}{}^2$	0.87	4.4	4.4

Furthermore, table 3 displays the cross validation results of the model. The rightmost column displays cross validated differences between predicted and observed biomass change. The mean residual over all plots was 0.5 t/ha with a corresponding standard deviation of 5.3 which means that the difference was not significantly different from zero. Moreover, mean residuals for the two parts of the data material distributed across the range of observed biomass change were also calculated. Table 3 shows that none of the differences were significant.

Table 3. Cross validated predicted biomass change, observed biomass change, and cross validated difference (Mean residual) for full range data and for groups according to observed biomass change.

Biomass range (*n*)	Predicted biomass change (*SD*)	Observed biomass change (*SD*)	Mean residual (*SD, significance*)
Full range (*27*)	17.1(*10.1*)	17.5 (*11.1*)	0.5(*5.3, ns*)
<12 t/ha (*9*)	7.7(*5.2*)	7.2(*2.7*)	–0.5*(4.9, ns)*
>12 t/ha (*18*)	22.0(*10.2*)	22.7(*10.1*)	1.0*(5.6, ns)*

4. Discussion

The graphical plots show that both the height variables (figure 1) and the density variables (figure 2) were affected differently across their range by growth during the four year period. Figure 1 shows that the low height percentiles were little affected, whereas the high percentiles ere quite much affected. The mean increase of the 90th height percentile between 2005 and 2008 (0.82 m), was in fact very close to the mean increase of field measured dominant height which was 0.92 m in the current data. Both the 90th percentile and the maximum laser height are close approximations for dominant height at one specific point in time, even though there will be some negative bias because the very apex of a tree top does not have enough mass to trigger an echo (e.g. Ørka *et al*. 2010). The bias may also be dependent on tree size (Ørka *et al*. 2010) which is reasonable since both the mass of the top and the allometry of a tree changes with growth. It is therefore to be expected that the delta values of both *p90* and *p100* will be biased estimates of dominant height growth. Furthermore, the delta values of height percentiles in increasing distance from the tree tops were decreasing, indicating that the low height percentiles were quite stable between 2005 and 2008. This is also logical since the relatively largest height growth is taking place in the upper part of the canopy. However, there may be an effect of sensor. Different sensors may produce point clouds with different properties (Næsset 2009) and variables derived may therefore be shifted between datasets. In this study, the sensor used in 2008 (ALTM-GEMINI) was an upgraded version of the sensor used in 2005 (ALTM-3100), and it is likely that they produce point clouds with slightly different properties. Figure 2 shows that the trend for the density variables is the opposite compared to the height variables. Thus, those variables that reflect variation at lower levels of the canopy were those that had the largest mean difference between acquisitions. The reasons for this are both that individual tree canopies get denser with growth, but also because the intersection of adjacent tree canopies increases. The pattern may also be affected by sensor effects.

The results presented in the graphical figures show some trends for the delta values of both the height and density variables. However, it is not straight forward to use the large delta values in a model intended for explaining the variation of biomass change just because they are large. The variables used in such a model must of course be relevant for the observed variation of the response variable. Table 2 shows the variables selected for the model all came from the height distribution. Most of the range of the distribution is represented. However, it was unexpected that no density delta value was selected. Other studies that have modeled biomass and timber volume using ALS, usually use both density and height variables (e.g. Næsset and Gobakken, 2008) and it would be natural to expect that delta values from both height and density variables to appear in a model of biomass change. Preliminary analyses based only on untransformed delta variables showed, however, that the delta density variable from the lower part of the canopy had little effect on the model. The single variable that was most correlated to the change in biomass was *δp80*.

The model was tested by means of cross validation. Table 3 shows that there were no statistically significant differences between predicted and observed biomass changes, neither for the whole data material, nor for the two groups of plots according to the magnitude of their observed biomass.

Sensor effects will always be something that need to dealt with in multi-temporal applications of ALS. However, in the current study where we utilized multi-temporal field- and laser data and the observed biomass change was modeled by means of delta laser variables, the sensor effects were accounted for by the model. A different combination of sensors would have resulted in a different model.

Delta values derived from multi-temporal ALS datasets seem to have potential for explaining the variation in biomass change based on the results of the current study. However, other and more complex relationships can be found for other forest types and for other types of changes. In the present dataset the only change was growth, but with harvest of different kinds, wind throws, insect attacks etc., the relationships may be different. Future studies should therefore include other types of change as well as explore other model forms and modeling techniques.

References

ANON, 2005, TerraScan user's guide. Terrasolid Ltd., Jyväskylä, Finland, 169 pp. www.terrasolid.fi. Accessed 2 Oct 2006

AXELSSON, P., 2000, DEM generation from laser scanner data using adaptive TIN models. International Archives of Photogrammetry and Remote Sensing **33**, pp. 110–117

GOBAKKEN, T. AND NÆSSET, E., 2004, Effects of forest growth on laser derived canopy metrics. In: International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Freiburg, Germany, Vol. XXXVI, Part 8/W2, pp. 224–227.

HÖHNE, N., WARTMANN, S., HEROLD, A. AND FREIBAUER, A., 2007, The rules for land use, land use change and forestry under the Kyoto Protocol – lessons learned for the future climate negotiations. *Environmental Science and Policy,* **10**, pp. 353–369.

HOLTMEIER, F. AND BROLL, G., 2007, Treeline advance-driving processes and adverse factors. Landscape Online, **1**, pp. 1–33.

HOPKINSON, C., CHASMER, L. AND HALL, R.J., 2008, The uncertainty in conifer plantation growth prediction from multi-temporal lidar datasets. *Remote Sensing of Environment,* **112***,* pp. 1168–1180.

KUPFER, J.A. AND CAIRNS, D.M., 1996, The suitability of montane ecotones as indicators of global climatic change. Prog. Phys. Geogr., **20**, pp. 253–272.

MARKLUND, L.G., 1988, *Biomassafunktioner för tall, gran och björk i Sverige*. [Biomass functions for pine, spruce, and birch in Sweden] (Rapport 45). Institutionen för skogtaxering, Sveriges Lantbruksuniversitet.. (In Swedish with English summary)

MONTGOMERY, D.C., PECK, E.A. AND VINING, C.G., 2001, Introduction to linear regression analysis (Third edition). Wiley, New York

NÆSSET, E., 2002, Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment,* **80**, pp. 88–99.

NÆSSET, E. AND GOBAKKEN, T., 2005, Estimating forest growth using canopy metrics derived from airborne laser scanner data. *Remote Sensing of Environment,* **96**, pp. 453–465.

NÆSSET, E. AND 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**, pp. 3079–3090.

NÆSSET, E. AND NELSON, R., 2007, Using airborne laser scanning to monitor tree migration in the boreal-alpine transition zone. *Remote Sensing of Environment,* **110**, pp. 357–369.

NÆSSET, E., 2004a, Effects of different flying altitudes on biophysical stand properties estimated from canopy height and density measured with a small-footprint airborne scanning laser. *Remote Sensing of Environment ,* **91**, pp. 243–255.

NÆSSET, E, 2004b, Practical large-scale forest stand inventory using small-footprint airborne scanning laser. *Scandinavian Journal of Forest Research*, **19**, pp. 164–179.

NÆSSET, E., 2007, Airborne laser scanning as a method in operational forest inventory: Status of accuracy assessments accomplished in Scandinavia. *Scandinavian Journal of Forest Research,* **22**, 433–442.

NÆSSET, E, 2009, Effects of different sensors, flying altitudes, and pulse repletion frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. *Remote Sensing of Environment*, **113**, pp. 148–159.

ØRKA, H.O., NÆSSET, E. AND BOLLANDSÅS, O.M., 2010, Effects of different sensors and leaf-onand leaf-off canopy conditions on echo distributions and individual tree properties derived from airborne laser scanning. *Remote Sensing of Environment,* **114**, pp. 1445–1461.

SAS INSTITUTE INC., 1990, SAS procedures guide for personal computers. Version 6. 5th ed. SAS Institute Inc. Cary, N.C.

SOLBERG, S., NÆSSET, E., HANSSEN, K.H., AND CHRISTIANSEN, E., 2006, Mapping defoliation during a severe insect attack on Scots pine using airborne laser scanning. *Remote Sensing of Environment,* **102**, pp. 364−376.

ST-ONGE, B., AND VEPAKOMMA, U., 2004, Assessing forest gap dynamics and growth using multi-temporal laser-scanner data. In *Laser-scanners for forest and landscape assessment. Proceedings of the ISPRS working group VIII/2*. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVI, Part 8/W2. (pp. 173–178). Germany' University of Freiburg.

YU, X., HYYPPÄ, J., KAARTINEN, H., AND MALTAMO, M., 2004, Automatic detection of harvested trees and determination of forest growth using airborne laser scanning. Remote Sensing of Environment, **90**, pp. 451–462.

YU, X., HYYPPÄ, J., KAARTINEN, H., HYYPPÄ, H., MALTAMO, M., AND RÖNNHOLM, P., 2005, Measuring the growth of individual trees using multi-temporal airborne laser scanning point clouds. In *Proceedings of the ISPRS Workshop Laser Scanning 2005*, 12-15 September 2005, Enschede, the Netherlands., 36, pp. 204–208.

YU, X., HYYPPÄ, J., KAARTINEN, H., MALTAMO, M., AND HYYPPÄ, H., 2008, Obtaining plotwise mean height and volume growth in boreal forests using multi-temporal laser surveys and various change detection techniques. *International Journal of Remote Sensing*, **29**, pp. 1367–1386.

YU, X., HYYPPÄ, J., KUKKO, A., MALTAMO, M., AND KAARTINEN, H., 2006, Change detection techniques for canopy height growth measurements using airborne laser scanner data. *Photogrammetric Engineering and Remote Sensing,* **72**, pp. 1339−1348.

YU, X., HYYPPÄ, J., RÖNNHOLM, P., KAARTINEN, H., MALTAMO, M. AND HYYPPÄ, H., 2003, Detection of harvested trees and estimation of forest growth using laser scanning. In *Proceedings of the ScandLaser Scientific Workshop on Airborne Laser Scanning of Forests*, 3-4 September 2003, Umeå, Sweden. Swedish University of Agricultural Sciences. pp. 115–124.

Biomass assessment of open woodland trees based on a 3D normalized cut segmentation of full waveform LiDAR data

MATTHIAS RENTSCH†, JOSEF REITBERGER†, ALFONS KRISMANN‡ and PETER KRZYSTEK†*

†Department of Geoinformatics, Munich University of Applied Sciences, 80333 Munich, Germany

‡Institute for Landscape and Plant Ecology (320), University of Hohenheim, 70593 Stuttgart, Germany

Abstract

The main goal of the BMU (German federal ministry for the environment, nature conservation and nuclear safety) funded project 'LiDAR based biomass assessment' is the precise and nationwide detection and calculation of the biomass potential coming from wood cuttings. For this purpose, first and last pulse LiDAR data serve as data basis and are available all over the country by 2013. A major issue of our study is to model the relationship between the vegetation volume derived from adequate LiDAR measurements and the above-ground biomass (AGB) for well-defined grove types. The mandatory field calibrations are performed in advance for selected reference groves (e.g. single trees, tree stands, hedges). For this purpose, full waveform LiDAR data were acquired in August 2009 in leaf-on condition. Furthermore, the groves are captured in April 2010 by a precise and full 3D TLS registration in leaf-off situation. Firstly, the paper is reporting about preliminary results from the 3D normalized cut segmentation adopted for open woodland trees and the follow-on vegetation volume calculation. Secondly, the focus is on the estimation of the above-ground biomass from the diameter at breast height (DBH) and the tree height. Finally, conversion factors are derived which relate the AGB to the LiDAR derived volume calculations.

1. Introduction

One of the main intensions of the "Erneuerbare Energien Gesetz (EEG)" which was implemented by the German federal government in 2009 is the enhancement of renewable energy resources for the electric power supply up to at least 30% until 2020, and continued afterwards. Adjacent to wind, water, the sun, geothermal energy, liquid and gaseous biomass, the EEG is focussed on the increased usage of compact biomass as energy resource. Beside biomass coming from woodland and forest fuel, the incorporation of wood cuttings from landscaping is getting more and more attractive. However, an intensive use of this primary energy source is not yet realized due to technical and logistic reasons. Nevertheless, recent studies have revealed that wood cuttings comprise an enormous energetic potential, representing the biomass resource that could be used most sustainably (NBBW 2008, Engel 2004).

In the last two decades, up-to-date remote sensing methods and techniques were successfully incorporated in subject areas of forest science, like the assessment of forest canopy fuel (e.g. Erdody and Moskal, 2010), and of woody biomass which is the scope of this paper. In this context, small-footprint LiDAR systems have established as an appropriate method for the area-wide and precise estimation of forest structural parameters which serve as basis for the estimation of the secondary quantities, e.g. vegetation volume and above-ground biomass. A successful segmentation-based method to retrieve stem volume estimates from 3D tree height models produced by airborne laser scanners (ALS) was already presented by Hyyppä et al. (2001). Herein, the resulting precision was found to be better than that in conventional standwise forest inventories. A refinement of timber volume estimation is given by Maltamo et al. (2004) by the use of a parameter prediction method to describe also smaller trees. Lim et al. (2003) derived laser height metrics from LiDAR which are capable to accurately estimate biophysical properties of tolerant northern hardwood forests with R^2=0.87 for total wood volume and R^2=0.85 for total aboveground biomass. A comprehensive estimation of aboveground biomass based on first and last pulse data (density 0.7-1.2 pts/m^2) was conducted by Næsset and Gobakken (2008) in Norway's boreal forests, leading to R^2=0.88 for AGB. An ongoing nationwide project, 'LASER-WOOD' is conducted by Hollaus et al. (2010) which aims to assess the growing stock and above-ground biomass potential for the Austrian forests using LiDAR and forest inventory data. Popescu (2007) reports from conclusions of recent studies 'which demonstrated that DBH is the most reliable variable for biomass estimation'. In airborne applications, the DBH could not be measured directly

and has to be derived by regression analysis or clustering methods. Nevertheless, in case of terrestrial laser scanning surveys, the direct measurement of DBH is feasible on a high level of accuracy (Bienert *et al.* 2007, Wezyk *et al.* 2007). Heurich (2008) describes the automatic recognition and measurement of single trees over the richly structured natural forests of the Bavarian Forest National Park. The results show that it was possible to identify 93.3% of the wood volume measured on the ground. The aim of the study of Kalliovirta and Tokola (2005) was the investigation of the relationship of DBH with maximum crown diameter, tree height and other possible independent variables by formulation of a total of 76 models. The authors also mentioned that the relations of DBH with crown width are different for forest-grown and open-grown trees. The study of van Aardt et al. (2008) is dedicated to evaluate the potential of an object-oriented approach to forest type classification (deciduous-coniferous, deciduous-coniferous-mixed) as well as volume and biomass estimation in the Virginia Piedmont, U.S.A. based on small-footprint, multiple return LiDAR data. The classifications yield an overall accuracy up to 89%, and the estimates considering the volume and biomass estimations exhibit differences of no more than 5.5% with respect to field surveys. A scale-invariant estimation approach aimed to forest biomass estimation from LiDAR remote sensing is given by Zhao et al. (2009) in a temperate mixed forest of eastern Texas. The results indicate that the models incorporated in this approach can accurately predict biomass with an R^2 ranging from 0.80 to 0.95.

Several studies also include the LiDAR intensity for biomass assessment. Kim et al. (2009) distinguish between live and dead standing tree biomass in a mostly mixed coniferous forest on the North Rim of Grand Canyon National Park, USA. The authors conclude that the estimation power for live biomass benefits from the consideration of the LiDAR intensity. The investigations of Morsdorf et al. (2010) were performed in a multi-layered Mediterranean forest ecosystem revealing that ALS height and intensity information are appropriate for separating between three different vegetation strata with obtained overall accuracies of 80 to 90%. Remarkably high correlations between the number of ALS returns per height bin and the field measured canopy volume profile were achieved for the oak (96%) and the pine plot (91%), revealing the significant high coherence of the LiDAR echo characteristics with the vegetation volume determination of the reference plots. Biomass carbon stocks for a Mediterranean forest in central Spain were estimated by García et al. (2010) using LiDAR height and intensity data. Herein, the use of normalized intensity-related variables led to more accurate estimates of the biomass for three different tree species.

All in all, the selected papers exhibit various and sophisticated approaches concerning the determination of forest metrics from LiDAR data and the estimation of volume and AGB. Nevertheless, these methods were more or less applied to closed forest stands. In contrast, open woodland groves are characterized by heterogeneous structures containing various species within a small area, complex shapes, and high fractions of bushy material in many cases which suppress the penetration of laser pulses. In addition, the standard methods for tree segmentation which lead to quite good results for forest stands are less effective in case of open woodland. These conditions make it even difficult to correctly assess the biomass for open woodland groves (Straub 2010). As consequence, only a small number of papers are dealing with this topic. Straub (2010) is reporting about some studies in Germany, which are characterized by high uncertainties in terms of the amount of biomass. One reason is that the validation is performed only with a small number of random tests coming from destructive sampling. Velázquez-Martí et al. (2010) have developed prediction models which allow the biomass determination of a bushy Mediterranean forest area near Valencia/Spain consisting of 5 species using a LiDAR derived canopy height model (CHM). If this models can be transferred to the conditions in Germany was not yet examined.

As described before, the main goal of the superior project ‘LiDAR based biomass assessment’ is the precise and nationwide detection and calculation of the biomass potential coming from wood cuttings. As the ignition of the project was in November 2009, our paper is focused on first results concerning the initial calibration efforts for a sample reference grove. (1) Estimation of the above-ground biomass on a tree-level basis using the manually measured DBHs and tree heights. (2) The volume determination based on the TLS point cloud. (3) The volume determination after a 3D normalized cut segmentation (Reitberger *et al.* 2009) for full waveform and first/last-pulse data. (4) Determination of a conversion factor to derive the above-ground biomass from independent estimation using method (1) based on the calculated vegetation volumes from methods (2) and (3).

Material

For the entire project, various calibration sites are distributed over six federal states of Germany (mostly in Baden-Württemberg). The two calibration areas which are concerned in our investigations are located in the

Bavarian Forest in southeastern Germany (Arnbruck and Regen). For this two areas, a sum of 13 reference groves were chosen by visual inspection comprising four hedges with trees, one hedge, two high groves, five single trees and one forest island. All of these groves were captured by full waveform LiDAR data (10 pts/m^2) in August 2009 by the Riegl LMS-Q560 scanner in leaf-on condition. Furthermore, the areas were covered by routine first/last-pulse ALS surveys on behalf of the Bavarian Survey Administration in April 2008 and October 2009. The acquisitions were performed under leaf-off conditions with a mean point density of around 1-2 pts/m^2. In addition, seven of the groves are measured by a precise and full 3D TLS registration in April 2010 using the Riegl LMS-Z390 in leaf-off situation. Finally, independently gathered tree heights using a laser rangefinder and manually measured caliper DBHs are included for two objects (status July 2010).

In the following, the methods are described and applied on one reference grove of type 'high grove' located in the area of Regen, named as '08_Reg_hohFg' in the following. The grove consists of around 200 trees and bushes and covers an area of approximate 0.4 ha. Tree species are mostly oak (61%) and birch (30%), with some Scots pines, cherry trees, hornbeams and maple (in all 9%). The bushes consist of hazel nut (96%), elder (4%) and whitethorn (1%). The grove is situated on slightly sloped terrain and is characterized by a densely closed canopy cover at a maximum level of around 22 m above ground. The grove (figure 1) was captured by five single TLS scans with 1 cm resolution which were merged by RiSCANPRO®. The scanner positions were determined by differential GPS with an absolute accuracy of 1-3 cm.

Figure 1. TLS point cloud of reference grove '08_Reg_hohFg'

In December 2009, all trees and bushes of reference grove '08_Reg_hohFg' with a DBH of >3.0 cm were captured. The DBHs were derived from the mean of at least two caliper measurements to account for the non-circular stem outlines. The heights of the outer trees could be acquired by a laser rangefinder. In case of missing laser returns, the heights of some of the inner trees were visually estimated according to their neighbors. Initially, the tree positions were registered by a hand held GPS receiver with a low horizontal accuracy of around 3-5 m. The coordinates were later on replaced by the centers of the tree stems which were discovered in the precise TLS point clouds.

2. Method

The main objective of the field calibrations is the determination of correlation factors which allow the appropriate conversion of the vegetation volume [m^3] of open woodland groves derived by LiDAR methods into the unit of above-ground biomass, the bulk stacked cubic meter [BCM]. As a first step, the simplified assumption is established by considering a linear correlation between vegetation volume and the corresponding above-ground biomass. Thus, the biomass is calculated on behalf of model assumptions and has to be validated against reference measures based on the DBH-tree-height-method and destructive sampling, respectively.

In recent studies concerning the volume-biomass-conversion, the vegetation volume is in many cases approximated by simple geometric elements like cuboids, leading to high uncertainties for the volume determination (e.g. 32%, Cremer 2007): Thus, the primary option of our work is to refine the estimation of the vegetation volume by means of a precise normalized surface model generation based on TLS data and

taking advantage of the 3D normalized cut segmentation applied to the full waveform and first/last pulse LiDAR data.

Usually, the merchantable wood volume is expressed as a function of DBH alone or DBH and tree height.

$$VOL_m = f(DBH,h_) \ [m^3] \quad (1)$$

According to the basic conditions concerning the biomass estimation of a corresponding project (Johst and Conrady 2009), we applied the stem volume equations published by Zianis et al. (2005). Unfortunately, the stem volume which is calculated by these equations is not commonly defined in Europe. Whereas in Nordic countries, the stem is dimensioned from the root up to the top, some other countries consider the stem as the main trunk up to a minimum diameter of 7 cm. Nevertheless, in our study, the following functions and coefficients are employed.

Table 1. Functions and coefficients applied for stem volume calculation

Tree species	Index acc. to Appendix B and C (Zianis *et al.* 2005)
Oak (quercus robur)	207
Birch (betula spp.)	33
Scots pine (pinus sylvestris)	149
Wild cherry (prunus avium)	192
Hornbeam (carpinus betulus)	40
Maple (acer pseudoplatanus)	10
Hazel nut	218
Elder/Whitethorn	207

For these functions there is no information if the bark is included or not, thus we assume the volume for the stems and the bark. The tree biomass resulting from the stem volume is deduced by

$$TBM = VOL_m * f_{BCM} \ [BCM] \quad (2)$$

f_{BCM} represents the biomass conversion factor and is approximated by 2.43 (FNR 2009). This factor is commonly valid for all tree species. The determination of the component biomass (branches and foliage), which has to be added to the TBM, is a difficult task depending on tree species, tree age and other parameters and, thus, is only roughly assumed by 20% of the tree biomass.

3. Results

Biomass estimation using DBH and tree height

The results of the stem volume calculations are displayed in table 2.

Table 2. Stem volume for reference grove '08_Reg_hohFg'

Tree species	Stem volume [m^3]
Oak (quercus robur)	89.907
Birch (betula spp.)	19.904
Scots pine (pinus sylvestris)	1.672
Wild cherry (prunus avium)	2.466
Hornbeam (carpinus betulus)	0.005
Maple (acer pseudoplatanus)	0.053
Hazel nut	1.197
Elder/Whitethorn	0.042
Sum	115.246

The tree biomass is calculated from equation (2) resulting in

TBM =280.048 BCM

As mentioned before, the estimation of the component biomass (branches and foliage) is affected by various parameters. As we derive a mean DBH of 29.4 cm for oak representing the dominant tree species (61%), the estimation of 20% component biomass is acceptable and corresponds quite well with the assumption given by Jenkins et al. (2003). Finally, we can calculate the total above-ground biomass for the reference grove '08_Reg_hohFg'

AGB = 336.058 [BCM]

Vegetation volume from TLS point cloud

Before generating a normalized surface model (nDSM), a high resolution (grid size 0.25 m) terrain model (DTM) was interpolated from classified ground points originating from the TLS points and the full waveform ALS points. The ALS last and single returns had to be included because the point coverage within the reference grove could not be sufficiently covered by the TLS data (due to occlusions in the inner zone). Contrary to the DTM, a grid size of 0.5 m was chosen for the surface model (DSM) to get a slightly better smoothing. The vegetation volume was calculated by product sum of the height of each nDSM cell (above a defined threshold) and the cell base area of 0.25 m^2 (table 3).

Vegetation volume after 3D normalized cut segmentation

Reitberger et al. (2009) have shown, that the new 3D normalized segmentation method leads to an improvement of the tree detection rate with respect to a conventional watershed segmentation, especially in the case of full waveform data. Figure 2 displays the result after the segmentation of full waveform ALS data of reference grove '08_Reg_hohFg'. Herein, the black lines represent the reference trees (list not yet completed). The positions were recovered from the TLS data, the tree heights are preliminary taken from the laser rangefinder measurements, however, in future will be reconstructed from the full waveform ALS data.

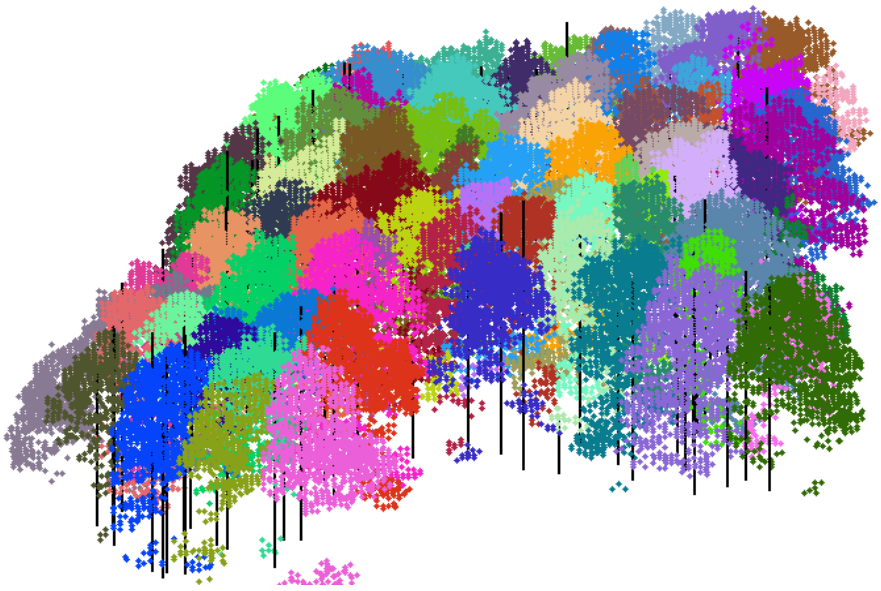


Figure 2. Segmentation result for full waveform ALS data for '08_Reg_hohFg'

The vegetation volume results from the product sum of the maximum height and the basal area of each segment (table 3). The 3D normalized cut segmentation was also applied to the first and last pulse ALS data, the result is presented in figure 3.

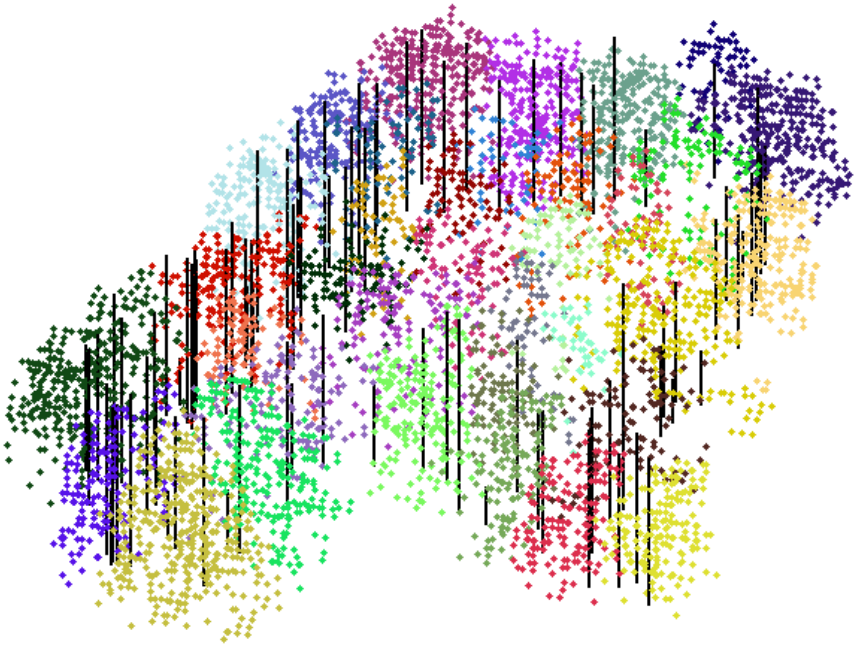


Figure 3. Segmentation result for first and last pulse ALS data for '08_Reg_hohFg'

As before, the vegetation volume results from the product sum of the maximum height and the basal area of each segment (table 3)

Derivation of biomass conversion factors

With the results from 4.1 to 4.3, the biomass conversion factors can be calculated as displayed in table 3.

Table 3. Results for vegetation volume, AGB and biomass conversion factors

LiDAR data type	Basal area [m^2]	Vegetation volume [m^3]	AGB from method (1) [BCM]	Conversion factor [BCM/m^3]
TLS	4106.75	67286.43	336.06	0.0049
ALS full waveform	3980.62	86662.15	336.06	0.0039
ALS first+last pulse	3399.50	72175.59	336.06	0.0047

4. Discussion

There are numerous methods in the literature for estimating the above-ground biomass of forest trees from forest structural parameters, mainly based on the DBH and the tree height. In our study, we used functions and coefficients from Zianis et al. (2005) in order to be in consensus with a corresponding project which is, amongst others, dedicated to derive the AGB from the DBH-tree-height-method and to validate it against reference biomass from destructive sampling. For the reference grove '08_Reg_hohFg', the DBHs of all trees with diameter > 3 cm were captured by caliper measures. The DBHs can also be reconstructed from height bins of the TLS point clouds. First random samples show a quite well level of coherence with maximum deviations of around 3.5 cm in diameter. The tree heights are acquired by a mobile stand fixed laser rangefinder. Although the heights of the outer trees could be measured quite well, the observations for

the inner trees are influenced by more uncertainties. For the next future, the tree heights will be recovered by interpolation using the DSM derived from full waveform ALS data. Nevertheless, the AGB estimation in our study for the reference grove is representing a reasonable value.

The determination of the vegetation volume for open woodland trees is referred in only a few publications and reports, in which most methods calculate the product of total basal area and maximum or mean tree height. For our purpose, we derived the volumes using an nDSM in case of TLS data and segmented trees and tree clusters respectively for ALS data. The variations of the results (table 3) are mainly caused by different point densities and LiDAR system properties. The total basal area is the largest one for the TLS data, followed by the full waveform ALS data. This can be explained by the capability of TLS to detect even the tiniest twigs which expands the basal area. This aspect accounts also for the full waveform ALS data with respect to the first and last pulse ones. Contrary, the volume from TLS is significantly below the volume from the ALS data. One explanation might be that mainly the canopies of the inner trees are weakly hit by the laser beams and thus, only a few echo returns could be detected. However, in case of the ALS data, the closed canopy layer is better represented by laser hits. On the other hand, the volume derived after the segmentation of the ALS data is definitely overestimated. We have to admit, that for the time being, this method exhibits no clear advantage with respect to the nDSM based approach, and thus, implies the need for a significant improvement.

Nevertheless, the resulting factors for the conversion of vegetation volume into above-ground biomass are ranging from 0.0039 to 0.0049 and are in good coherence with other investigations. Cremer (2007) is reporting about factors concerning open woodland groves from 0.0010 to 0.0040 with an average deviation of 32%. Trötschler (2008) presents values ranging from 0.0017 to 0.0059 for comparable grove types.

5. Conclusions and Outlook

The focus of our study is on preliminary calibration efforts concerning the biomass assessment of open woodland trees. For this purpose, we used data from field measurements (laser rangefinder, caliper, TLS) and ALS data (full waveform and first/last pulse). One objective was dedicated to the derivation of factors for the conversion of vegetation volume into above-ground biomass. For the volume calculations various methods were incorporated, e.g. the 3D normalized cut segmentation. The approach was applied on a reference grove of type ‘high grove’ consisting of around 200 hardwood trees.

It can be resumed, that our approach shows promising results which can be compared with finding from other studies. However, the results are very preliminary and by far not representative. The first directive for the next future must be the inclusion of more random samples, in order to account for all different reference groves at a significant amount. The next issue will be the refinement of the volume determination in order to benefit from the capabilities of ALS data (e.g. using the ‘alpha-shapes’ approach). In this context, the segmentation methods have to be adopted for the conditions of open woodland groves, e.g. more complex structures, higher vegetation density and divergent tree shapes with respect to forest trees. And finally, the relationship between volume and biomass has to be investigated in detail and comprehensively validated.

Acknowledgments

We like to thank Milan Geoservice GmbH and the Bavarian Office for Surveying and Geographic Information for providing the LiDAR data. This research has been funded by the German federal ministry for the environment, nature conservation and nuclear safety (BMU) under the reference 03KB037A.

References

BIENERT, A., SCHELLER, S., KEANE, E., MOHAN, F. and MAAS H.-G., 2007, Tree detection and diameter estimations by analysis of forest terrestrial laserscanner point clouds. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, **36**, pp. 50-55.

CREMER, T., 2007, Mobilisierung und wirtschaftliche Nutzung von Rohholz aus Wald und Landschaft zur Energieerzeugung. Final report (in German), Deutsche Bundesstiftung Umwelt (DBU), Osnabrück, Germany. Available online at: http://www.dbu.de/OPAC/ab/DBU-Abschlussbericht-AZ-22128.pdf (accessed 27 Jul 2010).

ENGEL, C., 2004, Biomasse: Nachwuchs für Deutschland. *Öko-Mitteilungen* (in German), Öko-Institut, Freiburg, Germany. Available online at: http://www.oeko.de/service/bio/ dateien/de/oemi_artikel_biomasse.pdf (accessed 27 Jul 2010).

ERDODY, T.L. and MOSKAL, L.M., 2010, Fusion of LiDAR and imagery for estimating forest canopy fuels. *Rem. Sens. Env.*, **114**, pp. 725-737.

FACHAGENTUR NACHWACHSENDE ROHSTOFFE (FNR), 2009, Bioenergie - Basisdaten Deutschland. Information broschure (in German), Gülzow, Germany. Available online at: http://www.fnr-server.de/ftp/pdf/literatur/pdf_92-basisdaten_bioenergie_2009.pdf (accessed 27 Jul 2010).

GARCIA, M., RIAÑO, D., CHUVIECO, E. and DANSON, F.M., 2010, Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data. *Rem. Sens. Env.*, **114**, pp. 816-830.

HEURICH, M., 2008, Automatic recognition and measurement of single trees based on data from airborne laser scanning over the richly structured natural forests of the Bavarian Forest National Park. *For. Ecol. Manage.*, **255**, pp. 2416-2433.

HOLLAUS, M., EYSN, L., SCHADAUER, K., JOCHEM, A., PETRINI, F. and MAIER, B., 2010, LASER-WOOD: Estimation of the above ground biomass based on laser scanning and forest inventory data. *Poster presentation at „11. Österreichischer Klimatag"*, Mar 11+12, 2010, Vienna, Austria, extended abstract 2p.

HYYPPÄ, J., KELLE, O., LEHIKOINEN, M. and INKINEN, M., 2001, A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. *IEEE Trans. Geosci. Rem. Sens.*, **39**, pp. 969-975.

JENKINS, J.C., CHOJNACKY, D.C., HEATH, L.S. and BIRDSEY, R.A., 2003, National-scale biomass estimators for United States tree species. *For. Scie,*. **49**, pp. 12-35.

JOHST, A. and CONRADY, D., 2009, Energieholz und Biodiversität – Die Nutzung von Energieholz als Ansatz zur Erhaltung und Entwicklung national bedeutsamer Lebensräume. Project proposal (in German), Naturstiftung David, Erfurt, Germany.

KALLIOVIRTA, J. and TOKOLA, T., 2005, Functions for estimating stem diameter and tree age using tree height, crown width and existing stand database information. *Silva Fennica*, **39**, pp. 227-248.

KIM, Y., YANG, Z., COHEN, W.B., PFLUGMACHER, D., LAUVER, C.L. and VANKAT, J.L., 2009, Distinguishing between live and dead standing tree biomass on the North Rim of Grand Canyon National Park, USA using small footprint lidar data. *Rem. Sens. Env.*, **113**, pp. 2499-2510.

LIM, K., TREITZ, P., BALDWIN, K., MORRISON, I. and GREEN, J., 2003, Lidar remote sensing of biophysical properties of tolerant northern hardwood forests. *Can. J. Remote Sensing*, **29**, pp. 658-678.

MALTAMO, M., EERIKÄINEN, K., PITKÄNEN, J., HYYPPÄ, J. and VEHMAS, M., 2004, Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. *Rem. Sens. Env.*, **90**, pp. 319-330.

MORSDORF, F., MÅRELL, A., KOETZ, B., CASSAGNE, N., PIMONT, F., RIGOLOT, E. and ALLGÖWER, B., 2010, Discrimination of vegetation strata in a multi-layered Mediterranean forest ecosystem using height and intensity information derived from airborne laser scanning. *Rem. Sens. Env.*, **114**, pp. 1403-1415.

NACHHALTIGKEITSBEIRAT BADEN-WÜRTTEMBERG (NBBW), 2008, Energie aus Biomasse: Potenziale und Empfehlungen für Baden-Württemberg. Press release (in German), Stuttgart, Germany. Available online at: http://www.nachhaltigkeitsbeirat-bw.de/mainDaten/dokumente/zsf_bioenergiegutachten.pdf (accessed 27 Jul 2010).

NÆSSET, E. and GOBAKKEN, T., 2008, Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. *Rem. Sens. Env.*, **112**, pp. 3079-3090.

POPESCU, S.C., 2007, Estimating biomass of individual pine trees using airborne lidar. *Biomass and Bioenergy*, **31**, pp. 646-655.

REITBERGER, J., SCHNÖRR, C., KRZYSTEK, P. and STILLA, U., 2009, 3D segmentation of single trees exploiting of full waveform LiDAR data. *ISPRS J. Photogramm. Remote Sens.* **64**, pp. 561-574.

STRAUB, C., 2010, Acquisition of the bioenergy potential and the availability in the forest and non-forest land using new remote sensing techniques. PhD thesis (in German), University of Freiburg, Germany. Available online at: http://www.freidok.uni-freiburg.de/volltexte/7632/ (accessed 20 Jul 2010).

TRÖTSCHLER, P., 2008, Regio-Energieholz: Potenzialstudie und Machbarkeitsstudie zur Ermittlung und naturschutzrechtlichen Bewertung der Energieholzreserven sowie modellhafte Umsetzung. Project proposal (in German), Bodensee-Stiftung, Radolfszell, Germany. Available online at: http://www.iln-singen.de/KurzRegioEnergieholz.pdf (accessed 27 Jul 2010).

VAN AARDT, J.A.N., WYNNE, R.H. and SCRIVANI, J.A., 2008, Lidar-based mapping of forest volume and biomass by taxonomic group using structurally homogeneous segments. *Photogramm. Eng. Remote Sens.*, **74**, pp. 1033-1044.

VELÁZQUEZ-MARTI, B., FERNÁNDEZ-GONZÁLEZ, E., ESTORNELL, J. and RUIZ, L.A., 2010, Dendrometric and dasometric analysis of the bushy biomass in Mediterranean forests. *For. Ecol. Manage.*, **259**, pp. 875-882.

WEZYK, P., KOZIOL, K., GLISTA, M. and PIERZCHALSKI M., 2007, Terrestrial laser scanning versus traditional forest inventory: First results from the Polish forests. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, **36**, pp. 424-429.

ZHAO, K., POPESCU, S. and NELSON, R., 2009, Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. *Rem. Sens. Env.*, **113**, pp. 182-196.

ZIANIS, D., MUUKKONEN, P., MÄKIPÄÄ, R. and MENCUCCINI, M., 2005, Biomass and stem volume equations for tree species in Europe. *Silva Fennica Monographs*, **4**, 63 p.

Woody biovolume extraction from laser scanned trees

*ALEXANDER BUCKSCH *† , STEFAN FLECK‡, SABIENE RUMPF‡,*

PETER RADEMACHER‡

†Department of Earth Observation and Space Systems (DEOS),

Delft University of Technology, Delft 2629HS, The Netherlands

‡ Departments of Forest Growth / Environmental Control, North-West German Forest Research Station, 37079 Göttingen, Germany

Abstract

Biomass extraction from indiviual trees measured by terrestrial laser scanning is a valuable tool for forestry and for orchard plantations to extract the economic and ecological parameters of trees.

The focus of this paper is to extract the woody biovolume from laser scanned trees based on the SkelTre-Skeleton, which is a skeletal structure description. State of the art research proposed skeletons to extract parameters like branch length and branch diameters from trees on the basis of the SkelTre-Skeleton (Bucksch et al. 2010), from trerrestrial LiDAR measurements. In this paper we report on the fully automatic and purely laser-scan derived biovolume extraction of two collected evaluation scenarios:

1.) six leafless orchard apple trees

2.) two European beech trees in a forest

Validation measurements on 6 apple trees comprise the lengths and diameters of all branch segments of their canopies, but not their specific weight so that biovolumes can be compared. The beech tree biovolumes are derived from an allometric relation of tree height, breast height diameter and the known density of the beech wood (Rademacher et al. 2009)

We apply this method in order to assess the biovolume fractions of branch diameter classes in the tree crowns of 3m high apple trees as well as those of 20m high beech trees in the leafless stage in winter. The results show that for the six apple trees the difference in biovolume between automatic and field measurement is 1.18 litres on average. For the beech trees huge differences were derived demanding better field data to draw actual conclusions from it.

1. Introduction

Biomass itself is an important parameter for economic assessment of a forest, since it gives the amount of wood available for fueling or other purely mass-based purposes. Other purposes like wood as building material or source of chemical compounds may require qualitative information on top of this, especially the available amount of wood in specific branch diameter classes. Beyond this, even the distribution of biomass in space may play a role e.g. for the physiological description of light use and transpiration of the trees that determines their ecological function.

The acquisition of detailed data on trees is possible by using a terrestrial laser scanner. New studies show that the diameter and the length of branches are extractable by utilizing a skeletonization (Bucksch and Fleck 2010). Length and diameters of all woody axes are sufficient to determine the woody biovolume of a tree. From the biovolume the biomass can be derived under the assumption of a known density of the wood. The results of this paper give an insight into the amount of biovolume captured by a laser scanner which is calculated on the basis of the skeletal structure extracted by the SkelTre-Algorithm (Bucksch et al. 2010).

We asses the biomass in two main scenarios. The first scenario uses apple trees growing on an orchard plantation. This first scenario contains leafless trees with dense crowns of approximately 3 meters height. The second scenario focuses on mature beech trees scanned in a forest before budbreak in spring.

Related work

Until now the majority of biomass estimations are carried out by airborne systems(e.g. Popescu 2007, Drake et al. 2002). We refer to a review of current techniques used to assess the forest structure with terrestrial laser scanning given by Van Leeuwen et al. (2010) . They also cover the topic of individual tree assessment..

We give in this section a few example studies, which investigate the biomass or tree parameters on an individual tree level close to the context of this paper.

Fleck et al. (2004) used a terrestrial laser scanner to reconstruct the structure of a tree manually in order to predict the physiological properties of a tree. Rossell et al. (2004) described a manual approach to quantify the tree parameters in orchards by using a CAD software and slices through the point cloud. They were able to derive the crop load and tree parameters by mounting the laser scanner on a vehicle. Parker et al (2004) determined the above ground biomass in forests on the basis of terrestrial laser scans and stated that the vertical structure in terrestrial laser scans is consistent with the one found in the field measurements.

Typical for most studies on individual trees is the estimation of tree parameters like diameters of trunk and branches, tree height or crown width with a high degree of accuracy(e.g. Thies et al. 2004, Henning and Radtke 2006).

The best comparable study to this paper is the one by Lefsky and Mc Hale (2008) who determined the biovolume of a complex branching systems of trees by representing the lidar measurments in a 3-dimensional voxel grid. The voxel grid was exploited to derive the biovolume of the tree. Further the sampling pattern of the sensor was taken into account by Lefsky and Mc Hale (2008), in order to derive estimates for the fine branches. A further motivation can be found in (Strahler et al. 2008), who suggested that laser scanning can serve as platform for biomass extraction from individual trees without using allometric relationships.

Our approach automatically calculates the woody biomass of an individual tree and does not use allometric relations. Mostly the evaluation of individual trees stops at the extraction of tree parameters, which can be used in conjunction with allometry to caculate the biomass. Our technique allows to determine the biomass without using allometric relationships and caculates the biomass directly from the point cloud on the basis of an extracted skeleton.

Study areas

In case of the apple trees the study was mainly conducted in apple orchards of the Annapolis Valley, Nova Scotia, Canada close to the city of Kentville (45°4’39’’ N, 64°29’45’’W). The six investigated apple trees (*Malus* x *domestica* Borkh. ‘Honeycrisp’) were located in two orchards that belong to the test sites of the Atlantic Food and Horticulture Research Centre (Fleck et al. 2010). Three apple trees grew on a trellis system and the other three trees stood as single trees in rows. The orientation of the rows is from North to South with a tree spacing of 3m within each row and a spacing of 5m between rows. Trees of comparable height were located next to the investigated trees. The manually measured trunk diameters ranged from 3.9cm to 8.1cm. The tree height of the six apple trees varied from 1.27m to 3.03m.

In case of the beech trees the study was carried out in a forest stand in Hundelshausen (51° 18′ N, 9° 51′ O) in Germany close to the city of Göttingen. The two investigated beech trees (*Fagus sylvatica* L.) were selected in a nutrient rich area of the forest. The manually measured trunk diameters at breast height (DBH) ranged from 21.6cm to 50.8cm. The tree height of the beeches varied from 20.1m to 23.4m.

Figure 1: *On the left one of the scanned apple trees (Apple 4) and on the right one of the scanned beech trees (Beech 2). Undersampling in the upper crown of the beech tree is visible.*

2. Methods

Field measurement

The six apple trees were scanned with the Imager 5006 laser scanner of Zoller and Froehlich. They were registered with the Neptan based software ZF-LaserControl. Each tree was scanned from 4 positions in 2 different heights to overcome occlusion effects. The branches of each tree were numbered for reconstructing the branch hierarchy and their length was measured following the elongation direction of the branch. The diameter of each branch was measured at its base and tip, about 1cm before the node or end bud. The diameters of branches were measured with a calliper in two directions and averaged. If both diameter measurements were more than 1mm apart, a third diameter measurement was taken and the average of three measurements was taken. Branch diameters thicker than 5cm were derived from circumference measurements with a meter tape assuming the trunk or main branch to have a circular cross-section. An example of a registered apple tree is shown in Figure 1 on the left.

The two beech trees were scanned with a Faro Photon system and registered with the Software Faro scene based on fixed targets on the surrounding trees. Each tree was recorded by 6 single scans. Targets were also placed 10 m above the floor in the surrounding tree crowns to assure a homogeneous distribution of targets for registration of the 20m high trees. An example of a registered beech tree is shown in Figure 1 on the right.

The biovolumes of the beech trees were estimated by calculating the biomass as described in Rademacher et al. (2009). Assuming a specific weight of 0.74 kg/l for the wood of beech trees the biovolume was obtained on the basis of the diameter at breast height (DBH). The formula for calculating the biovolume is therefore given as:

$$volume = \frac{0.001074 * DBH - 0.141885}{0.74}$$

Automatic measurement

Branch diameters and length were derived with the method of Bucksch and Fleck (2010). This method segments the tree into individual branches and utilizes the distances to the skeleton to determine the diameter of each individual branch. First a skeleton, as shown in Figure 2, represented as a graph is extracted from the point cloud. This graph is assumed to be centered within the point cloud representing the

tree. A graph contains vertices connected by edges. For every point cloud point the distance to the closest edge is taken. The cylinder length is given by the extracted branch length and the cylinder diameter as the maximum closest to the median of a binned histogram of the scan point distances to the branches represented by the segmented skeleton.

We calculate for every branch the volume of the half cone based on the diameters extracted at the branch base and branch end. For branches where only one diameter was extractable the cylindrical volume was taken. This case of one diameter can occur on finer branches when the branch end is strongly undersampled.

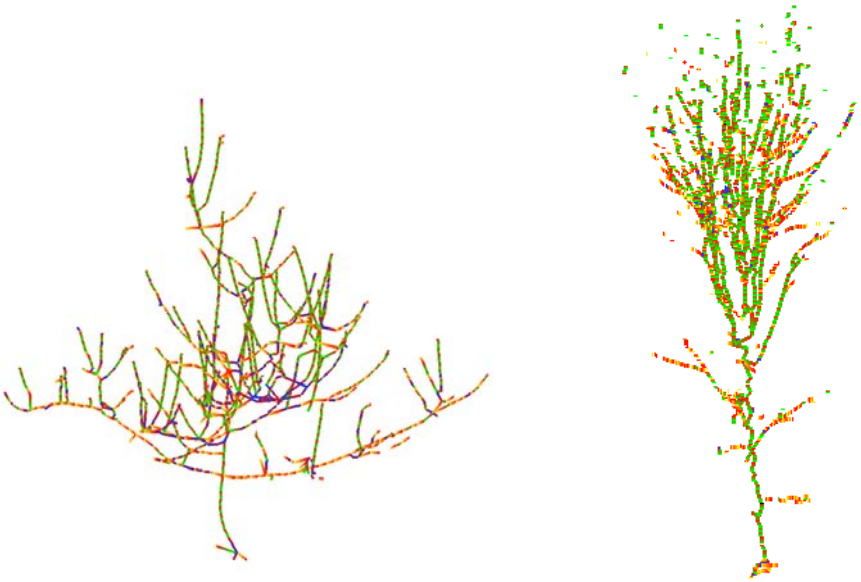


Figure 2: *Example Skeletons of an appletree (Apple 4) on the left and a beech tree on the right (Beech 2)*

It was shown in (Bucksch and Fleck 2010), which utilized the same six apple trees as this study, that the frequency distributions of the branch diameters show excellent correlation with manual measured field data while the detectable branch length is shorter than the manually measured one. For very thin branches (<1.5cm diameter) an insufficient amount of points for estimating the diameter was obtained. Because of that we assumed a constant diameter for all branches represented in the skeleton where no diameter could be obtained. In case of the six apple trees we assumed 1cm diameter. For the beech trees we assumed 5cm, because the distance between scanner and crown is much bigger than for the apple tree. This larger distance causes undersampling to be present on thicker branches (compare Figure 1 and Figure 2). These constant factors were applied to the difference between the total detected skeleton length and the skeleton length resulting in an estimated diameter. Noise can cause the skeletonization to fail and results in diameter estimations for finer branches in the upper crown much larger than the estimated trunk diameter. For this reason we rejected volume measurements bigger than the stem diameter.

3. Results

In Table 1 a one-to-one comparison of the collected field data and the automatic measurements is shown. The influence of the branch segmentation is minor. The results for the six apple trees show that the estimated biomass varies on average by 1.18 liters from the field measurements. The reason for the less good biomass extraction for apple can be found in the fact that a major branch was removed by pruning between laser and field measurement. For all apple trees the detected overall length of the branching system was between 60% and 85% of the manually measured length.

For the forest scenario no reliable biovolume estimation could be found. Though the used allometric relationship for beech is based on extensive measurements that were extremely well correlated (Rademacher et al. 2009), it cannot be excluded that it was not appropriate for the beech trees in our study, since allometric relationships are highly site specific and may vary substantially between regions. Also the estimated specific weight may not necessarily be accurate, especially for the fine branches of the investigated beech trees. The scanned beech trees were harvested after scanning and their biovoume was

additionally assessed using randomized branch sampling (RBS). The results of this assessment were not yet available for this summary, but will be reported on the conference.

Table 1: Comparison of the obtained biomass by manual field and automatic measurement. The measurements were corrected by assuming a 1cm diameter for the apple trees and 5cm for the beech trees where no diameter was estimatable.

Tree	Biovolume automatic measurement (uncorrected / corrected) in litres	Biovolume field measurement in litres and difference to the automatic measurement
Apple 1	8.5 / 11.8	10.3 (-1.5)
Apple 2	3.4/ 6.4	8.3 (+1.9)
Apple 3	1.2 / 2.4	2.8 (+0.4)
Apple 4	5.2 / 5.6	7.0 (+1.4)
Apple 5	18.6 / 20.4	18.7 (-1.7)
Apple 6	10.4 / 13.2	13.0 (-0.2)
Beech 1	939 / 1102	1446.0 (+344.0)
Beech 2	1394 / 1568	2740.0 (+1172.0)

4. Conclusions

From these prelimary results we conclude that skeletonization can be used to determine the biovolume of an individual tree. On average a differce of 1.18 litres to the manual measurement was found. Still, the results on forest trees have to be evaluated against manual field measurements that are carried out at the North-West German Forest Research Station. Further research will aim on the possibility to extract the branch diameters more robustly.

We found that the results of the biomass estimation on the apple trees may be further improved by investigating he influence of the resolution parameter in the skeleton extraction process.

References

BUCKSCH A., FLECK S., 2010, Automated detection of branch dimensions in woody skeletons of fruit tree canopies, Photogrammetric Engineering and Remote Sensing, Special Issue of Silvilaser 2009 (in print)

BUCKSCH, A., LINDENBERGH R., MENENTI M., 2010, SkelTre - Robust skeleton extraction from imperfect point clouds, The Visual Computer, DOI: 10.1007/s00371-010-0520-4

DRAKE, J.B., DUBAYAHA, R.O., CLARK D.B., KNOX, R.G., BLAIR, J.B., HOFTON, M.A., CHAZDON, R.L., WEISHAMPEL, J.F.,PRINCE, S.D., 2002

Estimation of tropical forest structural characteristics using large-footprint lidar

Remote sensing enviroment, 79 pp. 305-319

Fleck, S., C.G. Embree, D. Nichols, 2010. The systematic influence of Crop Load, Spur Type, 3D-Canopy Structure, and Leaf Zonal Chlorosis on Leaf Photosynthesis of 'Honeycrisp' Apple Trees. Acta Horticulturae (in print)

FLECK, S., VAN DER ZANDE, D, SCHMIDT, M, COPPIN, P., 2004. Reconstructions of tree structure from laser-scans and their use to predict physiological properties and processes in canopies. In Proceedings: ISPRSWG VIII/2 Workshop: Laser-Scanners for Forest and Landscape Assessment''. ISPRS International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVI-8/W2, Freiburg, Germany, 3–6 October, pp. 119–123

HENNING, J.G., RADTKE, P.J., 2006, Ground-based laser imaging for assessing three-dimensional forest canopy structure, Photogrammetric Engineering and Remote Sensing, 72(12), pp. 1349-1358

LEFSKY, M., MCHALE, M.R., 2008, Volume estimates of trees with complex architecture from terrestrial laser scanning. Journal of Applied Remote Sensing, 2(1).

PARKER, G., HARDING, D., BERGER, M., 2004. A portable LIDAR system for rapid

determination of forest canopy structure, Journal of Applied Ecology, 41, pp. 755–767.

POPESCU, S.C., 2007, Estimating biomass of individual pine trees using airborne lidar, Biomass and Bioenergy, 31, pp. 646–655

ROSELL, J.R., LLORENS, J., SANZ, R., ARNO, J., RIBES-DASI, M., MASIP, J., ESCOLA, A., CAMP F., SOLANELLES, F., GRACIA, F., GIL E., VAL, L., PLANAS, S., PALACIN, J.,2009, Obtaining the three-dimensional structure of tree orchards from remote 2D terrestrial LIDAR scanning, Agricultural and Forest Meteorology,149(9), pp. 1505-1515

Rademacher, P., Khanna, P.K., Eichhorn, J., Guericke, M., 2009, Tree Growth, Biomass, and Elements in Tree Components of Three Beech Sites. In: Brumme, R., Khanna, P. K. (Eds.): Functioning and Management of European Beech Ecosystems. Ecological Studies 208, 105-136

STRAHLER, A.H, JUPP, D.L.B., WOODCOCK, C.E., SCHAAF, C.B., YAO, T., ZHAO, F., YANG, X., LOVELL, J., CULVENOR, D., NEWHAM, G., NI-MIESTER, W., BOYKIN-MORRIS, W.,2008, Retrieval of Forest Structural Parameters Using a Ground-Based Lidar Instrument(Echidna®), .Canadian Journal of Remote Sensing, 34(2), pp.426-440.

THIES, M.; PFEIFER, N.; WINTERHALDER, D., GORTE, B.G.H., 2004, Three-dimensional reconstruction of stems for assessment of taper, sweep and lean based on laser scanning of standing trees, Scandinavian Journal of Forest Research, 19(6), pp. 571-581

VAN LEEUWEN, M., NIEUWENHUIS, M., 2010, Retrieval of forest structural parameters using LiDAR remote sensing, European Journal of Forest Research, 129 (4), pp. 749-770

The potential of airborne laser scanning for monitoring the subalpine birch forest ecotone

MATTIAS NYSTRÖM†, JOHAN HOLMGREN†, and HÅKAN OLSSON†

mattias.nystrom@srh.slu.se

† Department of Forest Resource Management, Swedish University of Agricultural Sciences, SE-90183 Umeå, Sweden

Abstract

Climate changes, insect damages, and browsing pressure by herbivores are causing changes of the sub arctic tree line, as well as in the presence of low shrubs of salix species in the mountain heaths above the tree line. These changes are however difficult to monitor with manual methods because of the complex mosaic pattern of the forest-tundra ecotone. The possibility to map this ecotone with airborne laser scanning (ALS) has therefore been tested at a site 6 km SE Abisko in northern Sweden (Lat N 68° 20', Long E 19° 1', 400-800 m a.s.l.). ALS point cloud data above ground and 10 m radius field plots were used to estimate regression models for maximum tree height, above ground tree biomass, and canopy closure, with relative RMSE 9.3%, 25.1% and 23.6%, respectively. Preliminary results show a strong potential of ALS as a future method to monitor changes in the arctic tree-line ecotone.

1. Introduction

The sub arctic and alpine tree lines in northern Scandinavia constitutes most often of birch forest dominated by mountain birch (*Betula pubescence spp. Czerepanovii*). These forests are characterized by low, often bent, trees with slow growth and umbrella-like canopies. The trees have limited commercial value and not much systematic monitoring of them have been done in the past (Tømmervik *et al.* 2009). There is however a growing interest especially for monitoring of the tree line, which is supposed to move to higher altitudes, and further north with a warmer climate (Kullman 1998, 2010). A sometimes predicted dramatic change of the tree line would affect animals and plants, as well as existing land use like reindeer herding and tourism. Tree line changes in the north would also influence global carbon budgets (Sjögersten and Wookey 2009). In addition to climate, the position of the tree line is also influenced by other factors, like browsing by moose and reindeer, human actions, and insect attacks. Especially the Autumnal Moth (*Epirrata autumnata*) will defoliate large areas with intervals of about a decade. The above factors, and in particular the local climate, will at many places contribute to a border between birch forest and mountain heath that has a very complex mosaic pattern. This border is difficult to monitor with pure field methods. The use of manual photo interpretation is also unreliable and dependent on the interpreter (Heiskanen *et al.* 2008). Optical satellite data could be used for large area overviews of the taiga-tundra tree line (Hill *et al.* 2007). Satellite data will however not provide a detailed assessment of the structure, biomass and height of the tree line ecotone, neither will single tree be detected with this method.

Early studies have already shown that scattered trees on the mountain heath above the forest can be detected by laser scanning (Næsset and Nelson 2007, Rees 2007). There is however also an interest to monitor the structure of the tree line ecotone, including biomass, tree height and canopy closure for the tree vegetation in the transition zone between open mountain heath and closed mountain birch forest. A multitude of earlier studies have showed that laser scanning works well for estimating the managed boreal forests in Scandinavia using empirical modeling of laser data features as explanatory variables for forest features measured on field plots (Næsset *et al.* 2004). The aim of this paper is to investigate how well a similar approach will work when applied to mountain birch forest in the ecotone towards the open mountain heath.

2. Materials and methods

2.1 Study area

The study area is located 6 km SE Abisko in the northern Sweden, centred around Lat N 68°20', Long E 19°1', 440-700 m a.s.l. (Figure 1). The forest in the study area is subalpine birch forest dominated by mountain birch *(Betula Pubescens ssp czerepanovii)* but a few Junipers, Rowens and Salix taller than 2 m could also be found. The birches in the area was of the multi-stem type with often several stems sharing the she same root system. Sample plots were placed in a smaller area (0.7 km^2) in the ecotone between birch forest and tundra. This area was characterized of a pattern of forest and mountain heath vegetation and located at an altitude of 500- 700 m a.s.l.

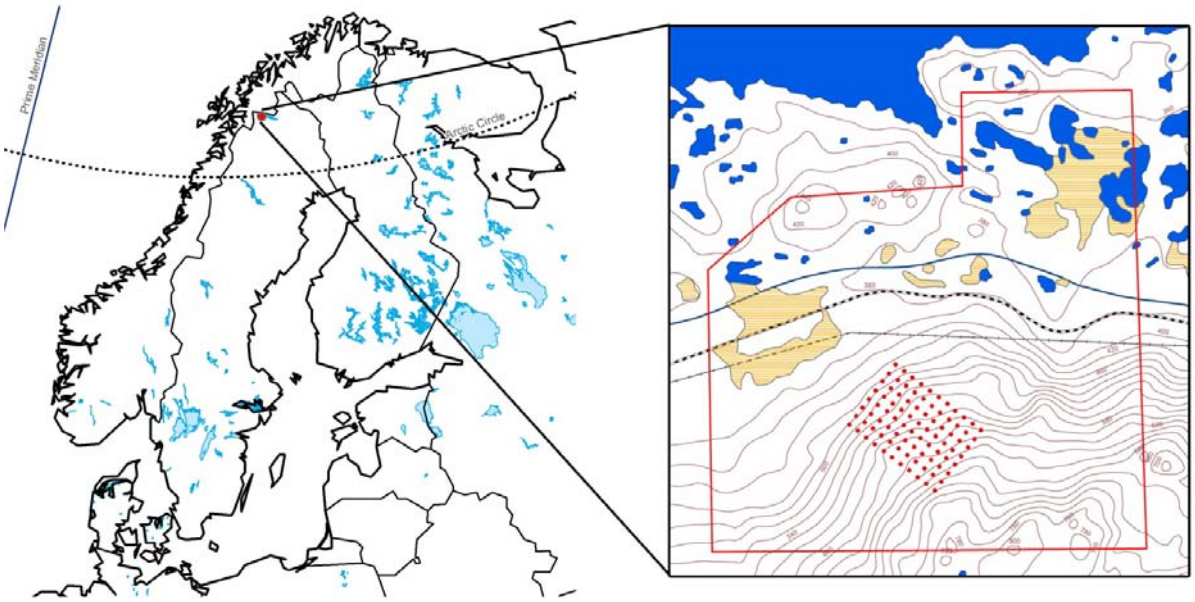


Figure 1. *Location of the test area, the red polygon in the right map is the laser scanned area and the regular pattern of dots is the sample plots.*

2.2 Field data

The field inventory was performed one year after the laser scanning, during four weeks in August 2009. A 100 m spaced grid of 88 sample plots was surveyed and marked permanently to enable future re-inventory. The coordinates for the north-west corner of the grid was randomly selected inside a 100x100 m rectangle. The grid was aligned orthogonal to the main slope direction. In total 31 of the sample plots had no trees at all taller than 2 m. The reason for field inventing empty sample plots was to document the whole transition zone between forest and heath, as well as to be able to come back in the future and detect if new trees have been established.

A Garmin eTrex Vista HCx GPS was used to navigate to the sample plots. When 10 m remained to the plot centre, navigation with GPS was cancelled and a measuring tape was used to navigate the last 10 m. This method was used to avoid bias caused by the surveyor. The plot centre position was measured with an RTK-GPS (Trimble R7) with the base station placed 6 km away, at the research station in Abisko.

The sample plots had a radius of 10 m and the border was measured with an ultrasonic positioning system (UPS). Trees with a common root system or overlapping crowns were grouped. For each group, position was measured with the UPS and maximum tree height was measured with a measuring pole. All tree stems taller than 2 m and with at least 30 mm DBH, i.e. stem diameter at 1.3 m along the stem from ground, was callipered and linked to the main tree in the tree group. The number of stems taller than 2 m but less than

30 mm at DBH was counted. Width and length of an ellipsoid describing the outline of the canopy for the tree group was measured for every fifth group with a measuring tape. A linear regression model was used to estimate the crown area for all tree groups using the following explanatory variables: tree basal area, number of trees in the group, tree height, and number of trees taller than 2 m but less than 30 mm DBH. The total canopy closure (CC) for a sample plot was calculated as the sum of the area of all ellipsoids divided by the sample plot area. Type and coverage of understory vegetation was noted for all plots.

Functions for estimating above ground tree biomass (AGB) from measures of the tree stems existed from previous studies in the Abisko area (Dahlberg *et al*. 2004). Two different biomass functions were used: for trees with field measured height, DBH and tree height was used as explanatory variables in the function, otherwise, where tree height not was measured, only DBH was used.

Table 1. *Summary statistics of the field data on the 57 non-empty sample plots.*

Definition	Unit	Abbr.	Mean	Min[a]	Max	N
Stem diameter	mm	DBH	65	30	266	1698
Stem density [b]	ha^{-1}		948	32	2165	57
Maximum tree height [c]	m	MH	6.6	2.4	10.5	57
Biomass [d]	$t\ ha^{-1}$	AGB	13.8	0.06	44.1	57
Canopy closure [c]	-	CC	0.392	0.005	1.031	57

1. [a] the lowest non-zero value
2. [b] Stems taller than 2 m and >30 mm DBH.
3. [c] On sample plot level.
4. [d] Above ground tree biomass on sample plot level.

2.3 Laser data acquisition

A TopEye MkII laser scanner, operated by Blom Sweden, mounted on a helicopter was used to collect laser data. The scanning was performed on August 1, 2008 under leaf-on conditions. Flight altitude was about 500 m above ground, which resulted in a foot-print size of 50 cm on ground. The pulse repetition frequency was 50 kHz, the scan frequency 35 Hz. The wavelength was 1064 nm and the length of each pulse was 4 ns (=1.2 m). The scan pattern was an ellipsoidal Palmer scan with swath ±20 degrees and ±14 degrees along the flight lines both measured from the vertical line. Flight lines were performed in north-west direction and overlapping data has not been modified. First and last echo were registered for each emitted pulse. The average point density on the sample plots was 13 m^{-2}, minimum 3.6 m^{-2}, and maximum 32 m^{-2}.

2.4 Laser data processing

The laser data were pre-processed by the contractor and delivered as LAS-files. TerraScan (Soininen 2010) from Terrasolid was used to classify the point cloud into ground and non ground points (Axelsson 1999, 2000). A maximum terrain angle of 88 degrees and an iteration angle of 6 degrees was used. The statistical software R and in house developed programs was used to further process the point cloud.

A Digital Elevation Model (DEM) with 0.5 meter grid size, representing the ground level, was created. A Digital Surface Model (DSM) was created by assigning each raster cell the maximum z-value for laser returns classified as vegetation by TerraScan. A normalised DSM (nDSM) was calculated by subtracting the DEM from the DSM. Canopy heights (CH) were, as an alternative, also calculated by subtracting the DEM from the z-value of each laser return.

Percentiles (P_{xx}) in steps of 5% were calculated on the nDSM, using values above a height threshold defined as the maximum value of 1 m and 10% of the maximum nDSM value inside the sample plot. Percentiles were also calculated on the CH-vector.

Vegetation ratio (VR) was determined from the CH-vector by dividing number of hits above 2 m with the total number of returns inside the sample plot. Standard deviation (SD) and mean vegetation height (MVH) was calculated on the nDSM and on the CH-vector.

2.5 Model development

The correspondence between field plot data and measures from the ALS point cloud above ground were used to estimate regression models for the dependent variables maximum height (MH), above ground tree biomass (AGB), and canopy closure (CC). Both dependent variables and independent variables were ln-transformed in the following way: $ln(X+1)$. The ln-transform was not used for the maximum height model.

Selection of independent variables to the models was in a first step done using best subset regression. In the second step the number of independent variables was reduced to not be correlated. Variance Inflation Factors (VIF) was used to check multicollinearity for the independent variables.

Candidates as independent variables were: percentiles (P_{xx}^{nDSM} or P_{xx}^{CH}), vegetation ratio (VR), mean vegetation height (MVH^{nDSM} or MVH^{CH}), and standard deviation (SD^{nDSM} or SD^{CH}) of vegetation height.

The models on sample plot level for maximum height (MH), above ground tree biomass (AGB), and canopy closure (CC) were estimated with equations (1)– (3).

$$MH = 0.575 + 1.106 \cdot P_{95}^{CH} \quad (1)$$

$$ln(AGB+1) = -0.200 + 2.114 \cdot ln(SD^{nDSM}+1) + 2.447 \cdot ln(VR_1+1) \quad (2)$$

$$ln(CC+1) = 0.127 + 0.449 \cdot ln(Mean^{nDSM}+1) - 0.180 \cdot ln(P_{20}^{CH}+1) \quad (3)$$

Equation (2) and (3) were back-transformed in prior to evaluation. Logarithmic bias was corrected by multiplying with a correction coefficient (Holm 1977):

$$q = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} \hat{y}_i} \quad (4)$$

where y_i are the field measured values and $\hat{y}_i$ the estimated values. RMSE was calculated using leave one out cross-validation on sample plot level.

2.6 Raster maps

Raster maps were calculated from the models (equation (1), (2), and (3)). The pixels were 10×10 m. Each pixel was calculated from laser data within a circular area with 10 m in radius, centred in the pixel centre.

3. Results

Regression models with low errors and high coefficients of determination could be constructed using as little as maximum two independent variables (Table 2). More than 50 independent variables, were tested for inclusion in the models, the few variables in the final models were selected based on their significance and limited correlation. The model for maximum height had, as expected, the lowest relative RMSE (Table 2). A

few outliers could be observed for all predicted variables (Figure 2). In the raster map one can observe higher biomass in steeper terrain and the mosaic patterned tree-line ecotone (Figure 3).

Table 2. *Summary of the models (equation (1), (2), and (3)).*

Dependent variable	Unit	Indep. variables, significance [a]		R^2 adj.	RMSE	rRMSE
Maximum height	m	P_{95}^{CH} ***		0.92	0.61	9.3 %
Biomass [b]	t ha^{-1}	SD^{nDSM} ***	VR ***	0.97	3.47	25.1 %
Canopy closure	-	MVH^{nDSM} ***	P_{20}^{CH} **	0.91	0.093	23.6 %

[a] Abbreviation of significance (p): *=0.1, **=0.01, ***=0.001

[b] Above ground tree biomass

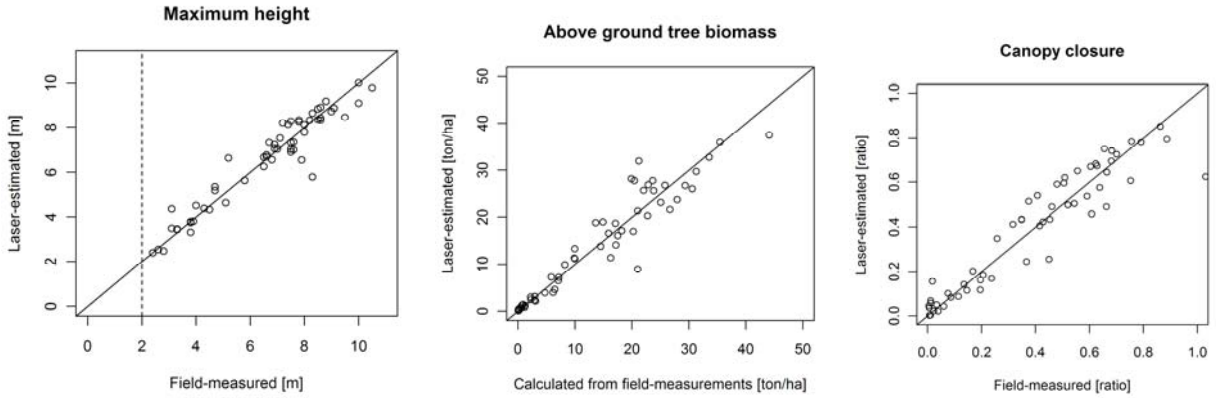


Figure 2. *One-to-one plots of field based data versus laser estimates for maximum height (left), above ground tree biomass (middle), and canopy closure (right). Sample plots with trees below two meters were removed (dashed vertical line in the left plot).*

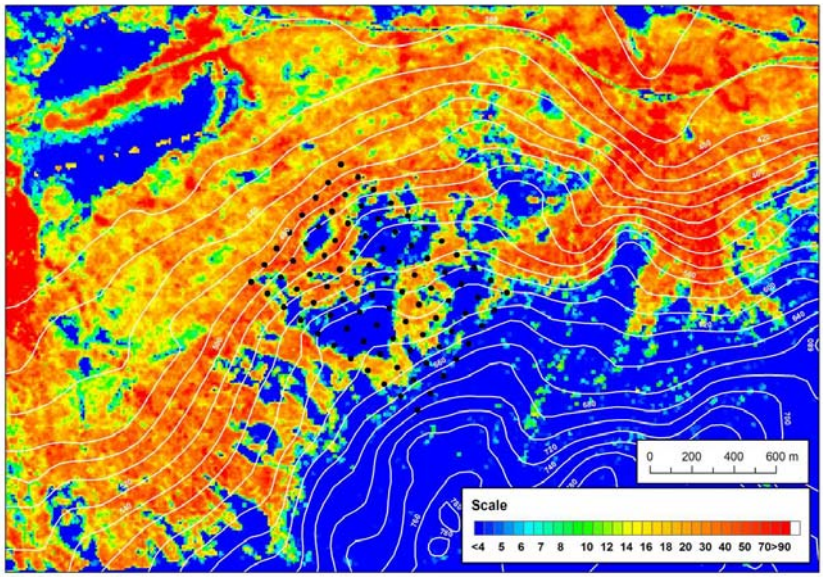


Figure 3. *Example of a raster map derived from the model (equation (2)) for above ground tree biomass (t ha^{-1}). The black dots are the sample plots. Remark: This map is extrapolated outside the area calibrated with field plots.*

4. Discussion

In this study, maximum height, above ground tree biomass, and canopy closure for mountain birch forest were sucessfully estimated from the ALS point cloud data. This is to our knowledge the first time that ALS based biomass estimates for these forests has been reported, since earlier ALS studies of mountain birch has concentrated on the existence of forested areas (Rees 2007), and the existence of pioneer trees (Næsset and Nelson 2007, Næsset 2009). In spite of the irregular shape of the stems which also grows in gropus and most often are leaning, and the the umbrella like tree canopies, surprisingly good results were achived. The plot level accuracy obtained is only slightly worse than that obtained in the managed boreal forest with taller and more regularly shaped trees.

In one sample plot, the canopy closure estimated from field data is slightly above 1. The reason is that canopy area was only measured for every fifth tree group and calculated using a regression model to the others. In this sample plot there were several trees with large DBH and therefore high assigned canopy area.

The model for maximum height has the lowest RMSE. This can be explained by the fact that heights are directly measured both by the laser scanner, and in field. The biomass estimates were based on the correlation between laser data point cloud features and an allometric biomass model driven by field surveyed stem diameter measurements. These allometric relationships are not as well behaved in the mountain birch forest as in the normal boreal forest (Dahlberg *et al.* 2004). In the case of canopy closure, it was a bit surprising that the best function not included the vegetation ratio; the largest error in the canopy closure model is however most likely due to uncertainties in the field estimates of canopy closure.

The raster map show the mosaic patterned transition between forest and mountain heath. It is evident from these maps that it would be difficult to draw a line defined as the tree-line. Another interesting notation in the raster map is that the trees appear to be taller and the forest appears to have higher biomass in steeper terrain.

As seen, airborne laser scanning have high potential to become a method to monitor the subalpine birch forest ecotone. In order to test a monitoring system for the mountain birch forest, there is a need to continue with multitemporal studies.

Acknowledgments

This study is part of the research programme Environmental Mapping and Monitoring with Airborne laser and digital images (EMMA), financed by the Swedish Environmental Protection Agency. The laser data was acquired through a cooperation with the university of Lund. The Abisko Scientific Research Station is acknowledged for supporting the work with field data acquisition.

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.

DAHLBERG, U., BERGE, T. W., PETERSSON, H., and 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.

HEISKANEN, J., NILSSON, B., MÄKI, A.-H., ALLARD, A., MOEN, J., HOLM, S., SUNDQUIST, S., and 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 managemenat and geomatics, Working report 242, 59 p.

HILL, R. A., GRANICA, K., SMITH, G. M., and 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).

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., 2010, A Richer, Greener and Smaller Alpine World: Review and Projection of Warming-Induced Plant Cover Change in the Swedish Scandes. Ambio, 39, 159-169.

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., GOBAKKEN, T., HOLMGREN, J., HYYPPA, H., HYYPPA, J., MALTAMO, M., NILSSON, M., OLSSON, H., PERSSON, A., and SODERMAN, U., 2004, Laser scanning of forest resources: The Nordic experience. Scandinavian Journal of Forest Research, 19, 482-499.

NÆSSET, E., and NELSON, R., 2007, Using airborne laser scanning to monitor tree migration in the boreal-alpine transition zone. Remote Sensing of Environment, 110, 357-369.

REES, W. G., 2007, Characterisation of Arctic treelines by LiDAR and multispectral imagery. Polar Record, 43, 345-352.

SJÖGERSTEN, S., and 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.

TØMMERVIK, H., JOHANSEN, B., RISETH, J. A., KARLSEN, S. R., SOLBERG, B., and HOGDA, 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.

Predicting needle losses of individual Scots pines from airborne laser point clouds

Kantola, T.[*1], *Vastaranta, M.*[1], *Xiaowei, Y.*[2], *Lyytikäinen-Saarenmaa, P.*[1], *Holopainen, M.*[1], *Talvitie, M.*[1], *Kaasalainen, S.*[2], *Solberg, S.*[3], *Hyyppä, J.*[2]

tuula.kantola@helsinki.fi

1University of Helsinki, Department of Forest Sciences, Finland, tuula.kantola@helsinki.fi, mikko.vastaranta@helsinki.fi, paivi.lyytikainen-saarenmaa@helsinki.fi, markus.holopainen@helsinki.fi., mervi.talvitie@helsinki.fi

2 Finnish Geodetic Institute, yu.xiaowei@fgi.fi, sanna.kaasalainen@fgi.fi, juha.hyyppa@fgi.fi.

3Norwegian Forest and Landscape Institute, Norway

Svein.Solberg@skogoglandskap.no

Abstract

Climate change has been observed to be related to the increase of forest insect damages in the boreal zone. The prediction of the changes in the distribution of insect-caused forest damages has become a topical issue in the field of forest research. The common pine sawfly (*Diprion pini* L.) (Hymenoptera, Diprionidae) is regarded as a significant threat to boreal pine forests. Defoliation by D. pini caused severe growth losses and tree mortality of Scots pine (*Pinus sylvestris* L.) (Pinaceae). Logistic regression is commonly used in modelling the probability of occurrence of an event. In this study the logistic regression was investigated for predicting the needle loss of individual Scots pines (pine) using the features derived from airborne laser scanning (ALS) data. The defoliation level of 164 trees was determined subjectively in the field. Statistical ALS features were extracted for single trees and used as independent variables in logistic regression models. Classification accuracy of defoliation was 87.8% as respective kappa-value was 0.82. For comparison, only penetration features were selected and classification accuracy of 78.0% was achieved (kappa=0.56). Based on the results, it is concluded that ALS based prediction of needle losses is capable to provide accurate estimates for individual trees.

Keywords: ALS, needle loss, Diprion pini, logistic regression, forest disturbances

1. Introduction

In Finland, approximately 76% of the land surface is forested, which is the highest percentage amongst the European countries. The forests have a high impact on the Finnish economy and environment. The Climate change has caused rising temperatures throughout the world in the present. Ecological balance has been interrupted, causing among other things pest damages in managed forests (Dale et al. 2001, Evans et al. 2002). The speed of change seems to be even higher at higher latitudes, as in Scandinavia.

Forest insects, formerly regarded as harmless species, are now chancing into pest status and causing serious damages in Finland (Lyytikäinen-Saarenmaa and Tomppo 2002, de Somviele et al. 2007). Economic losses caused by defoliators can be considerable, approximately 300–1000 eur perha, depending on intensity of needle loss. It can require over a one decade for a tree to fully recover (Lyytikäinen-Saarenmaa et al. 2006).

Systems for a rapid assessment of forest areas affected by hazardous events and the monitoring of those areas gain high importance (Lyytikäinen-Saarenmaa et al. 2008). The symptoms of forest damage caused by

insects are in many cases visible from far distances. Some types of forest damage such as crown discoloration, defoliation and dieback are even more visible from the bird's eye view than from the ground. Especially a better assessment of the dispersion and range extension pattern is possible based on remote sensing (RS) data. RS can produce data for large areas of remote, inaccessible forest lands quickly and at a much lower cost than ground surveys (Ciesla 2000, Hall et al. 2007).

Recent development of RS technologies in particular airborne laser scanning (ALS) techniques has given new perspectives to the forest inventory and monitoring. With the capability of directly measuring forest structure, including canopy height and crown dimensions, ALS is increasingly used for forest inventories at different levels. Previous studies have shown that ALS-data can be used to estimate a variety of forest inventory attributes including tree, plot and stand level estimates for tree height (Falkowski et al. 2006, Hyyppä and Inkinen 1999, Magnussen et al. 1999, Maltamo et al. 2004), biomass (Bortolot and Wynne 2005, Lefsky et al. 1999, van Aardt et al. 2008), volume (Hyyppä et al. 2001, Næsset 1997, Wallerman and Holmgren 2007), basal area (Lefsky et al. 1999, Means et al. 2000, Næsset 2002) and tree species (Brandtberg 2007, Holmgren and Persson 2004, van Aardt et al. 2008). ALS is also promising method for predicting, detecting and monitoring forest hazards, e.g. defoliation because of its capability of direct measurement of vegetation structure. There are several recent studies and method developments to achieve more accurate ALS-based biomass detection (Sohlberg et al. 2006, Zhao et al. 2009, Hawbaker et al. 2009). The accuracies for detection of defoliated trees has not been reported yet, although single trees biomass and defoliation level are highly correlated (see e.g. Hyyppä et al. 2009).

The aim of the present study was to test the accuracy of the needle loss predictions determined from the ALS data with single trees canopy-based statistical features. Logistic regression models were used in the estimation.

2. Material and methods

2.1. Study area

Study was performed in Ilomantsi district, in Eastern Finland (62°53´, 30°54´, fig. 1). The study area was total of 34.5 km^2. The study area was located in a region where *D. pini* has caused considerable damages in an area of 10 000 ha during past 10 years. The inventory was taken during May and June 2009. The forests in the study area are mainly pure pine stands growing on dry to dryish sandy soil sites. The majority of the stands are young or middle-aged forests having the average age of 53 years and diameter of 14.7 cm, respectively.

2.2. Field measurements

The visual assessment of defoliation percentage was carried out for every individual Scots pine. A sample of 164 of pines were selected to this study. The trees were located with a Trimble Pro XH (Trimble Navigation Ltd., Sunnyvale, CA, USA), which can reach up to 30 cm precision.The intensity of defoliation of a single tree was visually estimated from different directions according to Eichhorn (1998), comparing the tree under investigation to a reference tree with full healthy foliage in the same site type. In this study an accuracy of 10% was used. The defoliation percentage of 20 % was determined as a boundary of severe needle loss. The number of trees with severe needle loss was 84 as 80 trees were classified as healthy trees. Tree heights describing the size distribution with in classes in presented figure 2.

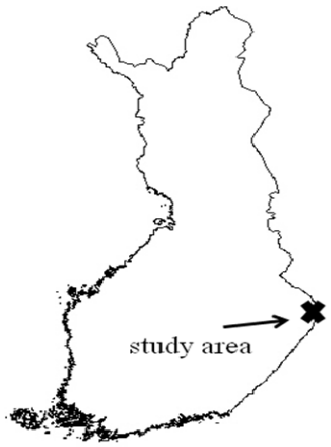


Figure 1:Map of the study area

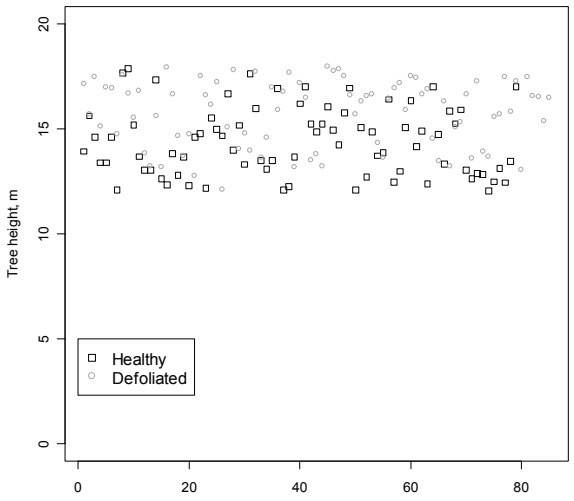


Figure 2. Plotted heights of healthy and defoliated trees. X-axel presents the number of a tree in class i and y-axel presents the height of a tree in class i.

2.3. Remote sensing materials

The ALS data were acquired on July 2008 with a Leica ALS50-II SN058 laser scanner (Leica Geosystem AG, Heerbrugg, Switzerland). The flying altitude was 500 m at speed of 80 knots, a field of view of 30 degrees, a pulse rate of 150 kHz and a scan rate of 52 Hz. The density of the returned pulses within the field plots was approximately 20 pulses per m^2.

ALS data were first classified into ground or non-ground points using the standard approach of the TerraScan based on the method explained by Axelsson (2000). A digital terrain model was created by using classified ground points. Laser heights above ground (normalized height or canopy height) were calculated by subtracting ground elevation from corresponding laser measurements. Canopy heights close to zero were considered as ground returns and heights greater than 2 m were considered as vegetation returns. The laser returns between them were considered as returns from ground vegetation or bushes. Only the returns from vegetation were used for tree feature extraction.

2.4. Tree detection and feature extraction

A raster canopy height model (CHM) was created from normalized data for individual tree detection and crown segmentation. Single tree segmentations were performed on CHM images using a minimum curvature-based region detector. During the segmentation processes, the tree crown shape and location of individual trees were determined. The procedure consisted of the following steps:

1. CHM was smoothed with a Gaussian filter to remove small variations on crown surface. The degree of smooth is determined by the value of standard deviation (Gaussian scale) and kernel size of the filter.
2. Minimum curvatures were calculated. Minimum curvature is one of the principal curvatures. For a surface like CHM, higher value of minimum curvature describes tree top.
3. The smoothed CHM image was then scaled based on the computed minimum curvature resulting in a smoothed yet contrast stretched image.

4. Local maxima were then searched in a given neighbourhood. They were considered as tree tops and used as markers in the following marker controlled watershed transformation for tree crown delineations .

Each segment was considered to present a single tree crown. Laser returns falling within each individual tree segment were extracted and canopy height of these returns were used for deriving tree features (Table 1).

Table 1. Features extracted from ALS data for trees

Feature	Description
Hmax	Maximum laser height
Hmean	Arithmetic mean of laser heights
Hstd	Proportion of vegetation hits
CH	Crown height
CA	Crown area as a convex hull
CV	Crown area as a convex hull in 3D
HP10-90	Heights 0th-90th percentile
DS10-90	Percentage of returns below 10-90% of total height
MaxD	Maximum crown diameter when crown was considered as a ellipse

2.5. Logistic regression

The probability of a defoliation class of a single tree was modelled with multiple logistic regression using the function glm in the R statistical package (R Development Core Team, 2007). Logistic regression is commonly used in modelling the probability of occurrence of an event. In logistic regression, logit transformation is used to make the relationship between the response probability and the explanatory variables linear. The multiple logistic regression model is expressed as:

$$\text{logit}(p) = \ln[p/(1-p)] = \beta 0 + \beta 1x1 + \beta 2x2 + \ldots\ldots\ldots + \beta nxn \quad , \qquad (1)$$

,where p is a probability for an event to occur and x1…xn are the variables explaining the probability. The predicted probabilities are calculated by transforming back to the original scale:

$$p = \exp(\text{logit}(p))/[1 + \exp(\text{logit}(p))]. \qquad (2)$$

For selecting the independent variables in the models, stepwise logistic regression was applied. Classification accuracy (%) and kappa-values were calculated for accuracy observation.

3. Results

The defoliation levels of single trees were predicted by using two different logistic regression models. In the first model all the features (table 1) were used as preliminary variables. The best features of them were chosen using stepwise logistic regression (table 2). The classification accuracy of the model was 87.8% and respective kappa-value was 0.82.

Table 2. Parameters used in the model were all the extracted ALS-features were used as a preliminary variable

	Estimate	Std. Error	z value	Pr(> |z|)
Intercept	-20.2010	3.5325	-5.72	0.0000
Hmax	-1.9063	0.4327	-4.41	0.0000
Hmean	3.0903	0.5702	5.42	0.0000
Hstd	2.4238	0.6193	3.91	0.0001
DS40	-15.1475	6.5149	-2.33	0.0201
DS60	29.7942	7.3697	4.04	0.0001

In the second model, only the percentages of returns below certain relative heights (10-90%) were used as preliminary variables. The assumption was that there is more laser pulse returns from lower relative heights as the defoliation level increases. The classification accuracy of the model (Table 3.) was 78.0% (kappa-value = 0.56). DS features and predicted defoliation probability are plotted in figure 3. Correlations with predicted defoliation probability and DS10, DS30, DS50, DS70 and DS90 were 0.62, 0.63, 0.64, 0.35 and -0.22 respectively.

Table 3. Parameters used in the model were only the percentages of returns below certain relative heights were used as a preliminary variable

	Estimate	Std. Error	z value	Pr(> |z|)
Intercept	-1.9754	0.9645	-2.05	0.0406
DS60	18.2783	2.8710	6.37	0.0000
DS80	-9.0043	1.9887	-4.53	0.0000

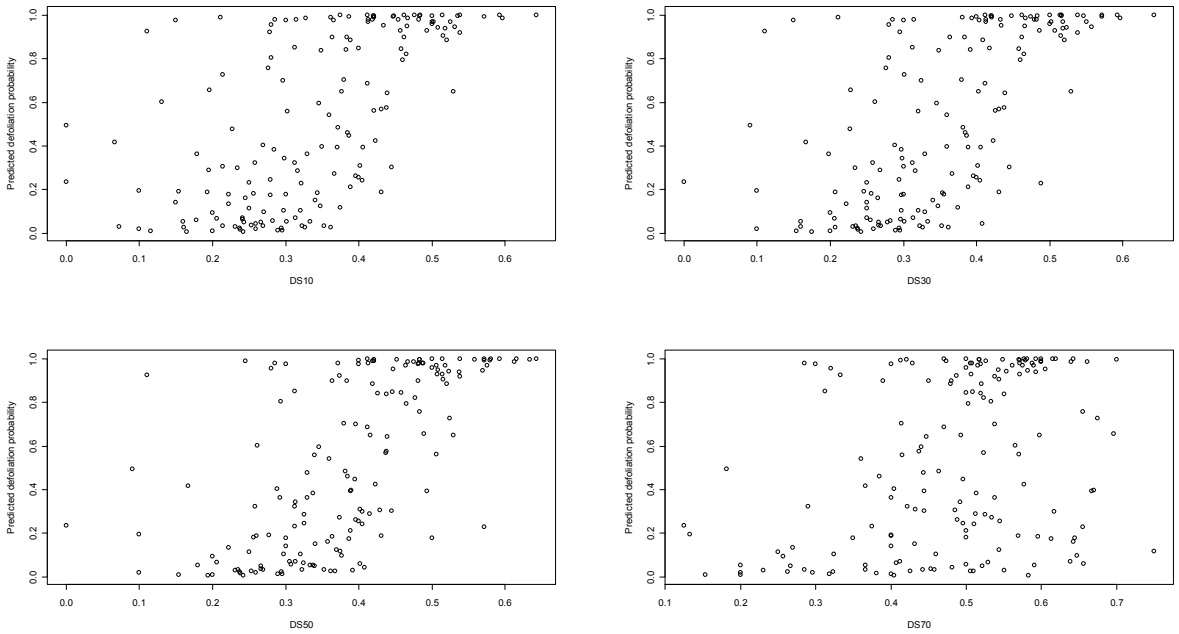


Figure 3. Plotted DS features (from 10%, 30%, 50% and 70% heights)

vs. predicted defoliation probability.

4. Discussion

In this study the statistical ALS features were tested in classification of defoliation of individual trees. The logistic regression was used in estimation. The analysed trees were classified into two different classes, defoliated and healthy trees. This classification enables to test if it is possible to detect defoliation from ALS data.

Both health and defoliated trees were collected from the same study area. It was typical in the study area that taller and older trees in dominant canopy strata were more defoliated than younger ones. To achieve reliable results, affect of tree size and age had to be eliminated. From that reason the analysis was limited to a smaller sample set (164 trees) to get same sized and aged trees to the both classes. In this study the defoliation level was estimated visually, which could easily cause deviation in results if the surveyor were not a professional. Naked-eye calibration is essential when two or more researchers are estimating the critical variable.

According to the study results it may be possible to detect the trees and forest stands with high level of defoliation. This may be important information for detecting and mapping insect damages which are usually rare and clustered phenomena. It may also be meaningful auxiliary information for improving the inventory of forest damages. For example with RS data, the stratification could be carried out by focusing on more plots in areas where pest damage could already be detected from the material.

Further studies will be focused on test the classification accuracy with more defoliation classes. Although, from practical point of view, it is most important to detect areas severe defoliation so the use of only two classes in our first tests were justified. When several defoliation classes are used, non parametric estimation methods will also be tested. The optimal feature extraction and selection for this kind of purpose should be studied further. In Finland, when ALS data is collected for practical forest inventory, same-date aerial images are also gathered. The use of spectral values extracted from the aerial images or ALS intensity could enhance the classification accuracy.

References

Axelsson, P. 2000. DEM generation from laser scanner data using adaptive TIN models. International Archives of Photogrammetry and Remote Sensing. Amsterdam, 16–23 July 2000. Vol. XXXIII(B4):110–117.

Bortolot, Z. and Wynne, R.H. 2005. Estimating forest biomass using small footprint LiDar data: An individual tree-based approach that incorporates training data. ISPRS Journal of Photogrammetry & Remote Sensing 59: 342–360.

Brandtberg, T. 2007. Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar. ISPRS Journal of Photogrammetry & Remote Sensing 61: 325–340.

Ciesla, W. 2000. Remote sensing in forest health protection. USDA Forest Service Remote Sensing Applications Center Salt Lake City, UT and Forest Health Technology Enterprise Team Fort Collins, CO.

Dale V.H., Joyce L.A., MCNulty S., Neilson R.P., Ayres M.P., Flannican M.D., Hanson P.J., Irland L.C., Lugo A.E., Peterson C.J., Simberloff D., Swanson F.J., Stocks B.J. and Wotton B.M. (2001): Climate change and forest disturbances. Bioscience 51: 723–734

De Somviele, B., Lyytikäinen-Saarenmaa, P. and Niemelä, P. (2007): Stand edge effects on distribution and condition of Diprionid sawflies. Agricultural and Forest Entomology 9:17–30.

Eichhorn, J. (1998): Manual on Methods and Criteria for Harmonized Sampling, Assessment, Monitoring and Analysis of the Effects of Air Pollution on Forests. Part II. Visual Assessment of Crown Condition and Submanual on Visual Assessment of Crown Condition on Intensive Monitoring Plots. United Nations Economic Commission for Europe Convention on Long-range Transboundary Air Pollution, Germany.

Evans, H., Straw, N. and Watt, A. (2002): Climate change: Implication for insect pests. In: M. roadmeadow (eds.) Climate Change: Impact on UK Forests. Forestry Commisssion Bulletin 125: 99–118

Falkowski, M.J., Smith, A.M.S., Hudak, A.T., Gessler, P.E., Vierling, L.A., Crookston, N.L. 2006. Automated estimation of individual conifer tree height and crown diameter via two-dimensional spatial wavelet analysis of lidar data. Canadian Journal of Remote Sensing 32(2): 153–161.

Hall, R.J., Skakun, R.S. & Arsenault, E.J. 2007. Remotely sensed data in the mapping of insect defoliation. In: Understanding Forest Disturbance and Spatial Pattern. Remote Sensing and GIS Approaches. (Michael A. Wulder & Steven E. Franklin, eds.). CRC Press, Taylor & Francis Group, Boca Raton. pp. 85–111.

Hawbaker, T.J., Keuler, N. S., Lesak, A.A., Gobakken, T., Contrucci, K., and Radeloff V.C. 2009. Improved estimates of forest vegetation structure and biomass with a LiDAR-optimized sampling design. *J. Geophys. Res.* 114, G00E04.

Holmgren, J., Persson, Å. 2004. Identifying species of individual trees using airborne laser scanner. Remote Sensing of Environment 90:415–423.

Hyyppä, J., Inkinen, M. 1999. Detecting and estimating attributes for single trees using laser scanner. The Photogrammetric Journal of Finland 16: 27–42.

Hyyppä, J., Kelle, O., Lehikoinen, M., Inkinen, M. 2001. A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. IEEE Transactions on Geoscience and Remote Sensing 39(5):969–975.

Hyyppä, J., Jaakkola, A., Hyyppä, H., Kaartinen, H., Kukko, A., Holopainen, M., Zhu, L., Matikainen, L., Chen, R., Chen, Y., Kaasalainen, S. Krooks, A. Litkey, P., Rönnholm, P., Vastaranta, M. & Lyytikäinen-Saarenmaa, P. 2009. Map Updating and Change Detection Using Vehicle-Based Laser Scanning, in proceedings of JURSE 2009, 20–22 May 2009.

Lefsky, M.A., Harding, D., Cohen, W.B., Parker, G., Shugart, H.H. 1999. Surface lidar remote sensing of basal area and biomass in deciduous forests of Eastern Maryland, USA. Remote Sensing of Environment 67:83– 98.

Lyytikäinen-Saarenmaa, P. and Tomppo E. (2002): Impact of sawfly defoliation on growth of Scots pine Pinus sylvestris (Pinaceae) and associated economic losses. Bulletin of Entomological Research 92: 137–140

Lyytikäinen-Saarenmaa, P., Varama, M., Anderbrant, O., Kukkola, M., Kokkonen, A.-M., Henderström, E., and Högberg, H.-E. 2006. Monitoring the European pine sawfly in maturing Scots pine stands with pheromone traps. *Agricultural and Forest Entomology*, 8:7–15.

Lyytikäinen-Saarenmaa, P., Holopainen, M., Ilvesniemi, S. and Haapanen, R. 2008. Detecting pine sawfly defoliation by means of remote sensing and GIS. Forstschutz Aktuell. pp. 14–15.

Magnussen, S., Eggermont, P., LaRiccia, V.N. 1999. Recovering tree heights from airborne laser scanner data. Forest Science 45:407–422.

Maltamo, M., Mustonen, K., Hyyppä, J., Pitkänen, J., Yu, X. 2004. The accuracy of estimating individual tree variables with airborne laser scanning in boreal nature reserve. Canadian Journal of Forest Research 34:1791–1801.

Means, J. E., Acker, S.A., Fitt, B.J., Renslow, M., Emerson, L., Hendrix, C.J. 2000. Predicting forest stand characteristics with airborne scanning lidar. Photogrammetric Engineering & Remote Sensing 66(11):1367–1371.

Næsset, E. 1997. Estimating timber volume of forest stands using airborne laser scanner data. Remote Sensing of Environment 61:246–253.

Næsset, E. 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sensing of Environment 80:88–99.

R Development Core Team. (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL www.R-project.org

Sohlberg, S., Næsset, E., Hanssen, K.H., and Christiansen, E. 2006. Mapping defoliation during a severe insect attack on Scots pine using airborne laser scanning. *Remote Sens. Environ.,* 102:364–376.

Van Aardt, J.A.N., Wynne, R.H., Scrivani, J.A. 2008. Lidar-based mapping of forest volume and biomass by taxonomic group using structurally homogenous segments. Photogrammetric Engineering & Remote Sensing 74(8):1033–1044.

Wallerman, J., Holmgren, J. 2007. Estimating field-plot data of forest stands using airborne laser scanning and SPOT HRG data. Remote Sensing of Environment 110(4):501–508.

Zhao, K., Popescu, S., and Nelson, R. 2009. Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. *Remote Sens. Environ.* 113(1):182–196.

Above the canopy, beneath the trees – LAI determination based on full waveform ALS, high resolution TLS, and litter trap data

Inga Mölder‡, Stefan Fleck‡, Johannes Eichhorn‡, Jeanne Fischer‡

‡ Department of Environmental Control, Northwest German Forest Research Station,
37079 Göttingen, Germany

Keywords: Leaf area index, terrestrial laser scanning, airborne laser scanning, Fagus, Quercus

1. Introduction

The potential of airborne LiDAR for tree vitality assessment is explored as one of many techniques in the framework of the EU project FutMon (Further Development and Implementation of an EU-level Forest Monitoring System). Leaf area index (L) has been frequently used as parameter for tree vitality assessment. Up to now, it has been virtually impossible to exactly determine L. Therefore, different approaches to estimate L have been developed. The measured objects are identical and theory behind the indirect approaches is similar, with gap fraction being the basic quantity for all calculations. In this study, we oppose a direct method (leaf area determination from litter traps) to indirect methods of L determination (LAI-2000, TLS, ALS after Solberg et al. 2006). The direct method is used as reference method, giving most likely the value which is closest to the real L.

While a homogeneous distribution of leaves throughout the canopy is assumed for all indirect methods, leaf biomass may be shifted to lower crown layers under conditions of stress (Leuschner et al. 2009). We hypothesize that these stress induced leaf distribution patterns affect the accuracy of indirect terrestrial methods more than the airborne method. Furthermore, we analyze the hypothesis that there are differences between the two different analyzed species.

2. Material and methods

We chose two study areas: one is dominated by oak (60 km², 51.487° North, 9.519° East), the other one by beech (18 km², 50.658° North, 8.653° East), both are situated in temperate forests in Central Germany.

We established 20 plots with a size of 50 by 50m in each study area, where we conducted comprehensive field measurements. Each plot had a grid of 16 points, where we measured the effective plant area index (PL_e) with the Plant Canopy Analyzer (LAI-2000 Li-Cor, Lincoln, Nebraska, USA). In order to gain PL_t, the plant area index defined as one half of the total surface area of leaves and supporting woody materials per unit ground surface, we corrected PL_e for clumping with the between-shoot clumping factor Ω_e, which we measured with TRAC (Tracing Radiation and Architecture of Canopies, (Chen et al. 1997) on a transect of 120m. The within-shoot clumping factor γ_e, was set to 1 (Leblanc et al. 2002).

On 5 plots per site, leaf area index was measure directly. For this, we installed 12 litter traps per plot, determined specific leaf area of a subsample and calculated total leaf area per ground area based on the dry weight.

Full waveform ALS-measurements were carried out on July 9th and 10th 2009. The airborne scans were done in a height of 300m with the TopEye system S/N 724 installed in a helicopter (SE-JIH). With a pulse frequency of 170kHz, a scanning frequency of 70Hz, and a maximum scan angle of ±12°, this resulted in a mean pulse density of about 60 returns per square meter. The xyz-data was classified into digital elevation model and digital surface model. In analogy to the effective plant area index (L_e) from the LAI-2000 measurements, we generated a cut-off level in 1.5m above ground to distinguish echoes from the tree

canopy cover (N_a) and echoes from the ground or plants below the canopy (N_b). The airborne effective plant area index (AL_e,) was then calculated as logarithm of the ratio between total number of echoes in the stand (N_{all}) and N_b (Solberg et al. 2006). A pre-study showed that the relationship between measurements of LAI-2000 and ALS got worse with increasing cutoff-height, except from stands with bracken (*Pteridium aquilinum*) (data not presented).

In addition to AL_e, which is based on the gap fraction of the whole canopy higher than 1.5m above ground, we distinguished between upper (AL_{eu}) and lower part of the canopy (AL_{el}). The boundary height between upper and lower canopy was chosen from histograms of the tree height model of every plot and set to the most abundant value of the tree height model, which we assume to be the boundary between shaded and sunlit crown.

Terrestrial laser-scanning was conducted with the 3D laser-scanner Imager 5006 (Zoller + Fröhlich GmbH, Germany). After combining all scans of a plot by registration of fixed control points lying in the scene with the Software Z+F LaserControl, the plots were cut out of the 3D-point cloud along their actual margins (Software Cyclone, Leica, Germany) based on posts marking the 4 corner points. The xyz-data were transferred to a voxel model. Each voxel had an edge length of 10cm and had to contain at least 5 echoes. Comparable to the processing of ALS data, we calculated an effective plant area index for the terrestrial data (TL_e) as the ratio of total number of voxels to number of voxels from beneath the canopy for the entire crown, and subsequently distinguished between TL_e for the upper crown (TL_{eu}) and for the lower crown (TL_{el}) using the same boundary height.

3. Results and discussion

As first results, we present general regressions between plant canopy analyzer and laser data. We found a strong relationship between PL_e and the AL_e (Fig. 1). If oak and beech stands are combined, a linear relationship with an R^2 of 0.88 can be observed ($p<0.001$, $y=0.51x+0.117$). If the forest types are analyzed separately, we can see a stronger relationship between PL_e and AL_e in beech stands ($R^2 = 0.81$, $p < 0.001$, $y=0.577x-0.28$) compared to oak stands ($R^2= 0.3$, $p=0.012$, $y=0.198x+1.01$). In contrast to the overall relationship between PL_e and AL_e, the intercept of PL_e and AL_e for oak had an intercept that significantly differed from zero ($p<0.001$).

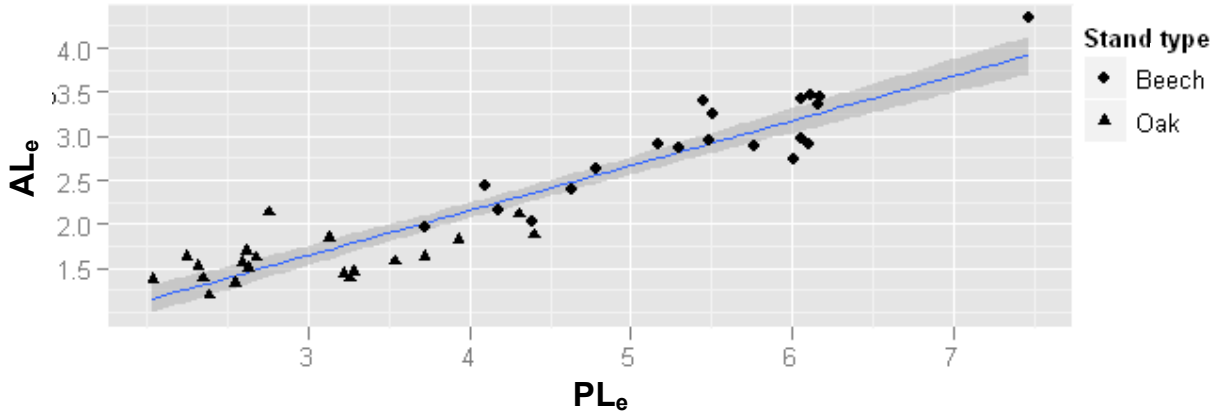


Fig. 1: *Regression ($R^2=0.88$, $p<0.001$) between effective plant area index measured with the LAI-2000 (PL_e) and effective plant area index derived from ALS (AL_e).*

The relationship between terrestrial LiDAR measurements (TL_e) and the PL_e was not as distinct as the correlation with the airborne data (Fig. 2). In fact, the relationship was positive, but the rate of explained variance was rather small ($R^2=0.21$, $p=0.012$, $y=0.030x+1.157$) and the intercept significantly different from

zero ($p<0.001$). If the datasets of both forest types are separated, they do not show a trend in beech forests but a positive trend is visible in the oak forests ($R^2=0.34$, $p= 0.022$, $y=0.083x+1.006$).

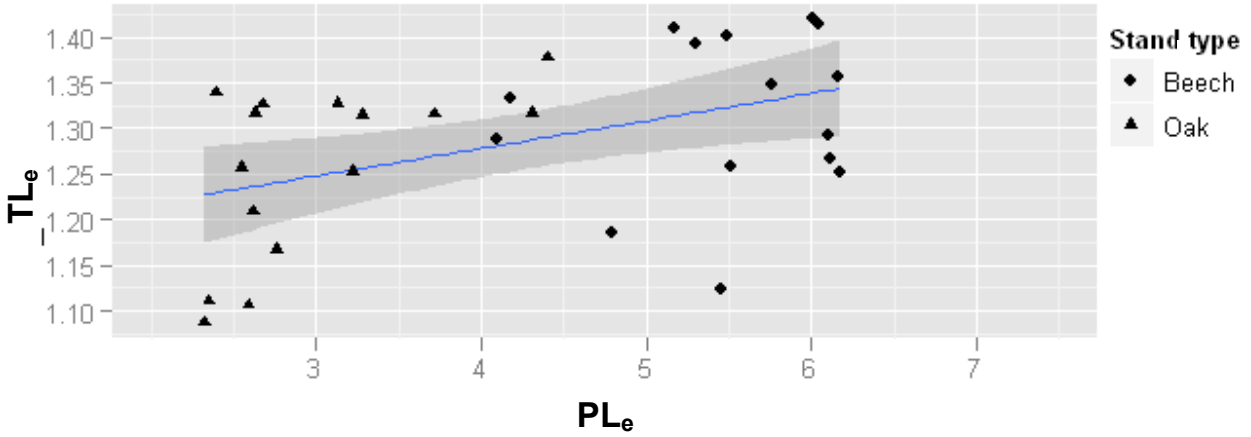


Fig. 2: *Regression ($R^2=0.21$, $p=0.012$) between effective plant area index measured with the LAI-2000 (PL_e) and effective plant area index derived from TLS (TL_e).*

The distinction between effective plant area index for upper canopy (L_{eu}) and lower canopy (L_{el}) led to contrary results when airborne and terrestrial data are compared to LAI-2000 measurements. In the airborne data, PL_e and AL_{eu}, show a significantly positive linear relationship in both forest types (Fig. 3). However, the relationship with AL_{el} did not show the same strength of correlation. In beech stands, AL_{el} correlated better with PL_e than it did in the upper canopy ($R^2=0.78$, $p<0.001$), while in oak stands the correlation between PL_e and AL_{el} ($R^2=0.16$, $p=0.078$) was much worse than that for the upper canopy.

For L_e equivalents from terrestrial laser scanning, we found corresponding trends. TL_{eu} did not correlate with PL_e in both forest types but TL_{el} was, at least in oak stands, positively correlated to PL_e ($R^2=0.33$, $p=0.027$).

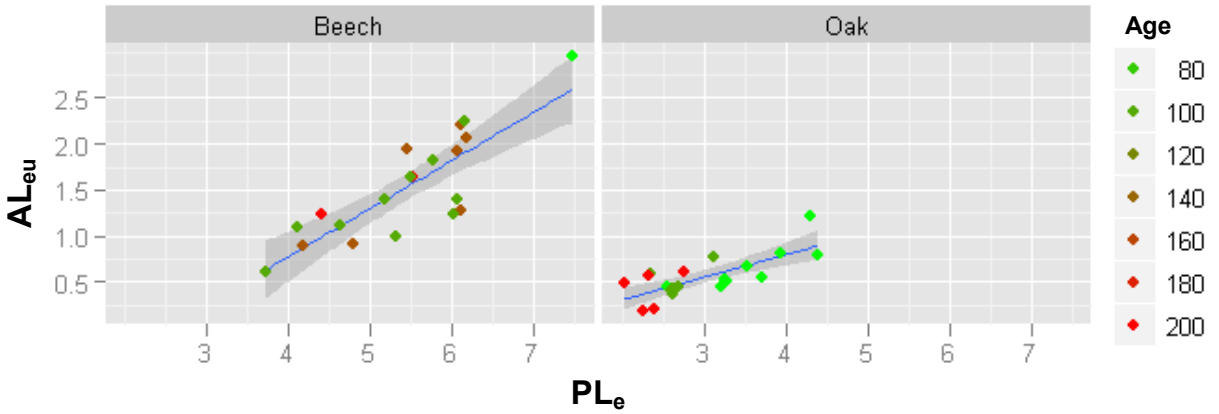


Fig. 3: *Regression between effective plant area index measured with the LAI-2000 (PL_e) and effective plant area index for the upper canopy based on ALS-measurements (AL_{eu}). The correlation was better for beech stands (left, $R^2=0.69$, $p<0.001$) than for oak stands (right, $R^2=0.55$, $p<0.001$).*

Our first results show species-specific differences when LiDAR and LAI-2000 measurements are compared. Similarly, also Riaño et al. (2004) found species-specific differences in the comparison between LiDAR and hemispherical pictures. Further results will be shown on the conference and will be discussed with respect to the different suitability of airborne and terrestrial measurements for representation of upper and lower part of the crown.

References

Chen, Jing M., Paul M. Rich, Stith T. Gower, John M. Norman, und Steven Plummer. 1997. Leaf area index of boreal forests: Theory, techniques, and measurements. *Journal of Geophysical Research* 102, 24: 29429-29443.

Leblanc, Sylvain G., Jing M. Chen, und Mike Kwong. 2002. Tracing Radiation and Architecture of Canopies TRAC MANUAL Version 2.1.3. Ottawa, Ont., September 30. http://faculty.geog.utoronto.ca/Chen/Chen%27s%20homepage/PDFfiles/tracmanu.pdf.

Leuschner, Christoph, Sylvia Voß, Andrea Foetzki, und Ina C. Meier. 2009. In welcher Beziehung stehen geschätzter Blattverlust und tatsächliche Blattflächenreduktion bei der Buche? *Forstarchiv* 80, 5: 229-235. doi:10.2376/0300-4112-80-229.

Riaño, David, Fernando Valladares, Sonia Condés, und Emilio Chuvieco. 2004. Estimation of leaf area index and covered ground from airborne laser scanner (Lidar) in two contrasting forests. *Agricultural and Forest Meteorology* 124, 3: 269-275. doi:10.1016/j.agrformet.2004.02.005.

Solberg, Svein, Erik Næsset, Kjersti Holt Hanssen, und Erik Christiansen. 2006. Mapping defoliation during a severe insect attack on Scots pine using airborne laser scanning. *Remote Sensing of Environment* 102, 3 : 364-376. doi:10.1016/j.rse.2006.03.001.

LiDAR data acquisition and processing protocols for forest resource inventories in Ontario, Canada

PAUL TREITZ†, KEVIN LIM‡, MURRAY WOODS§, DOUG PITT¶, DAVE NESBITT§, DAVE ETHERIDGE|*

† Department of Geography, Queen's University, Kingston, Ontario, K7L 3N6, Canada

‡ Lim Geomatics Inc., P.O. Box 45089, 680 Eagleson Road, Ottawa, Ontario, K2M 2G0, Canada

§ Ontario Ministry of Natural Resources, Southern Science & Information Section
3301 Trout Lake Road, North Bay, Ontario, P1A 4L7, Canada

¶ Canadian Wood Fibre Centre, Canadian Forest Service, 1219 Queen St. E.
Sault Ste. Marie, Ontario, P6A 2E5, Canada

| Ontario Ministry of Natural Resources, Northeast Science & Information Section
P.O. Bag 3020, South Porcupine, Ontario, P0N 1H0, Canada

Abstract

Over the past two decades there has been an abundance of research demonstrating the utility of airborne light detection and ranging (LiDAR) for predicting forest biophysical/inventory variables at the plot and stand levels. However, to date there has been little effort to develop a set of protocols for data acquisition and processing that would move governments or the forest industry towards cost-effective implementation of this technology for strategic and tactical (i.e. operational) forest resource inventories. The goal of this paper is to initiate this process by examining the significance of data acquisition intensity (i.e. point density) for modelling forest inventory variables. Field data for approximately 200 plots, sampling a broad range of forest types and conditions across Ontario, were collected for three study sites. Airborne LiDAR data, characterized by 3.2 points m^{-2} were systematically decimated to produce additional datasets with point densities of approximately 1.6 and 0.5 points m^{-2}. Models, derived using stepwise regression, were developed for each of the three LiDAR datasets to estimate several forest inventory variables including: (1) basal area (R^2=0.25-0.94); (2) gross total volume (R^2=0.46-0.95); (3) gross merchantable volume (R^2=0.37-0.94); (4) quadratic mean diameter (R^2=0.59-0.86); (5) total aboveground biomass (R^2=0.26-0.93); (6) average height (R^2=0.76-0.95); (7) top height (R^2=0.75-0.98) and density (R^2=0.20-0.89). Aside from a few cases, no decimation effect was found on the precision of the prediction of these forest variables, which suggests that a point density of 0.5 points m^{-2} is sufficient for plot and stand level modelling under these forest conditions.

1. Introduction

There has been a rapid growth in the application of airborne light detection and ranging (LiDAR) data for forestry, especially with respect to the potential production of an enhanced forest resource inventory (eFRI) and much improved land base feature delineation. Research into the application of airborne light detection and ranging (LiDAR) has demonstrated that accurate (and precise) forest inventory variables can be measured and modeled from LiDAR height metrics (van Leeuwen and Nieuwenhuis 2010; Wulder *et al*. 2008). These include critical parameters such as mean diameter at breast height (DBH), merchantable volume, fractional cover, leaf area index and biomass (Lefsky *et al*. 2005; Lim and Treitz, 2004; Morsdorf *et al*. 2006; Maltamo *et al*. 2006; Woods *et al*. 2008).

However, standards for the acquisition, processing and application of LiDAR data for forestry and natural resources inventory and management are not well defined, nor are they likely to be standardized across all inventory variables or forest types. For example, data acquisition standards that determine the optimal

acquisition of LiDAR data for forestry (in terms of forest variable estimation and cost efficiency) have not been universally defined, nor is there documentation of expert knowledge defining suitable acquisition criteria for estimating forest variables. These standards are required if the forest industry is to gain the best possible return from the technology across a range of forest conditions and for specific operational requirements. This deficiency must be addressed to provide the forest sector, both in industry and government with a distinct competitive advantage in achieving truly sustainable forest management that encompasses economic, ecological, and social values.

The overall goal of our research has been to examine acquisition standards for collecting, processing and analysing LiDAR data to derive forest inventory attributes that lead to the production of an eFRI for Ontario forests. Specifically, a key question that has yet to be fully addressed, and that the forest industry continues to ask as it considers operationalizing the use of LiDAR in forest resource inventories, is: *What is the optimal point density for predicting forest inventory variables*? Although research has been conducted on the impact of point density on LiDAR metrics (e.g. Hopkinson 2007, Lim *et al*. 2008; Morsdorf *et al*. 2008; Næsset 2009), it is still not clear how LiDAR data collected at different point densities impacts the estimation of a full range of forest biophysical variables for forests in Ontario. To investigate this question, we examined the impact of three point densities (3.2, 1.6, and 0.5 points m^{-2}) on the prediction of several forest inventory variables for forest types common to Ontario.

2. Methodology

2.1 Study areas

The three study areas include the Swan Lake Research Forest (SLRF), Petawawa Research Forest (PRF) and Romeo-Malette Forest (RMF) (figure 1).

Swan Lake Research Forest

The SLRF is a 2000 ha forest located 250 km north of Toronto within Algonquin Provincial Park (45° 28' N, 78° 45' W). The SLRF is situated within the Great Lakes–St. Lawrence Forest region and comprises mature stands of shade– and mid–tolerant hardwoods (sugar maple [*Acer saccharum* Marsh.], American beech [*Fagus grandifolia* Ehrh.], soft maple [*Acer rubrum* L.], yellow birch [*Betula alleghaniensis* Britt.], ironwood [*Ostrya virginiana* (Mill.) K. Koch]), conifers (eastern hemlock [*Tsuga canadensis* (L.) Carrière], eastern white pine [*Pinus strobus* L.], white spruce [*Picea glauca* (Moench) Voss], red spruce [*Picea rubens* Sarg.], eastern larch [*Larix laricina* (Du ROI) K. Koch], eastern white cedar [*Thuja occidentalis* L.], balsam fir [*Abies balsamea* (L.) Mill.]), and minor proportions of mid–tolerant and intolerant hardwoods (i.e. white birch [*Betula papyrifera* Marsh.], black cherry [*Prunus serotina* Ehrh.], white ash [*Fraxinus americana* L.], black ash [*Fraxinus nigra* Marsh.], and trembling aspen [*Populus tremuloides* Michx.]).

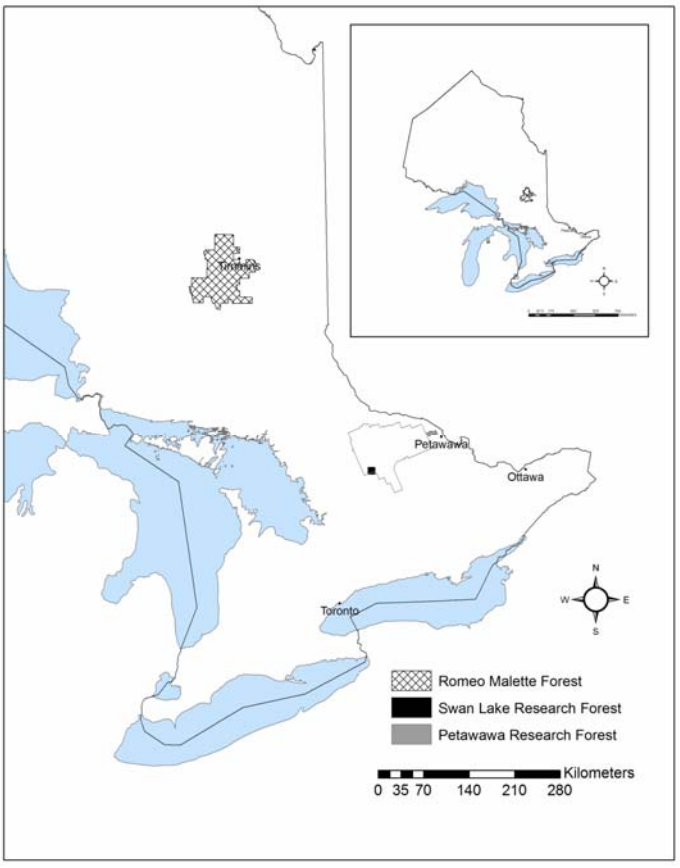


Figure 1. Locations of the three study sites in Ontario, Canada

Petawawa Research Forest

The PRF is located approximately 200 km west of Ottawa and 180 km east of North Bay, just east of Chalk River, Ontario. The research forest encompasses 10,000 ha of mixed mature natural and plantation forest that is representative of the Great Lakes–St. Lawrence Forest and is characterized by eastern white pine, red pine (*Pinus resinosa* Ait.), trembling aspen, and white birch. Red oak (*Quercus rubra* L.) dominates many poor, dry soils.

Romeo-Malette Forest

The RMF is located in the northeast portion of Ontario's Boreal Forest near Timmins, Ontario. It is an active forest management unit with approximately 532,000 productive forest hectares. The forest is characterized by extensive coniferous stands on poorly drained lowlands and gently rising uplands. The dominant species are black spruce [*Picea mariana* (Mill) B.S.P.], white birch, trembling aspen, jack pine [*Pinus banksiana* Lamb.], eastern white cedar, white spruce, eastern larch, and balsam fir.

2.2 LiDAR data

Airborne LiDAR data were collected in August 2007 for each of the study areas on a strip basis using an Optech ALTM 3100 mounted in a Cessna Grand Caravan aircraft. The base mission was flown at 1000 m with a 20° field of view (½ angle), scan rate of 54 Hz, and a maximum pulse repetition frequency of 100,000 Hz. This configuration resulted in a cross track spacing of 0.499 m, an along track spacing of 0.572 m, an average sampling density of 3.2 points m^{-2}, and a swath width of approximately 475 m. The LiDAR data were classified as ground or non-ground returns by the vendor using the TerraScan software and proprietary algorithms.

2.3 Ground reference data

The forest types sampled were: (i) tolerant hardwoods (i.e. sugar maple, beech, yellow birch) (Tol-Hwd); (ii) Great Lakes–St. Lawrence pine communities (i.e. white pine, red pine, jack pine) (GrtLks-Pine); (iii) Boreal black spruce (Boreal-SB); (iv) Boreal jack pine (Boreal-PJ); (v) Boreal intolerant hardwoods (i.e. white birch, trembling aspen) (Boreal-IH); and (vi) Boreal mixedwoods (Boreal-MW). Ground reference data were collected for the three study areas during the periods of November-December 2007 and May-October 2008. A circular, fixed area plot of 400 m^2 (11.28 m radius) was used for sampling all forest types except the tolerant hardwood group, where a 1000 m^2 (17.84 m radius) plot size was used to better represent the uneven-aged size class structures present. The centre of each circular plot was geo-referenced with a Trimble Pro XT™ kinematic GPS unit connected to a Hurricane™ antenna, mounted on a tripod. A minimum of 300 points were collected for each post position and later post-processed against a base station to achieve sub-meter accuracy.

Each plot had all trees larger than or equal to 10.0 cm measured for DBH with a diameter tape. Each tree was assessed for species, status (i.e. live or dead), crown class (i.e. dominant, co-dominant, etc.) and visual quality. A Vertex™ hypsometer was used to measure tree height for every tree in the plot. Heights of deciduous species were measured during leaf-off conditions to obtain the most accurate height measurements possible. The forest variables included in the analysis are presented in table 1.

Table 1. The forest variables considered for this analysis.

Variable	Alias
Basal Area (m^2) $SUMBA = DBH^2 * 0.00007854$ Per hectare value calculated by summing each tree per plot.	SUMBA
Gross Total Volume ($m^3\ ha^{-1}$) (Honer, 1983) $SUMGTV = ß_1 * DBH^{2*}(1-0.04365* ß_2)^2/(ß_3+(0.3048 * ß_4/Ht))$ Individual tree volume equation. Coefficients vary by species. Per hectare value calculated by summing each tree per plot.	SUMGTV
Gross Merchantable Volume ($m^3\ ha^{-1}$)¥ (Honer, 1983) $SUMGMV = SUMGTV * (ß_1 + ß_2(X) + ß_3(X))$ $X = [(1+HS/HT)(Dtop^2/DBH^2)]$; HS = Stump Height; HT = Total Height; Dtop = Minimum Top Diameter Individual tree volume equation. Coefficients vary by species. Per hectare value calculated by summing each tree per plot.	SUMGMV
Density (stems ha^{-1}) Number of live trees 10.0 cm and larger expressed per hectare.	DENSITY
Quadratic Mean DBH (cm) $\sqrt{\left[\left(\sum DBH^{\,2} / n\right)\right]}$, where n is stems per plot.	QMDBH
Average Height (m) Calculated as the average height of all trees 10.0 cm and larger.	AVGHT
Top Height (m) Calculated as the average height of the largest 100 stems per hectare.	TOPHT
Aboveground Biomass (Kg ha^{-1}) (Ter-Mikaelian and Korzukhin, 1997) $SUMBIO = ß_1 * DBH^{ß2}$ Individual tree above ground biomass equation. Coefficients vary by species. Per hectare value calculated by summing each tree per plot.	SUMBIO

¥ Merchantable limits for GMV for all species used a stump height of 20 cm and 10 cm top.

A total of 32 plots were established in the SLRF and assigned to the Tol-Hwd forest type. Similarly, 35 plots were established in the PRF and assigned to the GrtLks-Pine forest type while 136 plots were established in the RMF, with each plot assigned to one of four forest types (i.e. Boreal-IH, Boreal-MW, Boreal-PJ and Boreal-SB).

2.4 Data Processing

Normalization

All LiDAR returns were normalized against a triangulated irregular network (TIN) that was developed using the LiDAR returns that were classified as ground. The process involved subtracting from the original z-value of a return the z-value on the TIN matching its x-y coordinates. No height threshold (e.g. 2 m) was used to filter the LiDAR data.

Decimation

LiDAR data were decimated according to the methodology described by Raber *et al.* (2007). A decimation level 0 (D0) represented the original dataset characterized by a point density of approximately 3.2 points m^{-2}. The decimation level 1 (D1) LiDAR dataset was derived by taking alternating points along each scan line with each scan line retained, thereby increasing the cross track spacing by a factor of two. For the decimation level 2 (D2) LiDAR dataset, every fourth point along each scan line was retained, thereby increasing the cross track spacing by a factor of 4, and every other scan line was retained, resulting in an increase in the along-track spacing by a factor of 2. The systematic decimation resulted in the D1 and D2 LiDAR datasets possessing point densities of approximately 1.6 and 0.5 points m^{-2}, respectively.

LiDAR Predictors

Three types of predictors (i.e. statistical, canopy height, and canopy density) were derived from the normalized LiDAR data. The statistical group of predictors included mean height and standard deviation (*stddev*) of LiDAR height measurements. The canopy height predictors consisted of deciles of LiDAR canopy height (i.e. *p10*…*p90*) and the maximum (*max*) LiDAR height. For each plot, the range of LiDAR height measurements was divided into 10 equal intervals and the cumulative proportion of LiDAR returns found in each interval, starting from the lowest interval (i.e. *d1*), was calculated. Since the last interval always sums to a cumulative probability of 1, it was excluded resulting in 9 canopy density metrics (i.e. *d1* … *d9*). The remaining two canopy density metrics were calculated as the number of first returns divided by all returns intersecting a sample plot (*Da*) and the number of first and only returns divided by all returns intersecting a sample plot (*Db*).

2.5 Statistical analyses

Model Building

Multiple stepwise regressions with a significance level of 0.05 were used for model building. A diagnosis of each model was performed to determine if parametric statistical assumptions were satisfied. The Shapiro-Wilk Test was used to determine if residuals were normally distributed and the Brown-Forsythe Test was used to check for the presence of heteroscedasticity (i.e. unequal error variance). As LiDAR predictors have been reported to be highly correlated, the variance inflation factors (VIFs) for the predictors used in each model were examined. Candidate models where predictors exhibited VIFs greater than 10 were discarded, as these would suggest the presence of multi-collinearity in the predictor data (Neter *et al.* 1996).

Effect of Decimation Treatments

For each forest variable, predictions from the models constructed for the different decimation levels were compared to respective observed values to create a multivariate response vector of absolute prediction errors for each plot:

$$\mathbf{e}_{i(j)} = [e0_{i(j)}, e1_{i(j)}, e2_{i(j)}]$$

where, $e0_{i(j)}$ is the absolute error associated with the undecimated prediction for plot *j* within each forest/type *i* (Tol-Hwd, GrtLks-Pine, Boreal-PJ, Boreal-SB, Boreal-IH and Boreal-MW):

$$e0_{i(j)} = \left| y_{i(j)} - \hat{y}0_{i(j)} \right|,$$

and $e1_{i(j)}$ and $e2_{i(j)}$ denote similar errors for the decimated predictions, 1.6 and 0.5 points m^{-2}, respectively.

Prediction errors were then subjected to repeated measures analyses of variance (RMANOVA), treating the different forests/types as fixed effects, and making comparisons of the different decimation levels (within-subject effects) with multivariate tests (Wilks' lambda). RMANOVA is used when all members of a sample are measured repeatedly under a number of different conditions. Within the context of this study, it was a particular forest variable for a plot that was repeatedly predicted using LiDAR data of three varying point densities. Given repeated measurements, the use of a standard ANOVA is not appropriate (Popescu and Zhao, 2008). We contrasted the different decimation levels to test for *increases* in prediction error that were proportional and disproportional to the decimation levels applied (i.e. *linear* and *quadratic* contrasts, respectively, associated with increasing decimation). With this approach, the following null hypothesis was tested for each forest variable:

H_0: Decimation of the LiDAR point cloud from 3.2 to 1.6 and 0.5 points m^{-2} does not reduce prediction precision; versus

H_{a1}: Decimation of the LiDAR point cloud reduces prediction precision proportional to decimation level; or

H_{a2}: Decimation of the LiDAR point cloud reduces prediction precision disproportional to decimation level.

In cases where a decimation forest/type interaction was indicated, similar analyses by forest/type were used to reveal the source of interaction.

3. Results and discussion

For illustration purposes, the models developed for each variable for Boreal-SB plot data are presented in table 2. For black spruce, the models typically exhibit very high coefficients of determination (i.e. R^2) and RMSEs ranging from approximately 4-19%. Variable-by-stand type results indicate that height-related models (QMDBH; AVGHT; TOPHT) tended to perform very well (i.e. RMSE < 10%) and volume/biomass-related models (SUMBA; SUMGTV, SUMGMV, SUMBIO) performed moderately well (RMSE typically 10-20%). Density models tended to exhibit the highest RMSEs.

Table 2. Models (and associated statistical descriptors) developed for variables based on black spruce plots in the RMF. Models are presented for each of the decimations levels (i.e. D0, D1, D2). Similar sets of models were developed for SLRF (Tol-Hwd), PRF (GrtLks-Pine), and RMF (Boreal-PJ, Boreal-SB, Boreal-IH and Boreal-MW).

Variable	Dec. Level	Equation	R^2	R^2 Adj.	RMSE	RMSE (%)
SUMBA ($m^2 ha^{-1}$)	D0	58.9 - 20.2 d5 + 1.58 p50 - 0.379 Db	0.918	0.909	3.01	11.68
	D1	- 0.91 + 2.23 p50 + 0.622 Da	0.910	0.904	3.15	12.22
	D2	48.6 + 1.26 p60 + 1.27 p40 - 0.478 Db	0.935	0.929	2.66	10.33
SUMGTV ($m^3 ha^{-1}$)	D0	- 702 + 32.1 mean - 210 d6 + 873 d9 - 110 p20	0.949	0.942	17.96	11.04
	D1	- 66.3 + 42.2 mean - 5.59 p30	0.927	0.922	21.52	13.23
	D2	268 + 23.5 mean - 3.30 Db + 4.95 p40	0.942	0.936	19.14	11.77
SUMGMV ($m^3 ha^{-1}$)	D0	- 114 + 5.63 p40 + 17.4 p90	0.916	0.910	18.19	16.69
	D1	- 182 + 20.1 p80 + 7.02 p40 + 113 d4	0.915	0.907	18.25	16.74
	D2	- 195 + 41.5 mean + 164 d3 - 282 p10	0.939	0.933	15.46	14.18
DENSITY (stems ha^{-1})	D0	- 112 - 4076 d3 + 3254 p10 - 159 p30 - 180 p80 - 56.4 Db + 106 p40 + 9705 d9	0.888	0.857	231.84	14.11
	D1	5380 - 4678 d4 + 5580 p10 - 134 p30 - 95.9 p90	0.794	0.766	313.83	19.10
	D2	11212 - 2363 d4 - 204 p90 - 80.4 Db + 70.5 p40	0.861	0.842	257.74	15.69
QMDBH (cm)	D0	8.03 + 1.18 p90 - 0.327 Da - 0.138 p40	0.838	0.822	0.82	5.80
	D1	8.58 + 1.07 p90 - 0.308 Da	0.783	0.769	0.95	6.72
	D2	2.44 + 1.12 p90 - 0.234 Da + 5.33 d6	0.863	0.850	0.76	5.34
AVGHT (m)	D0	2.41 + 0.948 p90 + 2.80 d1 - 0.0821 Da	0.951	0.946	0.49	3.81
	D1	1.73 + 0.836 p90 + 2.99 d2	0.941	0.938	0.53	4.16
	D2	2.34 + 0.802 p90 + 2.26 d3	0.936	0.932	0.56	4.34
TOPHT (m)	D0	2.32 + 0.547 max + 0.436 p90	0.923	0.918	0.69	4.10
	D1	2.10 + 0.578 max + 0.429 p90	0.927	0.922	0.66	3.98
	D2	3.79 + 0.577 max + 0.326 p90	0.903	0.897	0.77	4.58
SUMBIO ($kg ha^{-1}$)	D0	-321894 + 15682 mean - 114298 d6 + 427669 d9	0.925	0.918	11 673	11.58
	D1	- 17265 + 21243 mean	0.905	0.902	13 199	13.09
	D2	157912 + 11931 mean - 1752 Db + 3131 p40	0.932	0.925	11 143	11.05

For all forest variables tested, overall model prediction precision varied strongly with forest/type ($p < 0.01$; Table 3). Boreal-SB variables tended to be predicted with the greatest precision (lowest mean absolute errors) and GrtLks-Pine variables the least precision. The Boreal-SB stands were quite similar in that the majority were upland sites with mature black spruce of natural origin, whereas the GrtLks-Pine communities were more diverse in terms of species, management and origin. The range of conditions sampled included unmanaged white pine, shelterwood white and red pine, thinned and unmanaged red pine plantations as well as some natural jack pine stands. In future work, this group will be subdivided further to better consider the volume/height relationships for these species and management conditions.

Table 3. Summary of mean absolute errors for the decimation levels tested and p-values from RMANOVA (Wilks' lambda) testing for decimation and forest/type effects and their interaction. (Note that values in bold suggest a loss of precision with increasing decimation for some forest/types).

	SUMGMV m^3 ha^{-1}	SUMGTV m^3 ha^{-1}	SUMBA m^2 ha^{-1}	DENSITY stems ha^{-1}	QMDBH cm	AVGHT m	TOPHT m	SUMBIO kg ha^{-1}
Overall mean absolute error:								
3.2 points m^{-2}	29.4	29.1	3.4	**176**	1.69	**0.65**	0.59	17285
1.6 points m^{-2}	28.5	29.1	3.3	**206**	1.69	**0.68**	0.59	17631
0.5 points m^{-2}	27.3	30.5	3.4	**201**	1.52	**0.76**	0.61	17688
Source of variation:								
Decimation	0.23	0.46	0.94	<0.01	0.09	0.02	0.30	0.62
linear	0.10	0.25	0.86	<0.01	0.05	<0.01	0.14	0.56
quadratic	0.64	0.29	0.82	0.07	0.08	0.17	0.57	0.80
Decimation x Forest/Type	0.87	0.72	0.96	**<0.01**	<0.01	**0.10**	0.98	0.32
Forest/Type	<0.01	<0.01	0.02	<0.01	<0.01	<0.01	<0.01	<0.01

With the exception of 2 of the 8 forest variables tested, we generally found little evidence to reject the null hypothesis that decimation of the LiDAR point cloud has no affect on model precision ($p > 0.10$; Table 3). In overall analyses, mean prediction errors tended to increase with decimation for the variables DENSITY and AVGHT ($p \leq 0.02$), but there was evidence to suggest that this situation was not consistent across all forest/types studied ($p \leq 0.10$). More specifically, DENSITY prediction errors for Boreal-PJ increased from 226 stems ha^{-1} at 3.2 points m^{-2}, to 299 and 324 stems ha^{-1} through decimation to 1.6 and 0.5 points m^{-2} (decimation linear, $p < 0.01$). To a lesser extent, prediction errors for GrtLks-Pine increased from 171 stems ha^{-1} at 3.2 points m^{-2}, to 211 and 192 stems ha^{-1} through decimation to 1.6 and 0.5 points m^{-2} (decimation quadratic, $p = 0.09$). Thinning treatments had been applied to these forest/types potentially giving rise to increased error as a function of insufficient sample size to account for a suitable range of density conditions. Similarly, AVGHT prediction errors for Boreal-MW and GrtLks-Pine tended to increase sharply (30 to 40%) with the highest level of decimation (i.e. 0.5 points m^{-2}) (decimation quadratic, $p \leq 0.10$). However, these examples appear rare in the context of the overall data set and one may argue that with a significance level of 10%, we might expect to observe trends that suggest rejection of H_0 up to 10 % of the time simply through random chance alone. Thus,

we feel that it is reasonable to conclude that decimation of the LiDAR point cloud from 3.2 to 1.6 and 0.5 points m^{-2} did not reduce the prediction precision of the forest variables tested

4. Conclusions

The results from this research demonstrate that a point density of 0.5 points m^{-2} is sufficient for the estimation of forest inventory variables at the plot and stand levels for the different forest types considered in this study. In cases where a decimation effect was observed for a forest variable, the effect may be attributed to differences in model form, specifically as it relates to number of predictors used. In short, these results indicate that low-density LiDAR based predictions offer significant potential for integration into tactical forest resource inventories for Ontario.

Acknowledgements

The authors gratefully acknowledge financial support from the Ontario Centre of Excellence for Earth and Environmental Technologies and the Forest Research Partnership (i.e. Tembec Inc., Ontario Ministry of Natural Resources, Canadian Wood fibre Centre). Dr. Treitz also acknowledges research support from the Natural Sciences and Engineering Research Council and the Premier's Research Excellence Award. Thanks to the following individuals for contributions to field data collection and logistics: Kelly Von Bargen, Paul Courville, Beth Denaburg, Michelle Desaulniers, Karin Van Ewijk, Stephanie Gagliardi, Anne Hagerman, Anthony Iserhoff, Katalijn MacAfee, Jem Morrison, Stefanie Phelan, Meg Southee, Stan Vasiliauskas.

References

Honer, T.G., Ker, M.F. and Alemdag, I.S., 1983, Metric timber tables for the commercial tree species of central and eastern Canada. Information Report M–X–140. Fredericton, New Brunswick, Maritimes Forest Research Centre, Canadian Forestry Service, 139 pp.

Hopkinson, C., 2007, The influence of flying altitude and beam divergence on canopy penetration and laser pulse return distribution characteristics. *Canadian Journal of Remote Sensing*, **29**, pp. 623-632.

Lefsky, M., Hudak, A.T., Cohen, W.B. and Acker, S.A., 2005, Geographic variability in LiDAR predictions of forest stand structure in the Pacific Northwest. *Remote Sensing of Environment*, **95**, pp. 532-548.

Lim, K. and Treitz, P., 2004, Estimation of aboveground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. *Scandinavian Journal of Forest Research,* **19**, pp. 558-570.

Lim, K., Hopkinson, C. and Treitz, P., 2008, Examining the effects of sampling point densities on laser canopy height and density metrics. *Forestry Chronicle*, 84, pp. 876-885.

Maltamo, M., Eerikäinen, K., Packalén, P. and Hyyppä, J., 2006, Estimation of stem volume using laser scanning-based canopy height metrics. *Forestry*, **79**, pp. 217-229.

Morsdorf, F., Koetz, B., Meier, E., Itten, K.I. and Allgöwer, B., 2006, Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction. *Remote Sensing of Environment*, **104**, pp. 50-61.

Morsdorf, F., Frey, O., Meier, E., Itten, K.I. and Allgöwer, B., 2008, Assessment of the influence of flying altitude and scan angle on biophysical vegetation products derived from airborne laser scanning. *International Journal of Remote Sensing*, **29**, pp. 2152-2163.

Næsset, E., 2009, Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. *Remote Sensing of Environment*, **113**, pp.148-159.

Neter, J., Kutner, M.H., Nachtsheim, C.J., and Wasserman, W., 1996, *Applied Linear Statistical Models*, Fourth Edition. Boston, WCB/McGraw-Hill.

Popescu, S.C. and Zhao, K., 2008, A voxel-based LiDAR method for estimating crown base height for deciduous and pine trees. *Remote Sensing of Environment*, **112**, pp. 767-781.

Raber, G. T., Jensen, J.R., Hodgson, M.E., Tullis, J.A., Davis, B.A. and Berglund, J., 2007, Impact of LiDAR nominal post-spacing on DEM accuracy and flood zone delineation. *Photogrammetric Engineering and Remote Sensing*, **73**, pp. 793-804.

Ter-Mikaelian, M.T. and Korzukhin, M.D. 1997, Biomass equations for sixty-five North American tree species. *Forest Ecology and Management*, **97**, pp. 1-24.

Van Leeuwen, M. and Nieuwenhuis, M., 2010, Retrieval of forest structural parameters using LiDAR remote sensing. *European Journal of Forest Research*, **129**, pp. 749-770.

Woods, M., Lim, K. and Treitz, P., 2008, Predicting forest stand variables from LiDAR data in the Great Lakes – St. Lawrence forest of Ontario. *Forestry Chronicle*, **84**, pp. 827-839.

Wulder, M.A., Bater, C.W., Coops, N.C., Hilker, T. and White, J.C., 2008, The role of LiDAR in sustainable forest management. *Forestry Chronicle*, **84**, pp. 807-826.

Tree top detection using local maxima filtering: a parameter sensitivity analysis

Jean-Matthieu Monnet†, Éric Mermin†, Jocelyn Chanussot‡ & Frédéric Berger†*

jean-matthieu.monnet@cemagref.fr

†UR EMGR, Cemagref, France

‡Gipsa-Lab, Grenoble Institute of Technology, France

Abstract

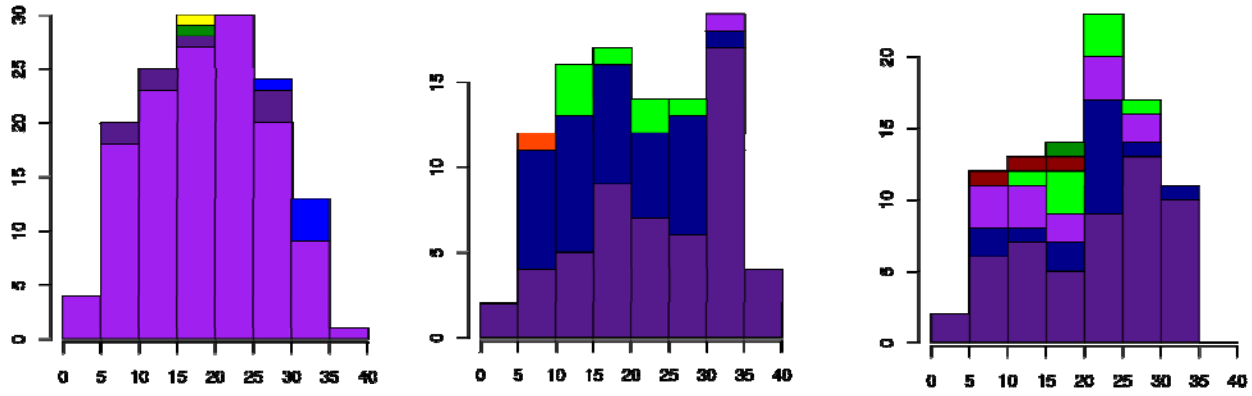
The sensitivity of a treetop detection algorithm is investigated by automated evaluation of detection performance for several parameter combinations. The algorithm consists in digital elevation models computation, morphological filtering, Gaussian smoothing and local maxima extraction and selection. The analysis is performed on three field plots located in the French Alps. One is a Norway spruce stand while the two others are dominated by Silver fir and European beech. Detection rates above 42.9% are achieved with less than 4.1% of false positives. Even though some similarities exist regarding resolution and morphological filtering, optimal settings determined on one plot performed uncertainly on the others. Besides, optimised parameters may depend on both the laser data — mainly point density — and on the forest structure and species.

1. Introduction

In French mountainous forests, timber harvesting is economically sustainable only in areas where tree density and individual volume result in sufficient standing value. Indeed, topographical constraints increase both harvesting and transport costs. In practical terms, silvicultural operations are reduced to rare operations designed to maintain stand structure and made profitable thanks to the felling of very large trees (Ancelin et al. 2006). However, there is a renewed interest in mountainous areas in order to exploit woody renewable resources and preserve the ecological and social functions of forests. Unfortunately, stand characteristics information required for efficient stand management is frequently absent or outdated.

Airborne laser scanning (ALS) is a remote sensing technique whose applications for forest inventory have been widely investigated in the past fifteen years. In particular, numerous studies have shown its efficiency for single tree identification (Persson et al. 2002, Brandtberg et al. 2003, Popescu and Wynne 2004, Chen et al. 2006, Kwak et al. 2007, Heurich 2008). Usual methods rely on the computation of raster elevation models which are then filtered by image processing techniques to extract tree tops and delineate tree crowns. Several algorithms have been proposed and tested on particular forest contexts. Such techniques are of high interest for planning silvicultural and harvesting operations in mountainous forests. However, when dealing with large scale forest inventory, few calibration plots and little time for algorithm parametrisation are available. Hence the need to control the robustness of algorithms to parameters and input data.

The objective of this study is to perform an exploratory evaluation of the sensitivity of a basic tree top detection algorithm to parametrisation. Parameters influence is assessed by automatically evaluating detection performance for a high number of combinations. Optimum settings obtained on three plots are then cross evaluated to test their robustness.



a) Plot 1 b) Plot 2a c) Plot 2b

Figure 1. Height distribution of trees inventoried in the three test plots (frequency in percent as function of height classes in metres). Colours refer to tree species (■ Abies alba, ■ Picea abies, ■ Fagus sylvatica, ■ Larix decidua, ■ Acer pseudoplatanus, ■ Betula sp., ■ Prunus avium, ■ Ilex aquifolium, ■ Sorbus aucuparia).

2. Material

Field data

The two forest stands investigated in this study are located in the French Alps. Plot 1 is an irregular stand dominated by Norway spruce (*Picea abies*) and located in the Chamonix valley (45°56'16"N, 06°53'55"E, altitude 1140 m). It was originally established to test new silvicultural strategies designed to improve mountainous stands stability and resilience (Mermin and Renaud 1996). The 0.25 ha square plot was first inventoried in 1994. The area was divided into squares with markers positioned every 12.5 m with a tape measure, a clinometer and a compass. Trees with diameter at breast height (DBH) larger than 7.5 cm had their positions to the nearest marker recorded with the same instruments. Plot corners were georeferenced using a Trimble GPS Pro XRS receiver. On June, 2nd 2009, the plot inventory was updated. DBH and tree heights were measured with a tape and a Vertex III hypsometer. 147 live trees were inventoried, with a majority of Norway spruce (89.8% of stems). Other species are silver fir (*Abies alba*, 5.4%), European larch (*Larix decidua*, 3.4%) and deciduous trees (1.4%). Figure 1(a)[5] displays the distribution of tree heights on the plot.

Plot 2 is a study area initially used for real-size rockfall experiments (Dorren et al. 2006), located at Vaujany (45°12'07"N, 06°03'00"E, altitude 1280 m). It is an irregular stand dominated by Silver fir (*Abies alba*) with a deciduous understory. In 2001 twelve markers were positioned in the hillside. They were georeferenced using a Trimble GPS Pro XRS receiver. It is noteworthy that plot boundaries were not previously delimited, which resulted in an irregular shape. Trees with a DBH larger than 7.5 cm had their positions to the nearest marker recorded with a clinometer and a compass mounted on a tripod and with a laser rangefinder. On May, 12th 2009, the plot inventory was updated. DBH and tree heights were measured with a tape and a Vertex III hypsometer. For the purposes of this study, the plot is divided into two subplots. In the upper part of the study area (subplot 2a, 0.33 ha), 98 live trees were measured. Silver fir and European beech (*Fagus sylvatica*) are dominant (respectively 55.1 and 35.7% of the stems). Sycamore maples (*Acer pseudoplatanus*) are also encountered (7.1%), as well as Norway spruce and European holly (*Ilex aquifolium*) with 1% each. 92 live trees were inventoried in the lower part (plot 2b, 0.34 ha), with a majority of

Silver fir (56.5%). European beech (16.3%) and Norway spruce (14.1%) are also present. The remaining stems are sycamores (8.7%) and other deciduous trees. Tree height distributions are shown on figure 1

Laser data

Laser data was acquired with a fullwave RIEGL LMS-Q560 scanner. Acquisition parameters are summarised in table 1. Echoes were extracted from the binary acquisition files and georeferenced with the RIEGL software suite. The contractor also classified the resulting point cloud into ground and non-ground echoes using the TerraScan software, which implements an algorithm based on iterative surface reconstruction by triangulated irregular network (Axelsson 2000). Plot 2 is located within the overlap of two adjacent flight strips which resulted in a high echo density (36.6 m^{-2}) whereas point cloud density was 7.8 m^{-2} for plot 1.

Table 1. Laser scanner acquisition parameters.

Plot number	1	2
Flight year	2007	2008
Wavelength (nm)	1550	1550
Pulse repetition rate (kHz)	200	200
Scan frequency (Hz)	104.4	73.3
Half scan angle (°)	30	30
Flight height (m)	550	500
Laser footprint (m)	0.29	0.27
Theoretical point spacing (m)	0.45	0.29
Final echo density (m^{-2})	7.8	36.6

3. Method

Workflow

For each test plot, the input data are the classified point cloud (easting, northing, altitude, classification,) and the inventoried trees list (easting, northing, height). Each step of the sequential procedure is explained in the following sections:

5 calculation of the raster images from the point cloud,
6 image processing with morphological and Gaussian filters,
7 maxima extraction and selection with variable window size,
8 performance assessment.

Raster calculation

The digital terrain model (DTM) is computed by bilinear interpolation of laser points classified as ground points over a regularly spaced grid of resolution *res*. The corresponding digital surface model (DSM) is based on the highest point recorded in each pixel. A basic, unfilled DSM ($DSM_{b,u}$) is calculated by affecting to each pixel the altitude of the highest point recorded in the cell. An interpolated, filled DSM ($DSM_{i,f}$) is calculated by bilinear interpolation of the highest points recorded in each cell, at the pixel centres. Void cells left in the basic DSM are interpolated from the neighbouring pixels to construct a filled basic DSM ($DSM_{b,f}$). On contrary pixels in $DSM_{i,f}$ without any laser points have their values erased to construct an unfilled interpolated DSM ($DSM_{i,u}$). Four corresponding crown height models (CHM) are computed by subtracting the DTM to the DSMs.

As plot extent was not previously delimited for plots 2a and 2b, raster masks are computed from the tree positions and heights. Buffers are constructed around each tree and merged. Remaining holes are filled by morphological closing with a 6 m radius structuring element. The previous external border is then reconstructed by morphological reconstruction. This procedure is designed to ensure that actual tree tops

are located within the plot mask and therefore reduce border effects during the automated linking of detected maxima with inventoried trees. Indeed, due totree tilting and GPS position errors, tree tops may be located several metres away from the georeferenced tree trunks. Buffer radius r_t for a tree of height h_t (m) is $r_t = 2.1 + 0.14 \times h_t$ (see paragraph 3.5 for parameters explanation).

Image processing

The scanning pattern usually leads to several low or void pixels in the surface models, due to the irregular sampling and to the shading effect of trees at the borders of flight strips. This can be seen as a salt-and-pepper noise and be treated by several morphological filters: median, adaptive median, closing, reconstruction. Filter windows (half width) or structuring element (disk radius) sizes are $r_m \in \{0.25,0.5,0.75,1,1.5,2\}$ (metres). Depending on the resolution *res*, sizes are approximated to the nearest integer number of pixels.

The final objective is to detect tree tops, i.e. the maxima of trees envelope. Local irregularities of branches and more generally of the canopy can be considered as high frequency noise. This can be filtered by Gaussian smoothing. Tested filter sizes in metres are $\sigma \in \{0,\ 0.1,0.2,0.3,0.4,0.5,0.75,1,1.25,1.5,2,3,4\}$. A discrete approximation of the continuous filter is used.

Maxima extraction and selection

Maxima are extracted with a sliding filter of variable size. Pixel values in the resulting maxima image correspond to the half width (metres) of the biggest centred square window where the corresponding pixel is global maximum in the original image. The plot mask is then applied to the maxima image to set the values of all pixels outside the plot to zero.

At this stage there are high chances that some of the local maxima are not actual tree tops but vertical branches, dual apexes of deciduous trees... A maxima selection procedure based on a minimum height and on the relationship between a maxima height h_m and its distance to the nearest higher pixel d_m is performed. Maxima which do not fulfil the following inequalities are discarded.

$$\begin{cases} d_m \geq d_{min} + d_{prop} \times h_m \\ h_m \geq h_{min} \end{cases} \quad \text{with } (d_{min}, d_{prop}, h_{min}) \text{ the selection parameters}$$

Their values in the maxima image are set to zero. Tested parameters are (d_{min} , d_{prop}) $\in$ {(0.5,0), (0.75,0), (1,0), (1.5,0), (1.75,0), (2,0), (2.25,0), (2.5,0), (1.5,1⁄80), (1, 1⁄40), (2⁄3,1⁄30), (10⁄25,1⁄25), (0,1⁄20)} and h_{min} $\in$ {0,2.5,5,7.5,10,12.5,15}.

The final list of selected maxima is extracted from the image. Maxima are assumed to be located at the centre of the pixels with non-zero values. Maxima heights h_m are estimated as the values of the corresponding pixels in the morphologically filtered image. Maxima distances to the nearest pixel of higher value d_m are the values in the maxima image.

Performance assessment

An automated procedure is used to link selected maxima to inventoried trees. It is assumed that GPS positioning planimetric error ϵ_{gps} is inferior to 1.5 m and that the slope of a tilted tree s_{tree} is inferior to 14% (value determined by measures performed on the 26 highest trees in plot 2). Height measures accuracy is estimated to be ϵ_h = 15%. Under these asumptions, the limit matching distance between a selected maxima and a field tree top assumed to be vertically located at h_t metres above the GPS position of tree *T* is:

$$d_{max}(h_t) = \frac{\epsilon_{gps}}{\cos(s_{terrain})} + s_{tree} \times (\epsilon_h + 1) \times h_t$$

with $s_{terrain}$ the terrain slope. For each pair constituted of a tree T and a maxima M of respective heights h_t and h_m, a matching index $I_{T,M}$ is computed as the ratio of the distance between the assumed tree top and detected maxima, and $d_{max}(h_t)$

$$I_{T,C} = \frac{d_{T,C}}{d_{max}(h_t)}$$

The potential pair with the lowest matching index is validated and the lists of remaining maxima and trees are updated before reiterating the procedure. Linking ends when all remaining values of $I_{T,C}$ are greater than 1

For accuracy assessment a trade-off between the true positives ratio (percentage of correctly detected trees $R_{TP} = N_{TP}/N$) and false positives ratio (ratio of unlinked detected maxima to field trees number $R_{FP} = NF_P/N$) has to be found. It is assumed that for every five additional true positives, one false positive is tolerated. The accuracy score S is thus computed as :

$$S = (w \times R_{FP})^2 + (1 - R_{TP})^2 \qquad \text{with } w = 5$$

which corresponds to the squared euclidean distance to the perfect match (N_{FP} = 0 and N_{TP} = N) with the false positives ratio weighted by a factor 5.

All computations are performed using Matlab®; and its Image Processing Toolbox. For each plot (1 and 2) and also for the subplots (2a and 2b) considered separately, all combinations of parameters (total number: 1381744) are tested and accuracy results recorded. To assess the precision of tree height estimation, linear models are fitted with field height as independent variable and maxima height as dependent variable.

4. Results

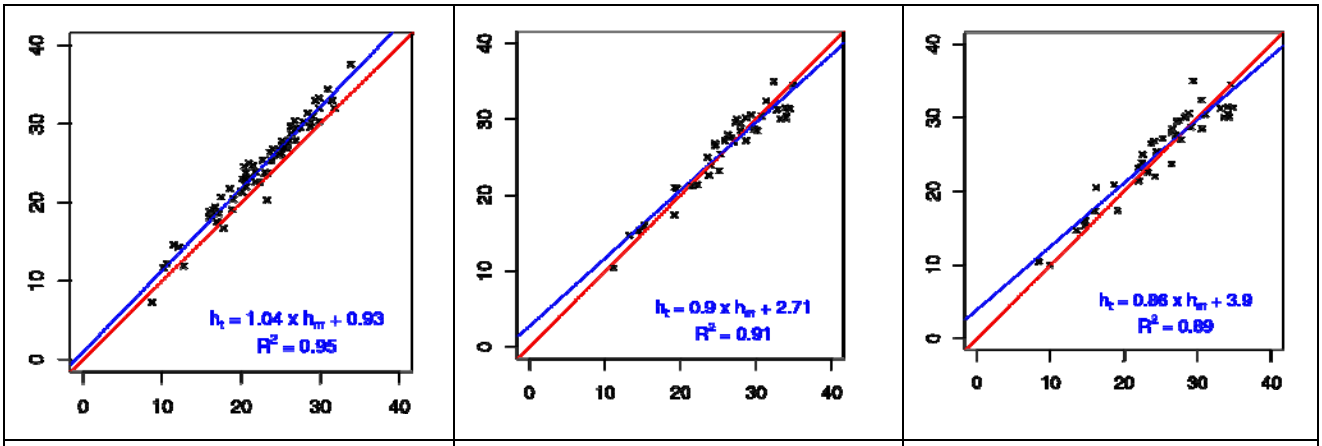
Optimal settings

The best scores and corresponding parameter settings for plots 1, 2 and subplots 2a and 2b are displayed in table 2. Detection rates above 42% are achieved with respectively 6, 2 and 2 false alarms for plots 1, 2a and 2b. These results are obtained with the smallest tested resolution (*res* = 0.2 m). A stronger morphological filtering is performed for the Vaujany subplots (2a and 2b) whereas maxima filtering is less selective. Indeed the best score is attained for several values of h_{min} and only maxima that are very close to their nearest higher pixel (d_{min} = 0.5 m) are removed. It is noteworthy that the optimum setting for the entire plot 2 is different from those of subplots 2a and 2b considered alone. Moreover, overall accuracy obtained in plot 2 is quite lower.

Table 2. Best detection results and corresponding parameter settings.

	Accuracy assessment			Height model		Morphological		Gaussian	Maxima selection[6]		
Plot	Score	R_{TP}	R_{FP}	*res*	Model	Type	r_m	σ	h_{min}	d_{min}	d_{prop}
1	0.32	46.9	4.1	0.2	$CHM_{b,u}$	closing	0.4	0.3	7.5	1	0.025 [7]
2a	0 .34	42 .9	2.0	0.2	$DSM_{i,f}$	median	1	0.1	5-15	0.5	0[8]
2b	0 .32	44 .6	2.2	0.2	$CHM_{i,u}$	median	0.8	1	0-7.5	0.5	0[9]
2	0 .39	38 .9	2.6	0.2	$CHM_{b,i}$	median	1	0.1	12.5	0	0.05

In figure 2, field heights of detected trees (h_t) ars plotted against corresponding maxima heights (h_m). Linear regressions show that on plot 1 tree heights are underestimated by about 1 meter (slope: 1.04 and intercept: 0.93 m). On the Vaujany site, heights of big trees are overestimated whereas small trees are slightly underestimated (slopes inferior to 0.91).



(a) Plot 1	*(b)* Plot 2a	*(c)* Plot 2b

Figure 2. Field height h_t plotted against laser estimated height h_m for trees correctly detected in the three test plots. Red line corresponds to y = x. Blue line is linear regression of h_t against h_m. Corresponding detection settings and accuracy are displayed in table 2

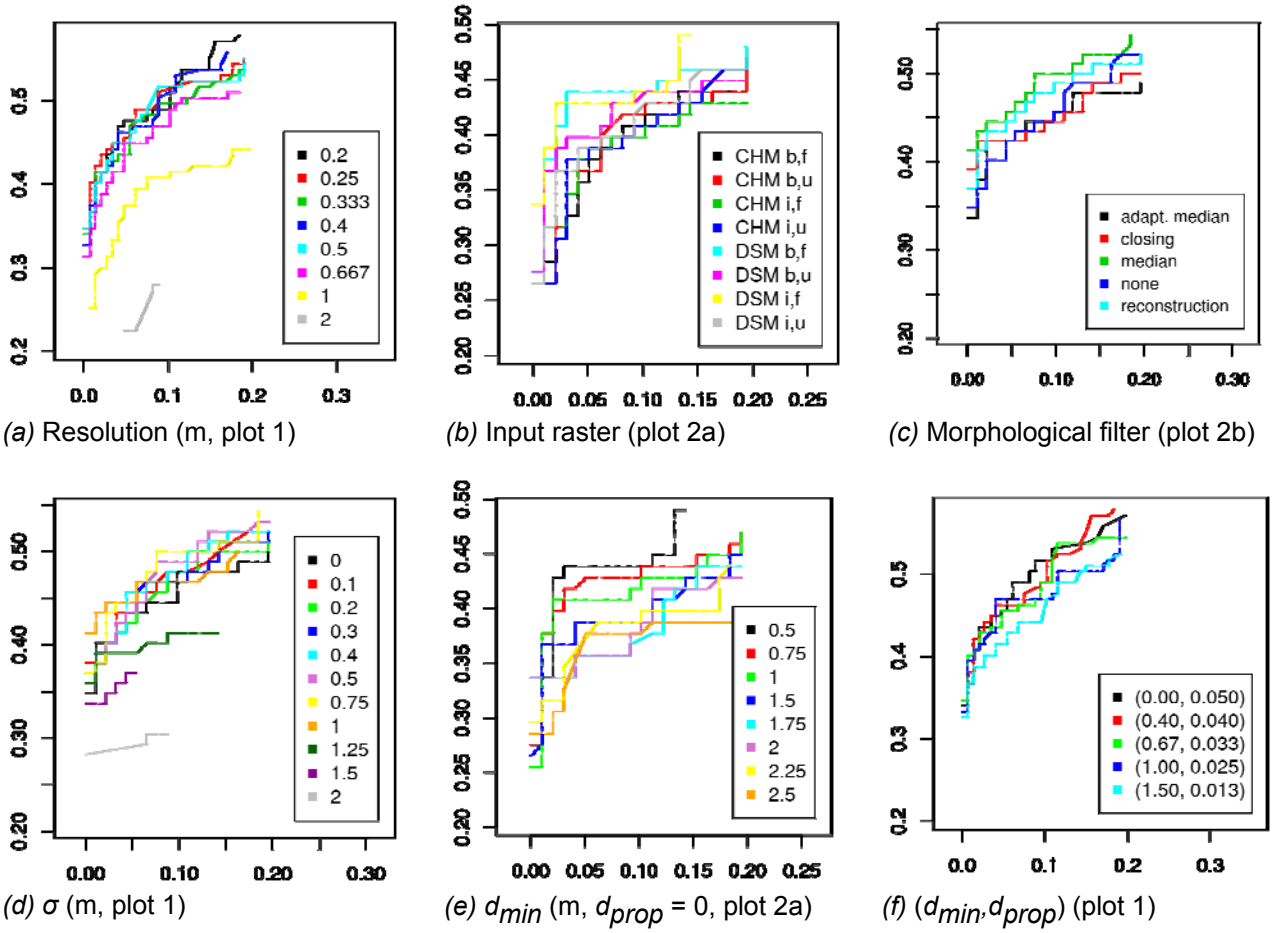
The effect of the various parameters on detection accuracy are displayed in figure 3. Due to the high number of combinations only the monotonically increasing border of the point groups envelopes are represented in the (R_{FP} ,R_{TP}) plane, with $R_{FP} \leq 0.2$ and $R_{TP} \geq 0.2$. For plot 1, resolutions below 0.5 m pixel size give similar accuracy (figure 3(a)), whereas for plots 2a and 2b parameter combinations with higher resolutions allow better results. On plot 2a, filled DSMs seem well adapted (figure 3(b)), contrary to plots 1 and 2a where unfilled CHMs obtain slightly better accuracy.

The median filter performs well in all cases (figure 3(c)) and is particularly superior to other morphological filters when employed in plot 2a on a two metres wide square window. Gaussian filter does not seem to improve performance in this plot, whereas moderate smoothing (respectively with σ around 0.8 and 0.3 m) is advantageous for plots 1 (figure 3(d)) and 2a.

Maxima selection by height thresholds does not enhance detection performance. In plot 2a, maxima selection with a fixed radius distance to the nearest higher pixel brings improvement (figure 3(e)), whereas selection based on a radius proportional to maxima height performs better with the other plots (figure 3(f)).

Stand sensitivity

Table 3 shows the detection accuracy obtained on the test plots when the parameter setting is calibrated on another plot. Plot 1 parameters result in high false positives rate in the other plots. Plot 2a settings achieve the highest number of true positives, but with a relatively large number of false detections in plot 2b. Parameters of plot 2b perform poorly. Finally global settings for plot 2 yield an acceptable compromise between detection rate and false alarms in all plots



(a) Resolution (m, plot 1) *(b)* Input raster (plot 2a) *(c)* Morphological filter (plot 2b)

(d) σ (m, plot 1) *(e)* d_{min} (m, d_{prop} = 0, plot 2a) *(f)* (d_{min},d_{prop}) (plot 1)

Figure 3. Plots in the (R_{FP},R_{TP}) plane of the monotonically increasing envelopes of points groups obtained with various settings for a given parameter.

Table 3. Accuracy of optimal settings determined on one test plot when used on another plot. Triplets are constituted of true positives rate R_{TP} (%), false positives rate R_{FP} (%) and height RMSE (m) for detected trees.

Plot	Plot 1 settings			Plot 2a settings			Plot 2b settings			Plot 2 settings		
1	46.9	4.1	2.28	38.8	4.1	3.14	4.1	4.1	1.97	35.4	2.7	3.09 [10]
2	36.3	14.7	2.03	42.6	9.5	1 .94	32.1	8 .4	2.29	38.9	3.2	1.9 [11]
2a	31.6	15.3	1.93	42.9	2.0	2.09	20.4	14.3	2.35	36.7	1.0	1.94 [12]
2b	41.3	12.0	2.11	42.4	16.3	1.76	44.6	2.2	2.26	41.3	4.4	1.86

5. Discussion

Detection results are similar to those obtained by Heurich (2008) in richly structured forests in Bavaria, with 76.9% of the trees in the upper layer identified and 5.4% detection error. On plot 2, 65.4% (resp. 84.8%) of trees above 20 m (resp. 25 m) are identified with 2.6% of false positives. In the same study, 46.8% of trees in high altitude spruce plots were determined with only 0.8% false detection, which is close to results obtained

on plot 1. However such comparisons must be handled with care due to differences in field protocols, detection methods and evaluation criteria.

Underestimation of tree heights in coniferous stands such as plot 1 is a common situation (Gaveau and Hill 2003). For mountainous areas, Hirata (2004) explained that the tilting of trees toward downslope results in an overestimation of tree height by scanner laser. Indeed, additional measures (data not shown) performed on the 26 taller trees in plot 2 show that the vertical distance between tree apex and ground is significantly greater than the vertical distance between tree apex and stem base (bias: 2.76 m, $p < 0.001$ in Student t-test).

Detection results obtained with settings calibrated on plots different from the tested plot show that in mountainous stands, algorithm performance highly depends on the parametrisation. Trends in parameter sensitivity of the algorithm may be related to both laser scanner acquisition parameters and stand characteristics. In plot 1 laser point density is lower which may explain why resolutions of pixel size lower than 0.5 m do not bring any major improvement. Indeed, there is no additional information in the point cloud to be transferred to the raster elevation models, as most of the additional pixels are empty. Regarding other parameters, plot 1 and 2b display similar tendencies. Plot 2a singularities may be explained by the high proportion of deciduous trees. Indeed, hole filling during raster computations and median filtering may have more importance when dealing with irregular, concave crowns of beech than with those of spruce or silver fir. Besides, in mixed complex stands with tree collectives, relations between tree height, diameter and crown surface are not straightforward. This may explain why basic distance thresholds perform better than diameter-height relationships for maxima selection.

It turns out from this exploratory analysis that accuracy of a basic tree top detection algorithm requires careful parametrisation. A better hindsight would be acquired by testing the algorithm on other stands and with different laser data (hardware, acquisition parameters) and also by investigating the combined influence of parameters. To reduce the range of investigated values, some parameters such as optimal raster resolution may be roughly estimated from input data characteristics. For example, Chen et al. (2006) determine raster resolution from the point cloud density and variable window size for maxima selection from the prediction intervals of the tree height relationship with crown size. Construction of a robust, generic parameters setting is also of interest when processing newly acquired laser data with no prior knowledge on the forest stands.

6. Conclusion

The analysis of the parameters influence on tree top detection in two forest stands show that identification accuracy highly depends on algorithm settings. Besides, optimal values may be linked to input data such as laser points density or forest stand characteristics (structure and species).

In order to design a semi-supervised procedure for wide area single-tree based inventories in complex areas, further research is required to investigate how prior stand knowledge or preparatory tree identification with generic settings could help determine optimal detection parameters.

Acknowledgements

J.-M. Monnet held a doctoral fellowship from Cluster de Recherche Energies (région Rhône-Alpes, France).

References

ANCELIN, P., BARTHELON, C., BERGER, F., CARDEW, M., CHAUVIN, C., COURBAUD, B., DESCROIX, L., DORREN, L., FAY, J., GAUDRY, P., GAUQUELIN, X., JOUD, D., LOHO, P., MERMIN, É., PLANCHERON, F., PROCHASSON, A., REY, F., RUBEAUD, D. and WLÉRICK, L., 2006, *Guide des sylvicultures de montagne: Alpes du Nord Françaises*.

AXELSSON, P., 2000, DEM generation from laser scanner data using adaptive TIN models. In *Proceedings of the XIXth ISPRS Congress, IAPRS*, XXXIII, pp. 110–117.

BRANDTBERG, T., WARNER, T.A., LANDENBERGER, R.E. and MCGRAW, J.B., 2003, Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in North America. *Remote Sensing of Environment*, **85**, pp. 290–303.

CHEN, Q., BALDOCCHI, D., GONG, P. and KELLY, M., 2006, Isolating individual trees in a savanna woodland using small footprint lidar data. *Photogrammetric Engineering and Remote Sensing*, **72**, pp. 923–932.

DORREN, L.K.A., BERGER, F. and PUTTERS, U.S., 2006, Real-size experiments and 3-D simulation of rockfall on forested and non-forested slopes. *Natural Hazards And Earth System Sciences*, **6**, pp. 145–153.

GAVEAU, D.L.A. and HILL, R.A., 2003, Quantifying canopy height underestimation by laser pulse penetration in small-footprint airborne laser scanning data. *Canadian Journal of Remote Sensing*, **29**, pp. 650–657.

HEURICH, M., 2008, Automatic recognition and measurement of single trees based on data from airborne laser scanning over the richly structured natural forests of the Bavarian Forest National Park. *Forest Ecology and Management*, **255**, pp. 2416–2433.

HIRATA, Y., 2004., The effects of footprint size and sampling density in airborne laser scanning to extract individual trees in mountainous terrain. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVI (Part 8/W2)* pp. 102–107.

KWAK, D.A., LEE, W.K., LEE, J.H., BIGING, G.S. and GONG, P., 2007, Detection of individual trees and estimation of tree height using LiDAR data. *Journal of Forest Research*, **12**, pp. 425–434.

MERMIN, É. and RENAUD, J., 1996, Installation de placettes permanentes dans les forêts résineuses des Alpes du Nord : objectifs, méthode, résultats. Technical report, Cemagref.

PERSSON, Å., HOLMGREN, J. and SODERMAN, U., 2002, Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering and Remote Sensing*, **68**, pp. 925–932.

POPESCU, S. and WYNNE, R., 2004, Seeing the trees in the forest: using lidar and multispectral data fusion with local filtering and variable window size for estimating tree height. *Photogrammetric Engineering and Remote Sensing*, **70**, pp. 589–604.

Delineation of individual tree crowns from small footprint, high point density LIDAR data

LINHAI JING†, BAOXIN HU†, JILI LI†, AND TOMAS NOLAND‡*

jinglinhai@hotmail.com

† Department of Earth and Space Science and Engineering, York University, Toronto, Ontario, Canada M3J 1P3

‡Ontario Ministry of Natural Resources, Ontario Forest Research Institute, 1235 Queen St. West, Sault Ste. Marie, Ontario, Canada, P6A 2E5

Abstract

A canopy height model (CHM) derived from LiDAR data can be segmented to obtain individual tree crowns (ITCs). However, branches, tree crowns, and tree clusters in a forest scene may have similar shapes, multiple scale levels, and overlapping scale ranges. The multi-scale structure of forests can cause the current ITC delineation methods for CHM to underperform, especially over deciduous forests and mixed forests. In this paper, an innovative method is proposed for ITC delineation from CHM. In this method, the CHM is reduced over multiple tree crown scale levels, the local maxima in each reduced CHM are taken as markers for the watershed segmentation of the CHM. Multiple tree crown scale levels result in multiple layers of markers and then multiple segmentation maps. After the hierarchical segmentation maps are morphologically processed, a map of multi-scale ITCs can be obtained. Tested on a mixed forest with close canopies, this method yielded an ITC map with high accuracy.

1. Introduction

LiDAR (Light Detection And Ranging) data provide height information for targets. From high-density and small-footprint LiDAR data, a canopy height model (CHM) with a small grid size can be derived and then segmented to delineate individual tree crowns (ITCs). Various ITC delineation methods have been developed for remotely sensed imagery and CHM (e.g. Gougeon 1995, Brandtberg and Walter 1998, Culvenor 2002, Pouliot *et al*. 2002, Schardt *et al*. 2002, Erikson 2003). Although successful for coniferous forests, the current ITC delineation methods for CHM underperform over deciduous forests and mixed forests.

Objects in a deciduous forest or a mixed forest, including branches, tree crowns, or tree clusters, have multiple scale levels, and the scale ranges of the objects overlap. Such a multi-scale structure causes the current ITC methods to underperform on deciduous forests and mixed forests. In the typical ITC delineation methods for CHM, treetops are first carefully localized and then used as reference points for crown delineation (Pouliot *et al*. 2005). However, a deciduous tree crown is relatively flat and its branches have peak-like shapes, and consequently, it is difficult to select a treetop from multiple local height maxima within a deciduous tree crown. In order to detect treetops, local maximum filtering with a fixed or variable-size window are widely employed (Dralle and Rudemo 1996, Wulder et al. 2000, Persson et al. 2002, Popescu et al. 2002). However, when the size of the local window varies with tree heights to cover varied-size crowns (Popescu and Wynne 2004, Chen et al. 2006), the weak correlation between tree height and crown size makes it difficult to derive accurate crown sizes and proper treetops. These issues largely are related to the multi-scale structure of forests and may be solved using multi-scale segmentation of CHM.

Multi-scale segmentation of CHM is presently widely used in ITC delineation from CHM (e.g. Falkowski *et al*. 2006, Wolf and Heipke 2007). Despite of the successes of the relevant methods, problems over forests with high canopy closure and density are often witnessed. In order to delineate multi-scale ITCs from CHM, an

innovative method is proposed in this study based on Multi-scale Segmentation and Hierarchical segment processing and is called MSH for brevity.

Study area and test data

The study area is located in Sault Ste. Marie, Ontario, Canada (46°33′56″N, 83°25′18″W), ranges from 300-306 m above the mean sea level, and is of Great Lakes-St. Lawrence forest. The most common species in the forest are jack pine (*Pinus banksiana*), black spruce (*Picea mariana*), aspen (*Populus tremuloides* Michx.), and white birch (*Betula papyrifera* Marsh.) with percentage of 50%, 10%, 20% and 10%, respectively. No cuttings occurred in recent decades and a number of stands have multi-layered canopy structures.

The LiDAR data over the study area were acquired in August 2009 with flight about 150 m high above the ground. The point density of the LiDAR data is 20-25 points/m^2. A 512×400 CHM with a grid size of 0.15 m was derived from the LiDAR data and is shown in figure 1. As demonstrated from the figure, the canopies of the forest are close and dense, and a number of between-crown gaps are as obscure as within-crown gaps.

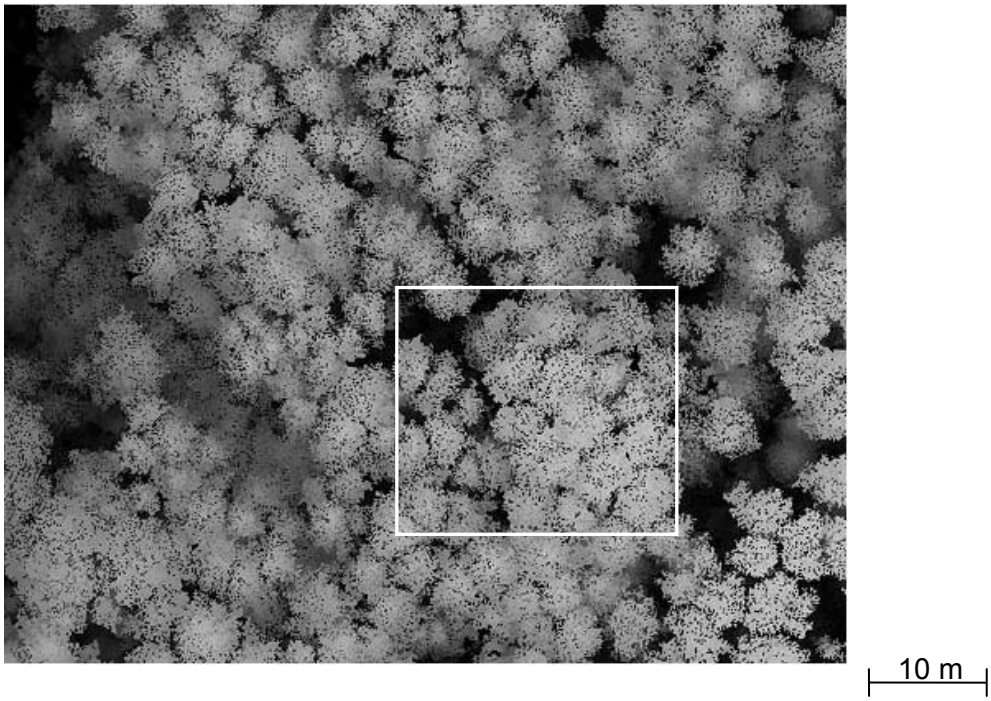


Figure 1. The CHM of the study area.

2. Methodologies

2.1Flowchart of the MSH method

As shown in figure 2, the MSH method includes three sections: scale exploration, multi-scale image segmentation, and segment processing. In the section of scale exploration, a LiDAR-derived CHM is explored using grain analysis to obtain the scale range and significant scale levels of the tree crowns in the CHM; in the section of multi-scale image segmentation, the CHM is smoothed to suppress branches and then reduced over multiple tree crown scale levels, and the local maxima of each reduced CHM are taken as extracted and act as markers for the watershed segmentation of the branch-suppressed CHM; in the third

section, the resulting multi-layered segments are morphologically merged, split up, and selected to generate an ITC map. The three sections are described in detail as follows.

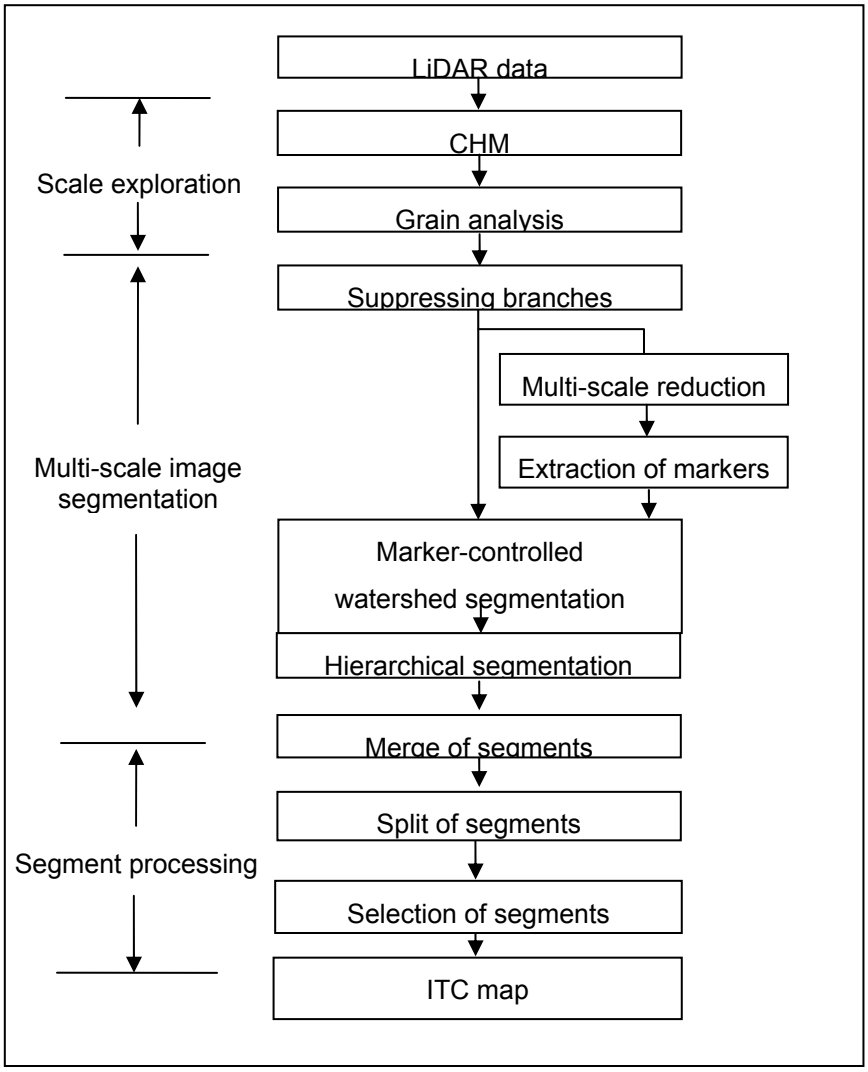


Figure 2. The flowchart of the MSH method.

2.2 Scale exploration

The CHM can be explored using grain analysis in order to obtain the scale range and the significant scale levels of tree crowns within the CHM. In this analysis, a morphological opening operation with a disk structuring element (SE) was applied to the CHM to detect whether each object within the CHM can contain the SE. In the resulting opened CHM, circular objects not smaller than the SE were retained, whereas smaller ones were sifted. As the disk diameter increased in an increment of two pixels, a series of opened CHMs were generated. For each two consecutive opened CHMs, the mean value of their difference image represents the first derivative of the first opened CHM. All of the first derivatives of the opened CHMs except the last one in the series are depicted in Fig. 3, revealing the size distribution of the objects in the forest scene. As indicated from this figure, there exist multiple significant scale groups of the objects within the CHM, including 3-5, 9-13, 17-19, and 23-27 pixels. The first group indicates the scales of the branches, and the three other groups represent the scales of small, medium, and large tree crowns, respectively. Selecting the minimum scale in each tree crown scale group, three significant tree crown scale levels of 9, 17, and 23 pixels were chosen to guide the design of the Gaussian filters used in the MSH method. Given a scale level k, the window size and σ of the Gaussian filter were set to be k and $0.5k$, respectively.

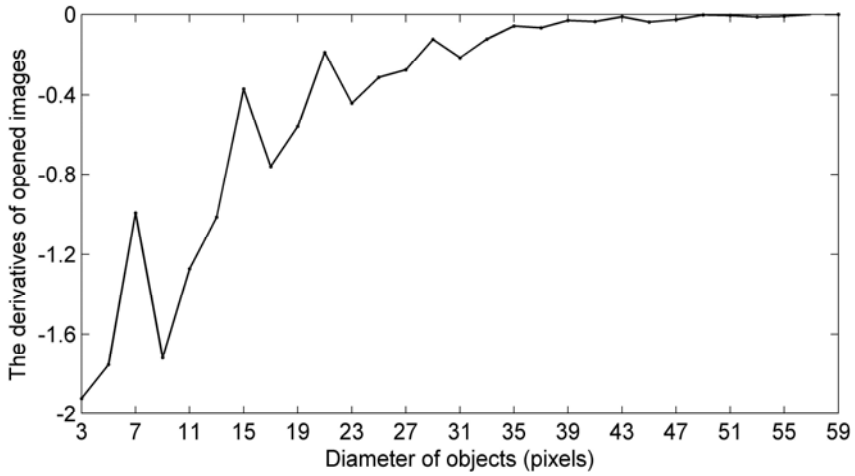


Figure 3. The size distribution of the objects within the CHM.

2.3 Multi-scale image segmentation

Gaussian filters are employed to smooth and reduce the CHM in the MSH method. In this study, after the CHM was applied with a Gaussian filter with window size equal to the minimum tree crown scale, branches were effectively suppressed, whereas multi-scale tree crowns were retained.

In order to illustrate the procedure of the MSH method, a CHM subset is taken as example (the rectangle in figure 1). This subset (figure 4(*a*)) is a mixed forest with close and dense canopies, in which the majority of trees are deciduous trees and several trees at the upper-right corner are conifers. In the branch-suppressed version of the subset (figure 4(*b*)), branches with bordering gaps were effectively eliminated, whereas multi-scale tree crowns and some sub-crowns with deeper and wider bordering gaps were remained.

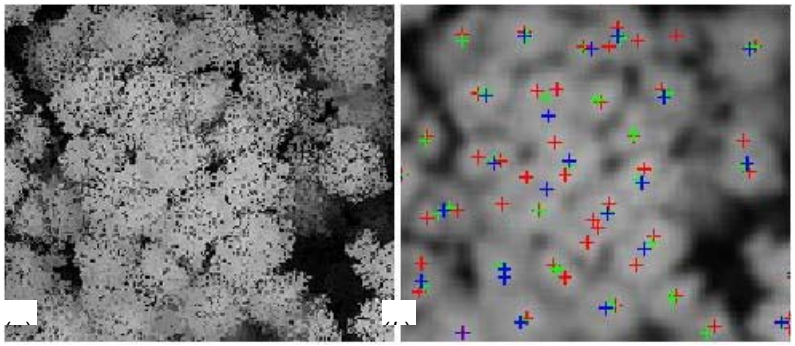


Figure 4. The CHM subset (a) and its branch-suppressed version with markers at scale levels of 9 (red crosses), 17 (green crosses), and 23 pixels (blue crosses) (b).

After branches were suppressed, the CHM was reduced over the multiple tree crown scale levels by being applied with a Gaussian filter with window size equal to the tree crown scale levels previously described. Local maximum points in each reduced CHM were taken as markers. The multiple crown scale levels resulted in multiple layers of markers at different tree crown scale levels, as shown in different colours in figure 4(*b*). It is clear that the three layers of markers are different in terms of marker count and locations.

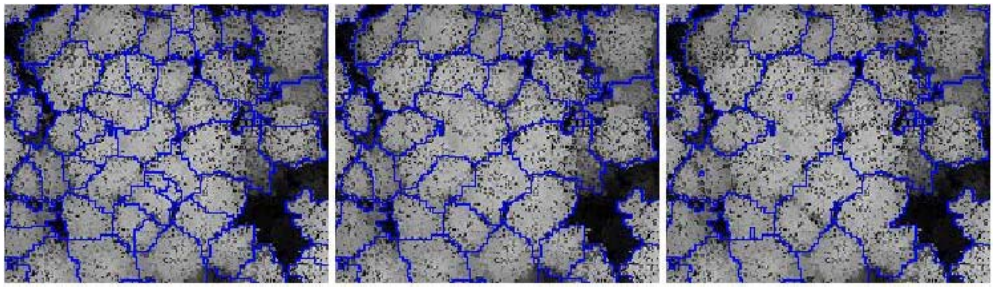
In the CHM, non-tree areas with low heights, such as bushes, grass, water, and gaps, can be delineated as follows:

$$H < T_1 E_H \tag{1}$$

where H is the CHM, E_H the mean value of the CHM, and T_1 a threshold. Threshold T_1 was set to 0.5 in this study. The resulting map consists of a number of regions. Only the regions larger than a threshold can be retained and the holes in regions need to be filled.

In the branch-suppressed CHM, the bordering pixels and inner pixels of each non-tree area were set to be zero and a maximum value of the CHM, respectively. This was done in order to isolate non-tree areas as individual segments from tree crowns in the image segmentation. Afterwards, the branch-suppressed CHM was inverted and then segmented using the marker-controlled watershed approach with each layer of markers (Meyer and Beucher 1990, Vincent and Soille 1991). The multiple layers of markers at different crown scale levels led to multiple segmentation maps. In each segmentation map, a number of segments contained non-tree areas. If a segment contained non-tree area more than a threshold, for instance, 30% of its area, it was removed.

For the branch-suppressed CHM in figure 4(*b*), its segmentation maps at crown scale levels of 9, 17, and 23 pixels are shown in figures 5(*a*)-(*c*), respectively, with segment boundaries in blue and the original CHM as background. As demonstrated from these figures, each map comprises segments of branches, tree crowns, and tree clusters. The hierarchical segments needed to be morphologically processed so that ITC segments could be extracted from the multiple maps and combined into an individual map.



(*a*) (*b*) (*c*)

Figure 5. The segmentation maps at tree crown scale levels of 9 (a), 17 (b), and 23 pixels (c), respectively.

2.4 Segment processing

Merge of segments In the previous series of segmentation maps, a coarse-scale segment (figure 6(*a*)) may superimpose multiple fine-scale segments (figure 6(*b*)). It is well known that tree crowns within CHM are visually recognized as circular peak-shaped objects. The coarse-scale segment can be refined to be as circular as possible by merging the fine-scale segment as follows:

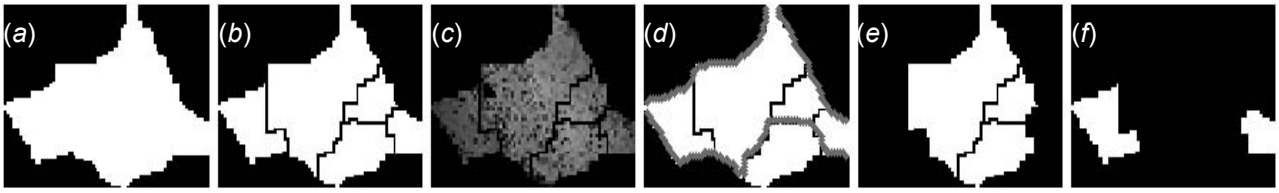


Figure 6. The refinement of a coarse-scale segment. (a) The coarse-scale segment. (b) The fine-scale segments superimposed. (c) The fine-scale segments contained in the coarse-scale segment with the CHM as background. (d) A combination of the fine-scale segments. (e) The refined coarse-scale segment. (f) The fine-scale segments excluded from the refined coarse-scale segment.

(1). Determine the fine-scale segments that are contained in the coarse-scale segment. If more than half of a fine-scale segment is covered by the coarse-scale segment, it is taken as a segment contained by the coarse-scale segment. Combine all of the contained fine-scale segments (figure 6(*c*)), together with the gaps between them, to be an intermediate segment with circularity (*c*) calculated by:

$$c = \frac{A}{\pi(\frac{D}{2})^2} \qquad (2)$$

,where *A* and *D* are the area and the equivalent diameter of the intermediate segment, respectively. The equivalent diameter of a segment is the diameter of a circle with the same area.

(2). If only one fine-scale segment is contained in the intermediate segment, substitute the coarse-scale segment with the unique fine-scale segment and exit the procedure.

(3). If more than one fine-scale segment are contained in the intermediate segment, randomly remove a fine-scale segment and assemble all the other fine-scale segments (those delimited by gray lines in figure 6(*d*)) as a combination. A series of combinations with different circularity values can be generated.

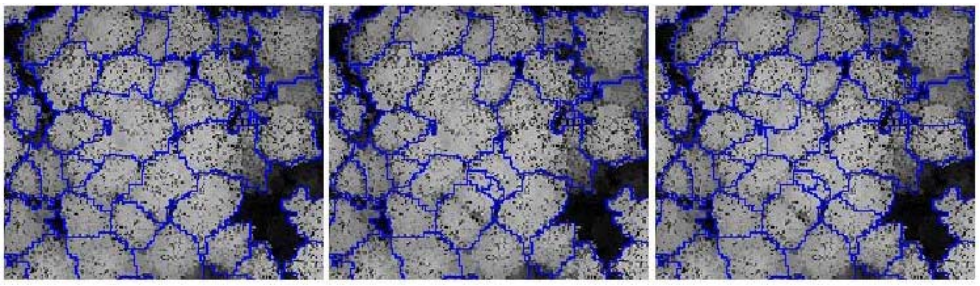
(4). Sort the preceding series of combinations regarding circularity and update the intermediate segment with the combination (figure 6(*e*)) with the maximum circularity and satisfying the requirements as follows:

$$(c_{\mathrm{c}} > c)\ \mathrm{AND}\ (A_{\mathrm{c}} > T_2 A_0) \tag{3}$$

where c_{c} and A_{c} are the circularity and area of the combination, respectively, c is the same as in equation (2), A_0 is the area of the coarse-scale segment, T_2 is a threshold, and 'AND' is a logical operator. Threshold T_2 was set to be 0.55 in this study to avoid updating the intermediate segment with a circular sub-crown segment.

(5). Repeat steps 3 and 4 until no update occurs and then substitute the coarse-scale segment with the intermediate segment. Copy the fine-scale segments (figure 6(*f*)) which are contained in the original coarse-scale segment but excluded from the updated coarse-scale segment into the updated map as individual segments.

Using the merge procedure described above, each coarse-scale segmentation map can be iteratively refined regarding all of the finer-scale segmentation maps to yield a refined version. The two refined coarse-scale maps at crown scale levels of 17 and 23 pixels were thus generated and combined with the original finest-scale segmentation map at crown scale level of 9 pixels to construct a new series of segmentation maps, which better represents the multi-scale structure of the forest scene. However, a few refined segments are tree clusters and needed to be split up.



(*a*) (*b*) (*c*)

Figure 7. The refined segmentation maps at tree crown scale levels of 17 (a) and 23 pixels (b), respectively, and the refined coarsest-scale segmentation map after tree clusters are split up (c).

Split of segments A refined coarsest-scale segment may superimpose multiple fine-scale segments and probably is a tree cluster. Such a coarse-scale segment (figure 8(*a*)) needs to be split up, if it meets the requirements as follows:

$$\begin{cases} c < T_{\mathrm{c}} \\ 2 \le n \le T_{\mathrm{n}} \end{cases} \tag{4}$$

where c is the circularity of the coarse-scale segment, n is the number of the fine-scale segments, and T_c and T_n are two thresholds. Threshold T_c was set to be 0.85 in this study due to a fact that a tree crown typically has a circularity value more than 0.85 (Wolf and Heipke 2007), and threshold T_n was set to be 5 based on a fact that a tree crown typically comprises more finer-scale objects than a tree cluster. The coarse-scale segment can be split up as follows:

(1) Determine all of the fine-scale segments (figure 8(*b*)) superimposed by the coarse-scale segment and crop them regarding the coarse-scale segment (figure 8(*c*)).
(2) Classify the fine-scale segments into two types. Type 1 is the fine-scale segments larger than an area threshold T_A. Threshold T_A was set in this study to be the area of a circle with diameter equal to the maximum tree crown scale level, i.e., $\pi(23/2)^2$. Type 2 is the fine-scale segments satisfying the following requirements:

$$\begin{cases} A_i > T_A \\ c_i > max(c_0, T_{cL}) \text{ OR } e_i > max(e_0, T_{eL}) \end{cases} \quad (5)$$

where A_i, c_i, and e_i are the area, circularity, and extent of the fine-scale segment i, respectively, c_0 and e_0 are the circularity and extent of the coarse-scale segment, respectively, T_{cL} and T_{eL} are two thresholds, operator '*max*' extracts the smaller one value from two, and 'OR' is a logic operator. Thresholds T_{cL} and T_{eL} were set to be 0.5 and 0.4, respectively. The extent of a segment is the ratio of its area to the area of its bounding box.

For a tree cluster composed of several trees of similar size, it can be assumed that all of the fine-scale segments are of the type 1 and the majority are of the type 2. These conditions can be expressed as follows:

$$\begin{cases} N_1 \geq 2 \\ N_1 - N_2 \leq 1 \end{cases} \quad (6)$$

,where N_1 and N_2 are the numbers of the fine-scale segments of types 1 and 2, respectively. If the coarse-scale segment satisfies these conditions, it can be updated with all of the cropped fine-scale segments (figure 8(*d*)).

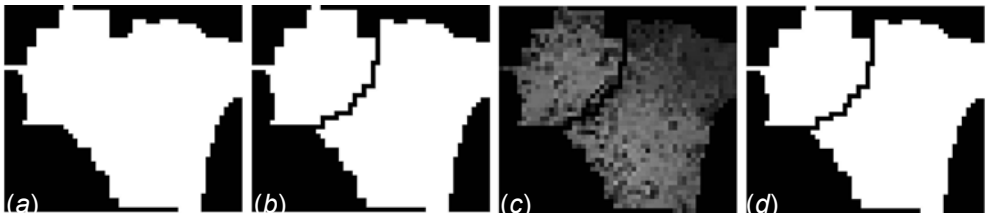


Figure 8. The split of a tree cluster segment. (a) The tree cluster segment. (b) The fine-scale segments superimposed. (c) The fine-scale segments cropped regarding the tree cluster segment and with the CHM as background. (d) The split tree cluster segment.

The refined coarsest-scale segmentation map previously described was iteratively applied with the split procedure above regarding all of the finer-scale maps to split up the tree cluster segments contained. The refined coarsest-scale segmentation map in figure 7(*b*) was thus processed, and the resulting map is shown in figure 7(*c*).

Selection of segments After tree cluster segments are split up, the refined coarsest-scale segmentation map comprises a number of segments with varied size and circularity. ITC segments can be extracted from the map as follows:

$$(A_i > T_{A1}) \text{ AND } (c_i > T_{c2}) \quad (7)$$

where A_i and c_i are the area and circularity of each segment, respectively, T_{A1} and T_{c2} are two thresholds, and 'AND' is the same as in equation (3). Thresholds T_{c2} and T_{A1} in this study were set to be 0.4 and the area of a circle with diameter equal to the minimum crown scale, i.e., $\pi(9/2)^2$, respectively.

3. Experimental results

The ITC map generated by the MSH method is shown in figure 9 with segment boundaries in blue and the original CHM as background. The CHM was also manually delineated by an independent researcher and the resulting ITC map is also shown in figure 9 with segment boundaries in red. In both maps, the segments adjacent to map boundaries were previously removed. The automatically and manually delineated maps can be taken as target and reference, respectively, for brevity. It can be seen from the figure that although the forest is dense and close, most of the target segments coincide with individual reference segments. However, it is obvious that a target segment may superimpose zero or more than one reference segment. The spatial relationships between the total 178 target and the total 177 reference segments needed to be quantitatively explored.

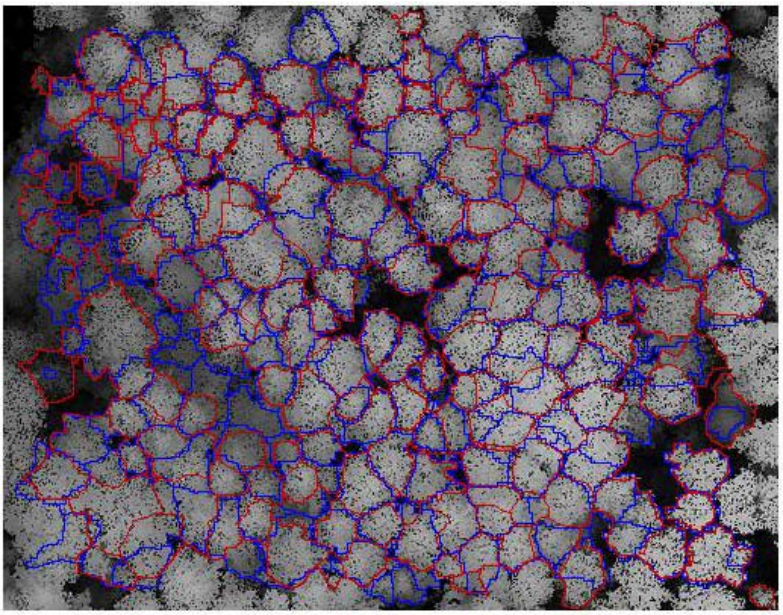


Figure 9. The overlapping target and reference maps.

Let us define that a target segment superimposes a reference segment, only if the former covers more than half of the latter. If a target segment superimposes *n* reference segments, the spatial relationship between them can be denoted as 1:*n*. Table 1 lists the statistics of the spatial relationships between the segments within the two maps. As indicated from this table, 48, 119, 8, and 3 segments have relationships of 1:0, 1:1, 1:2, and 1:3, respectively.

Table 1. The spatial relationships between the target and reference segments.

Relationship	1:0	1:1	1:2	1:3
Count	48	119	8	3
Percent (%)	27.4	66.5	4.5	1.7

About 27% of the target segments superimpose no reference segments and are shown in figure 10 and labelled. As indicated from the figure, segments 10a, 10b, 11b, 11c, 29a, 29b, and 29c are probably

individual crowns and were merged in manual delineation; segments 7, 8, 9, 18, 15, 20, 23, and 26 are covered by significantly larger reference segments containing large neighbouring between-crown gaps, and therefore are not taken as segments with 1:1 relationship; and segments 4, 6a, 6b, 14, 16a, 16b, 22, 28a, 28b, and 30 are probably small tree crowns, but they are ignored by manual delineation. Conclusively, the majority of the target segments with relationship of 1:0 are small tree crowns, but they are improperly delineated by manual interpretation. This demonstrates that the MSH method can effectively detect small tree crowns.

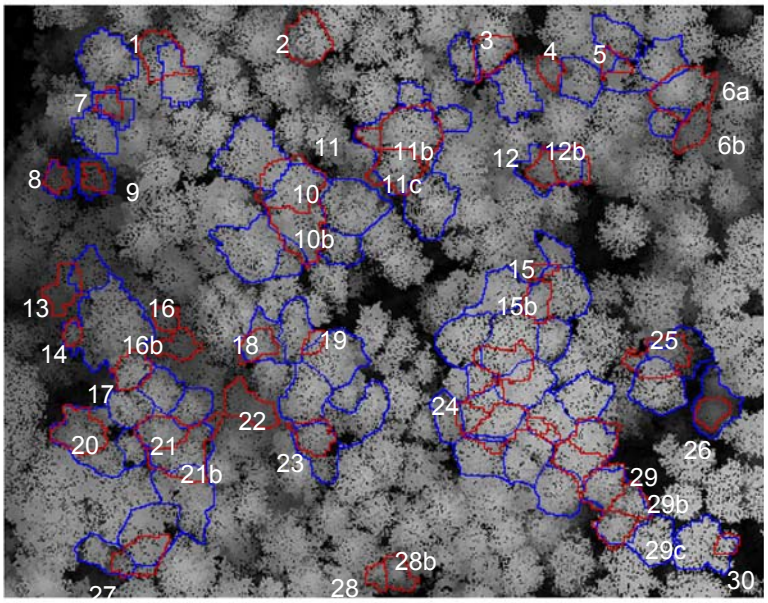


Figure 10. The target ITC segments with spatial relationship of 1:0.

4. Discussions

The bottleneck of the current ITC algorithms for CHM is to find the optimal seed pixels. In the MSH method proposed in this study, the CHM is reduced at multiple tree crown scale levels, the local maxima within each reduced CHM are taken as a layer of seed points, and then the branch-suppressed CHM is segmented using the marker-controlled watershed approach regarding each layer of seed points as markers to obtain a segmentation map. Multiple tree crown scale levels result in multiple layers of markers and then multiple segmentation maps. Since the segments in a map correspond to different objects, such as branches, crowns, and tree clusters, they need to be merged, split up, and selected regarding morphological measurements to obtain circular ITC segments.

Based on an assumption that tree crowns are circular, circularity and extent are used in the MSH method to avoid elongated tree crown segments being obtained. When a large tree crown and one or more adjoining small crowns can jointly construct a segment with higher circularity, the segment rather than the large crown will be taken as an ITC.

Acknowledgements

The authors are grateful for the financial support through the research grants provided by the Natural Sciences and Engineering Research Council (NSERC) of Canada, the Canadian Space Agency, and GeoDigital Inc. We would also like to thank GeoDigital Inc for providing the LiDAR data.

References

BRANDTBERG, T., and WALTER, F., 1998, Automated delineation of individual tree crowns in high spatial resolution aerial images by multiple-scale analysis. *Machine Vision and Applications*, **11**, pp. 64–73.

CHEN, Q., BALDOCCHI, D., GONG, P., and KELLY, M., 2006, Isolating individual trees in a savanna woodland using small footprint Lidar data. *Photogrammetric Engineering & Remote Sensing*, **72**, pp. 923-932.

CULVENOR, D.S., 2002, TIDA: an algorithm for the delineation of tree crowns in high spatial resolution remotely sensed imagery. *Computers & Geosciences*, **28**, pp. 33–44.

DRALLE, K., and RUDEMO, M., 2006, Stem number estimation by kernel smoothing of aerial photos. *Canadian Journal of Forest Research*, **26**, pp. 1228–1236.

ERIKSON, E., 2003, Segmentation of individual tree crowns in colour aerial photographs using region growing supported by fuzzy rules. *Canadian Journal of Forest Research*, **33**, pp. 1557-1563.

FALKOWSKI, M. J., SMITH, A. M. S., HUDAK, A. T., GESSLER, P. E., VIERLING, L. A., and CROOKSTON, N. L., 2006, Automated estimation of individual conifer tree height and crown diameter via two-dimensional spatial wavelet analysis of lidar data. *Canadian Journal of Remote Sensing*, **32**, pp. 153–161.

GOUGEON, F. A., 1995, A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images. *Canadian Journal of Remote Sensing*, **21**, pp. 274–284.

Meyer, F., and BEUCHER, S., 1990, Morphological segmentation. *Journal of Visual Communication and Image Representation*, **1**, pp. 21–46.

PERSSON, Å., HOLMGREN, J., and SÖDERMAN, U., 2002, Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering & Remote Sensing*, **68**, pp. 925–932.

POPESCU, S. C., and WYNNE, R. H., 2004, Seeing the trees in the forest: using lidar and multispectral data fusion with local filtering and variable window size for estimating tree height. *Photogrammetric Engineering & Remote Sensing*, **70**, pp. 589–604.

POPESCU, S. C., WYNNE, R. H., and NELSON, R. F., 2002, Estimating plot-level heights with lidar: local filtering with a canopy height based variable window size. *Computers and Electronics in Agriculture*, **37**, pp. 71–95.

POULIOT, D. A., KING, D. J., and PITT, D. G., 2005, Development and evaluation of an automated tree detection-delineation algorithm for monitoring regenerating coniferous forests. *Canadian Journal of Forest Research*, **35**, pp. 2332-2345.

POULIOT, D. A., KING, D. J., BELL, F. W., and PITT, D. G., 2002, Automated tree crown detection and delineation in high-resolution digital camera imagery of coniferous forest regeneration. *Remote Sensing of Environment*, **82**, pp. 322–334.

SCHARDT, M., ZIEGLER, M., WIMMER, A., WACK, R., and HYYPPAE, J., 2002, Assessment of forest parameters by means of laser scanning. *International Archives of Photogrammetry and Remote Sensing*, **XXXIV part 3A**, pp. 302–309.

SERRA, J., 1982, Image Analysis and Mathematical Morphology*, 1 (New York: Academic Press).*

SOILLE, P., 1999, *Morphological Image Analysis: Principles and Applications* (New York: Springer-Verlag).

VINCENT, L., and SOILLE, P., 1991, Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **13**, pp. 583-598.

WOLF, B. -M., and HEIPKE, C., 2007, Automatic extraction and delineation of single trees from remote sensing data. *Machine Vision and Applications*, **18**, pp. 317–330.

WULDER, M., NIEMANN, K. O., and GOODENOUGH, D., 2000, Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution imagery. *Remote Sensing of Environment*, **73**, pp. 103–114.

A Comparative Study of Single Tree Detection Algorithms Using Simulated Forest Plots

JUNJIE ZHANG†, GUNHO SOHN†*

† Department of Earth and Space and Engineering, York University, Toronto, Canada

Abstract

Single tree detection using small-footprint airborne LiDAR data has been extensively studied over the past few years. Researchers has proposed many different methods for this purpose, however, few of the methods has a chance to get thoroughly tested under different forest conditions, either it is because of lack of remote sensing data, or no reference data. In this paper, we present a method to simulate airborne LiDAR point cloud of forest plots with different degrees of overlapping. The point process theory was visited in the first part of the paper.

1. Introduction

Many tree detection and crown delineation using high resolution remote sensing data based on local maxima extraction from

A. The validation of the single tree detection algorithm may be troublesome because the availability of field survey data or the small sample size of field survey data the constrain of too few.
B. Make the test of the algorithm under control or in a controlled manner, to see the result of the algorithms how the algorithms work under different situation. and validation in
C. Without have to collect field data, if this data can be generated. Less expensive and time-consuming. Don't have to spend a lot of mony and time in field survey in order to get reference data and this methods is more accurate.
D. Make use of the existing data set. It is able to get value of the existing data set. If use tree model generated by software, in order to simulation the forest plot, quite usually complicated computer graphic techniques has to implemented such as ray tracing in order to get the simulated point cloud, which is very much constrained by software and interllecture reaseons.

2. Methods

Point Process for Forest Plot Simulation

This section details point process in forest statistics and how it is used to simulate forest plots with different of crown overlapping extend.

Point Process in Forestry Statistics

In forestry statistics, the distribution of trees in a forest plot can be regarded as point processes or marked point processes. The "points" are the location of the trees and the "marks" are tree characteristics such as crown diameter. Point process has been applied in the forestry statistics in two perspectives: for analyzing the spatial variability of forest stands to understand and quantify ecological relationship; and for simulating the forest pattern and predicting its dynamics (Stoyan & Penttinen, 2000).

Broadly speaking, there are two types of point process models. The basic "reference" model of a point process is the uniform Poisson point process, or homogeneous Poisson process. It corresponds to the hypothesis of Complete Spatial Randomness (CSR), which assumes points lie uniformly and are independent of each other. Such a model is implausible to in a forest ecological system, as in a typical forest,

trees growing in the neighborhood would competing for common resources and influence the growth and mortality of each other (Grabarnik & Särkkä, 2009). Although clustered patterns usually appear at the initial stage of forest development, it is observed that there is always a tendency towards regularity with the evaluation of a forest (Comas & Mateu, 2007; Stoyan & Penttinen, 2000).

All the other point process models fall within the other type of point process: inhomogeneous point process. In modern forestry statistics, two classes of such process are of particular interest: Cox and Gibbs processes. Cox processes used to be a popular class of point process models for forestry. A cox process is also called “double stochastic Poisson process” as the process can be seen as the result of a two-stage random mechanism (Stoyan & Penttinen, 2000). Matérn’s cluster process and Thomas process are two common Cox models. The drawback of the model is that there is no hard-core constraint between trees and to variable at short distances, which means it may produce locally severely clustered points.

Gibbs point process can model the pairwise interaction in the spatial pattern, which is proven to be very suitable for modeling the competition between neighboring trees. This type of process can be defined by the probability density function:

$$f(X) = \frac{1}{Z}\exp\left(-U(X)\right)$$

,where X represent the point process, Z is a normalizing constant and $U(X)$ denotes an energy function which contains a pairwise interaction function.

$$U(X) = n\alpha + \sum\sum_{i<j} \phi\left(\|x_i - x_j\|\right)$$

,where α is a local term, $\|x_i - x_j\|$ denotes the distance between point i and j and $\phi(\cdot)$ the pair potential function defines the pairwise interaction process.

According to the design of the interaction function, a number of models are further defined within Gibbs point process, such as the hard-core process, the Strauss process, the Strauss hard-core process and the soft-core process. In a hard-core process, there will be no two points located within a specified distance of each other.

The interaction between points in the Gibb point processes mentioned above are all symmetric. However, in the real case, the interactions between different trees are not symmetric. For example, bigger trees have more influence on the small trees in their neighborhood. Marked Gibbs point process can takes the “marks”, which could be both qualitative and quantitative characteristics of forests in this case, into account and provides a more powerful way to include the asymmetric competition between plants in a model with greater flexibility (Comas & Mateu, 2007; Cressie, 1992). For detailed definition of those point processes, one can reference to Neeff et al. (2005) or Comas and Mateu (2007).

Forest Plots Simulation

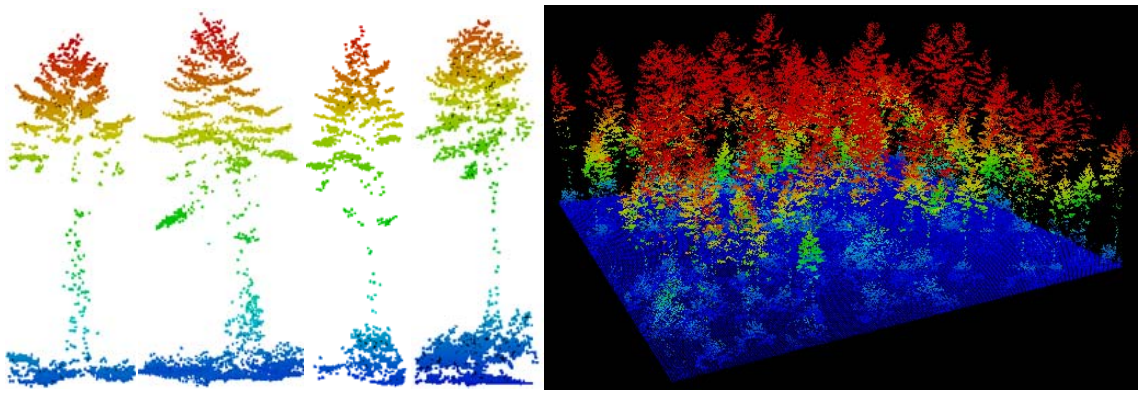
In our study, point process is employed to simulate forest plots with different extend of canopy overlapping, to test how the single tree detection algorithms work under different scenarios.

For this purpose, the “marks” of the point process are the crown diameters and we consider only the horizontal interaction of neighboring trees in the simulation, in this case the canopy overlapping. There are at least two choices for modeling the “points” - the location of trees and the “marks” - the crown diameters: they can be modelled jointly or considered as independent process and modelled randomly, and different choices

typically lead to different stochastic models. We here choose the later as a relative simple but effective way to do the simulation. We first model the location.

As rigidly quantification as to what extent the crown are overlapped is not need, we choose the hard-core process to model the spatial distribution of tree locations, and then attach the “marks”, which adhere to the empirical distribution, to the “points”. The empirical distribution of crown width is acquired from the field survey, and the marks adhere to this distribution can be generated. In the point process simulation, Markov chain Monte Carlo (MCMC) method combined with Hastings-Metropolis and spatial birth-and death algorithms are used to simulate the target point process. In the model, the interaction between trees is controlled by the distance specified. If the distance is set large, for example, as the mean plus the standard derivation of the crown width, the chance the tree crowns overlap will be small and vice versa. Three scenarios are simulated in this way, by setting the interaction distance equals to the mean plus the standard derivation of the crown width, the mean of the crown width and the mean minus the standard derivation of the crown width, representing forest plots with separating canopy, canopy slightly touch each other and overlapping canopy respectively.

To simulate the LiDAR point clouds of a forest plot, tree templates which can be clearly identified are manually selected and clipped from multiple return, high density, small-footprint airborne LiDAR dataset acquired in forest area (see Figure 1-a). Then tree characteristics of the tree templates, such as tree height and the crown width, are extracted. Afterwards, those tree templates are randomly picked and located on each site of the point process after being scaled according the ratio of the “marks” to its crown width and rotated at a random angle using automatic algorithm (see Figure 1-b). In the simulated forest plots, all the tree parameters are exactly known and are used as reference data to test the single tree detection algorithm.



(a) (b)

Figure 12. (a) Four tree templates clipped from LiDAR point clouds acquired in forest plots for illustration; (b) A simulated forest plot using point process and the tree templates library.

Single Tree Detection Algorithms

Two single tree detection algorithms are investigated about detection performance using the point clouds of the simulated forest plots with different degrees of canopy overlap. One algorithm is local maxima approach, and the other is a Markov Random Field based model for single tree detection presented by Zhang and Sohn (2010).

Local Maxima Approach

Local maxima (LM) approach has been extensively used to detect single trees from remote sensing imagery (Pouliot, King, Bell, & Pitt, 2002; Wulder, Niemann, & Goodenough, 2000) and airborne LiDAR data (Chen,

Baldocchi, Gong, & Kelly, 2006; Popescu, Wynne, & Nelson, 2002; Reitberger, Schnörr, Krzystek, & Stilla, 2009). A local maximum filter passes over an image to find the pixels having a larger value than all the pixels in the window and these local maxima are identified as tree tops.

Researches of Wulder et al. (2000), Popescu et al. (2002) and Chen et al. (2006) all suggested that the determination of the LM filter window size has influence of the detection results, and variable size window performs better than static size window in term of reducing both commission and omission errors, considering the relationship between the tree height and canopy width.

In our research, how the window size affects the LM detection results on LiDAR data is further explored using the three simulated forest plots. Circular LM filter is used and the ratio of the LM window size to the pixel value will be set larger, nearly equal and smaller than the mean crown width to tree height ratio of the tree templates, to see its effect on the detection rate.

Markov Random Field Model

This method highlights a novel Markov Random Field model to detect single trees from ALS data. It tries to find the optimal configuration of singles trees ALS through an energy minimization scheme.

To achieve this, we firstly overpopulate the local maxima from the CHM recovered from ALS data using a circular type of LM filter with variable window size. Next, trees are modelled as objects and the neighbourhood system is set up using TIN. Then, energy functions are carefully designed to incorporate constraints for penalizing false trees and favour true ones. Finally, the optimal tree models are obtained through an energy minimization process.

The strategy of this method is to minimize the omission error detecting tree tops using relatively small LM filter. Then a Markov Random Field model is defined on all the treetops. Energy formulation based on a data term which measures how features extracted from the data support the object as an individual tree, such as symmetry of the crown and valley depth between the tree crowns, and a contextual term which take into consideration some interactions between neighbouring objects are designed. As the false treetops are unfavoured in the model, they will be finally be pruned from the model during the optimization process. This method was applied on ALS data acquired from a coniferous forest and experimental results showed a good detection rate (Zhang & Sohn, 2010).

3. Experimental Results and Discussion

LiDAR Data and Field Survey Data

The ALS data used in this study was acquired in a coniferous forest area about 60 km east to Sault Ste. Marie, Canada by Riegl LMS-Q560 in August, 2009. The field survey was conducted during the same time. In the fieldwork, plot location, tree species, tree height and crown width and some other forest characteristics were noted down. These field survey data was used to estimate the empirical distribution of tree characteristics, such as crown width, which obeys a Gaussian distribution with mean of about 5.25m and the standard deviation of 1.25m. The field data is also helpful for identifying suitable tree templates from the ALS data.

Simulated Forest Plots

Three forest plots were generated using hard-core process. The sizes of the plots were all 100×100 m, and homogeneous Poisson processes with intensity of 0.025 were used as the starting points in the construction

of hard-core process. Figure 2(a)-2(c) show us the resulting point processes. The tree numbers of the three forest plot were 186, 234 and 261 respectively.

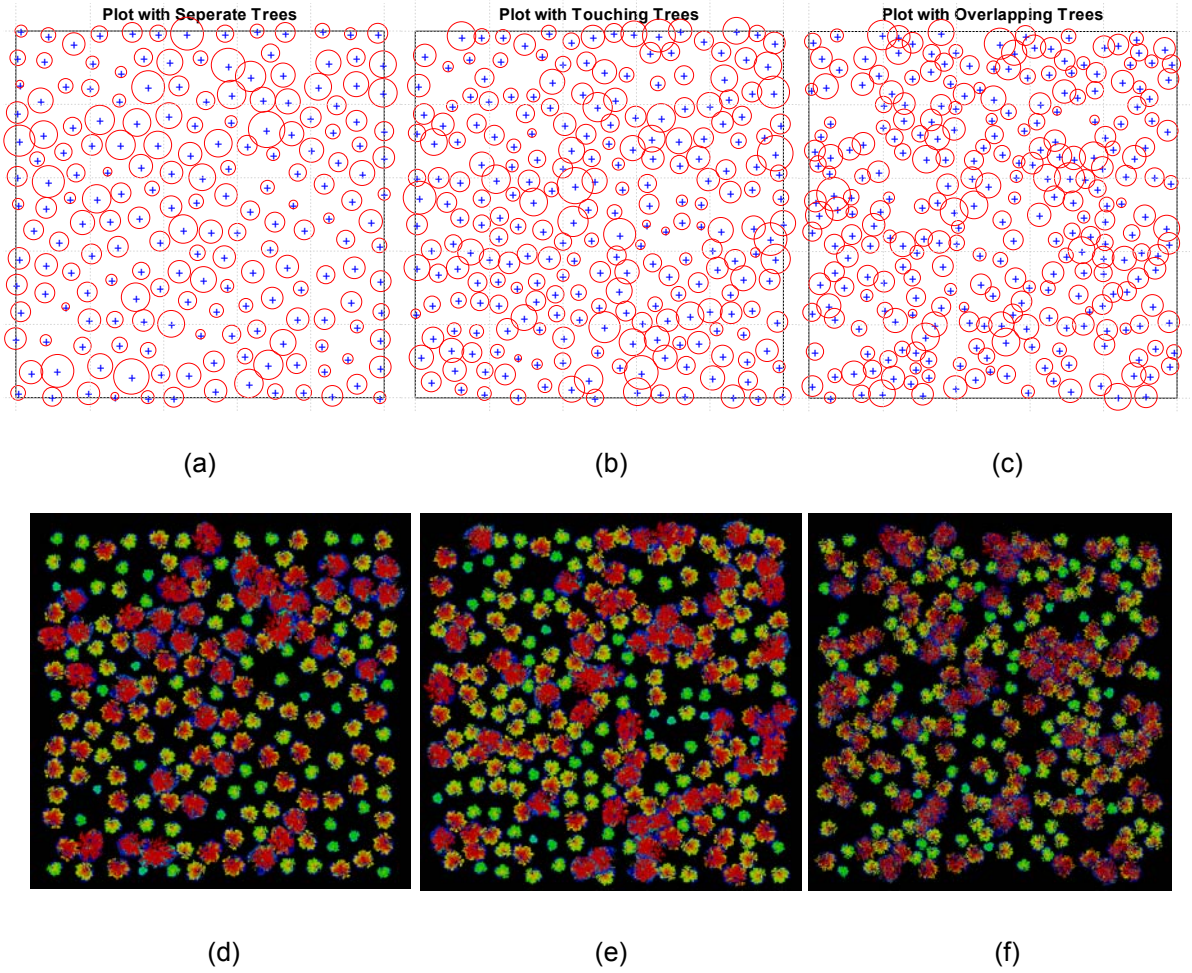


Figure 13. (a)-(c) Point process simulated forest plots with three different degrees of overlapping: (a) plot with separating crown; (b) plot tree crown slightly touching each other; (c) plot with overlapping trees. (d)-(f) the corresponding point clouds of the three forest plots generated.

The interaction distance used in the hard-core process to simulated forest plot with separate trees was set as the mean plus the standard derivation of crown width. As we can see from Figure 2(a), the trees are mostly separated and no overlaps. The interaction distances used in the hard-core process were set smaller and from Figure 2(b)-2(c), more trees were overlapped and the degree of overlapping increased. Figure 2(d)-2(e) show the corresponding point clouds of the three forest plot.

LM Detection Results

The minimum ratio of the crown width to tree height of the entire tree templates was 0.156. In this case, three ratios of variable size window to pixel value were used to test the performance of LM filter under three forest conditions with increasing degree of overlapping. Table 1 shows the LM detection results in the three forest plots. The three ratio used were 0.25, 0.20 and 0.15. The third ratio 0.15 is a little smaller than the minimum crown width to tree height ratio, which means the LM filter window size would always be smaller than the actual crown size, and we can see from the table, small window size produce a great number of commission errors; while the first ratio 0.25 is a little larger, we can see the omission error produced by large window size

is also the largest. Among the three forest plots, the detection rate of LM on the plot with separating canopy crowns is best. The omission errors were low, which means without canopy occlusion from surrounding trees, trees can be much easily detected. When the overlapping degree increases, it is very likely shorter trees occluded by taller tree in the neighborhood can't be detected. For the same variable window size ratio, in terms of omission error, LM always performs worse on plot which has a higher degree of overlapping. What can be concluded as well is that smaller window size is not always good. When the window size ratio of LM filter drop to a certain degree, the omission error reduces really slowly, but meanwhile, the commission error can increase exponentially and greatly degrade the overall detection quality, e.g. for separating plots, when window size ratio drop from 0.20 to 0.15, the omission error seldom change, while commission error increased from 37% to 209% and overall quality decreased from 72.2% to 32.1%.

Table 13: LM detection results of three forest plots

Ratio of Variable	Detected	Correct		Commission		Omission		Overall
Size Window	Trees	no	%	no	%	no	%	Quality
Plot 1 - separating stand; 186 trees								
0.25	192	184	98.9%	8	4.3%	2	1.1%	94.8%
0.20	253	184	98.9%	69	37.1%	2	1.1%	72.2%
0.15	575	185	99.5%	390	209.7%	1	0.5%	32.1%
Plot 2 - touching stand; 234 trees								
0.25	221	215	91.9%	6	2.6%	19	8.1%	89.6%
0.20	312	220	94.0%	92	39.3%	14	6.0%	67.5%
0.15	674	227	97.0%	447	191.0%	7	3.0%	33.3%
Plot 3 - overlapping stand; 261 trees								
0.25	218	214	82.0%	4	1.5%	47	18.0%	80.8%
0.20	311	239	91.6%	72	27.6%	22	8.4%	71.8%
0.15	697	245	93.9%	452	173.2%	16	6.1%	34.4%

MRF Detection Results and Comparison

MRF detection was then applied on the LM detection results with variable window size ratio of 0.20 to get rid of the commission errors. LM detection Results from ratio 0.20 were chosen for at this ratio, the omission error was not so high while the commission error in not so much. Figure 3 provides a clear impression detection results before and after MRF. Exact detection results can be checked in Table 2.

Figure 3(a)-(c) show the LM detection results with variable window size ratio of 0.2 for the three forest plots, from which we can see all for the three plots LM produced lots of commission errors (green circle). With the increase

of overlapping degree, more cyan dot line circle appear, which mean the trees were undetected and were omission errors.

Figure 3(d)-(f) show the detection result after MRF. As can be easily interpreted, most of the green circles have been successfully remove, which means this method could effectively remove the commission errors. As it is shown in Figure 4, the average commission error of about 35% dropped to about 1.6%. We can also see some yellow dot line circles show up in those three figures, which means those trees were over-pruned during the process. However, this didn't influence the final result very much. Averagely, the overall detection quality increased about 20% after MRF detection.

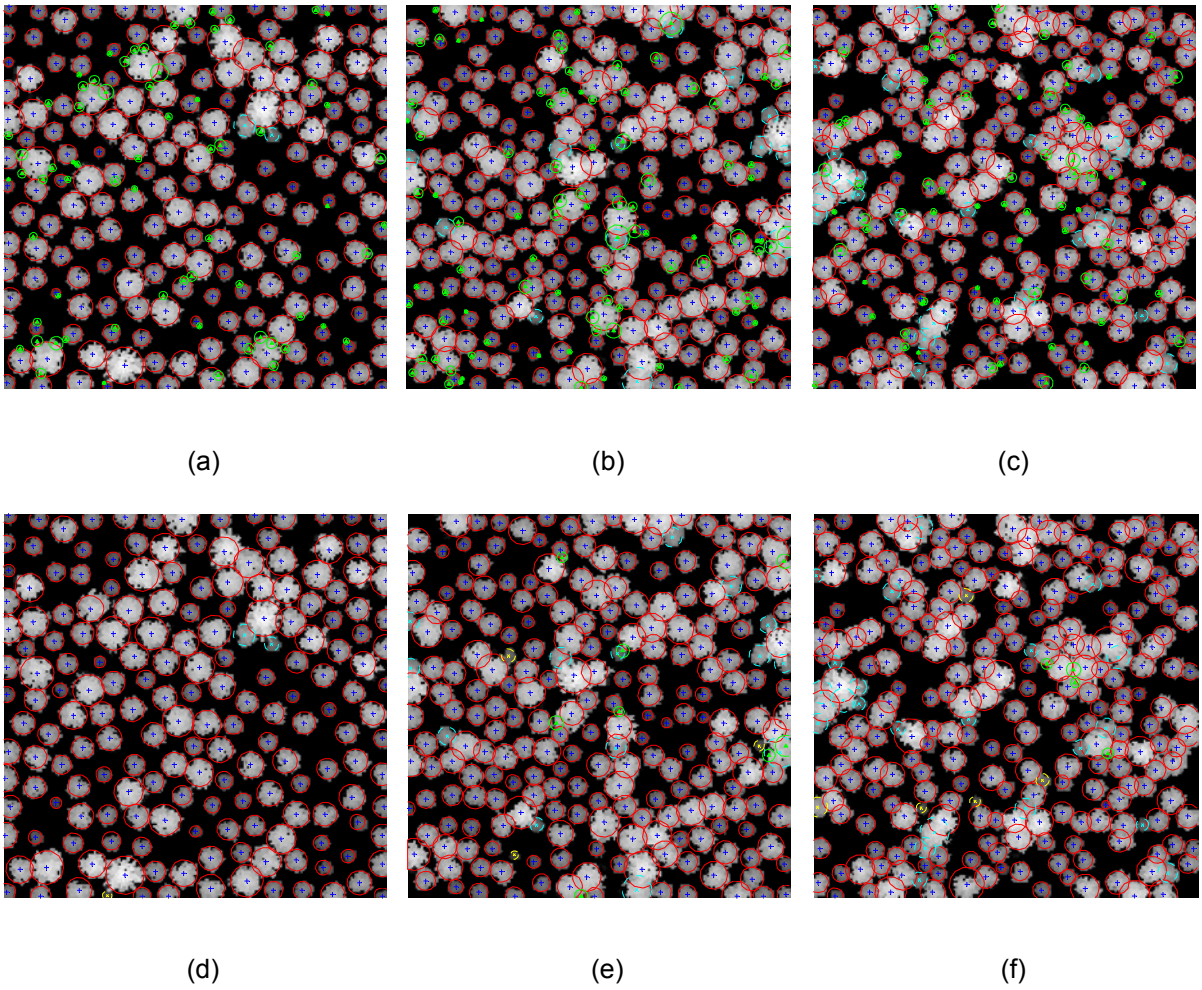


Figure 14. MRF detection results improve on LM detection results. (a)-(c) show the LM detection results on the three forest plots with variable window size ratio of 0.2; (d)-(f): show the MRF detection results using the corresponding LM detection as initial setting. (Green circles represent for commission errors; Cyan dot line circles represent for omission errors from LM detection, and yellow dot line circles represent for omission errors from MRF detection.)

Table 2: MRF detection results when compared with LM detection results

	Detected	Correct		Commission		Omission		Overall
	Trees	no	%	no	%	no	%	Quality
Plot 1 - separating stand; 186 trees								
VSW 0.2	253	184	98.9%	69	37.1%	2	1.1%	72.2%
MRF	183	183	98.4%	0	0.0%	3	1.6%	98.4%
Plot 2 - touching stand; 234 trees								
VSW 0.2	312	220	94.0%	92	39.3%	14	6.0%	67.5%
MRF	226	218	93.2%	8	3.4%	16	6.8%	90.1%
Plot 3 - overlapping stand; 261 trees								
VSW 0.2	311	239	91.6%	72	27.6%	22	8.4%	71.8%
MRF	238	234	89.7%	4	1.5%	27	10.3%	88.3%

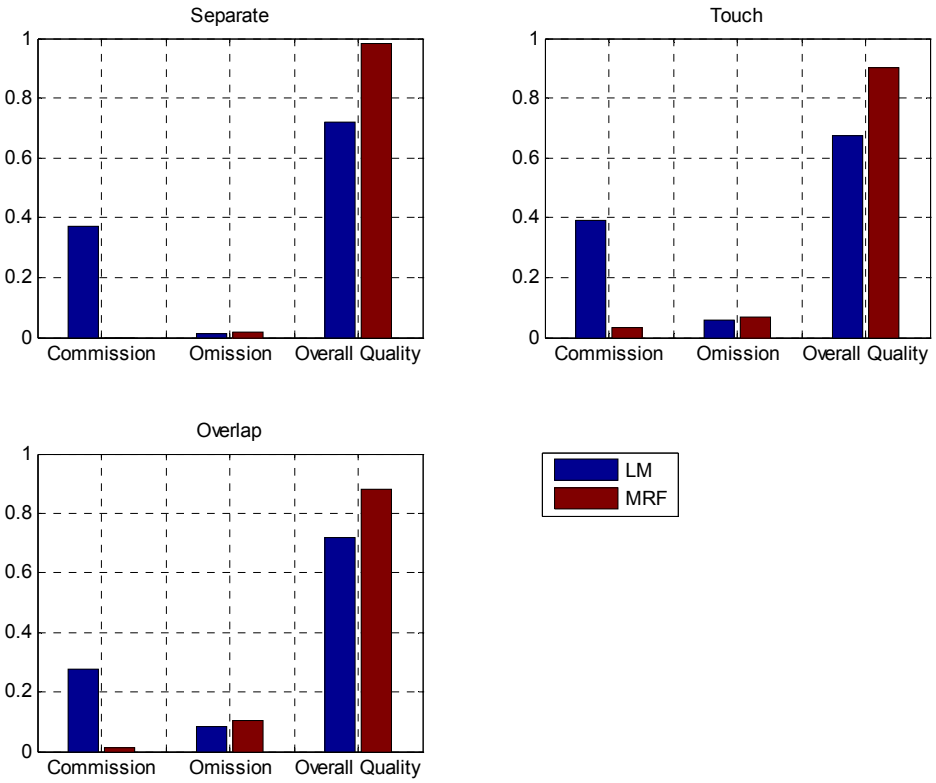


Figure 15. Comparison of LM detection results with that improved by MRF detection.

4. Conclusions

We have shown in this paper using point process to simulate airborne LiDAR data of forest plots with different degree of overlapping. The simulation technique using point process provides a powerful tool to the fields where the complex forest ecosystems need to be better understood. The simulated forest plots provide excellent reference dataset and fully controlled environment to test two single tree detection algorithms, which allow us to have a better understanding of how they work. The result showed that LM detection methods are both sensitive to filter window size as well as specific forest conditions, such as crown overlapping, which may result in commission and omission errors. These simulated forest plots also provide ideal dataset for the testing another single tree detection based on Markov Random Field model to see how it works under different forest situations. The drawback of the method is that it bases on LM approach and can't recover the omission error made by LM. This issue will be addressed in future research.

Acknowledgements

This research was funded by GeoDigital International Inc., Ontario Centre for Excellence (OCE) and NSERC.

References

Chen, Q., Baldocchi, D., Gong, P., & Kelly, M. (2006). Isolating Individual Trees in a Savanna Woodland Using Small Footprint Lidar Data. Photogrammetric Engineering & Remote Sensing, 72(8), 923-932.

Comas, C., & Mateu, J. (2007). Modelling Forest Dynamics: A Perspective from Point Process Methods. Biometrical Journal, 49(2), 176-196.

Cressie, N. A. C. (1992). Statistics for spatial data. New York, USA: John Wiley & Sons, 920 pp., ISBN 0-471-84336-9 (Vol. 14): Elsevier.

Grabarnik, P., & Särkkä, A. (2009). Modelling the spatial structure of forest stands by multivariate point processes with hierarchical interactions. Ecological Modelling, 220(9-10), 1232-1240.

Neeff, T., Biging, G. S., Dutra, L. V., Freitas, C. C., & dos Santos, J. R. (2005). Markov point processes for modeling of spatial forest patterns in Amazonia derived from interferometric height. Remote Sensing of Environment, 97(4), 484-494.

Popescu, S. C., Wynne, R. H., & Nelson, R. F. (2002). Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size. Computers and Electronics in Agriculture, 37(1-3), 71-95.

Pouliot, D. A., King, D. J., Bell, F. W., & Pitt, D. G. (2002). Automated tree crown detection and delineation in high-resolution digital camera imagery of coniferous regeneration. Remote Sensing of Environment, 82, 322-334.

Reitberger, J., Schnörr, C., Krzystek, P., & Stilla, U. (2009). 3D segmentation of single trees exploiting full waveform LIDAR data. ISPRS Journal of Photogrammetry and Remote Sensing, 64(6), 561-574.

Stoyan, D., & Penttinen, A. (2000). Recent Applications of Point Process Methods in Forestry Statistics. Statistical Science, 15(1), 61-78.

Wulder, M., Niemann, K. O., & Goodenough, D. G. (2000). Local Maximum Filtering for the Extraction of Tree Locations and Basal Area from High Spatial Resolution Imagery. Remote Sensing of Environment, 73(1), 103-114.

Zhang, J., & Sohn, G. (2010, 1-3 September). A Markov Random Field Model for Individual Tree Detection from Airborne Laser Scanning Data. Paper presented at the Photogrammetric Computer Vision and Image Analysis, Paris, France.

Experimental investigation of geometric features extracted from airborne LiDAR for tree species classification

CONNIE KO†, GUNHO SOHN† and TARMO K REMMEL‡*

cko@yorku.ca

†Department of Earth and Space and Engineering, York University, Toronto, Canada

‡Department of Geography, York University, Toronto, Canada

Abstract

We present a methodology to classify individual trees into a given list of species obtained by field surveying using airborne LiDAR data; targeted species include birch, maple, oak, poplar, white pine and jack pine at a study site northeast of Sault Ste Marie, Ontario, Canada. The average density of the LiDAR data is 40 points per m^2 and was acquired in August 2009. We investigate approaches involve extraction and derivation of 3D lines (branches and bole) and 3D polygons (tree crown) from a given point cloud. These features will be quantified and used as identifiers for tree crown shape, distribution and orientation of branches and the bole. The first goal of this project is to identify important, general, and unique geometric markers for the different tree species and to use these geometric markers to classify other trees by comparing their geometric traits with these training markers. Some characteristics that we consider include individual branch segment lengths and their respective connecting angles with the bole. Drawing those line and polygon features is not only useful for visualizing the structures of individual trees, they also permit the determination of biophysical parameters such as growth characteristics, tree age, and potentially species; each being useful in applications including biomass measurement (growth and accumulation) or timber harvest (wood volume) calculations. A second goal is to integrate data driven models into rule-based models that fit branching structures to 3D point clouds for classification purposes; the original idea was inspired by L system (*Lindenmayer* system), a re-writing language (a stepwise language that allows geometric features to be repeated at different orders) to create tree branching structures with two sets of user defined rules; an axiom (similar to a trunk) and production (similar to branches). This new method involves deriving rules from the LiDAR point data to reconstruct the internal branching structures.

1. Introduction

The use of Light Detection and Ranging (LiDAR) for forest applications is becoming an important method in retrieving attributes for forest inventories because of its ability in mapping objects in three dimensions (Gobakken and Naesset, 2004; Hilker et al., 2008; Korpela et al., 2007), common attributes retrieved in these studies include tree height and diameter. Among many of the forest inventory attributes, tree species is particularly valuable information to have because of the implication to biomass estimation, forest composition, wood volume, wildlife habitat, or fire assessments. Existing researches suggested that tree species identification can be done by using combining LiDAR and spectral data (Hilker et al., 2008; Holmgren et al., 2008; Koukoulas and Blackburn., 2005; Persson et al., 2004; Hill and Thomson, 2005) or by investigating the geometric and spatial distribution of the point clouds. Some of the geometry based studies includes Barilotti et al. (2009), they fit different curved surfaces to the tree tops of different species include spruce, fir larch and beech. Kato et al. (2009) used an isosurface algorithm to derive a surface that wraps around tree crowns for retrieving tree parameters such as crown volume, crown width and crown base height for douglas-fir, western red ceder, tulip tree, atlas ceder, oak and maple. Spatial (height and/or intensity) distribution related studies include Holmgren and Persson, 2004; Brandtberg., 2007; Ørka et al., 2009; Kim et al., 2009; Korpela et al., 2009 and Suratno et al., 2009. Common attributes that are used for species classification in these studies include percentage of first and single returns; mean, standard deviation for height for first return and all returns; mean, standard deviation for intensity for first return and all returns. These attributes can be obtained for the entire tree height profile (e.g., Brandtbert, 2007; Kim et al., 2009) or by height percentiles (e.g., Ørka et al., 2009; Holmgren and Persson, 2004; Korpela et al, 2009).

We would like to present a method that is based on geometric parameter derivation, specifically line features and these line features will be used to represent bole and branches. Reconstructing the line features forms the skeleton of the tree and therefore assigns contexts to a tree, e.g., branch angles, lengths, growth direction, tree crown shape, number of branches. The context information is useful in classifying tree because tree forms are alike within one species. Some species such as white pine, jack pine, balsam fir and white spruce have stronger linear features, as a result, we have better success with feature reconstruction with those speices; in this paper we will put our focuses in those species.

Study area and datasets

LiDAR dataset

LiDAR data was collected on August 7, 2009 at about 75 km east of City of Sault Ste. Marie, Ontario, Canada. Scanner for point cloud acquisition was Riegl LMS-Q560 at an altitude of about 122 m above ground. Point density is approximately 40 pulses per m^2 with multiple returns per pulse. In our study area, we identified tree species such as white birch, balsam fir, hard maple, soft maple, red oak, jack mine poplar, white pine and white spruce in the area by field surveying.

Field survey

A field survey was conducted from July 30 to Aug 12, 2009, the major task of the survey was to identify tree species and measure tree attributes of those trees. Item we have measured include tree height, tree crown base height, tree crown diameter (measured in two perpendicular directions) and diameter breast height (dbh). These measured values will be used to compare the values obtained or derived from the LiDAR dataset for validation purposes.

2. Methods and results

2.1 Bole derivation

The individual tree that has been field surveyed is identified in the LiDAR dataset and segmented manually. At this stage of the research we are interested in looking at features in the tree crown therefore the point cloud that is observed as understory is removed from the individual tree. The first level of line feature extraction is the bole, or the main growth direction. The bole of the trees is derived by the method described in Ko et al. (2009); it is by connecting the centriods of a downward moving voxel. Figure 1 shows some of the derived bole for tree sample.

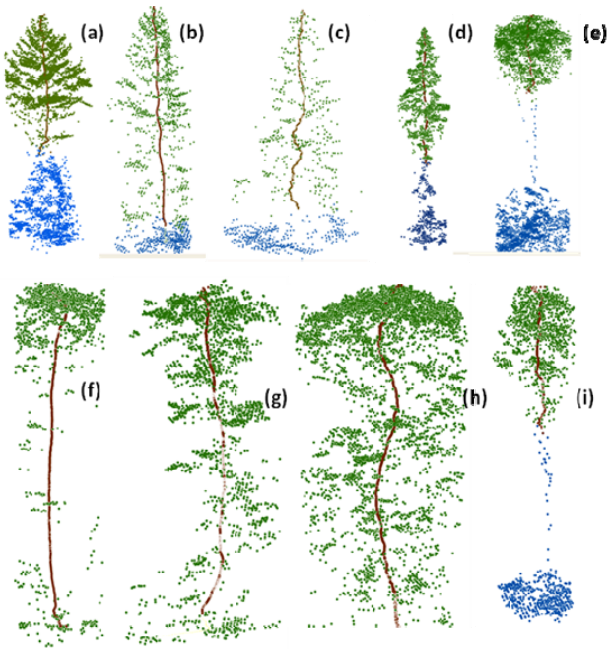


Figure 1. The derived bole by downward moving voxel for tree samples (a) white pine (b) jack pine (c) balsam fir (d) white spruce (e) white birch (f) soft maple (g) red oak (h) hard maple (i) poplar. For tree (a), (b), (c), (d), (e) and (i), blue points indicate understory and are those points were not included in the bole derivation algorithm.

2.2 Branch derivation

The second level of line feature extraction is the branches extraction; coniferous species such as white pine, jack pine, balsam fir and white spruce were observed to have strong sensation of branching line features whereas deciduous species such as hard maple, red oak, white birch and poplar does not. To extract those line features, we have chosen three methods as an initial attempt; 1) combining k-means clustering and principle component analysis (PCA) to extract those line features, 2) random sample consensus (RANSAC) and 3) Hough transformation.

k-means clustering and principle component analysis

Each tree will be clustered from k=2 to k=100 or until an empty cluster is produced, four validity tests are be used to identify four possible numbers of clusters and the most optimal number of clusters will be decided by visual interpretation. In the future, this issue will be re-addressed so that the optimal number of clusters can be calculated without the visual verification. The validity test is Davies-Bouldin Index (Davis and Bouldin, 1979), the second test is the by silhouette plot (Rousseeuw, 1987), the third is the Krzanowski-Lai index (Krzanowski and Lai, 1985) and the forth is *Calinski* and *Harabasz index (Calinski and Harabasz, 1974). Figure 2 shows the result of the four validity tests on one of the white pines in the study area and we have chosen to use 49 clusters for this particular tree.*

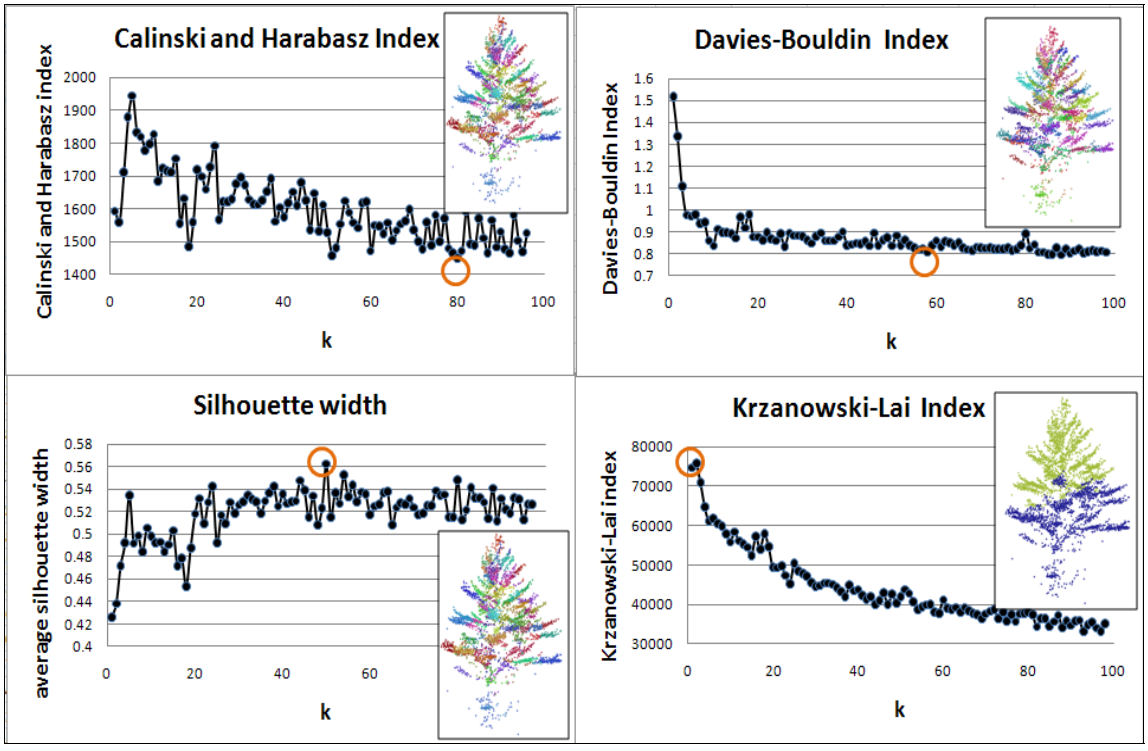


Figure 2. The sample white pine is being tested for the best number of cluster, the Calinski and Harabasz Index indicate 81 clusters is the optimum, Davies Bouldin Index indicate 58 clusters, Silhouette width indicate 49 clusters and Krzanowski-Lai Index indicate 2 clusters. The results of the clusters are located at each graph and the optimal number of cluster is circled in orange.

In Ko et al. (2009), branches (line features) were drawn by connecting closest point to the bole in each cluster to the furthest point. However, using this method can result in geometric misrepresentation; this is because the lines drawn this way does not represent the major direction of the clusters, and is very sensitive to the presence of outliers. For our case, the direction of the clusters is important (representing branch

growth direction), therefore we have implemented PCA for line feature extraction to overcome this problem. Figure 3 shows a schematic diagram of the advantage of using PCA method to extract line over connecting line by closest to furthest point within the cluster. Figure 3(*a*) shows that the lines can be misrepresented when joining the closest point to the furthest point and figure 3(*b*) shows that this problem can be avoided by using PCA analysis for line extraction. The PCA method calculates the principle components of each of the clusters in the tree, the first principle component will be used to define the direction vector of the line and the end points of the line is bounded by the extent of the cluster in Cartesian space. This way, the line extracted will represent the major direction of the cluster. The orthogonal distances from each point to the line is also calculated as an indicator of how well the line features fit into each cluster. Although the original rationale for connecting the closest point to the furthest point was to make sure the line features derived from the cluster will be able to connect to the bole whereas the branches derived by the PCA method may or may not intersect the bole. However, the deviation of the end points to the bole can be used to assess of the quality of the cluster. Figure 4(*a*) shows the result of the line features derived by PCA analysis for one of the tree samples (white pine, same tree as shown in figure 1(*a*)). Figure 4(*b*) shows the result of the line features derived by PCA analysis for another tree samples (birch, same tree as shown in figure 1(*e*). Conifer trees appear to have strong linear branch features whereas deciduous trees do not.

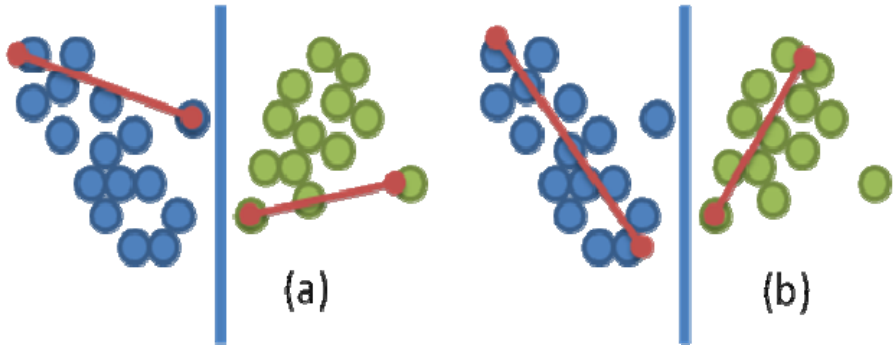


Figure 3. (a) shows the line features extracted by joining the closest point to the furthest point can be misrepresented and (b) shows that this problem can be avoided by using PCA analysis which detects the major direction of the point clusters.

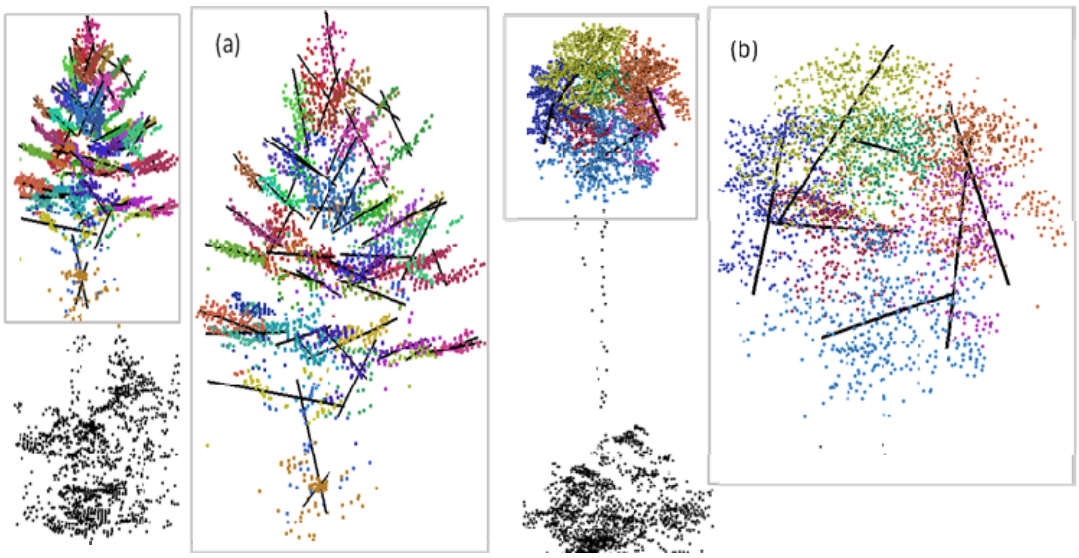


Figure 4. The result of the line features derived by PCA analysis for two of the tree samples. (a) is a White Pine tree (same tree as figure 1(a)); (b) is a White Birch (same tree as figure 1(e)). K-means clustering was only performed on the tree crown, black points in the diagram are the points that are not included in the clustering.

Random Sample Consensus (ransac)

Unlike using PCA to extract line features that require pre-processing of the data (rely on clustering results), RANSAC were able to detect line features automatically and RANSACis commonly used in situation for line extraction with the presence of outlier. For the case of tree, 3D sphere with a predefined radius is generated per each point for collecting its member points. Inliers are defined as the LiDAR points are located within a distance threshold from randomly hypothesized 3D line, while the other points located outside the threshold are considered as outliers. By setting different percentages of the ratio (inlier over total member points), we can achieve different line extraction results, figure 5 shows some of the results from four settings (10%, 15%, 20% 25% inliers in figure 5(*a*), (*b*), (*c*), (*d*) respectively), sample tree from the same tree as figure 1(*e*). We stopped at 25% inliers level because at that point, the orientation and location of line features extracted no long represent the shape of the tree.

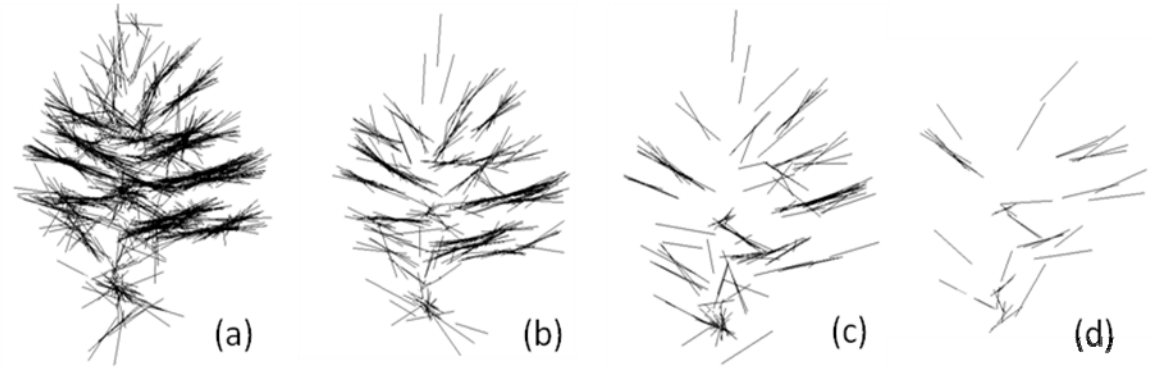


Figure 5. (a) shows line features extracted from ransac algorithm with 10% of inlier, 4026 line features were extracted; (b) has 15% of inlier, 1840 line features were extracted; (c) has 20% of inliers, 979 line features were extracted; (d) has 25% of inliers, 364 line features were extracted.

Hough transformation

The Hough transformation is a common technique in retrieving line features for image processing. Similar to ransac, Hough transformation were able to detect line features directly instead of relying on the result from clustering. However, this is done in two dimensions. First, the point cloud is projected into x-z plane; then, all the points are transformed into Hough space (Hough, 1962). Points that are collinear in the original image will intersect in Hough space, then, by accumulating those intersections, pixels that have accumulated with high values in Hough space will represent line features in the original image. As a result, lines can be delineated by selecting those peaks. Figure 6 shows the result of the line features derived by Hough transformation for the tree sample in figure 1(*a*).

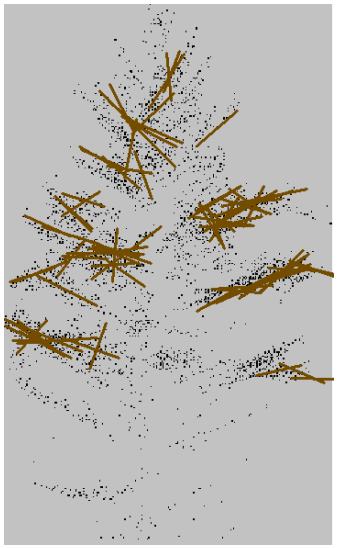


Figure 6. Line features derived by Hough transformation for tree sample in figure 1(a).

2.3 Manual digitized tree for validation

For validation purpose, the tree in figure 1(*a*) is clustered visually. Points that belong to one cluster are plot with the same colour and lines were drawn by manually by identifying the best fit end points of the clusters. Figure 7 shows the result of the visual interpretation of the tree. For this tree, there are 52 clusters formed and the smallest cluster contains 6 data points and the largest cluster contains 234 points. This tree structure will be used to validate the quality of line features derived from the above three methods.

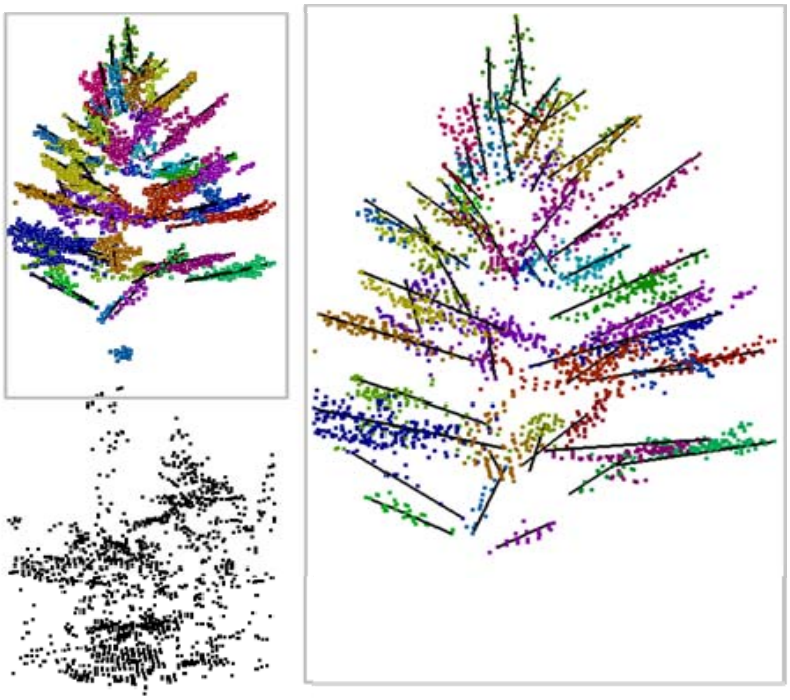


Figure 7. Figure showing the result of manual digitized tree for tree in figure 1(a).

2.4 Attribute information

The attribute information that can be retrieved from the line features extracted from the above methods includes:

1. Vertices of the bole (relate to main growth direction of the tree);
2. Vertices of the line feature, each line feature will have two end points, end points are categorized into two categories; i) end point that is closer to the trunk ii) end point that is further away from the trunk;
3. Tilt angle: angle between the line feature and the horizontal Cartesian plane (x-y), positive angle means the line is tilted up; negative angle means the line is tilted down;
4. Rotation angle: angle between the unit vector (1,0,0) and the line feature that is projected to the horizontal Cartesian plane (x-y);
5. Length of the line segment;
6. For PCA analysis, the summation of the orthogonal distances between all the points to the line is recorded, hence, the summation of distance per point is calculated (low value represent the cluster is well represented by a line);
7. For PCA analysis, the magnitude of the second principle component is calculated.

These attribute information is useful as input parameters for tree modeling. The first attempt is presented in Ko et al. (2010a), first, the line features were derived by k-means clustering (method described in section 3.2.1). Then, because we are focusing on conifer species, the tree structure can be simplified by grouping branches that are close together vertically into layers of branches to simplify the complexity of modeling (see

figure 8). With the calculated attributes such as rotation angles, tilting angles and length of each line segments, the simplified structure is then inversely represented by a series of symbols originally used by L-system (Prusinkiewicz and Lindenmayer, 1990). Figure 8(*a*) shows a schematic diagram of the procedures of the tree modeling described in Ko et al., (2010a) and figure 8(*b*) shows the result of a sample tree. We believe by performing the same procedure to the same species (similar to training a sample), we can obtain a generalization for tree has similar geometric structures (that represent same or similar species) and use those information for species classification purpose. In the discussion section, we would like to discuss some of the problems and limitations we are encountering and approaches to handle them in the future research.

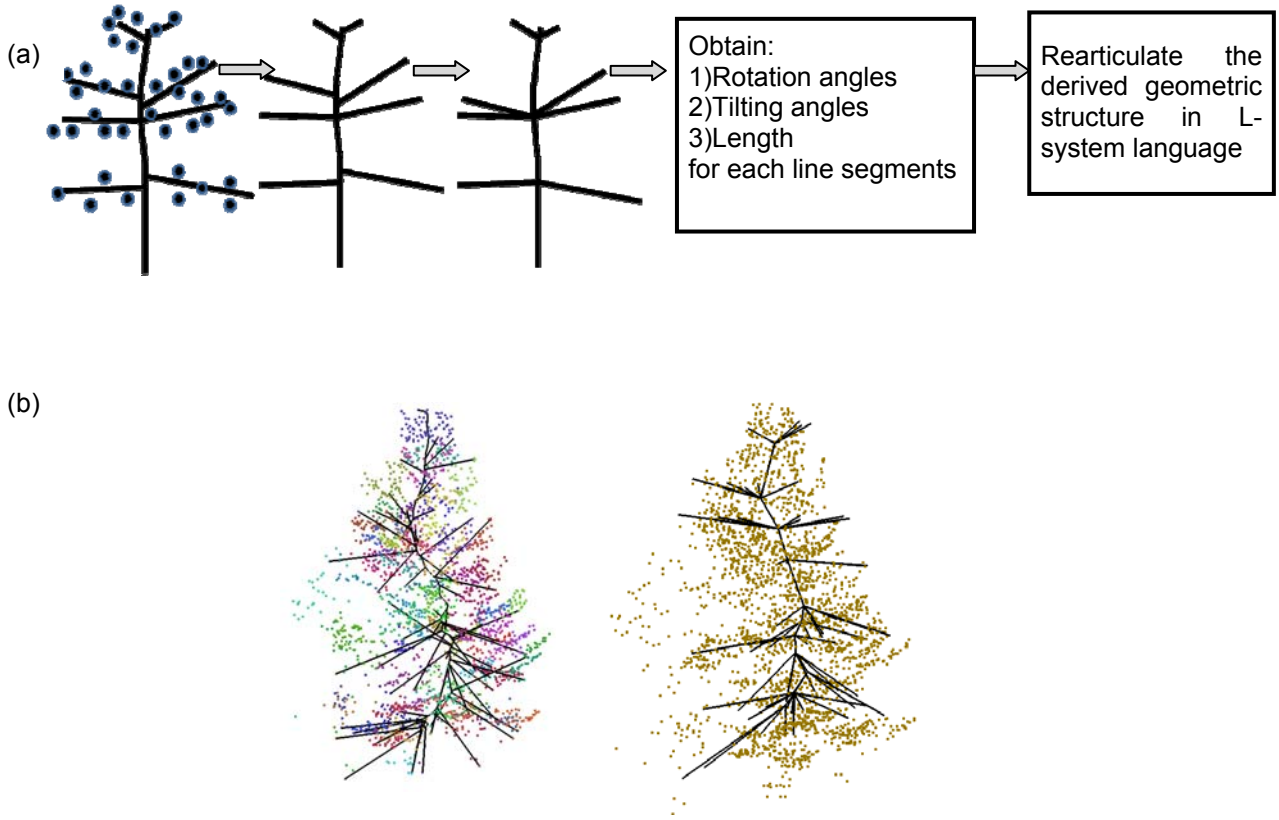


Figure 8. (a) shows the schematic diagram of the procedures of tree modeling described in Ko et al., (2010a). (b) shows one result for one of the tree samples, diagram on the left shows the clustering results (k-means, 65 clusters) and the derived line features. The diagram on the right shows the reconstructed tree branches by L-system.

3. Discussion and future research

One of the goals of this project is to develop an algorithm that can be applied to LiDAR data for tree species classification, the additional benefits include obtaining tree attributes such as branch length, branch orientation, shape of the tree crown, organization of the branch structures and height features. Although this paper displayed a framework of workflow and presents some of the preliminary results, we have not concluded which method of line features extraction is more efficient than the other. However, we believe the line features extracted from species with strong linear geometric features (high organization) such as white pine, jack pine, balsam fir and white spruce are representative. Species such as white birch, poplar, soft maple, red oak and hard maple can be identified by the low level of organization from the line features and might have to tackled with a different geometric feature such as ellipsoid.

The line features that we derived will be analyzed quantitatively, we are interested in looking at how the change of rotation and tilting angles with tree height; some species tends to have negative tilting angle towards the bottom of the tree crown and have large positive tilting angle towards the top of the tree crown;

whereas some species tends to have relatively uniform tilting angles throughout the tree crown. The change of ratio between branch length and tree crown height also inform us about the narrowness of the tree crown; if the crown is treated as a solid object, crown volume and wood volume can also be calculated from the derived structures. When the branches are grouped into layers, the derived number of layers can potentially infer the age of the tree. We believe all the information would be useful in species differentiation and research will be conducted in this direction in the future.

We would like to also addressed some of the limitations we are encountering, the algorithm we are suggesting is sensitive to shadow and occlusion due to overlap of objects and scan angle. In the future, we will address this issue by partitioning trees into sections, information derived from the high point density section will be use to supplement the occluded sections. Also the bole derivation algorithm will have to be improved, in the case of a tree that has more than one tree tops, instead of running one voxel starting from the tree top, several tree top voxels will have to be identified. The presence of understory is removed manually with a vertical threshold; a more automatic approach is presented in Ko et al. (2010b) and will be included in the future research.

Acknowledgement

RANSAC algorithm was written by Heungsik Brian Kim and we thank you his contribution to the paper. This research was funded by GeoDigital International Inc., Ontario Centre for Excellence (OCE) and NSERC.

Reference

BARILOTTI, A., CROSILLA, F. and SEPIC, F., 2009, Curvature analysis of LiDAR data for single tree species classification in alpine latitude forests. In *Laser scanning 2009, IAPRS, Vol. XXXVIII, Part 3/W8 – Paris, France*, Sep. 1-2, 2009.

BRANDTBERG, T., 2007, Classifying individual tree species under leaf-off and leaf-on conditions using airborne LIDAR. *ISPRS Journal of Photogrammetry and Remote Sensing*, **61**, pp. 325–340.

CALINSKI, T. and HARABASZ J., 1974, A dendrite method for cluster analysis. *Communications in statistics ,* **3**, pp. 1-27.

DAVIES, D.L. and BOULDIN, D.W., 1979, A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **1** (4), pp. 224-227.

GOBAKKEN, T. and NAESSET, E., 2004, Estimation of diameter and basal area distributions in coniferous forest by means of airborne laser scanner data. *Scandinavian Journal of Forest Research* , **19** (6), pp. 529-542.

HILKER, T., WULDER, M.A. and COOPS, N.C., 2008, Update of forest inventory data with lidar and high spatial resolution satellite imagery. *Canadian Journal of Remote Sensing,* **34** (1), pp. 5-12.

HILL, R.A. and THOMSON, A.G., 2005, Mapping woodland species composition and structure using airborne spectral and LiDAR data. *International Journal of Remote Sensing*, **26**, pp. 3763–3779.

HOLMGREN, J. and PERSSON, *Å*., 2004, Identifying species of individual trees using airborne laser scanner. *Remote Sensing of Environment,* **90**, pp. 415–423.

HOLMGREN, J., PERSSON, A. and SÖDERMAN, U., 2008, Species identification of individual trees by combining high resolution LIDAR data with multi-spectral images. *International Journal of Remote Sensing*, **29**, pp. 1537-1552.

HOUGH, P. V. C.,1962, Method and means for recognizing complex patterns. *U. S. Patent 3,069,654.*

KATO, A., MOSKAL, L., M., SCHIESS, P., SWANSON, M., E., CALHOUN, D. and STUETZLE, W., 2009, Capturing tree crown formation through implicit surface reconstruction using airborne lidar data. *Remote Sensing of Environment*, **113**, pp. 1148-1162.

KIM, S., MCGAUGHEY, R.J., ANDERSEN H. and SCHREUDER G., 2009, Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. *Remote Sensing of Environment*, **113,** pp. 1575–1586

KO, C., REMMEL, T.K. and SOHN, G., 2010b, A statistical partitioning of vegetated airborne laser scanning data towards understory and canopy separation. *Canadian Association of Geographers Annual General Meeting, June 1-5, Regina, Saskatchewan, Canada.*

KO, C., SOHN, G. and REMMEL, T.K., 2010a, Derivation of 3D geometric models of single tree using high density airborne LiDAR data. *AOLS (Association of Ontario Land Surveyors) 118th Annual General Meeting, Feb 17-19, 2010, Huntsville, Ontario*.

KO, C., SOHN, G. and REMMEL, T. K., 2009, Classification for Deciduous and Coniferous Trees Using Airborne LiDAR and Internal Structure. In *Proceedings of Silvilaser 2009, Oct 14-16, 2009, Texas A&M, USA*.

KORPELA, I., DAHLIN, B., SCHÄFER, H., BRUUN, E., HAAPANIEMI, F., HONKASALO, J., ILVESNIEMI, S., KUUTTI, V., LINKOSALMI, M., MUSTONEN, J., SALO, M., SUOMI, O. and VIRTANEN, H., 2007, Forest inventory using LiDAR and aerial images for 3D treetop positioning, species recognition, height and crown width estimation. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVI-3/W52, 12–14 September 2007, Espoo, Finland (The Netherlands: GITC bv)*, pp. 227–233.

KORPELA, I., TOKOLA, T., ØRKA, H.O. and KOSKINEN, M., 2009, Small footprint discrete-return LiDAR in tree species recognition. *Proceedings of the ISPRS Hannover Workshop 2009, Hannover, Germany*, June 2-5, 2009.

KOUKOULAS, S. and BLACKBURN, G. A., 2005, Mapping individual tree location, height and species in broadleaved deciduous forest using airborne LIDAR and multi-spectral remotely sensed data. *International Journal of Remote Sensing,* **26** (3), pp. 431-455.

KRZANOWSKI, W. and LAI, Y., 1985, A criterion for determining the number of groups in a data set using sum of squares clustering. *Biometrics*, **44**, pp. 23–34.

ØRKA, H. O., NÆSSET, E. and BOLLANDSÅS, O. M., 2009, Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. *Remote Sensing of Environment*, **113**, pp. 1163-1174.

PERSSON, Å., HOLMGREN, J., SÖDERMAN, U. and OLSSON, H., 2004, Tree species classification of individual trees in Sweden by combining high resolution laser data with high resolution near- infrared digital images. *The International Archives of the Photogrammetry*, *Remote Sensing and Spatial Information Sciences*, **36**, pp. 204–207.

PRUSINKIEWICZ, P. and LINDENMAYER, A., 1990. *The Algorithmic Beauty of Plants*. Springer Verlag, New York. 228p.

ROUSSEEUW, P.J., 1987, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, **20**, pp. 53-65.

SURATNO, A., SEIELSTAD, C. and QUEEN, L., 2009, Tree species identification in mixed coniferous forest using airborne laser scanning. *ISPRS Journal of Photogrammetry and Remote Sensing*, **64**(6), pp. 683-693.

Assessing laser pulse penetration in spruce canopies – combining field measured branch properties with discrete return airborne laser scanning data

Vegard Lien†, Johannes Breidenbach†, Terje Gobakken†, Erik Næsset†*

†Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences, 1430 Ås, Norway

Abstract

In forest inventory, airborne laser scanning (ALS) has long been used to estimate various biophysical properties, such as stand volume and mean tree heights. It is also possible to use ALS to detect and measure single trees allowing the derivation of even more detailed information. However, even if the laser measurements are accurate one specific challenge occurs with tree canopies – they are not solid objects, and the laser echoes are not necessarily returned from the surface of the canopy. Thus, naïve algorithms might not reproduce the spatial extent of each tree crown correctly. This can result in biased estimates if crown attributes are used in subsequent statistical analyses.

The aim for this study was to assess the laser penetration into tree crowns by comparing field measured branch structures with high density ALS data. The material consisted of measured branches from 15 trees, where branch diameter and branch base height had been measured. Precisely measured tree positions allowed the calculation of geographical coordinates for each measured branch tip. A crown surface assumed to be the true crown extent was constructed using a convex hull algorithm..

The ALS data were collected with a discrete return scanning system, Optech 3100EA, with known pulse level orientation parameters, such as aircraft position and mirror angle. Each pulse was intersected with the crown surface model and the distance from the intersection to the pulse return was calculated. This distance was then used to analyze patterns of the pulse penetration, given scan angle and angle of incidence when intersecting the canopy. The results revealed no strong influence of scan angle or angle of incidence on the penetration.

Keywords: Forest inventory, wood quality, ALS, lidar, canopy structure

1. Introduction

In forest inventory, airborne laser scanning (ALS) has long been used to estimate various biophysical properties, such as stand volume or mean tree heights (Naesset 1997a, b). It is also possible to use ALS to detect and measure single trees allowing the derivation of even more detailed tree parameters, such as diameter, biomass or crown base height (Persson et al. 2002, Popescu 2007, Maltamo et al. 2009). While most studies focused on dominant trees, methods for deriving unbiased estimates of forest attributes using ALS crown segments in an inventory context were described by Lindberg et al. (2010) and Breidenbach et al. (2010). However, even if the laser measurements are accurate one specific challenge occurs with tree canopies – they are not solid objects, and the laser echoes are not necessarily returned from the surface of the canopy.

For example, Persson et al. (2002) found difference of 1.3 meter between field measured tree height and the highest laser pulse return. The penetration of laser pulses into the crown has been affirmed by several other studies (e.g. Gaveau and Hill 2003, Andersen et al. 2006, Chasmer et al. 2006). Thus, naïve algorithms might not reproduce the spatial extent of each tree crown correctly. This can result in biased estimates if crown attributes are used in subsequent statistical analyses such as estimates of stem branchiness. While some studies have utilized simulation to assess pulse properties (Ni-Meister et al. 2001), not much work has been done to empirically investigate ALS-pulse properties on single trees. Thus the aim of this study was to

quantify the extent of ALS pulse penetration into spruce canopies and further to assess the effects of scanning angle and angle of incidence on the penetration using empirical data.

2. Material and methods

Study area and field measurements

Field measurements were performed in two different stands in the municipality of Aurskog-Høland, southeastern Norway (59° 44' N, 11° 29' E, 150-200 m a.s.l.). The area is dominated by coniferous trees, with spruce (*Picea abies* Karst.) and pine (*Pinus sylvestris* L.) as the two most common species.

A total of 15 spruce trees were randomly selected and measured in April-May 2007. All trees had a breast height diameter larger than 150 mm. Precise tree positions were collected using differential GNSS, (Topcon Legacy E+) and a total station (Sokkia SET5F). In addition to tree positions, the cardinal directions and breast height were marked on the stem surface while the trees were still standing. Each tree was then felled, and diameter measurements were taken for every meter along the stem. Between two subsequent diameter readings, the two living branches with the largest diameter were selected for further measurements. If any of the branches had snapped, or was not freely suspended, e.g. pushed towards the ground due to the felling, the third or fourth largest branch was selected. In addition to the branch diameter, the height of the branch base was registered. Then each branch was given a gentle lift to bring the branch into the same position as it would have had if the tree still had been standing upright. On the suspended branches, the direct distance from the outermost position of branch to the stem surface was measured, together with the vertical and horizontal angle of the branch. In addition to the extensive measurements on the larger branches, a subset of measurements was taken on the smaller branches in the same knot-whorls as the large branches. These measurements included branch diameter, height and horizontal angle. It was also registered if the branch was living or dead, i.e. having no foliage.

Laser data

Airborne laser data were collected 12 June 2006 by Blom Geomatics using an Optech 3100EA laser scanner mounted in a Piper 31 Navajo aircraft. Technical parameters are given in the laser scanning system gives up to four discrete returns, but for this study only first returns were used (single and first of many). In addition to pulse return positions, the aircraft position was also known for each and every pulse.

Table 14, Flight parameters

Instrument	Optech ALTM3100EA
Flight altitude	800 m above ground
Speed	75 ms^{-1}
Pulse rep. freq.	100 kHz
Mirror freq.	70 Hz
Scanning angle	± 5 degrees

Computations

Determining the outer crown surface from field measurements

To calculate branch length relative to stem centre it was necessary to construct taper models for each tree, as the measured branch length was relative to stem surface. This model was a simple linear regression model where the measured diameters along the stem were used to predict a diameter at an arbitrary height. To be able to model the crown extent it was necessary to use also the branches for which only diameter measurements were available. However, since they were lacking vertical branch angle and length measurements, two additional models had to be constructed. One described the vertical branch angle variations and another branch length. To model the branch angle, a mixed-effects model was chosen. In this model, relative branch height, i.e. the height above ground for the branch base, and the total tree height were used as predictors together with a random effect for each tree. As with the vertical angle model, a mixed-effects model was used to determine branch lengths. This model included branch diameter and relative height as predictors and a random effect on tree level.

Using angle and length measurements for the extensively measured branches and by applying length and angle models for the branches having only diameter recordings, the branch-tip positions relative to the stem

centre were calculated for all measured branches. Then by adding geographical coordinates for the stem centre, global coordinates were derived for each branch-tip. A convex hull algorithm was then used to create a surface model based on the branches of each tree. The convex hull is built as triangular faces which are connected to each other, sharing corners at the different branch-tip positions. As the name convex implies, the overall shape must be convex, thus not allowing local minima or sinks.

Modelling the intersection of ALS returns with the canopy surface

Having a convex hull model for each tree, the next step was to intersect the convex hull model with the laser pulses. As position of origin and returned pulse position was known, each pulse could then be described as a line intersecting the planes of the crown model. The result of a line intersecting a plane will be a single point unless the line and plane are parallel to each other. The last step was to check if the resulting point was inside one of the triangular faces. Because of the convex shape, there will always be two points, one inbound and one outbound through the convex hull. Only the inbound intersection was chosen, because this is where the penetration into the crown begins. For each intersecting pulse, two different metrics were derived.

1. Penetration depth; Euclidian distance from intersecting point to return position reported from the scanning system. A negative sign indicates that the reported return lies below the crown surface (position *R* minus position *I* in Figure 16).
2. Angle of incidence between pulse and surface model at the point of intersection (angle α_1 in Figure 1).

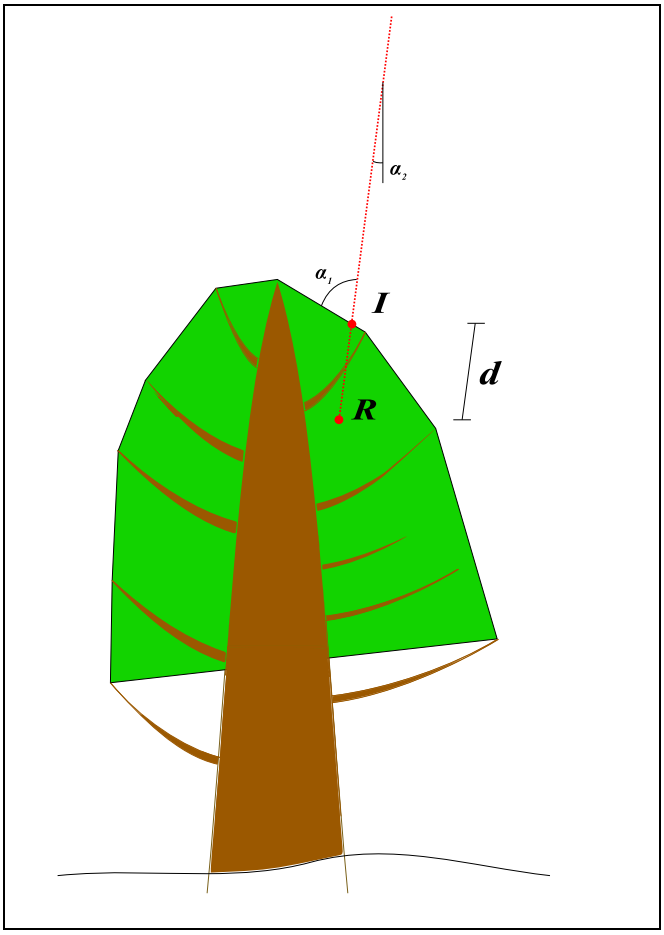


Figure 16, Schematic drawing of one single tree with surface model and an incoming laser pulse

Analysis

To reduce the influence of adjacent trees, only pulse returns inside the crown surface models of the selected trees were used. A classification of ground hits was done by selecting pulses with a penetration depth equal to the distance from the intersection and down to the ground. General statistics, such as mean values and standard deviation, were then derived to describe the penetration into the crown. Distance to stem centre, scan angle and angle of incidence were plotted against penetration depth to examine any influence.

3. Results and discussion

A general description of the trees used in this study is given in Table 15. Also the number of intersecting laser pulses used for calculations is displayed. As the table shows, the mean number of intersecting pulses per tree is 86, where the larger trees have more intersecting pulses than smaller trees due to larger crown extent.

Table 15, General description of measured trees

TreeID	Diameter (mm)	Total height (m)	Height of lowest green branch (m)	Number of intersecting pulses
11010	277	20.37	4.19	157
11130	346	24.97	2.80	145
11140	184	18.79	8.02	30
12155	367	23.97	3.09	95
12201	258	19.36	5.50	80
13034	183	19.80	4.10	31
13088	407	26.29	3.74	180
13162	286	22.77	6.27	84
21076	297	26.01	12.94	105
21083	245	23.30	8.86	54
22055	240	20.60	4.44	81
22065	326	26.74	10.72	115
22166	161	18.26	6.28	33
23067	260	23.60	6.82	74
23152	214	21.12	6.13	33
Average		22.40	6.26	86

The distribution of differences between intersected height and ALS-height is given in Figure 17 as box plots for each tree. Because only returns from within the crown surface model were selected, all values are smaller than 0. The figure shows less variation for some trees than other, but for all intersecting pulses, the mean value for pulse penetration is -3.11 meter, with a standard deviation of 2.71 meter.

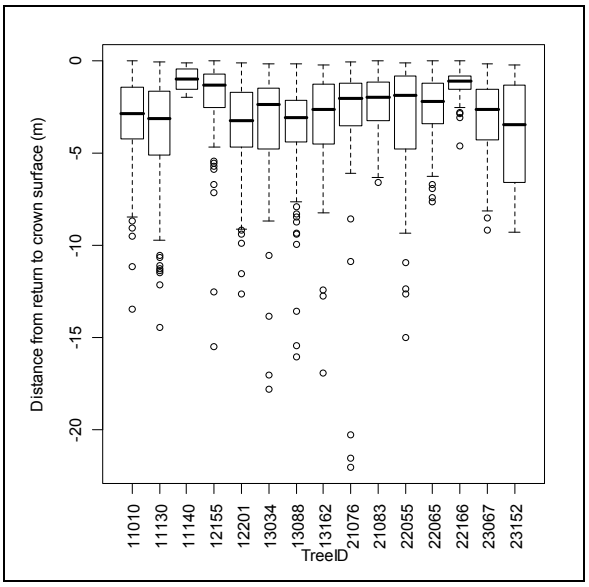


Figure 17, Penetration of returns into crown surface model.

The first factor analyzed that could influence penetration depth is the horizontal distance from the stem centre to the intersected pulse. As Figure 18 shows, the penetration depth increases with distance from the stem centre, as does the variation. An error in xy-position, both for the pulse and for crown surface near the edge of the crown, will give larger variation of penetration than near the top of the tree crown. This is supported by the large amount of ground hits at greater distances from the stem. Taking that into consideration, the observed influence of distance from stem centre is mainly an effect of increasing influence of errors in both the ALS coordinates and the crown surface model.

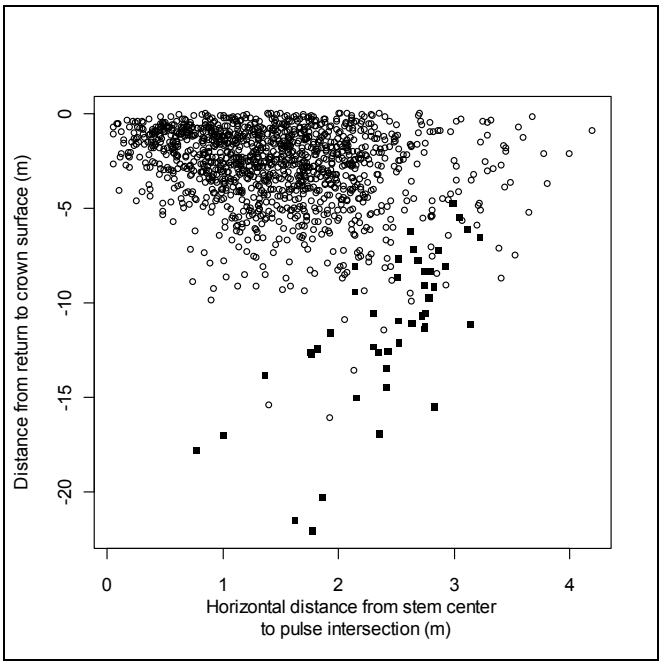


Figure 18, Distribution of pulses with respect to distance from stem centre. Ground classified pulses shown as filled boxes.

To analyze the penetration depth of returns close to the tree top, pulses less than 0.5 m from the stem centre were selected. All trees combined, the average penetration depth close to the stem centre was -1.70 m with a standard deviation of 1.05 m (Figure 19). This penetration depth is of the same magnitude as the reported

underestimation of total tree height versus maximum ALS-height in other studies (Persson et al. 2002, Chasmer et.al 2006).

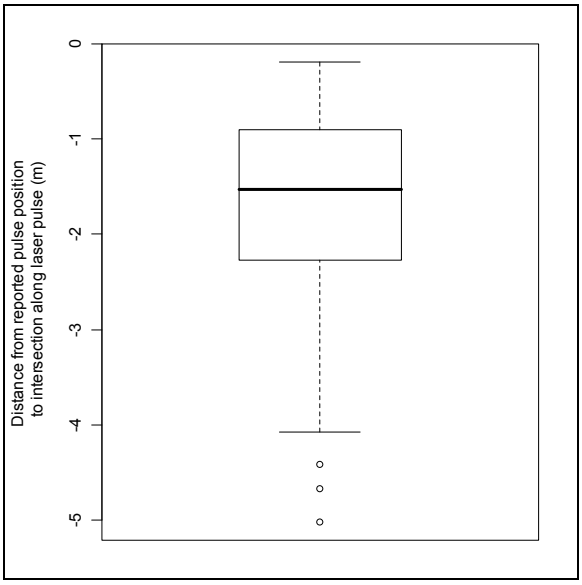
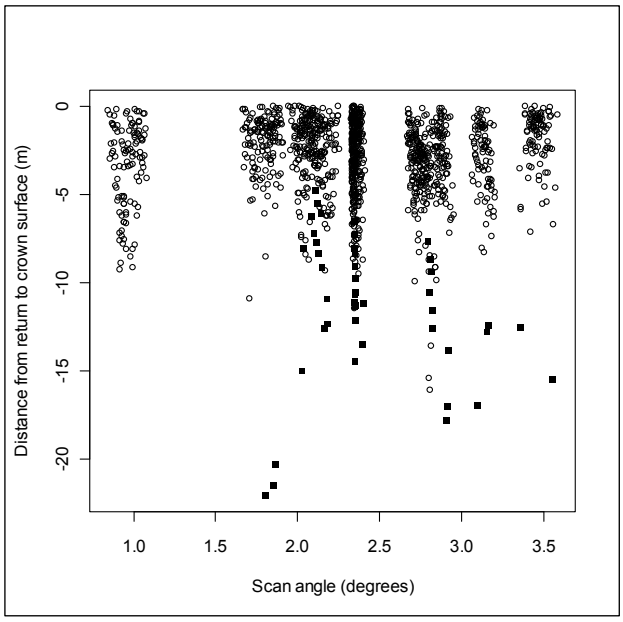


Figure 19, Differences in reported height versus intersected height for pulses <0.5 meter from stem centre.

Scan angle and angle of incidence appear to have only a minor influence on penetration depth (Figure 20, Scan angle versus pulse penetration Figure 21, Angle of incidence versus pulse penetration

and Figure 21). Scan angle show no effect on penetration depth, while angle of incidence may show a slight influence on penetration depth. This effect is probably an effect of increasing variation, just like the effect observed with distance from stem centre (Figure 18). The reason for this is that low angle of incidence only occur at the edge of the tree crown, where the variance and amount of ground hits are higher.



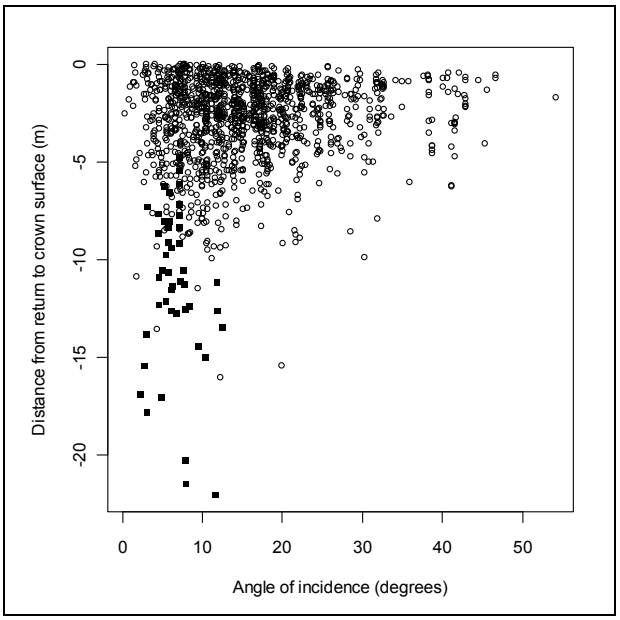


Figure 20, Scan angle versus pulse penetration Figure 21, Angle of incidence versus pulse penetration

As the maximum scan angle is very narrow for this particular flight, (only 5 degrees half angle), it is difficult to spot effects of scan angle. At higher scan angles, an effect could be noticeable. Also a larger dataset with

more measured trees would probably be necessary for further studies. It is also important to notice that different ALS-sensors or flight altitude could give significant effects on pulse penetration (Naesset 2009).

Our study helped us in understanding the interaction of ALS pulses and the canopy based on empirical data better. We conclude that neither distance from stem centre, scan angle nor angle of incidence had a strong effect on penetration depth in the data material that was analyzed.

References

ANDERSEN, H-E., REUTEBUCH, S. E., McGAUGHEY, R. J., 2006, A rigorous assessment of tree height measurements obtained using airborne lidar and conventional field methods. Can. J. Remote Sensing, Vol. 32, No. 5, pp. 355–366.

BREIDENBACH, J., NÆSSET, E., LIEN, V., GOBAKKEN, T. and SOLBERG, S., 2010, Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. Remote Sensing of Environment, 114, 911–924.

CHASMER, L., HOPKINSON, C., TREITZ, P., 2006, Investigating laser pulse penetration througha conifer canopy by integrating airborne and terrestrial lidar. Can. J. Remote Sensing, Vol. 32, No. 2, pp. 116–125.

GAVEAU, D. L. A., HILL, R. A., 2003, Quantifying canopy height underestimation by laser pulse penetration in small-footprint airborne laser scanning data. Can. J. Remote Sensing, Vol. 29, No. 5, pp. 650–657.

LINDBERG, E., HOLMGREN, J., OLOFSSON, K., WALLERMAN, J., OLSSON, H., 2010, Estimation of tree lists from airborne laser scanning by combining single-tree and area-based methods, International Journal of Remote Sensing, Vol. 31, pp. 1175-1192

MALTAMO, M., PEUHKURINEN, J., MALINEN, J., VAUHKONEN, J., PACKALÉN, P., TOKOLA, T., 2009, Predicting Tree Attributes and Quality Characteristics of Scots Pine Using Airborne Laser Scanning Data, Silva Fennica, Vol. 43, pp. 507-521

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, Vol. 39, pp. 1943-1958

NÆSSET, E., 1997a, Determination of mean tree height of forest stands using airborne laser scanner data, ISPRS Journal of Photogrammetry and Remote Sensing, 52, 49-56.

NÆSSET, E., 1997b, Estimating Timber Volume of Forest Stands Using Airborne Laser Scanner Data, Remote Sensing of Environment, 61, pp. 246-253

NÆSSET, E., 2009,Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data, Remote Sensing of Environment, 113, pp. 148–159.

PERSSON,Å., HOLMGREN, J., SÖDERMAN, U., 2002, Detecting and Measuring Individual Trees Using an Airborne Laser Scanner, Photogrammetric Engineering & Remote Sensing, 68, pp. 925–932.

POPESCU, S.C., 2007, Estimating biomass of individual pine trees using airborne lidar, Biomass and Bioenergy, Vol. 31, pp. 646-655

Comparative testing of single-tree detection algorithms

JARI VAUHKONEN †, LIVIU ENE‡, SANDEEP GUPTA§, JOHANNES HEINZEL§, JOHAN HOLMGREN¶, JUHO PITKÄNEN|, SVEIN SOLBERG¤, YUNSHENG WANG§, HOLGER WEINACKER§, KNUT MARIUS HAUGLIN‡, VEGARD LIEN‡, PETTERI PACKALÉN†, TERJE GOBAKKEN‡, BARBARA KOCH§, ERIK NÆSSET‡, TIMO TOKOLA† and MATTI MALTAMO†

jari.vauhkonen@uef.fi

†School of Forest Sciences, University of Eastern Finland, Joensuu, Finland

‡Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences, Ås, Norway

§Department of Remote Sensing and Landscape Information Systems, Albert-Ludwig-University, Freiburg, Germany

¶Department of Forest Resource Management, Swedish University of Agricultural Sciences, Umeå, Sweden

|Finnish Forest Research Institute, Joensuu, Finland

¤Norwegian Forest and Landscape Institute, Ås, Norway

Abstract

The aim of this study was to validate and compare single-tree detection algorithms under different forest conditions. Field data and corresponding airborne laser scanning (ALS) data were acquired from boreal forests in Norway and Sweden, coniferous and broadleaved forests in Germany, and pulpwood plantations in Brazil. The data represented a variety of forest types from pure Eucalyptus stands with known ages and planting densities to conifer-dominated Scandinavian forests and more complex deciduous canopies in Central Europe. ALS data were acquired using different sensors with pulse densities varying between the data sets.

Field data in varying extent were associated with each ALS data set for training purposes. Treetop positions were extracted using altogether six different algorithms developed in Finland, Germany, Norway and Sweden, and the accuracy of tree detection and height estimation was assessed. Furthermore, the weaknesses and strengths of the methods under different forest conditions were analyzed.

The results showed that forest structure and density strongly affected the performance of all algorithms. The differences in performance between methods were more pronounced for tree detection than for height estimation. The algorithms showed a slightly better performance in the conditions for which they were developed, while some could be adapted by different parameterization according to training with local data. The results of this study may help guiding the choice of method under different conditions and may be of great value for future refinement of the single-tree detection algorithms.

1. Introduction

Since around 1995, a large number of scientific studies have indicated a huge potential of airborne laser scanning (ALS) data to provide highly accurate estimates of important biophysical parameters of forests, like tree height, timber volume (e.g. Næsset 1997, Magnussen and Boudewyn 1998, Means *et al*. 2000), and other parameters related to the structure and distribution of the tree layer (e.g. Zimble *et al*. 2003, Maltamo *et al*. 2005). The results have been most encouraging for coniferous forests. Overviews are provided by Lim *et al*. (2003), Næsset *et al*. (2004) and Hyyppä *et al*. (2008). Currently, there are two main approaches for using ALS to characterize forest resources: (1) an area-based approach typically providing data at stand level and (2) a single-tree approach where individual trees are the basic unit of the assessment.

The single-tree approach requires dense laser scanning (at least 5–10 laser pulses m^{-2}) and provides direct measurements of the position, height, and canopy shape of the dominating trees. However, not all trees can be automatically detected and the success rate is mainly dependent on scanning density and forest structure. In Scandinavia and Central Europe, rates of correctly detected trees higher than 70% have been found (Hyyppä *et al*. 2001, Persson *et al*. 2002, Koch *et al*. 2006, Solberg *et al*. 2006). In deciduous forests, the success rate has generally been lower. In Germany, success rates of around 50–60% have been reported (Koch *et al*. 2006), which coincide well with similar studies in North America.

Different methods for single-tree detection have been developed by different research groups. Most studies so far have been carried out at limited test sites and are restricted to few tree species. Kaartinen and Hyyppä (2008) compared twelve different algorithms on two test sites located in southern Finland. According to them, the extraction method was the main factor influencing on the accuracy, the percentage of detected trees varying from 25 to 90% for the various methods. However, their conclusions were based on results from test sites with fairly simple forest conditions and thus overall suitable for single-tree detection. Forest conditions is an important factor affecting the performance of the algorithms.

The purpose of this study was to test and compare the accuracy of single-tree detection algorithms under different forest conditions. Compared to previous international comparisons (Kaartinen and Hyyppä 2008), there are fewer algorithms involved, but the variation in forest conditions is much greater.

2. Material and methods

2.1 An overview

The comparison performed in this study covered altogether six tree detection algorithms. ALS data and a small sample of field data were delivered to the operators of these, requesting to extract treetop positions of each test site for validation. Each ALS data set included at least XY-coordinates and height above ground values, which were calculated by data deliverers. The extent of the associated field data varied, with a purpose to adapt the algorithms to the local conditions by training. The accuracy of tree detection and height estimation was assessed for each algorithm and test site.

2.2 Test sites and data

The test sites considered in this study were located in Brazil, Germany, Norway and Sweden. The Brazilian test site was an even-aged pulpwood plantation growing Eucalyptus. The German data consisted of 11 plots pre-stratified to coniferous and deciduous. The dominant conifer species was pine, while the deciduous plots were composed of oak, beech and birch trees. The species composition in the Scandinavian data consisted of pine and spruce and to a lesser degree of deciduous trees, mainly birch. Of these two data sets, the Swedish data included notably larger trees than the Norwegian data. The main properties of the field data are given in table 1.

Table 1. Main attributes of the field data. Average and (standard deviation) of number of stems (N), basal area (BA), basal area-weighted mean tree diameter (D).

Site	N / ha	BA, m^2/ha	D, cm
Brazil	824 (41)	22.0 (7.0)	18.9 (3.1)
Germany, deciduous	733 (331)	25.2 (4.9)	35.8 (16.5)
Germany, coniferous	664 (347)	28.8 (3.3)	30.6 (9.5)
Norway	1093 (485)	24.3 (8.4)	22.4 (4.8)
Sweden	610 (253)	34.6 (10.6)	32.9 (7.7)

The ALS data used in this study is described in table 2.

Table 2. Main properties of the ALS data sets.

	Brazil	Germany	Norway	Sweden
Acquisition date	August 16, 2008	August 30, 2007	June 6, 2006	April 24, 2007
Instrument	Optech ALTM 3100	TopoSys Harrier 56	Optech ALTM 3100	TopEye MKII
Density (nominal), m^{-2}	1.5	16	7.4	30
Footprint, m	0.36	0.23	0.2	0.13
Mean altitude, m	1200	450	800	130
Field of view, °	30	22.5	10	20/14[1]

[1] Elliptic scan pattern: cross flight direction / forward-backward.

2.3 Tree detection methods

The applied methods are listed below, but for detailed description, the reader should consult the primary publications. In the following text, the algorithms will be referred to by the numbers of the subsections below.

Cluster formation using modified k-means approach (Gupta et al. 2010). The algorithm used the local height maxima as seed points. A height reduction factor was employed on the 3-D ALS data and the respective seed points to minimize the bias and better grouping of similar objects.

A voxel layer single tree modelling algorithm based on raw 3D laser spots (Wang et al. 2008). The method was extended such that input areas with different shapes and variable sizes can be handled. Furthermore, the method was integrated with an algorithm for merging split tree crowns, based on the horizontal distance of the two tree tops in relation to the crown radii, the vertical height differences of the two tree tops in relation to the crown length of the higher tree, and the difference of the two crown base heights.

Adaptive segmentation based on Poisson forest stand model (Ene et al. 2010). The method employs a twofold strategy: (1) controlling the amount of canopy height model (CHM) smoothing and (2) obtaining a CHM resolution suitable for representing the smallest tree crowns. It is assumed that raw estimates of the stem density or tree spacing can be obtained, and that the trees are randomly located within plots. The CHMs were first interpolated to various resolutions accommodated to each training dataset. Two runs of the pit-filling algorithm (Ben-Arie et al. 2009) were applied to each CHM, followed by low-pass filtering using a binomial kernel with size proportional to the expected nearest neighbor distance between trees.

Local maxima with residual height adjustment (Solberg et al. 2006). The method used here was modified by varying the amount of smoothing, i.e. first by varying the window size for initial filtering of nearby ALS echoes, and second by varying the number of Gaussian 3×3 filter runs. These smoothing factors were decided separately for each data set after testing with the test trees.

Segmentation based on geometric tree crown models (Holmgren et al. 2010; Holmgren and Wallerman 2006). This method is based on Gaussian smoothing of a correlation surface, which is calculated with respect to geometric models and CHM height values. The local maxima and the segments are defined by the correlation surface and the CHM, and for each segment, the geometric models are used to decide if the segment should be merged with a neighbouring segment.

Adaptive filtering based on CHM height values (Pitkänen et al. 2004). The CHMs used with this method were interpolated to a grid of 0.5 m by taking the maximum first return height value within a radius of 0.5 m. The empty cells were filled by taking the average from a 3×3 window, and the interpolation was repeated successively until every cell had a height value. The algorithm requires the determination of the kernel widths (sigma, σ) and the height classes for which the sigma are applied. These were selected separately for each data set based on training data.

2.4 Evaluation criteria and performance measures

The performance of the tree detection was evaluated by comparing

9 Estimated number of trees vs. the ground truth
10 RMSE and bias of stem number and mean height (plot level)
11 Number of field trees linked to treetop candidates (omission errors) and candidates not linked (commission)
12 RMSE and bias of tree height

The RMSE and bias were calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(\hat{x}_i - x_i\right)^2}{n}} \text{ , and} \quad (1)$$

$$bias = \frac{\sum_{i=1}^{n}\left(\hat{x}_i - x_i\right)}{n} , \quad (2)$$

where n is the number of observations, and x_i and $\hat{x}_i$ are the reference and estimated attributes, respectively, for the tree or plot i. In the case of individual tree height, only measured observations were used in the calculation of RMSE and bias.

The Brazilian field data set included no positioned trees, so that the evaluation was carried out at the plot level only. In order to evaluate two latter criteria, the treetop candidates detected from ALS data were linked to field trees according to the following procedure:

Estimate maximum crown width using diameter at breast height and tree height (*H*) in species-specific models by Pretzsch *et al*. (2002) and Nagel *et al*. (2002). In case *H* is not measured, predict it using local height curves.

Initially accept those treetop candidates that are located within the maximum crown width. From multiple candidates, link the one with the smallest Euclidean 3-D distance with the top of the field tree.

Remove gross linking errors by comparing the height values of the linked pairs. Remove those links with a residual of more than 2 × standard error of a simple regression model $H_{candidate} \sim f(H_{field})$ (see Breidenbach *et al*. 2010).

3. Results

Table 3. Quantitative success of tree detection by different methods. N – number of trees, H – mean tree height.

Site	Algorithm	$\hat{N}_{TOTAL}$	$RMSE_N$	$BIAS_N$	$\hat{\overline{H}}$	$RMSE_H$	$BIAS_H$
Brazil	3	792	252	-75	27.2	1.9	-0.5
N_{TOTAL} = 872	4	520	363	-331	28.4	2.5	0.7
$\overline{H}$ = 27.6	5	787	270	-80	26.5	3.7	0.2
	6	733	297	-131	26.7	1.9	-0.2
Germany	1	483	375	-214	19.1	2.5	2.1
N_{TOTAL} = 695	2	583	302	-113	19.5	4.6	3.3
$\overline{H}$ = 16.0	3	341	430	-358	19.5	4.9	3.6
	4	700	247	5	17.9	2.1	1.2
	5	557	276	-139	18.0	2.5	1.8
	6	454	300	-262	18.5	4.1	2.2
Norway	1	1978	628	-545	15.1	2.7	2.2
N_{TOTAL} = 4002	2	2052	622	-488	15.3	2.8	2.1
$\overline{H}$ = 12.8	3	2114	576	-483	15.0	2.7	1.9
	4	2274	605	-475	14.4	2.8	1.5
	5	2726	460	-316	14.2	2.2	1.2
	6	1810	685	-568	15.2	3.2	2.4
Sweden	1	3222	564	-275	22.6	3.7	1.8
N_{TOTAL} = 4150	2	2701	674	-358	22.7	5.3	4.1
$\overline{H}$ = 20.1	3	2858	637	-383	23.6	4.8	3.5
	4	2972	614	-349	22.2	3.9	2.0
	5	3560	415	-175	22.6	3.8	2.5
	6	2841	668	-388	23.5	4.4	3.2

On average, the number of trees found by the algorithms corresponded to 65% of the number of trees measured in the field. The average tree detection rate varied between the test sites, being 81% for Brazil, 70% for Germany, 54% for Norway, and 73% for Sweden. The average percentages of treetop candidates linked to field trees were 48% (Germany), 42% (Norway) and 60% (Sweden). Thus, the algorithms generally resulted in more commission errors in the German data.

Algorithm #5 usually found a number of trees most close to field observations, especially in the Scandinavian data (table 3). Algorithm #3 was slightly more accurate in Brazil and algorithms #4 and #2 in the German test site. Algorithms #3 and #6 usually resulted in lowest number of trees. The performance of algorithm #4

varied most between the data sets. Algorithms #2 and #1 performed well in a single data set, but at an average level in others.

Algorithms #5 and #4 resulted in most treetop candidates linked to field trees (table 4). Compared to these, algorithms #6 and #3 produced more omission errors, but generally resulted in the lowest number of commission errors. Considering a summation of these error components, the most accurate algorithms were #6 and #5 in the German data; algorithm #3 followed by #4–6 in the Norwegian data; and algorithms #6, #4, #5, in this order, in the Swedish data. Algorithms #1 and #2 performed poorer according to these metrics.

There were less differences between the algorithms in tree height estimation. Except for the Brazilian data set, the mean tree height was generally overestimated (table 3). The performance of all algorithms was almost equal in estimating height for the detected individual trees (table 4). In most cases, algorithm #4 resulted in lowest accuracies.

Table 4. Qualitative success of tree detection by different methods. N – number of trees, H – tree height, Om – omission, and Com – commission.

Site	Algorithm	N_{est}	N_{linked}	*Om., %*	*Com., %*	$RMSE_H$	$BIAS_H$
Germany	1	483	324	53.4	32.9	2.0	-0.8
	2	583	314	54.8	46.1	1.7	-1.1
	3	341	274	60.6	19.6	1.7	-0.9
	4	700	424	39.0	39.4	2.1	-1.3
	5	557	400	42.4	28.2	1.7	-0.8
	6	454	356	48.8	21.6	1.5	-0.8
Norway	1	1978	1523	61.9	23.0	1.9	-0.2
	2	2052	1437	64.1	30.0	1.8	-0.4
	3	2114	1763	55.9	16.6	1.7	-0.2
	4	2274	1809	54.8	20.4	1.9	-0.4
	5	2726	2034	49.2	25.4	1.9	-0.2
	6	1810	1558	61.1	13.9	1.7	-0.2
Sweden	1	3222	2269	45.3	29.6	1.3	-0.1
	2	2701	2086	49.7	22.8	1.2	-0.1
	3	2858	2447	41.0	14.4	1.2	0.0
	4	2972	2667	35.7	10.3	1.4	-0.8
	5	3560	2880	30.6	19.1	1.3	0.0
	6	2841	2646	36.2	6.9	1.3	0.0

Figures 1 and 2 demonstrate the effects of forest structure on the performance of the algorithms, indicating a considerable difference between conifer test sites and German deciduous forests (figures 1(*B*) and 2(*B*)). None of the algorithms differed significantly from their respective overall performance (table 4), when the

detection of trees of different sizes was considered (figure 1). Treetop positions extracted by algorithms #1 and #2 most often deviated from other the methods, while these were extracted mainly from the same locations (figure 2).

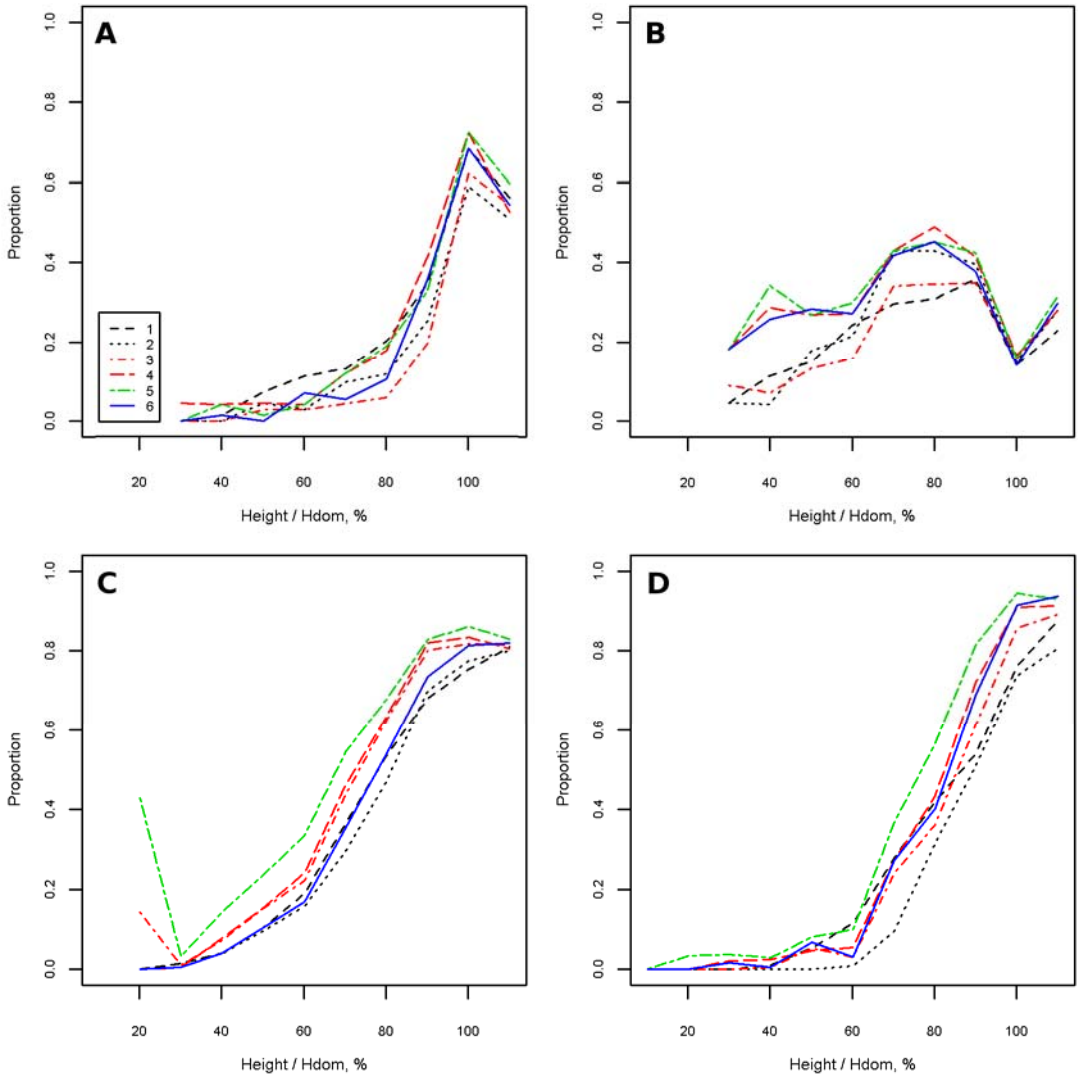


Figure 1. Proportion of treetop candidates linked to field trees as a function of tree height relative to plot-level dominant height. A – Germany (coniferous plots), B – Germany (deciduous plots), C – Norway, and D – Sweden.

4. Discussion

The results of this study showed in general less differences between the tested algorithms than the earlier comparison performed by Kaartinen and Hyyppä (2008). The principles of the algorithms tested here are obviously more similar than in the study of Kaartinen and Hyyppä (2008), which included a wider range of different algorithms. The results of the present comparison showed more remarkable differences between the methods in tree detection rates rather than in height estimation accuracies.

There were two basic differences between the algorithms tested in this study. First, algorithms #1 and #2 use raw point data, whereas the others require a CHM interpolation step. Second, algorithms differ in the ability to employ training data in setting proper parameters for the interpretation. Here local training data were available, and due to the training ability, the performance of some algorithms could be clearly improved. For example, by considering tree spacing, algorithm #3 was able to adapt well to the Brazilian test site.

Algorithms #1 and #2 were developed in Central Europe, and they resulted in a better performance in those conditions compared to Scandinavia. On the other hand, Scandinavian algorithms #3–6 performed fairly well even in the non-Scandinavian test sites.

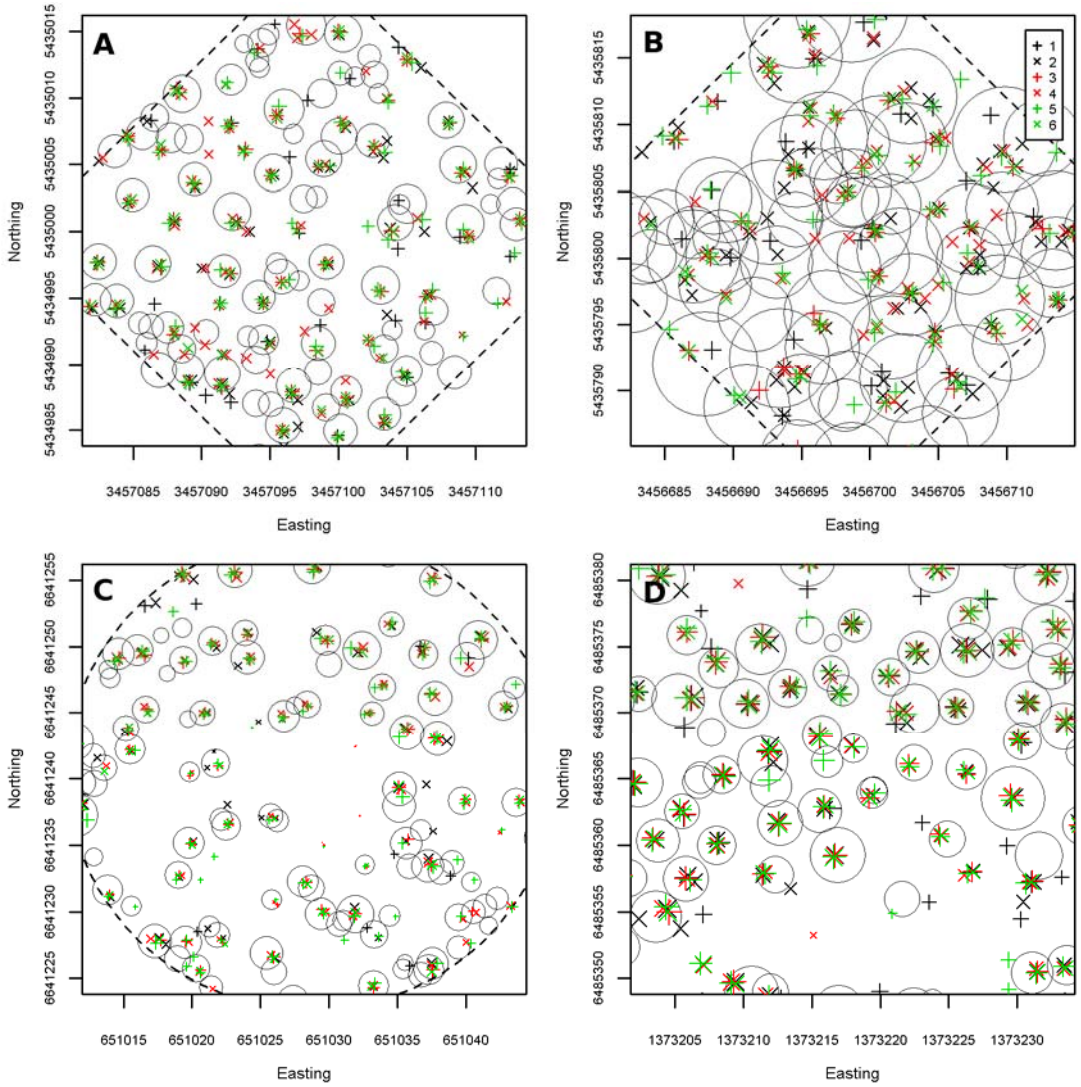


Figure 2. Interpretation results of example plots illustrated on top of a map of trees, where circles represent the predicted crown diameter based on field data. The plot borders, if visible, are drawn with a dashed line. A – Germany (coniferous), B – Germany (deciduous), C – Norway, and D – Sweden.

Both quantitative and qualitative success criteria were considered when evaluating the algorithms. In some cases, these criteria were found contradictory. Consider, for example, the plot-level mean tree height: the fewer trees an algorithm detected, the more inaccurate the height estimate, since the detection result is focused on the largest trees on the plot. This is still logical, however, as increasing commission often improves the accuracy of plot-level mean height since commission trees are normally not as large as properly detected trees. On the other hand, algorithms #3 and #6, which found lowest number of trees, were found to benefit from the qualitative criteria.

Linking treetops to field trees clearly succeeded in Scandinavian data sets, while it was known that the lateral difference between tree tops and stem foot positions could be up to 7m in the German data. Thus the commission error rates between different data sets cannot directly be compared. There are alternative measures for evaluating the success of single-tree detection (see, e.g., Solberg *et al*. 2006), and developing appropriate metrics remains a topic for future work. Furthermore, comparing algorithms with respect to using the delineated point clouds for predicting attributes such as tree species and stem volume should be considered in the future.

Finally, it was clearly observed that forest structure affected the performance of all algorithms. These effects will be more carefully analyzed in our future work.

Acknowledgements

This study is a contribution to the WoodWisdom-Net project “New Technologies to Optimize the Wood Information Basis for Forest Industries – Developing an Integrated Resource Information System (WW-IRIS)”.

References

BEN-ARIE, J.R., HAY, G.J., POWERS, R.P., CASTILLA, G. and St-ONGE, B., 2009, Development of a pit filling algorithm for LiDAR canopy height models. *Computers and Geosciences*, **35**, 1940–1949.

BREIDENBACH, J., NÆSSET, E., LIEN, V., GOBAKKEN, T. and SOLBERG, S., 2010, Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. *Remote Sensing of Environment*, **114**, 911–924.

ENE, L., NÆSSET, E. and GOBAKKEN, T., 2010, Single tree detection in heterogeneous boreal forests using airborne laser scanning and area based stem number estimates (submitted).

GUPTA, S., KOCH, B. and WEINACKER, H. 2010, Tree species detection using full waveform lidar data in a complex forest. In Wagner W., Székely, B. (eds.): Technical Commission VII Symposium – 100 Years of ISPRS, Vienna, Austria, July 5-7, 2010. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences (IAPRS), Vol. XXXVIII, Part 7B, pp. 249–254.

HOLMGREN, J. and WALLERMAN, J., 2006, Estimation of tree size distribution by combining vertical and horizontal distribution of LIDAR measurements with extraction of individual trees. In: T. Koukal and W. Schneider (Eds.). Workshop on 3D Remote Sensing in Forestry, 14-15 February 2006, University of Natural Resources and Applied Life Science, Vienna, Austria, pp. 168–173.

HOLMGREN, J., BARTH, A., LARSSON, H. and OLSSON, H., 2010, Prediction of stem attributes by combining airborne laser scanning and measurements from harvesting machinery. SilviLaser 2010.

HYYPPÄ, J., KELLE, O., LEHIKOINEN, M. and INKINEN, M., 2001, A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. *IEEE Transactions on Geoscience and Remote Sensing*, **39**, 969–975

HYYPPÄ, J., HYYPPÄ, H., LECKIE, D., GOUGEON, F., YU, X. and MALTAMO, M. 2008. Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing*, **29**, 1339–1366.

KAARTINEN, H. and HYYPPÄ, J., 2008, EuroSDR / ISPRS Commission II project: ”Tree Extraction” – final report. Official publication n:o 53. EuroSDR, Frankfurt am Main, Germany. 60 p.

KOCH, B., HEYDER, U. & WEINACKER, H., 2006, Detection of individual tree crowns in airborne lidar data. *Photogrammetric Engineering & Remote Sensing*, **72**, 357–363.

LIM, K., TREITZ, P., WULDER, M., ST-ONGE, B. and FLOOD, M., 2003, LiDAR remote sensing of forest structure. *Progress in Physical Geography*, **27**, 88–106.

MAGNUSSEN, S. and BOUDEWYN, P., 1998, Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. *Canadian Journal of Forest Research*, **28**, 1016–1031.

MALTAMO, M., PACKALÉN, P., YU, X., EERIKÄINEN, K., HYYPPÄ, J. and PITKÄNEN, J., 2005, Identifying and quantifying structural characteristics of heterogenous boreal forests using laser scanner data. *Forest Ecology and Management*, **216**, 41–50.

MEANS, J. E., ACKER, S. A., BRANDON, J. F., RENSLOW, M., EMERSON, L. and HENDRIX, C. J., 2000, Predicting forest stand characteristics with airborne scanning lidar. *Photogrammetric Engineering & Remote Sensing*, **66**, pp. 1367–1371.

NAGEL, J., ALBERT, M. & SCHMIDT, M., 2002, Das waldbauliche Prognose- und Entscheidungsmodell BWINPro 6.1. *Forst und Holz*, **57**, 486–493.

NÆSSET, E., 1997, Determination of mean tree height of forest stands using airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **52**, 49-56.

NÆSSET, E., GOBAKKEN, T., HOLMGREN, J., HYYPPÄ, H., HYYPPÄ, J., MALTAMO, M., NILSSON, M., OLSSON, H., PERSSON, Å. and SÖDERMAN, U., 2004, Laser scanning of forest resources: the Nordic experience. *Scandinavian Journal of Forest Research*, **19**, 482–499.

PERSSON, Å., HOLMGREN, J. and SÖDERMAN, U., 2002, Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering & Remote Sensing*, **68**, 925–932.

PITKÄNEN, J., MALTAMO, M., HYYPPÄ, J. and YU, X., 2004, Adaptive methods for individual tree detection on airborne laser based canopy height model. In: Theis, M., Koch, B., Spiecker, H. and Weinacker, H. (Eds.). Proceedings of ISPRS working group VIII/2: "Laser-scanners for forest and landscape assessment". University of Freiburg, Germany, pp. 187–191.

PRETZSCH, H., BIPER, P. and DURSKÝ, J., 2002, The single-tree based simulator SILVA: construction, application and evaluation. *Forest Ecology and Management*, **162**, 3–21.

SOLBERG, S., NÆSSET, E. and BOLLANDSÅS, O. M., 2006, Single tree segmentation using airborne laser scanner data in a heterogeneous spruce forest. *Photogrammetric Engineering & Remote Sensing*, **72**, 1369–1378.

WANG, Y., WEINACKER, H., KOCH, B., STERENCZAK, K., 2008, Lidar point cloud based fully automatic 3D single tree modelling in forest and evalutions of the procedure. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **XXXVII**, Part B6b, Beijing 2008, 45–51.

ZIMBLE, D.A., EVANS, D.L., CARLSON, G.C., PARKER, R.C., GRADO, S.C. and GERARD, P.D., 2003, Characterizing vertical forest structure using small-footprint airborne LiDAR. *Remote Sensing of Environment*, **87**, 171–182

Determination of stand first-thinning maturity using airborne laser scanning

Mikko Vastaranta[1], Markus Holopainen[1], Xiaowei Yu[2], Juha Hyyppä[2], Hannu Hyyppä[3], Risto Viitala[4]

mikko.vastaranta@helsinki.fi

1University of Helsinki, Department of Forest Sciences, Finland, mikko.vastaranta@helsinki.fi, markus.holopainen@helsinki.fi.

2 Finnish Geodetic Institute, yu.xiaowei@fgi.fi, juha.hyyppa@fgi.fi.

3 Research Institute of Modelling and Measuring for the Built Environment, Helsinki University of Technology, Finland, hannu.hyyppa@tkk.fi.

4 HAMK University of Applied Sciences, Finland,

risto.viitala@hamk.fi.

Abstract

Airborne laser scanning (ALS) provides direct information from the highest parts of the tree crown and structure of the forest canopy. Thus, information provided by ALS opens up new perspectives to assess stand first thinning maturity. In practical forest inventory, ALS has been mainly utilized in retrieval of stand characteristics. Still, much of the information needed for forest management planning must be collected in the field. In the present study, statistical features extracted from the ALS data were used in prediction of stand first-thinning maturity and timing with the *k*-most similar neighbour (*k*-MSN) and random forest (RF) estimation methods. The research material consisted of 127 treewise measured circular plots located in young thinning stands from the vicinity of Evo, southern Finland. Timing of the first thinning were determined in the field by an expert and classified into four classes: 1) immediately, 2) 1–5 years, 3) 6–10 years or 4) no operations during the next ten years. The overall prediction accuracies for these four classes were 54.3% and 74.0% with the *k*-MSN and RF methods respectively. This study demonstrated practical method how to utilize ALS data to predict stand thinning maturity.

1. Introduction

Commercial thinnings are management practices in which both the silvicultural and economic aspects are taken into account. From the silvicultural stand point, the goal of thinning is to provide sufficient growing space and thus improve the vitality of future crop trees. The timing and intensity of thinning are always affected by the previous management of the stand. Harvest schedules and timing of forest operations are selected so that utility of the decision maker get maximized. The decision maker is usually assumed to maximize the net present value (NPV) of the forest area or stand. General guidelines for timing of thinning and clear-cuts are presented as recommendations for good silviculture (Recommendations of Tapio…, 2006), which are based on growth and yield studies. In practical forest management planning, stand-thinning stages are determined in the field by a forest manager and decisions are based on the above-mentioned guidelines, spatial distributions of trees and vigour of the tree crowns. When forest management calculation systems are applied, the thinning stage is determined by the stands basal area and dominant height via thinning curves as the regeneration stage is determined by the age or mean diameter (Recommendations of Tapio…, 2006). When timing of harvests is suggested by computation, these suggestions include a degree of uncertainty and are often replaced by suggestions derived at in the field. This is mainly caused by the inability of forest management calculation systems to predict stand spatial distributions of trees and vigour of the living crowns, i.e. the silvicultural aspects of the thinning are ignored. Delayed thinning or incorrect harvest decisions may result in growth and income losses for forest owners (Haara & Korhonen, 2004).

In Finland, information for forest management planning is collected in two phases. In the first phase, inventory of the stand characteristics is carried out and in the second data are completed with information collected in the field. Information collected in the field includes forest site types, biodiversity targets and stand operational need. Recently, an area-based airborne laser-scanning (ALS) inventory method (Naesset, 2002) was utilized for practical forest stand characteristic inventories in Finland. In the method, percentiles of the distribution of laser canopy heights are used in addition to aerial photographs and the estimation is based on nonparametric estimation methods (Maltamo et al., 2006).

Information provided by ALS offers new perspectives for assessing stand-thinning stages and since ALS is becoming a widely popular method for forest inventory, these tools could be adopted for practical forest management planning. As a method, ALS provides direct information from the highest parts of the tree crown and structure of the forest canopy. ALS-based estimates for crown closure, leaf-area index (LAI) or crown coverage have been utilized in many studies (Means et al., 1999; Holmgren et al., 2003; Hopkinson & Chasmer, 2009). Basically, variation in the penetration of laser pulses describes the variation in density of the crowns (Hirata et al., 2009). For practical forest inventory, statistical features are extracted from the ALS data. These include penetration as vegetation laser pulse returns *versus* total returns, height percentiles of the distribution of canopy heights and canopy cover percentile as a proportion of laser returns below a given percentage of total height. From these, features such as penetration and canopy cover percentiles are used to predict LAI and crown closure with relatively good accuracy (e.g. Jensen et al., 2008; Hopkinson & Chasmer, 2009). These features probably correlate well with the stand-thinning stage.

The objective of this study was to test the accuracy of the first-thinning maturity predictions determined from the ALS data with canopy-based statistical features. Nonparametric *k*-most similar neighbour (*k*-MSN) and random forest (RF) estimations were used. Reference suggestions for stand-thinning maturity and the timing of thinning were made in the field.

2. Materials and methods

2.1 Study area

The study area is in an app. 2000-ha managed forest area located in the vicinity of Evo, Finland (61.19° N, 25.11° E, Fig 1.). The area is dominated by coniferous tree species, namely Scots pine (*Pinus sylvestris* L.) (52%) and Norway spruce (*Picea abies* (L.) H. Karst.) (31%). As classified by stand development class, the area consists mainly of young thinning stands (26%), advanced thinning stands (40%) and mature stands (23%). The corresponding proportions of forest site type are: grass-herb sites (8%), moist sites (71%), dry sites (19%) and poor sites (2%).

2.2 Determination of forest management actions in the field

The research material consisted of 127 tree-level measured fixed-radius (10 m) plots located in young thinning stands. The plots were located with Trimble's GEOXM 2005 Global Positioning System (GPS) device (Trimble Navigation Ltd., Sunnyvale, CA, USA), and the locations were postprocessed with local base station data, resulting in an average error of app. 0.6 m.

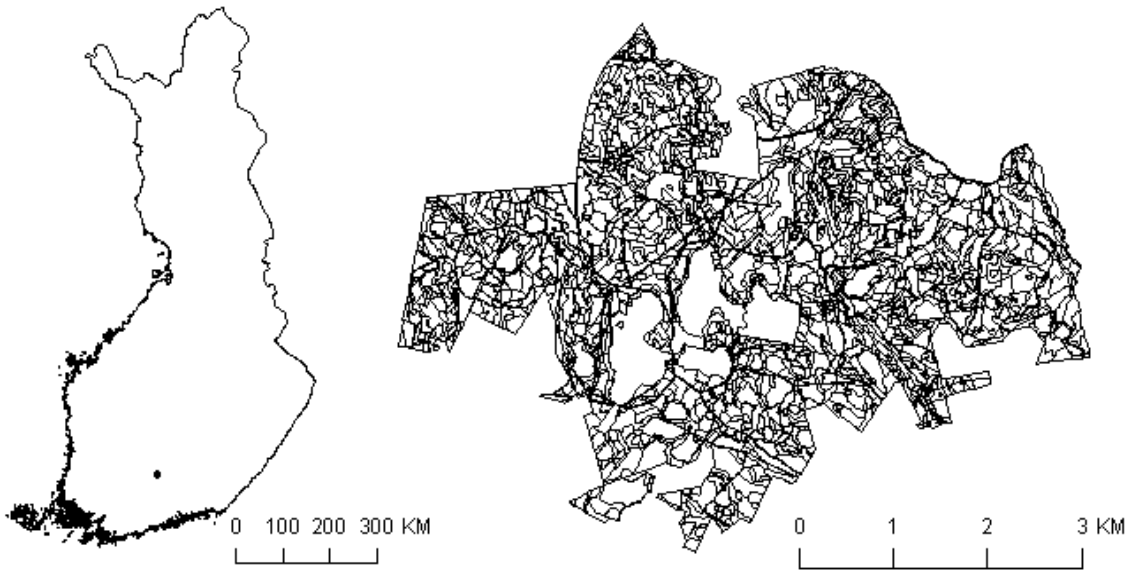


Figure 1. Location of the study area.

Proposals for forest management operations were determined in the field for the plots. Timing of the first thinning was classified as promptly, 1–5 years or 6–10 years. If there were no forest management operations determined, it was interpreted as a rest. For the next 10–year planning period, the operations determined in 82 plots as a count of the rest plots were 45. The count of immediately timed first thinnings was 35. In 23 plots, timing was from 1 to 5 years and in 24 plots from 6 to 10 years.

2.3 Acquisition and processing of ALS data

The ALS data were acquired in July 2009 with an Optech 3100 laser scanner (Optech Inc., Vaughan, Ontario, Canada). The flying altitude was 400 m. The density of the pulses returned within the field plots was approximately 10 points per m^2. The ALS data were first classified into ground and nonground points, using the standard approach of the TerraScan-based method explained in Axelsson (2000). A digital terrain model (DTM) was then developed, using classified ground points, and laser heights above the ground (normalized height or canopy height) were calculated by subtracting the ground elevation from the laser measurements. Canopy heights close to zero were considered as ground returns and those greater than 2 m as vegetation returns. The data intermediate between them were considered as returns from ground vegetation or bushes. Only vegetation returns were used for ALS feature extraction. Several features were extracted from the vegetation returns for sample plots. They included the maximum laser hit of the plot, mean, standard deviation and coefficient of variation of the canopy heights, penetration as vegetation returns *versus* total returns, height percentiles of the distribution of canopy heights from 10% to 100% at intervals of 10% and canopy cover percentile as a proportion of laser returns below a given percentage (from 10% to 100% at 10% intervals) of the total height.

2.4 *k*-MSN

Nonparametric estimation methods, such as *k*-MSN, are one alternative to predicting stand-thinning maturity if there is thinning maturity determined in the ground truth. The estimation method applied in practical forest stand characteristic inventory in Finland is *k*-MSN (Maltamo et al. 2006). In *k*-MSN, the similarity is based on canonical correlations and the Mahalanobis distance (Moeur & Stage, 1995). The benefit of the k-MSN method is that the similarity measure can be solved analytically. The *k*-MSN is described more detailed in the original paper by Moeur & Stage (1995). Automatic feature selection was carried out before *k*-MSN

imputations, using a simple genetic algorithm (GA) presented by Goldberg (1989) and implemented in the R GALGO library (Trevino & Falciani, 2006).

2.5 Random forest

The RF algorithm, proposed by Leo Breiman in 2001, is also a nonparametric estimation approach and composed of a set of regression trees that are constructed from bootstrapped training data, which in general are sets of samples taken randomly, with replacement from the original training set. A regression tree is built for each of the bootstrap sets. RF is created by averaging over the trees. A regression tree is a sequence of rules that split the feature space into partitions that have similar values for the response variable. A method based on classification and a regression tree is usually adopted to generate regression trees. At each node of a regression tree, the data are split until the leaf nodes contain fewer samples than some preselected value or the sum of the squares of the distances to the mean value of the respective group is less than a threshold. RF was described more detailed and used for estimation of forest or tree variables e.g. in Breidenbach et al. (2010), Vauhkonen et al. (2010) and Yu et al. (2010). The R statistical computing environment (R Development Core Team, 2007) and the yaImpute library (Crookston & Finley, 2007) was applied in the *k*-MSN and RF imputations.

3. Results

Table1.Classification accuracies for first thinnings (%).

	k-MSN	RF
Promptly	71.4	77.1
1-5 years	30.4	60.9
6-10 years	37.5	75.0
Rest	62.2	77.8
Overall	54.3	74.0

The *k*-MSN and RF estimations were used to predict the stand first-thinning maturity, with extracted ALS features used as predictors. Predictions classified the thinning maturity phase into four classes. The thinning phases (classes) studied were promptly, thinning during the next 1–5 years, thinning during the next 6–10 years or no thinning during the next 10 years (rest). The reliability of the predictions was examined by calculating the classification accuracy using cross validation.

The classification accuracy varied within the single classes from 60.9% to

77.8% with RF and from 30.4% to 71.4% with *k*-MSN (Table 1). The overall prediction accuracy for the four classes was 54.3% and 74.0% with the *k*-MSN and the RF methods, respectively. From thinnings that should be done immediately, *k*-MSN and RF classified correctly 71.4% and 77.1%, respectively

In Figure 2, the ALS features that describe the proportional canopy densities in the 20%, 40%, 60% and 80% height quantiles (lower values mean denser canopy) are plotted. Plot thinning probabilities were estimated with logistic regression in this case. Clear negative correlations (cor = -0.78, cor = -0.55) between the predicted probability of the thinning and the canopy density can be seen, especially in the height quantiles of 20% and 40%.

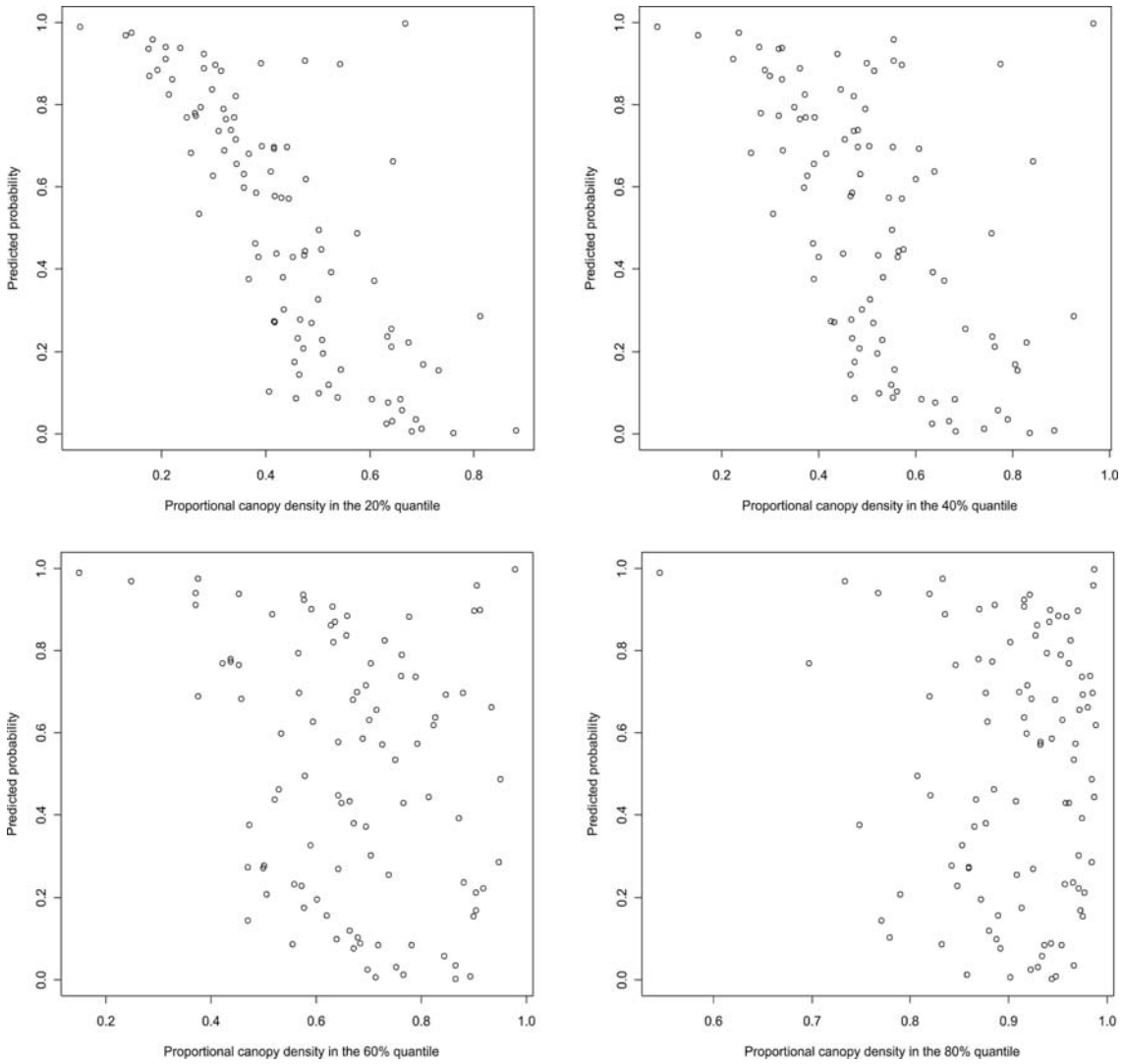


Figure 2. Proportional canopy densities plotted against predicted probability of thinning during the next 10 years.

4. Discussion

In the present study, *k*-MSN and RF estimations were used to determine stand first-thinning maturity. The overall accuracies achieved were 54.3% and 74.0% with *k*-MSN and RF, respectively. It should be noted that the "right" timing of thinning is a subjective decision. When decisions are done in the field (as in practice and also here), they always include a degree of uncertainty. From the practical point of view, it is important to detect targets that need first thinning during the next 10-year period, which is the time interval used in forest management planning in Finland. Especially, immediate thinnings should be detected.

ALS data collected for practical forest resource inventory could be used in the manner presented to enhance the quality of computationally suggested forest management operations or to locate stands that should be checked in the field. If this type of procedure is used in practice, suggestions for forest management operations should be made in the field at the same time as the ground truth data are collected for estimation of stand characteristics.

All the reference thinning suggestions were determined in the field at the plot level by an expert and the decisions were based on silvicultural recommendations. In practice, forest management operations are always carried out at the stand level. In forest inventory, plot-level field measurements are generalized to a

grid or segments and stand-level characteristics are calculated from grid cells (segments) that are within stand boundaries. Forest management operations could be generalized in the same way. The accuracy and usability of this procedure need further investigation.

The ALS features are highly correlated with the forest canopy structure and thus provide information on both the silvicultural and economic stand points, when the thinning stage is assessed. Therefore, direct models could also be developed to predict thinning maturity. Better understanding of the capabilities of ALS could provide means for removing subjective field references or oversimplified harvest decisions, as is the case with computational forest management operations.

A new practical procedure for predicting stand first-thinning maturity was presented in this study. Further research is needed to test the method at the stand level and also to develop models that could utilize single-tree-level information. Our results can be used in linking ALS-based forest inventory with practical forest management planning.

Acknowledgements

This study was made possible by financial aid from the Finnish Academy project Improving forest supply chain by means of advanced laser measurements.

References

Axelsson, P. 2000. DEM generation from laser scanner data using adaptive TIN models. International Archives of Photogrammetry and Remote Sensing, 33(B4), 110–117.

Breidenbach, J., Naesset, E., Lien, V., Gobakken, T. & Solberg, S. 2010. Prediction of species spesific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. Remote Sensing of Environment 114, 911-924.

Breiman, L., 2001. Random forests. Machine Learning 45, 5-32.

Crookston, N. & Finley, A. 2007. yaImpute: An R Package for k-NN Imputation. Journal of Statistical Software, 23(10), 1–16.

Goldberg, D.E. 1989. Genetic algorithms in search, optimization, and machine learning. Addison-Wesley Publishing Company, Reading, Massachusetts, 412 p.

Haara, A. & Korhonen, K.T. 2004. Toimenpide-ehdotusten tuottaminen laskennallisesti ajantasaistetusta kuvioaineistosta [Determination of forest management actions from updated standwise field inventory data]. Metsätieteen aikakauskirja, 2/2004, 157–173. (in Finnish).

Hirata, Y., Furuya, N., Suzuki, M. & Yamamoto, H. 2009. Airborne laser scanning in forest management: Individual tree identification and laser pulse penetration in a stand with different levels of thinning. Forest Ecology and Management, 258, 752–760.

Holmgren, J., Nilsson, M. & Olsson, H. 2003. Estimation of tree height and stem volume on plots using airborne laser scanning. Forest Sciense, 49, 419–428.

Hopkinson, C. & Chasmer, L. 2009. Testing LiDAR models of fractional cover across multiple forest ecozones. Remote Sensing of Environment, 113, 275–288.

Jensen, J., Humes, K., Vierling, L. & Hudak, A.T. 2008. Discrete return lidar based prediction of leaf area index in two conifer forests. Remote Sensing of Environment, 112, 3947–3957.

Maltamo, M., Malinen, J., Packalén, P., Suvanto, A. & Kangas, J. 2006. Nonparametric estimation of stem volume using airborne laser scanning, aerial photography, and stand-register data. Canadian Journal of Forest Research, 36, 426–436.

Means, J., Acker, S., Harding, D., Blair, J., Lefsky, M., Cohen, W., et al. 1999. Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the western cascades of Oregon. Remote Sensing of Environment, 67, 298–308.

Moeur, M. & Stage, A.R. 1995. Most similar neighbor: an improved sampling inference procedure for natural resource planning. Forest Sciense, 41, 337–359.

Næsset, E. 1997. Determination of mean tree height of forest stands using airborne laser scanner data. ISPRS Journal of Photogrammetry and Remote Sensing, 52, 49–56.

R Development Core Team. 2007. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL www.R-project.org

Recommendations of Tapio for good silviculture. 2006. Metsäkustannus 100 p. (in Finnish).

Trevino, V. & Falciani, F. 2006. GALGO: an R package for multivariate variable selection using genetic algorithms. Bioinformatics, 22(9), 1154–1156.

Vauhkonen, J., Korpela, I., Maltamo, M. & Tokola, T. 2010. Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. Remote Sensing of Environment 114(6): 1263–1276.

Yu, X.; Hyyppä, J.; Holopainen, M.; Vastaranta, M. 2010 Comparison of Area-Based and Individual Tree-Based Methods for Predicting Plot-Level Forest Attributes. Remote Sensing, 2, 1481-1495.

Applying LiDAR data and alpha shape methodology to estimating forest growth with process-based summary model

Sanna Härkönen†, Timo Tokola‡, Jari Vauhkonen‡, Petteri Packalen‡, Annikki Mäkelä§

Email: sanna.harkonen@metla.fi

† Finnish Forest Research Institute, P.O. Box 68, FI-80101 Joensuu

‡ University of Eastern Finland, Faculty of Science and Forestry, P.O. BOX 111, FI-80101 Joensuu

§ Department of Forest Sciences, University of Helsinki, Latokartanonkaari 7 (P.O. Box 27), FI-00014, University of Helsinki, Finland

Abstract.

This study introduces the use of airborne LiDAR data as input information for process-based process models for forest growth estimation. Alpha shape methodology was applied to estimate the required stand-level input variables. The tested model was a combination of a light use efficiency model and a process-based summary model for estimating the stand-level growth. LidAR based metrics were used for estimating the mean tree height, crown length and crown volumes at the plot level, which were further converted to leaf biomasses to be utilised by the process model. The modelled growths were compared with the diameter growths measured in the field. The test data contained 59 Scots pine dominated sample plots located in the Eastern Finland.

The approach proved to be promising way to apply LiDAR data to the process-based growth estimations. However, especially the biomass estimation method needs to be improved in the future.

1. Introduction

Photosynthesis-based forest growth models, so called process models, offer a relevant alternative for current forestry as they allow simulating growth response to new kind of thinning regimes or changing climate, without being dependent on the data measured in the past as the traditional empirical growth models (e.g. MELA model by Hynynen *et al*. 2002, SILVA simulator by Pretsch *et al*. 2002). To be practically utilisable, the most complicated process-based models remain still out of the range due their difficult parameterization. Instead, applicable solutions are either hybrid models combining best parts of empirical and process-based models (e.g. Valentine and Mäkelä 2005, Nuutinen *et al*. 2006) or summary models, which are simplified versions of the detailed ones e.g. FOREST-BGC by Running (1994) and 3-PG by Landsberg and Waring (1997).

The key input variables for the process-based models are connected to crown leaf biomass and crown structure. In traditional field inventory these variables have been difficult to measure and they have been derived from basic field measurements using allometric equations. However, nowadays airborne light detection and ranging (LiDAR) data enables detailed definition of crown structure of the forests, as well as producing other relevant input data for growth models (Waring *et al*. 2009). In addition to traditional stand characteristics, LiDAR data has been applied for estimating e.g. leaf area index or light interception (e.g. Lefsky *et al*. 2002, van Aardt *et al*. 2008, Lee *et al*. 2009), but rarely in the whole estimation chain by process-based models (Kotchenova *et al*. 2004).

Crown structure can be defined based on LiDAR data using different approaches, for example by fitting a certain crown shape model to the data point cloud (Holmgren and Persson 2004), by using voxel-based approach (Popescu and Zhao 2008) or by with three-dimensional (3D) alpha shapes by Edelsbrunner and Mücke (1994), applied to e.g. to tree-species recognition by Vauhkonen *et al*. (2009) with promising results. In this study the alpha shape technique is applied for producing input data for process-based growth model chain. The LiDAR data based simulations are compared both with field measured growth and with growth estimations using basic field input data with the same process-based model (Härkönen *et al*. 2010b) focusing on diameter growth.

2. Methods and material

Used growth models

The process-based model utilised here is a combination of summary model approach by Härkönen *et al*. (2010b), which defines the level of total carbon production based on LUE model (Mäkelä *et al*. 2008), and an approach by Valentine & Mäkelä (2005) which describes the dynamic growth of the tree dimensions. The detailed model descriptions can be found in the above mentioned articles and in a study by Härkönen *et al*. (manuscript). The approach is based on pipe theory and the parameters are mainly from studies by Mäkelä *et al*. (1986), Vanninen and Mäkelä (2000), Vanninen and Mäkelä (2005) and Mäkelä and Vanninen (2001). The tree growth is estimated at the stand level, based on carbon production and respiration in different components of the trees. In the LiDAR based approach the crown volume at the stand level are estimated based on the alpha shapes and used for deriving the leaf biomasses for the stands. In the field data approach the leaf biomasses are derived from the basic field measurements (mean tree height and crown length) using allometric equations by Vanninen and Mäkelä (2001).

Statistical analysis

The model performance was assessed with the root mean squared error (RMSE), the relative root mean squared error ($RMSE_{\%}$), the absolute model bias, the relative model bias ($BIAS_{\%}$) and the coefficient of determination (R^2) (table 1). The residuals were also examined.

Table 1. Statistical equations used in the analysis. y_i is the field measured mean value on a plot i, $\hat{y}_i$ is the simulated mean value on a plot i, $\bar{y}$ is the arithmetic average of the y values, and n is the total number of the plots.

Statistics	Equation
Root mean squared error	$RMSE = \sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 / n}$
Relative root mean squared error	$RMSE_{\%} = RMSE / \bar{y} \times 100$
Absolute bias	$BIAS = \sum_{i=1}^{n}(y_i - \hat{y}_i) / n$
Relative bias	$BIAS_{\%} = BIAS / \bar{y} \times 100$
Degree of determination	$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$

Field data

A total of 59 Scots pine dominated sample plots (circular with 9 m radius) from Heinävesi located in Eastern Finland from years 2004 and 2009 were used in this study (table 2). Tally tree data contained diameter and tree species. Tree height and crown base height were measured of the sample trees (crown base was measured only in 2009). The sample trees were defined as the median basal area trees per tree species strata. The sample plots were located in managed forest, 37% of the plots being of *Myrtillus* site type, 60 % of *Vaccinium* type and <1 % of *Calluna* type (classification by Cajander, 1925).

For the field input data the tree-wise crown base heights for year 2004 were estimated using the tree-species-wise ratios of crown base to tree height measured in 2009 on the same plot. The mean crown radius was estimated based on the crown length using equation by Mäkelä and Vanninen (2001).

2. **The drilled growth samples (5-years growth in radius at 1.3 m height) taken of all the sample trees (n=187) representing the dominant layer in 2009, were used for generalising the tree-wise five-year basal area growth $B_{growth5,\ tree}$ (m^2) for all the tally trees from 2004 using the following model (n=187, R^2=0.5063, p-value<0.05):**

3. $\ln(B_{growth5,\ tree}) = -6.65906 + 0.8055 \ln(D_{1.3,tree}) - 0.27386 \ln(B_{tot})$

4. $+ 0.07109\ SITE_{omt} - 0.02855\ SITE_{mt} - 0.25233\ SITE_{vt}$

5. $+ 0.08392\ SP_{spruce} - 0.14205\ SP_{deciduous}$, (1)

6. **where $D_{1.3,tree}$ (cm) is the tree diameter in 2004, B_{tot} is the stand basal area (m2 ha-1) in 2004, $SITE_i$ is dummy variable representing site type (omt=*oxalis-myrtillus* type, mt=*myrtillus* type, vt=*vaccinium* type) and SP_i is dummy variable representing tree species. The generalised growths were calibrated by multiplying the predicted growths with plot-wise ratio of the mean measured basal area growth to the mean predicted basal area growth of the dominant trees on the plot.**

Table 2. Stand level basic information of the sample plots in 2004.

	Mean height (m), basal-area weighted	Basal area (m^2 ha^{-1})	Mean diameter (cm)	Age (years)	Number of trees ha^{-1}
Minimum	6.6	11.6	10	14	550
Mean	15.2	21.5	18.45	47	1315
Maximum	24.1	46.6	26.9	79	3026

LiDAR data

The LiDAR data was gathered on night 4.8.2004 using Optech ALTM 2033 at 1500 m height with 30° angle the width of each laser strip being 800 m and the pulse density 0.7 per m^2. The footprint was 45 cm. Seven strips were scanned with overlap of 35 %, resulting to the total laser-scanned area of 20 km^2 (Packalen & Maltamo 2007). The DEM (Digital elevation model) of resolution 1 m produced based on the laser scanned data was used for defining the tree height values. The canopy height model was build using an interpolation procedure by Pitkänen *et al*. (2004).

The plot data was segmented using a water-shed segmentation algorithm. The segmented data was then given as an input to alpha shape procedure (Edelsbrunner and Mücke 1994, Vauhkonen *et al*. 2009) using the open source Computational Geometry Algorithms Library (Da and Yvinec, 2007). The alpha shape procedure provided estimates of mean height, mean crown base height and crown volume per segment. The procedure was run using only first-return points above 15 % of the tree height. The plot-wise mean height and crown base height were calculated as segments' ground area weighted averages. The plot-wise total crown volumes were determined as a sum of all segments' volumes on the plot.

Estimating of number of trees per plot was tested by using an equation fitted with Matalansalo LiDAR data by Suvanto *et al.* (2005). The mean tree volume was defined as a total crown volume divided by the estimated amount of trees per plot and used for defining the mean tree's leaf biomass. Leaf biomass and crown dimension data of Scots pine measured in Southern Finland (Vanninen & Mäkelä 2000, Vanninen & Mäkelä 2005) was used for plotting an equation between tree crown volume and leaf biomass as

$$\ln(W_{F,tree}) = 0.9537 \ln(V_{tree}) - 1.6268, \quad (2)$$

where $W_{F,tree}$ is the tree-wise leaf biomass (kg DW) and V_{tree} is the crown volume (m^3) (n=99, R^2 = 0.7907). The equation was used for converting the mean tree volume to mean leaf biomass per tree. Further, the stand leaf biomass was determined as the mean tree's leaf biomass multiplied by number of trees per hectare estimated from the LiDAR. The mean crown width was determined from

the LiDAR-based mean tree crown length and the estimated crown volume of the mean tree assuming the crowns as ellipsoids.

Weather data

The weather data for the year 2004 was available as grid of 10 x 10 km from the Finnish Meteorological Institute. The daily radiation, vapour pressure deficit and temperature data nearest to the sample plots were used in photosynthesis calculations.

3. Results and discussion

The mean diameter growth was simulated with 1) the process-based model with the stand-level field input data and 2) the process-based model with the LiDAR-based stand-level input data for five years and compared with the field observations. The diameter growth was overestimated with both of the approaches, the higher overestimation being with the field input version (table 3, figure 1). Otherwise the both approaches behaved rather similarly (figure 2), but the LiDAR input version produced less biased results than the model with field input. A tendency to overestimate growth of small trees and to underestimate the growth of bigger trees was detected with both field and LiDAR data (figures 3 & 4).

*Table 3. RMSE, bias and R^2 of plot-wise mean diameter (cm) estimations in the end of the simulation period (2009) compared with the mean diameter observed in the field.**denotes p-value<0.05*

	RMSE		**Bias**		**R^2**
Stand basal area in the end of simulation (2009)	**m^2 ha^{-1}**	**%**	**m^2 ha^{-1}**	**%**	
Process-based model, field input	1.4	6.9	-1.2**	-6.0	0.88
Process-based model, LiDAR input	1.2	5.8	-0.5**	-2.4	0.92

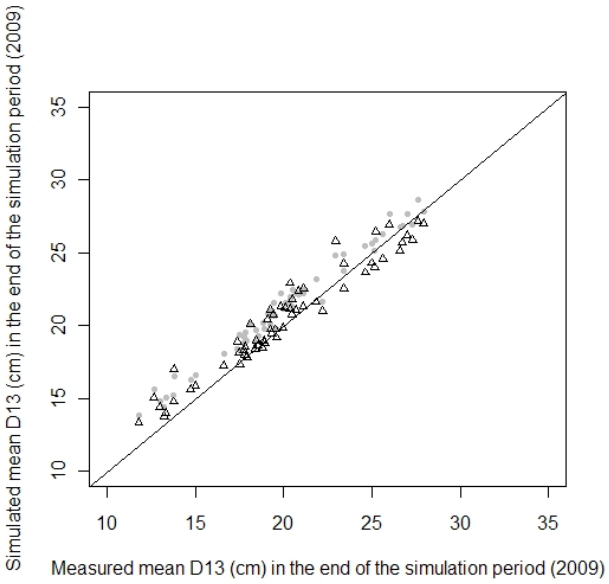


Figure 1. Measured stand mean diameter (cm) in the end of the simulation period (2009) plotted against the simulated mean diameters. Grey dots denote the simulations with field input data and triangles the simulations with LiDAR data.

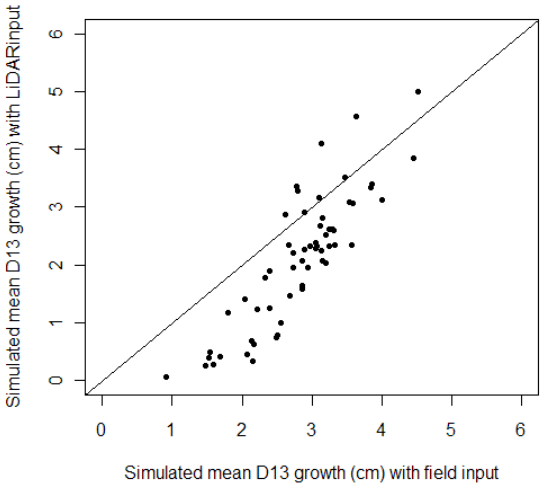


Figure 2. Stand mean diameter growth (cm) in 5 years simulated with field input plotted against the growth simulated with LiDAR input.

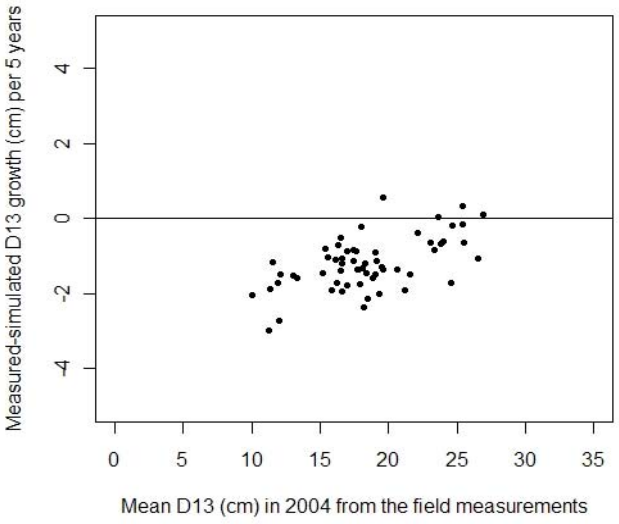


Figure 3: Residuals of 5-year diameter growth (cm) estimations with field input plotted against stand mean diameter (cm) in the field

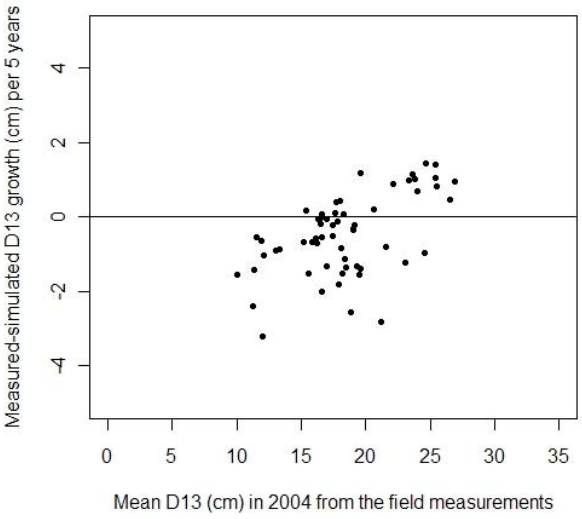


Figure 4. Residuals of 5-year diameter growth (cm) estimations with field input plotted against stand mean diameter (cm) in the field.

The results of this study indicate that even this kind of low-density LiDAR data is suitable to be used as direct input for process based growth models. However, there were still problems with the current approach, e.g. related to estimation of crown volumes and conversion of crown volumes to leaf biomasses. Firstly, the crown base height estimation using the alpha shape method did not work as well as e.g. with tree species detection (Vauhkonen *et al*. 2009) and lead to overestimated total stand crown volume, and further to overestimated leaf biomasses. This was obviously due to low density of the used LiDAR data, but as the estimates based on the field data contain also uncertainty, it is difficult to verify the real amount of estimation error. The equation fitted with tree-wise biomass data from Finland (Vanninen and Mäkelä 2001) was used for estimating the mean crown volume per tree. In order to determine the tree-level mean crown volume, number of trees per hectare were estimated using a statistical function fitted with field and LiDAR data from Matalansalo (Suvanto *et al*. 2004). With denser LiDAR data the local maxima method would obviously be a more universal approach for estimating the number of trees. As the field data used was from small set of sample plots, it is rather difficult to assess performance of the approach in larger scale.

Accuracy of the estimates seems to be at the same level that those from previous studies conducted in Finland. In a study by Härkönen *et al*. (2010a) empirical model estimations (Hynynen *et al*. 2002) were evaluated against large data set from national forest inventory plots in Finland, and in that study $RMSE_{\%}$ and $bias_{\%}$ of estimated stand mean diameter were of 7.7% and 2.1%, respectively. Also the growth estimates obtained using the most similar neighbour method with Finnish data have been at the similar level (Sironen *et al*. 2008).

Based on the findings of this study, the LiDAR-based approach produced reasonable results despite of tendency to overestimate crown volumes. The weakest part in the approach is related to estimation of the crown volume and converting it to leaf biomasses, which requires improving in the future. Further testing of the approach with larger range of site types, tree species and geographical area remains as the future tasks. Alco comparing different methods for estimating the crown structure would be necessary. In general, the approach seems promising and it could be applied directly also for example to estimating carbon fluxes, while the model already contains components for estimating gross- and net primary production as well as the whole net ecosystem production.

Acknowledgements

The study was funded by a project “Applying forest planning and environmental informatics to developing of bioenergy economics” of the University of Eastern Finland and finalized in the Finnish Forest Research Institute’s Carb-Bal project funded by the Academy of Finland. We want to thank both the field and laboratory teams at the University of Eastern Finland for measuring the data. We are grateful to Mr. Aki Suvanto and Dr. Remko Duursma for helping with the data and to Prof. Benoit St-Onge for providing valuable comments on the study. Finally, we want to thank the study area owner for good cooperation and the Finnish Meteorological Institute for providing the weather data.

References

VAN AARDT, J.A.N., WYNNE, R.H., SCRIVANI, J.A. 2008. Lidar-based Mapping of Forest Volume and Biomass by Taxonomic Group Using Structurally Homogenous Segments. *Photogrammetric Engineering and Remote Sensing*, **74**, pp. 1033–1044.

CAJANDER, A.K. 1925. Metsätyyppiteoria. *Acta Forestalia Fennica*, **1**, pp. 1-84. (Theory of forest site types, in Finnish).

EDELSBRUNNER, H, MÜCKE, E.P. 1994. Three-dimensional alpha shapes. *ACM Transactions on Graphics*, **13**, pp. 43-72.

HÄRKÖNEN, S., MÄKINEN, A., TOKOLA, T., RASINMÄKI, J., KALLIOVIRTA, J. 2010a. Evaluation of forest growth simulators with NFI permanent sample plot data from Finland. *Forest Ecology and Management,* **259**, pp. 573-582.

HÄRKÖNEN, S., PULKKINEN, M., DUURSMA, R., MÄKELÄ, A. 2010b. Estimating annual GPP, NPP and stem growth in Finland using summary models. *Forest Ecology and Management,* **259,** 524-533.

HÄRKÖNEN, S, TOKOLA, T., VAUHKONEN, J., PACKALÉN, P., MÄKELÄ, A. Linking airborne LiDAR data to climate-adaptive forest growth model. Unpublished manuscript.

HOLMGREN, J., PERSSON, Å. 2004. Identifying species of individual trees using airborne laser scanner. *Remote Sensing of Environment*. **90**, pp. 415-423.

HYNYNEN, J., OJANSUU, R., HÖKKÄ, H., SIIPILEHTO, J., SALMINEN, H., HAAPALA, P. 2002. Models for Predicting Stand Development in MELA System. *Finnish Forest Research Institute Research papers,* **835**, 116 p.

KOTCHENOVA, S.Y., SONGA, X., SHABANOVA, N.V., POTTER, C.S., KNYAZIKHIN, Y., MYNENI, R.B. 2004. Lidar remote sensing for modeling gross primary production of deciduous forests. *Remote Sensing of Environment*. **92**, pp. 158-172.

LANDSBERG, J.J., WARING, R.H. 1997. A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. *Forest Ecology and Management,* **95**, pp. 209-228.

LEE, H. SLATTON, K.C., ROTH, B.E., CROPPER, W.P. 2009. Prediction of forest canopy light interception using three-dimensional airborne LiDAR data . *International Journal of Remote Sensing,* **30**, pp. 189 – 207.

LEFSKY M.A.,COHEN W.B., PARKER G.G., HARDING, D.J. 2002. Lidar remote sensing for ecosystem studies. *BioScience*, **52**, pp. 19–30.

LIM, K., TREITZ, O., WULDER, M., ST-ONGE, B., FLOOD, M. 2003. LiDAR remote sensing of forest structure*. Progress in Physical Geography*, **27**, pp. 88-106.

MÄKELÄ, A.1986. Implications of the pipe model theory on dry matter partitioning and height growth in trees. *Journal of Theoretical Biology,* **123**, pp. 103-120.

MÄKELÄ A., VANNINEN P. 2001. Vertical structure of Scots pine crowns in different age and size classes. *Trees ,* **15**, pp. 385-392.

MÄKELÄ, A., PULKKINEN, M., KOLARI, P., LAGERGREN, F., BERBIGIER, P., LINDROTH, A., LOUSTAU, D., NIKINMAA, E., VESALA, T., HARI, P. 2008. Developing an empirical model of stand GPP with the LUE approach: analysis of eddy covariance data at five contrasting conifer sites in Europe. *Global Change Biology,* **14**, pp. 92-108.

NAESSET, E., OKLAND, T. 2002 Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. *Remote Sensing of Environment*. **79**, pp. 105-115.

NUUTINEN, T., MATALA, J., HIRVELÄ, H., HÄRKÖNEN, K., PELTOLA, H. VÄISÄNEN, H., KELLOMÄKI, S. 2006. Regionally optimized forest management under changing climate. *Climate Change,* **79**, pp. 315–333.

PACKALÉN, P., MALTAMO, M. 2007. The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. *Remote Sensing of Environment*. **109**, pp. 328-341.

PITKÄNEN, J., MALTAMO, M., HYYPPÄ, J., YU, X., 2004. Adaptive methods for individual tree detection on airborne laser based canopy height model. In: M. Theis, B. Koch, H. Spiecker, H., Weinacker (Eds.). *Proceedings of ISPRS working group VIII/2 "Laser-Scanners for Forest and Landscape Assessment"*. Freiburg, University of Freiburg, pp. 187-191.

POPESCU, S.C., ZHAO, K. 2008. A voxel-based lidar method for estimating crown base height for deciduous and pine trees. *Remote Sensing of Environment*. **112**, pp. 767-781.

PRETZSCH, H., BIBER P., DURSKY, J. 2002. The single tree-based stand simulator SILVA: construction, application and evaluation. *Forest Ecology and Management,* **162**, pp. 3–21.

RASINMÄKI, J., KALLIOVIRTA, J., MÄKINEN, A. 2009. An adaptable simulation framework for multiscale forest resource data, *Computers and Electronics in Agriculture*, **66**, pp. 76–84.

RUNNING, S.W. 1994. Testing Forest-BGC Ecosystem Process Simulations Across a Climatic Gradient in Oregon. *Ecological Applications,* **4**, pp. 238-247.

SIRONEN, S., KANGAS, A., MALTAMO, M., KALLIOVIRTA, J. 2008. Localization of growth estimates using non-parametric imputation methods, *Forest Ecology and Management,* **256**, pp. 674-684.

SUVANTO, A., MALTAMO, M., PACKALÉN, P., KANGAS, J. 2005. Kuviokohtaisten puustotunnusten ennustaminen laserkeilauksella. *Metsätieteen aikakauskirja,* **4/2005**, pp. 413–428. (Predicting stand-wise tree characteristics using laser scanning, in Finnish).

VALENTINE, H., MÄKELÄ, A. 2005. Bridging process-based and empirical approaches to modeling tree growth. *Tree Physiology*, **25**, pp. 769-779.

VANNINEN, P., MÄKELÄ A. 2000. Needle and stemwood production in Scots pine (Pinus

sylvestris L.) trees of different age, size, and competitive status. *Tree Physiology*, **20**, pp. 527-533.

VANNINEN, P., MÄKELÄ, A. 2005. Carbon budget for Scots pine trees: Effect of size, competition and site fertility on growth allocation and production. *Tree Physiology*, **25**, pp. 17-30.

WARING, R.H., COOPS, N.C., LANDSBERG, J.J. 2009. Improving predictions of forest growth using the 3-PGS model with observations made by remote sensing. *Forest Ecology and Management,* **259**, pp. 1722-1729.

VAUHKONEN, J., TOKOLA, T., PACKALÉN, P., MALTAMO, M. 2009. Identification of Scandinavian commercial species of individual trees from airborne laser scanning data using alpha shape metrics. *Forest Science,* **55**, pp. 37-47.

Integrating airborne LiDAR and Landsat ETM+ data for large area assessment of forest canopy height in Amazonia

ROSS A. HILL†, DOREEN S. BOYD‡, and CHRISTOPHER HOPKINSON§*

†School of Applied Science, Bournemouth University, Talbot Campus, Poole, Dorset, BH12 5BB, UK

‡School of Geography, University of Nottingham, University Park, Nottingham, NG7 2RD, UK

§Applied Geomatics Research Group, Nova Scotia Community College, Annapolis Valley Campus, Lawrencetown, Nova Scotia, B0S 1P0, Canada

Abstract

Quantifying structure is essential for modelling tropical forest ecosystem processes and services. This paper investigates the ability to assess canopy height of tropical rainforest from Landsat Enhanced Thematic Mapper (ETM+) data by performing a detailed assessment of the within-pixel variation of forest canopy height characteristics and the influence this has on recorded spectral response. Forest canopy height is derived from airborne small-footprint LiDAR data acquired using a Leica ALS50 II system. Forest types studied include regenerating, floodplain and *terra firme* forest. Spectral reflectance in all optical ETM+ bands was shown to be negatively correlated with the range, maximum, mean and standard deviation of canopy height per pixel. The strongest overall correlation coefficient (-0.83, $p < 0.01$) occurred between ETM+ Band 4 and maximum canopy height. It was shown that either near or middle infrared (ETM+ Bands 4 and 5) in a simple ratio or normalised index with either visible green or red (ETM+ Bands 2 and 3) also correlated strongly with canopy height (mean and maximum) and could be used to predict height across the ETM+ scene. In particular, the NDVI has been demonstrated as a potential cost-effective tool for large-area assessment of tropical rainforest canopy height appropriate for UN-REDD (Reduced Emissions from Deforestation and Forest Degradation) monitoring.

1. Introduction

There is a long history of the use of Landsat data (MSS, TM and ETM+) to assess tropical forest. Published studies cover a range of applications, but can loosely be categorised into; i) spectral separability and mapping of forest types (e.g. Singh 1987, Roy *et al.* 1991, Salovaara *et al.* 2005), ii) assessment of mature forest biomass and carbon pools (e.g. de Moraes *et al.* 1998, Foody *et al.* 2003, Meng *et al.* 2007), iii) mapping of successional forest stages and associated biomass (e.g. Sader *et al.* 1989, Boyd *et al.* 1996, Espirito-Santo *et al.* 2005); iv) assessment of biodiversity and species abundance (e.g. Foody and Cutler 2003, Gillespie *et al.* 2009, Hernandez-Stefanoni *et al.* 2009); and v) deforestation assessment and monitoring (e.g. Nelson *et al.* 1987, Lung and Schaab 2010). The over-riding themes are thus identifying where tropical forest occurs, where it is being disturbed/removed and where it is being allowed to regenerate, with a focus on accounting for biomass and carbon modelling.

In more recent years, both airborne LiDAR (e.g. Laser Vegetation Imaging Sensor data) and space-borne LiDAR (e.g. ICESat GLAS data) have been demonstrated to be of direct value for estimating structure and biomass in tropical forests (Drake *et al.* 2002, 2003, Lefsky *et al.* 2005). However, as pointed out by Hudak *et al.* (2002) whilst airborne LiDAR can provide detailed and accurate measurements over the area of acquisition, this typically is of limited spatial coverage compared with satellite data. They advocated the integrated use of LiDAR and satellite multi-spectral data (in their case Landsat ETM+) to assess forest biomass over large areas. This approach has subsequently been applied for tropical forests, typically deriving relationships for forest biophysical characteristics from LiDAR data and then extrapolating these across landscapes by forest type as assessed from spectral data (e.g. Asner 2009, Helmer *et al.* 2009).

This paper investigates the ability to assess canopy height of tropical rainforest directly from Landsat Enhanced Thematic Mapper (ETM+) data by performing a detailed and unique assessment of the within-pixel variation of forest canopy height characteristics and the influence this has on recorded spectral response. This study builds on previous work mapping primary forest types in Landsat TM imagery of Peruvian Amazonia and investigating spectral response in relation to field plot data (Hill and Foody 1994, Foody and Hill 1996, Hill 1999). These previous studies were successful but limited by the difficulties of coupling ground and satellite data which were compromised by their comparative scale. This study benefits from extensive forest structure data derived from airborne, small-footprint LiDAR data.

Study area

The field site is situated along the Rio Tambopata in the Department of Madre de Dios in south east Peru (Figure 1). In this region, the foothills of the Andes meet the western Amazon Basin creating a hotspot of biodiversity (Myers *et al.* 2000). The biological significance of the forests in this region is reflected in their status as one of the most protected regions in Amazonia (Lloyd 2004). The study area covers 19 x 7.5 km and is centred approximately on the Explorer's Inn ($12^0$50'20" S, $69^0$17'45" W), where the rivers Tambopata and La Torre meet. This area lies at the boundary between the Tambopata National Reserve and its protective Buffer Zone, which at this site is within the Native Community land of Infierno. Sustainable use of forest resources is granted within both the National Reserve and its Buffer Zone, unlike the Bahuaja Sonene National Park to the south, which together make up the ca. 1.4 million ha area previously known as the Tambopata-Candamo Reserve Zone (TCRZ).

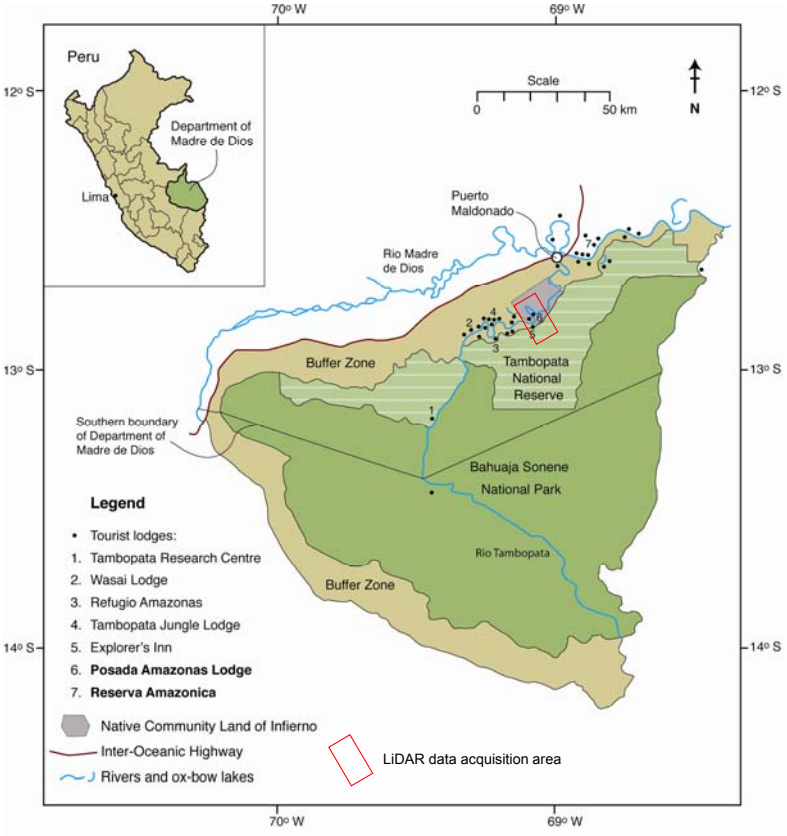


Figure 1. The study site in the Tambopata National Reserve, southeastern Peru.

The forests of this area are predominantly undisturbed and mature. According to the typology of Philips (1993) the forests in this area can be categorised as terra firme forest (on both clay and sand soils),

floodplain forest (on alluvial or silty clay soils with a flooding regime which declines in frequency with distance from the Tambopata), swamp forest (on either permanently or seasonally flooded soils) and regenerating forest on abandoned farmland next to the Tambopata. Forest structure and composition varies significantly between these categories.

Remote sensing data

Airborne LiDAR data were acquired on June 7, 2009 using a Leica ALS50 II system flown at an average altitude of 1600 m. This system has a 200 kHz repetition rate and uses a 1064 nm wavelength laser. The along/across track sampling rate was one hit per m^2, resulting in an average density of 2.1 laser hits per m^2 on the ground with flightline overlap. The nominal footprint size was 24 cm. All laser point data were geo-registered relative to a GPS base station located at Puerto Maldonado Airport, less than 40 km from the study area.

The Leica ALS50 II system records multiple (i.e. up to four) returns per laser pulse. These were separated into ground and non-ground (i.e. vegetation) hits based on the algorithm of Axelsson (2000) which is implemented in TerraScan software (Terrasolid, Finland). The ground returns were subsequently interpolated into a Digital Terrain Model (DTM) at 2m spatial resolution (Figure 2A), applying an Inverse Distance Weighting algorithm. A Digital Surface Model (DSM) was created of the maximum return height per 2 m grid cell and this was normalised using the DTM to create a Canopy Height Model (CHM) (Figure 2B).

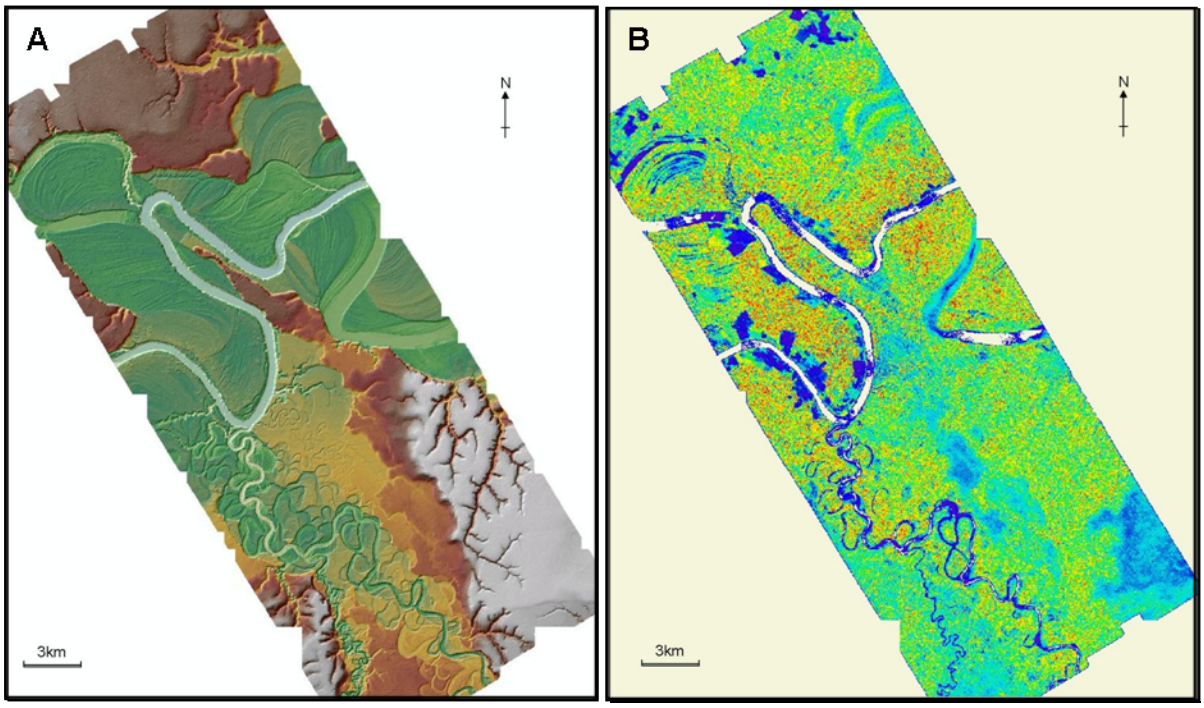


Figure 2. A) Digital Terrain Model and B) Canopy Height Model of the study site at 2 m resolution derived from airborne LiDAR data

A Landsat 7 ETM+ scene for path 002 row 069 acquired on July 28, 2009 was downloaded from the USGS Earth Explorer website (http://edcsns17.cr.usgs.gov/EarthExplorer/). This image was acquired in SLC-off mode, due to the failure in 2003 of the Scan Line Corrector which compensates for the forward motion of the Landsat 7 satellite. The result is that outside of the central 26 km of the scene, there are narrow strips of missing data. Landsat 7 SLC-off data are pre-processed by the USGS to what is termed Standard Terrain Correction (Level 1T), using ground control points (from the GLS2005 data set) and elevation data (typically SRTM). Radiometric calibration is part of the Level 1 Product Generation System (LPGS) applied to Landsat 7 ETM+ data, using detector gain parameters which are updated on a 90 day cycle. The post-calibration dynamic range for each band is scaled over an 8-bit range. No atmospheric correction is applied in this process and the no-data areas are not in-filled.

2. Remote sensing data analysis

The ETM+ image was subset down to an area incorporating the LiDAR data acquisition. In spite of the geometric correction procedure applied to the ETM+ scene in the Level 1 Product Generation System, a spatial offset remained between the two data products. The ETM+ scene was registered to the LiDAR CHM using 10 ground control points manually identified in both datasets. A second order polynomial transformation was applied and nearest neighbour resampling of the ETM+ image, to achieve data geo-registration with an average overall accuracy of 3.2 m.

A 30 x 30 m grid matching the spatial framework of the ETM+ scene was placed over the CHM. Each 30 m grid cell contained 225 individual CHM grid cells (each of 2 x 2 m) and from these the minimum, maximum, range, mean and standard deviation of canopy height was derived. The output of this was a series of 30 m resolution raster layers. The 30 m resolution datasets (i.e. ETM+ and CHM block statistics) were then stacked together to create a 12-band dataset. This enabled per-pixel comparison of canopy spectral reflectance and height characteristics. Data were extracted and compared for individual pixels of different forest types across the study site; *terra firme* (162 pixels), floodplain forest (124 pixels), and regenerating forest of at least 25 years (134 pixels). Relationships were examined, using Pearson's correlation, between forest structure (range, maximum, mean and standard deviation of canopy height) and spectral characteristics (in the optical bands only) for the forest types combined into a single data set. In addition to focussing on individual band values, simple band ratios and normalised difference band indices (such as the NDVI) were also calculated per-pixel and compared with the canopy height data.

For the spectral band or band ratio/index with the strongest correlation with mean canopy height, least squares linear regression was performed to derive a predictive equation. This was used to predict mean canopy height across the subset ETM+ scene (excluding no-data areas). The predictive accuracy was assessed based on an independent set of 'pixels' extracted from the 12-band 30 m resolution dataset. These independent data totalled 100 'pixels' for each of *terra firme* and floodplain forest, and 50 'pixels' for regenerating forest. The difference between the observed and predicted mean canopy height per 30 m cell was calculated both as an average difference and RMSE per forest type.

3. Results

From Table 1 it is clear that spectral reflectance in all ETM+ bands was negatively correlated with all four measures of forest canopy height examined (range, maximum, mean and standard deviation of canopy height). All of these correlation coefficients were significant at the 0.01 level with the exception of Band 3 with mean canopy height. It was noticeable that the strongest correlation coefficients for all four canopy structure variables occurred with Band 4, whilst the strongest correlation coefficients per band occurred with maximum canopy height for all but Band 3. The strongest overall correlation coefficient therefore occurred between Band 4 and maximum canopy height (-0.83).

For the band ratios calculated, with the exception of the 4/5 ratio, all had significant (at the 0.01 level) negative correlation coefficients with all four canopy structure measures examined, and all were most strongly correlated with maximum canopy height (Table 1). The strongest correlation coefficient with maximum and mean canopy height occurred for the Band 4/2 ratio (-0.81) and the Band 4/3 ratio (-0.78) respectively. It is worth noting that the Band 4/2 ratio had the highest correlation coefficient with three out of the four forest canopy measures.

A similar pattern was observed with the normalised difference band indices. Hence, once again with the exception of ND45, all calculated normalised difference band indices showed significant (at the 0.01 level) negative correlation coefficients with all four canopy structure measures examined, and all were most strongly correlated with maximum canopy height (Table 1). As with the band ratios, the strongest correlation coefficient for the normalised difference band indices with maximum and mean canopy height, involved Bands 4 and 2 (-0.81) and Bands 4 and 3 (-0.78) respectively.

Table 1. Correlation coefficients between selected Landsat ETM+ bands, band ratios and normalised difference band indices and forest canopy structure variables.

(Coefficients shown in italics are not significant at the 0.01 level).

Band, Ratio, Index	Canopy Height Range	Canopy Height Maximum	Canopy Height Mean	Canopy Height Standard Deviation
ETM+ 1	-0.17	-0.22	-0.19	-0.12
ETM+ 2	-0.43	-0.56	-0.48	-0.41
ETM+ 3	-0.22	-0.21	*-0.08*	-0.20
ETM+ 4	-0.60	-0.83	-0.77	-0.54
ETM+ 5	-0.60	-0.76	-0.68	-0.53
ETM+ 7	-0.52	-0.68	-0.62	-0.45
4/2	-0.58	-0.81	-0.75	-0.52
4/3	-0.54	-0.79	-0.78	-0.49
4/5	*0.10*	*-0.02*	*-0.08*	*0.06*
4/7	-0.20	-0.32	-0.32	-0.20
5/2	-0.58	-0.73	-0.66	-0.50
5/3	-0.57	-0.76	-0.71	-0.50
5/7	-0.34	-0.40	-0.33	-0.31
7/2	-0.46	-0.61	-0.56	-0.39
7/3	-0.47	-0.66	-0.65	-0.41
ND42	-0.57	-0.81	-0.77	-0.51
ND43 (NDVI)	-0.53	-0.79	-0.78	-0.49
ND45	*0.09*	*-0.03*	*-0.09*	*0.05*
ND47	-0.21	-0.33	-0.33	-0.22
ND52	-0.57	-0.73	-0.66	-0.49
ND53	-0.56	-0.76	-0.72	-0.56
ND57	-0.34	-0.40	-0.38	-0.32
ND72	-0.45	-0.61	-0.60	-0.38
ND73	-0.46	-0.66	-0.65	-0.40

Band 4 was therefore highly correlated with both the maximum and mean canopy height, and this remained the case when this band was used in either a simple ratio or a normalised difference index with either Band 2 or Band 3 (Figure 3). As the strongest relationship with mean canopy height occurred with the NDVI, this was used to predict canopy height per-pixel across the ETM+ scene based on the regression equation:

$$\text{Mean canopy height} = 72.19 - (108.3 * \text{NDVI}) \quad (1)$$

For the independent test areas, this resulted in an average error for mean canopy height prediction of 0.65 m, 3.60 m and 3.62 m respectively for *terra firme*, floodplain and regenerating forest. Thus in general, there was a positive bias (i.e. over-estimation) in mean canopy height estimation. The RMSE for mean canopy height prediction was 4.34 m, 5.59 m and 5.57 m respectively for *terra firme*, floodplain and regenerating forest.

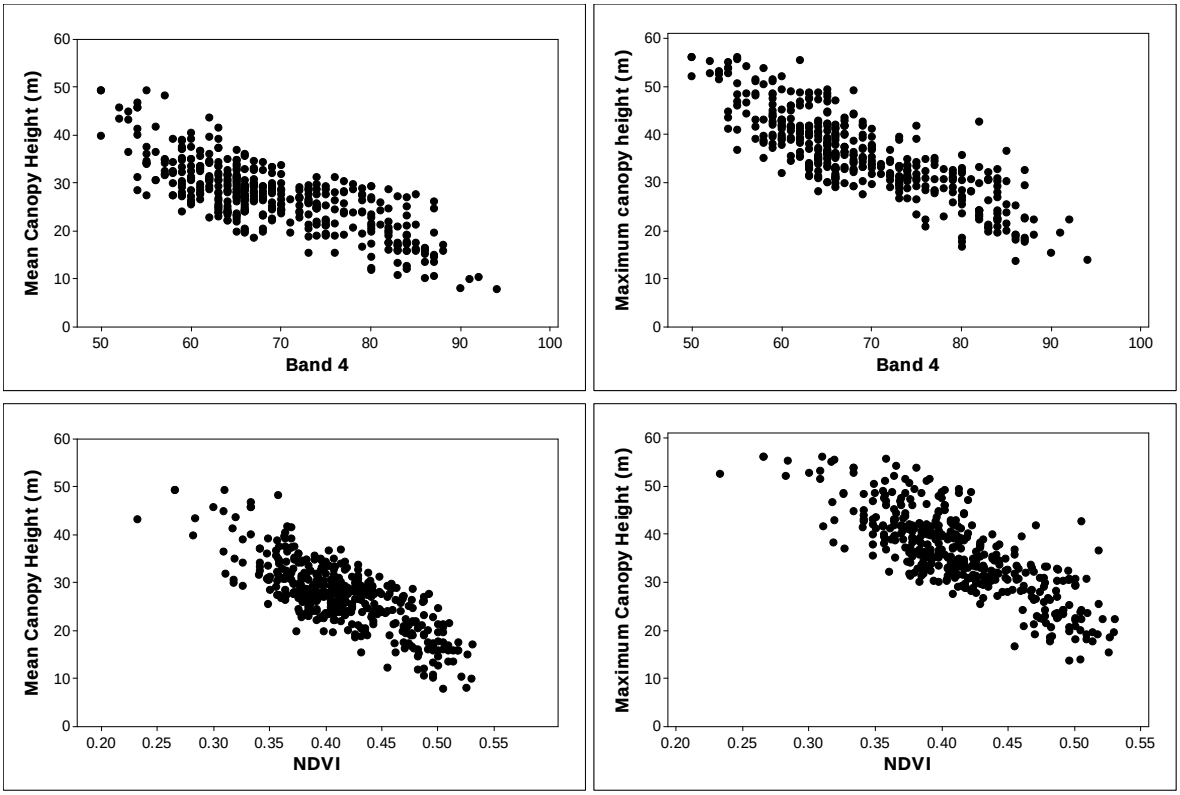


Figure 3. Plots of Band 4 and NDVI against mean and maximum canopy height for terra firme, floodplain and regenerating forest.

4. Discussion

The results of this project build on those reported elsewhere, although these have often focussed on the influence of canopy structure on spectral reflectance characteristics for regenerative forest. Thus, for regerenarative forest on abandoned pastures near Manaus in Brazil, Steininger (1996) reported that both near infrared reflectance and the NDVI (derived from Landsat TM data) increased for the first four years of forest growth and then either declined (near infrared) or stabilised (NDVI) thereafter, reaching a point that was indistinguishable from mature forest at just 13 years. They related these spectral patterns to temporal changes in canopy geometry and leaf area. Looking at seven age classes regenerative forest (between 0 and >14 years) plus primary forest in Brazilian Amazonia, Boyd *et al.* (1996) demonstrated significant negative relationships (correlation coefficient of -0.74 or stronger) for all TM bands (except Band 4) and a weak negative relationship (-0.26, significant at the 0.05 level) for the NDVI with regenerative stage. In this case, regenerative stage was taken as a surrogate measure for canopy height and structural complexity.

Lu *et al.* (2004) provided a detailed investigation of Landsat TM spectral reflectance and derived band ratios and indices against field-recorded forest structure measurements (including height) for three sites of heavily degraded and regenerating forest in Brazilian Amazonia. They found strong negative correlation coefficients (-0.66 or stronger) between reflectance in all bands and canopy height, but weaker positive relationships for the NDVI (0.38 to 0.63). They related these relationships to the effects of canopy shadow associated with increasingly complex stand structure. In particular, they identified the presence of emergent trees with high diameter at breast height (DBH) and tall canopy as having a negative impact on reflectance due to their canopy shadowing, which is strongest in the near infrared wavelength (Band 4). As a result, Lu *et al.* (2004) concluded that Landsat TM reflectance in tropical rainforests relates to tree height distribution rather than biomass or LAI differences.

Ingram *et al.* (2005) demonstrated significant negative relationships (correlation coefficient of -0.54 or stronger) for Landsat ETM+ Bands 3, 4 and 5 with basal area, and for NDVI with stem density (correlation coefficient of -0.69) for degraded and pristine littoral rainforest in southeastern Madagascar. These results are most comparable with those of the current study, and build on the results of previous work (Hill and

Foody 1994, Foody and Hill 1996) in which negative relationships were shown between TM Bands 2 and 4 (and their simple ratio) with tree stem density for different forest types in this study area. Asner and Warner (2003) used IKONOS imagery to investigate the role of sub-pixel shadow fractions in tropical forest in Brazilian Amazonia. They demonstrated that a 10% increase in shadow fraction resulted in a 3% and 10% decrease in red and near infrared response respectively. However, they reported the NDVI to be only weakly sensitive to changes in canopy shadow fraction.

In the present study, where relationships between spectral response (per band and for derived ratios and band indices) and canopy height variables were significant, the fact that all showed strongest correlations with maximum canopy height demonstrates the importance of this structural variable. For both floodplain and *terra firme* forest at this study site, the tallest canopy is directly associated with emergent trees and these areas have the greatest local variation in canopy height. Whilst the spatial distribution of emergent trees in Tambopata is typically greater than the 30 m grid size examined here, the maximum height per 30 m grid cell is nonetheless more sensitive to the presence of emergent trees and greater canopy complexity than mean canopy height. This therefore supports the conclusions of Lu *et al.* (2004) and Asner and Warner (2003).

For predicting mean canopy height in this study, the NDVI was the most strongly correlated of the band ratios and normalised difference band indices examined. This contrasts sharply with numerous published studies which have demonstrated weak or no relationship between NDVI and tropical forest structure (e.g. Sader *et al.* 1989, Foody and Curran 1994, Steininger 2000, Foody *et al.* 2003). This may be explained by the range of forest types examined in the present study. Ingram *et al.* (2005) commented on how the NDVI is sensitive at detecting what they termed ‘extremes in forest condition’ rather than structural differences between forests of ‘moderate condition’. They therefore concluded that the NDVI is valuable for detecting tropical rainforest loss or severe degradation, but not for detecting subtle variations in forest structure across a landscape. The present study, however, has demonstrated that the NDVI is sensitive to differences in forest structure when focussing on a range of forest type and status, with a mean canopy height ranging over 8-50 m.

The results of this study also contrast with numerous published studies which have highlighted the benefit of using middle infrared bands in preference to near infrared-based indices to assess tropical forest structure (e.g. Boyd *et al.* 1999; Steininger 2000; Foody *et al.* 2001; Twele *et al.* 2008). It is notable that these studies related to successional, logged-over or montane forests which are unlikely to contain tall emergent trees and thus the shadowing effect discussed above.

Whilst it has been shown possible in this study to predict canopy height across a single ETM+ image, based on the relationship identified between the NDVI and LiDAR-derived canopy height, a final note of caution is required. Foody *et al.* (2003) demonstrated that predictive relations developed at one tropical rainforest location are not readily transferable to another, whilst Miura *et al*. (2006) demonstrated that NDVI values over tropical forest are sensor dependent. Using Hyperion data they simulated NDVI from satellite sensors that could be used for monitoring tropical rainforests (MODIS, AVHRR, ETM+) and demonstrated the need for cross-calibration.

5. Conclusions

By focussing on a range of tropical rainforest types (regenerative, floodplain, *terra firme*) with structure varying over a mean height range of 8 m to 50 m, it has been shown that in particular near and middle infrared response is negatively related to canopy height. Furthermore, it has been shown that near infrared in a simple ratio or normalised index with either visible red or green also correlates strongly with canopy height (mean and maximum). The NDVI has therefore been demonstrated as a potential cost-effective tool for large-area assessment of tropical rainforest canopy structure appropriate for UN-REDD (Reduced Emissions from Deforestation and Forest Degradation) monitoring.

Acknowledgements

We are grateful to Dr. Bryan Mark of Ohio State University for providing the LiDAR data and to the USGS for access to Level 1 T Landsat ETM+ data.

References

ASNER, G.P., 2009. Tropical forest carbon assessment: integrating satellite and airborne mapping approaches. *Environmental Research Letters*, 4(3), article number 034009.

ASNER, G.P., and WARNER, A.S., 2003. Canopy shadow in IKONOS satellite observations of tropical forests and savannas. *Remote Sensing of Environment*, **87**, pp. 521-533.

AXELSSON, P., 2000. DEM generation from laser scanner data using adaptive TIN models. IAPRS 2000; **33**(B4), pp. 110-117.

BOYD, D.S., FOODY, G.M., CURRAN, P.J., LUCAS, R.M., and HONZAK, M., 1996. An assessment of radiance in Landsat TM middle and thermal infrared wavebands for the detection of tropical forest regeneration. *International Journal of Remote Sensing*, **17**, pp. 249-261.

BOYD, D.S., FOODY, G.M., CURRAN, P.J., 1999. The relationship between the biomass of

Cameroonian tropical forests and radiation reflected in middle infrared wavelengths

(3.0-5.0 um). *International Journal of Remote Sensing*, **20**, pp. 1017-1023.

DE MORAES, J.F.L., SEYLER, F., CERRI, C.C., and VOLKOFF, B., 1998. Land cover mapping and carbon pool estimates in Rondonia, Brazil. *International Journal of Remote Sensing*, **19**, pp. 921-934.

DRAKE, J.B., DUBAYAH, R.O., KNOX, R.G., CLARK, D.B., and BLAIR, J.B., 2002. Sensitivity of large footprint lidar to canopy structure and biomass in a neotropical rainforest. *Remote Sensing of Environment*, **81**, pp. 378-392.

DRAKE, J.B., KNOX, R.G., DUBAYAH, R.O., CLARK, D.B., CONDIT, R., BLAIR, J.B., and HOFTON, M.A., 2003. Above-ground biomass estimation in closed canopy Neotropical forests using lidar remote sensing: factors affecting the generality of relationships. *Global Ecology and Biogeography*, **12**, pp. 147-159.

ESPIRITO-SANTO, F.D.B., SHIMABAKURO, Y.E., and KUPLICH, T.M., 2005. Mapping forest successional stages following deforestation in Brazilian Amazonia using multi-temporal Landsat images. *International Journal of Remote Sensing,* **26**, pp. 635-642.

FOODY, G.M., and CUTLER, M.E.J., 2003. Tree biodiversity in protected and logged Bornean tropical rain forests and its measurement by satellite remote sensing. *Journal of Biogeography*, **30**, pp. 1053-1066.

FOODY, G.M., and CURRAN, P.J. 1994. Estimation of tropical forest extent and regenerative stage using remotely sensed data. *Journal of Biogeography*, **21**, pp. 223-244.

FOODY G.M., and HILL R. A., 1996. Classification of tropical forest classes from Landsat TM data. *International Journal of Remote Sensing*, **17**, pp. 2353-2367.

FOODY, G.M., CUTLER, M.E.J., MCMORROW, J., PELZ, D., TANGKI, H., BOYD, D.S., and DOUGLAS, I., 2001.Mapping the biomass of Bornean tropical rain forest from remotely sensed data. *Global Ecology and Biogeography*, **10**, pp. 379-387.

FOODY, G.M., BOYD, D.S., and CUTLER, M.E.J., 2003. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between groups. *Remote Sensing of Environment*, **85**, pp. 463-474.

GILLESPIE, T.W., SAATCHI, S., PAU, S., BOHLMAN, S., GIORGI, A.P., and LEWIS, S., 2009, Towards quantifying tropical tree species richness in tropical forests. *International Journal of Remote Sensing*, **30**, pp. 1629-1634.

HELMER, E.H., LEFSKY, M.A., and ROBERTS D.A., 2009. Biomass accumulation rates of Amazonian secondary forest and biomass of old-growth forests from Landsat time series and the Geoscience Laser Altimeter System. *Journal of Applied Remote Sensing*, **3**, article number: 033505.

HERNANDEZ-STEFANONI, J.L., DUPUY, J.M., and CASTILLO-SANTIAGO, M.A., 2009. Assessing species density and abundance of tropical trees from remotely sensed data and geostatistics. *Applied Vegetation Science*, **12**, pp. 398-414.

HILL R.A., and FOODY, G.M., 1994. Separability of tropical rain-forest types in the Tambopata-Candamo Reserved Zone, Peru. *International Journal of Remote Sensing*, **15**, pp. 2687-2693.

HILL, R. A., 1999, Image segmentation for humid tropical forest classification in Landsat TM data. *International Journal of Remote Sensing*, **20**, pp. 1039-1044.

HUDAK, A.T., LEFSKY, M.A., COHEN, W.B., and BERTERRETCHE, M., 2002, Integration of lidar and ETM+ data for estimating and mapping forest canopy height. *Remote Sensing of Environment*, **82**, pp. 397-416.

INGRAM, J.C., DAWSON, T.P. and WHITTAKER, R.J., 2005. Mapping tropical forest structure in southeastern Madagascar using remote sensing and artificial neural networks. *Remote Sensing of Environment*, **94**, pp. 491-507.

LEFSKY, M.A., HARDING, D.J., KELLER, M., COHEN, W.B. CARABAJAL, C.C., ESPIRITO-SANTO, F.D., HUNTER, M.O., and DE OLIVIERA, R., 2005. Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical Research Letters*, **32**(22), article number: L22S02.

LLOYD, H. (2004) Habitat and population estimates of some threatened lowland forest bird species in Tambopata, south-east Peru. *Bird Conservation International* **14**, pp. 261-277.

LU, D., MAUSEL, P. BRONDIZIO, E., and MORAN, E., 2004. Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. *Forest*

Ecology Management, **198**, pp. 149-167.

LUNG T., and SCHAAB, G., 2010. A comparative assessment of land cover dynamics of three protected forest areas in tropical eastern Africa. *Environmental Monitoring and Assessment*, **161**, pp. 531-548.

MENG, Q., CIESZEWSKI, C.J., MADDEN, M., and BORDERS, B., 2007. A linear mixed-effects model of biomass and volume of trees using Landsat ETM+ images. *Forest Ecology and Management*, **244**, pp. 93-101.

MIURA, T., HUETE, A., and YOSHIOKA H., 2006. An empirical investigation of cross-sensor relationships of NDVI and red/near-infrared reflectance using EO-1 Hyperion data. *Remote Sensing of Environment*, **100**, pp.223-236.

MYERS, N., MITTERMEIER, R.A., MITTERMEIER, C.G., DA FONSECA, G.A.B. and KENT, J., 2000. Biodiversity hotspots for conservation. *Nature,* **403**, pp. 853-858.

NELSON, R.F., HORNING, N., and STONE, T.A., 1987. Determining the rate of forest conversion in Mato Grosso, Brazil using Landsat MSS and AVHRR data. *International Journal of Remote Sensing,* **8**, pp. 1767-1784.

PHILLIPS, O., 1993, *Comparative Valuation of Tropical Forests in Amazonian Peru.* Unpublished PhD thesis, Missouri Botanic Gardens, St. Louis, Missouri, USA.

ROY, P.S., RANGANATH, B.K., DIWAKAR, P.G., VOHR, T.P.S., BHAN, S.K., SINGH, I.J., and PANDIAN, V.C., 1991. Tropical forest type mapping and monitoring using remote sensing. *International Journal of Remote Sensing,* **12**, pp. 2205-2225.

SADER, S.A., WAIDE, R.B., LAWRENCE, W.T., and JOYCE, A.T., 1989. Tropical forest biomass and successional age class relationships to a vegetation index derived from Landsat

TM data. *Remote Sensing of Environment*, **28**, pp. 143-156.

SALOVAARA, K.J., THESSLER, S., MALIK, R.N., and TUOMISTO, H., 2005. Classification of Amazonian primary rain forest vegetation using Landsat ETM plus satellite imagery. *Remote Sensing of Environment*, **97**, pp. 39-51.

SINGH, A., 1987. Spectral separability of tropical forest cover classes. *International Journal of Remote Sensing,* **8**, pp. 623-627.

STEININGER, M.K., 1996. Tropical secondary forest regrowth in the Amazon: age, area and change estimation with Thematic Mapper data. *International Journal of Remote Sensing,* **17**, pp. 9-27.

STEININGER, M.K., 2000. Satellite estimation of tropical secondary forest above-ground biomass: data from Brazil and Bolivia. *International Journal of Remote Sensing,* **21**, pp. 1139-1157.

TWELE, A., ERASMI, S., and KAPPAS, M., 2008. Spatially explicit estimation of leaf area index using EO-1 Hyperion and Landsat ETM+ data: implications of spectral bandwidth and shortwave infrared data on prediction accuracy in a tropical montane environment. *GIScience and Remote Sensing*, **45**, pp. 229-248.

Fusion of full waveform Lidar and hyperspectral remote sensing data for the characterization of forest stands

Henning Buddenbaum and Joachim Hill

Remote Sensing Department, University of Trier, Trier, Germany

Abstract

Full waveform small footprint laser scanning und airborne hyperspectral image data from the sensor HyMap of two forest areas in Germany were fused for a detailed forest stand characterization. Fusing active laser scanning data with passive hyperspectral data increases the information content without adding much redundancy.

Hyperspectral data offer the maximum spectral reflectance information available from remote sensing. Full waveform laser scanning records the vertical distribution of reflected laser energy and offers the maximum of structural information about forest stands.

In order to combine both datasets, we defined voxels above the HyMap pixels, containing the mean laser intensity in slices of 50 cm or 100 cm height for the area of each HyMap pixel. These datasets yield a detailed impression of the vertical structure of forest stands and can be used, among other things, to derive tree height, crown base height, and biomass. In addition, the joined images performed better in classifying tree species and age classes than each of the single images.

1. Introduction

Active and passive remote sensing systems inherently differ in how they measure the earth. While passive remote sensors measure the reflected solar radiation (or the emitted radiation), active sensors detect the echoes of their own radiation source. The highest information content among passive sensors is achieved by hyperspectral imaging, also known as imaging spectroscopy (IS, Goetz 2009). IS provides quasi-continuous reflectance spectra in many narrow bands for each pixel (Vane & Goetz 1988), but only for the surface of the observed object. This data can be used for detailed quantitative analyses, e.g. determination of chlorophyll or water content in leaves, or for difficult classification tasks, e.g. tree species determination (Le Maire *et al.* 2008, Sims & Gamon 2003, Ustin & Xiao 2001). However, in forested areas, discerning between different age classes of the same tree species is still error-prone (Buddenbaum *et al.* 2005, Niemann 1995). In addition, IS is not well suited for the derivation of structural forest parameters like tree height or crown volume (Chopping *et al.* 2009). Airborne Laser Scanning (ALS), as an active remote sensing technique, offers different information about the study area (Koetz *et al.* 2007). The three-dimensional position of each reflecting point is measured. Full waveform small footprint ALS records the temporal distribution of reflected laser energy, thus enabling the description of the vertical structure of vegetation stands. Laser scanner data can be used to differentiate between coniferous and deciduous trees either by analysing crown shapes that lead to different surface roughness or by exploiting the intensity information of laser echoes from the crowns (Ørka *et al.* 2009, Kim *et al.* 2009). However, a more detailed determination of tree species based solely upon Laserscanning is not feasible (Holmgren *et al.* 2008). A combination of both data sets provides more information on canopies than any single remote sensing system.

This abundance of spectral and structural information can be used for the characterization of forest stands, e.g. for species and age class classification or for the determination of structural or biochemical properties. Many studies have investigated the information extent of ALS alone (e.g. Koch *et al.* 2009, Hyyppä *et al.* 2008) and of IS alone (e.g. Schlerf *et al.* 2010, Ghiyamat & Shafri 2010, Schlerf & Atzberger 2006, Zarco-Tejada *et al.* 2004) for forest applications. There are also several studies on the combination of passive image data and ALS data (e.g. Verrelst *et al.* 2009, Reitberger *et al.* 2008, Anderson *et al.* 2008, Dalponte *et al.* 2008) but few, if any, researchers have examined the fusion of IS and small-footprint waveform ALS.

Study areas and data

The study was conducted in two areas: The Idarwald forest (49°40'N, 7°10'E) near Morbach and a section of Pfälzerwald forest (49°16'N, 7°47'E) near Merzalben, both located in Rhineland-Palatinate, Germany. Both are managed forests with mixed and pure stands of beeches, oaks, Norway spruces and Douglas firs as dominant tree species. Active forestry measures in the areas include plantation and selective thinning.

In 2003 a HyMap image with 5 m ground sampling distance was recorded over Idarwald forest. In 2005 a full waveform small footprint ALS dataset of the same area was recorded using a Riegl Litemapper 5600 (Hug *et al.* 2005) with a Laser pulse density of about 4 m^{-2}.

IS and ALS images of the Pfälzerwald study area were both recorded in 2009, using the sensors HyMap and ALTM 3100, respectively. The spatial resolutions were similar to the Idarwald datasets. All data sets were recorded under leaf-on conditions in late summer.

HyMap is an airborne opto-mechanic IS sensor that covers the wavelength range from 0.45 to 2.48 µm with a radiometric resolution of 13–17 µm in about 125 bands with a signal to noise ratio of typically over 500:1. There are gaps in the wavelength regions affected by water vapour around 1450 and 1950 nm. The sensor's field of view is 60° (Cocks *et al.* 1998).

2. Methods

The HyMap images were radiometrically, atmospherically and geometrically corrected and geocoded. Bad bands were identified and excluded from the datasets resulting in a 122 band image of the Idarwald site and a 125 band image of the Pfälzerwald site.

With a HyMap pixel size of 5 m × 5 m and an average Laser pulse density of 4 m^{-2} there are about 100 Laser shots on every pixel. To create a dataset that integrates the waveform information for the area of each HyMap pixel, we defined voxels above ground level of the HyMap pixels, containing the mean laser intensity in slices of 50 cm (Idarwald) or 100 cm (Pfälzer Wald) height (Popescu & Zhao 2008). Every hit with non-zero intensity in a voxel was counted and the intensities were added to each other so that the mean intensity in each voxel could be calculated. Ground height for each pixel was taken from a DEM created from the same ALS flight. In order to get a uniform spatial distribution of brightness, the intensities were normalised using a $1/r^2$ relationship where r is the distance from the sensor to the ground. The resulting images do not show the flight lines or areas covered by multiple flight lines which is taken as evidence for an adequate intensity correction.

The mean values in each slice above ground were stored as multi-band image files. The highest trees in the Idarwald site were 38 m, so at 0.5 m vertical resolution we have 76 bands; the highest trees in the Pfälzerwald site are 40 m, so at 1 m vertical resolution we have 40 bands. Because these images have the same geometry as the HyMap images they were easily joined to form multi-band images with 196 channels (Idarwald) and 165 channels (Pfälzerwald). Since most normalised mean ALS intensities have values below 100, they were combined with IS reflectance scaled as percent.

3. Results

Figure 22 depicts some combined IS/ALS waveform spectra of the Idarwald study site. The left part of each plot contains the 122 HyMap reflectance values (in percent). The right part contains the mean ALS intensity at the given height above ground. All pictured spectra show a distinct peak at ground level. But pixels without an echo from the ground, i.e. completely closed canopies, can also be found in the images. Spectra *a* and *b* show similar coniferous stands with nearly identical IS spectra. The waveform spectra reveal their difference: stand *a* is nearly closed (only a low ground echo) and up to 30 m high with crowns between 12 and 30 m, stand *b* is more open (a high peak at ground level) and up to 38 m high with a crown space about 18 to 38 m above ground.

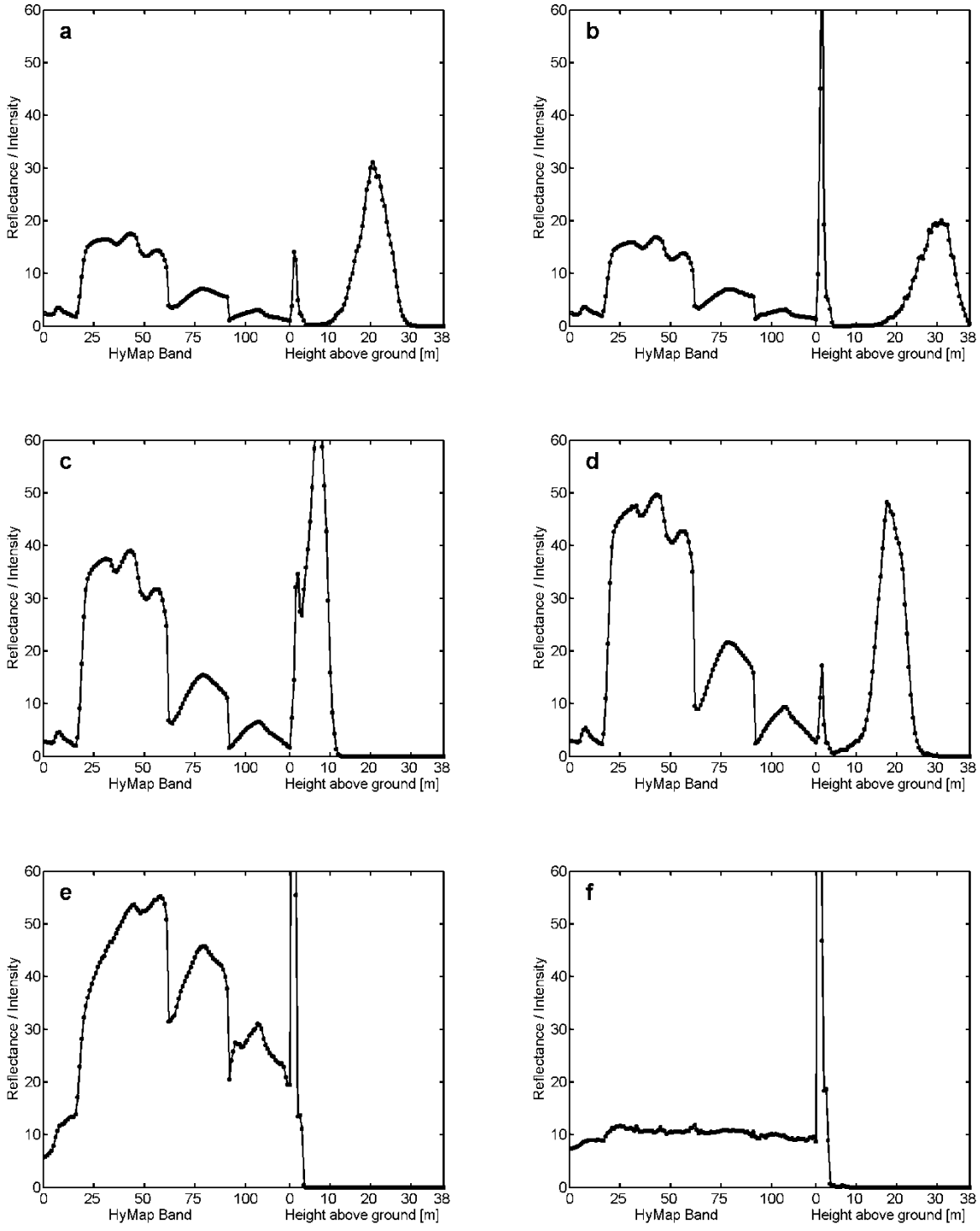


Figure 22. Combined hyperspectral and ALS waveform spectra from Idarwald forest.

Stands *c* and *d* also display similar IS signals (deciduous trees), but clear differences in the ALS signal. Stand *c* is very young, only 10 m high with crowns touching ground, while stand *d* is about 25 m high. In stand *d* only a small portion of the laser light is reflected from the ground. Plots e and f are examples for spectra without three-dimensional structure. Plot e is a mostly uncovered soil; plot f is an asphalt road.

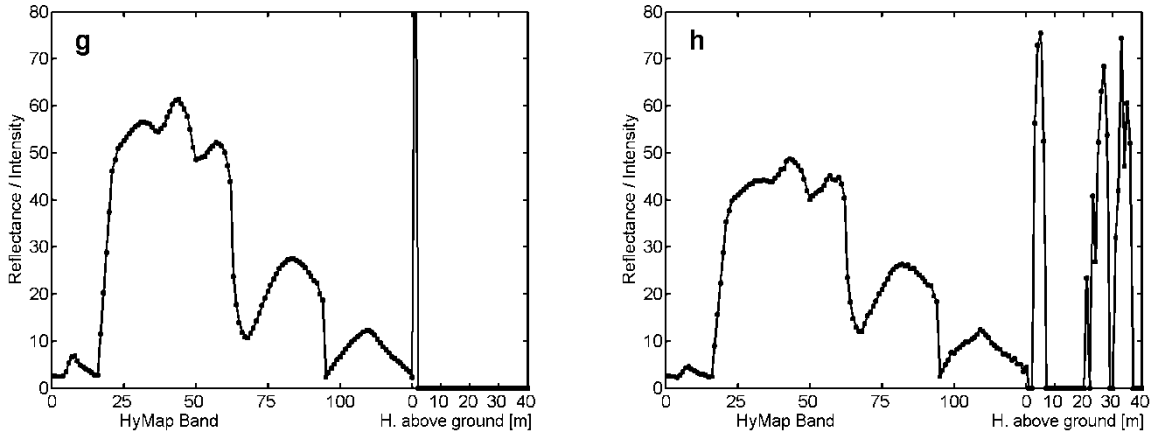


Figure 23: Combined ALS and waveform spectra from Pfälzerwald.

Figure 23 shows combined spectra from the Pfälzerwald site. In this case discrete point data (including intermediate echoes) instead of the waveform ALS data was used to create the right parts of the combined spectra, leading to less smooth shapes. This approach is similar to the one described in Riaño *et al.* (2003). Plot *g* is a grassland spectrum, plot *h* is a multi-storey forest spectrum.

Possible applications for the fused dataset are numerous. For instance, difficult classification tasks can be tackled. A test classification of coniferous stands from the Idarwald data set with four age classes of Norway spruce and two age classes of Douglas fir was conducted using HyMap data alone, HyMap data combined with waveform data and HyMap data combined with several height percentiles derived from the ALS data. The classification setup was similar to the one described in Buddenbaum *et al.* (2005), but different input data was used. The combined IS/ALS waveform data set led to the highest classification accuracy, followed by the IS/ALS percentile data.

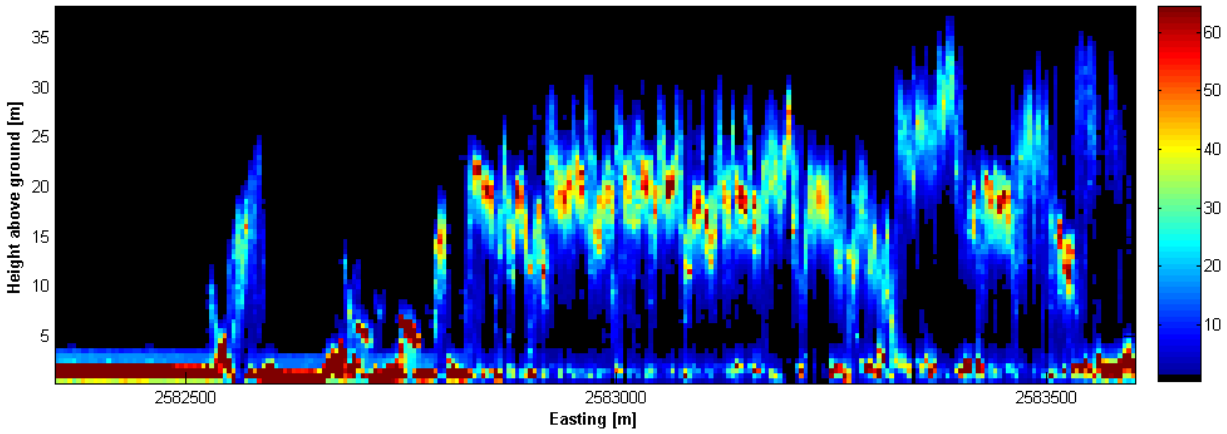


Figure 24: Vertical slice through ALS waveform voxels.

Further applications include the possibility of easily creating vertical and horizontal slices through the data. Figure 24 shows a vertical slice through a transect of about 1.2 km of forest and clearing areas. The reflected intensities are colour-coded. Forest-free areas, tree heights and vertical crown extents are well visible. Figure 25 depicts horizontal slices through a part of Idarwald forest at 10, 20, and 30 m above ground as an RGB composite. Stands with dense crowns at the respective heights are shown in high intensity at their colour. Because few crowns span more than 10 m with high reflectance, most areas are dominated by one colour and mixed colours are rare. Areas displayed in black have no trees or only trees smaller than 10 m.

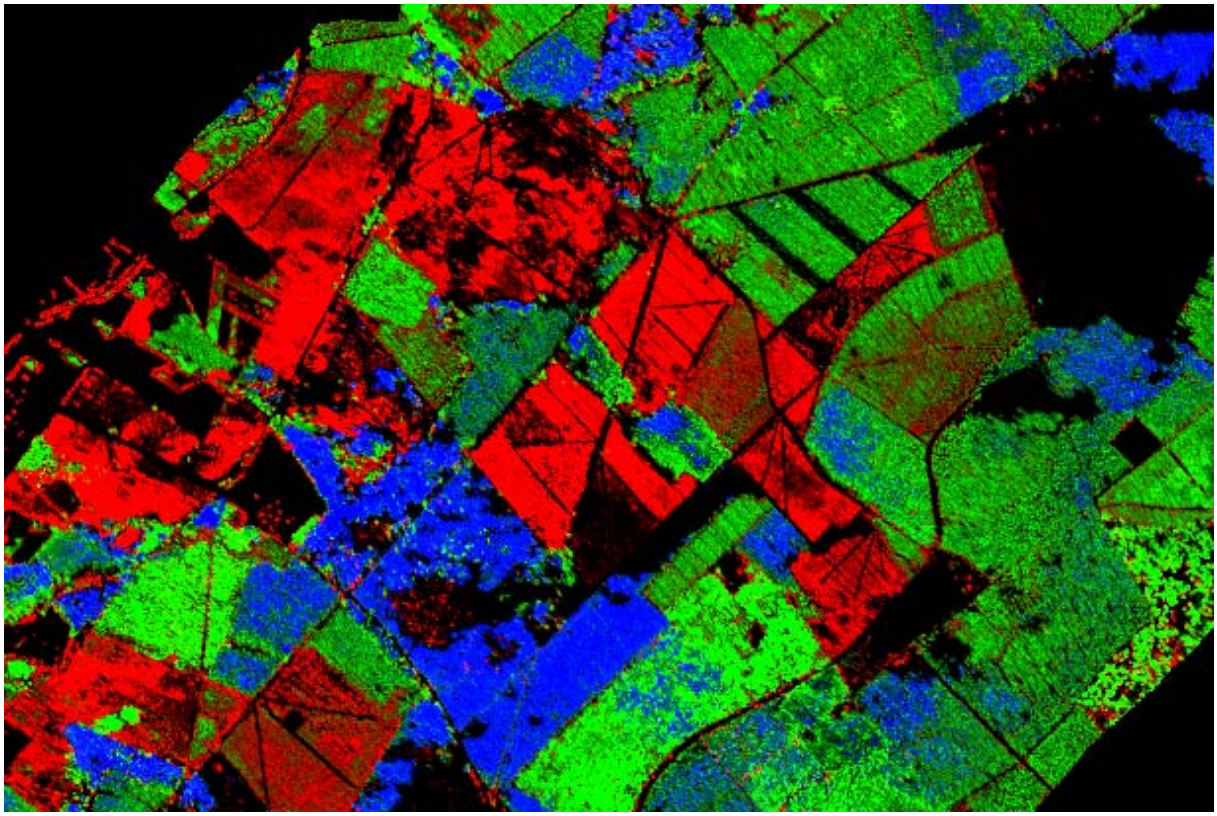


Figure 25: Horizontal slices at 10, 20, and 30 m above ground depicted in red, green, and blue, respectively.

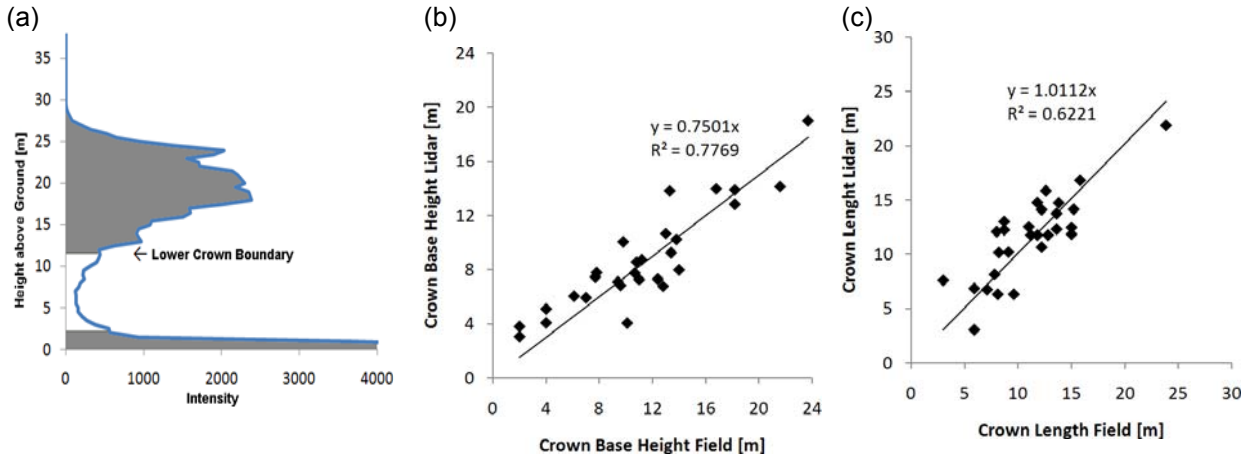


Figure 26: (a) The lower crown boundary is defined as the 10th intensity percentile above 2 m above ground, (b) Crown base heights from field measurements and from ALS, (c) crown lengths from field measurements and from ALS.

Another application of the combined ALS wavelengths is a physically-based measurement of the crown base heights and the vertical crown extent in one-storey forest stands. The crown base height is an important characteristic for forest health, timber quality and fire behaviour (Vauhkonen 2010). Following an approach by Riaño *et al.* (2003) and similar to an approach by Næsset & Økland (2002), the lower crown boundary was defined as a percentile above a fixed height of ground vegetation echoes. We chose the 10th percentile

of intensity above 2 m above ground (Figure 26a). Figure 26b shows the relationship between field-measured and ALS-derived crown base heights for 28 stands of beech and spruce measured in the Idarwald forest in late summer 2005. The ALS-derived values are a little too low, but there is a good correlation. The corresponding relationship for crown lengths, calculated as the difference of tree heights and crown base heights, is shown in Figure 26c. The RMSE for crown base heights is 3.6 m, the RMSE for crown lengths is 2.3 m.

4. Conclusions

Combining IS and ALS data, i.e. passive and active remote sensing, for forest applications increases the dimensionality without adding much redundancy. The vertical intensity profiles obtained from combining several small-footprint waveforms correspond to large-footprint waveforms of exactly the desired pixel size, in this case the pixel size given by the HyMap image. These intensity profiles provide a detailed impression of crown closure and the crown structure, especially the vertical crown extent.

The combined datasets have proven valuable as classification input and as a detailed description of the reflectance spectrum and the vertical structure of each pixel. The composite ALS profiles alone can be used for a physically-based derivation of structural parameters like crown base height and crown length. While single waveform profiles are rather noisy, the composite waveforms show the vertical crown extent well, although the lower crown regions are hit by much less laser energy than the upper regions and are thus represented worse.

An even more detailed stand characterization through the combination of remote sensing data sources would be possible using multi-temporal or multi-angle image data or higher spatial resolution data, e.g. terrestrial laser scanning. Both, multi-temporal hyperspectral data acquisition and terrestrial laser scans of some forest stands, are planned in the near future. An expected problem is that in these approaches the amount of data per observed area grows very large, but it is unknown if the amount of useful information also grows. A lot of the extra data is expected to be redundant, in contrast to the combination of active and passive data sets presented in this paper.

References

ANDERSON, J.E., PLOURDE, L.C., MARTIN, M.E., BRASWELL, B.H., SMITH, M.-L., DUBAYAH, R.O., HOFTON, M.A. & BLAIR, J.B., 2008, Integrating waveform Lidar with hyperspectral imagery for inventory of a northern temperate forest. *Remote Sensing of Environment*, **112**, pp. 1856–1870.

BUDDENBAUM, H., SCHLERF, M. & HILL, J., 2005, Classification of coniferous tree species and age classes using hyperspectral data and geostatistical methods. *International Journal of Remote Sensing*, **26**, pp. 5453–5465.

CHOPPING, M., NOLIN, A., MOISEN, G.C., MARTONCHIK, J.V. & BULL, M., 2009, Forest canopy height from the Multiangle Imaging SpectroRadiometer (MISR) assessed with high resolution discrete return lidar. *Remote Sensing of Environment*, **113**, pp. 2172–2185.

COCKS, T., JENSSEN, R., STEWART, A., WILSON, I. & SHIELDS, T., 1998, The HyMap Airborne Hyperspectral Sensor: The System, Calibration and Performance. *1st EARSEL Workshop on Imaging Spectroscopy*, Zurich, October 1998.

DALPONTE, M., BRUZZONE, L. & GIANELLE, D., 2008, Fusion of hyperspectral and LIDAR Remote Sensing Data for Classification of Complex Forest Areas. *IEEE Transactions on Geoscience and Remote Sensing*, **46**, pp. 1416–1427.

GHIYAMAT, A. & SHAFRI, H.Z.M., 2010, A review on hyperspectral remote sensing for homogeneous and heterogeneous forest biodiversity assessment. *International Journal of Remote Sensing*, **31**, pp. 1837–1856.

GOETZ, A.F., 2009, Three decades of hyperspectral remote sensing of the Earth: A personal view. *Remote Sensing of Environment*, **113**, pp. 1–13.

HOLMGREN, J., PERSSON, Å. & SÖDERMAN, U., 2008, Species identification of individual trees by combining high resolution LiDAR data with multi-spectral images. *International Journal of Remote Sensing*, **29**, pp. 1537–1552.

HUG, C., ULLRICH, A. & GRIMM, A., 2005, Litemapper-5600 – A Waveform-digitizing Lidar Terrain and Vegetation mapping System. In: M. Thies, B. Koch, H. Spiecker & H. Weinacker (eds., *Proc. ISPRS Workshop Laser Scanners for Forest and Landscape Assessment,* Freiburg, Germany, 3–6 October 2004. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVI - 6/W2.

HYYPPÄ, J., HYYPPÄ, H., LECKIE, D., GOUGEON, F., YU, X. & MALTAMO, M., 2008, Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing*, **29**, pp. 1339–1366.

KIM, S., MCGAUGHEY, R.J., ANDERSEN, H.-E. & SCHREUDER, G., 2009, Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. *Remote Sensing of Environment*, **113**, pp. 1575–1586.

KOCH, B., STRAUB, C., DEES, M., WANG, Y. & WEINACKER, H., 2008, Airborne laser data for stand delineation and information extraction. *International Journal of Remote Sensing*, **30**, pp. 935–963.

KOETZ, B., SUN, G., MORSDORF, F., RANSON, K.J., KNEUBÜHLER, M., ITTEN, K. & ALLGÖWER, B., 2007, Fusion of imaging spectrometer and LIDAR data over combined radiative transfer models for forest canopy characterization. *Remote Sensing of Environment*, **106**, pp. 449–459.

LE MAIRE, G., FRANÇOIS, C., SOUDANI, K., BERVEILLER, D., PONTAILLER, J.-Y., BRÉDA, N., GENET, H., DAVI, H. & DUFRÊNE, E., 2008, Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass, *Remote Sensing of Environment*, **113**, pp. 3846–3864.

NÆSSET, E. & ØKLAND, T., 2002, Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. *Remote Sensing of Environment*, **79**, pp. 105–115.

NIEMANN, K. O., 1995, Remote Sensing of Forest Stand Age Using Airborne Spectrometer Data. *Photogrammetric Engineering & Remote Sensing*, **61**, pp. 1119-1127.

ØRKA, H.O., NÆSSET, E. & BOLLANDSÅS, O.M., 2009, Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. *Remote Sensing of Environment*, **113**, pp. 1163–1174.

POPESCU, S.C. & ZHAO, K., 2008, A voxel-based lidar method for estimating crown base height for deciduous and pine trees. *Remote Sensing of Environment*, **112**, pp. 767–781.

REITBERGER, J., KRZYSTEK, P. & STILLA, U., 2008, Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees. *International Journal of Remote Sensing*, **29**, pp. 1407–1431.

RIAÑO, D., MEIER, E., ALLGÖWER, B., CHUVIECO, E. & USTIN, S.L., 2003, Modeling airborne laser scanning data for the spatial generation of critical forest parameters in fire behavior modeling. *Remote Sensing of Environment*, **86**, pp. 177–186.

SCHLERF, M. & ATZBERGER, C., 2006, Inversion of a forest reflectance model to estimate structural canopy variables from hyperspectral remote sensing data. *Remote Sensing of Environment*, **100**, pp. 281–294.

SCHLERF, M., ATZBERGER, C., HILL, J., BUDDENBAUM, H., WERNER, W. & SCHÜLER, G., 2010, Retrieval of chlorophyll and nitrogen in Norway spruce (*Picea abies* L. Karst.) using imaging spectroscopy. *International Journal of Applied Earth Observation and Geoinformation*, **12**, pp. 17–26.

SIMS, D.A. & GAMON, J.A., 2003, Estimation of vegetation water content and photosynthetic tissue area from spectral reflectance: a comparison of indices based on liquid water and chlorophyll absorption features. *Remote Sensing of Environment*, **84**, pp. 526–537.

USTIN, S.L. & XIAO, Q.F., 2001, Mapping successional boreal forests in interior central Alaska. *International Journal of Remote Sensing*, **22**, pp. 1779–1797.

VANE, G. & GOETZ, A.F., 1988, Terrestrial Imaging Spectroscopy. *Remote Sensing of Environment*, **24,** pp. 1–29.

VAUHKONEN, J., 2010, Estimating crown base height for Scots pine by means of the 3D geometry of airborne laser scanning data. *International Journal of Remote Sensing*, **31**, pp. 1213–1226.

VERRELST, J., GEERLING, G.W., SYKORA, K.V. & CLEVERS, J.G.P.W., 2009, Mapping of aggregated floodplain plant communities using image fusion of CASI and LiDAR data. *International Journal of Applied Earth Observation and Geoinformation*, **11**, pp. 83–94.

ZARCO-TEJADA, P.J., MILLER, J.R., MORALES, A., BERJÓN, A. & AGÜERA, J., 2004, Hyperspectral indices and model simulation for chlorophyll estimation in open-canopy tree crops. *Remote Sensing of Environment*, **90**, 463–476.

Laserscanning for forest inventory & planning – an operational project for the Swiss Canton of Appenzell – Ausserrhoden

Manuela Hirschmugl†, Hubertus Schmidtke‡, Roland Wack†, Martin Ofner†, Mathias Schardt†, Klaus Granica†*

† Institute of Digital Image Processing, Joanneum Research, Graz, Austria

‡ SILVACONSULT, Winterthur, Switzerland

Abstract

This paper describes an operational method of forest stand and parameter mapping for the Swiss Canton of Appenzell – Ausserrhoden (243 km² with 80 km² of forest), ranging from 400 m up to 2500 m above sea level. The method is operational in terms of its commercial use. It is robust against varying data qualities and forest conditions. There is well defined product at fixed costs. LiDAR first and last pulse data from a Leica ALS50-II scanner and a SPOT 5 multispectral satellite image with a geometric resolution of 10 m were used as the basic data sets. In the first step, automatic image segmentation was used to generate forest stands, which were then adjusted according to user needs. Subsequently, the following forest parameters were calculated per stand: age class (based on tree height); species mixture; crown cover percentage; dominant tree height; mean tree height; standing timber volume and living tree biomass.

1. Introduction and Literature Review

The conventional technology for forest stand mapping in Switzerland is delineation and visual interpretation of aerial photographs (stereo or ortho). This guaranties a high quality of forest stand maps to a certain price. Customers are per se not interested in technology but in the final product. Their question is: What is the advantage of the a forest stand map derived from LiDAR data against the well known and accepted conventional forest stand maps from aerial photographs in terms of content, accuracy and costs? To offer results from LiDAR technology to customers it is necessary to very accurately define the final product. This product is a digital forest stand map with clearly defined attributes at a fixed prize. In a previous pilot project SILVACONSULT AG and Joanneum Research had developed a procedure for forest stand mapping using LiDAR data and multispectral data [Schmidtke et al., 2008]. In this pilot project financed by four Swiss cantons and the Swiss federation, three test sites of 1000 ha each were processed. They represented different data situations, different conditions of forests and geomorphology. Target was to develop a robust procedure that could be used under different conditions and on large areas. The aim of this pilot project was not to implement all scientifically published possible methods and results derived from LiDAR data, neither new basic research. Instead, the aim was to collect the robust components of LiDAR data applications available and fit them together into a cost-efficient production process. Based on the results of this pilot project, the Swiss Canton of Appenzell Ausserrhoden ordered the first forest stand mapping product using LiDAR data with our procedure under commercial conditions. This paper describes the procedure.

Airborne LiDAR (Light Detection and Ranging) is an active remote sensing technology emitting laser pulses in the visible or near infrared wavelength and measuring the time lag between the emission and the return of the reflected pulse(s). If a laser pulse is send out over a vegetated surface such as forest, multiple reflections can occur. Typically the first reflection (first pulse) represents the height of the canopy, while part of the beam penetrates the canopy and is reflected as a last pulse from the ground. Filtering techniques are used to separate ground and canopy signals [Wack and Wimmer, 2002]. This kind of data has proven to be very useful to derive main forest attributes, as a large amount of scientific papers have been dealing with this issue over the past decade [Hyyppä and Hyyppä, 1999, Hyyppä et al., 2000, Hyyppä et al., 2002,Ziegler et al.,2002]. There are basically two different ways of deriving forest parameters using first and last pulse data: either on an individual tree basis [Koch et al., 2006, Pitkänen et al., 2004] or on stand level [Andersen et al.,

2003, Barbati et al., 2009, Koch et al., 2009, Næsset, 2002, Wack and Stelzl, 2005]. For individual tree measurements, the most frequently derived forest attributes are tree position, height, crown width, crown base height and as secondary products diameter at breast height (dbh), basal area and timber volume of the individual trees. Few studies have been trying to extract species information, e.g. [Donoghue et al., 2007]. Stand-level forest attributes are most often timber volume or above-ground biomass [Barbati et al., 2009, Hollaus et al., 2009].

This paper describes a combined single-tree and stand-wise approach to derive the following forest parameters at a stand level: age class, species mixture, crown cover percentage in total and of the upper canopy, dominant tree height, standing timber volume and living tree biomass. In this regard, the individual tree detection process is only an intermediate result for the derivation of the stand-wise attributes. The aim of this development was to generate robust and operational approach of the use of airborne LiDAR data in combination with multispectral satellite images for a large area forest mapping exercise done in the Swiss Canton of Appenzell – Ausserrhoden. The idea behind this development was to significantly reduce the amount of both field work and manual digitizing work and thus to reduce costs for the forest inventory.

Data

Two main data types were used for the implementation of this work: airborne LiDAR and multispectral satellite data from SPOT V satellite. From an airborne LiDAR campaign, first and last pulse data from a Leica ALS50-II scanner at a frequency of 150.000 Hz provided a digital terrain as well as surface model. The second main data set was a SPOT 5 multispectral satellite image with a geometric resolution of 10 m mainly used for tree species classification. In addition, airborne CIR images have been used for selection of training data, for the adjustment of the segment border and for final checks. The properties of the used data sets are summarized in Table 16. Since no mayor disturbances like storms occurred between 2007 and 2009, the time difference between airborne and space borne data sets is not critical.

Table 16: Data properties

Data	Acquisition date	Sensor	Geometric resolution
LiDAR data	14.6.–1.8. 2009	Leica ALS50-II	Average: 11 points/m²
multispectral satellite data	15.8.2007	SPOT V	10 m pixel spacing
CIR orthophotos	15.–16.8. 2009	UltracamX digital frame camera	10 cm pixel spacing

2. Methods

The overall process is sketched in Figure 27 with the inputs in light gray and the main processing steps in dark gray. First, the LiDAR DSM and DTM are used to calculate a vegetation height model (VHM). Faults in the VHM are to be identified and eliminated: power poles, bridges over gullies, overhanging trees on rocks etc. The clean VHM is used for the tree top detection (see Chapter 3.1.). In parallel, the CIR orthophotos are used to identify ground control points (GCPs) in the SPOT scene which are required to orthorectify the satellite image. This orthorectified SPOT image and the VHM are used for the segmentation of forest stands (see Chapter 3.2.). For the classification of the tree species, a standard pixel-based maximum likelihood classification is performed. Finally, all intermediate results (tree tops, forest stands and species information) and auxiliary information on yield are used for the derivation of the stand-wise forest parameters (see Chapter 3.3.).

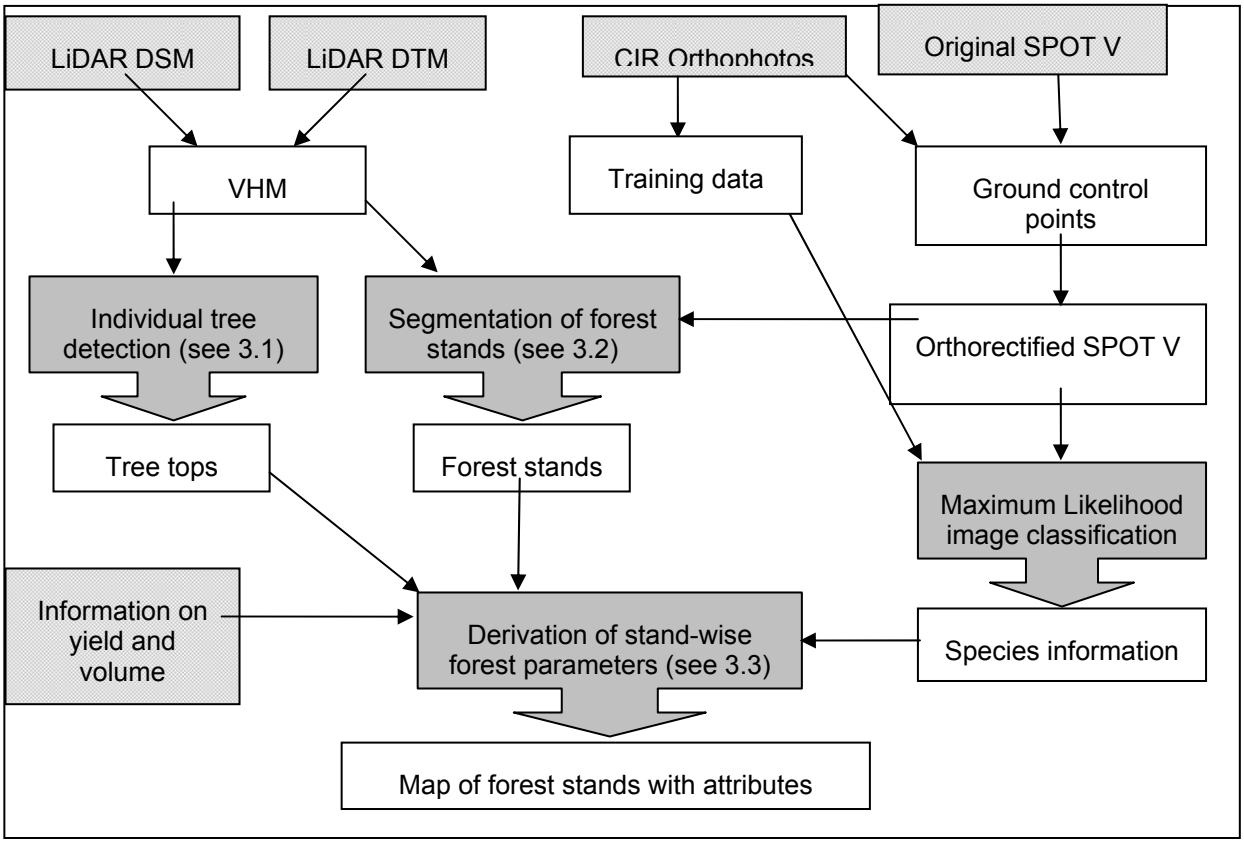


Figure 27: Overall process description

Individual tree detection

The method was developed at the Institute of Digital Image Processing, Joanneum Research [Wack and Stelzl, 2005] and is based on Laplacian-of-Gauss (LoG) filtering. For mathematical details on this filtering approach, see e. g. [Gonzalez and Woods, 2002]. The procedure consists of the following steps; intermediate results are shown in Figure 2.

- **The LoG is used to blur the image, with the degree of blurring being determined by the value of the standard deviation _. The procedure used here involves three scales of LoG filtering based on three different sigma values (2, 3, 4) in order to detect trees of different sizes. The results of the LoG filtering with different sigma values are depicted in Figure 2 b, c and d. The dependence of the tree detection success from a single chosen sigma has been discussed by [Chen et al., 2006].**

- **A local maximum approach is performed on the original VHM, see Figure 2 e.**

- **The LoG images are weighted according to their respective level and then added (Figure 2 f).**

- **From this summation image, intensity maxima are detected again using LMA; the result is shown in Figure 2 g.**

- **Finally, these intensity maxima are dragged to their nearest height maximum (result from step 2).**

- **Result is a vector file with points representing the tree tops detected with coordinates and height. The final result is visualised in Figure 2 h.**

There are some uncertainties related mainly to dense deciduous forests. If trees are standing close together, especially in young stands, there may be some omission error. Big broadleaf trees without one clear tree top may have several superfluous tree top detections. For visual interpretation, the distinction of individual deciduous trees of the same species is equally difficult and erroneous. In the case of Appenzell Ausserrhoden there are 85 % conifers, which lead to quite precise detections.

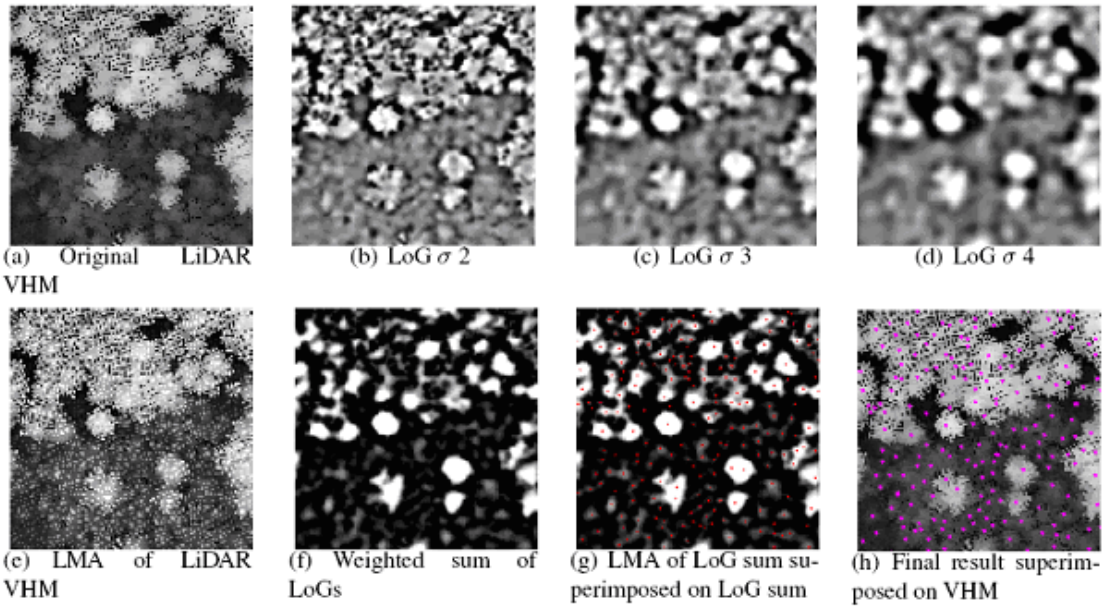


Figure 28: Processing steps and intermediate results for the LoG approach based on LiDAR data (from [Hirschmugl, 2008])

Segmentation of forest stands

A forest stand is typically defined by properties such as age and age distribution, species, density, yield, necessity of measures, site quality etc. These properties are traditionally assessed through field work and through visual interpretation of aerial (stereo) images. In this project, the use of automatic segmentation is assessed in order to save time for manual delineations. A processing chain of several filtering, segmentation and merging steps was set up to generate homogeneous segments. The main input data sets used are again the VHM and the SPOT image. In addition, existing information on infrastructure such as roads and forest roads, which are generally considered as fixed stand borders, can optionally also be integrated.

Not all properties typically used for forest stand delineation can be derived from remote sensing data, examples are local yield or site conditions. However, some main characteristics can be used:

- the spectral signature of the SPOT image has a strong correlation with the tree species (especially the NIR and SWIR bands for coniferous and deciduous differentiation);
- the tree height (VHM) is typically correlated with the age of a stand (with some restrictions);
- tree density and structure are well respresented in the LiDAR VHM.

Thus, the first step for the forest stand segmentation is the generation of an artificial stack of three bands consisting of

1. the first principal component image of the multispectral SPOT image
2. the mean height information generated from the LiDAR VHM
3. a structure feature, also calculated from the LiDAR VHM with a so-called ‘sector-statistics’ approach

All three inputs were resampled to a common resolution of 5m. This three-band image was then integrated with existing forest roads as fixed stand borders and segmented using a region growing approach. In a post-

segmentation step, segments below the minimum mapping unit were merged with the adjacent, spectrally most similar segment. The automatically generated segments of the forest stands were finally revised visually where necessary.

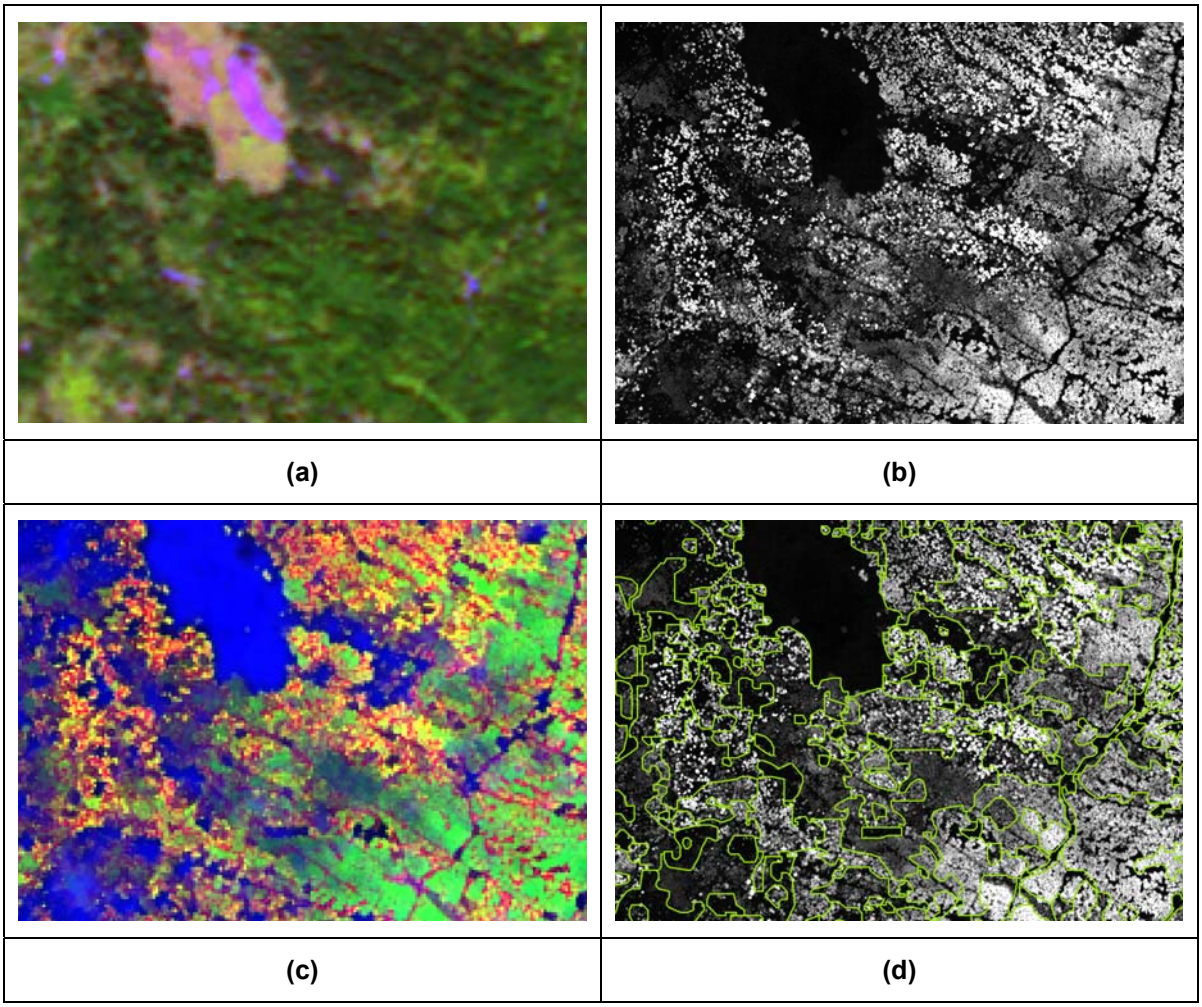


Figure 29: (a) SPOT image; (b) VHM; (c) artificial stack of properties; (d) VHM overlaid with segment borders

Derivation of stand-wise forest parameters

Number of trees

The individual tree tops were detected (see Section 3.1). The number of trees per ha (stem density) and the total number of trees are the output stand parameters.

Height information:

Based on the individual tree detections, three different segment-wise height values are estimated: dominant height, mean height and dominant height of the suppressed trees. These three values are calculated as follows:

Dominant height = Mean height of the 20% highest detected trees of the segment

Mean height = mean height of all detected trees within the segment

Dominant height of the suppressed trees = mean height of the 20% highest detected trees smaller than 2/3 of the dominant height.

Crown cover percentages:

For the estimation of the crown cover percentage of each segment, the VHM was cut off at a user-defined threshold (in the current study at 1.3m) and all area above this threshold are considered as covered. By intersecting this info with the segments, the crown cover percentages can be calculated. In addition to the overall crown coverage that one of the upper canopy layer (upper third of the stand height) was derived from the LiDAR data. This is an important silvicultural parameter.

Stage of stand development:

In Switzerland the age of the forest stands normally is not known and often stands are uneven aged and mixed. Main parameter is the stage of development. This is defined through the stem diameter at breast height (dbh) of the dominant trees which means the 100 biggest trees per hectare (ddom). According to the yield tables from [Badoux, 1983], the stage of stand development is defined as given in Table 17. The diameter at breast height correlates with the tree height depending on tree species and yield class. For each yield class the related tree height classes can be taken to calculate the stage of development of each stand out of the tree heights derived from the LiDAR data. Because in Appenzell Ausserrhoden there is no standwise information on soil types or productivity, an average yield class was taken for the whole Canton.

Table 17: Definitions for stages of stand development

Structure	Stage of development	Crown cover	Diameter of dominant layer (d_{dom})	Dominant height (h_{dom})	Code
homogeneous	Young stands	> 20%		<= 1.3 m	1
	Thicket	> 20%	<12 cm	> 1.3 m -	2
	Pole timber 1	>= 20%	12-20	Relation between h_{dom} - d_{dom} according to yield tables from Badoux	3
	Pole timber 2	>= 20%	21-30		4
	Timber 1	>= 20%	31 -40		5
	Timber 2	>= 20%	41 -50		6
	(Timber 3 - strong timber)	>= 20%	> 50		7
heterogeneous	mixed	>= 20%	mixed	Threshold through standard deviation of height values	8
N/A	Not interpretable		-	-	99

Tree species

From the classification of the SPOT V data we classified four classes of tree species mixtures and one non-forest class:

- pure conifer stand (>80% conifers)
- mixed conifer stand (50 – 80% conifers)
- mixed broadleaf stand (20 - <50% conifers)
- pure broadleaf stands (<20% conifers)

This grid was the basis for the stand wise species mixture calculation.

Timber volume and total living tree biomass:

The timber volume of the whole project area was known from a terrestrial sample plot inventory. This amount is then distributed to the stands according to each stand's integral of the VHM per species. This measure can be visualized in the form of a wave as shown in Figure 30. To create such a waveform, all laser points of a stand were accumulated according to their height above ground. Furthermore each data set was scaled by the stand area, thus making it possible to directly compare the waveform like distribution (see Figure 30). The variations between the vertical forest structures are caused by different age classes, species and growing conditions at each site. The sum beyond the waveform is calculated and used to divide the total timber volume to the stands. The hypothesis is that we are much closer to the real standing timber volume per stand than any stratum mean. This was confirmed by the client. The parameter was standing timber volume in m^3/ha and total m^3 for each stand. In the final step, standard procedures were used to calculate the living tree biomass in tonnes of CO2 and tonnes of carbon from the standing timber volume [IPCC 2006].

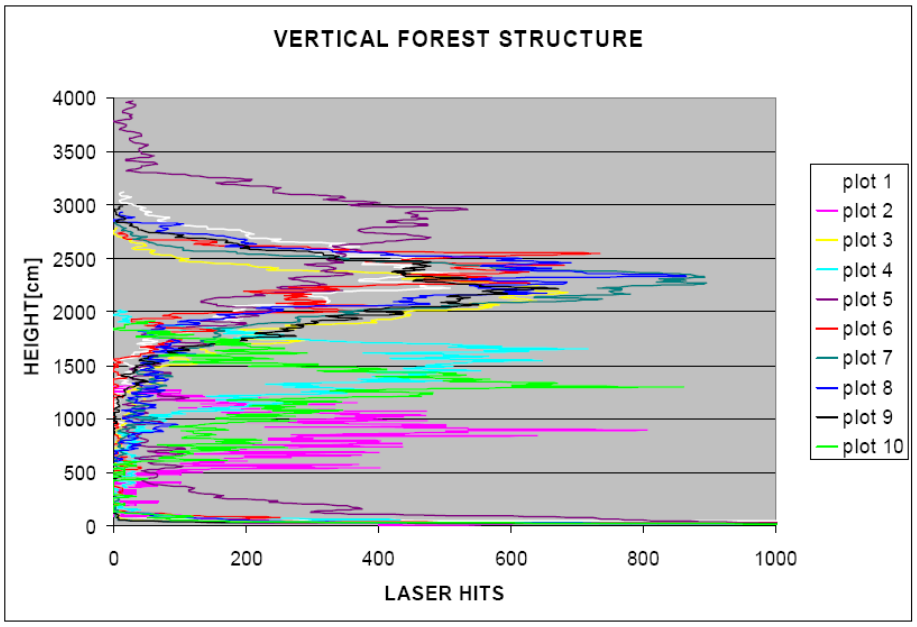


Figure 30: Vertical vegetation structure from LIDAR data (from [Wack, 2006])

3. Results

Result of the project is a digital forest stand map. This consists of the digital vegetation model and of a polygon coverage of the forest stands with attribute data. In addition there is a point coverage with the tree tops and a SPOT V classification of the main species combination. The processing has been done for the whole Canton of Appenzell - Ausserrhoden. The results are used by the forest administration of the Canton and thus undergo a usability assessment by practical usage. Due to missing ground truth data, no typical accuracy assessment could be performed. Accuracy assessments in previous projects under similar conditions showed the following results:

The quality of the pixel-based species classification of the four classes deciduous, coniferous and two mixed classes based on SPOT data has been around 85%. [Wack, 2006] showed a correlation coefficient between the calculated and field-measured timber volume of between 0.89 and 0.99 depending on the tree species.

An example of the final forest map including all calculated parameters is shown in Figure 31. A simple height class coded visualisation of the VHM (Figure 32) shows all trees with their specific height. Such maps are useful for further planning within the forest enterprises and forest administrations, such as for example to plan harvest operations or to find anchor tree candidates for cable way systems in the mountainous terrain.

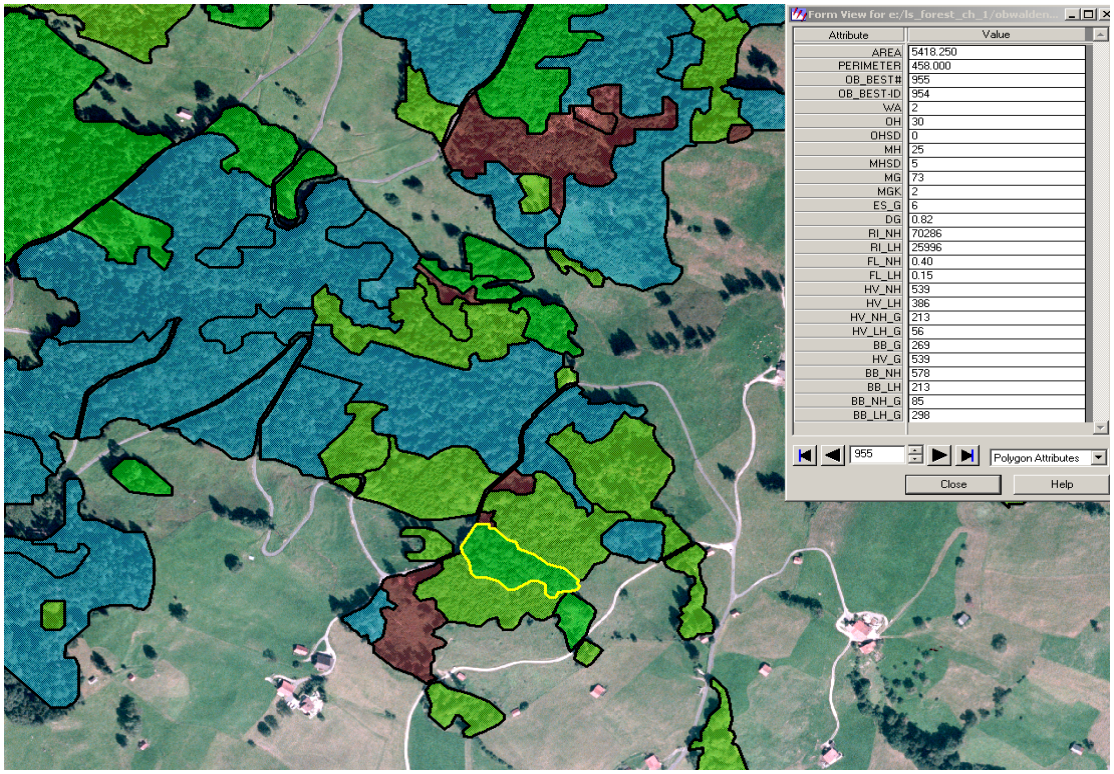


Figure 31: Example of final result a GIS forest stand map indicating stage of development with all parameters

Figure 32: VHM colour-coded with tree tops and stand borders. The numbers indicate standing timber volume per ha and in total for each stand.

4. Discussion and Conclusions

This project is a showcase for practical usability of remote sensing based assessments for forestry. LiDAR data is an important source of information especially for forest parameters like tree height, stem density and structure of the stand. Species determination is the main weak point of LiDAR data processing compared with aerial photo interpretation, thus the information is completed by using multispectral satellite imagery to classify species. Minimum information on the proportion of conifers and deciduous trees is crucial for the estimation of biomass, since different parameters apply for the two species groups. The procedure used is a combination of satellite imagery with all steps, of LiDAR data evaluation and orthophoto interpretation. The procedure is robust because it is focussed on parameters that can be derived with high accuracy under varying data situations and forest conditions. The project proofed that the procedure is fit for commercial use and that the technology transfer was successful.

In addition to the standard product of a digital forest stand map, other analysis can be done on request: forest canopy gap detection, critical gaps in canopy related with danger of avalanches, stability of forest stands against storm damages, forest road detection and other types of DTM evaluation.

Further developments of the standard product digital forest stand map will include:

- the improvements of the automatic stand segmentation procedure to be closer to the traditional forester's stand borders;
- the use of surface models from stereo-photogrammetry to replace the costly LiDAR campaign for frequent updates;
- the calculation of parameters giving an indication on the biomass usable for energy purposes, which is an important topic in the current EU-2020 (energy directive) discussion

References

[Andersen et al., 2003] Andersen, H.-E., Reutebuch, S. E., and Schreuder, G. F. (2003). Bayesian Object Recognition for the Analysis of Complex Forest Scenes in Airborne Laser Scanner Data. In Proceedings of ISPRS Workshop '3-d Reconstruction from Airborne Laserscanner and INSAR Data'.

[Badoux, 1983] Badoux, E. (1983). Ertragstafeln. Eidgenössische Anstalt für forstliches Versuchswesen. 3. Auflage.

[Barbati et al., 2009] Barbati, A., Chirici, G., Corona, P., Montaghi, A., and Travaglini, D. (2009). Area-based assessment of forest standing volume by field measurements and airborne laser scanner data. Int. Journal of Remote Sensing, 30(19):5177–5194.

[Chen et al., 2006] Chen, Q., Baldocchi, D., Gong, P., and Kelly, M. (2006). Isolating Individual Trees in a Savanna Woodland Using Small Footprint Lidar Data. Photogrammetric Engineering & Remote Sensing, 72(8):923–932.

[Donoghue et al., 2007] Donoghue, D., Watt, P., Cox, N., and Wilson, J. (2007). Remote Sensing of species mixtures in coniferous plantations using LiDAR height and intensity data. Remote Sensing of Environment, 110:509–522.

[Gonzalez and Woods, 2002] Gonzalez, R. C. and Woods, R. E. (2002). Digital Image Processing. Prentice Hall, Inc., Upper Saddle River, New Jersey, second edition. 793 p.

[Granica et al., 2004] Granica, K., Schardt, M., Hirschmugl, M., and Gallaun, H. (2004). The Observation of Protection Forests in Critical Zones Using New Satellite Technology. In Proceedings of INTERPRAEVENT 2004, volume V, pages 37–48.

[Hirschmugl, 2008] Hirschmugl, M. (2008). Derivation of Forest Parameters from UltracamD Data. PhD thesis, Graz University of Technology. 220 p.

[IPCC 2006] 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Prepared by the National Greenhouse Gas Inventories Programme, Eggleston H.S., Buendia L., Miwa K., Ngara T. and Tanabe K. (eds). Published: IGES, Japan.

[Hollaus et al., 2009] Hollaus, M., Dorigo, W., Wagner, W., Schadauer, K., Hoefle, B., and Maier, B. (2009). Operational wide-area stem volume estimation based on airborne laser scanning and national forest inventory data. Int. Journal of Remote Sensing, 30(19):5159–5175.

[Hyyppä and Hyyppä, 1999] Hyyppä, H. and Hyyppä, J. (1999). Comparing the accuracy of laser scanner with other optical remote sensing data sources for stand attributes retrieval. The photogrammetric journal of Finland, 16:5–15.

[Hyyppä et al., 2000] Hyyppä, J., Hyyppä, H., Inkinen, M., Engdahl, M., Linko, S., and Zhu, Y.-H. (2000). Accuracy comparison of various remote sensing data sources in the retrieval of forest stand attributes. Forest Ecology and Management, 128:109–120.

[Hyyppä et al., 2002] Hyyppä, J., Schardt, M., and Koch, B. (2002). Highscan: New Methods in Forest Inventory and Mapping. Geoinformatics, 5.

[Koch et al., 2006] Koch, B., Heyder, U., and Weinacker, H. (2006). Detection of Individual Tree Crowns in Airborne Lidar Data. Photogrammetric Engineering & Remote Sensing, 72(4):357–363.

[Koch et al., 2009] Koch, B., Straub, C., Dees, M., Wang, Y., and Weinacker, H. (2009). Airborne laser data for stand delineation and information extraction. International Journal of Remote Sensing, 30:935–963.

[Næsset, 2002] Næsset, E. (2002). Predicting Forest Stand Characteristics with Airborne Scanning Laser Using a Practical Two-Stage Procedure and Field Data. Remote Sensing of Environment, 80:88–99.

[Pitkänen et al., 2004] Pitkänen, J., Maltamo, M., Hyyppä, J., and Yu, X. (2004). Adaptive Methods for Individual Tree Detection on Airborne Laser Based Canopy Height Model. In Thies, M., Koch, B., Spiecker, H., and Weinacker, H., editors, International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences. Proceedings of the ISPRS working group VIII/2: Laser-Scanners for Forest and Landscape Assessment, volume XXXVI, pages pp. 187–191. Institute of Forest Growth, Department of Remote Sensing and Landscape Information Systems, University of Freiburg.

[Schmidtke et al., 2008] Schmidtke, H., Götz, M., Wack, R., Schardt, M. (2008). Halbautomatisiertes Verfahren zur Erstellung von Waldbestandeskarten mit Luftbildern und Oberflächenmodellen. Forschungsprojekt. Projektbericht. Winterthur und Graz, November 2008. http://www.bafu.admin.ch/wald/01234/01238/08413/index.html?lang=de#sprungmarke0_18. 15.07.2010

[Wack, 2006] Wack, R. (2006). Combined use of satellite imagery and laserscanner data for the assessment of forest stand parameters. In Proceedings of Workshop on 3D Remote Sensing in Forestry, Vienna.

[Wack et al., 2003] Wack, R., Schardt, M., Barrucho, L., Lohr, U., and Oliveira, T. (2003). Forest inventory for eucalyptus plantations based on airborne laserscanner data. In Proceedings of ISPRS Workshop on Laserscanning. available at: http://www.isprs.org/commission3/wg3/workshop_laserscanning/papers/Wack_ALSDD2003.pdf; accessed Jan. 2008.

[Wack and Stelzl, 2005] Wack, R. and Stelzl, H. (2005). Assessment of forest stand parameters from laserscanner data in mixed forests. In Proceedings of ForestSat 2005, pages 56–60, Borås.

[Wack and Wimmer, 2002]Wack, R. and Wimmer, A. (2002). Digital Terrain Models from Airborne Laserscanner Data - a Grid based approach. In Proceedings of the ISPRS Commission III Symposium, pages 293–296, Graz.

[Ziegler et al., 2002] Ziegler, M., Granica, K., Schmitt, U., Gallaun, H., and Schardt, M. (2002). Satellite Remote Sensing and Laserscanning - Forest Inventory and Monitoring in the Alps. Geoinformatics, 5:28–31.

The intensity return behavior of a high resolution terrestrial 3D laser scanner for foliage structure measurements in pear orchards

Mathilde A. F. Balduzzi1, Dimitry Van der Zande1,2, Willem W. Verstraeten1 & Pol Coppin1

mathilde.balduzzi@biw.kuleuven.be

1 Katholieke Universiteit Leuven, Biosystem Department, M3-BIORES

Willem de Croylaan 34, BE-3001 Leuven, Belgium,

2 Management Unit of the North Sea Mathematical Model (MUMM), Department VI of the Royal Belgium Institute of Natural Science

Abstract

Light Detection and Ranging (LiDAR) technology can be used to describe and quantify vegetation structure by measuring the return time and intensity of light beams. As such, this technology may help to unravel the impact of the contribution of woody and foliar structures to the light intercepted by a spectral sensor.

In this study the LiDAR FARO LS880 is used. This device also measures an Intensity Return (Int-Re). A preliminary study is conducted to test how Int-Re can be used to determine the beam incidence angle to the leaf. First, the distance effect was corrected. Then 10 leaves of the Conference Pear (*Pyrus communis*) are scanned on a goniometric platform to provide the averaged Int-Re from a scatter plot selection on those leaves: this aims to determine the relationship between Int-Re and beam incidence angle. Finally, this relationship is tested on a selection of leaves that still are attached on the tree. The results show that the incidence angle versus Int-Re law can help to determine the beam incidence angle for pear trees with a precision of 10° for an incidence angle >45°, whereas the average Int-Re for an incidence angle <45° is more variable: in this case, the undulation effect on the averaged Int-Re appears to be the reason of such a variability.

1. Introduction

Leaf inclination (elevation, roll and azimuth) gives precious information on the photosynthesis process as they are directly linked. Leaf movements have two principal effects on the environmental aspect of photosynthesis: first, they provide a mechanism whereby the plant is able to achieve favorable photosynthesis rates at specific times during the day. Secondly, they allow the leaf to avoid high incidence photon irradiance at the times of the day that are not favorable for photosynthesis (Ehleringer & Forseth, 1989).

Traditionally, leaf elevation is determined directly with a protractor (Takenaka *et al.* 1998, Norman & Campbell, 1989) or with the angle determined by point's positions given by an electromagnetic digitizer (Thanisawanyangkura *et al.* 1997). However those two methods are time-consuming and labor intensive.

A more general index that describes leaf inclination is the leaf angle distribution (LAD). This index is an essential parameter to characterize canopy structure and plays a crucial role in the simulation of radiation transmission (Wang *et al.* 2007). In such studies, canopies could be represented by two kinds of objects: either a turbid medium or discrete scatterers defined by a Leaf Normal Distribution Function or LNDF (Otto & Trautmann, 2008, Chelle, 2006). However, in the case of a complete simulation of radiative transfer on object like trees, a complete description of the tree would be more relevant because the leaf orientation is not random in most cases, but is directly linked to the position of the leaf in the tree (Stuckens *et al.* 2009, Biswas, 3007).

Another important index for vegetation is the Leaf Area Index or LAI which is used in any flux transfer study as gases exchange e.g. CO2 (Arkebauer *et al.* 2009) or radiation (de Castro & Fetcher, 1998). In regard to the radiation interception, it is defined as the total one sided leaf area per unit ground surface area (Montheit & Unsworth, 1973). However, the research of Chen and Black (1992) noticed the problem within the LAI definition of the one-side area. It proposes a new definition of LAI that takes into account curvature and undulation.

TLiDAR (Terrestrial – Light Detection And Ranging) is an optical remote sensing technology that measures properties of scattered light to find the distance of a target. This principle uses Time Of Flight (TOF), phase shift or frequency modulated continuous wave technology (Nejad & Olyaee, 2006). Combined with one or two axial rotations (azimuth and elevation), TLiDAR provides a 3D picture of the scene surrounding it. This picture is a 3D (x,y,z) Cartesian scatter plot of scanned points, but also could be conceived as a 2D (elevation, azimuth) spherical view of scanned points with their distance values, d. Tanks to such information, TLiDAR became a very common tool in archeology, architecture and topography (eg. 3D-RiskMapping, 2006-2008), but also in agriculture and forestry. In forestry, TLiDAR has been used to determine forest metrics such as the diameter at breast height (dbh), tree height, stem density and volume estimation (Hopkinson *et al*. 2004), gap fraction (Danson *et al.* 2007) and Leaf Area Index (Zheng *et al.* 2009). In agriculture, this device has been used to estimate the vegetative volume and surface area (Rossel *et al.* 2008, Rossel *et al.* 2009).

To get an accurate geometric description of the position or angle, the number and density of points will be determinant and for scanned tree. Many problems may occur, especially for foliage scanning. For instance, the shadow effect (van der Zande *et al.* 2008) could give a lack of information on leaves. Instead of having the complete leaves, those leaves have a few points that may avoid any angle analysis. In addition to that, the lack of information due to the quality of the scanned scene caused by wind and the variable accuracy of the scatter plot due to object reflectance, device calibration, range (Kersten *et al.* 2005) and incidence angle (Soudarissanane *et al.* 2009, Linderbergh *et al.* 2005), could provide scans of low quality. In addition, the fact that lasers are spherical range finders means that the distance between two points on a flat surface will increase with the range to the beam aperture (van der Zande *et al.* 2006) determines also the quantity and quality of the scan. Finally, light ambiance for large distance will have an impact on the quality of the data as well (Voegtle *et al.* 2008).

In recent TLiDAR scanners the intensity return (Int-Re), which is the ratio between the outcome and income of emitted light, is an extra value provided by the device. This Int-Re is influenced by the range (Kaasalainen *et al.* 2008), material (Voegtle *et al.* 2008) and incidence angle with the material (Voegtle *et al.* 2009). Avoiding the range for a specific material and making an analysis of the relationship between Int-Re and the beam incidence angle with the material, would give extra information to determine the leaf angle. However, if the beam angle is not in line with the NADIR orientation, this Int-Re/ incidence angle relationship would be relative to the TLiDAR scanners position and to get the absolute leaf angle, extra measurement or information is needed. (see figure 1.)

In this case of curved and undulating leaves, the combination of distance measurement and the Int-Re could help to determine a complete mapping of the leaf angles: as the undulation/curvature range has in general a small diameter, the linear regression problem will be the same as in a case of a lack of data. At this point, flux transfer studies have a powerful understanding of flux behavior within tree canopies: working with reconstructed virtual trees (Coté *et al.* 2009), accurate description of leaf curvature and undulation would give better simulations for flux transfer studies.

The Int-Re is affected by the distance and reflectance of the scanned material combined with the actual incidence angle with the laser beam. Those variations can be corrected (Kaasalainen *et al.* 2009). Thus, to get comparable data, it is necessary to avoid the distance effect on the Int-Re as a first step and then to compare the Int-Re given by the TLiDAR for different incidence angles of the leaf material.

The Int-Re is affected by the distance and reflectance of the scanned material combined with the actual incidence angle with the laser beam. Those variations can be corrected (Kaasalainen *et al.* 2009). Thus, to get comparable data, it is necessary to avoid the distance effect on the Int-Re as a first step and then to compare the Int-Re given by the TLiDAR for different incidence angles of the leaf material.

In addition, new opportunities in the use of Int-Re to detect the leaf physiological aspect like the chlorophyll content with a LEICA ScanStation2 (532nm) (Eitel *et al.* 2010) or the water content as moisture have an impact on FARO LS880HE80 laser Int-Re (785nm) (Kaasalainen *et al.* 2010).

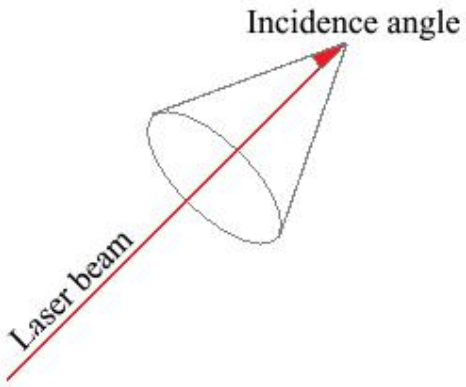


Figure 33. *The incidence angle will provide a set of isotropic objects that have this incidence angle with the beam: this set could be represented as a cone. This incidence angle coincide with leaf elevation at NADIR.*

2. Materials and method

TLiDAR and Int-Re parameters

The TLiDAR FARO LS880 was used in this study: it is a panoramic scanner which combines horizontal and vertical rotations created by the rotation of a mirror placed at 45° to the laser beam (horizontal rotation) and by the rotation of the trunnion of the TLiDAR (vertical rotation). This device employs a Class 3R of continuous wave (CW) semiconductor laser operating at λ =785 nm (near-Infrared) and with a maximal range of 76 m. The amplitude modulated technique related to CW aims to measure ranges: amplitude of the laser is modulated and an analysis of the frequencies of the signal provides the distance. The beam power is 20 mW, it has a diameter (circular) of 3 mm and its divergence is 0,014°. The sensor field of view is approximately of 3 mrad. Between the mirror and the photodiode of the scanner, optical elements deserve to reduce the light for small distances and in addition to that, the electric-converted signal passes through a logarithmic converter that gives the logarithmic relationship between different reflectance. Thus the relationship between Int-Re and distance won't follow the inverse square law or any linear function and small differences between scanners in the Int-Re behavior through range would appear. In conclusion the Int-Re and distance relationship needs to be studied. Light ambiance has few impacts on Int-Re: indeed a light signal is added (in term of union) to the laser's signal that doesn't fade out the presence of frequencies that would be in the signal without other light sources. However, there is more noise with increasing range and sometimes even no data at all, especially for low Int-Re. Thanks to FARO Company, the impact of the sun on the quantity of data is perceptible for a range higher than 10 meters and/or low reflectance object.

This Int-Re is influenced by the range (Kaasalainen *et al.* 2008), the material (Voegtle *et al.* 2008) and the angle (Voegtle *et al.* 2009) which could be consider as the expression of the brightness of the material. Indeed, the Int-Re/incidence angle relationship study is in some way finding the bidirectional reflectance factor with the same view and illumination direction as defined in Schaepman-Strub *et al.* (2006) for a variable incidence angle. The Int-Re is not influenced by the light ambiance in our case because the TLiDAR FARO LS 880 used in this study is an active device using phase shift technology, and according to the constructor, the Int-Re and distance data provided by the TLiDAR are not influenced by the light ambiance at range <10 m.

The TLiDAR data could be viewed as a 3D scatter plot in a Cartesian XYZ-basis or a 2D matrix with (i,j) = (elevation, azimuth) as indices (spherical view) and a range value. Each point has an extra value that is the

Int-Re with dimensionless value (from 0 to 2047). The scans are proceeded with an angle resolution of 0.018° for both azimuthal and elevation rotation.

In Kaasalainen *et al.* (2008), the Int-Re relationship with range for the FARO LS80, which is a similar device like the FARO LS880, present a shift between two Int-Re of different reflectance value at a fixed distance. However, because each device has a specific behavior for its Int-Re at small range due to a physical filter in front of the sensor, this relationship should be studied for each particular device. Thus, to make this study, a preliminary distance effect study should be done. As summed up in the flow chart below (figure 2.), a first study to set the relationship between Int-Re and distance is proceeded.

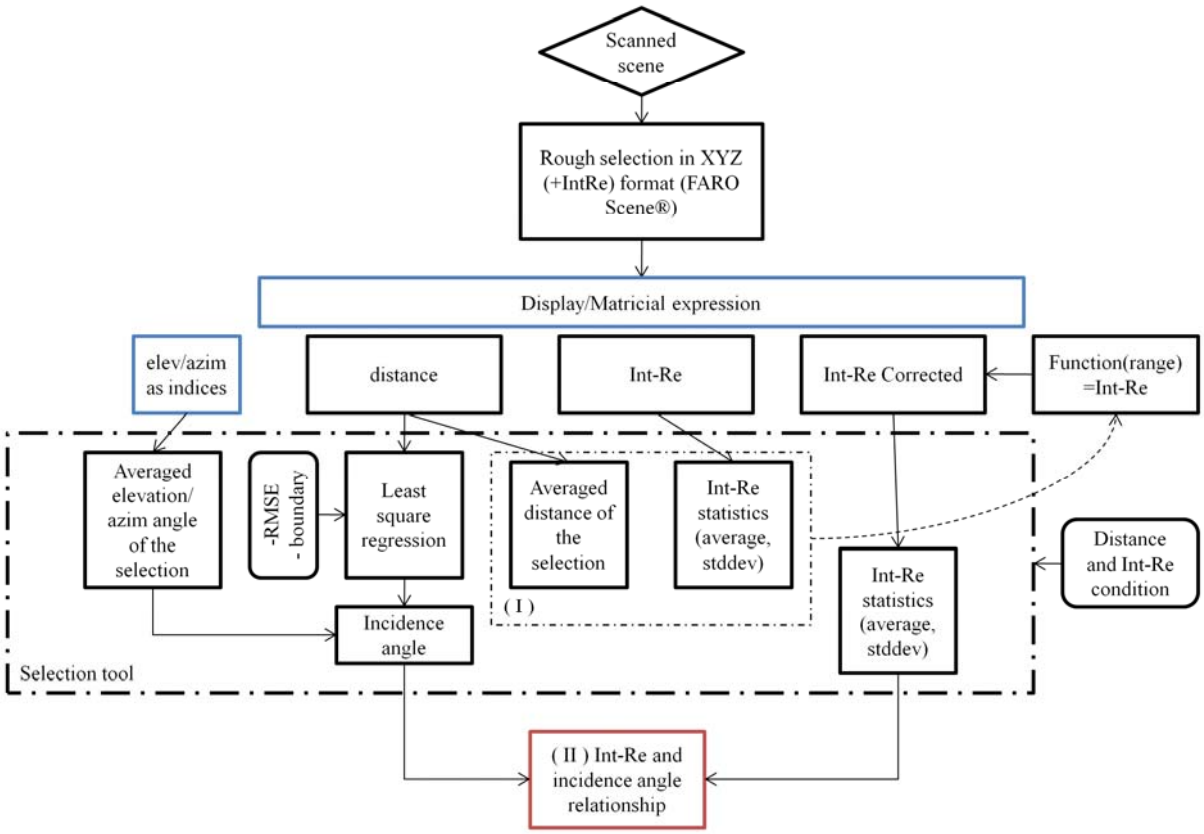


Figure 34. *Selection tool and analysis flowchart: (I) The Int-Re/distance relationship is determined by a polynomial regression of a graph with different (distance; Int-Re) points. This relationship is determined with different mate Canson® paper and a 99%-Spectralon®; (II) Once the Int-Re is corrected, each selection on leaf materials with a different beam incidence angle provides an averaged Int-Re with its standard deviation.*

Leaf angle relationship with Intensity Return

After testing the accuracy of the incidence angle with the goniometric platform, the incidence angle between the beam and the leaf will be compared to the averaged Int-Re on the leaves to deduce an angle and Int-Re relationship for pear tree leaves.

Those pear tree leaves are picked up from 24 Conference Pear trees of 2 years old in pot of approximately 40 000 cm³ placed in 2 rows of 12 trees in an East-West direction with a distance of 30 cm between the trees and 360cm between the rows. The soil mixture has not been specified by the supplier.

Pear tree leaves are curled (Baugher, 2004), so before the scanning procedure, they are cut in two parts along their central vein to be flattened as much as possible. They are fixed (back and front side) on the goniometric platform with black painted strings and are scanned within the hour they were picked up from the tree. The goniometric platform is placed in a position such that the center of its platform is facing the beam at φ,θ=0. An incrementation of 20° around vertical and then horizontal axis is applied for each scan to get incidence angles between 90° and 10°.

After the data have been corrected, many scans for those 10 pear trees leaves are proceed. For the selection of the leave material, the Int-Re boundary is fixed at 80° to avoid a selection of the goniometric platform, undulations that are still present and border effect of the leaf (Eitel *et al.* 2010). The Int-Re and incidence angle ω graph will give the relationship between both of them. A comparison between an azimuthal and an elevation rotation will be done. To get a reference relationship between the incidence angle and Int-Re, the incidence angles are rounded to get a resolution of 5°. The overall average (the average of each averaged Int-Re) for each 5° is taken as reference. The maximum Int-Re standard deviation at each 5° is also taken for the validation analysis as a condition of selection.

The two trees in the middle (n° 5 and 6) are taken as validation trees: this relationship will be tested by selecting many samples from leaves in the scene in respect to the maximum Int-Re standard deviation we found. The Int-Re and distance bound are changed thanks to a rough estimation of the incidence angle. For small incidence angles, a Int-Re limit of 100 with a distance boundary of 0.02 m are used, whereas for big incidence angles, those limitation are the same as in the case where leaves are flattened.

3. Results

Distance correction

Between each graph for each reflectance, a constant shift appears as in Kaasalainen *et al.* (2008). The correction with the n-polynomial avoids the distance effect for range >1m. Indeed, the correlation between the n-polynomial regression of the 99%-spectralon graph, used as reference (raw Int-Re corrected data: 1781), and the other graph gives a r^2 of 1 for reflectance >48% (raw Int-Re corrected data: 1619) and of 0.98 +/- 0.01 for reflectance below 22 % (raw Int-Re corrected data: 1408). The slopes of the regression of the correlation graph equals 1 +/- 0.3 for reflectance > 48 % with intercept < 33 (raw data). For reflectance <22 %, the slope of the correlation graph regression is bellow 0.85 with an intercept > 150.

Leaf Int-Re/angle relationship

The incidence angle provided by the least square regression gives a very good result on the goniometric platform. The correlation graph between the beam incidence angle on the goniometric platform manually calculated and the one given by the least square regression provide a regression with a r^2 of 1, a slope of 1 and an intercept of 1° for azimuthal angle and 2.8° for elevation angle.

For pear tree leaf material, in a first phase the variability on Int-Re was tested with two scans of the leaves at the same goniometric position. No difference appears between those two scans of one hour of time difference (graph not shown). The analysis also shows that there were no clear differences between the azimuthal or elevation rotation of the leaves that means that those type of leaves have an isotropic behavior for this experimental set up and at this laser wavelength (785nm). In addition, the differences between back and front leaves Int-Re are also not perceptible in those conditions. Figure 3. summarizes all the averaged Int-Re of a leaf selection from the scanned scene. There is a clear increase of the Int-Re with the incidence angle, and we could expect to see this with this method, an angle of +/-10°. For an incidence angle >45°, the averaged Int-Re are within d, the averaged standard deviation whereas the dispersion of the averaged Int-Re is greater for the incidence angle <45°.

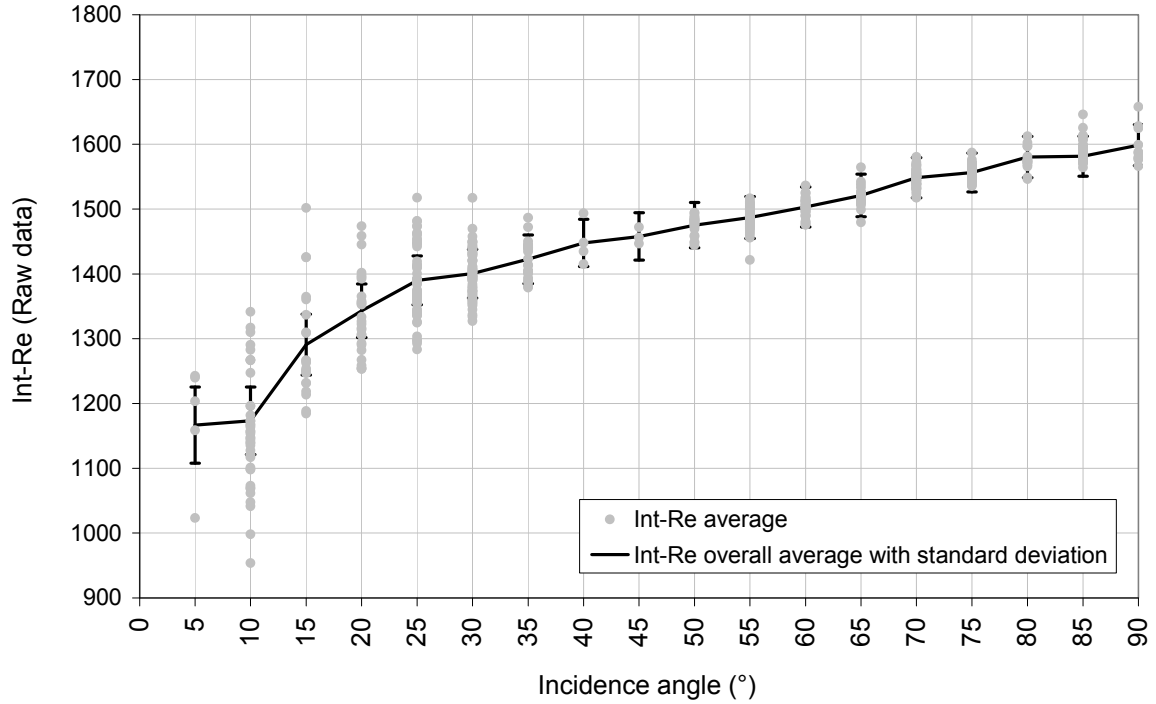


Figure 35. *Int-Re and incidence angle relationship with (1) averaged Int-Re measurement for each 5°-angle and (2) Int-Re overall average at each 5°-angle with the averaged standard deviation (on the scatter plot selection) for each 5°-angle.*

Validation of the Int-Re and angle relationship

The validation shows good result for angle ≥ 45° (Figures 4. and 5.). However, it gets a lot variability for an angle below 45°. With this method we could expect to have an accuracy of +/- 10° for angle ≥45°. The bad correlation for angle < 45° may be due to undulation, curvature and quality of the least square regression that is influenced by the noise and ghost point that are not eliminated by the selection and more present in such cases (Table 4.). The low quality of the result for an incidence angle < 45° could be explained by the presence in the selection, of a part of the leaves that are curveted (Figure 6.): those are more difficult to avoid for such angles as in this case, the problem of number of points occurs. The selection tool should have a more large selection scale for Int-Re and distance to get enough points for the least square regression.

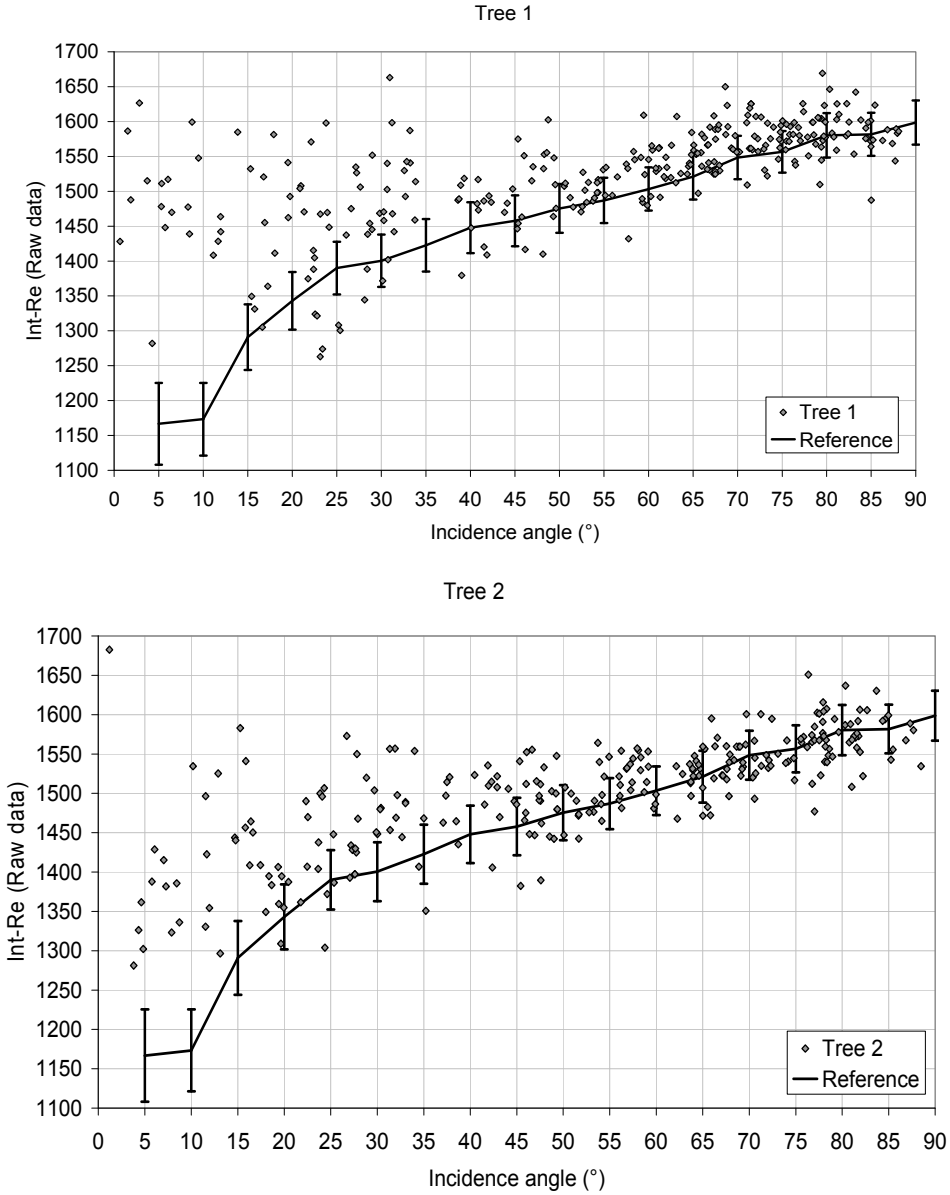


Figure 36 and 5. *Validation of the Int-Re/incidence angle relationship for two pear trees. The bold line represent the reference curve as the average value of the Int-Re at each 5°-angle with their standard deviation.*

Table 4. *Statistical result for the validation case for two conference pear trees with case of an incident angle smaller than 45° and a case with an incidence angle bigger than 45.*

	Slope	Intercept	r^2	Average residuals	RSME
Tree 1 (<45°)	-0.01	50.87	0.00	-26.23	23.16
Tree 1 (≥45°)	0.83	16.15	0.47	-4.71	10.08
Tree 2 (<45°)	0.69	25.98	0.17	-18.40	17.45
Tree 2 (≥45°)	0.77	17.04	0.49	-1.74	9.84

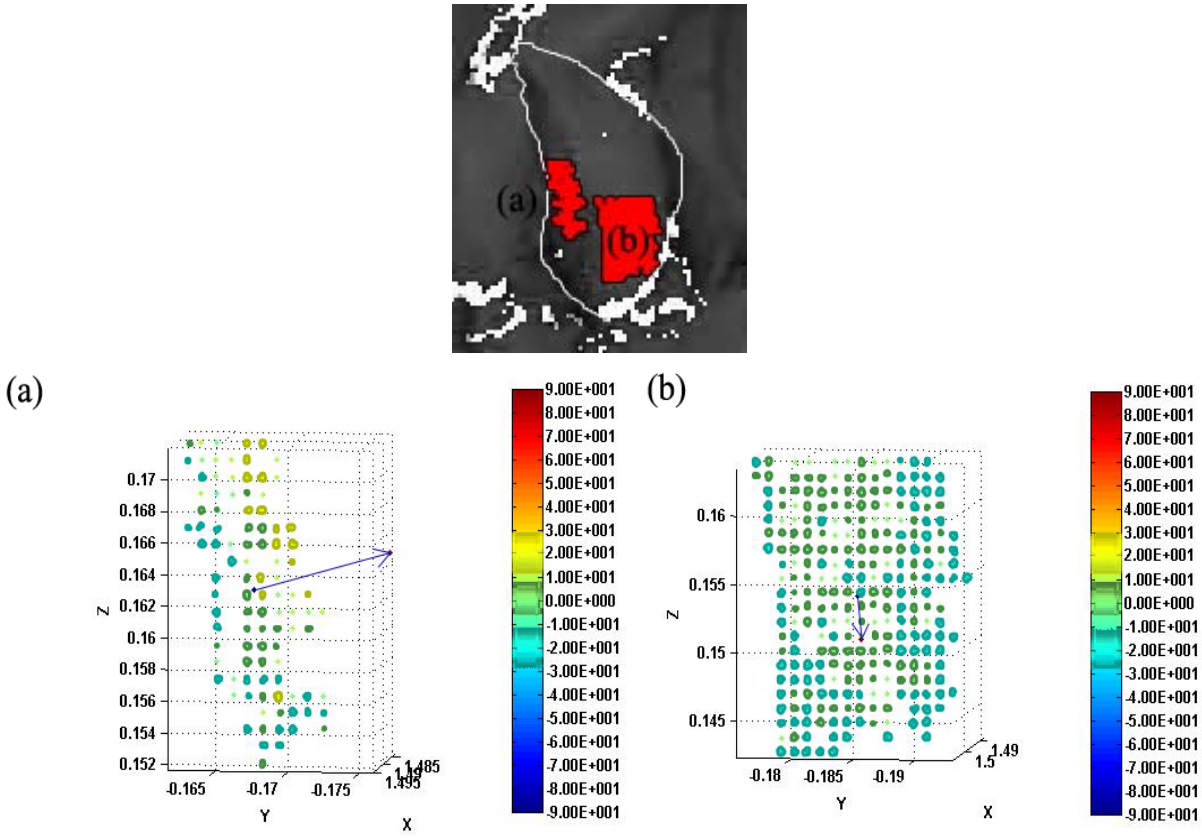


Figure 6. *Comparison between two partial leaf selections with their normal vector found by the least square regression. Color represents the difference between the Int-Re angle for each point of the selection and the Int-Re angle on the average Int-Re. A negative value would mean that the Int-Re angle from the averged Int-Re is higher than the Int-Re angle of the point. (a) A low incidence angle with a geometric incidence angle of 17.8° and an Int-Re angle of 40°. In this case, the spread of the Int-Re angle difference goes from -20° to +20° (+/- 5°). The curvature of the leave is clearly present. (b) A high incidence angle with a geometric angle of 72.8° and a Int-Re angle of 85°. In this case, most of the points have an Int-Re angle difference of 0° or -20° (+/- 5°)*

4. Discussion

As in Soudarissanane *et al.* (2009), the incidence angle decreases the accuracy of the scatter plot. Combined with the density of the scatter plot at this incidence angle (Lindenbergh *et al.* 2009) the least square regression may present some problems the correct representation of the incidence angle.

In addition to that, the distance correction that gives an Int-Re below 1619 will present errors. Kaasalainen *et al.* (2009) suggests in that case that a repeatable method of calibration should be set up. This error may have an impact on the validation because the leaf samples were not at the same range, whereas the case with the goniometric platform.

For practical matter, this could be reduced with the measurement set-up (Van der Zande *et al.* 2006): indeed, finding a measurement geometry that increases the number of leaves that face the beam reduces then the possibility (low incidence angle) that the Int-Re couldn't provide any results.

The leaf undulations and curvature may also have a great impact on the quality of the validation in the scope of this research as it has been shown in point 3.3. Further research should be done in this scope to know the quality of the information that can be provided by the Int-Re.

5. Conclusions

Finding angles with Int-Re without taking into account undulation and curving of the leaves is very difficult. To get better validation, this study should be done with species with more flat and uniform leaves. In general,

the linear regression on flat material is enough to get the incidence angle with the beam. To improve this method, a more complete probabilistic analysis could be done: instead of having a simple Int-Re and incidence angle relationship, the probability to get a specific angle with a given Int-Re could be given. This study would imply to determine which distribution of Int-Re appears for each angle. A simple assumption would be to use a normal distribution with the overall average and standard deviation.

Though we can see a future for the use of Int-Re in the study of those two parameters: indeed, using the incidence angle between the laser beam and the linear regression of curveted/undulating leaves and comparing it with the different angle due to those undulation/curvature of the leaves would give an estimation of them. This method would be more efficient in terms of computer processing because it would be a matter of simple trigonometric computing instead of a heavy non-linear fitting algorithm. In addition, some undulations that are present on leaves could be detected by a change in the Int-Re whereas it can't be with a simple geometric approach. Even the incidence angle has a visible variation, the surface of this undulation can be not big enough, or within the range resolution of the laser to be seen in the XYZ scatter plot.

With other leaves and/or laser wavelength, some study could be done on the spread of diseases or other stress in the tree.

In general, this method could be used for scanned trees with low quality due to the distance or with scans where data are missing, e.g. due to wind. However, this method gives an incidence angle with the beam which is not the elevation. One of the most obvious cases where we can easily find the elevation angle is when the laser beam is at nadir: in this case, the elevation angle is the same as the incidence angle.

Acknowledgement

We would like to acknowledge the useful technical information about the FARO LS880 provided by FARO, Stuttgart, Germany. A particular thank to Jo Janssen for his precious help in C# development and to Jan Stuckens for his advices. Funding support for this project has been provided by FWO Flanders.

Bibliography

Arkebauer, T.J., Walter-Shea, E.A., Mesarch, M.A., Suyker, A.E., Verma, S.B., 2009, Scaling up of CO2 fluxes from leaf to canopy in maize-based agroecosystems, *Agricultural and Forest Meteorology*, 149, pp 2110-2119.

Baugher, T.A., 2004, Anatomy and Taxonomy, Chap 1. In: *Concise Encyclopedia of Template Tree Fruit*, Eds: Baugher, T.A., Singha, S., Publisher: Food Product Press, The Haworth Press, New-York, pp 3-10.

Biswas, G.kr, 2007, Some new leaf normal-angle distribution models and their influence on geometry functions and area scattering phase functions related to radiative transfer problems in vegetative canopies, *Journal of Quantitative Spectroscopy & Radiative Transfer*, 108, pp 197-219.

Chelle, M., 2006, Could plant leaves be treated as Lambertian surfaces in dense crop canopies to estimate light absorption?, *Ecological Modelling*, 198, pp 219-228.

Chen, J.M., Black, T.A., 1992, Defining leaf area index for non-flat leaves, *Plant, Cell and Environment*, 15, pp 421-429.

Côté, J.F., Widlowski, J.L., Fournier, R.A., Verstraete, M.M., 2009, The structural and radiative consistency of three-dimensional tree reconstructions from terrestrial LiDAR, *Remote sensing of environment,* Vol.113, Issue 5, pp 1067-1081.

De Castro, F., Fetcher, N., 1998, Three dimensional model of the interception of light by a canopy, *Agricultural and Forest Meteorology*, 90, pp 215-233.

Danson, F.M., Hetherington, D., Morsdorf, F., Koetz, B., Allgöwer, B., 2007, Forest canopy gap fraction from terrestrial laser scanning, *IEEE geoscience and remote sensing letters*, vol.4, n°1, pp 157-160.

Eitel, J.U.H., Vierling, L.A., Long, D.S., 2010, Simultaneous measurements of plant strucutre and chlorophyll content in broadleaf sapling with a terrestrial laser scanner, *Remote sensing of environment.*

Ehelringer, J.R., Forseth, I.N., 1989, Diurnal leaf movement and productivity in canopies. In: *Plant Canopies: their Growth, Form and Function*; G. Russel, B. Marshall and P.G. Parvis, Eds; Publisher: Cambridge University Press, Cambridge, pp 129-142.

Hopkinson, C., Chasmer, L., Young-Pow, C., Treitz, P., 2004, Assessing forest metrics with a ground-based scanning LiDAR, *Canadian Journal of Forest Research*, 34, pp 573-583.

Kaasalainen, S., Kukko, A., Lindroos, T., Litkey, P., Kaartinen, H., Hyyppä, J., Ahokas, E., 2008, Brightness measurements and calibration with airborne and terrestrial laser scanners, *IEEE transaction on geoscience and remote sensing*, vol.46, N°2, February 2008, pp528-534.

Kaasalainen, S., Krooks, A., Kukko, A., Kaartinen, H., 2009, Radiometric calibration of terrestrial laser scanners with reference targets, *Remote Sensing*, 1, pp144-158.

Kaasalainen, S., Niittymäki, H., Krooks, A., Koch, K., Kaartinen, H., Vain, A., Hyypä, H., 2010, Effect of target moisture on laser scanner intensity, *IEEE transaction on geoscience and remote sensing*, vol.48, N°4, April 2010, pp 2128-2136.

Kersten, T.P., Sternberg, H., Mechelke, K., 2005, Investigation into the accuracy behavior of the terrestrial laser scanning system MENSI GS100, *Optical 3D Measurement Techniques VII*, Gruen/Kahmen (Eds.), Vienna 2005, Vol.I, pp 122-131.

Linderbergh, R., Pfeifer, N., Rabbani, T., 2005, Accuracy analysis of the LEICA HDS 3000 and feasibility in tunnel deformation monitoring. In: *IAPRS (ed.) Proc. In the ISPRS Workshop, Laser Scanning 2005, Vol. XXXVI(3/W3)*, Enschede, The Netherlands, pp 24-29.

Monteith, J.L., Unsworth, M.H., 1973, *Principles of Environmental Physics*, 2nd Edn. Edward Arnold, London.

Nejad, S.M, Olyaee S., 2006, Comparison of TOF, FMCW and Phase-Shift Laser Range Finding Methods by Simulation and Measurement., *Quatarly Journal of Technology and Education***,** 1, pp 11-18.

Norman, J.M., Campbell, G.S., 1989, Canopy Structure. In: *Plant Physiology Ecology*; R.W. Pearcy, J.R. Ehleringer, H.A. Mooney & P.W. Rundell, Eds; Publisher: Chapman & Hall, London, pp 301-325.

Otto, S., Trautmann, T., 2008, A note on G-Functions within the scope of radiative transfer in turbid vegetation media., *Journal of Quantitative Spectroscopy & Radiative Transfer***,** 109, pp 2813-2819.

Rossel, J.R., Sanz, R., Llorenz, J., Arno, J., Escola, A., Ribes-Dasi, M., Masip, J., Camp, F., Gracia, F., Solanelles, F., Palleja, T., Val, L., Planas, S., Gil, E., Palacin, J., 2008, A tractor-mounted scanning LiDAR for the non-destructive measurement of vegetative volume and surface area of tree-row plantations : a comparison with conventional destructive measurements, *Biosystems engineering*, pp I-7.

Rossel, J.R., Llorens, J., Sanz, R., Arno, J., Ribes-Dasi M., Masip, J., Escola, A., Camp, F., Solanelles, F., Gracia, F., Gil, E., Val, L., Planas, S., Palacin, J., 2009, Obtaining the three-dimensional structure of tree orchards from remote 2D terrestrial LiDAR scanning, *Agricultural and Forest Meteorology*, 149, pp 1505-1515.

Schaepman-Strub, G., Schaepman, M.E., Painter, T.H., Dangel, S., Martonchik, J.V., 2006, Reflectance quantities in optical remote sensing – definitions and case studies, *Remote sensing of Environment*, 103, pp27-42.

Soudarissanane, S., Lindenbergh, R., Menenti, M., &Teunissen, P., 2009, Incidence angle influence on the quality of terrestrial laser scanning points. In: *Proceedings ISPRS Workshop Laser Scanning, September 1-2, Volume XXXVIII, Part 3/W8*, Paris, France, pp 183- 188.

Stuckens, J., Somers, B., Delalieux, S., Verstraeten, W.W., Copin, P., 2009, The impact of common assumptions on canopy radiative transfer simulations: A case study in Citrus orchards, *Journal of Quantitative Spectroscopy & Radiative Transfer,* 110, pp 1-21.

Takenaka, A., Inui, Y., Osawa, A., 1998, Measurement of three-dimensional structure of plants with a simple device and estimation of light capture of individual leaves. *Functional Ecology* **,** 12, pp 159-165

Thanisawanyangkura, S., Sinoquet, H., Rivet, P., Cretenet, M. , Jallas, E., 1997, Leaf Orientation and sunlight leaf area distribution in cotton. *Agricultural and Forest Meteorology* , 86, pp 1-15.

Van Der Zande, D., Hoet, W., Jonckheere, I., van Aardt, J., Coppin, P., 2006, Influence of measurement set-up of ground-based LiDAR for derivation of tree structure, *Agricultural and forest meteorology*, 141, pp 147-160.

Van Der Zande, D., Jonckheere, I., Stuckens, J., Verstraeten, W., Coppin, P., 2008, Sampling design of ground-based LiDAR measurements of forest canopy structure and its effect on shadowing. *Canadian journal of remote sensing*, 34 (6), pp 526-538

Voegtle, T., Schwab, I., Landes, T., 2008, Influences of different materials on the measurements of a terrestrial laser scanner (TLS), *The International archives of the photogrammetry, remote sensing and spatial information sciences*, Vol.XXXVII. Part B5, Beijing, pp 1061-1066.

Voegtle, T., Wakaluk, S., 2009, Effects on the measurement of the terrestrial laser scanner HDS 6000 (LEICA) caused by different object materials, In: *Proceedings ISPRS Workshop Laser Scanning, September 1-2, Volume XXXVIII, Part 3/W8*, Paris, France, pp. 68-74.

Wang, W.-M., Li, Z.-L., Su H.-B., 2007, Comparison of leaf angle distribution functions: Effects on extinction coefficient and fraction of sunlight foliage. *Agricultural and Forest Metrology*, 143, pp 106-122.

Zheng, G., Moskal, L.M., 2009 Retrieving leaf area index (LAI) using remote sensing: theories, methods and sensors, *Sensors*, 9, pp 2719-2745.

3D-RiskMapping, 2006-2008, Theory and practice on Terrestrial Laser Scanning – Training material based on practical applications. Vlaams Leonardo Da Vinci Agentschap v.z.w.

Optical/LiDAR feature search for nonparametric prediction of stratified forest attributes using an improved genetic procedure

HOOMAN LATIFI† ARNE NOTHDURFT‡ CHRISTOPH STRAUB† BARBARA KOCH†

hooman.latifi@felis.uni-freiburg.de

†Dept. of Remote Sensing and Landscape Information Systems, University of Freiburg, Tennenbacherstraße. 4, D-79106 Freiburg-Germany.

‡Dept. of Biometry and Informatics, Forest Research Institute Baden Württemberg, Wonnhaldestraße 4, 79100 Freiburg- Germany

Abstract

The airborne laser scanner information has often been applied for stand/plot scaled prediction of forest attributes in Nordic areas; whereas its integration in forest inventories is still in experimental stage in central Europe. In a survey in a mixed temperate forest in southwestern Germany, a high dimensional predictor variable set was formed of various plot-level metrics extracted from TM data and height/intensity values of small footprint laser scanner information. Aiming to build parsimonious models, a genetic search algorithm was applied for variable selections. Optimal variable subsets of different cardinalities were derived. An automated process was used to stratify coniferous, deciduous and mixed forest strata using colour infrared orthoimages in a two-dimensional feature space. The selected variable subsets were applied for nonparametric predictions of structural stand attributes using the regression tree-based method "Random Forest" (RF). The model performance was assessed based on leave-one-out cross validations on bootstrap-resample data. Furthermore, the gain of the prior stratification was examined. The results indicate that the stratification of forest areas significantly improved the results of predictions. Moreover, the predictions were of higher accuracy for the forest type-related attributes compared to the total rates. The accuracy follows a slightly increasing trend along with an increased number of selected variables. However, the privilege of using parsimonious models shall be taken into account for practical applications.

1. Introduction

During the last few years, the use of regional scale remote sensing plays a key role in small scale forest inventory in Europe. For instance, the Airborne Laser Scanner (ALS) data have been extensively used as auxiliary information for predictions of attributes on forest stand, plot or single tree level. The ALS data have been applied for the characterization of highly variable forest canopy structures rather than passive optical imagery (Koukoulas and Blackburg, 2005; Koch *et al.,* 2009).

In the Federal state of Baden-Württemberg in southwest Germany, the classification of forest stands into forest types is usually accomplished during terrestrial surveys for mensurational descriptions, in which the personal expertise of the forestry technician plays a major role. Due to the high number of dominant species in forest stands, the stratification is often a time-consuming task in terrestrial surveys. Therefore, a straightforward automated method of forest type stratification by means of remote sensing data would provide a valuable assistance in the current data processing workflow. The near infrared (NIR) spectral information has been assessed to be useful elements of remote sensing data for stratifying forest areas into deciduous and coniferous forest types (e.g. Stibig *et al*., 2004; Straub *et al*., 2009). Here, an automated procedure is applied for the classification of forest types, which is based on spectral information from Colour Infrared (CIR) orthoimages (Straub *et al.* 2009).

The prior classification is used here feed the nonparametric predictions of a set of important forestry variables on plot level. In contrast to parametric regression models, the nonparametric methods may enable more flexible facilities for using the unknown regression relationships (Härdle, 1990). Yet, they often require larger sample sizes than parametric regression models, as the data itself serves as model. Nearest Neighbour (NN) imputation methods select units from the reference set to serve as surrogates for members of the target sets using a measure of similarity based on predictor variables (Stage and Crookston, 2007). These methods have been used for predictions of stand characteristics in Scandinavia (Packalén and Maltamo, 2006; Packalén and Maltamo, 2007), the US (Hudak *et al.*, 2008), and central Europe (Nothdurft *et al.*, 2009; Latifi *et al*., 2010). Here, the Random Forest (RF) method was used, which consists of regression and classification trees for resampled predictor variable sets (Breiman, 2001). The RF method allows for the simultaneous usage of various variables (e.g. multiple remote sensing features) as predictors. This may theoretically improve its precision in a regression; yet the model is always threatened by the risk of overfitting, meaning that the random effects, possible local optima, or noises are being estimated rather than the expected values. This may occur particularly when the number of samples is relatively low or when a high dimensional predictor set is in use. Although the RF method is proved to be more robust against overfitting (Breiman, 2001), there is always a common interest to build parsimonious models, as the models should often be valid beyond the underlying region of parameterization.

A method is used here to reduce high-dimensional remote sensing data and select the optimal candidates for RF models. While the deterministic methods such as stepwise selection are proved to have shortcomings when being applied on highly-correlated variable sets, stochastic search algorithms proved to be more efficient (Barros and Rutledge, 1998; Tomppo and Halme, 2004). Genetic Algorithm (GA) is an approach which works by evolving sets of variables to meet a certain optimization (fitness) criterion. Tomppo and Halme (2004) applied the GA method for selecting predictor variables amongst a set of optical features, whereas Latifi *et al.* (2010) applied it on discretized response variables for selecting variables amongst CIR orthoimages, TM and ALS– extracted features. Although the latter study evaluated the GA to efficiently reduce the relative root mean squared error (RMSE) of total volume and biomass at plot-level compared to stepwise selection methods, the GA was shown to produce unstable results in multiple runs. More appropriate hypothesis tests were applied here to show that GA based search may result in stable and parsimonious variable sets for building efficient NN regression models.

By applying the GA for subset selections, the aim is here to enhance the nonparametric RF models for a simultaneous modelling of a set of forest structural attributes. It is also studied whether the prior classification into forest type strata may lead to more accurate predictions of 6 common forestry attributes. The statistical metrics extracted from medium resolution TM imagery and normalized ALS points were used for the analysis. The model performance was assessed by means of randomly bootstrapped leave-one-out cross validations, and the RMSE was used as the diagnostic tool.

2. Material and methods

Study site

The study site is located in the north of Karlsruhe in the state of Baden- Württemberg, Germany (49° 03' 37" N and 8° 24' 09" E to 49° 01' 15" N and 8° 25' 49" E) and covers nearly 900 ha of managed forests. The stands are dominated by Scots Pine (*Pinus sylvestris* L.) (56.3%), European Beech (*Fagus sylvatica* L.) (17.8%) Sessile Oak (*Quercus petraea* Liebl.) and Pedunculate Oak (*Q. rubra* L.) (jointly 14.9%). Other species only play a minor role. The mean standing timber volume is 264.8 m^3 ha^{-1} which is less than the average in Baden-Württemberg (362 m^3 ha^{-1}). Approximately, the stands have 492 trees ha^{-1} featuring the mean basal area of 25.7 m^2 ha^{-1}.

Forest inventory data

The reference data were collected in summer 2006 on permanent circular sample plots arranged in a regular 100×200 m grid, with each plot consisting of four concentric circles featuring 2, 3, 6 and 12 m of radii.

Various characteristics were measured on the plots. The heights of trees without measurement were predicted by uniform stand height curves. The single tree timber volume was calculated using the taper functions of Kublin (2003). The plot-level total timber volume in m^3 ha^{-1} was derived by summing up the single tree volumes weighted by the inverse of the corresponding sample plot area, and the single tree biomass (i.e. total above-ground biomass including bark, branches, and needles) was estimated with Zell (2008) parameters for the allometric equation:

$$B = \beta_0 d^{\beta_1} h^{\beta_2} + \epsilon$$
$$d\text{: diameter in breast height}$$
$$h\text{: tree height} \qquad (1)$$
$$\beta_0, \beta_1, \beta_2\text{: tree specific function parameters}$$

The total biomass on each sample plot was derived by summing up all single tree biomass estimates. The following attributes were selected as the response variables: total standing timber volume [m^3 ha^{-1}], volume of coniferous trees [m^3 ha^{-1}], total volume of trees featuring > 40cm DBH [m^3 ha^{-1}], stem number of trees featuring >25 cm DBH, total biomass [tons ha^{-1}], and biomass of coniferous trees [tons ha^{-1}]. The stems with DBH > 25 cm were applied to correspond the first-pulse ALS metrics, as they mostly characterize the overstory of the stands. The volume of trees with DBH > 40cm was also used as the information on the portion of mature wood is of major relevance for the forest utilization planning.

Remote sensing data

A Harrier56 LiDAR system was mounted on a Helicopter in August 2007 to acquire Full wave ALS data by Toposys GmbH, in which a Riegl LMS-Q560 scanner was used. The flight was carried out twice over the study site to enable collecting a high point density of 16 points per m^2. A large side lap of over 50% was used in scanning. The “RiANALYZE 560” software was used by the data provider to process the collected full-wave data which was finally delivered as a pulse-form data with a point cloud featuring height, echo signal intensity values and pulse information.

The CIR optical data in 4 spectral bands were recorded in July 2008 using a RGB/NIR line scanner installed on a “Falcon II” system. The data covers an approximate range of 450 nm in the electromagnetic spectrum including visible (RGB) and near infrared (NIR) domains. This data were used to automatically classify coniferous and deciduous forest areas based on a two-dimensional feature space (Straub *et al.,* 2009). In order to keep the predictions independent from the forest stratification process, the CIR data was excluded from the underlying predictor data set of the NN models.

The medium resolution TM images were utilized to extract spectral predictor variables as to provide features of visible and near/mid-infrared domains. Spectral metrics from TM data could be used as surrogate for high resolution CIR images in plot scale predictions, without causing a considerable reduction in prediction accuracy (Latifi *et al.*, 2010). Thus, the cloud-free satellite images recorded by TM sensor in July 2006 were acquired for the study area. The thermal band was excluded from the analysis due to the coarse spatial resolution. The technical parameters of remote sensing data are listed in Table 1.

Table 1.Flight and technical parameters of LiDAR system "Harrier 56", RGB/NIR line scanner "Falcon II", and Landsat- TM satellite scanner. Measurement units in []

Harrier 56 LiDAR system	
Measurement rate	100 [kHz]
Field of View	45 [°]
Flying height	450 m above ground level
Flying speed	30 [m/s]
Point density	16 [points m^2]
Vertical/horizontal accuracy	< 0.20 [m] / < 0.50 [m]
RGB/NIR Falcon II line scanner	
Flying height	700 [m]
Spectral channels	B: 450-490 [nm] G: 500-580 [nm] R: 580-660 [nm] NIR: 770-890 [nm]
Viewing angle	21.6 [°]
Line rate	Up to 330 [Hz]
Pixel per line	682
Ground sampling distance	0.4 [m]
TM satellite data	
Spacecraft	Landsat 5
Acquisition date	04.07.2006- Day
Sun elevation	59.42506336
Sun azimuth	140.94471286
Spectral channels	Band 1: 0.45-0.52 [nm] Band 2: 0.52-0.60 [nm] Band 3: 0.63-0.69 [nm] Band 4: 0.76-0.90 [nm] Band 5: 1.55-1.75 [nm] Band 7: 2.08- 2.35[nm]

Methods

Forest stratification

An unsupervised classification technique was developed for stratification of the study site into coniferous, deciduous and mixed forest applying the CIR data. To this aim, coniferous and deciduous trees in the overstory were segmented using the two-dimensional pixel classification technique described in Straub *et al.* (2009). Near infrared and green channels were selected to define the feature space. Clusters were delineated within the feature space using the following procedure:

The feature space was iteratively smoothed until only two maxima were found. One maximum represented the centre of the "coniferous cluster" and the other one the centre of the "deciduous cluster". The defined cluster centres were used as the initial starting values to delineate the cluster boundaries with the help of a "pouring algorithm" which considers the feature space as topographical surface (MvTec 2010). Starting from the maxima, regions were grown downwards until "valley bottoms" were reached. Then each pixel of the image domain was assigned to one of the clusters. Additionally, ground pixels were extracted using the

height information from ALS. Afterwards, the inventory plots were assigned to one of the Coniferous (C) forest, Deciduous (D) forest or Mixed (M) forest strata. Finally, the plots were assigned to a stratum using the following definition:

$$\text{stratum} = \begin{cases} C \text{ if } P_C \geq 70\% \\ D \text{ if } P_D \geq 70\% \\ M \text{ if } P_C \text{ and } P_D < 70\% \end{cases} \quad (2)$$

Processing of remote sensing data

The ALS point cloud was normalized using a Digital Terrain Model (DTM) to represent the height above ground. The DTM was created using the filtering algorithm adapted from Kraus and Pfeifer (1998) and implemented in LiDAR analysis software "Fusion" (McGaughey, 2009), which is based on linear prediction implemented as an iterative process. Following the creation of a surface with equal weights for all ALS points, the distance and direction to the surface is used to compute weights for each ALS point using the following weight function:

$$\rho_\iota = \begin{cases} 1 & \upsilon_\iota \leq g \\ \frac{1}{1+(a(\upsilon_\iota - g)^b)} & g < \upsilon_\iota \leq g + w \\ 0 & g + w < \upsilon_\iota \end{cases} \quad (3)$$

Where a and b determine the steepness of the weight function, and g determines which points are assigned a maximum weight of 1.0 (assigned to the points below the surface by more than g). The above ground offset parameter w is used to establish an upper limit of points, which results in assigning a weight of 0 to the points above the level ($g + w$). υ_ι is the residual value, which is the oriented distance from the prior surface to the measured point. Following the iteration, the points satisfying the first two conditions of the weight function are considered bare-earth points (McGaughey, 2009).

As the ALS data was acquired during summer time, both first and last pulse data showed approximately similar patterns because of high density of the recorded point cloud. Therefore, the last pulse data was kept out of the analysis. The near-ground laser hits were excluded by taking the points higher than 2 m (i.e. canopy hits) into account when calculating canopy metrics (Packalén and Maltamo, 2006). The height metrics included measures of central tendency, measures of dispersion, height percentiles and height proportions which were computed directly from the normalized first pulse point cloud. The similar statistics (except proportions) were calculated for the ALS intensity values.

The spectral bands of the TM imagery were first used to derive the mean gray values at sample plot level. In addition, the Normalized Difference Vegetation Index (NDVI) IR/R ratio, two main components of Principal Component Analysis (PCA), and the Tasseled Cap components were extracted at plot level.

Variable screening

Though heuristics may also be used when dealing with highly-correlated variable sets, application of appropriate variable selection methods has also become an important issue. The performance of forward/backward stepwise selection procedures has often been criticized compared to the local search heuristics, which have shown an improved performance in feature selection (Barros and Rutledge, 1998; Cadima and Joliffe, 2001).

GA is a variable search procedure which is based on the principle of evolution by natural selection. The procedure works by evolving sets of variables (called chromosomes) that fit certain criteria from an initial random population via cycles of differential replication, recombination and mutation of the fittest chromosomes. The procedure is now adaptable to different optimization scenarios including classification, regression and survival analysis (Trevino and Falciani, 2006). For further information on the theory of GA-search, the reader may refer to Mitchell (1996).

The currently implemented GA optimization procedure (Cerdeira *et al*., 2009) essentially consists of the following steps (Fig 1): The procedure starts with creating a number of random variable sets (initial chromosomes), forming a population of chromosomes (niche). Couples are formed as half the size of the population. Those parents are mated to produce children, inheriting the parent's properties. The children are consisted of the variables belonging to both parents, but also of the variables with equal probability from the parents' symmetric difference. The number of generations (200 generations here) is used as a stopping rule, through which cross-over and mutations may be randomly occurred in the chromosomes. The procedure is iterated until an accurate chromosome is obtained.

Based on the GA a subset of regressor variables is searched, which leads to a maximum of the Tau squared index τ^2 as fitness criterion, which is analogous to the minimization of the Wilk's Lambda criterionΛ. For the matrix of response variables$Y$, and any GA selected subset X_{-} of the full regressor matrix X , Wilk's Lambda is calculated by $\Lambda = \frac{det\,(E)}{det\,(H+E)}$. The GA was implemented in the context of canonical correlation analysis. Hereof, the hypothesis matrix H is calculated as $Corr(Y, X_{-})Corr(X)^{-1}Corr(X_{-}, Y)$ and the error matrix E as$Corr\,(Y) - H$. The Tau squared index is then $T^2 = 1 - \Lambda^{1/r}$, with r being the rank of H. In addition, a local improvement algorithm was applied to the GA selected optimal variable subset (details in Cadima *et al.*, 2004; Cerdeira *et al.,* 2009).

A high solution rate of 1000 bootstraps with a random split into training and test sets was set to stabilize the GA results in repeated runs. The variable selection algorithm was applied for a sequence of different given subset sizes ranging from 5 to 18 variables per subset. The performance of the NN model was assessed for each of the optimal variable subsets of given cardinalities.

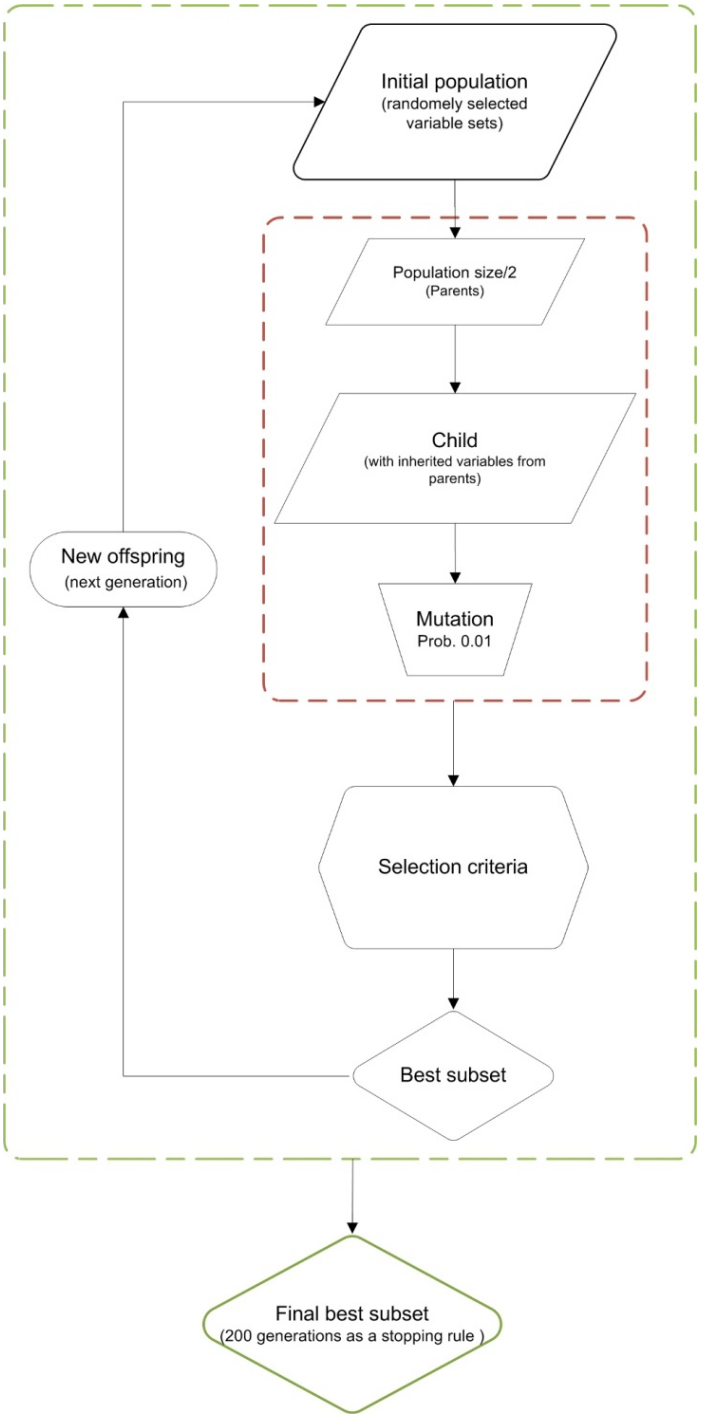


Fig. 1. GA-search in its current implementation

NN Predictions

The classification and regression tree based method of RF was used for simultaneous predictions of the response variables using the selected optical/ALS features. The RF algorithm works briefly as follows (Liaw and Wiener, 2002): It draws bootstrap samples from the original data. For each bootstrap sample, an *unpruned* regression tree is grown, in which the best splits are chosen from the randomly-sampled variables at each node. Then, the new predictions will be made by aggregating the predictions of the total number of trees. The distance between the target and reference units is calculated as one minus the proportion of

terminal nodes from all regression trees, where the target observation is in the same terminal node as the specific reference unit (Crookston and Finley, 2008).

One NN for predictions of each cell was used here by setting $k = 1$. Increasing k leads to a stronger shift of the predictions towards the sample mean, which may cause serious biases particularly in cases where the distribution of observations is skewed. However, an increase of *k* reduces the estimation error (Hudak *et al.*, 2008). The RF prediction was applied on both sets of stratified and unstratified sample plots using the variable subsets of different sizes selected from GA search procedure. The performances of the predictions were compared amongst different variable cardinalities, as well as different response variables.

Accuracy assessment

Performance of variable predictions was assessed by means of leave-one-out cross validation. The results of cross validation were reported as diagnostics including relative RMSE% and performance enhancement $Enh_{RMSE\%}$. If the absolute RMSE of the unstratified samples would be $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$, where n is the total number of sample plots, y_i is the observed response variable on plot i, and $\hat{y}_i$ is the prediction of the response variable on corresponding plot, the relative RMSE% is obtained by $RMSE\% = RMSE/\left(1/n\sum_{i=1}^{n} y_i\right) 100\%$.

For stratified predictions, given the RMSE of the corresponding stratum $RMSE^{(str)}$ as $RMSE^{(str)} = \sqrt{\frac{1}{N}\sum_{c=1}^{3} n_c \times RMSE_c^{\,2}}$, where N is total number of samples, n_c is number of sub-samples for each forest stratum, and $RMSE_c$ is absolute RMSE of the respective stratum, its percentage can be calculated here as $RMSE^{(str)}\% = \frac{RMSE^{(str)}}{\frac{1}{n}\sum_{i=1}^{n} y_i} \times 100$. The performance enhancement was derived as $Enh_{RMSE\%} = \frac{mean(RMSE^{(str)} - RMSE)}{mean(RMSE)} \times 100$. The above-mentioned statistics were calculated for 1000 randomly-bootstrapped sub samples, and the mean values of the entire runs were reported as final rates.

3. Results

Most of the GA-selected variables were those from ALS height metrics (Table 2). While some of the variables are repeated as the cardinality increases, some are replaced by the variables being highly correlated with them. As revealed as an example in Fig.2, the stratification of the entire regression space into forest types led to notably enhanced performances of the RF predictions for all of the six attributes. Using the prior classified data for predictions proved to be superior compared to the application of the unclassified data for all the subset sizes. The improvement is more obvious in two strata-related response variables including coniferous volume and coniferous biomass.

Table 2. GA -selected predictor variables from remote sensing data. The superscripts present the size of variable subsets in which the corresponding metric is included

Data source	Selected predictors	Description
Normalized LiDAR points - First pulse	Height. Mean [5,16]	Height mean
	Height. StdDev [5,18]	Height Standard Deviation
	Height. InterquartileDistance [12,16,18]	Height interquartile distance
	Height. Skewness [7,11,13,14,15,18]	Height skewness
	Height.P10 [8,13]	Height 10th percentile
	Height.P20 [6,12,16,18]	Height 20th percentile
	Height.P30 [18]	Height 30th percentile
	Height.P60 [10,12,16]	Height 60th percentile
	Height.P80 [9,12,13]	Height 80th percentile
	Height.P90 [10]	Height 90th percentile
	Height.P95 [6,8]	Height 95th percentile
	Percentage.FR.2-5m [9,10,11,12,15,16,18]	Percentage of first pulses >2 and <5m
	Percentage.FR.10-12.5m [7,11,13,14,15]	Percentage of first pulses >10 and <12.5m
	Percentage.FR.12.5-15m [9,14,16,18]	Percentage of first pulses >12.5 and <15m
	Percentage.FR.15-17.5m [16]	Percentage of first pulses >15 and <17.5m
	Percentage.FR.17.5-20m [14,18]	Percentage of first pulses >17.5 and <20m
	Percentage.FR.20-25m [6,7,8,9,11,12,13,14,15,16,18]	Percentage of first pulses >20 and <25m
	Percentage.FR.25-30m [5,6,7,8,9,10,11,12,13,14,15,16]	Percentage of first pulses >25 and <30m
	Int. Mean [8,18]	Intensity mean
	Int. Mode [10,11,13,14,15,16,18]	Intensity mode
	Int. Median [7,15]	Intensity median
	Int. StdDev [18]	Intensity Standard Deviation
	Int. InterquartileDistance [9]	Intensity interquartile distance
	Int. Skewness [6]	Intensity skewness
	Int.P10 [14,15,16,18]	Intensity 10th percentile
	Int.P20 [11,15,16]	Intensity 20th percentile
	Int.P30 [7,10,11,14,18]	Intensity 30th percentile
	Int.P40 [10,11,12,14,16,18]	Intensity 40th percentile
	Int.P60 [5,12]	Intensity 60th percentile
	Int.P70 [8,13]	Intensity 70th percentile
	Int.P95 [9,10,12,13,14,15,16]	Intensity 95th percentile
TM data	TM2 [10,14,15,18]	TM band 2
	TM4 [13]	TM band 4
	TM5 [6,13,14]	TM band 5
	TM7 [8,9,11]	TM band 7
	TM NDVI [13,15]	NDVI
	TM IR/R [12,13,18]	IR/R band Ratio
	TM_B457_PC1 [8,9,10,11,12,15]	2nd principal components of infrared bands
	TM_B457_PC2 [7]	2nd principal components of infrared bands
	TASSEL.BRIGHTNESS [15]	Tasseled Cap transformation- Brightness
	TASSEL.GREENNESS [5,15,16]	Tasseled Cap transformation- Greenness
	TASSEL.WETNESS [14,16,18]	Tasseled Cap transformation- Wetness

The RMSE% of the RF predictions decreased for all the response variables along with an increasing number of selected predictors. Whereas both total timber volume and total biomass proved to gain the lowest RMSE, the standing volume of trees > 40cm DBH yielded the worst RMSE compared to other variables under prediction (Fig. 3). Fig. 4 and Table 3 demonstrate a comparison of the performance enhancement rates for all the predicted variables across the variable subset sizes. While the improvement were the lowest for the

total stem number of trees having DBH > 25cm, the enhancements were apparently higher for the predictions of those strata-related response variables (i.e. coniferous volume and coniferous biomass).

Furthermore, the highest RMSE-enhancement in most of the response variables was observed when the size of subset was 12 variables.

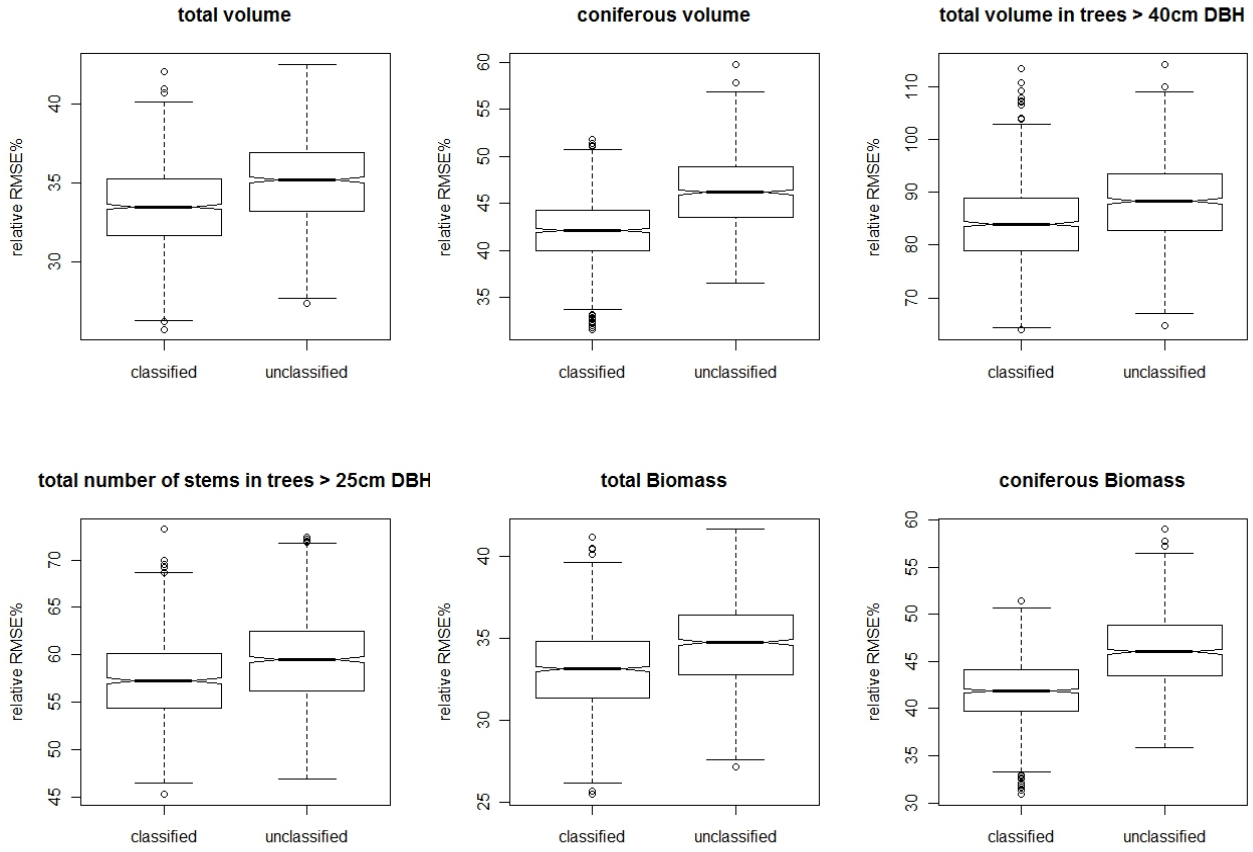


Fig 2. The relative RMSE% of RF predictions of the six response variables for the classified and the unclassified variant from 1000 bootstrap resample for subset of 12 variables. Non-overlapping notches between boxplots provide a strong evidence of significant difference of medians (Chambers et al., 1983).

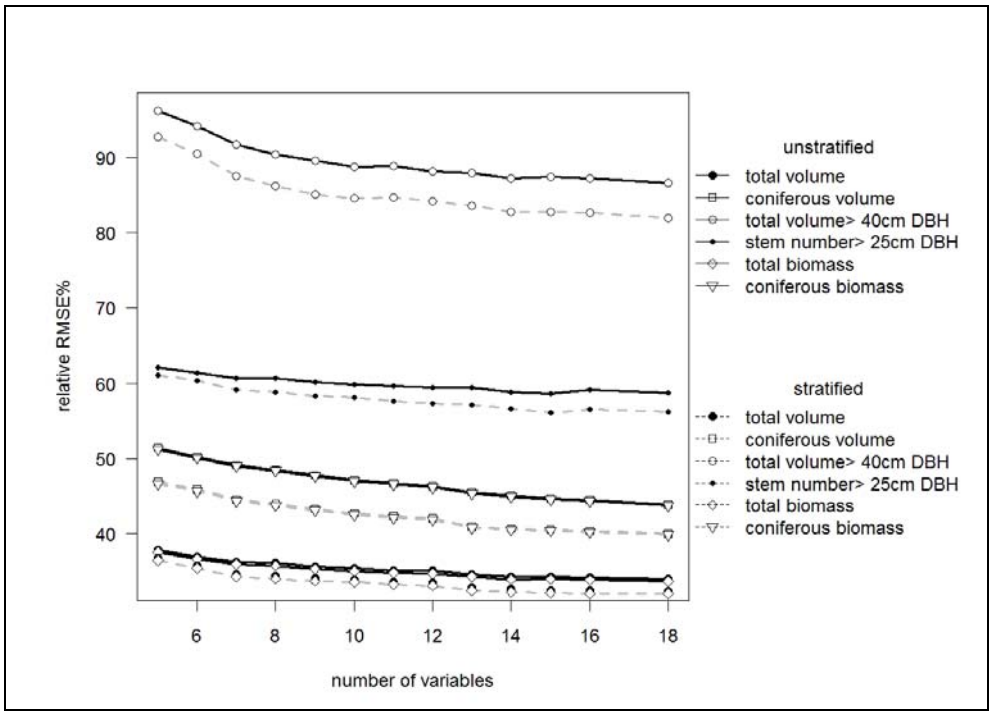


Fig. 3. Relative RMSE% of RF predictions for growing size of the optimal variable subsets for the unstratified (solid lines) and the stratified variant (dashed lines).

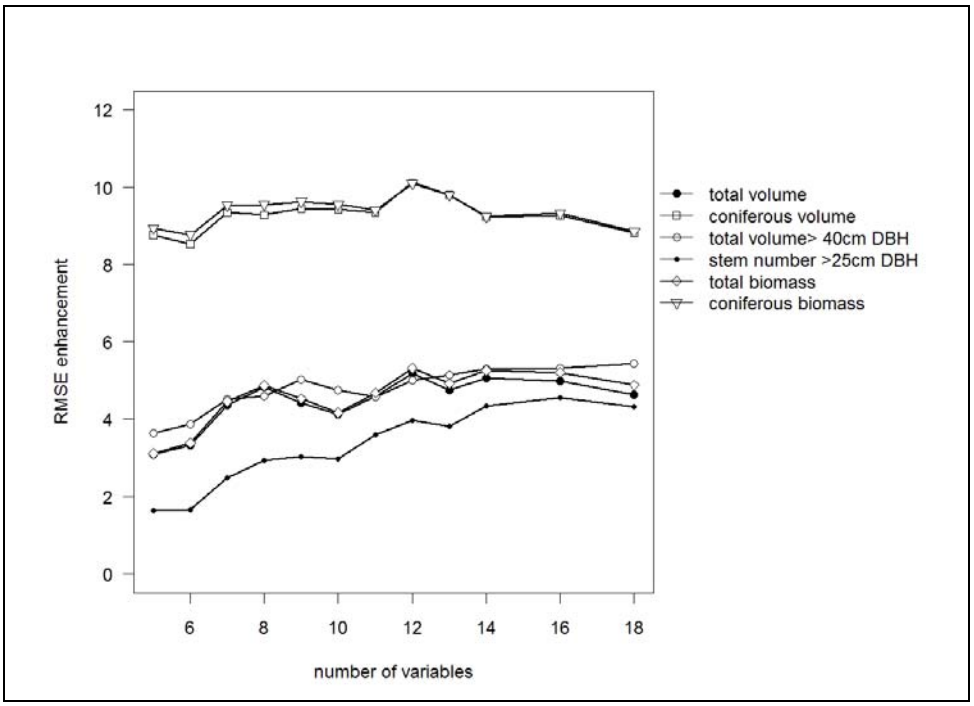


Fig. 4. Performance enhancement rates induced by the prior forest stratification for RF predictions of six response variables in different GA-selected subset sizes.

Table. 3. Performance enhancement rates in RMSE% induced by the prior forest type stratification for RF predictions of six-fold response variables in different GA-selected subset sizes.

Number of Variables	Performance enhancement rates [%]					
	Total volume	Coniferous volume	Total volume in trees> 40cm DBH	Number of stems in trees> 25cm DBH	Total biomass	Coniferous biomass
5	3.08	8.76	3.63	1.64	3.11	8.94
6	3.33	8.53	3.88	1.66	3.38	8.77
7	4.37	9.34	4.52	2.48	4.46	9.53
8	4.82	9.28	4.60	2.93	4.87	9.55
9	4.42	9.44	5.02	3.04	4.53	9.63
10	4.12	9.42	4.76	2.97	4.16	9.56
11	4.59	9.34	4.57	3.60	4.68	9.41
12	5.18	10.12	5.00	3.97	5.32	10.10
13	4.76	9.80	5.15	3.82	4.92	9.80
14	5.07	9.23	5.29	4.35	5.26	9.24
16	4.98	9.26	5.32	4.55	5.21	9.32
18	4.63	8.82	5.43	4.32	4.88	8.86

4. Discussion and Conclusion

Here, an efficient variable pruning scheme was carried out to reduce the dimensionality of correlated high-dimensional predictor variable set. The use of GA resulted in parsimonious NN models for plot-level predictions of relevant forest attributes. Whereas some studies (e.g. LeMay *et al.* 2008) used the canonical correlations as a standalone variable selection measure, we used it as an input to the evolutionary GA procedure to take all the possible combinations of predictors into account. Here, the τ^2 optimization criterion based on Wilk's Lambda was applied to obtain a tentative solution, which can then be fed to a local improvement algorithm for refinement. The latter enabled further enhancement of the selected subset, as also stated by Cadima *et al.* (2004) and Cerdeira *et al.* (2009).

The prior stratification used here does not require any user-defined training regions. However, further studies will be required to test the possibility of a more detailed classification of trees species. The results of RF modelling showed an increase in prediction accuracy across growing subset cardinalities. For majority of response variables, the improvement was maximized at the cardinality of 12 variables. Our optimal number of predictors is in line with previous studies in the field (Hudak *et al.*, 2008, Breidenbach *et al.*, 2010). The majority of the selected variables consist of first pulse height metrics extracted from ALS data, which reveals the dominance of topographic LiDAR data when predicting the forest structural attributes (e.g. Maltamo *et al.*, 2006; Hudak *et al.*, 2008). Yet, few optical metrics were also present in all the variable subsets.

RF takes simultaneously both predictor and response variables into account. Because of the Law of Large Numbers, it is reported to be less prone to overfit as more trees are added to the analysis (Breiman, 2001). Although increasing the number of regression trees built for each response variable may stabilize the estimates, the RF predictions may still diverge slightly from run to run. Therefore, the predictions were carried out numerous times as a resolve to stabilize the RF prediction results. The results of predictions using different subset cardinalities showed a stable increment of accuracy, meaning that the predictions are consequently protected of being trapped in possible local optima. The volume of trees > 40 cm DBH was

predicted least accurately (i.e. featuring > 80% RMSE), which is commonly due to the low frequency of old stems across the study area.

Both volume and biomass showed nearly equal performances, which was evident as the biomass is estimated by allometric equations with the standing volume as independent variable. Our results also support the results already obtained by Packalén and Maltamo (2007) and Maltamo et al (2009), in both of which the predictions of stem number are generally worse than those of plot/stand volume. Coniferous volume and biomass achieved far better performance enhancements compared to other attributes. This revealed the gain of the prior classification which led to more precise predictions by means of homogeneous strata.

This study yielded promising results in terms of applicability of the nonparametric methods for making area-based predictions of stratified forest structural attributes in the study area. It showed how an evolutionary GA search can be applied to a high dimensional remote sensing dataset to produce parsimonious variable subsets for NN regressions models. According to the results, the methodology is expected to produce promising results across similar areas, where the problems of small sample size and multiple variables seem to be challenging. Yet, supplementary studies are still required.

Acknowledgements

We would like to express their gratitude to the German Academic Exchange Service (DAAD) for financial support of this research within a doctoral scholarship. The authors are grateful to the Forest Service of Baden-Württemberg for providing the required forest inventory data.

References

BARROS, A. S., RUTLEDGE, D. N. 1998. Genetic algorithm applied to the selection of principal components. *Chemometrics and Intelligent Laboratory Systems* 40:65–81.

BREIDENBACH, J., NÆSSET, E., LIEN, V., GOBAKKEN, T., SOLBERG, S. 2010. Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. *Remote Sensing of Environment* 114: 911-924.

BREIMAN,L. 2001. Random Forests. *Machine Learning* 45(1): 5-32.

CADIMA, J. and JOLLIFFE, I.T. 2001. Variable Selection and the Interpretation of Principal Subspaces. *Journal of Agricultural, Biological and Environmental Statistics* 6: 62-79.

CADIMA, J., CERDEIRA, J. Orestes and Minhoto, M. 2004. Computational aspects of algorithms for variable selection in the context of principal components. *Computational Statistics & Data Analysis* 47: 225-236.

CERDEIRA, J.O., DUARTE SILVA, P., CADIMA, J., MINHOTO, M. 2009. Package "Subselect" usergide. R Development Core Team. Available at: www.cran.r-project.org/web/packages/subselect/index (accessed 19.12.2010)

CHAMBERS, J. M., CLEVELAND, W. S., KLEINER, B., TUKEY, P. A. 1983. *Graphical methods for data analysis*. Belmont, CA: Wadsworth and Boston,MA: Duxbury.

CROOKSTON, N. L., FINLEY, A. 2008. yaImpute: an R package for KNN Imputation. *Journal of Statistical Software* 23(10): 16p.

HÄRDLE, W. 1990. *Applied nonparametric regression* (In series: Econometric society monographs). (Cambridge University Press). ISBN 0-521-42950-1.

HUDAK, A., CROOKSTON, N., EVANS, J., HALL, D., FALKOWSKI, M. 2008. Nearest neighbor Imputation of species-level, plot-scale forest structure attributes from LiDAR data. *Remote Sensing of Environment* 112: 2232-2245.

KOCH, B., STRAUB,C., DEES, M., WANG,Y., WEINACKER, H. 2009. Airborne Laser Data for Stand Delineation and Information Extraction, *International Journal of Remote Sensing* 30(4) : 935-963.

KOUKOULAS, S., BLACKBURN, G.A. 2005. Mapping individual tree location, height and species in broadleaved deciduous forest using airborne LiDAR and multi-spectral remotely sensed data. *International Journal of Remote Sensing* 26: 431-455.

KRAUS, K., PFEIFER, N. 1998. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing* 53: 193-203.

KUBLIN, E. 2003. Einheitliche Beschreibung der Schaftform – Methoden und Programme – BDATPro. *Forstwissenschaftliches Centralblatt* 123: 183-200 (in German with English summary).

LATIFI, H., NOTHDURFT, A., KOCH, B. 2010. Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: application of multiple optical/LiDAR –derived predictors. *Forestry*. doi:10.1093/forestry/cpq022 13p.

LEMAY, V., MAEDEL, L., COOPS, N. C. 2008. Estimating stand structural details using nearest neighbor analyses to link ground data, forest cover maps, and Landsat imagery. *Remote sensing of Environment* 112: 2578-2591.

LIAW, A., WIENER, M. 2002. Classification and regression by randomForest. *R News* 2:18−22.

MALTAMO, M., MALINEN, J., PACKALÉN, P., SUVANTO, A and KANGAS, J. 2006.Non-parametric estimation of stem volume using airborne laser scanning, aerial photography, and stand-register data. *Canadian Journal of Forest Research* 36: 426-436.

MALTAMO, M., PACKALÉN, P, SUVANTO, A., KORHONEN, K.T., MEHTÄTALO, L., HYVÖNEN, P. 2009. Combining ALS and NFI training data for forest management planning: a case stud in Kuortane, western Finland. *European Journal of Forest Research* 128: 305-317.

MCGAUGHEY.R. J. 2009. *FUSION/LDV: Software for LIDAR Data Analysis and Visualization (Ver. 2.70)*. USDA Forest service- Pacific Northwest Research Station. 127p.

MITCHEL, M. 1996. *An introduction to genetic algorithms*. (The MIT Press-USA), ISBN 0-262-13316-4.

MVTEC. 2010. HALCON/.NET Reference Manual. Available online at: http://halcon.cn/download/documentation/reference-8.0/dotnet/pouring.html

NOTHDURFT, A., SABOROWSKI, J., BREIDENBACH, J. 2009. Spatial prediction of forest stand variables. *European Journal of Forest Research* 128(3): 241-251.

PACKALÉN, P., MALTAMO, M. 2006. Predicting the plot volume by tree species using airborne laser scanning and aerial photographs. *Forest Science* 52(6): 611-622.

PACKALÉN, P., MALTAMO, M. 2007. The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. *Remote Sensing of Environment* 109: 328-341.

STAGE, A. R., CROOKSTON, N. L. 2007. Partitioning error components for accuracy-assessment of near- neighbor methods of imputation. Forest Science 53(1): 62−72.

STIBIG, H.J., ACHARD, F., FRITZ, S. 2004. A new forest cover map of continental southeast Asia derived from SPOT-VEGETATION satellite imagery. *Applied Vegetation Science* 7: 153-162.

STRAUB, C., DEES, M., WEINACKER, H., KOCH, B. 2009. Using Airborne Laser Scanner Data and CIR Orthophotos to Estimate the Stem Volume of Forest Stands. *Photogrammetrie, Fernerkundung, GeoInformation* 3: 277-287.

TOMPPO, E., HALME, M. 2004. Using coarse scale forest variables as ancillary information and weighting of variables in k-NN estimation: a genetic algorithm approach. *Remote Sensing of Environment* 92: 1-20.

TREVINO, V., FALCIANI, F. 2006. GALGO: An R Package for Multivariate Variable Selection Using Genetic Algorithms. *Bioinformatics* 22(9):1154-1156.

ZELL, J. 2008. Methoden für die Ermittlung, Modellierung und Prognose der Kohlenstoffspeicherung in Wäldern auf Grundlage permanenter Großrauminventuren. Ph.D Dissertation, Faculty of Forest and Environmental Studies, University of Freiburg. 162p (in German with English summary).