A generic approach in estimating vegetation density for hydrodynamic roughness parameterization using high density airborne laser scanning data
M. Z. A. Rahman, B. Gorte, M. Menenti and A. L. Ibrahim

ABSTRACT
Vegetation density is among the important parameters required for determination of hydrodynamic roughness over vegetated areas. High density airborne light detection and ranging (LiDAR) data offer several potentials to improve estimation of vegetation density. Available methods in estimating vegetation density based on regression models did not take into account understorey vegetation and were not tested under different forest conditions. We present a method to develop and validate a generic regression model by using simulations of airborne laser scanning. The results show that available indices failed to produce good estimation which leads to a new predictor called low points index ($LP$). The vegetation density of trees is estimated using the FLI-MAP 400 data based on a regression model and estimated tree diameter at breast height. Finally, vegetation density is estimated at different spatial resolutions, which is useful for the estimation of multi-resolution and spatially distributed hydrodynamic roughness.

Key words | composite hydrodynamic roughness, diameter at breast height, high density airborne laser scanning, vegetation density

INTRODUCTION
Hydrodynamic roughness of vegetation is especially important in understanding flood-risk implications of restoring woody vegetation in riparian zones and floodplains (Darby 1999; Anderson et al. 2006; Antonarakis et al. 2008). Extracting information on the vegetation density in a dense riparian forest is rather challenging. Riparian forest might consist of different types of trees with different structures, densities, and most likely accompanied by understorey vegetation. In case of a flood, the impedance of flow is highly dependent on the density of dominant trees as well as on the understorey vegetation (Forzieri et al. 2011). Hydrodynamic roughness or resistance is the sum of the forces acting against the motion of fluid, which consequently limits the velocity and exerts control over flow depth and discharge. Therefore, the hydrodynamic roughness values of the overland vegetated area have an influence on the flow pattern, water depth, and flow velocity (Vieux 2004; Pleumeekers 2010).

Petryk & Bosmajian (1975) defined vegetation density, $D_v$ ($m^{-1}$) as the ratio of the total frontal area of vegetation in the flow direction per unit volume of water. They considered the stem as rigid cylinders and the required parameters to estimate hydrodynamic roughness are vegetation density and drag coefficient of stem. Several studies have been done to estimate hydrodynamic roughness of vegetation in forest areas using airborne light detection and ranging (LiDAR). Straatsma & Baptist (2008) developed a method to automatically estimate vegetation hydrodynamic roughness using airborne LiDAR and compact airborne spectral imager (CASI) data. In their study, vegetation was divided into forest and herbaceous vegetation. The estimation of $D_v$ for both types of vegetation was done by using a calibrated regression model with the percentage index ($P$) as a predictor. However, the regression model did not take into account the presence of understorey...
vegetation, and the predictor has not been evaluated for different forest conditions. Antonarakis et al. (2008) used airborne LiDAR to extract information on tree trunks to aid determination of hydrodynamic roughness using a rigid stem approach. The stem diameter and frontal area were estimated using allometric equations derived from field measurements. Trees were detected and delineated using the tree detection and crown segmentation (TDCS) method and the tree measurement method was implemented using the TreeVaW software (Popescu & Wynne 2004). This information was further used to estimate the diameter at breast height (DBH) of each tree. The $D_v$ was estimated by taking a ratio between the average inundated DBH and the average distance between trees.

During a flood, the understorey vegetation and tree trunks are the main factors that affect the flow (Vieux 2004; Antonarakis et al. 2008). However, there is still little research devoted to taking into account understorey vegetation in the estimation of $D_v$ due to several problems. Measuring density of understorey vegetation in the field is rather difficult, since $D_v$, for example different densities, properties of tree trunks as well as conditions of forest canopies should be considered. This inevitably leads to an insufficient quantity of field data that can be used to develop the required regression model. In addition, the performance of different indices used as predictors for $D_v$ for different forest conditions is still open to debate. It has been shown that intensive investigation of their capabilities is hampered by a lack of field data (Straatsma 2005), which supports a method based on simulated data.

Due to difficulties in obtaining field-measured vegetation density, $D_v$, we decided to use simulated airborne LiDAR observations over tree models. With this method, $D_v$ can be determined directly from the tree models, while the required indices (predictors) can be obtained from the generated point clouds. The airborne LiDAR simulations will be carried out for various conditions of forest patches and isolated trees. The regression models for the $D_v$ estimation are quite conveniently applied on an area with a group of trees with understorey vegetation. For isolated trees, the vegetation density, $D_v$, can be estimated from the tree trunk diameter (Arcement & Schneider 1989). The objective of this study, to estimate $D_v$ for trees in a forest patch and for isolated trees using high density airborne LiDAR data, is articulated in the following sub-objectives:

1. To simulate airborne LiDAR observations over different conditions of trees in forest patches and isolated trees.
2. To investigate the performance of several predictors for the estimation of $D_v$ by taking into account various forest conditions.
3. To investigate the impact of point density of the airborne LiDAR on the estimation of $D_v$ for trees in a forest patch.
4. To evaluate the method of tree DBH measurement introduced by Rahman et al. (2009) on isolated trees.

**METHODOLOGY**

The methodology is divided into six parts (see Figure 1). The first part deals with the development of simulated laser pulses of an airborne LiDAR system by considering the original FLI-MAP 400 specifications. The second part is devoted to the development of tree models in forest patches and isolated trees. The tree models are developed to mimic the trees during leaf-off conditions. The tree models consider a range of tree properties, such as size of tree trunk and branches and direction of branches. In addition, the dominant trees in the forest patch are also accompanied by understorey vegetation. The third step is the simulation of LiDAR observations using the tree models for both forest patches and isolated trees. The simulation of airborne LiDAR observations generates point clouds for different forest conditions and with different densities of laser pulses. The simulation is based on the concept of intersections between lines (laser beams) and cylinders (tree elements).

The fourth step involves computations of indices using simulated point clouds of trees in a forest patch. There are three indices used in this study, two of which (laser interception index ($li$) and percentage index ($P$)) are adopted from previous studies (MacArthur & Horn 1969; Straatsma 2007). In addition, we introduce a new index, called the low points index ($LP$). The fifth step focuses on the development and evaluation of regression models and the evaluation of the DBH estimation method. The final regression model will be generated by using simulated...
point clouds with a laser pulse density of 75 points per m². In the final step, the most suitable regression model from the previous step is applied on the real FLI-MAP 400 airborne LiDAR dataset to estimate $D_v$ for trees in a forest. The final $D_v$ maps are generated by combining $D_v$ values calculated either using the regression model, DBH information of isolated trees, or a combination of these methods. The $D_v$ maps are generated with different spatial resolutions.

**Study area**

The study area is located in the surroundings of the Duursche Waarden floodplain, The Netherlands. It covers several parts of the Den Nul, Veessen, and Fortmond regions. Trees are either present in a forest patch or isolated and scattered around the study area. Forest patches, for instance in the Fortmond region, consist of several tree species, such as softwood trees like willow and poplar, hardwood trees such as oak, ash and a small pine stand on a river dune, and also other vegetation, like reed marshes, and herbaceous vegetation (Straatsma 2007). In our field observation, we have found that individual forest patches contain several species of trees. The size and age of the trees vary within a patch, from small and young to large mature trees. Throughout the study area, in most cases different densities of understorey vegetation occur with the trees.

**Development of tree models**

Development of tree models aims at creating forest patches and isolated trees that closely imitate various conditions of deciduous trees during the leaf-off season. In a natural forest, trees would be accompanied by other plants, for instance understorey vegetation. In addition to tree trunks, understorey vegetation has a significant impact on water flow especially when its density becomes larger. In this study, we realize that it is very difficult to exactly model a natural forest condition. In a forest area, trees would be different in size, trunk diameter, canopy cover, direction and size of branches. Our tree models will only consider a limited number of properties of the main trees and understorey vegetation.
The development of tree models is divided into two parts. The first part concerns the development of tree models for dominant trees and the second part contributes to additional development of models of understory vegetation. The latter is considered only for trees in a forested area. Tree models of both dominant trees and understory vegetation are represented by sets of cylinders with different length, diameter, origin, and direction (see Figure 2). A dominant tree is designed to have four canopy layers. Each canopy layer contains six major branches each of which consists of another eight sub-branches. Understorey vegetation contains three canopy layers with a branch structure similar to the dominant tree.

There are several other factors of the branch properties that have been taken into account to mimic the actual conditions of dominant tree and understory vegetation. The first factor is that the branches always point to the sky and each branch is allowed to rotate randomly in three directions (X, Y, and Z axes). A tree branch is divided into a major branch and sub-branches. The rotation of the major branches affects the position and rotation of the corresponding sub-branches. The tree trunk diameter is chosen at random for each tree, which will finally define the size of the tree, i.e., diameter and length of branches, and the height of the tree. This leads to different gaps between trees and subsequently adds more variation on laser penetration and interception over the tree models in a forest patch. For each scenario of trees in forest patches, 25 dominant trees were created and further combined with understory vegetation. The branching scheme for the understory vegetation is similar to the dominant trees except that each plant has a fixed size, the location is random underneath the trees, and the density varies from 0 to 100%. The maximum number of understory vegetation is 300. Figure 3 depicts the entire procedures in developing an individual forest patch.

In order to cover various conditions of forest, the set of simulated forest patches should take into account two important aspects, namely, first, vegetation density at each level particularly above the maximum height of the water level (2.0 m) and, second, the variation of $D_v$ between 0.1 and 2.0 m height. The vertical profiles of canopy cover for all forest patch scenarios should be more or less distributed evenly over the forest height as shown in Figure 4(a).

![Figure 2](https://iwaponline.com/jh/article-pdf/15/2/446/386937/446.pdf)  
**Figure 2** | Scheme of tree branches for dominant trees and understory vegetation. The size (length and diameter) of the sub-branches becomes smaller when we move away from the tree trunk.

![Figure 3](https://iwaponline.com/jh/article-pdf/15/2/446/386937/446.pdf)  
**Figure 3** | Sequence of process required in defining the tree models for the dominant trees and understory vegetation. The trunk diameter for understory vegetation is fixed, thus changes in tree properties only take place in the rotation of the branches.

![Figure 4](https://iwaponline.com/jh/article-pdf/15/2/446/386937/446.pdf)  
**Figure 4** | Vertical distribution of $D_v$ for simulated forest patches of three different cases. The thin black lines represent $D_v$ for each forest patch scenario and the thick line defines the overall shape of the $D_v$ distribution: (a) vertical distribution of $D_v$ is almost evenly distributed among different height levels, (b) denser canopy at the top, and (c) denser canopy cover in the middle and lower part of the forest stand.
In another cases as shown in Figure 4(b) and Figure 4(c), most of the LiDAR interception will be from the top of the forest and the middle part of the forest. The vertical profiles of canopy cover as shown in Figure 4(a) can be obtained by having enough forest patch scenarios for different tree trunk diameters that define variations in tree size and tree height.

**Development of simulated laser pulses of an airborne LiDAR system**

The simulation of the airborne LiDAR laser pulses considers several specifications of the actual FLI-MAP 400 system (see Table 1). It is assumed that the system is able to record 75 laser pulses per m², which is the average point density of the FLI-MAP 400 system. This is also the nominal point density of the FLI-MAP 400 data set acquired on 28 March 2007 at 100-m flying height by Fugro Aerial Mapping Company for a forest patch in Fortmond. In this study the point density of an airborne LiDAR system is defined as the number of laser pulses reflected by a specific area on a flat surface.

The simplified system does not consider complex interactions between laser pulses and vegetation parts as well as the ground surface. Therefore, we do not consider any signal losses. The terrain beneath the tree is assumed to be flat. According to the specification of the actual system, the FLI-MAP 400 system would only be able to collect about 50% of the incoming pulses when the look angle increases to 7° forwards and backwards. Based on this, our simulated system reduces the number of laser pulses linearly to 50% from 1 to 7° for both forward and backward angles. The horizontal and vertical accuracies of the simulated laser pulses are set according to the specifications. Random horizontal and vertical errors are assumed and added to each point. The impact of point density is further studied by a reduction of laser pulses from 100% (75 points per m²) to 10% (8 points per m²).

The simulated laser pulses are a set of lines originating from one source located above the tree models and spread over the entire flat ground surface. Here, we ignore the effect of the moving platform on the distribution of laser pulses.

<table>
<thead>
<tr>
<th>Property</th>
<th>FLI-MAP 400 specification</th>
<th>Simulated system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple return capability</td>
<td>Maximum 4 returns per pulse</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Laser intensity capture</td>
<td>For all laser returns</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Laser point density (single pass)</td>
<td>Typically between 50 and 100 points per square meter, depending on data collection parameters</td>
<td>75 points per square meter, decreasing linearly up to 50% when tilted 7° forward and 7° backward</td>
</tr>
<tr>
<td>Laser ranging accuracy</td>
<td>Better than 1 cm</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Laser scan angle</td>
<td>60°</td>
<td>60°</td>
</tr>
<tr>
<td>Laser look angles</td>
<td>Nadir (50% of pulses)</td>
<td>Nadir (50% of pulses)</td>
</tr>
<tr>
<td></td>
<td>Forward looking (7°) (25% of pulses)</td>
<td>Forward looking (7°) (25% of pulses)</td>
</tr>
<tr>
<td></td>
<td>Backward (7°) (25% of pulses)</td>
<td>Backward (7°) (25% of pulses)</td>
</tr>
<tr>
<td>Laser footprint</td>
<td>Approximately 5.7 cm at 100 m altitude</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Maximum operating height</td>
<td>400 m</td>
<td>Operating height is 100 m</td>
</tr>
<tr>
<td>Total system accuracy (absolute)</td>
<td>8 cm horizontal at 95% confidence interval</td>
<td>8 cm horizontal at 95% confidence interval</td>
</tr>
<tr>
<td>Digital still imagery</td>
<td>5 cm vertical at 95% confidence interval</td>
<td>5 cm vertical at 95% confidence interval</td>
</tr>
<tr>
<td></td>
<td>11.00 megapixel</td>
<td>Not applicable</td>
</tr>
<tr>
<td></td>
<td>Forward and down looking perspectives</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Forward view at 30° to the horizon</td>
<td></td>
</tr>
<tr>
<td>Line scan imagery</td>
<td>Integrated line scan camera fitted to laser scanner used to generate RGB values for each laser return</td>
<td>Not applicable</td>
</tr>
</tbody>
</table>
hits. The platform is assumed static at particular locations above the simulated forest, which is possible for the case of a helicopter.

Airborne LiDAR observations are simulated at five different positions, namely, nadir and four off-nadir views above trees in a forest patch and above isolated trees. The off-nadir views are generated by shifting the X and Y coordinates of the nadir view by ±40 and ±15 m, respectively. This is to make sure that the simulated system will have optimum positions that allow good laser penetration through tree canopies, thus increasing the chances of getting laser hits from the lower part of the tree.

Simulation of airborne LiDAR observation over tree models

Simulation of airborne LiDAR interceptions over simulated forest patches and isolated trees is based on the concept of intersections between a set of lines and a set of cylinders. Figure 5 illustrates how the intersection between a single line and a set of cylinders yields several intersection points. The valid point is the highest hit, while the other points are occluded. If no intersection was found, the return pulse is assigned to the ground surface. In order to obtain a good regression model for vegetation density \( D_v \) estimation, the simulation should be able to generate a point cloud that covers almost all forest conditions. In this study, we generated 686 forest patch scenarios (tree trunk diameter varied between 0.3 and 2.0 m) which were combined with 10 classes of laser pulses densities, that is, in total we generated 6,860 sets of point clouds. As in the case of validating the method for DBH estimation, the properties of simulated isolated trees need to be varied. The distribution of laser hits over a single tree is very useful, especially for the evaluation of a DBH estimation method that directly measures the tree DBH on the point clouds.

The main aim of the airborne LiDAR campaign simulation over a single tree is to validate the DBH estimation method and to investigate the performance of this method against different point cloud densities. The DBH estimation method was previously introduced by Rahman et al. (2009) and the estimated values were validated based on a small number of field data. In this study there are 72 single tree scenarios with tree trunk diameters varying from 0.3 to 2.0 m. With 10 sets of laser pulse densities, this adds up to 720 sets of generated point clouds. Figure 6 shows the whole experimental setup of the simulations.

Estimation of vegetation density for trees in forest patches based on a regression model

Instead of relying on the field data, generated point clouds and the tree models are an ideal solution to develop robust regression models for vegetation density \( D_v \) estimation. As previously explained, this approach allows us to develop as many as needed different scenarios of tree properties in combination with the understorey vegetation to cover different conditions of trees in nature. Now, \( D_v \) and the corresponding predictors can be easily calculated from the tree models and generated point clouds, respectively. In this study, the value of \( D_v \) is estimated based on a method by Petryk & Bosmajian (1975). The value of \( D_v \) is defined as the ratio of total frontal area \( A \) \((m^2)\) of each plant stem and branch \( i \) to \( n \) number of stems and branches in the direction of flow per unit flow volume \( (m^3) \), \( ah \), where \( a \) is the horizontal area \((m^2)\) where the plants stand and \( h \) is equal to the height of the flow (m).

\[
D_v = \frac{1}{ah} \sum_{i=1}^{n} A_i
\]

The calculation method of \( D_v \) is applied on tree models in different scenarios of forest patches. In this case, the total frontal area for the tree models of dominant trees and understorey vegetation is equivalent to the total area of sliced
cylinders with four possible scenarios (see Figure 7): (1) the total cylinder is inundated, (2) conditions where part of the cylinder is inundated, and (3) the last condition where part of a tilted cylinder is inundated by water. In the first case, the frontal area $A$ for the cylinder is calculated as $dh_{cyl}$ where $d$ (m) is the cylinder diameter and $h_{cyl}$ (m) is the height of the cylinder. In the second case, $A$ is calculated as $2.0d$ and in the third and fourth cases as $dh_{cyl} - A_d$, where $A_d$ (m$^2$) is the non-submerged area marked by the shaded area (see Figure 7(b)).

The calculation of the total frontal area of each plant stem $\Sigma A_i$ used here is adopted from a method by Arcement & Schneider (1989), which defines $\Sigma A_i$ as:

$$\sum_{i=1}^{n} A_i = h \sum_{i} q_i d_i$$

(2)

where $h$ is the maximum flood level (2.0 m), $q_i$ is the number of tree trunks, and $d_i$ (m) is the trunk diameter. Apart from the calculation of $D_v$ on the tree models, the predictors of the regression models are calculated using the point clouds generated from the same tree models. The linear regression models are developed by using three different predictors, namely percentage index ($P$), laser interception index ($LI$) and low points index ($LP$). The percentage index $P$ (Straatsma & Baptist 2008) is defined as the percentage of laser hits that fall within the height range $[h_1, h_2]$ in the canopy (see Equation (3)):

$$P(h_1, h_2) = \frac{1}{h_2 - h_1} \frac{N_{h_1-h_2}}{N_{tot}}$$

(3)

where $N_{h_1-h_2}$ is the number of laser hits reflected by vegetation between heights 1 and 2 above the ground surface ($h_2 > h_1$), $N_{tot}$ is the total number of laser hits including vegetation and ground surface. In this study, $h_1$ is set to 5 cm and $h_2$ to the maximum water level of 2.05 m that account for the maximum vertical error of the FLI-MAP 400 system. The first part of the equation is used to compensate for higher vegetation that increases $N_{tot}$ but does not necessarily increase the vegetation density.

The laser interception index, $LI$ is based on a method by MacArthur & Horn (1969) which was used to estimate the leaf area index ($LAI$) of a horizontal canopy layer. For the
LAI computation, let \( D(h) \) denote the density of foliage at height \( h \) (m) above ground, whereas \( \varphi(h) \) denotes the probability of a point quadrant passing through the lowest \( h \) meter without being intercepted. Equation (4) can be used to compute LAI for the canopy layer between \( h_1 \) and \( h_2 \) (MacArthur & Horn 1969):

\[
LAI(h_1, h_2) = \int_{h_1}^{h_2} D(h) dh = \ln \frac{\varphi(h_1)}{\varphi(h_2)}
\] (4)

Equation (4) is further simplified to Equation (5) by removing the common denominator of \( \varphi(h_1) \) and \( \varphi(h_2) \):

\[
LAI(h_1, h_2) = \ln \left( \frac{N_{h_1}}{N_{h_2}} \right)
\] (5)

where \( N_{h_1} \) and \( N_{h_2} \) are the numbers of laser hits whose height exceeded \( h_1 \) and \( h_2 \) with \( N_{h_1} \geq N_{h_2} \). The first part of Equation (3) is used in Equation (6) to make LI independent of the height interval:

\[
LI(h_1, h_2) = \frac{1}{h_2 - h_1} \ln \left( \frac{N_{h_1}}{N_{h_2}} \right)
\] (6)

We propose a new predictor called the low points index, \( LP \) which is defined as the ratio between the total number of laser hits between \( h_1 \) and \( h_2 \) \((N_{h_1-h_2})\) and the number of laser hits on the ground surface \((N_{grd})\).

\[
LP(h_1, h_2) = \frac{1}{h_2 - h_1} \frac{N_{h_1-h_2}}{N_{grd}}
\] (7)

The performance of each predictor in estimating \( D_v \) is evaluated through the regression coefficient values of regression models constructed using generated point clouds with 75 points per m². In addition, the consistency of the predictors on the data set with density lower than 75 point per m² is inspected by using the regression coefficient of regression models constructed with lower density data set. The final regression model will be constructed using the best predictor with the data set of 75 points per m², which is the nominal point density of the FLI-MAP data. The performance of this model on the lower point density data set is analysed using the mean absolute error or difference (MAE) (Equation (8)) and root mean squared error (RMSE) values (Equation (9)) of each lower density data \((D_{v\text{estimate}})\) as:

\[
\text{MAE}_{D_v} = \frac{1}{n} \sum_{i=1}^{n} |D_{v\text{estimate}} - D_{v\text{treemodel}}| \] (8)

\[
\text{RMSE}_{D_v} = \sqrt{\frac{\sum_{i=1}^{n} (D_{v\text{estimate}} - D_{v\text{treemodel}})^2}{n}}
\] (9)

where \( D_{v\text{treemodel}} \) is the \( D_v \) measured over the tree models with different densities of understorey vegetation, \( D_{v\text{estimate}} \) is the \( D_v \) estimated using the regression model developed using the generated point clouds of 75 points per m² that is applied on the lower density datasets, and \( n \) is the number of observations.

**DBH estimation directly from the point clouds for isolated trees**

The estimation of tree DBH is based on the method explained in Rahman et al. (2009). It was shown that estimating tree DBH directly on the point clouds for trees located in a forest area is rather challenging. In a forest area, dominant trees might be accompanied by understorey vegetation and very close neighbouring trees. These two factors raise two different problems. The first problem is that the very close-by trees will intercept laser pulses and eventually decrease the number of laser hits, especially from the lower part of the tree we want to observe. The second problem is that the laser hits from the neighbouring trees and understorey vegetation at locations very close to the tree trunk are considered as noise that should be removed prior to the DBH estimation.

In this study, the DBH estimation is carried out on isolated trees. Point clouds for isolated trees with understorey vegetation will not be filtered and the estimation of \( D_v \) is based on the regression model developed for trees in forest areas. However, as explained before, for the area outside the forest patch in Fortmond, the point density is rather low, i.e., about 15 points per m² in which the performance of the tree DBH estimation for isolated trees should be
evaluated. We will investigate the performance of this method based on the generated point clouds for isolated trees by using the MAE of DBH (Equation (10)) and the RMSE value to indicate the precision of the estimation (Equation (11)):

\[
\text{MAE}_{\text{DBH}} = \frac{1}{n} \sum_{i=1}^{n} |\text{DBH}_{\text{estimate}} - \text{DBH}_{\text{treemodel}}| \quad (10)
\]

\[
\text{RMSE}_{\text{DBH}} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\text{DBH}_{\text{estimate}} - \text{DBH}_{\text{treemodel}})^2} \quad (11)
\]

where \(\text{DBH}_{\text{estimate}}\) is the estimated DBH values from the generated point clouds, \(\text{DBH}_{\text{treemodel}}\) is the DBH values of tree models, and \(n\) is the number of trees. Furthermore, the investigation will also consider different densities of laser pulses, bias in the DBH estimation, and the influence of number of laser hits over tree trunks on the estimated DBH.

Since the DBH estimation is done only on the single trees without understorey vegetation we will skip the tree filtering and noise removal steps. Nonetheless, several conditions should be defined to select suitable trees for the DBH estimation process.

In this study, suitable isolated trees are defined through several requirements. First, each individual tree needs to be delineated by identifying its crown (Rahman & Gorte 2009). The products of this process are tree crown segments and tree locations. The point cloud under the tree crown is selected for each tree segment. For each tree segment a histogram is created with the point clouds divided into several height classes at 0.2 m height intervals. The histogram is filtered with a 1D Gaussian filter to produce a rather smooth histogram surface and to this surface multiple Gaussian functions are fitted (see Figure 8). Instead of having one Gaussian-like curve for the tree crown, the detailed histogram construction with 20-cm height bins would exhibit multiple Gaussian-modes for the tree crown. Furthermore, the frequency distribution of point clouds for the understorey vegetation would also have multiple Gaussian-modes. As shown in Figure 8, the ground surface exhibits a Gaussian-mode at the left end of the histogram. The start level of the tree trunk is defined by using the lowest Gaussian-mode of the tree crown and the starting level is calculated as \(\mu - 3\sigma\). The end of the tree trunk is defined using the Gaussian-mode of the ground surface and the end level is calculated as \(\mu + 3\sigma\).

We impose three conditions to select suitable isolated trees for the DBH measurement. The first condition is that the length of the exposed tree trunk should be at least 2.0 m. Furthermore, a vertical error of point clouds belonging to the ground surface beneath the trees is assumed at 0.5 m. The second condition is that the End level of the tree trunk should be lower than 0.5 m, otherwise the tree segment is considered to contain understorey vegetation and \(D_v\) is estimated as for a forest patch. The final condition deals with the minimum number of points required and this is defined later in this study.

On the left side of Figure 8 are the corresponding histograms for the trees in which the thick lines represent the smoothed histogram surface and fitted Gaussian functions, and the dotted lines represent the boundary that marks the end and start of a tree trunk. For example, a tree suited to DBH measurement is shown in Figure 8(a). The End is lower than the Start, the End line is lower than 0.5 m, and the length of the exposed tree trunk is more than 2.0 m. The tree trunk is hardly seen in Figure 8(b), which is clearly caused by the canopy of the tree almost touching the ground. In Figure 8(c) the tree is accompanied by rather dense understorey vegetation which results in a less visible tree trunk. In both examples (Figure 8(b) and Figure 8(c)) the automatically derived Start and End of the tree trunk are not fulfilling the requirements as Start should always be lower than End.

**Calculation of the final vegetation density**

After the best regression model is selected and the performance of the tree DBH estimation method is evaluated, the final vegetation density is calculated using these approaches. The FLI-MAP 400 data are first normalized by subtracting the elevation value of each laser hit to the corresponding digital terrain model (DTM) value. The DTM has been created by interpolating collected laser hits of the ground surface that are obtained based on the triangulation irregular network (TIN) densification algorithm (Axelsson 2000).
The inverse distance weighted (IDW) method is used to interpolate the ground points into a DTM with a spatial resolution of 1.0 m.

At first, the areas containing trees in the FLI-MAP 400 data should be separated into forest patches and isolated trees. This is done manually. After that the values of vegetation density ($D_v$) for trees in a forest patch are calculated by applying the regression model on the FLI-MAP 400 data. The estimation of $D_v$ for isolated trees requires additional processing steps. The tree crown

![Figure 8](image_url)
The delineation method introduced by Rahman & Gorte (2009) is applied on the trees area marked as isolated trees. The polygons of the tree crowns are used to select the corresponding point cloud for each tree segment which is tested according to the rules introduced earlier in this study. The estimation of $D_v$ for isolated trees that do not comply with the set of rules is done with the method for trees in a forest patch. For the isolated trees that comply with the rules, the DBH information is used to calculate the vegetation density using Equation (1).

The calculation of $D_v$ (m$^{-1}$) for each segment of isolated trees requires additional processing steps as shown in Figure 9 for the case of Tree1, Tree4, Tree7 and Tree8. The tree crown segment is converted into a raster with 1.0 m spatial resolution. The calculation of $D_v$ takes into account the total projected area ($\Sigma A$) of the tree crown and also the water depth (see Figure 9(a)), and therefore $D_v$ is defined as:

$$D_v = \frac{1}{a_v b_v h} \Sigma_a$$  \hspace{1cm} (12)

where $a_v$ (m) and $b_v$ (m) are defined in Figure 9, $\Sigma A$ is rewritten as $hd$, where $h$ is the water depth and $d$ is the tree trunk diameter. Vegetation density $D_v$ maps should be generated at different spatial resolutions depending on the resolution of the composite hydrodynamic roughness maps. There are six possible cases, where in each case $D_v$ is calculated for each pixel either using the regression model, Equation (12), or a combination of these approaches. The latter is more likely to occur when the spatial resolution of the $D_v$ map becomes coarser. Figure 9 shows the concept used to calculate the final $D_v$ for each pixel, for example, at a spatial resolution of 20 m.

In the case of pixels A and E, the value $D_v$ for each pixel with a spatial resolution of 20 m is calculated using Equation (12). For pixel B, the value of $D_v$ is calculated using the regression model. Next in pixel C though, there are two separate tree areas in a single pixel, in this study we assume that these areas are connected. The point clouds for Tree5 and Tree6 are combined and the $D_v$ value in pixel C is calculated using the regression model. For pixel D, the value of $D_v$ is a combination between the $D_v$ value obtained through the regression model for Tree3 and the DBH value for Tree1. Once again as for pixel C, we assume that Tree5 and Tree1 in pixel D are connected. In pixel D, the value of $hd$ for Tree1 is $(N_n/N_t)(hd)$, where $N_n$ is the number of pixels of Tree1 within pixel D and $N_t$ is the total number of pixels for Tree1. The total frontal area of vegetation ($\Sigma A$) for Tree3 is calculated as $a_v b_v h D_v$. Therefore, the value of $D_v$ for pixel D can be calculated using Equation (12) with $\Sigma A = (N_n/N_t)(hd) +$.

**Figure 9** Example illustrating the estimation of vegetation density, $D_v$, for each pixel in the vegetation density maps. Tree crown segments for trees with DBH values are converted to a raster with 1.0-m resolution (a). The spatial resolution of the vegetation density map is assumed to be 20 m (b).
Finally, the $D_v$ value for pixel $F$ is calculated using Equation (13):

$$D_v = \frac{h}{a_v b_v h} \sum_{i=1}^{n} d_i$$

where $a_v b_v h = h(N_{T,7} + N_{T,8})$, where $N_{T,7}$ indicates the number of pixels for Tree7, and $N_{T,8}$ the number of pixels for Tree8.

**RESULTS AND DISCUSSION**

Simulated point clouds for the tree models in forest patches

The generated point clouds derived from the simulated airborne LiDAR observations over various conditions of forest patches gave good results (see Figure 10). The results have successfully shown the effect of different layers of

![Figure 10](https://iwaponline.com/jh/article-pdf/15/2/446/386937/446.pdf)

Figure 10 | Examples of simulated point clouds of the tree models in forest patches with understorey vegetation density varies from 20 to 100%.
vegetation, including the forest canopies and understorey vegetation on the laser pulse interception. It was also shown that another portion of the laser pulses travels below the tree canopies and understorey vegetation and hits the ground surface. The vertical structure of the forest stand is clearly visible in the point clouds including tree crown, tree trunk, understorey vegetation and the ground surface. This is a typical property of airborne LiDAR observations during leaf-off conditions in a forest area. With large variations of trees in the simulated forest patches, the generated point clouds are quite different from one scenario to another.

Development of regression models

Estimation of vegetation density $D_v$ for trees in a forest patch is based on a linear regression model, which relies on the tree models to reconstruct forest patches, while the predictors are calculated using the generated point clouds. The regression models take different densities of understorey vegetation and variations in the tree properties into account that might closely represent the real conditions of a natural forest. In this section, the performance of $LI$, $P$ and $LP$ as predictors are evaluated and the best model is selected to estimate the final $D_v$ using the real FLI-MAP 400 data set. For the calculation of the $LI$, $P$ and $LP$ using the FLI-MAP 400 data, height at level 1, $h_1$, is set to 0.5 m and height at level 2, $h_2$, is set to 2.5 m. Only point clouds generated with a laser pulse density of 75 points per m$^2$ are used to generate the final regression model. Meanwhile, the generated point clouds are used to assess the impact of different densities of laser pulses on the performance of all predictors. Furthermore, the impact of applying the regression model constructed with a laser pulse density of 75 points per m$^2$ on lower density data sets is also evaluated. Figure 11 shows the results of the regression analysis on $D_v$ and $LP$, $LI$ and $P$. The $LP$ has the highest regression coefficient value with 0.94. This value is significantly better than $P$ and $LI$ with a regression coefficient of only 0.66 and 0.46, respectively.

Figure 11 | Linear regression models for low point index ($LP$), laser interception index ($LI$) and percentage index ($P$).
Further investigation has been carried out on the effect of the variation of airborne LiDAR point density from 10% (8 points per m²) to 100% (75 points per m²) on the resulting regression models. The effect of applying the regression model of LP on the generated point clouds with lower point densities is shown in Figure 12. The RMSE and the MAE values are rather low, which indicates that the regression models developed using 75 points per m² generated point clouds can be used on any area and data set with lower point density. The difference between these values for all data sets with different point densities is small, which indicates that the magnitudes of the individual errors are almost the same. Therefore, it can be concluded that the LP can be used to estimate $D_v$ values rather accurately and consistently even when the regression models are generated using a lower density point cloud.

DBH estimation from point clouds

Figure 13 shows six examples of the generated point clouds for each isolated tree, where each isolated tree scenario has rather unique tree properties reflected by the random properties of the cylinders. The results show that each tree scenario has different distributions of point clouds over the entire tree.

The RMSE for the simulated laser pulse density of 75 points per m² suggests that with a perfect round shape of the tree trunk and without noise in the point clouds (no reflected laser pulses from the tree branches along the tree trunk), the tree DBH can be estimated with 0.25 m accuracy. The MAE for the simulated laser pulse density of 75 points per m² is 0.18 m. The RMSE value is higher than the MAE, which suggests that the individual errors have a rather different magnitude and this applies to the rest of the data sets with lower laser pulse densities. However, by decreasing the laser point density, the accuracy of the DBH estimation does not change much up to a density of 8 points per m² (see Figure 14(a) and Figure 14(b)). There is no strong correlation ($R = 0.046$) between the number of laser hits on the tree trunk and the absolute difference between the measured tree trunk diameter ($DBH_{tremodel}$) and the estimated diameter ($DBH_{estimated}$) values for tree models (see Figure 14(c)). The relationship is expected to be strongly negative. The low correlation value is due to the large variation of laser hits’ pattern on the tree trunk. The result as shown in Figure 14(c) does not demonstrate any useful pattern that can help us to define the minimum number of laser hits on the tree trunk required for the DBH measurement.

Therefore, in this study, the minimum number of laser hits on the tree trunk required for the DBH estimation is set to 20. This is an additional requirement to select suitable trees for the DBH estimation as defined earlier. The mean value of $DBH_{tremodel} - DBH_{estimated}$ also can be called as ‘bias’, is equal to 0.17 m, which suggests that the underestimation of the tree DBH is significant.

Estimation of vegetation density using the FLI-MAP 400 data

The estimation of vegetation density, $D_v$, from the FLI-MAP data set begins with the extraction of point clouds representing isolated trees and forest patches. The area is separated from the other landcover classes through a landcover classification process. Next, the forest patches and the isolated trees are manually separated. Individual tree crown delineation is done on the isolated trees using the method described by Rahman & Gorte (2009). The minimum and maximum tree crown radius is set to 2.0 and 6.0 m, respectively. In order to speed up the tree crown
Figure 13 | Comparison of 3D line fitting on the points representing the tree trunk for simulated isolated trees (first and second row) and trees found in the FLI-MAP 400 data set (third row). In both cases, the 3D line in the DBH estimation method is fitted over various patterns of laser hits on the tree trunk.
delineation, the point clouds lower than 5.0 m are excluded from the data set. The individual tree crown delineation has identified about 5,538 isolated trees. With the information of the individual tree crown segments, we can extract the corresponding point clouds from the airborne LiDAR data. The point clouds are then used to select suitable trees for the DBH estimation using the method and conditions as discussed earlier in this study. Only 23 isolated trees were used for the DBH estimation process and this is due to the following reasons:

1. Low number of laser hits on the tree trunk.
2. Dense understorey vegetation.
3. Tree canopy almost touching the ground, which makes the tree trunk less visible.

For the rest of the isolated trees, the $D_v$ values are estimated using a regression model as given in Equation (14). Vegetation density, $D_v$, for trees in forest patches is estimated using a regression model with the $LP$ as the predictor:

$$D_v = 9.75LP + 0.05 \quad (14)$$

The $LP$ calculation is made on each pixel over the entire airborne LiDAR data. In this study, the $D_v$ maps are generated with different spatial resolutions. Therefore, the $LP$ calculation should also be made with different spatial resolutions (5, 10, 20, 30, 40 and 50 m). On the other hand, for the selected isolated trees with the measurable tree DBH, the $D_v$ is estimated using Equation (12). However, as the spatial resolution of the $D_v$ maps decreases, there are several situations where both methods are combined to calculate the final $D_v$ value. The $D_v$ maps are generated with different spatial resolutions (5, 10, 20, 50, 40 and 50 m) (see Figure 15). The productions of multi-resolution $D_v$ maps are required to generate multi-resolution composite hydrodynamic roughness maps.

**CONCLUSION**

We have introduced a method for vegetation hydrodynamic roughness estimation that accounts for different densities of understorey vegetation. Instead of relying on the field data, we have proposed a simulation of airborne LiDAR observations over various forest conditions. The simulator is developed based on the simplified FLI-MAP 400 system and the simulation is made over simulated forest patches and isolated trees. The results show that high density airborne LiDAR can be used to estimate the hydrodynamic vegetation density with two different approaches.

The results have also shown that one of the key features of the airborne LiDAR systems that enable this kind of research is the off-nadir viewing capability. With this feature we have a better chance of getting laser hits from the lower part of the forest and from tree trunks (Straatsma 2005). In general, a higher number of laser hits on the tree trunks can be obtained with larger observation angles of laser pulses. We admit that modelling the real structure of a tree and understorey vegetation is rather difficult. What has been shown in this study, however, is that the simulated trees would represent the simple form of the tree structure in reality. But it has been also proved that with some degree of simplification, the generated point clouds gave quite reasonable results that we might have expected from an airborne LiDAR observation during a leaf-off season. Further improvement on simulation of the detailed structure of trees and understorey vegetation is highly recommended. In addition, a better approach is still required to account for the impact on the laser hits'
pattern and its density on the DBH estimation. It is shown that there is no strong correlation between the number of laser hits of the tree trunk and the error in DBH estimation. This certainly requires information on the distribution of the laser hits to be part of the correlation analysis. Finally, vegetation density $D_v$ can be estimated and mapped at any desired spatial resolution. This is useful for the estimation of multi-resolution and spatially distributed hydrodynamic roughness values.

**REFERENCES**


Arcement, G. J. & Schneider, V. R. 1989 Guide for Selecting Manning’s Roughness Coefficients for Natural Channels and

---

*Figure 15* | Vegetation density ($D_v$) generated with different spatial resolutions, namely, 5, 10, 20, 30, 40 and 50 m.


Straatsma, M. 2007 Hydrodynamic roughness of floodplain vegetation; Airborne parameterization and field validation. Faculty of Geoscience, Utrecht University, Utrecht, 175 pp.


First received 3 December 2011; accepted in revised form 9 August 2012. Available online 10 October 2012