Skeleton-based botanic tree diameter estimation from dense LiDAR data

Alexander Bucksch, Roderik Lindenbergh, Massimo Menenti and Muhammad Z. Raman
Delft Institute of Earth Observation and Space Systems,
Delft University of Technology, Delft, The Netherlands

ABSTRACT
New airborne LiDAR (Light Detection and Ranging) measurement systems, like the FLI-MAP 400 System, make it possible to obtain high density data containing far more information about single objects, like trees, than traditional airborne laser systems. Therefore, it becomes feasible to analyze geometric properties of trees on the individual object level. In this paper a new 3-step strategy is presented to calculate the stem diameter of individual natural trees at 1.3m height, the so-called breast height diameter, which is an important parameter for forest inventory and flooding simulations. Currently, breast height diameter estimates are not obtained from direct measurements, but are derived using species dependent allometric constraints. Our strategy involves three independent steps: 1. Delineation of the individual trees as represented by the LiDAR data, 2. Skeletonization of the single trees, and 3. Determination of the breast height diameter computing the distance of a suited subset of LiDAR points to the local skeleton. The use of a recently developed skeletonization algorithm based on graph-reduction is the key to the breast height measurement. A set of four relevant test cases is presented and validated against hand measurements. It is shown that the new 3-step approach automatically derives breast height diameters deviating only 10% from hand measurements in four test cases. The potential of the introduced method in practice is demonstrated on the fully automatic analysis of a LiDAR data set representing a patch of forest consisting of 49 individual trees.

Keywords: ALS, LiDAR, Skeleton extraction, tree delineation, stem diameter estimation, FLI-MAP 400

1. INTRODUCTION
In forestry the breast height diameter (stem diameter at 1.3m height) of individual botanical trees is an important parameter to estimate further parameters like the leaf area index, [1], or to obtain insight in the carbon water relations in forests, [2]. Furthermore, the breast height diameter is often used for hydrological applications, because single standing trees are a main influence on the water flow in flooded areas, [3]. Recent high density LiDAR data give a new possibility to measure the breast height diameter of single standing trees directly. In [4], it is shown that the FLI-MAP 400 system obtains high density LiDAR data with approximately 50 points per square meter. Successful estimation methods to calculate the stem diameter directly within the FLI-MAP data, rather then making a species depended estimation, as described in [5], are not known until now. HARPER - Histogram Analysis & Retraction of Points Estimated Radii - is a method to extract the breast height diameter directly from the data. After describing the FLI-MAP 400 system, an overview of the current estimation process for breast height diameters is given. As last part of the introduction an overview of existing methods for tree skeletonisation is given. In Section 2 we describe the three principles of the suggested three step strategy:

1. Delineation of the individual trees;
2. Skeletonisation of single trees and;
3. Histogram analysis of the point distances to the skeleton.

Further author information: (Send correspondence to A.K. Bucksch)
A.K. Bucksch.: E-mail: a.k.bucksch@tudelft.nl, Telephone: +31 (0)15 2784992
R.C. Lindenbergh: E-mail: r.c.lindenbergh@tudelft.nl, Telephone: +31 (0)15 278 7649
M.Menenti: E-mail: m.menenti@tudelft.nl, Telephone: +31 (0)15 278 4244
M. Z. Raman: E-mail: M.Z.AbdRahman@tudelft.nl, Telephone: +31 (0)15 278 8337
An extensive result section shows all analyzed test cases and discusses the results before taking conclusions from them.

1.1 FLI-MAP 400

The FLI-MAP 400 System, which is the follow up of the original FLI-MAP System, [4], is an airborne LiDAR system that is able to obtain approximately 50 height points per m² when flying at 100m above the ground at 20 m/s speed (see Table 1). The FLI-MAP 400 System is built as a helicopter system, that enables lower flying height compared to an airplane. The lower flight height is resulting in data sets that are 10 times as dense as traditional laser altimetry. According to [4], the main characteristics of the FLI-MAP 400 System are:

1. Operating at very low altitudes (50-150 meters);
2. Four integrated photo and video cameras;
3. Double laser system to eliminate shadowing.

The two lasers in the system are reflector less range finders, firing 150,000 laser pulses per second at a 60° angle perpendicular to the flight path. The reflected laser intensity of the terrain and objects also gives information on the type of surveyed material. In the Netherlands, airborne laser altimetry has been used for the production of a detailed elevation model of the whole country. The project, known as AHN (Actual Height model of The Netherlands), [6], is developed to provide detailed information about elevation, highly demanded by water boards, provinces and the national government. A comparison between the new FLI-MAP 400 system, its predecessor FLI-MAP and systems used for obtaining the AHN, [6], results in the values as listed in Table 1. The FLI-MAP 400 system is one system used to obtain the data for the upcoming AHN2, [6]. AHN2 is the successor of the AHN project. Therefore new data processing methods employing the strength of the FLI-MAP 400 system will have applications at large scale.

<table>
<thead>
<tr>
<th></th>
<th>FLI-MAP</th>
<th>FLI-MAP 400</th>
<th>AHN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aircraft height</td>
<td>50-150m</td>
<td>50-400m</td>
<td>1000m</td>
</tr>
<tr>
<td>Aircraft speed</td>
<td>50-80 km/h</td>
<td>50-80 km/h</td>
<td>250 km/h</td>
</tr>
<tr>
<td>Pts/m² 100m</td>
<td>10-25</td>
<td>ca. 50</td>
<td>1</td>
</tr>
</tbody>
</table>

1.2 Breast Height measurement

Before estimating the breast height diameter, the extraction of single trees from the data set is necessary, e.g. [7, 8, 9]. This extraction process is commonly called delineation. Delineation of trees is for example possible by looking at the density distribution of height points. After computing the neighborhood of every datum point, the local density maxima are extracted to locate the tree tops. The assumption of a species dependent shape model of a tree supports the allocation of points belonging to one individual tree, [10]. Estimations of the crown size and tree height as input for a species depended diameter estimation are described in [11, 12, 13]. This assumes an allometric relation between crown size, tree height and stem diameter. The state of the art way, described in [14], to estimate the stem diameter at breast height is first to estimate the crown radius as function of the tree height incorporating three species dependent parameters. The tree height is derived from the LiDAR data. The breast height diameter is then estimated as a function of tree height and maximum crown width also incorporating three species dependent diameters. Further examples of the allometric approach can be found in [15, 16, 17]. The major disadvantage of the allometric approach is, that the species of the tree has to be known on forehand, and that the breast height diameter is not measured directly from the LiDAR observations.

The approach introduced in this paper will not use species depended information to estimate the stem diameter. The estimation is done under support of a skeleton, describing the structure of the tree. Therefore the next section gives an overview of methods of existing skeletonisation methods of trees.
Rough Text:

Figure 1. Principle of the skeletonization: (a) point cloud of a delineated tree, (b) the octree organization, (c) the extracted octree-graph, (d) the resulting skeleton

1.3 Tree Skeletonisation

During the last years publications targeting on tree skeletonization of range data became available. Until now no publication was aiming at the skeletonization of range data obtained from an airborne platform. Therefore, we describe here skeletonization methods originally developed for terrestrial laser scan data. These methods can be grouped into four principle classes.

Morphological thinning methods organize the point cloud in a 3D raster of equally sized cells. From this raster the outer layer is removed until the skeleton remains. Removing the outer layer uses the morphological operations opening and erosion [18]. This class of algorithms requires a defined inner volume of the object to produce a centered skeleton. A linear time algorithm using 6 sub-iterations was introduced by [19]. Its application on tree point clouds was proposed in [20] and later extended to achieve better connectedness of the skeleton in case of undersampling in [21].

Medial axis based approaches use approximations of the medial axis from the point cloud, e.g. [22]. This class of algorithms derives a skeleton from a Voronoi space division. The medial axis is in general a set of surfaces in 3D, but can be reduced to a skeleton, [23]. Another possibility to extract the medial axis is the distance transform, [24], but this approach does not guarantee a connected skeleton. All cells are marked by their distance to the object boundary. The set of neighboring maximal distances form the skeleton. The main disadvantage of the medial axis is its sensitivity to irregularities on the object surface. These irregularities may occur because of noise or unsampled parts within the point cloud. An application to trees is not known to us.

Geometric methods use a function that is shifted over the represented surface for the extraction of a skeleton. The height function is often used to extract the level sets from a given point cloud, e.g. [25]. Placement of additional vertices in the centroid of the extracted level sets detail the extracted skeleton. The resulting graph is often referred to as a Reeb-graph, [26]. The biggest problems arising with these approaches are the rotational dependency of the height function and the sensitivity of the level set extraction to the sampling density, [27]. Except for one minor example in [27], no application to trees is known to us.

Clustering methods produce clusters of points from a graph describing the scan point order. Some criterion is used to produce the clusters. Neighboring clusters are connected to a skeleton. In [28] a neighboring graph is

![Figure 1](image-url)
used and points with the same quantized distance from the root are considered as belonging to a cluster. The approach shows good results until $\frac{2}{3}$ of the tree height. The remaining skeleton is produced by using species dependent allometries. Another promising clustering approach was presented by [29]. They used a $kd$-tree and used a $k$-means clustering to produce the cluster from which the skeleton is derived.

**Graph reduction** based approaches, as introduced in [30], extract an initial graph from a spatial subdivision. This initial graph is reduced by a set of rules to a skeleton. These rules consider the connectivity between different parts of the point cloud. Several advantages of such a approach could be demonstrated: a high robustness to noise on imperfect data, a good centeredness and a good connectivity. Centeredness is achieved by embedding the graph into the point cloud. Topological correctness is achieved by choosing a proper decision criterion to place connections between the different point cloud parts and the careful design of the reduction rules. A second strongly improved approach based on graph reduction, called the SkelTre algorithm, was presented in [31]. The SkelTre algorithm is based on a merging scheme of an initial octree-graph incorporating the elongation of the skeletonized object part. Furthermore it pays special attention to undersampling and point cloud gaps.

### 2. METHODOLOGY

#### 2.1 Tree delineation and pre-filtering of the data
Delineation describes the process of extracting single trees from a given data set. Here we used the delineation method described in [32, 33]. In this delineation process the tree tops are detected by a region growing approach, starting from a set of seed points. After identification of the tree tops, a histogram is utilized to detect the tree shape. This histogram is calculated as histogram of the distances between the data points and a line perpendicular to the ground starting at the detected tree top. The bins of this single tree histogram are assumed to represent consecutively the components tree crown, tree stem, undergrown vegetation and ground surface. The histogram boundary marks each part of the tree histogram automatically as one of the components. The delineation process uses eight input parameters. Two of these parameters, the expected minimum and maximum crown radius, seem to be most sensitive with respect to the delineation result. As shown later, this delineation approach contains trees that are identified by the algorithm but could not be found back in the field (overdelineation) and trees which are present in the field, but not delineated properly (underdelineation).

#### 2.2 Skeletonisation of the trees
The delineated trees are skeletonized with the SkelTre algorithm, [31], originally developed for terrestrial laser scan data. A single delineated tree (Figure 1a) is divided by an octree into small point cloud parts (Figure 1b). From this octree organization a graph is extracted (Figure 1c), which is retracted to a one-dimensional skeleton (Figure 1d). The SkelTre algorithm requires only the input of one user defined parameter for the purpose of airborne data. The required input parameter is the minimum cell size. An additional criterion to decide, if the two neighboring cells are connected, is not explicitly necessary, because ALS data is sparse.

#### 2.3 Diameter measurement
The diameter measurement is a 3-step approach: first the stem is extracted from the skeleton, secondly a representative bin is chosen by evaluating the histogram of distances of the point cloud points to the skeleton and third a geometric criterion decides whether a diameter estimation is valid.

##### 2.3.1 Stem extraction
The skeleton graph consists of vertices and edges (see Figure 1d). The extraction of the stem from the skeleton graph follows a simple rule. The stem is extracted by evaluating the incoming and outgoing edges of the skeleton graph vertices, starting at the root. The root is defined as the vertex assigned to the smallest z-coordinate of the whole skeleton graph. In case of more then one outgoing edge, the edge forming an angle closest to 180 degree with the incoming edge is selected. The stem is saved as an one dimensional list of vertices, ordered by their adjacency.

In order to select point cloud points relevant for obtaining the stem diameter, the stem is followed from the root to the stem segment that covers the breast height diameter at 1,30m height. Every vertex is belonging to
Figure 2. Four test cases marked at breast height. For every tree the manual measurement, the HARPER estimation and a cylinder fitting is given as a diameter estimation: 1.) A single standing tree, 2.) a single standing tree with noise around the stem, 3.) a tree from the border of a forest with huge shadowing parts and 4.) a tree delineated from inside a forest.
a subset of the original point cloud, initially derived from the octree subdivision. A minimum number of points along the stem is needed to evaluate the stem diameter. Here a minimum of 40 points is chosen. Assuming that the stem is not significantly changing in its diameter, the stem is followed upwards until sufficient point cloud points are present to evaluate. A simple threshold is used to determine a valid stem extraction.

2.3.2 Histogram evaluation

The calculation of the distances between point cloud points belonging to the stem and the skeleton graph, is done by determining for each point the distance to the skeleton edge at the height of the point.

\[ d(p) = \frac{|(x_2 - x_1) \times (x_1 - p)|}{|x_2 - x_1|}, \]

where \( d \) denotes the distance to a line defined by two points \( x_1 \) and \( x_2 \), which are the coordinates of the two vertices of a skeleton edge. Here \( p \) denotes a point cloud point, such that \( z(x_1) \leq z(p) \leq z(x_2) \), where \( z(.) \) denotes the z-coordinate, i.e. the height.

From the calculated distances to the skeleton, the histogram of the distances is calculated. For diameter evaluation, the peak bin closest to the median is chosen. The average of the distances in the peak bin are assumed as the radius of the tree. In the data set evaluated here a bin size of 5cm distance was suited.

2.3.3 Validity criterion

The selected bin is evaluated by a simple criterion, to enhance the robustness of the method. We take two cases into account:

1. Shadowing may not allow the tree to be scanned from all sides, which will greatly influence the centeredness of the skeleton, because the point cloud contains not sufficient geometric information.

2. Blunders can have a strong influence on the skeleton generated from the sparse airborne data, because they force the embedding of the skeleton outside the stem center.

Shadowing or blunders can lead to bad embedding of the skeleton graph. Bad embedding can lead to an unreliable centering of the skeleton and therefore lead to significantly smaller distances to the skeleton. These small distances to the skeleton result in an underestimation of the diameter. In our approach we used the relation between the radius of the smallest enclosing circle of a point set and the smallest diameter of a set of imprecise points, as given in [34], to detect an underestimated diameter. If underestimation is detected we assume the smallest diameter of a set of imprecise points as the stem diameter. For details the reader is referred to [34].

3. RESULTS AND VALIDATION

The validation of the HARPER method is done by comparing the automatically extracted diameters from the FLI-MAP 400 data to field measurements in the “Duursche Waarde” in the Netherlands as shown in Figure 3. First four representative test cases are analyzed in detail, and secondly the analysis of 49 delineated forest trees is shown. All delineated trees from the patch are illustrated in Figure 5.

3.1 Four relevant test cases

Four test cases were chosen to discuss the HARPER method (Figure 2). A simple single standing tree, a single standing tree containing noise on the stem, a tree on the border of a forest with huge shadowing and a tree delineated within a forest. For every tree the point cloud with its corresponding skeleton, the extracted stem, the histogram to be analyzed and the measurement results are given. We give in total three measurements for the test cases. The manual measurement as a ground truth, the result obtained by the HARPER method and the result of a fitted cylinder. The cylinder was fitted with the point cloud package Cyclone 5.7 of Leica, [35] by selecting a visually suited subset of the point cloud. The four test cases were delineated by hand in order to eliminate the influence of the delineation method. The first tree can be seen as a standard case giving good results. The second tree is an example where insufficient points to estimate the diameter are present. A tree
Figure 3. The test area in the Netherlands, marked by a red box and the investigated forest patch, indicated by a yellow cross.

Figure 4. The trees delineated from the data set. Note the missing stems of some trees.

Figure 5. The delineated trees from the forest patch in the Dutch RD coordinate system. Note that this is the delineation result and not the actual map of the trees. Compare with Table 2.
suffering from huge shadowing effects (Tree 3) still gives an acceptable result, even if the skeleton extraction was far from optimal. 13cm difference to the validation measurement was obtained. The tree from the interior of the forest (tree 4) shows a clear peak and a huge difference in the measuring result between the fitted cylinder and the automatically estimated diameter. In this case the stem was not represented properly in the point cloud. By applying the introduced validity criterion an estimation of the breast height diameter could be obtained that only deviates 10cm from the manual measurements. In all four test cases the histogram analysis performed better than the cylinder fitting with a standard software.

### 3.2 49 delineated forest trees

49 trees were delineated from a FLI-MAP 400 data set in Figure 4. The data set is a mixed forest stand containing leaf and pine trees in the “Duursche Waarde” in the Netherlands. The problems arising in this particular area are mostly connected to finding back the trees in the forest. The dense canopy roof does not allow precise GPS measurements to localize tree positions reliably. Furthermore, depending on the chosen input parameters of the delineation algorithm [32, 33] trees may be wrongly classified as undergrown vegetation. In the patch in Figure 5 a manual inspection revealed five trees that were not detected by the delineation method. One tree that was delineated from the data set was lying on the ground at time of the validation measurement. Some trees could not be found back in the forest. In total it was concluded that 49 trees are suited for validation. From these 49 trees 14 had not sufficient stem points available for analysis after undergrown vegetation filtering.

The diameter of the 49 trees with in the selected forest patch was extracted with the HARPER method. The overall standard deviation associated with our estimation method is 0.19m. The diameters show an average deviation of 28cm. Especially the delineated trees in the interior of the forest lead to strong underestimation.

<table>
<thead>
<tr>
<th>Tree number</th>
<th>HARPER in cm</th>
<th>Manual in cm</th>
<th>Tree number</th>
<th>HARPER in cm</th>
<th>Manual in cm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>43</td>
<td>39</td>
<td>36</td>
<td>19</td>
<td>49</td>
</tr>
<tr>
<td>2</td>
<td>nan</td>
<td>52</td>
<td>37</td>
<td>16</td>
<td>44</td>
</tr>
<tr>
<td>3</td>
<td>45</td>
<td>37</td>
<td>38</td>
<td>nan</td>
<td>34</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>36</td>
<td>39</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>35</td>
<td>43</td>
<td>40</td>
<td>nan</td>
<td>39</td>
</tr>
<tr>
<td>7</td>
<td>35</td>
<td>47</td>
<td>41</td>
<td>83</td>
<td>32</td>
</tr>
<tr>
<td>8</td>
<td>64</td>
<td>47</td>
<td>42</td>
<td>nan</td>
<td>50</td>
</tr>
<tr>
<td>9</td>
<td>66</td>
<td>48</td>
<td>45</td>
<td>20</td>
<td>45</td>
</tr>
<tr>
<td>10</td>
<td>66</td>
<td>56</td>
<td>46</td>
<td>47</td>
<td>19</td>
</tr>
<tr>
<td>11</td>
<td>28</td>
<td>38</td>
<td>50</td>
<td>nan</td>
<td>37</td>
</tr>
<tr>
<td>14</td>
<td>24</td>
<td>40</td>
<td>53</td>
<td>25</td>
<td>46</td>
</tr>
<tr>
<td>15</td>
<td>20</td>
<td>43</td>
<td>55</td>
<td>65</td>
<td>39</td>
</tr>
<tr>
<td>16</td>
<td>44</td>
<td>37</td>
<td>57</td>
<td>nan</td>
<td>54</td>
</tr>
<tr>
<td>17</td>
<td>32</td>
<td>45</td>
<td>59</td>
<td>nan</td>
<td>46</td>
</tr>
<tr>
<td>18</td>
<td>nan</td>
<td>37</td>
<td>60</td>
<td>25</td>
<td>39</td>
</tr>
<tr>
<td>19</td>
<td>nan</td>
<td>172</td>
<td>61</td>
<td>nan</td>
<td>34</td>
</tr>
<tr>
<td>20</td>
<td>14</td>
<td>40</td>
<td>62</td>
<td>nan</td>
<td>54</td>
</tr>
<tr>
<td>21</td>
<td>39</td>
<td>37</td>
<td>63</td>
<td>nan</td>
<td>47</td>
</tr>
<tr>
<td>22</td>
<td>56</td>
<td>33</td>
<td>65</td>
<td>20</td>
<td>45</td>
</tr>
<tr>
<td>23</td>
<td>nan</td>
<td>69</td>
<td>68</td>
<td>32</td>
<td>38</td>
</tr>
<tr>
<td>24</td>
<td>16</td>
<td>30</td>
<td>69</td>
<td>67</td>
<td>34</td>
</tr>
<tr>
<td>25</td>
<td>36</td>
<td>52</td>
<td>70</td>
<td>30</td>
<td>28</td>
</tr>
<tr>
<td>28</td>
<td>28</td>
<td>46</td>
<td>71</td>
<td>15</td>
<td>28</td>
</tr>
<tr>
<td>30</td>
<td>17</td>
<td>51</td>
<td>74</td>
<td>120</td>
<td>41</td>
</tr>
</tbody>
</table>

Table 2. Comparison between the HARPER diameter estimation and the manually measured diameter in the field. The tree numbers correspond to the numbers in the tree map of Figure 5.
Table 2 shows the manual and HARPER estimations of the breast height diameter of trees found back in the forest within two days. The tree number is corresponding to the numbers given in Figure 5.

3.3 Conclusions

We presented a proof of concept for a method to automatically extract the breast height diameter from trees sampled by the high density airborne laser data system FLI-MAP 400. On four relevant test cases the principle behavior of the Harper method was discussed. Furthermore, this HARPER method was validated against 49 trees of a forest patch in the “Duursche Waarde” in the Netherlands. The results indicate strongly, that there is need for new delineation methods, in order to make the HARPER method feasible for complex forests. Several trees were found in the delineation, which contain no stem data. Our further studies will focus on a possible delineation of the trees based on the skeleton-graph. Further a strong criterion to reject a computed diameter is needed.

Because of the difficulty to find back individual trees in a complex forest, it was not possible to get a large and reliable validation data set based on a delineation. Furthermore, activity will aim on the data collection of more isolated validation trees, because it was shown that the extracted diameter shows good results on the manually delineated trees. It is not clear, what causes the huge amount of diameter underestimations. Two reasons will be investigated, the influence of the delineation process and the influence of the FLI-MAP 400 system.

ACKNOWLEDGMENTS

This research was supported by the ICT Datadienst of the Rijkswaterstaat in the Netherlands.

REFERENCES


