Estimating crown base height for Scots pine by means of the 3D geometry of airborne laser scanning data

Jari Vauhkonen

University of Joensuu, Faculty of Forest Sciences, P.O. Box 111, 80101 Joensuu, Finland – jari.vauhkonen@joensuu.fi

Abstract

Crown base height (CBH) is an important factor in relation to several characteristics of the tree stock. This paper introduces approaches for estimating tree-level CBH from airborne laser scanning (ALS) data that employ features of computational geometry. For that purpose, the concepts of Delaunay triangulations and alpha shapes were applied and compared with approaches based on analysing return frequencies and predicting CBH by linear regression. These approaches were evaluated using test data on a total of 133 sawlog-sized Scots pine trees detected and delineated from ALS data with a density of approx. 4 returns m^{-2}. The results suggest that variables based on cross-section area and the frequencies of crown returns within predefined height bins are the most accurate for estimating CBH. By combining the best CBH estimate with the estimated tree height in linear regression, an RMSE of 1.5 m (14%) was achieved. Although the accuracy of estimating CBH was lower using the 3D geometry approaches presented here, they were considered to have potential for further development.

Keywords: LIDAR; computational geometry; Delaunay triangulation; alpha shape

1. Introduction

Previous studies have shown that it is possible under Scandinavian conditions to detect 40-70% of all trees using the individual tree delineation approach with airborne laser scanning (ALS) data (Persson *et al.* 2002; Maltamo *et al.* 2004). Although the tree detection rate and segmentation accuracy are highly dependent on forest structure, the trees that are detected are highly representative of the dominant tree layer, which is the most significant part of the forest for many applications. The individual tree information could be used either as such or as a complement to area-based approaches in applications such as characterising the growing stock (Persson *et al.* 2002; Holmgren and Persson 2004), monitoring its development (Yu *et al.* 2004) and planning timber procurement (Peuhkurinen *et al.* 2007). The rapid developments currently taking place in ALS technology will enable the necessary point density to be achieved with lower data acquisition costs in the near future, whereupon more interest will obviously be shown in single-tree methods.

Besides tree height, the height of the tree crown is an important characteristic obtainable from tree-level ALS data, since the crown base height (CBH) is related to tree growth, forest health, timber quality, the need for silvicultural operations and their optimal timing, for example. Due to the laborious measurements involved, CBH is seldom measured in the field, however, and it is also difficult to model using more commonly recorded field observations (e.g. Hynynen 1995). Some approaches that employ ALS data for estimating CBH have been presented in the last few years. Maltamo *et al.* (2006), comparing variables based on both ALS and field measurements, concluded that the accuracy of predicting the CBH was in practice similar at both the tree and plot level regardless of the source of the predictors. Plot-level approaches have usually been based on analysing the height distribution of laser returns (Næsset and Økland 2002; Andersen *et al.* 2005; Maltamo *et al.* 2006), whereas at the tree level various properties of the delineated 3D point clouds have been used to derive CBH (Pyysalo and Hyyppä 2002; Holmgren and Persson 2004; Maltamo *et al.* 2006; Solberg *et al*. 2006; Holmgren *et al.* 2008; Popescu and Zhao 2008).

Pyysalo and Hyyppä (2002) developed polygon models for extracting features from individual tree ALS data with a density of app. 10 returns m^{-2} and claimed that the upper canopy could be described in detail but parameters extracted from the lower parts were less accurate. The CBH was overestimated by 3 m. Holmgren and Persson (2004) divided a dataset of approx. 5 returns m^{-2} into 0.5 m height layers and defined the CBH as the distance from the ground to the lowest point above the highest layer that contained less than 1% of the non-ground points. This gave a correlation of 0.84 between estimated and measured CBH and led to an overestimation of 0.75 m. In a later study, Holmgren *et al.* (2008) approximated the tree crown using the alpha shape technique, calculating the area of the shape in voxel height layers and determining the CBH as the smallest area, within certain limits, below the maximum shape area. They now achieved a higher correlation between the estimated and measured CBH values (0.91), but were not able to determine whether this was because of the more efficient algorithm or the considerably higher density of approx. 50 returns m^{-2}. In this case the CBH was underestimated by 0.61 m. Solberg *et al*. (2006), for one, examined the deciles calculated from the first return height distribution. CBH was set at the upper of two neighbouring deciles having the largest height difference in between. This approach resulted in an overestimation of 3 m (RMSE 3.5 m). Recently Popescu and Zhao (2008) used ALS data with a density of 2.6 returns m^{-2} to extract pulse frequency and intensity data with height bins defined in terms of 0.5 x 0.5 x 1.0 m voxels. They fitted the resulting vertical profile to a polynomial and defined the CBH as the height corresponding to an inflection point in the polynomial. Overestimations of 0.36 and 0.12 m in the frequency and intensity approaches, respectively, were removed in Popescu and Zhao (2008) by obtaining the final CBH estimate by means of linear regression. In that way the RMSE was approx. 2 m for the frequency approach and slightly more for the intensity-based approach.

Most of the previous approaches seem to require the operator to define the essential parameters, such as the division of the height bins. As an individual tree is quite a small sample unit, it is crucial for estimation accuracy that the predefined parameters should be in the correct relation to the density of the data, for example. An approach that was capable of adapting to the properties of the source data could also reduce the need for field reference material. Computational geometry is a branch of computer science that deals with the study of algorithms and data structures for solving problems stated in terms of basic geometrical objects, such as points, line segments and polygons. As major attention is paid to the computational efficiency of the algorithms, the use of these could be advantageous for dealing with high density ALS data. So far it is mainly the concept of Delaunay triangulation that has been used in preprocessing the data (e.g. Hyyppä *et al.* 2001), while alpha shapes, for example, have recently been introduced for the later analysis (Holmgren *et al.* 2008; Vauhkonen *et al.* 2008). Here it is assumed that by employing suitable computational features it could be possible to estimate CBH independently of the properties of the ALS data.

The purpose of this study was to apply the concepts of Delaunay triangulations and alpha shapes to the prediction of the CBH for sawlog-sized Scots pine trees (*Pinus sylvestris* L.), bearing in mind the importance of this metric with respect to several attributes of this tree species in particular. The methods developed here were compared with alternative approaches adopted from earlier studies.

2. Material

Altogether 14 30 x 30 m square plots typically located in pure Scots pine stands on less fertile soils were established in the southern part of Koli National Park (lat. 63°1'19"N, long.

29°53'10"E) in North Karelia, eastern Finland, during the spring of 2006. All trees with a diameter at breast height of more than 5 cm were mapped and the attributes of each were recorded. Differentially corrected Global Positioning System measurements were used to determine the positions of the four corners of each of the plots. The accuracy of the positioning in the XY direction was approx. 1 m. Tree locations within a plot were measured using one corner as the origin and projecting the trees onto the same coordinate system as in the ALS data by affine transformation using the measured corner positions as reference points.

Georeferenced ALS point data were collected from an area of approx. 2500 ha in Koli on July 13, 2005, using an Optech ALTM 3100 scanner. Three Differential Global Positioning System receivers were employed to record the carrying platform position: one on the aircraft and two on the ground (the first as the base station and the second for back-up). ALS data was acquired using a mean altitude of 900 m above ground level, resulting in a nominal sampling density of about 4 returns m^{-2}. Elevations within the test area varied from 95 m to 350 m (local zero sea level), resulting in a varying sampling density across the target. The divergence of the laser beam (1064 nm) was 0.26 mrad. The data were captured using a scanning angle of ±11 degrees, which resulted in a swath width of about 350 m. The last pulse data were employed to generate a digital terrain model (DTM) by the method explained in Axelsson (2000) using a grid of 1 m. Height values for the laser points were obtained by subtracting the corresponding DTM values. Points with a value over 0.5 m were classified as vegetation hits. The canopy height model (CHM) was interpolated to a grid of 0.5 m using canopy heights by taking the maximum value of the laser measurements within a radius of 0.5 m.

The individual trees were detected from the CHM using a method described by Pitkänen *et al.* (2004) in which the CHM was first low-pass filtered using Gaussian kernels with the size of the smoothing window and the intensity of the smoothing increasing as a stepwise function of the heights of the CHM. Local height maxima were searched for from the filtered CHM (Pitkänen *et al.* 2004), and all the pixels were classified in the binarization as belonging either to the tree canopy or to the background area. Finally, watershed segmentation was performed to create the crown segments. A crown segment was linked to a field-measured tree if 1) only one field tree was met inside the segment and 2) the difference between the maximum height value within the segment and the field height was less than 2 m. The study was further focused into sawlog-sized (diameter at breast height over 17 cm) Scots pine trees (N=133). The heights measured in the field for the study trees ranged from 12.1 to 27.2 m, with an average of 19.5 m, and the CBH values from 4.8 to 18.8 m, with an average of 11.1 m.

3. Methods

The main attention here was focused on extracting information from tree-level ALS data using Delaunay triangulation (Figure 1c), a widely known technique in the literature of computer science, while another computational geometry technique that was also applied was the concept of alpha shapes (Edelsbrunner and Mücke 1994). An alpha shape can be regarded as a weighted Delaunay triangulation from which all the simplices which have an empty circumsphere with a squared radius larger than the defined alpha value have been removed. Although illustrated in 2D (Figure 1c), the computations regarding both concepts were performed in 3D using the functionality of the Open Source library CGAL (http://www.cgal.org). The reference methods employed information on the vertical profiles of trees (Figure 1a and b). The methods based on return frequencies (Holmgren and Persson 2004; Solberg *et al.* 2006; Popescu and Zhao 2008), cross-section area (Holmgren *et al.* 2008) and linear regression (Maltamo *et al.* 2006; Popescu and Zhao 2008) were adapted slightly for the present purpose. Thus, the variables considered in the estimation were obtained by

1. constraining Delaunay triangulation by average triangle size (*avgtri*) and alpha value (*alphatri*);

2. extracting connected alpha shape components from the tree base (*comp*);
3. analysing layered (*A_bins*) and incremental (*A_incr*) accumulation of 2D area;
4. analysing vertical profile based on return frequencies (*freq*); and
5. linear regression based on tree height (*h*) and the previous variables.

The first approach was based on detecting discontinuities in the 3D triangulation in terms of large tetrahedra (see the 2D representation in Figure 1c). Two alternative methods were applied for classifying a tetrahedron as unacceptably large. In the average triangle (*avgtri*) method, the highest 50% returns were first triangulated and the volume of an average tetrahedron was used as this criterion. In addition to the volume, metrics such as circumradii and different edge lengths were considered for the same purpose with more or less the same result. In the *alphatri* approach, a predefined alpha value was used for the same purpose. Efforts were made to link an alpha value with the tree size by means of the estimated tree height, but as the same result could be obtained using different alpha values, this was found troublesome. Here α=4 was chosen with respect to the reference data. In the actual algorithm, the neighbouring cells of the highest tetrahedron were traversed and if a cell was considered small by the given criterion, its neighbours were also traversed, this being repeated for as long as possible. The crown base height was then defined as the height of the lowest vertex in the obtained structure.

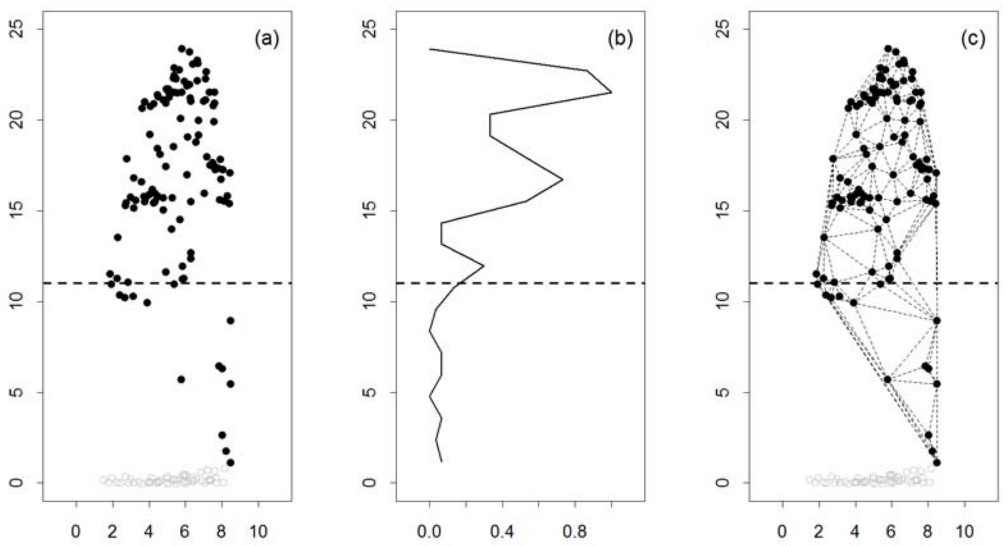

Figure 1: Possibilities for extracting information from ALS data. The ALS profile of an example tree (a), normalised return frequencies within 10% height bins (b), and Delaunay triangulation performed using the point data (c), illustrated in 2D for ease of visualization. The field-measured CBH is illustrated using a dashed line and ground hits using grey circles.

In the *comp* approach, connected components were removed from the lowest parts of an alpha shape generated with the full point data (Figure 2). An alpha value with one connected component was used as a starting point and the alpha values were traversed in descending order until a new component was split or the minimum height value of the highest component was changed. The first split component was allowed to intersect the previous, but otherwise the removal was accepted only if the component was located below the current main component. If not, the procedure was stopped and the CBH was defined as in the previous paragraph.

Third, the vertical profile of the point cloud (Figure 1b) was analysed. A 2D area approximated by the convex hull technique (*A_bins*) and return frequencies (*freq*), both extracted using overlapping 10% height bins with bin values of 5, 10, …, 95% of the tree height, were considered in the analysis. The value within a height bin was normalized using the largest bin value, resulting in values between 0 and 1. Values less than 0.1 were considered to be zeros. The CBH was defined as the lowest point within the first of 2 sequential zero bins below the maximum. The height bins had some overlap and 2 zero classes were considered in order to be sure not to detect a false CBH caused by the height bin division and the low pulse density. In a variation utilizing the 2D area (*A_incr*), the maximum area of the point cloud was first calculated and the point cloud was then traversed from the 20% tree height towards the top. While traversing, the area including the traversed point was calculated, and the crown base was defined at the point where the area calculated in this way exceeded a threshold of 20% of the maximum area.

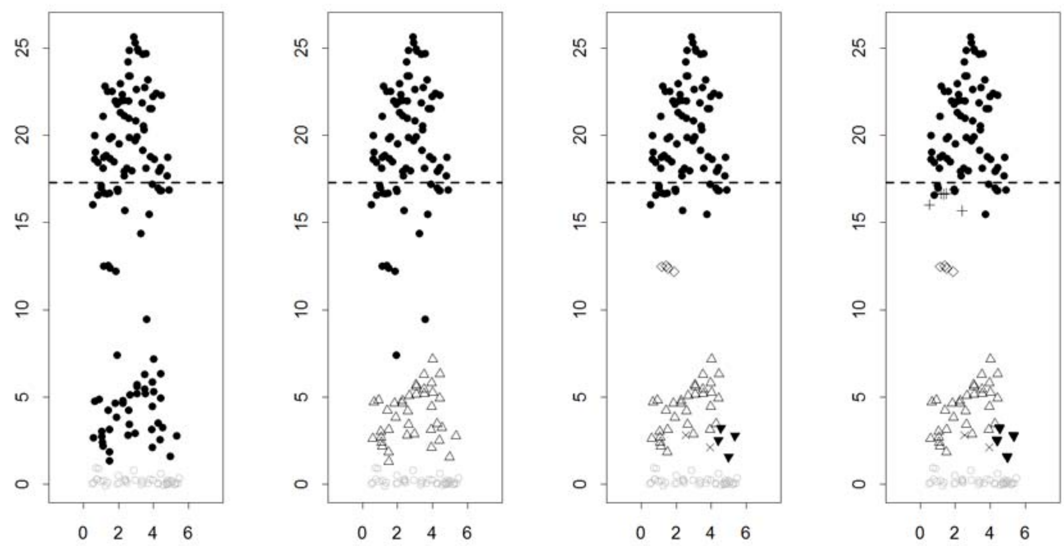

Figure 2: The first and last two steps of the *comp* algorithm in the case of an example tree. From left to right, 1, 2, 5 and 6 connected components illustrated using different symbols were separated from the point cloud during the traverse of alpha values. When the last split component (cross (+) symbols in the right-hand figure) overlapped the main component, the procedure was stopped and the previous main component was output as the result of the algorithm. The CBH measured in the field is illustrated using a dashed line. Note that the result was obtained in 3D and, thus, differs from 2D interpretation.

Finally, the CBH was predicted using linear regression. First the tree height alone was used as the independent variable, second all previous estimates were added to the model and any variables that were insignificant at the 95% confidence level according to the *t*-test scores were removed. The final CBH prediction was produced using leave-one-out cross validation, i.e. by fitting the regression model to all the observations except for the target tree itself.

The reliability of estimating the CBH was measured in terms of RMSE and bias:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(CBHfield_i - CBHpred_i\right)^2}{n}}, \text{ and} \qquad (1)$$

$$bias = \frac{\sum_{i=1}^{n}\left(CBHfield_i - CBHpred_i\right)}{n}, \qquad (2)$$

where n is the number of trees, $CBHfield_i$ the reference and $CBHpred_i$ the predicted CBH for tree i. The relative RMSEs were calculated by dividing the absolute RMSE values by the mean reference CBH.

4. Results

The 3D geometry methods presented here were slightly poorer than the reference methods with respect to estimation accuracy (Table 1), with RMSE values ranging from 2.77 to 3.88 m as opposed to approx. 2 m for the other methods. The estimation accuracy with all methods seems to slightly decrease as tree size increases (Figure 4).

The differences in output between the methods are illustrated in Figure 3. The 3D methods actually produce a subset of the initial point cloud that is adjusted from the lower parts of the crown, whereas the other methods only output a vertical CBH boundary. Compared to other approaches, the 3D methods are slightly more vulnerable to gaps in the vertical profile above the field-measured CBH (e.g. Figure 3b). Thus, the lower accuracy of the 3D methods is mainly caused by several clear outliers (Figure 4), although the average results (Table 1) are, however, fairly close to each other.

Table 1: Estimation accuracies of the different methods.

CBHpred =	RMSE, m	RMSE, %	bias, m
avgtri	3.88	34.9	0.522
alphatri	2.77	24.9	0.115
comp	2.83	25.4	-0.945
A_bins	1.84	16.6	0.411
A_incr	2.17	19.5	-0.738
freq	1.81	16.3	0.004
$-0.68 + 0.62 \times h$	2.03	18.2	-0.001
$-0.51 + 0.29 \times h + 0.57 \times A\_bins$	1.54	13.9	-0.006

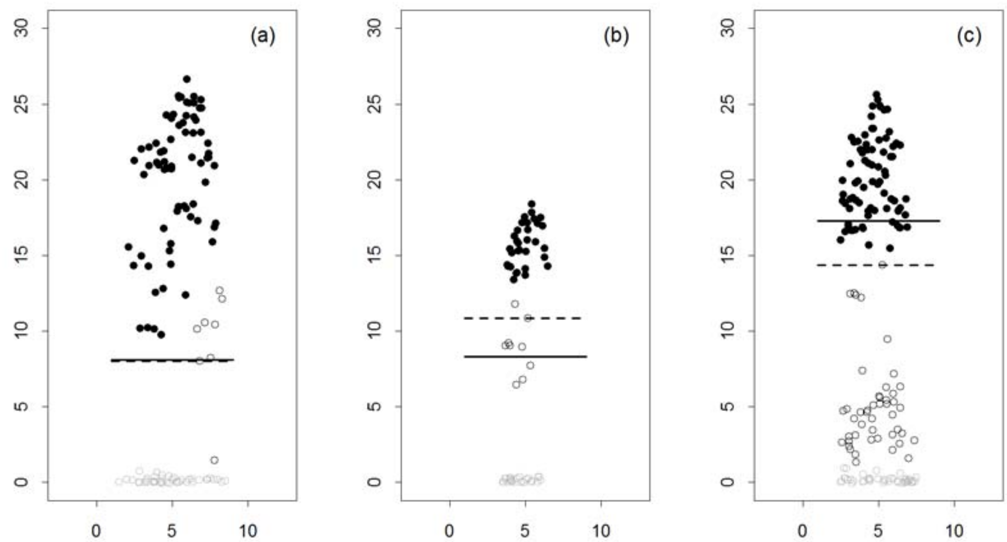

Figure 3: Differences in output between the methods – CBH as measured in the field (thick line), CBH estimate with the *freq* approach (dashed line), output of the *comp* method (black filled dots), and other laser returns (circle symbols).

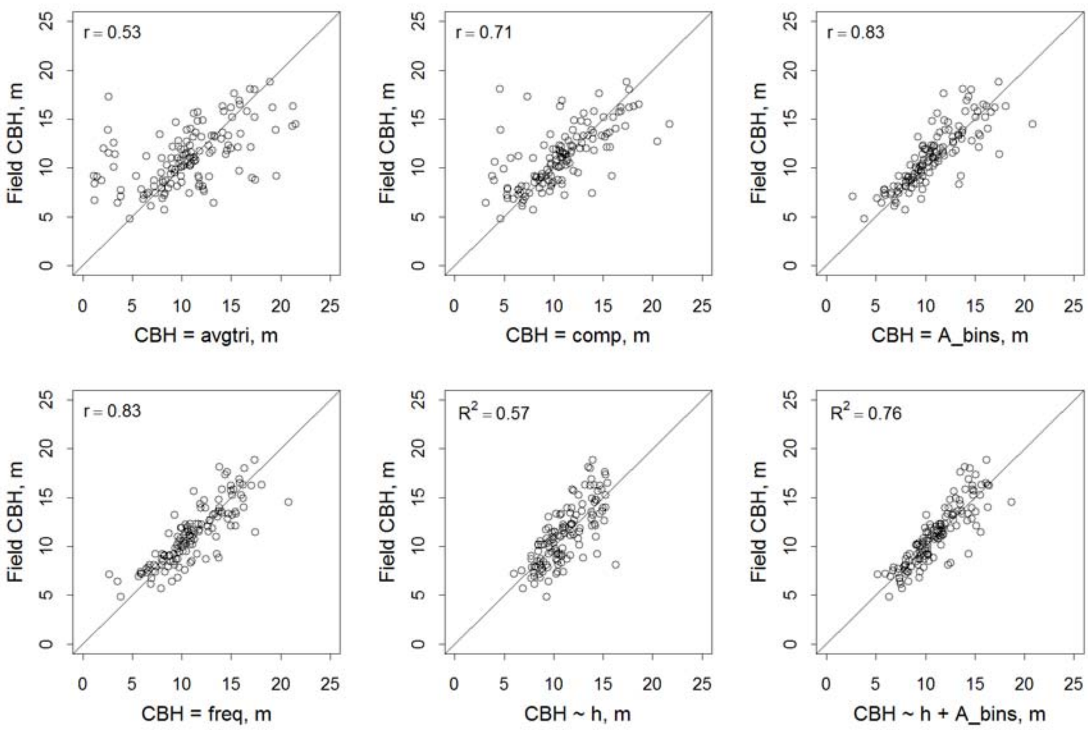

Figure 4: CBH estimates produced by different methods vs. CBH measured in the field.

5. Discussion

It was assumed here that by employing a 3D computational geometry approach new information could be obtained for estimating CBH. The results show that this estimation could be done more accurately using simpler methods. In previous studies, for example, Popescu and Zhao (2008) obtained RMSEs of around 2 m with a very similar sample arrangement. In fact, when predicting from tree height, their results practically equalled those obtained here, although the best result obtained here (RMSE 1.5 m) was slightly more accurate. The present results, in common with those of Popescu and Zhao (2008), show that it is possible to obtain accurate CBH estimates with modest pulse densities. Considering the pulse densities used in estimation, the accuracy of the developed 3D methods is also comparable with figures reported elsewhere (e.g. Pyysalo and Hyyppä 2002, Holmgren and Persson 2004). It should be noted, however, that only Scots pine trees were considered here. In the study of Popescu and Zhao (2008), for example, the difference between pine and deciduous trees was of significance.

An objective for this study was to develop methods that could be adapted to the properties of ALS data. The methods introduced here make use of 3D triangulations of point clouds, which in principle do not need *a priori* knowledge. The main error source consisted of several outliers, so that it is likely that by using higher density data the true discontinuities in the point clouds could be detected more accurately and, thus, the number of potential outliers could possibly be reduced. On the other hand, *a priori* knowledge of the extremes in the data, e.g. not allowing the CBH to be closer than 15% to the tree height at both ends, could possibly reduce the inaccuracy with lower density data. Speaking of pulse-level analysis, a higher density would increase the computational burden, which on the other hand could be constrained by using the efficient structures of computational geometry. Computation was very straightforward with the density considered here, although the code had not yet been optimized at all.

It can also be noted that, interpreted visually, the CBH values measured in the field may differ considerably from the point profile (Figure 3b-c). One reason for this is that the field measurement is made to the base of the lowest branch, whereas the scanner records reflections mainly from the top of the branch. The level of correspondence between the ALS-based CBH estimate and the ground truth is thus dependent on the technical definition of the CBH (see also Solberg *et al*. 2006). In view of this, the 3D measurements may in fact also produce a fair approximation of the living crown, but as the linear models are fitted to the field reference data, they logically generate better results in terms of the RMSE and bias figures.

With respect to the accuracy of certain applications, such as species recognition based on the properties of the point cloud (Holmgren *et al*. 2008; Vauhkonen *et al*. 2008), it may be crucial that the tree crown returns are accurately separated from those representing other trees and the undergrowth. The tree detection rate may vary in mature forests (Persson *et al.* 2002; Maltamo *et al.* 2004), and where the canopy is closed, better results may in general be gained using area-based approaches (e.g. Næsset and Økland 2002). However, given suitable conditions for the single-tree approach, the new methods presented here could be further developed towards 3D segmentation, which would be of help in providing accurate input point data for the applications of the ALS technique.

Acknowledgements

The support by the Academy of Finland (project 12193: "High Resolution Remote Sensing Potential to Measure Single Trees and Site Quality") is gratefully acknowledged. Additionally, the author would like to thank Mr. Jussi Peuhkurinen for having performed the preprocessing of the tree data, and him, Prof. Matti Maltamo, Prof. Timo Tokola, Mr. Malcolm Hicks and two anonymous reviewers for their help in preparing the manuscript.

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, 441-449.

Axelsson, P., 2000. DEM generation from laser scanner data using adaptive TIN models. *International Archives of Photogrammetry and Remote Sensing*, 54, 138-147.

Edelsbrunner, H. and Mücke, E.P., 1994. Three-dimensional alpha shapes. *ACM Transactions on Graphics*, 13, 43-72.

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

Holmgren, J., Persson, Å. 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, 1537-1552.

Hynynen, J., 1995. Predicting tree crown ratio for unthinned and thinned Scots pine stands. *Canadian Journal of Forest Research*, 25, 57-62.

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

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, 319-330.

Maltamo, M., Hyyppä, J. and Malinen, J., 2006. A comparative study of the use of laser scanner data and field measurements in the prediction of crown height in boreal forests. *Scandinavian Journal of Forest Research*, 21, 231-238.

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, 105-115.

Persson, Å., Holmgren, J. and Söderman, U., 2002. Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering & Remote Sensing*, 68(9), 925-932.

Peuhkurinen, J., Maltamo, M., Malinen, J., Pitkänen, J. and Packalén, P., 2007. Preharvest measurement of marked stands using airborne laser scanning. *Forest Science*, 53, 653-661.

Pitkänen, J., Maltamo, M., Hyyppä, J. and 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, 187-191.

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, 767-781.

Pyysalo, U. and Hyyppä, H., 2002. Reconstructing tree crowns from laser scanner data for feature extraction. *ISPRS Commission III Symposium*. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Science, vol. XXXIV, 302-309.

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*, 72(12), 1369-1378.

Vauhkonen, J., Tokola, T., Packalén, P. and Maltamo, M., 2008. *Identification of Scandinavian commercial species of individual trees from airborne laser scanning data using alpha shape metrics*. Submitted manuscript.

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, 451-462.