Effect of canopy structure on the performance of tree mapping methods in urban parks

Tanhuanpää, T., Yu, X., Luoma, V., Saarinen, N., Raisio, J., Hyyppä, J., Kumpula, T., Holopainen, M.

doi:https://doi.org/10.1016/j.ufug.2019.126441

Abstract

Urban forests consist of patches of recreational areas, parks, and single trees on roadsides and other forested urban areas. Large number of tree species and heterogeneous growing conditions result in diverse canopy structure. Large variation can be found both at the level of single tree crowns and canopy characteristics of larger areas. As urban forests are typically managed with small-scale, even tree-level operations, there is a need for detailed forest information. In this study, the effect of varying canopy conditions was tested on nine tree mapping methods. All methods utilized canopy height models (CHM) derived from dense airborne laser scanning data. The mapping methods utilized modified watershed segmentation differed in the intensity of CHM filtering. The performance of mapping methods was compared in different canopy conditions. The results showed considerable variation between the methods when tested in varying canopy conditions. Especially, presence of large broadleaved trees affected the accuracy of detecting individual trees.

Key words: LiDAR, Terrestrial laser scanning, Urban forest, Trees outside forests, Airborne laser scanning

Introduction

There is a growing interest in more detailed information on urban trees. Typically, tree-level data are needed for maintaining the urban tree reserve (allocating tree maintenance etc.) but during recent years, the interest has grown to apply tree-level information for defining the ecosystem services (ES) provided by the trees. Mapping of ES such as storm water catchment or air pollutant removal would likely benefit from detailed tree-level data. Quantifying and mapping the ES provided by urban trees helps in justifying the costs needed for tree management as well as assessing the ES-related benefits in regional level. As manual generation and maintenance of tree-level data is time consuming and thus costly, methods utilizing various high-resolution remote sensing (RS) material have been implemented for detecting and monitoring urban trees.

During the last decade, airborne laser scanning (ALS) datasets have become a common tool for urban mapping and planning (e.g., elevation models and building delineation in urban areas). Therefore, they have a high potential in operative mapping of urban vegetation (Lee et al., 2016; Saarinen et al., 2014; Moskal et al., 2011). The main advantage of ALS data is the ability to capture the vertical structure of forest stands with high detail and accuracy (e.g., White et al., 2015; Suárez et al., 2005). Potentially high spatial resolution, i.e., point density, enables detection of individual tree crowns (e.g., Tanhuanpää et al., 2016; White et al., 2013; Hyyppä and Inkinen, 1999). On the other hand, the spectral resolution has typically been limited to a single wavelength, which makes species interpretation challenging (Kim et al., 2011; Vauhkonen et al., 2009). Hence, ALS datasets have often been accompanied by spectral information from another RS data source (e.g., Alonzo et al., 2014; Koukoulas and Blackburn, 2005).

Delineating individual trees from ALS data has been widely studied, especially in the field of forestry (Mongus and Zalik, 2015; Duncanson et al., 2014; Hyyppä et al., 2012). Individual tree detection (ITD) aims at mapping all trees from a given area and estimating the trees' key attributes. Depending on the attribute of interest

and the materials used, tree attributes can be modeled with combined use of ALS and field data (e.g., Tanhuanpää et al., 2014; Vauhkonen et al., 2010; Persson et al., 2002) or measured directly from ALS data (Tanhuanpää et al., 2015; Vauhkonen, 2010; Popescu and Zhao, 2008). A central hindrance in ALS-based ITD becoming an operational state-of-the-art for producing tree maps for large areas has been the sensitivity of the approach to the spatial structure of the canopy. Depending on the density and spatial distribution of the canopy (i.e., forest structure), ITD methods often suffer from either omission or commission errors (Eysn et al., 2015; Vauhkonen et al., 2011). Omission errors typically occur in multi-layered canopy conditions, where crowns of larger and older trees cover smaller trees. On the other hand, wide and forked crowns of old broadleaved trees are likely to cause commission errors.

In the last decade, the interest towards utilization of point clouds from digital aerial photogrammetry (DAP) in providing information about forest resources has increased (White et al. 2016). This is partly due to the long tradition of utilizing aerial images in forest inventories but mostly because of the cost of DAP has been estimated to be one-third to one-half of the cost of ALS data (White et al. 2013). Although the capability of DAP point clouds in penetrating the canopy is not comparable with ALS (Vastaranta et al. 2013) comparable results between ALS and DAP in identifying individual trees has been obtained (Rahlf et al. 2015, St-Onge et al. 2015). Tanhuanpää et al. (2016) concluded that although plot-level mean height can be similar to ALS-based, identifying individual trees and estimating their height with DAP remains challenging. Additionally, DAP point clouds tend to require an accurate digital terrain model (DTM) that is usually produced with ALS data due to its capability of providing observations also below the canopy.

Several ALS-based tree delineation methods have been published during the last two decades (see review, Lindberg and Holmgren, 2017). Basically, the methods can be divided into two core approaches where the ALS point clouds are analyzed either directly (e.g., Lu et al., 2014; Wang et al., 2008) or by first simplifying them into canopy height models (CHM) that describe the outer envelope of the canopy (e.g., Yu et al., 2011). The direct analysis of point clouds enables the use of all the data in the delineation process, whereas in the CHM approach most of the data are typically discarded. The main benefit from simplifying the data into a CHM is that further processing of the data is computationally less intensive.

Previous studies have shown that the performance of different ITD methods varies between different canopy conditions (e.g., Eysn et al., 2015; Vauhkonen et al., 2011). It has also been reported that, due to their simpler crown structure, coniferous trees can be detected more accurately than deciduous trees (Jing et al., 2012; Koch et al., 2006). However, majority of the ALS-based ITD studies have been conducted in forest environment. Urban forest and trees are often interwoven with buildings, poles, wires, and other built objects common in urban environment. Because of the special qualities, urban areas are challenging environments for classification of trees and other objects (e.g., Rottensteiner et al., 2014). In terms of forest characteristics, urban forests are typically dominated by deciduous tree species and characterized by heterogeneous and fragmented canopy. Both CHM and point cloud-based ITD methods have been studied in urban surroundings. In Holopainen et al. (2013), dense ALS data were utilized in CHM-based method for detecting single trees in urban park area. The automatic method found 65.5% of the trees with DBH of 5 cm or above. Zhang et al. (2015), detected urban park trees directly from ALS point clouds, thus finding 90.6 % of the trees with DBH over 10.16 cm (4 inches). To our best knowledge urban comparisons on ITD methods' performance have not been made. Hence, there is little information available on the effects of canopy characteristics on ITD or the superiority of ITD methodologies in urban surroundings.

Watershed segmentation (WS) and its variations are commonly used in CHM-based approaches for delineating individual tree crowns (Duncanson et al., 2014; Koch et al., 2006, Hyyppä and Inkinen, 1999). In WS, the canopy envelope is typically inverted and treated as the floor of a water body. Local minima (i.e., the sinks of the CHM) are then used as seed points to delineate the basins, i.e., the individual tree crowns. As a tree crown can result in several local minima in a CHM, a central part of the WS process is to reduce the

number of possible seed points. The reduction can be done by using a threshold distance between neighboring seed points (e.g., Yu et al., 2011). The assumption is that the proportion of tree height and crown width is rather constant within the study area. The other option is to smooth the CHM, using either varying or fixed intensity smoothing kernel (e.g., Tanhuanpää et al., 2016; Hyyppä and Inkinen, 1999). For both approaches, finding the optimal number of seed points requires a priori knowledge about the tree crowns' height-width ratio. This is problematic in urban surroundings were high spatial variation of species and growth environments complicates choosing the correct number of seed points.

Earlier studies have pointed out the significance of deciduous trees for the accuracy of ITD (e.g., Koch et al., 2006). The large trees affect the crown delineation in two levels. Firstly, the crown shape affects how the tree itself appears in the data, and secondly the crown size and shape affect the detectability of the neighboring trees. It is typical for old trees to have several distinguishable tops within the crown. This type of crowns often cause commission errors in crown delineation, as each top act as a seed point in segmentation process. Here, threshold values are used to control the number of resulting tree segments. The value is used in determining the minimum distance between detectable trees (Yu et al., 2011). However, the growth pattern of urban open-grown trees has been reported to differ from that of closed canopy conditions, as bigger share of biomass is allocated in tree crown (e.g., Riikonen et al., 2017; Tanhuanpää et al., 2017). In varying light and other growth conditions the relation between tree height and crown width varies considerably, which complicates choosing the correct threshold value. If the value is defined too low, the single crown is split into several crown segments within the delineation process. On the other hand, too high threshold value causes losing the smaller tree crowns. The same logic also affects the smoothing of CHMs. Choosing the level of smoothing according to the large deciduous trees is likely to results in losing the smaller trees (Tanhuanpää et al., 2014).

Because of the heterogeneous nature of urban forest canopy, a widely applicable urban ITD method should be able to adapt to a wide range of variation in crown shapes and sizes. In this study, performance of nine ITD approaches was tested in heterogeneous urban park areas. The aim of the study was threefold.

Firstly, we wanted to investigate the effect of heterogeneous urban canopy conditions on the performance of ITD methods. Assumption was that park areas with lower canopy cover would offer more favorable conditions for ITD than denser areas. Secondly, as large trees have been found to complicate especially urban ITD, we explored their effect on plot-level accuracy of the ITD methods. Thirdly, expecting that none of the methods would outperform others in all canopy conditions, we wanted to define the best methods for different conditions and thus clarify the requirements of a universal ITD methodology oriented to heterogeneous urban park areas.

Material and methods

Study area

The study was conducted in the city of Helsinki, located in southern Finland (60°10'10.27"N, 24°56'7.62"E, Figure 1). The study area consisted of 2000 hectares of parks and recreational urban forests. The area was separated from other urban areas according to a property classification maintained by the City of Helsinki. The canopy structure of the selected areas varied from isolated individual tree crowns to multi-layer recreational areas amidst the city. The species composition ranged from pure deciduous parks to conifer dominated forest-like conditions.

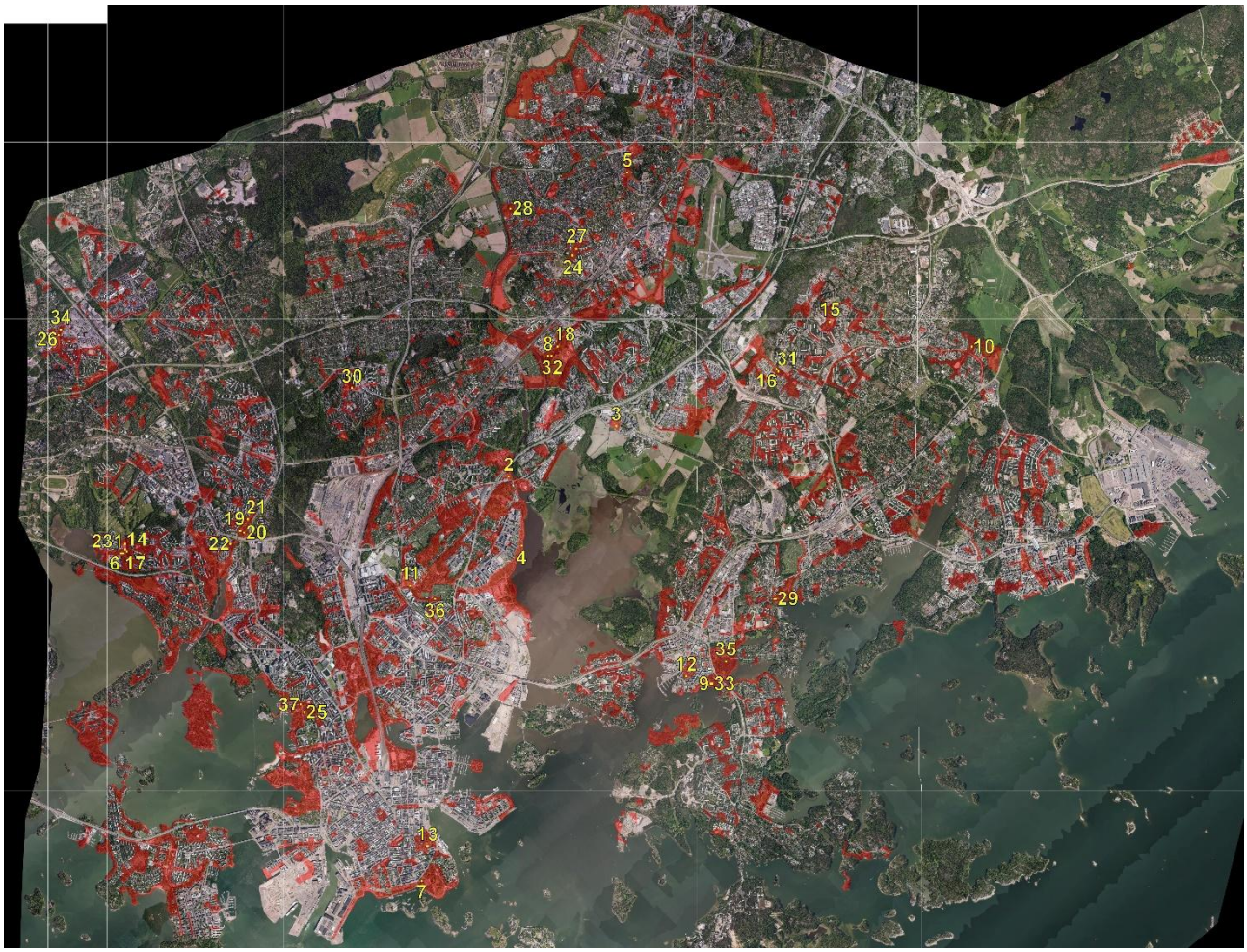


Figure 1: Aerial image of the study area (in red) and the sample plot locations (yellow numbering).

ALS data

The City of Helsinki supplied The ALS data. The data were collected in eight separate flights between May 20th and July 7th, 2015 using Leica ALS-70 HP and ALS-70 CM scanners. Scanning altitude was 500 m above ground level and side overlap between neighboring flight lines was 50%. Pulse density at nadir was 20 p/m^2.

CHM generation

ALS data were simplified into a CHM for the needs of park stratification and tree delineation. All calculations were done in Terra Scan software (Terrasolid, Helsinki, Finland). The CHM was created by first triangulating a digital elevation model (DEM) (see, Axelsson, 2000) and generating a digital surface model (DSM), representing the outer envelope of the canopy, from the highest echoes. The cell sizes were 1.0 m x 1.0 m and 0.5 m x 0.5 m for the DEM and DSM, respectively. Smaller cell size was used for DSM for capturing the fine-scale variation in tree crowns. The CHM was generated by subtracting the DEM values from the values of DSM. The spatial resolution of the final CHM was 0.5 m. ALS-based CHM is visualized in Figure 2.

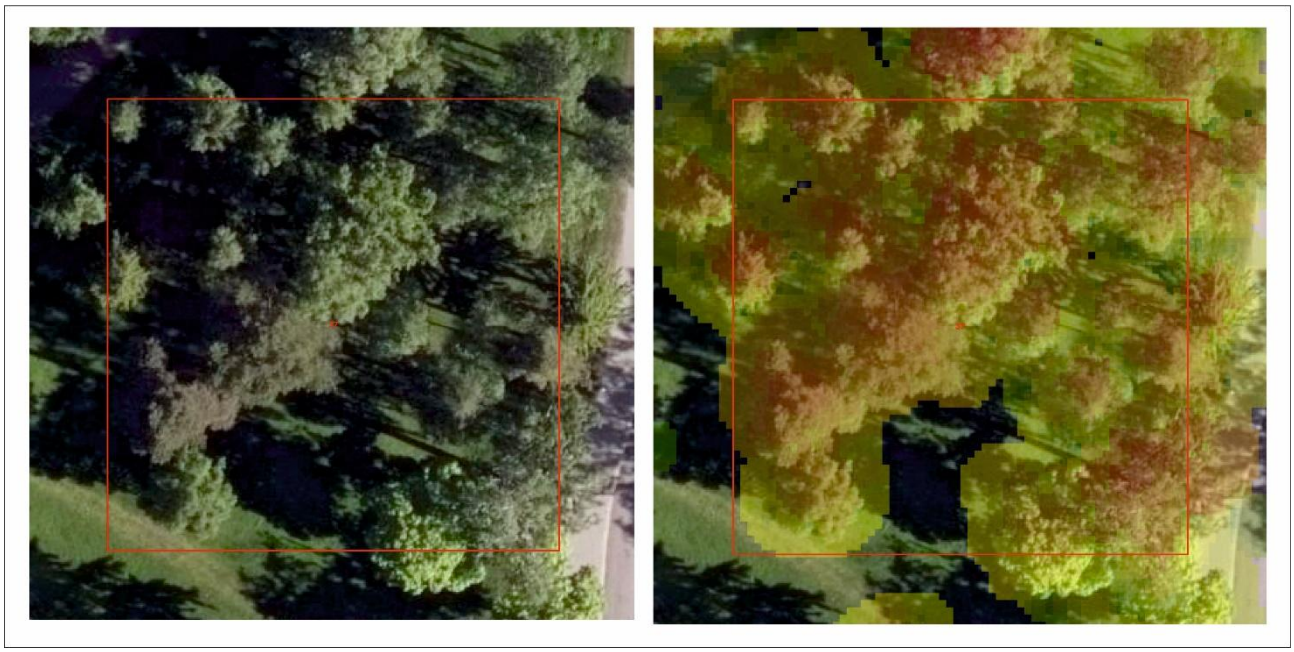


Figure 2: Plot number 37 visualized in an aerial image (left) and an ALS-based canopy height model (right) over an aerial image.

Stratification of park areas

For capturing the variation within the park areas, the whole study area was first divided in three strata in terms of canopy cover and mean height. As open grown trees have typically wider crowns than forest trees, we used 32 m by 32 m grid for the stratification. This resulted in four times larger plots compared to operational forest attribute mapping carried out by the Finnish Forest Centre (Suomen metsäkeskus). The grid was placed over the park areas and each grid cell fully inside park area was characterized by ALS-derived mean height (H_{mean}) and canopy cover-% (CC). Both figures were achieved by applying simplified WS (i.e., all peaks in the CHM were considered seed points) with cut height of 2 m (i.e., CHM values under 2 m were ignored) to the unfiltered ALS-based CHMs. CC was calculated as the proportional area of canopy polygons inside each 32 m by 32 m grid cell. Similarly, H_{mean} was calculated for each grid cell as a mean of height maxima of the canopy polygons. Hence, H_{mean} represents the height of vegetation peaks in CHM, not the mean height of ALS points. The stratification was used for allocating field plots over the study area proportionally to the total area of each strata (Table 1).

Table 1: Definition and key characteristics of the field plots within the park strata. n_{plot} stands for the number of plots, n_{tree} for the total number of trees, CC_{mean} for the ALS-derived mean canopy cover, H_{mean} for the ALS-derived mean height, DBH_{mean} for the mean, DBH_{min} for the minimum, and DBH_{max} for the maximum field-measured DBH within each stratum.

Stratum	Definition	n_{plot}	n_{tree}	CC_{mean}	H_{mean}	DBH_{mean}	DBH_{min}	DBH_{max}
1	CC < 50 %	12	82	26 %	8.0 m	20.6	6.0	108.2
2	CC > 50 % and H_{mean} < 15 m	15	453	74 %	10.5 m	19.6	6.0	65.1
3	CC > 50 % and H_{mean} > 15 m	10	277	83 %	18.0 m	25.6	6.2	116.2
Total	All plots	37	812	63 %	12.2 m	21.7	6.0	116.2

In terms of individual trees' size and shape, the variation between the three strata was substantial. Stratum 1 (Figure 3) consisted largely of open or semi-open grown trees, whereas most trees in strata 2 and 3 (Figures 4 and 5) were growing under significant light competition. Although H_{mean} in stratum 1 was the lowest within the study, it included some of the thickest trees.

Figure 3: Two 32 m x 32 m sample plots describing the canopy conditions in stratum 1. Yellow circles represent the position and relative size (DBH) of the field reference.

Figure 4: Two 32 m x 32 m sample plots describing the canopy conditions in stratum 2. Yellow circles represent the position and relative size (DBH) of the field reference.

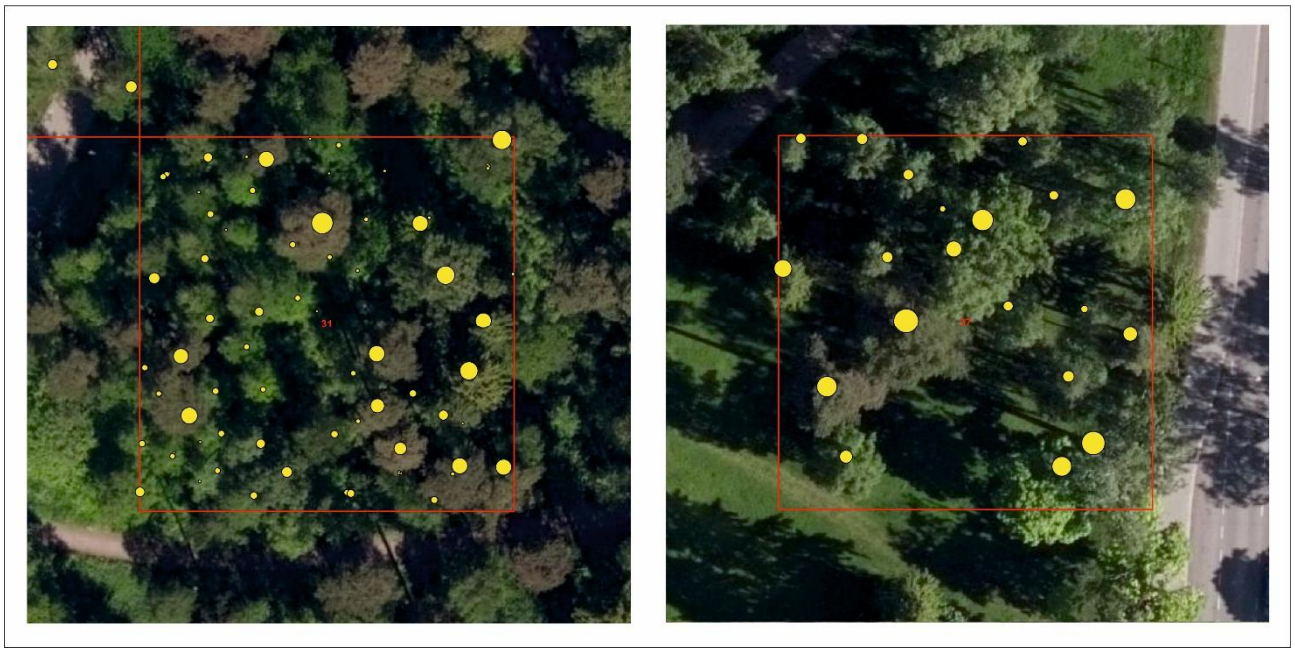


Figure 5: Two 32 m x 32 m sample plots describing the canopy conditions in stratum 3. Yellow circles represent the position and relative size (DBH) of the field reference.

Field measurements

In total, 37 grid cells were field measured as sample plots. All plots were scanned in June-July 2016 using Leica HDS 6100 (Leica Geosystems AG, Heerbrugg, Swizerland) terrestrial laser scanner (TLS). The field crew scanned each plot from five locations and co-registered the scans using spherical reference targets. The resulting point clouds were georeferenced by positioning the scan locations with Trimble R8s GNSS receiver (Trimble Navigation Ltd., Sunnyvale, CA, USA). Following the work of Saarinen et al. (2014), the georeferenced point clouds were used for creating tree maps for each field plot. The initial tree maps were first generated from TLS data by excerpting a 20 cm slice from the point clouds at the height of 1.3 m. In case of multi-stem trees, all stems that had branched below 1.3 m (i.e., DBH could be measured) were considered individual trees. The TLS-derived tree maps were field-checked for error of omission and commission in April 2017. In total, there were 812 trees on the field plots. Diameter at breast height (DBH) was determined for each tree by measuring them from two directions with a caliper. All trees with DBH of 6 cm or above were recorded. Based on total basal area, the most common tree species in stratum 1 were Salix sp. group (28%), Acer platanoides (25%), and Betula pendula (17%). The most common tree species for stratum 2 were Betula pendula (21%), Salix sp. group (19%), and Ulmus glabra (12%) and for stratum 3 Betula pendula (25%), Quercus robur (23%), and Acer platanoides (11%).

Tree delineation methods

Altogether, nine different crown delineation methods were tested for detecting the trees from ALS data (Table 2). All the applied methods were based on WS of smoothed CHMs.

Table 2: Description of the tree delineation methods tested in the study. WS stands for watershed segmentation.

Method	Filter description	Relative weight of center cell	Kernel size: pixels (effective)	Delineation method
$G_{0.2}$	Gaussian, σ = 0.2	84.4%	3 x 3 (1.5 m x 1.5 m)	simple WS
$G_{0.4}$	Gaussian, σ = 0.4	24.9%	5 x 5 (2.5 m x 2.5 m)	simple WS

$G_{0.5}$	Gaussian, σ = 0.5	15.9%	7 x 7 (3.5 m x 3.5 m)	simple WS
$G_{0.7}$	Gaussian, σ = 0.7	8.1%	9 x 9 (4.5 m x 4.5 m)	simple WS
$G_{0.9}$	Gaussian, σ = 0.9	4.9%	11 x 11 (5.5 m x 5.5 m)	simple WS
$G_{1.0}$	Gaussian, σ = 1.0	4.0%	13 x 13 (6.5 m x 6.5 m)	simple WS
F_1	Fixed kernel	25.0%	3 x 3 (1.5 m x 1.5 m)	simple WS
F_2	Fixed kernel	42.9%	3 x 3 (1.5 m x 1.5 m)	simple WS
G_{adapt}	Gaussian, σ = 0.2 – 2.0	4.0% – 96%	3 x 3 – 13 x 13 (1.5 m x 1.5m - 6.5 m x 6.5 m)	marker controlled WS

The first eight of the methods presented in Table 2 (i.e., $G_{0.2}$ – F_2) utilized WS without any pre-assumptions on crown size. Here, the segmentation process was changed only by using different filter when smoothing the CHM prior to applying WS. A Gaussian filter was tested using six different values of parameter σ, determining the size and weights of the filter ($G_{0.2}$ – $G_{1.0}$). The size of Gaussian filters ranged from 1.5 to 6.5 meters and the relative weight of the center cell varied from 84.5% to 4%, respectively. Methods F_1 and F_2 were adapted from earlier studies (Hyyppä et al. 2001 and Kaartinen et al. 2012, respectively). They can be considered as modifications of $G_{0.2}$ filter giving the center cell a lower relative weight with the same spatial extent.

G_{adapt} was the most sophisticated method tested in the study. It used an automatic algorithm of local maxima finding with a moving window of varying size, followed by crown delineation. The details of the algorithm are: firstly, the CHM was smoothed with a Gaussian filter with varying window size and value of σ was adjusted according to the CHM values, so that the higher the CHM value was the higher the value of σ. All in all, the widow size varied between 1.5 m x 1.5 m and 6.5 m x 6.5 m; Secondly, local maxima were searched for and considered as tree tops. Relying on the presumed relationship between crown width and tree height, the search radius was determined with Equation 1, which is adapted from an earlier study (Pitkänen et al, 2004)

$$C_w = 1.2 + 0.16h, \quad (1)$$

were C_w is crown width and h tree height in meters. Finally, single tree crowns were delineated using marker controlled WS with the tree tops found in the previous step as control markers. Afterwards, tree locations were obtained from the CHM pixels having the highest value within each tree segment.

Modelling of DBH

DBH was modeled using non-parametric Random Forest (RF) methodology (see, e.g., Yu et al., 2011). The RF method classifies target observations by number of randomly selected predictors. Best predictors are chosen by repeating the classification procedure several hundred times. In this study, sample trees for the training data were selected among the field measured trees. In this step most of the small trees (DBH < 20 cm) standing next to larger ones were discarded because their crowns were not represented properly by the crown segments. Depending on the tested ITD method, the number of trees in the training set varied between 356 and 442. The variation between methods was caused from matching the predicted tree crowns with field measurements. Prediction models for DBH were created using 26 tree-level ALS-derived height and density metrics as predictors. The DBH estimates were obtained by repeating the classification procedure 300 times.

Accuracy assessment

Stem count, minimum (DBH_{min}), maximum (DBH_{max}), and mean DBH (DBH_{mean}) were used as criteria in defining the plot-level accuracy for each method. For DBH, minimum and maximum were included to show

whether there are differences in the detection of small and large trees, whereas mean characterized whole plots. For each criterion, best-case accuracies were determined for each stratum. All assessments were made first at plot-level and combined into stratum-level results. Estimates were compared using root mean square error (RMSE) (Equation 2), bias (Equation 3) and their relative values (Equations 4 and 5).

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

$$bias = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{n}, \quad (3)$$

$$RMSE_{rel} = \frac{RMSE}{\bar{y}}, \quad (4)$$

$$bias_{rel} = \frac{bias}{\bar{y}}, \quad (5)$$

where *n* is the number of trees, *y* the observed value, $\hat{y}$ the modeled value, and $\bar{y}$ the mean of observations.

In addition, DBH estimates were transformed into DBH distributions with 3 cm bin size. DBH distributions were compared to the field-measured distributions using Reynold's Error index (Reynolds et al., 1988) (Equation 6) and relative error index (Packalén and Maltamo, 2008) (Equation 7). As both the DBH and the tree count are present in the approach, the distribution-based methods illustrate the overall correctness of plot-level observations.

$$EI = \sum_{i=1}^{k} w_i \left| f_i - \hat{f}_i \right|, \quad (6)$$

$$EI_{rel} = \sum_{i=1}^{k} 0.5 \left| \frac{f_i}{N} - \frac{\hat{f}_i}{\hat{N}} \right|, \quad (7)$$

where *k* is the number of DBH classes, w_i is the weight of class *i*, f_i is the observed tree count in class *i*, $\hat{f}_i$ is the predicted tree count in class *i*, N is the observed, and $\hat{N}$ the predicted tree count in the plot. Here, weight parameter w_i was set to 1, giving equal weight for all DBH classes. *EI* shows the absolute difference between the observed and predicted distribution, whereas EI_{rel} is its modification and can be interpreted as a proportion of incorrect observations in terms of the shape of the reference distribution. In this study, *EI* was used in comparing the within stratum results, whereas EI_{rel} was utilized when comparing the performance of each method in different strata.

Results

Tree count

The performance of all ITD methods showed considerable variation between the three strata. Depending on the stratum and the ITD method used, the $RMSE_{rel}$ of tree count varied between 0.36 – 1.79 and $bias_{rel}$ between -1.23 – 0.63 (Table 3).

In terms of $RMSE_{rel}$ stratum 2 had the highest accuracies. For $bias_{rel}$, stratum 3 showed the best results. Methods $G_{0.5}$ and F_1 resulted in the best accuracies in all three strata in nearly all criteria. An exception to this were RMSE and $RMSE_{rel}$ in stratum 2, were G_{adapt} resulted in the best results.

Table 3. RMSE and bias for the number of detected trees in all three strata. Column *Total* shows the overall accuracy without stratification. The accuracies of the best methods within each stratum are highlighted.

	Stratum 1	Stratum 2	Stratum 3	Total

	RMSE, trees/plot	*RMSE$_{rel}$*	*bias, trees/plot*	*bias$_{rel}$*	*RMSE, trees/plot*	*RMSE$_{rel}$*	*bias, trees/plot*	*bias$_{rel}$*	*RMSE, trees/plot*	*RMSE$_{rel}$*	*bias, trees/plot*	*bias$_{rel}$*	*RMSE, trees/plot*	*RMSE$_{rel}$*	*bias, trees/plot*	*bias$_{rel}$*
$G_{0.2}$	12.3	1.79	-8.4	-1.23	20.9	0.69	-14.1	-0.47	24.1	0.78	-15.3	-0.50	19.3	0.81	-11.9	-0.50
$G_{0.4}$	5.3	0.78	-3.2	-0.46	12.9	0.43	7.1	0.23	19.3	0.63	7.4	0.24	14.8	0.62	4.8	0.20
$G_{0.5}$	3.7	0.54	-1.2	-0.17	15.2	0.50	10.9	0.36	21.0	0.68	11.8	0.38	16.7	0.70	8.3	0.35
$G_{0.7}$	4.7	0.69	1.6	0.23	18.5	0.61	15.3	0.51	23.8	0.77	15.7	0.51	19.1	0.80	12.0	0.50
$G_{0.9}$	5.3	0.77	1.7	0.24	19.2	0.64	15.7	0.52	24.0	0.78	16.2	0.53	19.8	0.83	12.4	0.52
$G_{1.0}$	6.2	0.90	3.0	0.44	22.9	0.76	18.9	0.63	27.1	0.88	19.4	0.63	22.2	0.93	14.9	0.63
F_1	6.3	0.92	-3.9	-0.57	11.1	0.37	1.9	0.06	18.6	0.60	1.0	0.03	13.7	0.57	0.8	0.03
F_2	5.8	0.84	-3.7	-0.54	13.0	0.43	5.3	0.18	19.7	0.64	4.9	0.16	14.8	0.62	3.3	0.14
G_{adapt}	4.7	0.69	-2.4	-0.35	11.0	0.36	3.2	0.11	21.2	0.69	12.3	0.40	15.0	0.63	4.8	0.20

When comparing the accuracy of different methods, method F_1 resulted in the best results overall. F_1 had the lowest RMSE and bias within stratum 3 (18.6 and 1.0, respectively) and within the unstratified data (column *Total*, 13.7 and 0.8, respectively). Overall, the methods performing the best resulted in low bias in all strata. However, the RMSE remained rather high for all methods and strata.

DBH estimates

As the tree candidates were not matched to the field reference, the accuracy of DBH estimates was investigated at plot-level. The accuracies of estimated minimum, maximum, and mean DBHs were investigated and the plot-level results are presented in Table 4. The accuracy of minimum and maximum DBH illustrates the methods' ability to detect the smallest and the largest trees, whereas mean DBH demonstrates the methods' ability to capture the overall structure of the plot.

Table 4. Accuracy of plot-level estimates of minimum, maximum, and mean DBH. Column Total shows the accuracy for the entire data set. The bolded figures highlight the most accurate methods within each stratum for each DBH measure (min, max, and mean).

Method	DBH measure	*Stratum 1*				*Stratum 2*				*Stratum 3*				*Total*			
		RMSE, cm	$RMSE_{rel}$	*bias, cm*	$bias_{rel}$	*RMSE, cm*	$RMSE_{rel}$	*bias, cm*	$bias_{rel}$	*RMSE, cm*	$RMSE_{rel}$	*bias*	$bias_{rel}$	*RMSE, cm*	$RMSE_{rel}$	*bias, cm*	$bias_{rel}$
$G_{0.2}$	*min*	29.3	4.88	11.9	1.98	2.2	0.37	-1.5	-0.25	6.3	1.01	-2.6	-0.41	17.0	2.84	2.5	0.42
	max	16.8	0.16	1.2	0.01	19.8	0.30	-8.9	-0.14	16.9	0.15	0.3	0.00	17.9	0.15	-3.2	-0.03
	mean	23.3	1.13	8.7	0.42	3.9	0.20	-2.8	-0.14	11.0	0.43	-8.4	-0.33	14.8	0.70	-0.8	-0.04
$G_{0.4}$	*min*	27.2	4.54	10.4	1.73	4.2	0.71	-3.6	-0.59	8.0	1.29	-5.5	-0.89	16.3	2.72	0.3	0.05
	max	17.0	0.16	2.7	0.02	6.6	0.10	1.5	0.02	21.3	0.18	3.8	0.03	15.0	0.13	2.1	0.02
	mean	22.4	1.09	8.2	0.40	4.4	0.23	-3.6	-0.18	11.9	0.47	-9.8	-0.38	14.7	0.69	-1.7	-0.08
$G_{0.5}$	*min*	27.6	4.61	9.7	1.61	6.4	1.07	-5.1	-0.85	9.3	1.50	-5.3	-0.86	17.0	2.83	-0.5	-0.09
	max	17.7	0.16	1.2	0.01	5.6	0.09	3.2	0.05	17.8	0.15	8.2	0.07	13.9	0.12	3.8	0.03
	mean	23.4	1.13	7.5	0.37	5.5	0.28	-4.4	-0.22	10.0	0.39	-7.5	-0.29	14.9	0.70	-1.6	-0.08
$G_{0.7}$	*min*	27.4	4.56	8.7	1.44	7.9	1.32	-6.4	-1.06	9.9	1.60	-8.5	-1.37	17.2	2.87	-2.2	-0.36
	max	17.6	0.16	7.4	0.07	4.3	0.07	2.3	0.04	19.3	0.17	12.9	0.11	14.1	0.12	6.6	0.06
	mean	23.8	1.15	8.6	0.42	6.9	0.35	-6.0	-0.30	12.4	0.48	-9.3	-0.36	15.8	0.74	-2.4	-0.11
$G_{0.9}$	*min*	26.8	4.47	4.9	0.82	11.5	1.91	-10.7	-1.78	20.8	3.36	-17.2	-2.78	19.9	3.32	-7.2	-1.21
	max	10.2	0.09	0.0	0.00	8.2	0.13	2.9	0.05	17.6	0.15	6.4	0.06	11.7	0.10	2.8	0.02
	mean	18.1	0.88	3.6	0.18	9.3	0.48	-8.6	-0.44	17.8	0.70	-14.5	-0.57	15.2	0.72	-6.5	-0.30
$G_{1.0}$	*min*	22.5	3.75	4.5	0.75	11.7	1.96	-10.6	-1.77	12.3	1.99	-10.1	-1.64	16.2	2.69	-5.7	-0.94
	max	19.0	0.18	6.9	0.06	9.5	0.15	4.9	0.08	16.4	0.14	10.8	0.09	14.8	0.13	6.9	0.06
	mean	20.2	0.98	6.1	0.30	8.3	0.42	-7.4	-0.38	15.2	0.59	-11.4	-0.45	15.1	0.71	-4.4	-0.21
F_1	*min*	28.7	4.78	10.4	1.73	3.0	0.51	-2.7	-0.46	8.2	1.32	-5.4	-0.87	17.0	2.84	0.7	0.11
	max	14.0	0.13	3.7	0.03	7.4	0.11	5.1	0.08	17.0	0.15	8.5	0.07	12.5	0.11	5.4	0.05
	mean	23.8	1.15	8.9	0.43	3.3	0.17	-2.6	-0.13	8.5	0.33	-5.9	-0.23	14.6	0.69	-0.1	0.00
F_2	*min*	27.6	4.59	10.4	1.73	3.8	0.63	-3.3	-0.55	7.8	1.26	-4.2	-0.68	16.4	2.73	0.8	0.14
	max	16.9	0.16	0.4	0.00	7.9	0.12	3.3	0.05	15.9	0.14	6.6	0.06	13.5	0.12	2.8	0.02
	mean	23.7	1.15	8.5	0.41	3.8	0.19	-3.0	-0.15	11.2	0.44	-8.4	-0.33	15.2	0.71	-1.0	-0.05
G_{adapt}	*min*	26.8	4.47	8.9	1.48	4.3	0.72	-4.0	-0.67	7.9	1.28	-4.5	-0.73	16.0	2.67	0.1	0.02
	max	12.9	0.12	3.9	0.04	9.5	0.15	8.2	0.13	20.9	0.18	15.3	0.13	14.1	0.12	8.6	0.07
	mean	23.6	1.14	9.1	0.44	2.2	0.11	-0.7	-0.04	7.5	0.30	-3.0	-0.12	14.0	0.66	1.8	0.08

In stratum 1, all tested methods underestimated the minimum DBH and method $G_{1.0}$ resulted in the most accurate results. The relative bias remained high for all methods, varying from 0.75 – 1.98. The relative RMSE ranged between 3.75 and 4.88. For maximum and mean diameters, the results were more accurate, method $G_{0.9}$ giving the most accurate results. The relative bias of maximum and mean diameters ranged from 0.00 – 0.07 and 0.18 – 0.44, respectively. The best results in stratum 1 were achieved by using rather heavy filtering of CHM, i.e., using methods $G_{0.9}$ and $G_{1.0}$.

As the minimum diameter was systematically underestimated in stratum 1, the situation was the exact opposite in stratum 2. All methods overestimate the minimum DBH, method $G_{0.2}$ giving the most accurate results. Overall, the relative bias of minimum DBH in stratum 2 ranged from -0.25 to -1.78. Relative RMSE ranged between 0.37 and 1.96. Except for $G_{0.2}$, maximum DBH was underestimated by all methods. Method $G_{0.4}$ delivered the lowest $bias_{rel}$ (0.02) but method $G_{0.7}$ resulted in the smallest $RMSE_{rel}$ (0.07). Relative bias

and RMSE for maximum DBH ranged from -0.14 to 0.13 and from 0.07 to 0.30, respectively. For mean DBH, method G_{adapt} resulted in the most accurate results ($bias_{rel}$ = -0.04 and $RMSE_{rel}$ = 0.11). Overall, all methods overestimated the mean DBH. The relative bias and RMSE for mean DBH ranged from -0.44 to -0.04 and from 0.11 to 0.48, respectively.

Also, in stratum 3, all methods overestimated the minimum DBH. $G_{0.2}$ was the most accurate method for minimum DBH. The relative bias ranged from -2.78 to -0.41 and relative RMSE from 1.01 to 3.36. For maximum DBH method $G_{0.2}$ resulted in the lowest bias ($bias_{rel}$ = 0.00) and F_2 the lowest RMSE ($RMSE_{rel}$ = 0.14). Overall, all methods underestimated the maximum DBH also in stratum 3. Relative bias and RMSE ranged from 0.00 to 0.13 and 0.14 to 0.18, respectively. The mean DBH was best captured using method G_{adapt} ($bias_{rel}$ = -0.12 and $RMSE_{rel}$ = 0.30). All methods overestimated the mean DBH. The relative bias and RMSE for mean DBH ranged from -0.57 to -0.12 and from 0.30 to 0.70, respectively.

When the study area was not stratified, G_{adapt} was the most accurate method for determining minimum DBH ($bias_{rel}$ = 0.02 and $RMSE_{rel}$ = 2.67). The relative bias and RMSE ranged from -1.21 to 0.42 and from 2.67 to 3.32, respectively. For maximum DBH, all methods except $G_{0.2}$ resulted in slight overestimates. $G_{0.4}$ ($bias_{rel}$ = 0.02) and $G_{0.9}$ ($RMSE_{rel}$ = 0.10) resulted in the most accurate estimates, whereas the relative bias and RMSE overall ranged from -0.03 to 0.07 and from 0.10 to 0.15, respectively. The most accurate methods for estimating mean DBH were F_1 ($bias_{rel}$ = 0.00) and G_{adapt} ($RMSE_{rel}$ = 0.66). The relative bias and RMSE of mean DBH estimates ranged from -0.30 to 0.08 and from 0.66 to 0.74, respectively.

When comparing the accuracy of the DBH estimates from the three strata, stratum 1 clearly stands out. With few exceptions, in stratum 1 the $RMSE_{rel}$ is at its highest for all methods. Also, it is only in stratum 1 that the methods utilizing heavy filtering (i.e., $G_{0.9}$ and $G_{1.0}$) outperform the others. In strata 2 and 3, heavy filtering typically resulted in lowest accuracies in all three DBH estimates. For these two strata, $G_{0.2}$ was found the most accurate for minimum DBH, which reflects the methods sensitivity to small scale variation in CHM. For the same two strata, G_{adapt} was found the best method for acquiring the mean DBH. For maximum DBH, the differences between methods were relatively small and none of the methods clearly stood out. Except for $G_{0.2}$, all methods underestimated the maximum DBH in all three strata.

DBH distributions

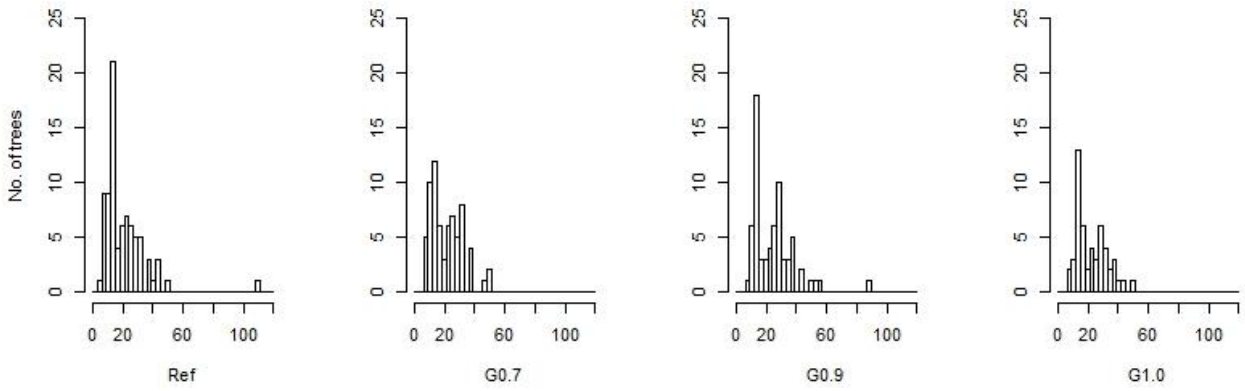
Error indices describing the correctness of the ITD-derived DBH distributions are presented in Table 5. As the tree count varies between the strata, *EI* is used to compare the performance of the methods within each stratum, whereas EI_{rel} is used for assessing the performance of each method over all strata.

Table 5. Error index (*EI*) and Relative error index (EI_{rel}) of the ITD-derived DBH distributions in all three strata. Column *Total* shows the overall accuracy without stratification. Best *EI* within each stratum are bolded.

Method		Stratum			
		1	*2*	*3*	*Total*
$G_{0.2}$	*EI*	160	506	428	1094
	EI_{rel}	0.590	0.386	0.573	0.501
$G_{0.4}$	*EI*	116	339	302	757
	EI_{rel}	0.594	0.409	0.616	0.522
$G_{0.5}$	*EI*	98	350	276	724
	EI_{rel}	0.607	0.499	0.602	0.561
$G_{0.7}$	*EI*	**75**	343	274	**692**
	EI_{rel}	0.644	0.502	0.655	0.588
$G_{0.9}$	*EI*	82	356	265	703
	EI_{rel}	0.715	0.543	0.673	0.633
$G_{1.0}$	*EI*	82	373	265	720
	EI_{rel}	0.725	0.611	0.681	0.666
F_1	*EI*	115	378	343	836
	EI_{rel}	0.621	0.440	0.613	0.543
F_2	*EI*	116	**338**	328	782
	EI_{rel}	0.563	0.402	0.618	0.509
G_{adapt}	*EI*	89	374	**257**	720
	EI_{rel}	0.475	0.439	0.548	0.478

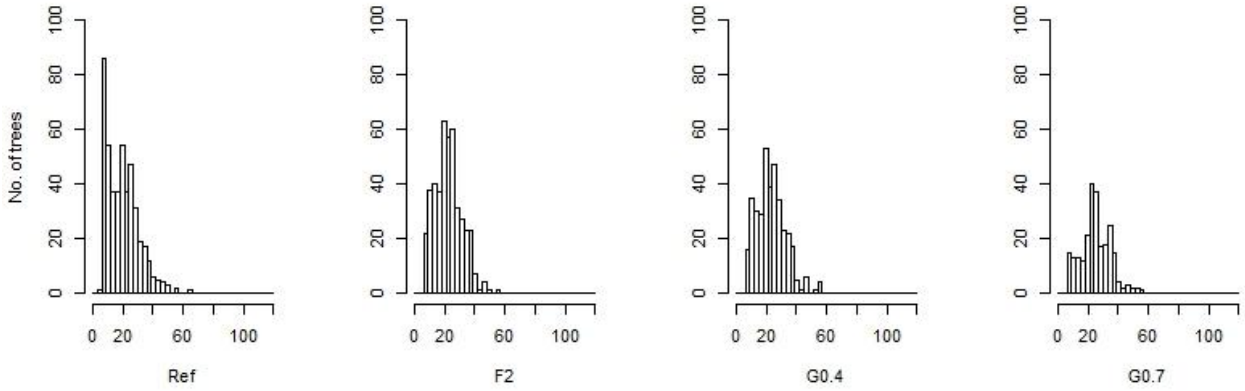
When examining *EI* within stratum 1, method $G_{0.7}$ results in the lowest number of errors in classification (75 misclassifications) followed by $G_{0.9}$ and $G_{1.0}$ (82 misclassifications). Large filter size seemed to increase the accuracy in terms of correctness of DBH distributions. In stratum 2, method F_2 had the lowest number of classification errors. However, the differences between $G_{0.4}$, $G_{0.7}$, and F_2 are small (339, 343, and 338, respectively). Unlike in stratum 1, there is no clear pattern between filter size and classification accuracy. In stratum 3, method G_{adapt} results in the best classification accuracy (257). Still, methods' $G_{0.9}$ and $G_{1.0}$ classification accuracies (265 for both) are close to that of G_{adapt}. Hence, it seems that, also in stratum 3, methods with larger filter size result in more accurate estimates for plot-level DBH distributions. Looking at column *Total*, the effect of different methods is clear. The intensity of CHM filtering increases gradually for methods $G_{0.2}$-$G_{1.0}$, which decreases the number of detected tree candidates. Method $G_{0.2}$ strongly overestimates, whereas $G_{0.4}$-$G_{1.0}$ underestimate the tree count. The best results were achieved with method $G_{0.7}$.

Differences between the three best methods within each stratum were visualized through DBH distributions. In stratum 1 (Figure 6), the methods with the best EI values resulted in two-peaked DBH distributions, which resembled that of the field reference. However, none of the three methods were able to detect the highest DBH values. Also, G_{adapt} overestimated the tree count both in the smallest and the largest DBH classes. In stratum 2 (Figure 7), all three methods underestimated the number of trees with DBH <10 cm, whereas method F2 overestimated the number of trees with DBH between 20-30 cm. The results are similar in stratum 3 (Figure 8), where the number of trees with DBH <20 cm is underestimated. Considering all three strata, the amount of detected small trees decreased as canopy cover and plot mean height increased.



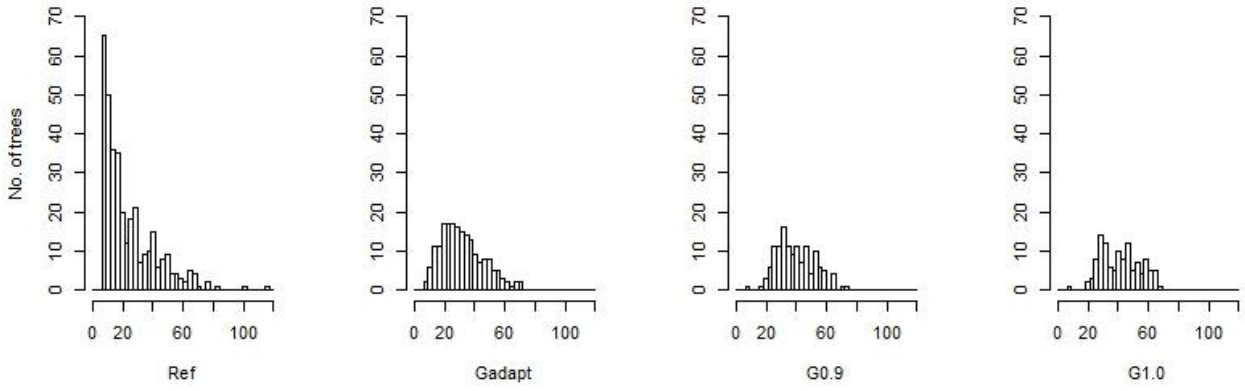
354

355 Figure 6. DBH distributions of the field reference and three best performing methods ($G_{0.7}$, $G_{0.9}$, and G_{adapt}) in stratum 1.



356

357 Figure 7. DBH distributions of the field reference and three best performing methods (F_2, $G_{0.4}$, and $G_{0.5}$) in stratum 2.



358

359 Figure 8. DBH distributions of the field reference and three best performing methods (G_{adapt}, $G_{0.9}$, and $G_{1.0}$) in stratum 3.

360

361 *Overall performance of the best methods*

The best performing methods were selected on grounds of EI as it summarized the predicted information of number of trees and DBHs. Table 6 summarizes the accuracy for the best methods in terms of EI, DBH, and number of trees.

Table 6. A summary of the best-performing ITD methods' accuracies for each stratum.

Stratum	Best method (in terms of EI)	Number of trees		DBH_{min}		DBH_{max}		DBH_{mean}		EI	EI_{rel}
		$RMSE_{rel}$	$bias_{rel}$	$RMSE_{rel}$	$bias_{rel}$	$RMSE_{rel}$	$bias_{rel}$	$RMSE_{rel}$	$bias_{rel}$		
1	G_{07}	0.69	0.23	4.56	1.44	0.16	0.07	1.15	0.42	75	0.644
2	F_2	0.43	0.18	0.63	-0.55	0.12	0.05	0.19	-0.15	338	0.402
3	G_{adapt}	0.69	0.40	1.28	-0.73	0.18	0.13	0.30	-0.12	257	0.548
all	G_{07}	0.80	0.50	2.87	-0.36	0.12	0.06	0.74	-0.11	692	0.588

Discussion

In this study, we investigated the effect of heterogeneous canopy conditions on the performance of nine CHM-based ITD methods in urban parks. To do so, the park areas were stratified into three groups according to ALS-derived canopy mean height and canopy cover. In addition, the effect of large trees on plot-level accuracy of ITD methods was explored. Accuracy was evaluated through the correctness of plot-level stem count and three DBH-related criteria (DBH_{min}, DBH_{max}, and DBH_{mean}). All criteria were examined separately to define the best-case methods for each stratum and criterion. Finally, special requirements of urban ITD methodology were brought up by defining the best ITD methods for each stratum, by using Reynold's Error index. As expected, the canopy conditions had a considerable effect on the performance of ITD methods. Contrary to our expectations, areas with lowest canopy cover turned out as the most demanding for ITD approach. This was largely due to heterogeneity of the stand structure, but also large open-grown trees. Finally, we demonstrated that selecting stratum-specific ITD methods improves the overall results of urban ITD.

Although the estimation of DBH was not in the core of this study, the issue is addressed here because DBH was a central parameter when assessing the accuracy of ITD methods. The light conditions varied from nearly open areas to closed canopy, which is known to affect the growth pattern of trees (e.g., Mäkinen, 2003; Niemistö, 1995). Variation in growth pattern weakens the relation between DBH and height and thus lowers the accuracy of DBH estimates (Tanhuanpää et al., 2017). Also, as the species of the detected trees were not identified, the species information could not be used in DBH modelling. As the growth patterns differ between different tree species, including species information to the process would likely enhance the overall accuracies. Still, as the DBH modelling procedure was the same for all methods, we presume that the method-related variation of DBH estimates remained fairly constant in all ITD methods tested, and that the main differences between methods resulted from differences in crown delineation. Considering the estimates for DBH_{min} and DBH_{max}, the number of trees is rather small, which decreases the robustness of these two figures. Since only one tree per plot was selected as the smallest or the biggest, the figures rely on 10-15 trees (i.e., one tree per plot) in stratum-level and 37 trees when no stratification was done. As all trees in each plot are used to achieve DBH_{mean}, this figure is more robust for describing the methods' performance.

As expected, the performance of ITD methods varied substantially between the three park strata. Table 6 shows how the relative RMSE of stem count and DBH-related accuracy indicators increase considerably while shifting from stratum 2 to stratum 3, towards denser and higher canopy. As similar results have been reported before (e.g., Koch et al., 2006), similar results were also expected to be reported in urban

surroundings. What was surprising in the results was that the poorest overall performance of ITD methods was found in stratum 1, representing the areas with sparse canopy. Table 6 shows how the performance of the best ITD methods is at the lowest in stratum 1 considering most of the studied criteria. Similarly, looking at the best-case results in stem count and DBH criteria (Tables 3 and 4), the results in sparse canopy areas are mainly weaker than in the other two strata. In the open park areas of stratum 1, variation in canopy structure is wide. The stratum contains the smallest and some of the largest trees in the study area. Also, the spatial distribution of the trees varied substantially within the stratum as clustering of the trees was not taken into consideration. Hence, similar composition in terms of canopy cover and mean height may result from highly different spatial patterns. Evenly distributed crowns are an ideal subject for delineation of individual trees as every tree stands out from the CHM. Then again, clustered crowns result in challenging crown objects for crown delineation, because tree crowns are interlocking, and tall trees cover the smaller ones.

As shown in earlier studies in managed rural forests, deciduous trees differ from conifers in crown shape (e.g., Kwack et al., 2007; Koch et al. 2006) and thus pose special requirements for ITD methods. The same also holds for urban forests where manmade growing conditions result in altered growth patterns (Tanhuanpää et al., 2014, 2017). As expected, similar findings were also made in this study. The observed effect of large deciduous trees was twofold. Because of their wide and fragmented crowns, old deciduous trees tended to be oversegmented with all methods. In addition to visual observations during the process, the phenomenon was also supported indirectly by the results. Stratum 2, which had the most accurate ITD results with respect to all criteria (Tables 3-5), had lower DBH_{mean} and DBH_{max} than strata 1 and 3 (Table 1). This means that the relative number of large trees in stratum 2 was lower and that the biggest trees were notably smaller than in the other two strata. Still, as the crowns were not investigated in detail, only indirect interpretations can be made. In addition to commission errors, large deciduous crowns also resulted in omission errors. Especially smaller trees located under or even close to large crowns were easily left undetected. We found this was especially problematic in case of open-grown crowns. To be correctly delineated, the wide multi-top crowns required either heavy filtering of CHM or, depending on the method used, large threshold value for seed point selection. In both cases small tree tops were ignored. Brandtberg et al. (2001) reported similar findings on delineating unevenly sized trees simultaneously.

The effect of large crowns can be seen in the distributions describing the trees in strata 2 and 3 (Figures 4 and 5). Looking at the reference distributions, the largest DBH classes in stratum 2 (Figure 7) consisted of trees with DBH <10 cm. However, with all the ITD methods tested, the trees in the largest classes had DBH of 20 cm and above. We presume that this was caused mainly by two factors. Firstly, small trees remained undetected under the taller ones which lead to omission errors in small diameter classes. On the other hand, fragmentation of large tree crowns caused detection of non-existing trees, which lead to commission errors in larger diameter classes. However, as dense dominant canopy layer is known to suppress shorter trees in CHM-based ITD regardless of whether there are single large trees or not (Wang et al. 2016), the finding can be only partly supported by the results. The effects of high canopy cover and large tree crowns are very similar and thus challenging to separate. In both cases, smaller DBH classes suffer from omission errors. The results lead to a conclusion that many of the large open-grown trees are likely misdelineated, resulting in several smaller crown segments, but the effect of higher canopy cover cannot be fully excluded from the analysis.

The effect of large trees was most evident in stratum 1, which turned up to be the most challenging environment for ITD (see, Tables 3-5). Then again, stratum 2 that consisted of areas with the lowest mean DBH and the canopy was formed mainly from young trees with rather consistent crowns, seems to represent the most favorable conditions for ITD in the study area.

Comparing the results from stratified and non-stratified data shows that there are no methods that would be the best in all strata. Despite the differences in smoothing of CHM and applying WS, simple WS and marker

controlled WS resulted in similar results. Applying the methods with the lowest EI (i.e., $G_{0.7}$, F_2 and G_{adapt}) leads to total of 670 misclassifications, whereas the best performing method ($G_{0.7}$) for the unstratified dataset leads to 692 misclassifications.

All the best three methods had similar trend of performance between the three strata (Table 5). In terms of EI_{rel}, stratum 2 had the most accurate results. For strata 1 and 3, the EI_{rel} values were typically rather similar between the methods. For the best three, results for stratum 1 were slightly better that for stratum 3. This means that the shape of the DBH distributions were more accurate in stratum 1. Overall, all the best methods underestimated the plot-level tree count (Table 6). From the three strata, the biggest underestimations were found in stratum 3 where majority of the small undergrowth was missed due to dense cover of the suppressing canopy (Figure 8).

In general, ITD approach performed the best in stratum 2. When comparing the results from the best-performing methods, the accuracy in stratum 2 was the highest in nearly all the criteria presented in Table 6. The only exception to this was DBH_{mean}, which had slightly higher relative bias than in stratum 3. We suggest that the reason for having the best results in stratum 2 lies in the low number of old open-grown trees. This is supported by the mean characteristics of the strata. On average, the trees in stratum 2 are growing much denser than in the other two strata. Also, the mean diameter is the lowest and mean height close to that of stratum 1. Hence the canopy in stratum 2 consists of younger trees with narrow and hence distinct crowns.

Stratification of park areas served two purposes. Firstly, it could be means for reaching more accurate results, as sites can be mapped with methods adjusted for those specific conditions. Secondly, considering operational applications, it allowed assessing the accuracy of the methods tested separately for each stratum. Knowing that the accuracy of ITD varies over the mapped area, the stratum-level accuracy could be used as a quality indicator for the created tree maps. However, as only ALS-derived metrics (CC and H_{mean}) were used for the stratification, the actual structure of the resulting strata varied considerably. The within-strata variation seemed to affect the accuracy of ITD. In terms of tree size, the park areas in stratum 1 and 3 are more diverse than in stratum 2. High variation in the size of trees lowered the accuracy of ITD regardless of the method used. In the light of this study, urban ITD methods would benefit from better and broader description of canopy. Further research effort should be put on meaningful stratification of park areas for enhancing the performance of ITD in park areas.

Overall, applying a single method for all three strata leads to lower overall accuracy with all three indicators tested (i.e., number of trees, DBH, and Error index). This underlines that applying ITD for heterogeneous urban forests requires specialized methods. Compared to the ITD methods developed for forest conditions, urban forests require algorithms that are more agile to cope with the high variation in tree size and shape.

Conclusions

The intense management of park areas creates growing conditions that are rather unnatural. This holds especially for light conditions. Altered growing conditions affect the formation of individual tree crowns and thus the characteristics of the whole canopy. In this study, we pursued the issue by stratifying the park areas by ALS-derived features describing the canopy structure. The results show that urban, highly heterogeneous, and deciduous-dominated forests are a challenging environment for CHM-based detection of individual trees. In the light of this study, the key elements limiting the accuracy of urban ITD are the presence of large open-grown deciduous trees and, on the other hand, the heterogeneity of the canopy. Large deciduous trees result in both omission and commission errors and are thus a single key component in forming a successful ITD application. In addition, it seems that canopy cover type-specific methods would increase the accuracy. However, the specialized methods need to tackle a large variation of crown characteristics even within small

park areas. To achieve the best possible results, stratifying the park areas into more homogeneous areas is a viable option. At least for the relatively open park areas, basing the stratification solely on canopy cover and mean height is not enough for defining the most suitable crown delineation method. As for adaptive delineation methods, this means that controlling the segmentation algorithm with solely CHM height does not provide enough information for achieving good ITD accuracy.

Acknowledgement

The project was done in co-operation with the City of Helsinki and funded by the City of Helsinki Innovation fund and the Academy of Finland (The Centre of Excellence in Laser Scanning Research, project number 2721955). Additional funding was also received from the Academy of Finland's Strategic Research Council (IBC-Carbon, project number 312559).

References

Alonzo, M., Bookhagen, B., & Roberts, D. A. (2014). Urban tree species mapping using hyperspectral and lidar data fusion. Remote Sensing of Environment, 148, 70-83.

Axelsson, P. (2000). DEM generation from laser scanner data using adaptive TIN models. International archives of photogrammetry and remote sensing, 33(4), 110-117.

Brandtberg, T., Warner, T. A., Landenberger, R. E., & 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(3), 290-303.

Duncanson, L. I., Cook, B. D., Hurtt, G. C., & Dubayah, R. O. (2014). An efficient, multi-layered crown delineation algorithm for mapping individual tree structure across multiple ecosystems. *Remote Sensing of Environment*, *154*, 378–386.

Eysn, L., Hollaus, M., Lindberg, E., Berger, F., Monnet, J. M., Dalponte, M., Kobal, M., Pellegrini, M., Lingua, E., Mongus, D., & Pfeifer, N. (2015). A benchmark of lidar-based single tree detection methods using heterogeneous forest data from the alpine space. *Forests*, *6*(5), 1721–1747.

Holopainen, M., Kankare, V., Vastaranta, M., Liang, X., Lin, Y., Vaaja, M., Yu, X., Hyyppä, J., Hyyppä, H., Kaartinen, H., Kukko, A., Tanhuanpää, T., & Alho, P. (2013). Tree mapping using airborne, terrestrial and mobile laser scanning–A case study in a heterogeneous urban forest. Urban forestry & Urban greening, 12(4), 546–553.

Hyyppä, J. (1999). Detecting and estimating attributes for single trees using laser scanner. Photogramm J 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., Yu, X., Hyyppä, H., Vastaranta, M., Holopainen, M., Kukko, A., Kaartinen, H., Jaakkola, A., Vaaja, M., Koskinen, J., Alho, P. (2012). Advances in forest inventory using airborne laser scanning. *Remote Sensing*, *4*(5), 1190–1207.

527 Jing, L., Hu, B., Li, J., & Noland, T. (2012). Automated delineation of individual tree crowns from LiDAR data
528 by multi-scale analysis and segmentation. Photogrammetric Engineering & Remote Sensing, 78(12), 1275–
529 1284.

530 Kaartinen, H., Hyyppä, J., Yu, X., Vastaranta, M., Hyyppä, H., Kukko, A., Holopainen, M., Heipke, C.,
531 Hirschmugl, M., Morsdorf, F., Næsset, E., Pitkänen, J., Popescu, S., Solberg, S., Wolf, B. M., & Wu, J-C. (2012).
532 An international comparison of individual tree detection and extraction using airborne laser scanning.
533 Remote Sensing, 4(4), 950-974.

534 Kim, S., Hinckley, T., & Briggs, D. (2011). Classifying individual tree genera using stepwise cluster analysis
535 based on height and intensity metrics derived from airborne laser scanner data. *Remote Sensing of*
536 *Environment*, *115*(12), 3329–3342.

537 Koch, B., Heyder, U., & Weinacker, H. (2006). Detection of individual tree crowns in airborne lidar data.
538 *Photogrammetric Engineering & Remote Sensing*, *72*(4), 357–363.

539 Koukoulas, S., & Blackburn, G. A. (2005). Mapping individual tree location, height and species in broadleaved
540 deciduous forest using airborne LIDAR and multi-spectral remotely sensed data. International Journal of
541 Remote Sensing, 26(3), 431–455.

542 Kwak, D. A., Lee, W. K., Lee, J. H., Biging, G. S., & Gong, P. (2007). Detection of individual trees and estimation
543 of tree height using LiDAR data. Journal of Forest Research, 12(6), 425-434.

544 Lee, J. H., Ko, Y., & McPherson, E. G. (2016). The feasibility of remotely sensed data to estimate urban tree
545 dimensions and biomass. *Urban forestry & urban greening*, *16*, 208–220.

546 Lindberg, E., & Holmgren, J. (2017). Individual Tree Crown Methods for 3D Data from Remote Sensing. *Current*
547 *Forestry Reports*, *3*(1), 19–31.

548 Lu, X., Guo, Q., Li, W., & Flanagan, J. (2014). A bottom-up approach to segment individual deciduous trees
549 using leaf-off lidar point cloud data. ISPRS Journal of Photogrammetry and Remote sensing, 94, 1-12.

550 Mäkinen, H. (2002). Effect of stand density on the branch development of silver birch (Betula pendula Roth)
551 in central Finland. Trees, 16(4-5), 346-353.

552 Mongus, D., & Žalik, B. (2015). An efficient approach to 3D single tree-crown delineation in LiDAR data. ISPRS
553 Journal of Photogrammetry and Remote Sensing, 108, 219-233.

554 Moskal, L. M., Styers, D. M., & Halabisky, M. (2011). Monitoring urban tree cover using object-based image
555 analysis and public domain remotely sensed data. *Remote Sensing*, *3*(10), 2243–2262.

556 Niemistö, P. (1995). Influence of initial spacing and row-to-row distance on the crown and branch properties
557 and taper of silver birch (Betula pendula). Scandinavian Journal of Forest Research, 10(1-4), 235–244.

558 Packalén, P., & Maltamo, M. (2008). Estimation of species-specific diameter distributions using airborne laser
559 scanning and aerial photographs. Canadian Journal of Forest Research, 38(7), 1750-1760.

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

562 Pitkänen J., Maltamo M., Hyyppä J., Yu X., 2004. Adaptive methods for individual tree detection on airborne
563 laser based canopy height model. International Archives of Photogrammetry, Remote Sensing and Spatial
564 Information Sciences, Vol. XXXVI, 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 (3), 767–781.

Rahlf, J., Breidenbach, J., Solberg, S., Astrup, R. (2015) Forest parameter prediction using an image-based point cloud: A comparison of semi-itc with ABA. Forests, 6, 4059–4071.

Reynolds, M. R., Burk, T. E., & Huang, W. C. (1988). Goodness-of-fit tests and model selection procedures for diameter distribution models. Forest Science, 34(2), 373–399.

Riikonen, A., Pumpanen, J., Mäki, M., & Nikinmaa, E. (2017). High carbon losses from established growing sites delay the carbon sequestration benefits of street tree plantings–A case study in Helsinki, Finland. Urban Forestry & Urban Greening, 26, 85-94.

Rottensteiner, F., Sohn, G., Gerke, M., Wegner, J. D., Breitkopf, U., & Jung, J. (2014). Results of the ISPRS benchmark on urban object detection and 3D building reconstruction. ISPRS Journal of Photogrammetry and Remote Sensing, 93, 256–271.

Saarinen, N., Vastaranta, M., Kankare, V., Tanhuanpää, T., Holopainen, M., Hyyppä, J., & Hyyppä, H. (2014). Urban-tree-attribute update using multisource single-tree inventory. *Forests*, *5*(5), 1032-1052.

St-Onge, B., Audet, F.-A., and Bégin, J. (2015). Characterizing the height structure and composition of a boreal forest using an individual tree crown approach applied to photogrammetric point clouds. Forests, 6, 3899–3922.

Suárez, J. C., Ontiveros, C., Smith, S., & Snape, S. (2005). Use of airborne LiDAR and aerial photography in the estimation of individual tree heights in forestry. Computers & Geosciences, 31(2), 253–262.

Tanhuanpää, T., Kankare, V., Setälä, H., Yli-Pelkonen, V., Vastaranta, M., Niemi, M. T., Raisio, J., & Holopainen, M. (2017). Assessing above-ground biomass of open-grown urban trees: A comparison between existing models and a volume-based approach. Urban forestry & urban greening, 21, 239–246.

Tanhuanpää, T., Saarinen, N., Kankare, V., Nurminen, K., Vastaranta, M., Honkavaara, E., Karjalainen, M., Yu, X., Holopainen, M., & Hyyppä, J. (2016). Evaluating the performance of high-altitude aerial image-based digital surface models in detecting individual tree crowns in mature boreal forests. *Forests*, *7*(7), 143.

Tanhuanpää, T., Kankare, V., Vastaranta, M., Saarinen, N., Holopainen, M., & Raisio, J. (2015). Deriving canopy metrics of urban trees from airborne laser scanning data. In *Joint Urban Remote Sensing Event (JURSE), 2015* (pp. 1-4). IEEE.

Tanhuanpää, T., Vastaranta, M., Kankare, V., Holopainen, M., Hyyppä, J., Hyyppä, H., Alho, P., & Raisio, J. (2014). Mapping of urban roadside trees–A case study in the tree register update process in Helsinki City. *Urban forestry & urban greening*, *13*(3), 562–570.

Vastaranta, M., Wulder, M.A., White, J.C., Pekkarinen, A., Tuominen, S., Ginzler, C., Kankare, V., et al. 2013. Airborne laser scanning and digital stereo imagery measures of forest structure: comparative results and implications to forest mapping and inventory update. Canadian Journal of Remote Sensing, 39, 382–395.

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., & Maltamo, M. (2011). Comparative testing of single-tree detection algorithms under different types of forest. *Forestry*, *85*(1), 27-40.

603 Vauhkonen, J., Korpela, I., Maltamo, M., & Tokola, T. (2010). Imputation of single-tree attributes using
604 airborne laser scanning-based height, intensity, and alpha shape metrics. *Remote Sensing of Environment*,
605 *114*(6), 1263–1276.

606 Vauhkonen, J., Tokola, T., Packalén, P., & Maltamo, M. (2009). Identification of Scandinavian commercial
607 species of individual trees from airborne laser scanning data using alpha shape metrics. *Forest Science*, *55*(1),
608 37–47.

609 Wang, Y., Hyyppä, J., Liang, X., Kaartinen, H., Yu, X., Lindberg, E., Holmgren, J., Qin, Y., Mallet, C., Ferraz, A.,
610 Torabzadeh, H., Morsdorf, F., Zhu, L., Liu, J., & and Alho, P. (2016). International benchmarking of the
611 individual tree detection methods for modeling 3-D canopy structure for silviculture and forest ecology using
612 airborne laser scanning. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(9), 5011–5027.

613 Wang, Y., Weinacker, H., & Koch, B. (2008). A lidar point cloud based procedure for vertical canopy structure
614 analysis and 3D single tree modelling in forest. Sensors, 8(6), 3938-3951.

615 White, J.C., Coops, N.C., Wulder, M.A., Vastaranta, M., Hilker, T., Tompalski, P. (2016). Remote sensing
616 technologies for enhancing forest inventories: Areview. Canadian Journal of Forest Research 42, 619-641.

617 White, J., Stepper, C., Tompalski, P., Coops, N., & Wulder, M. (2015). Comparing ALS and image-based point
618 cloud metrics and modelled forest inventory attributes in a complex coastal forest environment. Forests,
619 6(10), 3704-3732.

620 White, J., Wulder, M., Vastaranta, M., Coops, N., Pitt, D., & Woods, M. (2013). The utility of image-based
621 point clouds for forest inventory: A comparison with airborne laser scanning. Forests, 4(3), 518–536.

622 Yu, X., Hyyppä, J., Vastaranta, M., Holopainen, M., & Viitala, R. (2011). Predicting individual tree attributes
623 from airborne laser point clouds based on the random forests technique. *ISPRS Journal of Photogrammetry*
624 *and remote sensing*, *66*(1), 28–37.

625 Zhang, C., Zhou, Y., & Qiu, F. (2015). Individual tree segmentation from LiDAR point clouds for urban forest
626 inventory. Remote Sensing, 7(6), 7892–7913.