

Estimating Tree Height Distribution Using Low-Density ALS Data With and Without Training Data

Lauri Mehtätalo, Anni Virolainen, Jukka Tuomela, and Petteri Packalen

Abstract—This study applies an approach based on stochastic geometry for retrieval of forest characteristics from airborne laser scanning (ALS) in two situations: 1) without ground-measured training data and 2) with training data. The applied model treats the ALS echo heights as an outcome of a random process, expressing the observed heights of canopy envelope as a function of stand density, the parameters of the tree height distribution, and the shape of the individual tree crown. The model was applied to a eucalyptus plantation dataset with known spacing, where the main interest was to estimate the plot-specific tree height distribution. Estimation without training data resulted in RMSEs of 2.9 and 0.9 m for mean and dominant heights, respectively. Estimation using training data resulted in RMSEs of 1.4 and 0.8 m, respectively. In both cases, the estimates of dominant height were more accurate than with the reference method, but the estimates of mean height were less accurate (area-based approach; RMSEs 1.1 and 0.9 m, respectively). The model-based method was robust to substantial decrease in echo density from 1.4 echoes/m² to 0.14 echoes/m².

Index Terms—Airborne laser scanning (ALS), eucalyptus, forest inventory, recovery, height, stochastic geometry.

I. INTRODUCTION

AIRBORNE laser scanning (ALS) provides three-dimensional (3-D) information on forest canopy, and the use of it in forest inventory has increased rapidly (e.g., [1], [2]). There are two main approaches to recover stand attributes. In a single-tree detection [3], individual trees are detected from the point cloud to produce tree-specific information. Only trees visible from above are detected and trees lying under the dominant tree layers often remain undetected [4]. This approach requires high echo density but does not necessarily need ground measured training data for tree height recovery. In the area-based approach [5], the data derived from laser scanning are used to estimate different forest attributes for plots and stands. It uses low echo density (usually about one echo per square meter). A regression approach, either parametric [5] or nonparametric [6], is used to generalize 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 data, such as

quantiles of the echo heights, or proportion of echoes below a given threshold height. Currently, ALS data are routinely used in forest management inventories of seminatural forests in boreal conditions (e.g., [7]), but studies on the use of it on other forest types, such as Eucalyptus plantations, have been conducted as well (e.g., [8]–[10]).

The benefit of ALS data over other remote sensing materials, such as aerial and satellite images, is that ALS data include direct measurements of the dimensions of forest canopy. Therefore, the distribution of ALS echo heights has a mathematical relationship with the tree heights. Motivated by this fact, the distribution of ALS echo heights is often used as a measure of the canopy structure of the stand (e.g., [11]–[15]). However, the dependence of the canopy structure on the stand structure, measured, e.g., with the distribution of tree heights or diameters, is complex due to several reasons. These include the effects of stand density, crown overlap, individual crown shape, and variation in the crown projection area (see [16] for a detailed description of these issues). Such situation is naturally approached by a stochastic geometric approach [17], by recognizing that ALS provides measurements of the union of tree crowns, whereas the primary interest in forest inventory is on the individual trees that form this union. Magnussen *et al.* [18] developed such models for tree height recovery that take into account the effect of individual crown shape and the crown projection area. More sophisticated and even better justified models were published by Sun and Ranson [19], and Ni-Meister *et al.* [20] to model the energy of ALS echoes for a given forest stand (see also [11], [21]). Mehtätalo and Nyblom [22], [23] treated the forest stand as an outcome of a stochastic geometric process, parameters of which were the characteristics of interest in forest inventory.

To summarize, even though ALS has proved to be very useful in forest inventory applications, the relationship between the laser measurements and tree attributes is not very well known. Therefore, selection of the predictors for the models of the area-based approach is based mostly on intuition and empirical findings, and not on well-established theory about the processes generating the laser echo. In this study, we take a step toward this direction by applying the model of Mehtätalo and Nyblom [23] for estimation. We call this approach model-based approach in contrast to the conventional (empirical) area-based method.

Our approach is based on the concept of *canopy envelope*. The canopy envelope of a sample plot is a solid, moderately smooth surface that covers all trees like a hood when seen from

Manuscript received April 10, 2014; revised August 12, 2014, September 19, 2014, and December 12, 2014; accepted March 23, 2015. Date of publication April 15, 2015; date of current version May 26, 2015. This work was supported in part by Suomen Luonnonvarain tutkimussäätiö and in part by the University of Eastern Finland.

The authors are with the Faculty of Science and Forestry, University of Eastern Finland, 80101 Joensuu, Finland (e-mail: lauri.mehtatalo@uef.fi).

Digital Object Identifier 10.1109/JSTARS.2015.2418675

above, and has the minimum volume below it [16]. We use a model that expresses the probability distribution of the canopy envelope height as a function of stand density, the parameters of the height distribution of trees, and the shape of individual tree crowns. The model has been developed for three different spatial patterns of tree locations: 1) random locations [22]; 2) square grid [23]; and 3) rectangular grid pattern of locations [24]. Applications to empirical datasets have been previously reported [23], [24]. Especially, Mehtätalo *et al.* presented a two-stage approach, where the first stage included model training using training sample plots and the second stage included estimation for evaluation plots [24]. The model also enables estimation of all its parameters without field-measured training dataset using low-density data. However, such formulation of the model has not been presented before nor evaluated with empirical data.

This study applies the rectangular grid model [24] to the situation where no field measured training data are available. In addition, two different procedures utilizing ground-measured training data are presented and evaluated. The robustness of the model-based approach to decreased echo density is also evaluated. Three small improvements to model formulation and estimation procedures are implemented. First, the difference between the top surface of the canopy and echo height is allowed through a fixed shift parameter in the model. Second, the individual tree shape is modeled using the Lamé curve [25] instead of the ellipse used in the previous publications. Third, a new approximation of the likelihood is presented for model fitting.

II. MATERIAL

The study material includes sample plots from the pulpwood plantation growing eucalyptus in Bahia state, Brazil. This study used a dataset of 37 sample plots, of which 18 were used for training and 19 for evaluation. The plots were measured in August–September 2008. The plots represented the whole range of mean canopy heights in the plantation. Plots were located in 13 different stands. Trees in the plantation are planted in rows, with fixed square grid spacing. On each circular sample plot of radius 13 m, all trees were callipered for diameter. Every seventh tree was measured for height so that height was measured typically from six or seven trees per plot. Näslund's height curve [26] was fitted by stands and used to predict heights for trees without height measurement. The RMSE of predicted height of the tree level was 1.14 m and bias -0.004 m, and the RMSE of mean height of the plot level was 0.54 m and bias -0.002 m. However, predicted heights from this model are hereafter regarded as the true heights, and the prediction error in them is ignored. A summary of the training and evaluation plots is shown in Table I.

The ALS data were collected on August 16, 2008 using an Optech ALTM 3100 laser scanning system. The flight altitude was approximately 1200 m above ground level and the field of view was 30° . The nominal sampling density of about 1.5 measurements/m², totalling 471–1209 echoes per each of the 530 m² sample plot; the large variation in echo density results in that some plots were covered by one and others by two flight lines. The footprint diameter was about 35 cm. A digital

TABLE I
BASIC INFORMATION ON THE TRAINING AND EVALUATION DATASETS

	Training	Evaluation
Mean height (m)	20.24–34.78	20.24–34.73
Dominant height (m)	20.90–40.38	20.77–38.97
Row spacing (m)	4.0 or 5.0	4.0 or 5.0
Column spacing (m)	2.4, 2.75, or 3.0	2.4, 2.75, or 3.0

elevation model (DEM) was generated from the ALS data. There was no understory complicating the DEM modeling. First, laser echoes were classified as ground and nonground echoes using the method reported by Axelsson [27]. Then, a raster DEM with a 1-m pixel size was interpolated using ground echoes. Finally, the raster DEM was subtracted from the ellipsoidal heights of laser echoes to scale the ALS data to the above-ground level (hereafter denoted by z). We used solely echo categories first of many and only. The observations below 10 m were taken as ground hits and treated as zeroes in the analysis. All trees were over 15 m tall, so a 10-m threshold was regarded as sufficient.

III. METHODS

A. Model for a Tree

For simplicity of model formulation, it is assumed that the forest stand is viewed directly from above, even though this assumption is not exactly met in practice. The crown envelope of an individual tree is handled in two pieces. In the upper part above the level of maximum radius, the radius is decreasing toward the tree top. In the lower part, the radius is the constant describing the maximum radius although it would in reality decrease toward the ground. The upper part of the envelope is modeled using the Lamé curve [25]

$$\left(\frac{z - qh}{h - qh}\right)^t + \left(\frac{r}{R}\right)^t = 1 \quad (1)$$

where z is the envelope height, h is the tree height, $q =]0, 1[$ the relative crown length, $h - qh$ the distance between tree top and the maximum crown radius, r the crown radius, R the maximum crown radius, and t a shape parameter of the curve (see Fig. 1). When t increases, the curve approaches the rectangular shape, and when t approaches 0, the curve approaches the shape of inverted letter T. Case $t = 1$ produces a conical shape and $t = 2$ an ellipsoidal shape.

Solving (1) for h , and allowing an additional constant shift parameter P between the envelope height and the height of echo yields the following expression for the height of a tree with crown envelope radius r at height z

$$h(z, r) = \frac{z}{\left(1 - \left(\frac{r}{R}\right)^t\right)^{1/t} (1 - q) + q} + P. \quad (2)$$

The constant shift is justified by the assumption that the pulse does not return at the top surface but penetrates into the crown. For simplicity, all pulses are assumed to penetrate a similar fixed amount, which is quantified in the parameter P . Thus, function h has four parameters: 1) the maximum crown radius R ; 2) the relative height of the maximum radius q ; 3) the

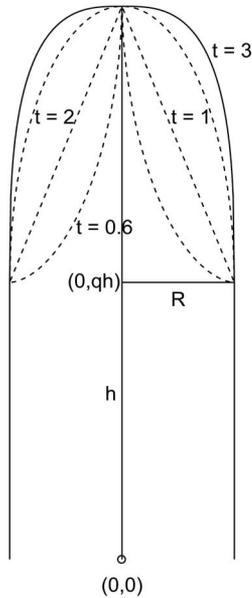


Fig. 1. Lamé curve for a tree with the total height of h and maximum crown radius R at height qh using four different values of parameter t .

shape parameter t ; and 4) the shift P . These parameters of the tree crown are denoted by $\theta = (R, t, q, P)$, and assumed to be common for all trees of a sample plot.

B. Model for Canopy Height of a Stand

Let F be the distribution function of tree heights. It can be described, e.g., by the cumulative form of the Weibull function as

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

where $\xi = (\alpha, \beta)$ are Weibull parameters. Other functions than Weibull could be used as well, but this function was chosen because of its flexibility and established use in modeling tree size [28], [29].

Let $Z(v)$ be the height of the stand-specific canopy envelope at a fixed point v , i.e., the vertical distance between the ground level and the canopy envelope. The canopy envelope height at a given point can be thought of as the maximum over the heights of the tree-specific envelopes at that point. Therefore, the probability of having $Z(v)$ below the height z above ground is the probability that none of the neighboring tree crowns at height z extend to point v [23]. Under independence of tree heights, it can be written as

$$P(Z(v) \leq z) = \prod_{i=1}^N Fh(z, \|v - u_i\||\theta)|\xi\} \quad (3)$$

where N is the number of neighboring trees at point v , θ includes the crown shape parameters, ξ includes the parameters of height distribution, and $\|v - u_i\|$ is the distance between the fixed point v and the location u_i of tree i .

Equation (3) assumes that the horizontal cross section of a tree crown is circular. The N neighboring trees include those trees that are located so close to the point v that their crowns

have a nonzero probability to reach to the point v . For a randomly selected point within set A , the cumulative distribution function of Z is the mean of (3) over the set. In a rectangular grid pattern of tree locations, however, this mean is equal for each grid cell (we ignore the possible effects of stand edge). Therefore, for a grid spacing with distance l between rows and distance m between trees of a row, the cumulative distribution function of echo height is

$$G(z|\theta, \xi) = \frac{1}{lm} \int_0^l \int_0^m \prod_{i=1}^N F\{h(z, \|u_i - v\|)|\theta)|\xi\} dv \quad (4)$$

$$z \geq 0.$$

The corresponding probability density function is defined in two parts. When $z > 0$, the probability density function is the differential of the cumulative distribution function

$$g(z|\theta, \xi) = \frac{1}{lm} \int_0^l \int_0^m \sum_{i=1}^N \left[f\{h(z, \|u_i - v\|)|\theta)|\xi\} \cdot \frac{d}{dz} h(z, \|u_i - v\|) \cdot \prod_{j=1, j \neq i}^N F\{h(z, \|u_j - v\|)|\theta)|\xi\} \right] dv \quad (5)$$

where $f = F'(x)$ is the probability density function of tree heights. When $z = 0$, the probability mass is

$$g(z|\theta, \xi) = G(z|\theta, \xi).$$

C. Estimation

Estimation of parameters θ and ξ is done with the method of maximum likelihood. Assuming that the laser observations are an independent and identically distributed sample from the distribution (4) (see [23] for a discussion on these assumptions), the log likelihood is the sum of logarithmic densities at the observed values of z

$$\ell(\theta, \xi) = \sum_{j=1}^M \ln g(z_j|\theta, \xi) + M_0 \ln G(0|\theta, \xi) \quad (6)$$

where M is the number of echoes from crowns and M_0 is the number of ground echoes.

Computing the likelihood function using (6) is computationally intensive for large datasets because of the numerical area integral of (5). To decrease the computing time, we classified the observed nonzero values of z into a small number of classes. We used 15 classes of variable width (average 1–2 m), so that each class had an equal number of observations. For each class i , the mean of z values was calculated as z_i . The zero echo heights formed an additional, 16 class. The resulting approximate likelihood was

$$\ell(\theta, \xi) \approx \sum_{i=1}^{16} n_i \ln g(z_i|\theta, \xi)$$

where n_i is the number of observations in class i .

The main parameters of interest are the Weibull parameters ξ , which specify the plot-specific distribution of tree heights. In addition, the model includes the crown-shaped parameters θ . Both parameter vectors can be simultaneously estimated using ALS data, but known values of either ξ or θ can be used as well. This enables different estimation procedures, which are detailed in the following sections. Especially, we present a one-stage procedure, where all parameters are estimated from ALS data without using training data at all, and two versions of a two-stage procedure, where information on measured tree heights of the training plots is utilized.

1) *Estimation Without Training Data*: In this procedure called *nofield*, the crown shape parameters and Weibull parameters of the evaluation plot in question are all estimated by maximizing the likelihood function (6) jointly with respect to θ and ξ . However, the shift parameter P is assumed to be known in advance for a technical reason: the solution without this assumption is not unique. We set the value of P to 2.1 m, according to the experience gained from the two-stage approach described below; and recognizing that a similar value was previously reported for broadleaved tree data [30].

2) *Estimation By Utilizing Training Data*: The procedures using training data include two stages. First, the crown shape parameters θ are estimated using the training dataset (Stage A). Second, the Weibull parameters of the height distribution ξ are estimated for the evaluation data using the estimated crown parameters from stage A and laser echo heights from the plot of question (Stage B). Two different two-stage procedures were implemented: 1) a *basic*; and 2) a *regress* procedure.

The two stages of the *basic* procedure are as follows.

- A) The Weibull distribution is fitted to the true tree heights of each of the 18 training plots using the method of maximum likelihood to get the estimates of Weibull parameters $\hat{\xi}$ for each plot. These estimates are used in the likelihood function (6), and the crown shape parameters $\hat{\theta}$ are estimated for the training plots. The correlation of the resulting plot-specific estimates $\hat{\theta}$ with the plot-specific means of the canopy echo heights is explored graphically. If needed, a regression line is fitted to explain the observed trend; otherwise, the mean of plot-specific estimates is computed.
- B) The crown shape parameters $\tilde{\theta}$ are predicted for the 19 evaluation plots using the regressions or means from stage A, and used in the likelihood (6), which is now maximized with respect to the Weibull parameters ξ . This results in an estimated height distribution for each evaluation plot.

In the *regress* procedure, the estimated height distribution is forced to have the mean height given by a predictive model based on the training data. The procedure is similar to the *basic* procedure, except for the following changes.

- a) In addition to estimation of θ , mean height \bar{h}_k for each training plot k is computed using the known tree heights and a regression of plot-specific mean height is fitted. The model for plot k is [31]

$$\bar{h}_k = \frac{1}{p_0 + p_1 z_{90,k} + p_2 z_{95,k}^2 + p_3 \sqrt{z_k}} + e_k \quad (7)$$

where p_0, \dots, p_3 are coefficients, $z_{90,k}$ and $z_{95,k}$ are the percentiles of ALS echo heights, \bar{z}_k is the mean of the laser echo heights, and e_k is the residual for plot k . The coefficients p_0, \dots, p_3 are estimated using nonlinear least squares.

- b) In addition to the crown shape parameters, also the mean height is predicted for each evaluation plot using the model of stage A. This enables profiling parameter β from the Weibull distribution by using equation $\beta = \bar{h}/\Gamma(1 + \frac{1}{\alpha})$, where \bar{h} is the predicted mean height; the equation is based on the expected value of the Weibull distribution. The resulting profile likelihood (6), is maximized with the only remaining parameter, i.e., the shape parameter α of the Weibull distribution.

3) *Area-Based Method*: The empirical area-based method is also implemented for the dataset. In that implementation, the model of the form of (7) is fitted for both the mean and dominant heights (mean height of 100 tallest trees/ha) in the training dataset of this study. These models are used for prediction in the evaluation data, and served as a reference method in evaluation of the model-based approaches.

D. Evaluation

The fit of estimated Weibull distribution was evaluated using Kolmogorov–Smirnov (KS) test on each plot. In addition, the mean and dominant heights of the evaluation plots were computed using the estimated distribution of tree heights as

$$\begin{aligned} \text{HMEAN}_k &= \int_0^\infty u f(u|\xi) du \\ \text{HDOM}_k &= \frac{N}{100} \int_q^\infty u f(u|\xi) du \end{aligned}$$

where $f(u|\xi)$ is the density of the estimated height distribution, q is the $(N - 100)/N$ th quantile of it, and N is the stand density (trees/ha). The accuracy of estimates compared to the true data was evaluated in terms of root mean square error (RMSE), bias, and correlation coefficient

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{\sum_{k=1}^n (h_k - \hat{h}_k)^2}{n}} \\ \text{bias} &= \frac{\sum_{k=1}^n (\hat{h}_k - h_k)}{n} \\ r^2 &= \text{cor}(h_k, \hat{h}_k)^2 \end{aligned}$$

where n is the number of plots, h_k is the true mean or dominant height of the plot k , \hat{h}_k is the estimated mean or dominant height of the plot.

We also analyzed the influence of decreased echo density on the results. For this purpose, the echo density was decreased from 1.4/m² to 0.35/m² and 0.14/m² and the estimation with method *basic* was conducted using these low density datasets. The thinning was conducted by taking systematically every 4th and every 10th echo from the original data, where the echoes were ordered according to time stamp of the echo.

TABLE II

SUMMARY OF THE ESTIMATES OF THE CROWN SHAPE PARAMETERS IN THE 19 EVALUATION PLOTS FOR THE ONE-STAGE PROCEDURE AND ON THE 18 TRAINING PLOTS IN THE TWO-STAGE PROCEDURES

	Range	Mean	s.d.
<i>One-stage procedure</i>			
R	1.87–2.58	2.22	0.23
q	0.001–0.99	0.46	0.48
t	0.42–15.00	5.78	3.83
<i>Two-stage procedures</i>			
R	2.00–2.67	2.27	0.19
q	0.001–0.99	0.42	0.45
P	0.23–3.86	2.09	1.04
t	0.05–15.00	$-1.07 + 0.25\bar{z}$	4.14

\bar{z} is the mean height of the canopy hits. In the one-stage procedure, the penetration parameter was set at $P = 2.1$ m.

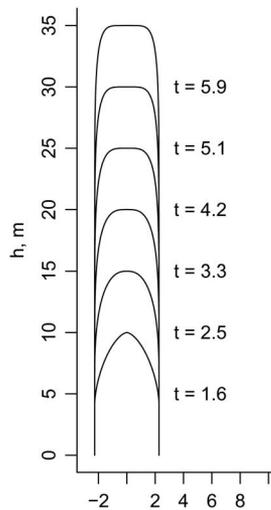


Fig. 2. Shape of the crown as tree height increases.

IV. RESULTS

The results from the estimation of the crown shape parameters are shown in Table II. The maximum radius R is similar for both one- and two-stage methods and does not vary much. The mean relative height of the maximum radius q is also similar for both methods, but varies much between sample plots. The range of the parameter t is also wide. With the one-stage procedure, the estimates of t are over two in all but two plots, indicating a shape that is more convex than an ellipsoid. In the two-stage approach, a correlation between t and average echo height was found. The fitted regression indicates increasing convexity as tree height increases (Fig. 2).

The nonlinear regression models of the reference method (Table III) had the RMSE of 1.097 and 0.896 m for the mean and dominant height in the evaluation data, respectively. The estimation using no field data (method *nofield*, Table IV) provided lower RMSE of dominant height than the reference method (0.86 m) but much higher RMSE of mean height (about 2.9 m). From among the two two-stage procedures, which have similar data needs than the reference method, the *basic* method provided a lower RMSE of dominant height than the methods *reference* and *nofield* (0.8 m), but a higher RMSE of mean

TABLE III

COEFFICIENTS AND THEIR STANDARD ERRORS FOR THE NONLINEAR REGRESSION MODEL OF MEAN HEIGHT AND DOMINANT HEIGHT

	p_0	p_1	p_2	p_3
Mean height	9.47e-2	-3.48e-3	4.08e-5	1.08e-3
SE	8.37e-3	4.36e-4	7.55e-6	1.55e-3
Dominant height	9.11e-2	-2.80e-3	3.11e-5	-1.08e-3
SE	1.13e-2	6.05e-4	1.02e-5	2.07e-3

The models provide the predictions of the reference method; however, the model for mean height was used also in method *regress*.

height (1.4 m). Forcing the mean of the height distribution to be equal to the prediction of the reference method (method *regress*) decreased the RMSE of mean height to the level of the reference method but increased the RMSE of dominant height to 0.99 m. The biases of mean and dominant height were not significantly different from zero (Table IV) for any of the methods due to low number of sample plots.

To enable comparison of the estimated height distributions with the true ones, the “true” Weibull parameters were estimated for each plot by fitting the Weibull distribution to the tree height data of each plot using maximum likelihood. KS tests on these fits did not indicate any lack of fit. Fig. 3 plots the ALS-based estimates of the Weibull parameters using methods *nofield*, *basic*, and *regress* on these true parameter values of the plot. The estimated scale parameter (β) is strongly correlated with true scale parameter with both methods, but the correlation between estimated and true shape parameter (α) is weak. The estimation of the shape parameter was especially inaccurate with the method *nofield*.

For a further insight into the results, the Weibull distributions based on methods *nofield*, *basic*, and *regress* were compared to the empirical height distribution of the plot using KS-test. Of the 19 evaluation plots, Weibull distribution fitted well with method *nofield* in two plots, with method *basic* in four plots, and with *regress* in seven plots. However, despite this lack of fit, the estimated height distribution was still many times able to provide good estimates of mean and dominant height. For example, in the plot 19 with *regress* method, the KS test indicated severe lack of fit (p -value is 0.0008), but a graph of the fitted density (Fig. 4, right, middle) does not show a big difference between the shape of estimated and true Weibull distribution. Fig. 4 shows the empirical height histograms and the fitted height distributions for three different plots using all three methods. The evaluation plots, marked in Fig. 3, were chosen to illustrate an example of a badly underestimated, well estimated, and badly overestimated shape parameter α . These fits are clearly illustrated with Fig. 4.

To illustrate the distribution of ALS echoes and the fitted model, Fig. 5 shows the fitted density function $g(z|\theta, \xi)$ for sample plot 3 (same plot as in the left panel of Fig. 4). The estimated density function shows good fit to the observed laser data.

The results with moderate and heavy thinned ALS data showed only small differences compared to the results with complete data (Table V). The RMSE of mean height even slightly decreased, whereas that of dominant height increased.

TABLE IV
VALUES OF RMSE, BIAS, AND COEFFICIENT OF DETERMINATION FOR THE ESTIMATED MEAN HEIGHT AND DOMINANT HEIGHT

	Mean height				Dominant height			
	RMSE	Bias	p -value	r^2	RMSE	Bias	p -value	r^2
Nofield	2.886	0.030	0.99	0.959	0.864	-0.232	0.15	0.994
Basic	1.417	-0.349	0.22	0.982	0.792	0.218	0.52	0.992
Regress	1.097	-0.004	0.92	0.967	0.990	0.497	0.054	0.987
Reference	1.097	-0.004	0.92	0.967	0.896	0.367	0.38	0.991

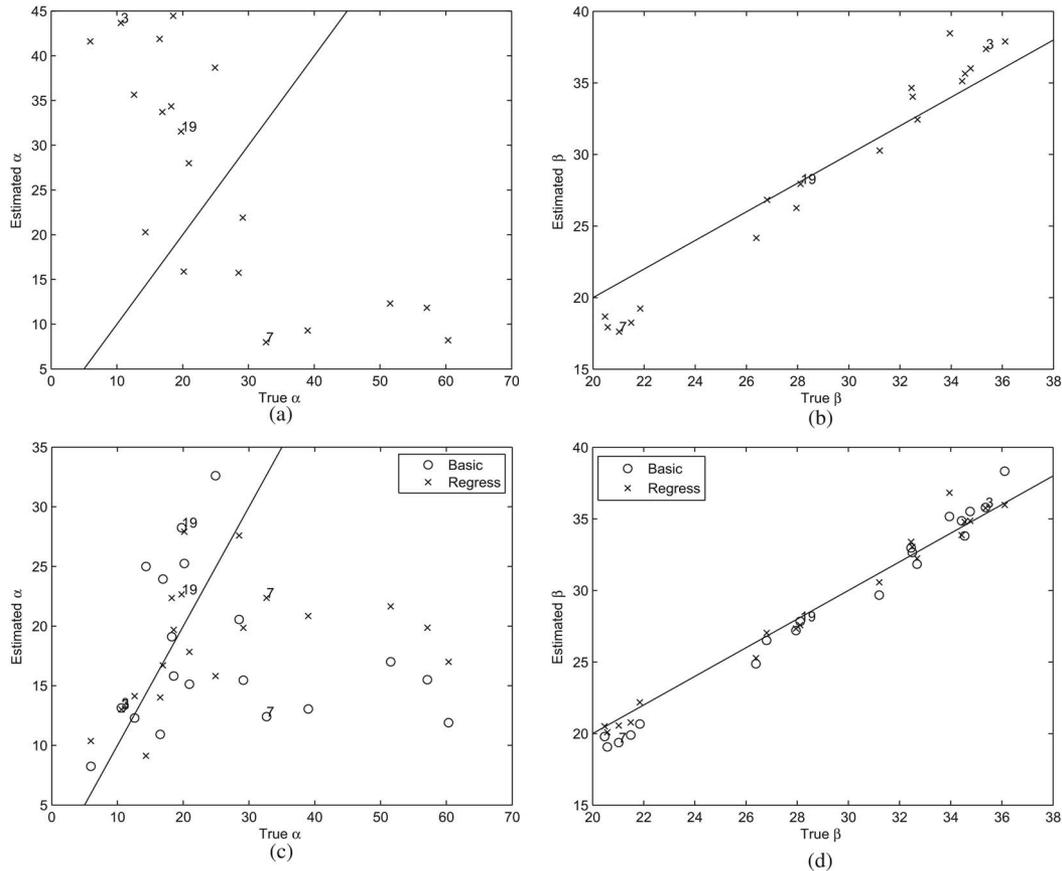


Fig. 3. Estimated Weibull parameters of height distribution with method *nofield* (top) and with the two-stage approaches (bottom). The example plots of Fig. 4 are labeled using the plot identification.

V. DISCUSSION

The study applied the model of Mehtätalo and Nyblom [23] to the estimation of forest attributes in two situations: 1) in a situation where no field calibration data are used; and 2) in a situation where field measured data are used for model training. Especially, we presented and evaluated a new area-based method that did not utilize the ground measurements of the training plots. The model for rectangular grid pattern of tree locations was used [24]. However, some improvements in the model and the estimation procedures were implemented, including a constant shift parameter to take into account the penetration of the pulse into tree crowns, replacing the ellipsoidal crown shape with a shape based on Lamé function, and introducing a two-stage approach that is able to utilize a fitted model for mean height in estimation. For model fitting, an approximate of the likelihood was presented to make the computations faster. In addition, the robustness of the model-based

approach to substantial decrease in the echo density was evaluated.

In the procedure that did not use training data, all parameters of the model were estimated in one stage by using the ALS echo heights. However, the value for the shift parameter was taken as given, because estimation of it using ALS data without measured tree height information was not possible. The approach has similar data needs than individual tree detection approaches. However, it can utilize low echo density data, too, as demonstrated in the evaluation where the echo density was reduced to 0.14 echoes/m².

The second applied approach was a two-stage approach, where the first stage included estimation of the four tree crown-related parameters using training plot data, and the second stage includes the estimation of the distribution of tree heights for other plots using the parameter estimates from the first stage and ALS data from the plots of interest. With respect to data needs, this approach is comparable to the area-based

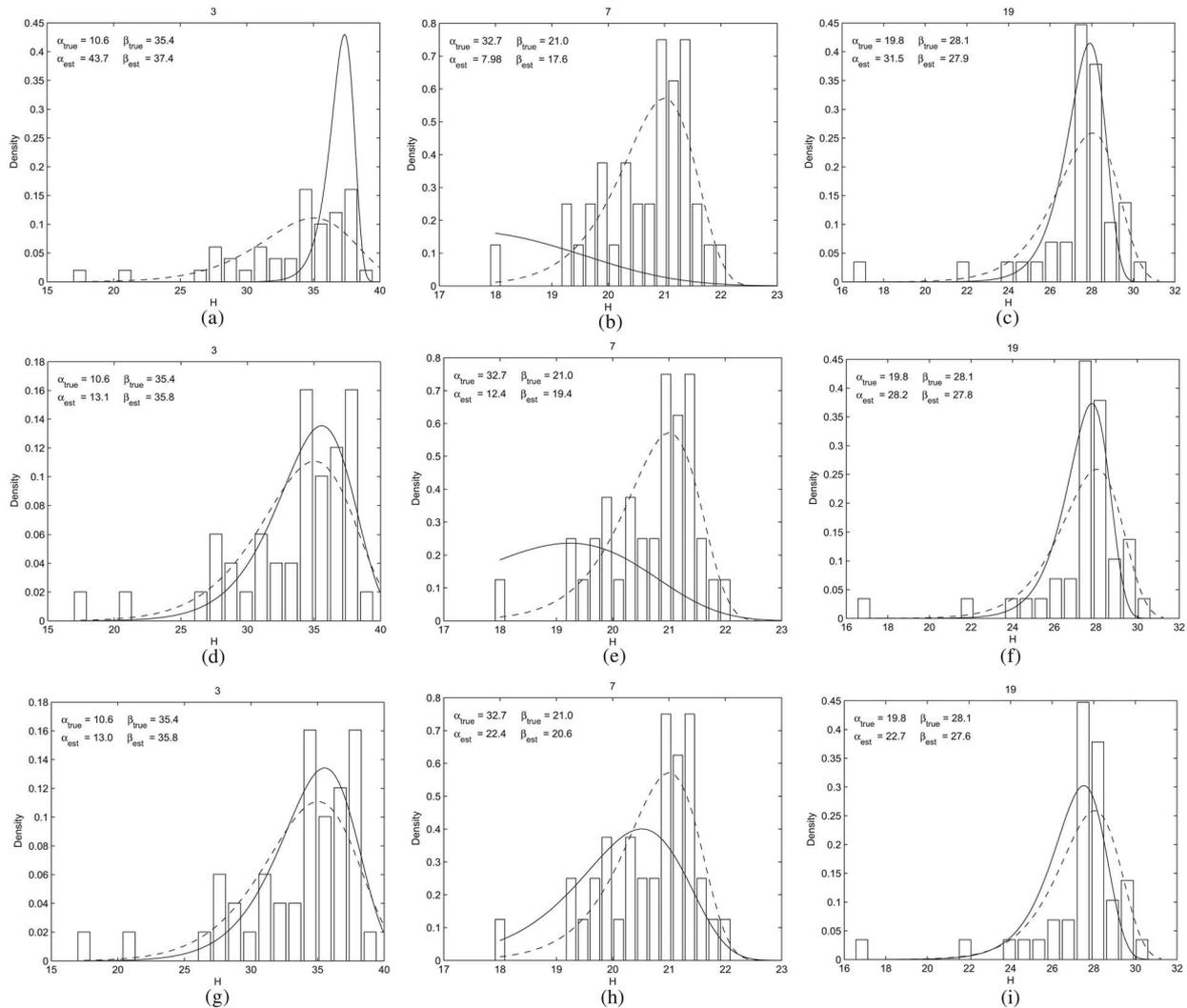


Fig. 4. True tree heights (histogram), the true Weibull density fitted to the tree heights (dashed line) and the estimated Weibull height distribution (solid line) on three example sample plots. The estimates are based on procedures *nofield* (top) *basic* (middle), *regress* (bottom) for sample plots 3 (left), 7 (middle), and 19 (right).

approach, even though the detailed field measurements needed from the training plots differ. The practical applications of these two approaches were demonstrated with empirical data from eucalyptus plantations with known spacing.

A specific property of our plantation dataset was that the rectangular spatial pattern, and therefore also the stand density were known in advance. This information was utilized implicitly in the model, and the main interest focused on the estimation of the plot-specific distribution of tree heights. However, knowledge of the density or spacing of trees is not necessary, and the methods presented in this study could be tailored for plantations with unknown density as well. In that case, the stand density would just be an additional parameter of interest in the model.

The fit of the estimated height distribution was evaluated using the RMSE and bias of mean and dominant heights. Both the one-stage and two-stage procedures provided more accurate estimates of dominant height than the empirical area-based method, which was used as a reference method. However, the estimates of mean height were less accurate. As an attempt to

increase the accuracy of mean height estimation, we also proposed the *regress* procedure, which ensured as good accuracy of mean height as with the reference method. Unfortunately, while significantly improving the accuracy of mean height, this restriction reduced the accuracy of dominant height.

We also evaluated the effect of echo density in the accuracy of prediction. Decreased echo density had only minor effect on the accuracy of results. Therefore, both procedures are robust to significant decrease in the echo density. This is an interesting option especially with the *nofield* approach: no other methods have been proposed that could be used for estimation using low-density data without field measurement data.

The true heights of the data were based on predictions of local regressions from a subsample of trees. The related prediction error was ignored in accuracy assessment. In addition, tree heights were measured only to the nearest 0.5 m. The RMSE of height prediction was 0.54 m which is almost half of the observed RMSE of mean heights of 1.1–1.4 m. Therefore, the reported RMSEs of the mean height and dominant height may be overestimates. In addition, the predicted tree heights

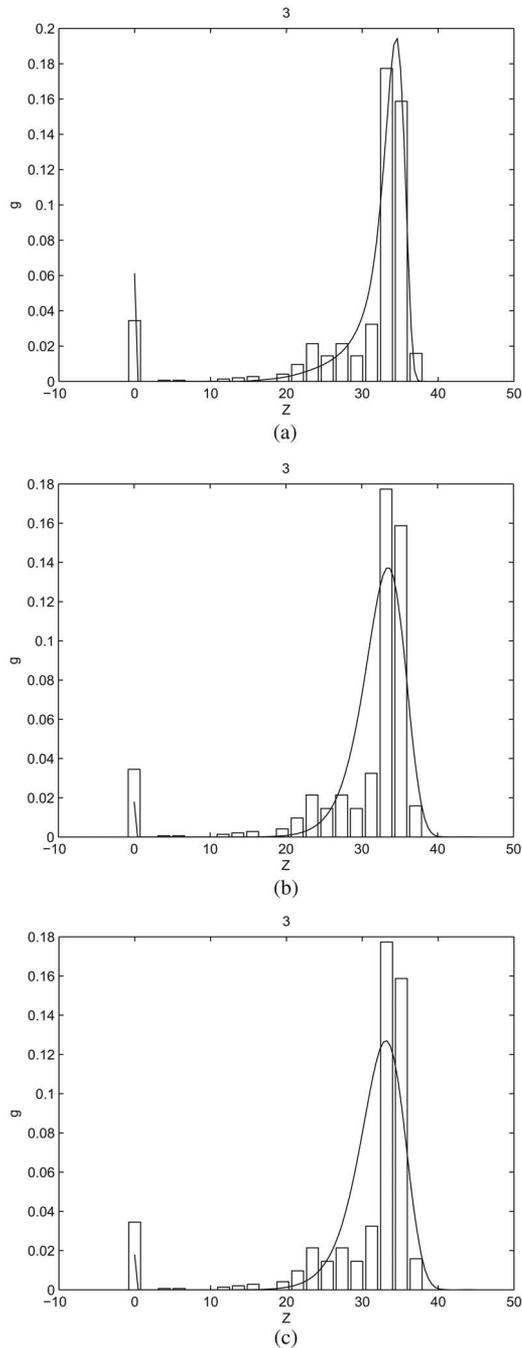


Fig. 5. Graph of distribution of echo heights Z (histogram) and fitted density $g(z|\theta, \xi)$ (line) in the one-stage procedure *nofield* (top) and in the two-stage procedure after step B5. The middle graph shows the result from method *basic* and the bottom graph from method *regress*.

have lower variation than true tree heights have; therefore, the height distributions regarded as true are too narrow. To conclude, ALS-based prediction of tree heights with our dataset is so accurate with all these methods, that quite large proportion of the observed errors may actually be due to the inaccuracy of evaluation data.

There are also some previous research on the use of ALS in Eucalyptus plantations [8]–[10]. Especially, Goncalves-Seco *et al.* [32] predicted forest attributes in high-density eucalyptus plantations using ALS data. They used linear and multiplicative

regression models to estimate attributes such as mean height, dominant height, mean diameter, and volume. They had the average echo density of about four echoes per square meter. The RMSE of prediction error was 1.33 m for mean heights and 1.18 m for dominant height, which are higher than with our *reference* method and comparable to the results of our *basic* method.

With respect to the data needs, the method *nofield* is comparable to the individual tree detection approach, where a row detection algorithm was used with the data collected from the same eucalyptus plantations [33]. Compared to their results, the method *nofield* of this study gave more accurate predictions for dominant height (RMSEs were 0.86 and 1.03 m, respectively), but less accurate results for mean height (RMSEs of 2.86 and 1.49 m, respectively). However, as already noted, the model-based method can be used also with low-density data, whereas the ITD cannot.

The penetration parameter was estimated to be about 2 m. The result is realistic if it is compared to Gaveau and Hill [30], where the laser pulse penetration was 2.12 m for a tree canopy of broadleaved trees dominated by oak and ash. The reasons for penetration can be divided to two separate causes: 1) the laser pulse may hit an opening between the branches of a tree crown and 2) the sensor needs some accumulated energy before recording the reflected echo. The constant penetration for a certain plot and inventory campaign might be suitable model for the second cause, whereas the first cause is clearly a random process even within a sample plot. Based on the second cause above, we hypothesized that a fixed value within a plot would provide a satisfactory model for the penetration. However, the problems noticed in this study indicate that a random component is needed to the model of penetration. A natural starting point for the random penetration is provided by the exponential distribution, which is based on the widely applied Beer–Lambert’s law [34]. Assuming randomness in the penetration would make it technically possible to estimate also the parameters related to it in the method *nofield*.

The model-based approach provided accurate estimates of the dominant height, which is computed using the right tail of the height distribution. However, the prediction of mean height was less accurate, and efforts to increase the accuracy always led to decrease in the accuracy of dominant height. A similar tradeoff between the RMSE of mean and dominant height occurred also in the evaluation with decreased echo density. This may indicate that the shape of the predicted height distribution was inconsistent with the observed heights of the plot.

One possible reason for this inconsistency may be the above-mentioned problems related to the estimation of true height. Another reason, which we regard as the most important, is that the model may not include all necessary components, or that they are not parameterized flexibly enough. Alternatives for improvements are, e.g., allowing randomness to the crown shape and penetration. We have already outlined the development of the model to this direction [16]. However, such formulation introduces a bunch of computational problems into the estimation, and the solution might be possible only by using approximate bayesian methods [35]. A third possible problem

TABLE V
EFFECT OF DECREASED ECHO DENSITY ON THE ACCURACY OF HEIGHT PREDICTION WITH METHODS *BASIC* AND *NOFIELD*

Echo density $1/m^2$	RMSE	Mean height			RMSE	Dominant height			
		Bias	p value	r^2		Bias	p value	r^2	
<i>Method nofield</i>									
1.4	2.886	0.030	0.99	0.959	0.864	-0.232	0.15	0.994	
0.35	2.767	0.081	0.90	0.954	0.957	-0.304	0.17	0.987	
0.14	2.817	-0.165	0.81	0.955	1.004	-0.269	0.25	0.989	
<i>Method basic</i>									
1.4	1.417	-0.349	0.22	0.982	0.792	0.218	0.52	0.992	
0.35	1.391	-0.296	0.37	0.979	0.840	0.089	0.66	0.987	
0.14	1.374	-0.245	0.45	0.978	1.011	0.421	0.07	0.985	

is that the estimation in current form is just too instable. A Bayesian estimation with highly informative priors has also been used as the solution to a similar problem [36]; however such informative prior information was not available on the Eucalyptus trees of this study. A fourth possible reason may be the assumption that trees are seen directly from above, which was quite badly violated for some sample plots in our dataset where the field of view was 30° . Taking into account the scanning angle in the approach would require a new definition for the whole concept of crown envelope, where the cross section of observable tree canopy is assumed to increase monotonically toward the ground.

The dataset of this study was an even-aged eucalyptus forest plantation. Such data were expected to be an easy dataset because of the known spacing and stand density. However, the results of this study were a disappointment in this respect. In most practical situations, the spatial pattern varies among plots and stand density is unknown. These issues provide additional difficulties in applications. On the other hand, the crowns of eucalyptus may actually be more variable and porous than, e.g., slowly growing ever-green conifer crowns, for which very good results were reported for one old-growth Norway spruce sample plot [23]. In addition, another approach based on stochastic geometry showed promise in a recent study [37] for Scots pine stands with canopy cover below 0.9. That approach uses only two-dimensional (2-D) data of crown segments and has, therefore, no assumptions on the vertical profile of tree crowns.

To summarize, this paper described the main principles on how an area-based approach for estimation of forest attributes could be formulated, starting from a stochastic geometric 3-D model that treats forest canopy as the union of tree crowns. The approach allowed estimation of tree height distribution using low-density ALS data even in the absence of field measurements. Even though the approach is intuitive and appealing, the empirical results were unsatisfactory. Therefore, and as outlined in the text above, further development of the model and estimation methods is needed before the method can be used in practice.

REFERENCES

- [1] C. Babcock, J. Matney, A. Finley, A. Weiskittel, and B. Cook, "Multivariate spatial regression models for predicting individual tree structure variables using LiDAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 1, pp. 6–14, Feb. 2013.
- [2] Z. Fang and C. Cao, "Estimation of forest canopy height over mountainous areas using satellite LiDAR," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 7, pp. 3157–3166, Jul. 2014.
- [3] J. Hyypää and M. Inkinen, "Detecting and estimating attributes for single trees using laser scanner," *Photogramm. J. Finland*, vol. 16, pp. 27–42, 1999.
- [4] A. Persson, J. Holmgren, and U. Söderman, "Detecting and measuring individual trees using airborne laser scanner," *Photogramm. Eng. Remote Sens.*, vol. 68, pp. 925–932, 2002.
- [5] E. Næsset, "Estimating timber volume of forest stands using airborne laser scanner data," *Remote Sens. Environ.*, vol. 61, pp. 246–253, 1997.
- [6] P. Packalén and M. Maltamo, "The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs," *Remote Sens. Environ.*, vol. 109, pp. 328–341, 2007.
- [7] E. Næsset *et al.*, "Laser scanning of forest resources: The nordic experience," *Scand. J. For. Res.*, vol. 19, pp. 482–499, 2004.
- [8] R. Wack, M. Schardt, U. Lohr, L. Barrocho, and T. Oliveira, "Forest inventory for eucalyptus plantations based on airborne laser scanner data," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 34, pp. 40–46, 2003.
- [9] P. K. Tickle, A. Lee, R. M. Lucas, J. Austin, and C. Witte, "Quantifying australian forest and woodland structure and biomass using large scale photography and small footprint LiDAR," *For. Ecol. Manage.*, vol. 223, pp. 379–394, 2006.
- [10] S. Tesfamichael, J. van Aardt, and F. Ahmed, "Estimating plot-level tree height and volume of Eucalyptus grandis plantations using small-footprint, discrete return LiDAR data," *Prog. Phys. Geogr.*, vol. 34, no. 4, pp. 515–540, 2010 [Online]. Available: <http://ppg.sagepub.com/content/34/4/515.abstract>
- [11] N. Coops *et al.*, "Estimating canopy structure of douglas-fir forest stands from discrete-return LiDAR," *Trees*, vol. 21, pp. 295–310, 2007.
- [12] T. J. Dean, Q. V. Cao, S. D. Roberts, and D. L. Evans, "Measuring heights to crown base and crown median with LiDAR in a mature, even-aged loblolly pine stand," *For. Ecol. Manage.*, vol. 257, pp. 126–133, 2009.
- [13] D. Jaskierniak, P. Lane, A. Robinson, and A. Lucieer, "Extracting LiDAR indices to characterise multilayered forest structure using mixture distribution functions," *Remote Sens. Environ.*, vol. 115, pp. 563–585, 2011.
- [14] A. Ferraz *et al.*, "3-D mapping of a multi-layered mediterranean forest using ALS data," *Remote Sens. Environ.*, vol. 121, pp. 210–223, 2012.
- [15] M. Melin *et al.*, "Moose (Alces alces) reacts to high summer temperatures by utilizing thermal shelters in boreal forests: An analysis based on airborne laser scanning of the canopy structure at moose locations," *Global Change Biol.*, vol. 20, no. 4, pp. 1115–1125, 2014 [Online]. Available: <http://dx.doi.org/10.1111/gcb.12405>
- [16] L. Mehtätalo, J. Nyblom, and A. Virolainen, "A model-based approach for the recovery of forest attributes using ALS data," in *Forestry Applications of Airborne Laser Scanning—Concepts and Case Studies*, M. Maltamo, E. Næsset, and J. Vauhkonen, Eds. New York, NY, USA: Springer, 2014, pp. 193–211.
- [17] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications*, 3rd ed. Hoboken, NJ, USA: Wiley, 2013.
- [18] S. Magnussen, P. Eggermont, and V. N. LaRiccía, "Recovering tree heights from airborne laser scanner data," *For. Sci.*, vol. 45, pp. 407–422, 1999.
- [19] G. Sun and K. J. Ranson, "Modeling LiDAR returns from forest canopies," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 6, pp. 2617–2626, Nov. 2000.
- [20] W. Ni-Meister, D. L. B. Jupp, and R. Dubayah, "Modeling LiDAR waveforms in heterogeneous and discrete canopies," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 9, pp. 1943–1958, Sep. 2001.

- [21] J. L. Lovell, D. L. B. Jupp, D. S. Culvenor, and N. C. Coops, "Using airborne and ground-based ranging LiDAR to measure canopy structure in Australian forests," *Can. J. Remote Sens.*, vol. 29, no. 5, pp. 607–622, 2003.
- [22] L. Mehtätalo and J. Nyblom, "Estimating forest attributes using observations of canopy height: A model-based approach," *For. Sci.*, vol. 55, no. 5, pp. 411–422, 2009.
- [23] L. Mehtätalo and J. Nyblom, "A model-based approach for ALS inventory: Application for square grid spatial pattern," *For. Sci.*, vol. 58, pp. 106–118, 2012.
- [24] L. Mehtätalo, A. Virolainen, J. Tuomela, and J. Nyblom, "A model-based approach for estimating the height distribution of eucalyptus plantations using low-density ALS data," in *Proc. SilviLaser*, 2010, pp. 295–324.
- [25] M. Rautiainen, M. Möttöus, P. Stenberg, and S. Ervasti, "Crown envelope shape measurements and models," *Silva Fennica*, vol. 42, no. 1, pp. 19–33, 2008.
- [26] M. Näslund, "Skogsförsöksanstaltens gallringsförsök i tallskog (Forest research institute's thinning experiments in Scots pine forests)," *Meddelanden från statens skogsförsöksanstalt Häfte 29*, 1937, (in Swedish).
- [27] P. Axelsson, "DEM generation from laser scanner data using adaptive TIN models," *Int. Arch. Photogramm. Remote Sens.*, vol. 54, pp. 138–147, 2000.
- [28] R. Bailey and T. Dell, "Quantifying diameter distributions with the Weibull function," *For. Sci.*, vol. 19, pp. 97–104, 1973.
- [29] J. Siipilehto and L. Mehtätalo, "Parameter recovery vs. parameter prediction for the Weibull distribution validated for Scots pine stands in Finland," *Silva Fennica*, vol. 47, no. 4, pp. 1–22, 2013 [Online]. Available: <http://www.silvafennica.fi/article/1057>
- [30] D. L. Gaveau and R. A. Hill, "Quantifying canopy height underestimation by laser pulse penetration in small-footprint airborne laser scanning data," *Can. J. Remote Sens.*, vol. 29, no. 5, pp. 650–657, 2003.
- [31] P. Packalén, L. Mehtätalo, and M. Maltamo, "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," *Ann. For. Sci.*, vol. 68, pp. 1085–1092, 2011.
- [32] L. Goncalves-Seco *et al.*, "Assessing the attributes of high-density eucalyptus globulus stands using airborne laser scanner data," *Int. J. Remote Sens.*, vol. 32, no. 24, pp. 9821–9841, 2011.
- [33] J. Vauhkonen, L. Mehtätalo, and P. Packalén, "Combining tree height samples produced by airborne laser scanning and stand management records to estimate plot volume in eucalyptus plantations," *Can. J. For. Res.*, vol. 41, no. 8, pp. 1649–1658, 2011.
- [34] S. Solberg *et al.*, "Mapping LAI in a Norway spruce forest using airborne laser scanning," *Remote Sens. Environ.*, vol. 113, pp. 2317–2327, 2009.
- [35] J.-M. Marin, P. Pudlo, C. Robert, and R. Ryder, "Approximate Bayesian computational methods," *Stat. Comput.*, vol. 22, pp. 1167–1180, 2012.
- [36] T. Lähivaara *et al.*, "A model for predicting the growth of eucalyptus globulus seedling stands in bolivia," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2690–2699, 2014.
- [37] J. Vauhkonen and L. Mehtätalo, "Matching remotely sensed and field measured tree size distributions," *Can. J. For. Res.*, vol. 45, no. 3, pp. 353–363, 2015.



Lauri Mehtätalo received the M.Sc. and D.Sc. degrees in forest sciences from the University of Joensuu, Joensuu, Finland, in 2000 and 2004, respectively.

Currently, he is an Associate Professor of Applied Statistics with the University of Eastern Finland, Joensuu, Finland. Previously, he worked as a Researcher with Finnish Forest Research Institute, Helsinki, Finland, as a Postdoctoral Researcher with Yale School of Forestry and Environmental Studies, New Haven, CT, USA, as a Senior Researcher in forest planning with the School of Forest Sciences, University of Eastern Finland, Joensuu, Finland, and as an Acting Professor of Forest Planning with the University of Helsinki, Helsinki, Finland. He has authored over 50 peer-reviewed publications. His research interests include applied statistics for forest inventory, forest growth and yield, and remote sensing.

Anni Virolainen received the M.Sc. degree in mathematics from the University of Eastern Finland, Joensuu, Finland, in 2010.

From 2010 to 2013, she worked as a Researcher with the University of Eastern Finland. She has authored three reviewed scientific papers.

Jukka Tuomela was born in Imatra, Finland, on February 10, 1959. He received the M.S. degree in electrical engineering from Helsinki University of Technology, Espoo, Finland, in 1986, and the Ph.D. degree in mathematics from the Universit de Paris, Paris, France, in 1992.

From 1992 to 1998, he was a Lecturer with Helsinki University of Technology. Currently, he is a Professor of Mathematics with the University of Eastern Finland, Joensuu, Finland. His research interests include differential equations and their computational analysis.

Prof. Tuomela is a Member of Finnish Mathematical Society, European Mathematical Society, and American Mathematical Society.



Petteri Packalén was born in Rauma, Finland, in 1973. He received the M.Sc., Lic.Sc., and D.Sc. degrees in forestry from the University of Joensuu, Joensuu, Finland, in 2002, 2007, and 2009, respectively.

Currently, he is a University Researcher in forest mensuration with the School of Forest Sciences, Faculty of Science and Forestry, University of Eastern Finland, Joensuu, Finland. Previously, he has been an Assistant, Senior Assistant, and Professor with the Faculty of Forestry, University of Joensuu. From August 2011 to July 2012, he was a Visiting Research Scientist at Oregon State University, Corvallis, OR, USA. Since 2007, he has also been a Consultant for remote sensing-based forest inventory. He has authored over 70 peer-reviewed research articles. His research interests include both practical and theoretical aspects of utilizing remote sensing data in the monitoring and assessment of the forest environment, nearest neighbor imputation, combined use of ALS and spectral data in forest inventory, and the use of ALS in wildlife management.