# A model-based approach for estimating the height distribution of eucalyptus plantations using low-density ALS data

LAURI MEHTÄTALO*†, ANNI VIROLAINEN†, JUKKA TUOMELA† and JUKKA NYBLOM

† Faculty of Science and Forestry, University of Eastern Finland, P.O. Box 111, FI-80101 Joensuu, Finland.

‡Department of Mathematics and Statistics, University of Jyväskylä, P.O. Box 35 (MaD), 40014 University of Jyväskylä, Finland

Under some simplifying assumptions, observations collected by airborne laser scanners can be regarded as observed heights of the canopy surface of the stand. The canopy surface is a surface generated by individual tree crowns. Assuming that individual tree crowns have a solid surface at the top side, and that the stand is viewed directly from above, the height of canopy surface at a given point is defined as the maximum over the tree-specific surfaces at that point. Our earlier studies have shown how the probability of having the canopy surface above a given height depends on (i) the stand density, (ii) the distribution of tree heights, (iii) the shape of individual tree crowns, and (iv) the spatial pattern of tree locations (either random or strictly regular). Based on this result, we have proposed a method based on maximum likelihood for estimating the stand density and height distribution using the observations on the canopy surface. This paper applies the model to the situation of the area-based approach, estimating the mean and dominant height of eucalyptus plantations (E. urograndis in Bahia state, Brazil) with known stand density and rectangular pattern of tree locations. First, a set of 18 training sample plots with known tree heights are used to estimate the parameters specifying the shape of individual tree crowns. These estimates are then used to estimate the distribution of tree heights for another set of 18 evaluation sample plots. The results indicated poorer performance when compared to the widely used empirical area-based approach. The reason was most likely an unrealistic model for individual tree shape. Possible improvements will be studied in the future.

## 1. Introduction

The area-based approach to the forest inventory using airborne laser scanning (ALS) is rapidly developing to an operational tool for forest inventory (e.g. Naesset et al 2004). The idea of the approach is to combine information measured on ground sample plots with wall-to-wall information collected by low-pulse density (about one pulse per $m^2$) discrete return airborne laser scanners. A regression approach, either parametric or non-parametric, is used in generalizing the relationship of ALS data and ground measurements from sample plot locations to other locations. The predictors of the regression model are different characteristics of the laser hits, such as quantiles of the laser data, or proportion of laser hits below a given threshold height. We call this approach as the *empirical area-based approach* in contrast to the *model-based approach* proposed in this paper

The ALS data collection procedure provides data that includes direct measurements of the canopy surface. This is the main difference to other remote sensing methods, such as to the use of satellite images and aerial photographs, and explains the good performance of the method. However, the relationship between the

*Corresponding author. Email address: lauri.mehtatalo@uef.fi

laser data and tree characteristics is not very well known, and selection of the predictors to the models of the empirical area-based approach is based mostly on intuition and empirical findings, not on well-established theory on how laser observations are generated by individual tree crowns.

From a theoretical point of view, there are two main factors that have an effect into the ALS observations: (i) the vertical profile of the forest canopy, and (ii) the physical properties of the laser measurements. A good theoretical model would integrate both of these into a single model. A starting point for such a model was proposed in Mehtätalo and Nyblom (2009). Their model expressed the vertical profile of laser observations for a stand as a function of stand density, and the parameters of the height distribution of trees and the shape of individual tree crowns. The model was used to estimate the stand density and tree height distribution in simulated forest stands. However, the model included several simplifications, such as the assumption of random tree locations within the stand. Mehtätalo and Nyblom (2010) developed the model into a systematic, square grid spatial pattern, and evaluated the model both using simulated data and by using pre-processed empirical data. However, both above-mentioned studies assumed that the canopy surface is solid, the shape of a single tree as a function of tree height is known in advance, and the laser pulses are taken directly from above at an infinitesimal point. In general, the model satisfactorily considered the process that generates the vertical profile of the forest canopy, but did not tackle the physical properties of the laser measurements.

The aim of this study was to apply the model of Mehtätalo and Nyblom (2010) into the situation of the area-based forest inventory, and to test it with empirical data. The idea is to first estimate the parameters specifying individual tree crowns using laser data from plots with known tree heights. The information on the crown shape is then used with another set of plots to estimate the distribution of heights. The stand density was not estimated, because the rectangular pattern of tree locations was known in advance.

## 2. Material

The study material includes sample plots from the pulpwood plantation growing eucalyptus in Bahia state, Brazil. On each sample plot, all trees were callipered for diameter ($d$). Every seventh tree was measured for height ($h$). Näslund's $h$-$d$ curve was fitted by stands and used to predict heights for trees without height measurement. Predicted heights from this model are hereafter regarded as the true heights, and the prediction error in them is ignored.

ALS data were collected on August 16, 2008 using an Optech ALTM 3100 laser scanning system. The flight altitude was 1200 m above ground level using a field of view of 30 degrees. The pulse density at the ground was about 1.5 measurements per square meter and the footprint diameter was about 35 cm. However, to reduce the computational burden, we used only every $6^{th}$ pulse of the original data. After this, we had on average 122 pulses for each of the 530 $m^2$ sample plot (0.23 pulses per $m^2$). The observations below 5 meters were taken as ground hits and treated as zeroes in the analysis.

The plots of this study are a subsample of the original data described in Packalen et al (2010). A total of 18 pairs of sample plots were selected systematically according to the canopy height properties so that they would represent the whole range of mean canopy heights in the plantation. The plots of each pair were then randomly assigned to the training and evaluation datasets. The same set of sample plots was used in Vauhkonen et al. (2010).

## 3. Methods

### *3.1. The model for canopy height*

Let $Z(v)$ be the height of canopy surface at a fixed point $v$, i,.e., the vertical distance between the ground level and the canopy surface. The canopy surface at a given point can be thought as the maximum over the tree-specific surfaces at that point.

Mehtätalo and Nyblom (2009, 2010) showed that the probability of having $Z(v)$ below the height $z$ above ground is the probability that none of the neighbor tree crowns at height $z$ extends to point $v$,

$$P(Z(v) \leq z) = \prod_{i \in \mathcal{N}} F\{w(z, \|v - u_i\| | \boldsymbol{\theta}) | \boldsymbol{\xi}\}, \quad (1)$$

where

$\mathcal{N}$ is the set of neighboring trees of point $v$ (see below)

$F(h|\boldsymbol{\xi})$ is the cumulative distribution function of tree height (e.g., the Weibull cdf),

$w(z,r|\boldsymbol{\theta})$ is a function that gives the total height for a tree that has crown radius $r$ at height $z$ above ground,

$\|v - u_i\|$ is the distance between the fixed point $v$ and the location $u_i$ of tree $i$

$\boldsymbol{\theta}$ includes the parameters of the height distribution and

$\xi$ includes the parameter that specify the shape of individual tree crown as the function of tree height

Equation (1) assumes independent and identically distributed tree heights within the stand. Furthermore, it assumes that tree crowns are solid so that the observations of $Z(v)$ do not penetrate into tree canopies. The horizontal cross section of a tree crown is assumed to be circular or at least the orientation is assumed to be uniform between 0 and 360 degrees so that trees are circular on average. The set of neighboring trees $\mathcal{N}$ includes the trees that are located so close to the point $v$ that their crowns (practically) have a nonzero probability to extend to the point $v$. In this study, we used a neighborhood of 16 trees (4 trees from 4 rows).

When laser observations are used, no information on the tree locations $u_i$ with respect to $v$ is available. However, it is natural to assume that the observations are placed uniformly over the stand area. Then it is justified to estimate $P(Z \leq z)$ as the mean of $P(Z(v) \leq z)$ over the stand area. However, if trees are planted into rows with distance $l$ between rows, and having distance $m$ between trees of a row, then the stand consists of $N$ cells of size $m$ by $l$, which all have equal mean of $P(Z(v) \leq z)$ (we forget the cells that are at the stand edge). Then it is enough to take the mean of $P(Z(v) \leq z)$ over only one cell. This mean is the integral of the right hand side of the equation 1 over the cell, divided by the cell area

$$G(z|\boldsymbol{\theta},\boldsymbol{\xi}) = P(Z \leq z) = \frac{1}{lm} \iint_0^{l,m} \prod_{i \in \mathcal{N}} F\{w(z, \|v - u_i\| | \boldsymbol{\theta}) | \boldsymbol{\xi}\} \, \mathrm{d}v_1 \mathrm{d}v_2. \quad (2)$$

By the definition of the cumulative distribution function (cdf, Casella and Berger 2001), this is the cdf of the laser observations, if the observations are taken at infinitesimal points directly from above and the other previously specified assumptions on the stand structure are met. Function (2) depends on two parameters: $\boldsymbol{\xi}$, which specifies the distribution function of the tree heights, and $\boldsymbol{\theta}$, which specifies the shape of an individual tree crown. In contrast to Mehtätalo and Nyblom (2010) the stand density is assumed to be known. If it were unknown, then the cell area, the bounds of the integral, and tree locations $u_i$ would be functions of stand density.

### *3.2 Estimation*

The probability density function (pdf) of the non-zero laser observations is obtained by differentiating (2) with respect to *z* as

$$g(z|\boldsymbol{\theta},\boldsymbol{\xi}) = G'(z|\boldsymbol{\theta},\boldsymbol{\xi}) \text{ for z>0.} \tag{3}$$

In addition, the probability to have canopy height of zero is $P(Z \leq 0) = G(0|\boldsymbol{\theta},\boldsymbol{\xi})$.

Assuming that the laser observations are an i.i.d. sample from the distribution specified by (2), the log likelihood is the sum of logarithmic densities over the observations. In addition, the likelihood has a term for the ground hits (see Mehtätalo and Nyblom 2009, 2010),

$$\ell(\boldsymbol{\theta},\boldsymbol{\xi}) = \sum_{j=1}^{M} I(z_j > 0) \ln g(z_j|\boldsymbol{\theta},\boldsymbol{\xi}) + M_0 \ln\ G(0|\boldsymbol{\theta},\boldsymbol{\xi}), \tag{4}$$

where $M_0$ is the number of hits ground hits. The assumption of independence is likely violated because there are several observations per tree in the dataset. Nevertheless, we keep calling $\ell(\boldsymbol{\theta},\boldsymbol{\xi})$ the likelihood.

Estimation of parameter $\boldsymbol{\xi}$ or $\boldsymbol{\theta}$ is based on maximizing the likelihood $\ell(\boldsymbol{\theta},\boldsymbol{\xi})$ with respect to the parameter of interest. The variance-covariance matrix of the estimation errors can be approximated by the inverse of the negative Hessian matrix at the solution (e.g. Casella and Berger, 2002). However, the variances are expected to be underestimates due to the lack of independence among the observations (Mehtätalo and Nyblom 2009).

### *3.3 The assumed functions for crown shape and height distribution*

The height distribution of trees was assumed to be of the Weibull form

$$F(h|\boldsymbol{\xi}) = 1 - \exp\left\{-\left(\frac{h}{\beta}\right)^{\alpha}\right\}, \tag{5}$$

where $\boldsymbol{\xi} = [\alpha \quad \beta]'$.

The height of a tree that has radius *r* at height *z* above ground was assumed to be of the form

$$w(z,r|\boldsymbol{\theta}) = \frac{-ry_0 - \frac{b^2}{a^2}z^* x_0 + \frac{b}{a}\sqrt{b^2 z^{*2} + a^2 r^2 - (x_0 r - y_0 z^*)^2}}{b^2 - \frac{b^2}{a^2}x_0^2 - y_0^2} \tag{6}$$

where

$$a = \sqrt{b^2 \frac{(1-x_0)^2}{b^2 - y_0^2}} \text{ and } z^* = \max\left(z, \frac{x_0 r}{y_0 + b}\right),\ a,b>0,\ x_0<1,\ y_0 \le 0.$$

This function is based on assuming the crown shape of a tree of height $H$ to be an ellipsoid that is centered at ($x_0H$, $y_0H$) and has the half axes $aH$ and $bH$ (see Figure 1); this function has been found to fit well to old-growth Norway spruce and Scots pine trees in Finland (unpublished). However, requiring that the ellipsoid passes through the tree top gives a condition, which was used to eliminate parameter $a$. The function $w$ above results from solving this equation of ellipse for tree height $H$. The function $w$ has three parameters: the relative height of the maximum radius ($x_0$), the relative maximum width of the crown ($y_0+b$) and a parameter controlling the shape of the crown ($b$).

To be specific, the ellipsoid shape is assumed only for the part above the maximum crown width. Below that height the crown is assumed to be cylinder with the diameter equal to the maximum crown width. This assumption is realistic because this is how a solid crown looks like when seen from above.

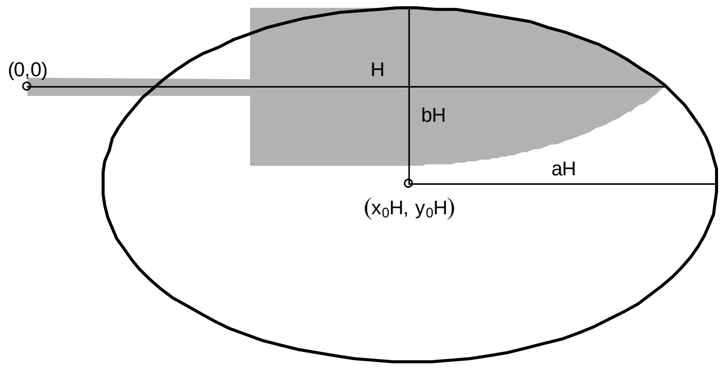



**Figure 1.** Graphical representation of the applied function of crown shape. The shaded object demonstrates a fallen tree of height $H$.

### *3.4 The estimation procedure*

The estimation procedure had the following five steps.

1. The Weibull distribution was fitted to the known heights to get the estimate $\hat{\boldsymbol{\xi}}$ for each of the 18 training plots.
2. The estimates of $\boldsymbol{\xi}$ from step 1 were used in the likelihood (3), which was maximized with respect to the crown shape parameter vector $\boldsymbol{\theta}$. This resulted in estimates of vector $\boldsymbol{\theta} = [x_0, y_0, b]'$ for each training plot.
3. The plot-specific estimates of the crown shape parameters were explored with respect to the mean height of the canopy hits, and a regression was fitted to explain the observed trend.
4. The fitted regressions were applied to the plots of the evaluation data to predict the parameters $\boldsymbol{\theta}$.
5. The predictions of $\boldsymbol{\theta}$ from step 4 were used in the likelihood (3), which was maximized with respect to the Weibull parameters $\boldsymbol{\xi} = [\alpha, \beta]'$. This resulted in an estimated height distribution for each plot of the evaluation data.

In a pairwise fitting approach, the steps 3 and 4 above were omitted. Instead, the estimated crown shape of the pair of the plot under consideration was used in step 5.

For comparison, also the empirical area-based method was implemented for the dataset. In that implementation, the model of Packalen et al (2010) was fitted for both the dominant and mean heights in the training dataset of this study.

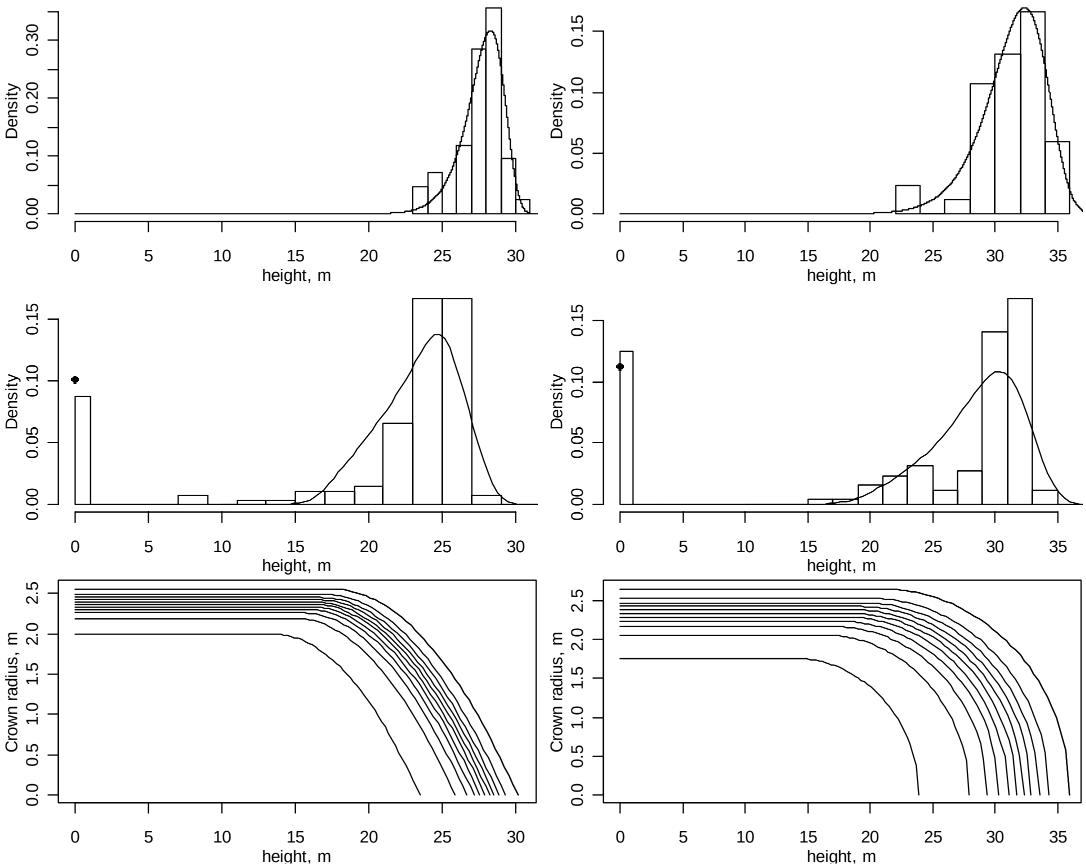



Figure 2. Examples of the fit into the modeling data on two plots. The upper graphs show the true tree heights (histogram), and the fitted Weibull density (line). The middle graphs show observed distribution of $Z$ (histogram), and the fitted density $g(\mathrm{z}|\boldsymbol{\theta},\boldsymbol{\xi})$ (line) and $G(0|\boldsymbol{\theta},\boldsymbol{\xi})$ (point) of step 2 of the estimation procedure. The lowest graphs demonstrate the estimated crown profiles for trees with heights at the $0.01^{th}$, $0.1^{st}$, $0.2^{nd}$, …, $0.9^{th}$, and $0.99^{th}$ quantiles of the fitted Weibull height distribution.

### *3.5. Evaluating the methods*

To evaluate the estimated height distributions, estimated and true mean and dominant heights (mean height of 100 tallest trees per ha) were computed for the evaluation plots. The true values were based on the true trees of the plot.

The estimated values for the model-based approach were based on numerical evaluation of the expected value of the height distribution of all and dominant trees. 95 % confidence intervals for these estimates were computed by using a Monte Carlo approach. In that approach, the approximate asymptotic estimation errors of the plot-specific Weibull parameter estimates were used to generate 1000 realizations of the Weibull parameters for each evaluation plot. The mean and dominant heights were numerically evaluated for each realization. The interval between the $25^{th}$ smallest and $25^{th}$ largest values of the 1000 $H$ and $H_{dom}$ values was then taken as the 95% confidence limit for the plot.

The predictions of the empirical area-based approach were obtained by direct application of the fitted regression models into the evaluation data. Confidence intervals could not be computed for this approach.

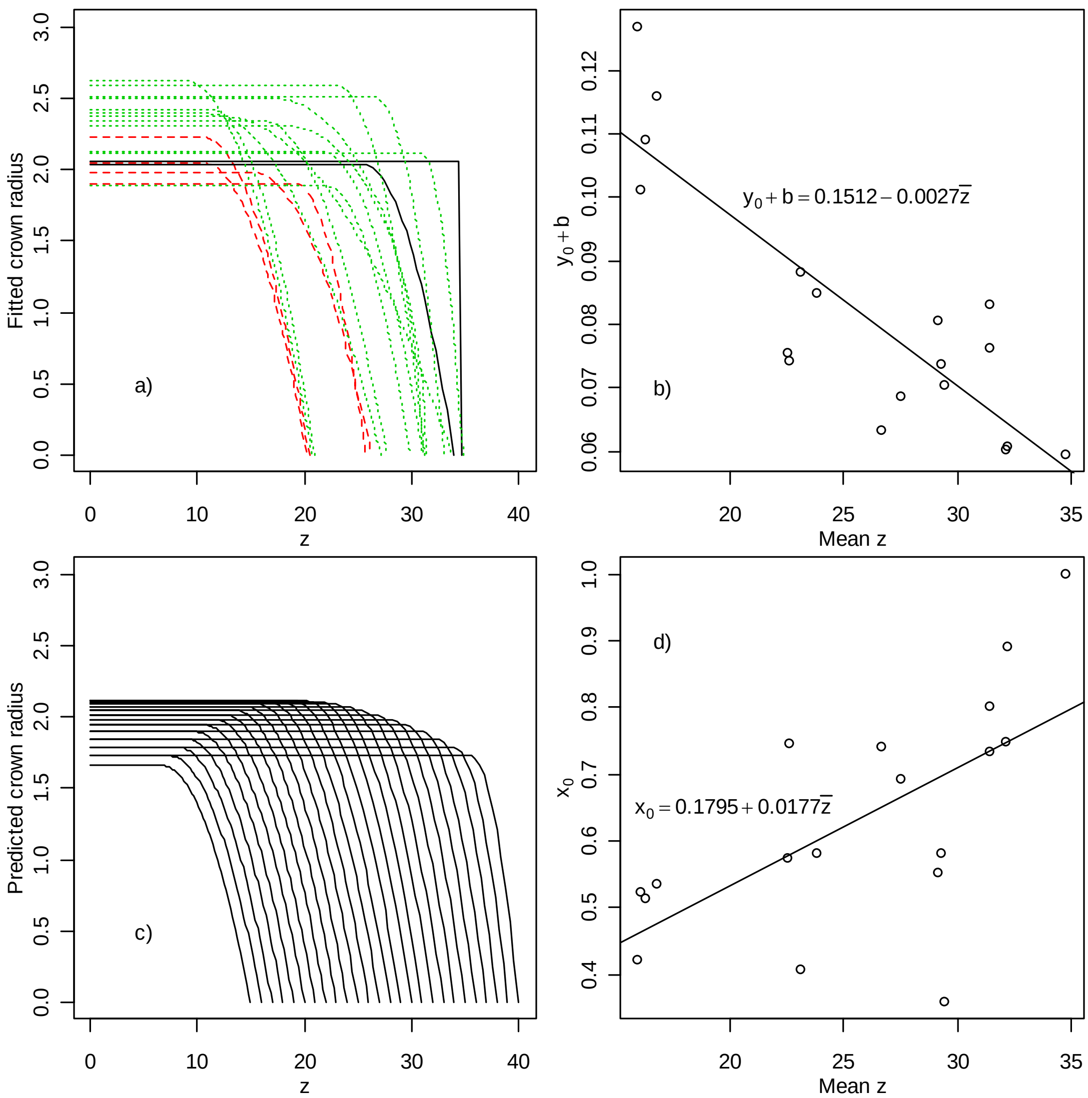
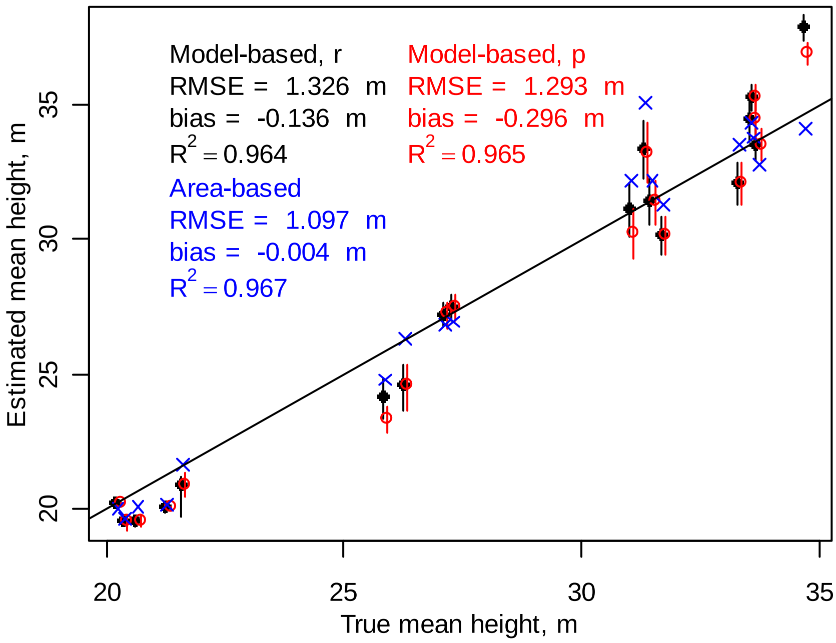



Figure 3. Demonstration for the modeling of individual tree crown shape. Graph (*a*) shows the estimated crown profiles for a tree of a typical height on each plot of the training dataset. Graphs (*b*) and (*d*) show the estimated values for the parameters of a crown shape as a function of average nonzero *z*- values of the plot, and trendlines fitted to these data. Graph (*c*) demonstrates the predicted crown profiles when average nonzero *z*- value varies from 15 to 40 meters.

## 4. Results

Figure 2 shows examples of the fitted Weibull diameter distribution on two sample plots of the training dataset. The Weibull function fits quite well to the tree heights, but the resulting fit of function (3) is not that good. The bottom graphs show that the estimated crown shape may be very different on different plots. This can be seen also in Figure 3(*a*). Figures 3(*b*) and 3(*d*) show that the relative crown width and length are smaller on plots of tall trees than on plots of short trees. Regression models explaining these trends are shown in Figures 3(*b*) and 3(*d*). Predictions of these models in Figure

3 (*c*) show that the absolute shape of tree crowns does not much vary among plots of different mean canopy height, but the crown is at a higher level on plots of larger trees.



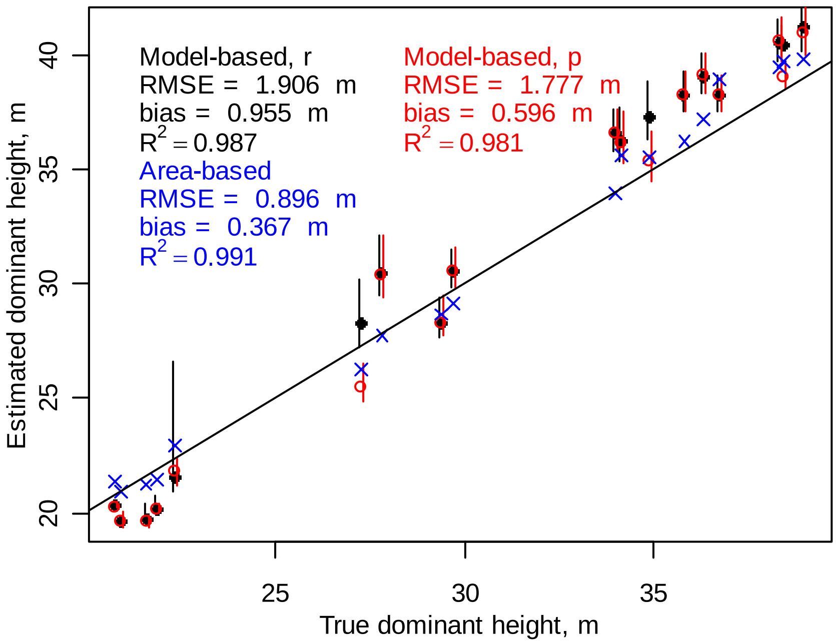




Figure 4. The estimated mean and dominant height of the validation plots on the values based on ground measurements and their confidence intervals. The predictions are based on the model-based method with predicted crown shape (black solid), the model-based method with crown shape of the corresponding pair of the training dataset (red open), and on the empirical area-based approach (blue cross).

Figure 4 shows the estimated mean heights (top) and dominant heights (bottom) for the evaluation plots using the three alternative methods. The RMSE of mean height was 1.33 m for the model based regression approach, and 1.29 m for the model-based pairwise approach. These values were clearly higher than the RMSE of the empirical area-based approach. Similar, but even stronger differences were observed in the case of dominant height. The mean and dominant heights were underestimated for plots of small trees. For the plots of large trees, no noticeable bias can be observed for the mean height, but the dominant height was overestimated. The dominant height had quite strong positive bias.

Figure 5 shows examples of the fits from steps 4 and 5 for typical plots of small (left), medium sized (middle) and large trees (right). Function (3) does not fit very well to any of the plots. Especially, the fitted proportion of ground hits is much higher than the observed value. The estimated distribution of tree heights seems, however, good for the plot of middle-sized trees. For the plot of large trees, the height distribution has a heavy right tail. For the plot of small trees, the maximum predicted height is equal to the minimum of the measured heights, because all laser observations on this plot (middle left graph) were below the minimum measured tree height (bottom left graph). These observations largely explain our earlier notes on Figure 4.

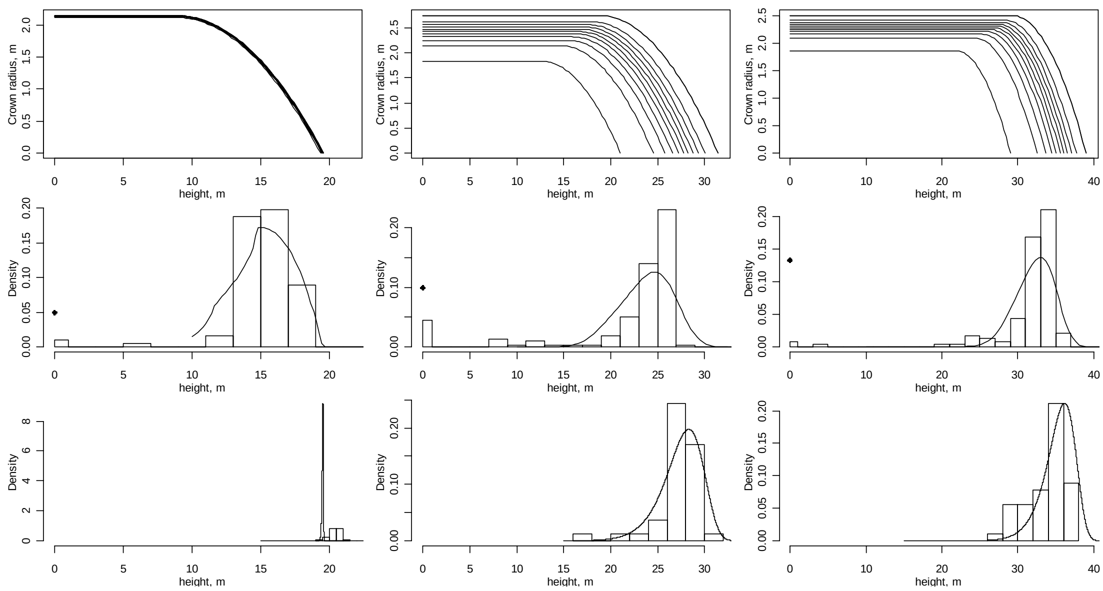



Figure 5. Example fits of step 5 of the estimation procedure for typical stands having small (left), middle-sized (middle), and large (right) trees. The upper graph shows the predicted crown shape for different tree heights, the middle graphs laser observations (histogram) and the fitted density $g(z|\boldsymbol{\theta}, \boldsymbol{\xi})$ (line) and $G(0|\boldsymbol{\theta}, \boldsymbol{\xi})$ (point), and the bottom graphs true tree heights (histogram) and the estimated Weibull height distribution (line).

## 5. Discussion

This paper developed and evaluated a new, model-based approach for the ALS forest inventory. The approach is based on a model that expresses the distribution of canopy height for a given stand density, height distribution, spatial pattern of tree locations, and individual tree crown shape. Detailed description of the model can be found in Mehtätalo and Nyblom (2009, 2010). In this study, the model was generalized for a rectangular spatial pattern. A special feature of the applied situation was that the stand density was known in advance. In addition, this was the first trial to use the model to estimate the parameters of individual tree crown shape.

The results were not as good as one could have expected. We believe that the main reason is an unrealistic assumed model for the individual tree crown. Especially, we assumed that the tree crown has a solid top surface. However, the laser return seems practically always to return from the inner parts of the crown, because the returning energy needs to exceed some limit before it is interpreted as a return (Gatziolis et al 2010). This is most likely the reason for the observed discrepancy between the true tree heights and laser observations in the left graph of figure 5. This effect should be included in our model, for example, by introducing an additional parameter for the penetration into the model individual tree crown.

Another problem was the assumption of similar relative shape among the trees. This assumption was found unrealistic. We relaxed this assumption by allowing different relative shape for different plots. However, a better approach could be a model that allows variation in relative shape also within a plot. In addition, alternative functions could be tested for the crown shape (e.g., Rautiainen et al, 2007).

The proposed method is theoretically justified because the parameters have interpretations arising from the stand structure and tree crown properties. In addition, the applied method of maximum likelihood has a strong theoretical basis, providing justified means to construct confidence intervals for the estimates, and to make likelihood-based tests between different models. However, these intervals may be too narrow because of the lack of independence among observations. In addition, they are valid only if the assumed model is correct. In this study, this was evidently not true, leading to too narrow confidence intervals for the mean and dominant heights.

The definition of the distribution function (Eq. 2) includes an integral over a two-dimensional plane. This function is differentiated with respect to $z$ and the derivative evaluated repeatedly at each observation to get the likelihood. For asymptotic inference, the likelihood needs still to be differentiated twice with respect to the parameters. This computational burden currently restricts the applicability of the method. This is also the main reason for that we did not search for the best possible models for tree canopies in this study. Instead, we just went through the chain of computations to demonstrate how this estimation could be done. Comparison and evaluation of the model with different assumed models is left to the future. In addition, possible approximations are searched for to decrease the computational requirements of the method.

## References


CASELLA, G. and BERGER, R.L. 2002. *Statistical inference. 2nd edition.* (Pacific Grove: Duxbury).

GATZIOLIS, D., FRIED J.S., AND MONLEON, V.S. 2010. Challenges to estimating tree height via LiDAR in closed-canopy forests: a parable from Western Oregon. *Forest Science,* **56(2)**, pp. 139–155.

MEHTÄTALO, L AND NYBLOM, J. 2009. Estimating forest attributes using observations of canopy height: a model-based approach. *Forest Science,* **55(5)**, pp. 411–422.

MEHTÄTALO, L AND NYBLOM, J. 2010. A model-based approach for ALS inventory: application for square grid spatial pattern. *Submitted MS.*

NÆSSET, E., GOBAKKEN, T., HOLMGREN, J, HYYPPÄ, H:, HYYPPÄ, J., MALTAMO, M, NILSSON M., OLSSON, H., PERSSON, Å. and SÖDERMAN, U. 2004. Laser scanning of forest resources: the Nordic experience. *Scandinavian Journal of Forest Research*, **19**, pp. 482–499.

PACKALÉN, P., MEHTÄTALO, L. and MALTAMO, M., 2010. ALS based estimation of plot volume and site index in a eucalyptus plantation with a nonlinear mixed effect model that accounts for the clone effect. Proceedings of Silvilaser 2010, the 10th International Conference on LiDAR Applications for Assessing Forest Ecosystems. September 14th - 17th, 2010, Freiburg, Germany.

RAUTIANEN, M., MÕTTUS, M., STENBERG, P. and ERVASTI, S. 2007. Crown envelope shape measurements and models. *Silva Fennica,* **42(1)**, pp. 19–33.

VAUHKONEN, J., ENE, L, GUPTA, S., HEINZEL, J., HOLMGREN, J., PITKÄNEN, J., SOLBERG, S., WANG, Y., WEINACKER, H., HAUGLIN, K.M., LIEN, V., PACKALÉN, P. GOBAKKEN, T., KOCH, B., NÆSSET, E., TOKOLA, T. and MALTAMO, M., 2010 Comparative testing of single-tree detection algorithms. Proceedings of Silvilaser 2010, the 10th International Conference on LiDAR Applications for Assessing Forest Ecosystems. September 14th - 17th, 2010, Freiburg, Germany.