Estimating forest stand density and structure using Bayesian individual tree detection, stochastic geometry, and distribution matching*

Kasper Kansanen[a,b,*], Jari Vauhkonen[c], Timo Lähivaara[d], Aku Seppänen[d], Matti Maltamo[e], Lauri Mehtätalo[a]

[a] *School of Computing, University of Eastern Finland, Postal Box 111, 80101 Joensuu, Finland*
[b] *Department of Forest Sciences, University of Helsinki, Postal Box 27, 00014 Helsinki, Finland*
[c] *Natural Resources Institute Finland (Luke), Bioeconomy and Environment Unit, Yliopistokatu 6, 80101 Joensuu, Finland*
[d] *Department of Applied Physics, University of Eastern Finland, Postal Box 1627, 70211 Kuopio, Finland*
[e] *School of Forest Sciences, University of Eastern Finland, Postal Box 111, 80101 Joensuu, Finland*

Abstract

Errors in individual tree detection and delineation affect diameter distribution predictions based on crown attributes extracted from the detected trees. We develop a methodology for circumventing these problems. The method is based on matching cumulative distribution functions of field measured tree diameter distributions and crown radii distributions extracted from airborne laser scanning data through individual tree detection presented by Vauhkonen and Mehtätalo (2015). In this study, empirical distribution functions and a monotonic, nonlinear model curve are introduced. Tree crown radius distribution produced by individual tree detection is corrected by a method taking into account that all trees cannot be detected. The evaluation is based on the ability of the developed model sequence to predict quadratic mean diameter and total basal area. The studied data consists of 36 field plots in a typical boreal managed forest area

*Declarations of interest: none

*Corresponding author
Email addresses: kasperkansanen@gmail.com (Kasper Kansanen), jari.vauhkonen@luke.fi (Jari Vauhkonen), timo.lahivaara@uef.fi (Timo Lähivaara), aku.seppanen@uef.fi (Aku Seppänen), matti.maltamo@uef.fi (Matti Maltamo), lauri.mehtatalo@uef.fi (Lauri Mehtätalo)

in eastern Finland. The suggested enhancements to the model sequence produce improved results in most of the test cases. Most notably, in leave-one-out cross-validation experiments the modified models improve RMSE of basal area 13% in the full data and RMSE of quadratic mean diameter and basal area 69% and 11%, respectively, in pure pine plots. Better modeling of the crown radius distribution and improved matching between crown radii and stem diameters add the operational premises of the full distribution matching.
Keywords: histogram matching, forestry, forest inventory, airborne laser scanning, Light Detection And Ranging (LiDAR)

1. Introduction

The distribution of tree diameters at breast height (DBH, measured outside bark at 1.3 m aboveground) characterizes the economic and ecological values of a forest. Predicting the diameter distribution is an important task for forest inventories, because it can be used to calculate further statistics such as basal area, volume and biomass. Predicting the diameter distribution has therefore been studied based on both of the most prevalent approaches to utilize remote sensing (especially airborne laser scanning, ALS, data), i.e., the area-based and individual tree detection (ITD) approaches.

In the area-based approach, statistics of the ALS return height distribution are used to explain forest attributes of interest with parametric models or non-parametric prediction techniques. To obtain diameter distribution, these techniques are applied to predict or recover theoretical distribution function parameters (e.g. Gobakken and Næsset, 2004; Mehtätalo et al., 2007; Thomas et al., 2008) or impute tree lists using k-nearest neighbor methods (e.g. Packalén and Maltamo, 2008; Shang et al., 2017; Lamb et al., 2017). Also more theoretical approaches to link the ALS return height distribution to the diameter distribution have been experimented (Magnussen and Renaud, 2016; Spriggs et al., 2017). Although improvements to area-based diameter distribution predictions are still possible, the methods have already been established in operationally run inventories

(Maltamo and Packalen, 2014) and successfully applied to forest types ranging from regular plantations (Arias-Rodil et al., 2018; Maltamo et al., 2018) to tropical forests with more variation in their structure (Rana et al., 2017).

In ITD, on the other hand, individual tree crowns are algorithmically detected from the data, leading to tree-level attributes such as height and crown radius (e.g. Persson et al., 2002). The diameter distribution is obtained by predicting the DBHs of the detected trees by using the tree-level attributes, possibly together with other ALS features, as model predictors. Recent studies have especially applied multi-layered or fully three-dimensional ITD methods (Reitberger et al., 2009; Li et al., 2012; Duncanson et al., 2014; Lähivaara et al., 2014; Lindberg et al., 2014; Lu et al., 2014; Vega et al., 2014). Lähivaara et al. (2014) assessed the number of trees detected based on two approaches in an area that is also studied by us. They reported an increase from 53% to 70% of trees detected by shifting from image analysis of interpolated surface models (Pitkänen et al., 2004; Pitkänen, 2005) to the developed three-dimensional framework. Both algorithms produced insignificant rates (<1%) of commission errors. However, even the most advanced ITD algorithms cannot be expected to correctly detect and delineate all trees, especially the proportion of them with crowns covered by or interlaced with neighboring trees. These limitations of ITD also have an effect on the diameter distribution estimate (e.g. Vauhkonen and Mehtätalo, 2015).

Knowledge on marked point patterns has been employed to compensate for undetected treees in ITD based on very-high resolution satellite image data (Zhou et al., 2013; Gomes et al., 2018). On the other hand, Mehtätalo (2006) and Kansanen et al. (2016) presented methods for estimating the true, field measured stand density from tree crown objects produced by ITD on ALS data. These methods were based on approximating the probability of detecting individual trees – the detectability – through stochastic geometry (Chiu et al., 2013). Mehtätalo (2006) estimated the detectability assuming that smaller trees are left undetected if their center points are inside the crown of a bigger tree. The method assumed the crowns to follow a Boolean model, with complete spa-

tial randomness of locations and independent identically distributed crown radii. Kansanen et al. (2016) reformulated this estimator to rely on fewer assumptions on the forest structure. An empirical detectability was based on a morphological transformation of the union of detected crowns larger than the tree whose detectability was being calculated. The developed Horvitz-Thompson type estimator (Kansanen et al., 2016) outperformed the one based on the theoretical area fraction of the Boolean model (Mehtätalo, 2006) in 36 field plots used for validating the method. These methods can also correct the biased crown radius distribution by adjusting it using the estimated detectability.

Predicting tree stem attributes for all trees would require a tree-level matching between the field measured and remotely sensed tree attributes, which cannot be achieved in the case of tree detection errors. To circumvent this problem, Vauhkonen and Mehtätalo (2015) proposed that stem diameter distributions and crown radii distributions derived through ITD could be directly related by building upon a histogram matching technique frequently used in digital image processing (Gonzalez and Woods, 2008). The developed distribution matching method avoids the problem of tree matching by matching the percentiles of the distributions in question as pseudo data and modeling the transformation from crown radius to stem diameter using these data points. Vauhkonen and Mehtätalo (2015) also showed that it was beneficial to use corrected crown radius distributions for the distribution matching. However, they used the correction method of Mehtätalo (2006) in data where only less than half of the plots met the stated assumptions on the spatial randomness and independence of the crown radii. The correction failed especially in forests with regular tree patterns, and although the method is promising, it is not operational because of the restrictive assumptions.

Based on the text above, it could be possible to improve the results from Vauhkonen and Mehtätalo (2015) by critically re-examining their methodological choices. First, because an accurate stand density estimate was crucial also with respect to the accuracy of the diameter distribution predictions, either an improved ITD algorithm or a better estimator for the detectability of the trees

could improve the results. Second, Vauhkonen and Mehtätalo (2015) modeled both the crown radii and stem diameter distributions as having Weibull forms to produce smooth transformations from one distribution to the other. However, since assuming a parametric distribution form is not fundamentally required by the method, a more flexible modeling approach could be beneficial to describe more complex forms of the diameter distribution. Finally, the ITD-detected tree heights were not utilized although they were available. The distribution of the detected heights could be assumed useful for predicting the diameter distribution of trees.

In this study, we investigate whether distribution matching (Vauhkonen and Mehtätalo, 2015) could be improved by enhancing the modeling chain for both the ITD and plot-level matching. Especially, we test a more sophisticated ITD algorithm (Lähivaara et al., 2014), density correction (Kansanen et al., 2016), and matching function for the transformation from tree crown radius to stem diameter. The proposed changes are hypothesized to improve the accuracy of the diameter distribution predictions, but also the operational feasibility of the full method chain, because of reducing a number of assumptions made regarding spatial point patterns and distributional forms of the stem diameters and crown radii.

2. Material

The study area is a typical boreal managed forest area in eastern Finland (62° 31′ N, 30° 10′ E) with Scots pine (Pinus sylvestris L.) as the dominant tree species. It represents 73% of the volume, Norway spruce (Picea abies [L.] H. Karst.) 16% of the volume and deciduous trees altogether about 11% of the volume. The same area was previously studied by Packalén et al. (2013), who describe the measurements carried out in more detail.

The ALS data for the area were collected on 26 June 2009 using an Optech ALTM Gemini laser scanning system from approximately 720 m above ground level with a field of view of 26°. The side overlap of 55% in the data acquisition

Attribute	**n**	**mean**	**sd**	**min**	**max**	**20**	**25**	**30**
λ, stems $\cdot$ ha^{-1}	36	1218.8	538.0	466.7	2560	6	20	10
	18	1121.4	487.8	544.4	2250	4	11	3
	43	1291.8	592.3	512	2875	14	23	6
	20	1204.2	582.6	512	2225	8	10	2
QMD, cm	36	17.2	4.3	10.2	29.0			
	18	16.9	3.5	11.2	23.6			
	43	16.4	3.5	11.5	27.2			
	20	16.7	3.5	11.5	23.4			
BA, m^2 $\cdot$ ha^{-1}	36	24.9	6.3	15.4	40.1			
	18	22.6	4.4	15.4	32.4			
	43	24.4	6.2	13.8	36.2			
	20	23.5	6.6	13.8	35.1			

Table 1: Mean, standard deviation, minimum and maximum of stand density (λ), quadratic mean diameter (QMD) and basal area (BA) in Kiihtelysvaara. The full data usable in our analysis contains 36 field plots, of which 18 have $> 95\%$ of basal area Scots pine (*Pinus sylvestris L.*). The training set needed by the tree detection algorithm (see Section 3.1.2) contains 43 field plots, of which 20 have $> 95\%$ of basal area Scots pine. The columns "20", "25" and "30" show the numbers of plots having that side length in metres.

means that each location was covered from two flight lines in order to increase the probability that trees have ALS hits each side. Pulse repetition frequency was set to 125 kHz, and when the instrument was operated in a multipulse mode, the nominal sampling density was 11.9 pulses/m^2.

The field measurements were carried out in May–June 2010. Altogether 79 field plots were placed subjectively, attempting to record the species and size variation over the area. The plot size varies between 20×20 m^2, 25×25 m^2 and 30×30 m^2. Trees were chosen under the criterion of either DBH $\geq$ 5 cm or height $\geq$ 4 m. Location, DBH and height were measured and species was registered. The full plot data were distributed to training and validation data sets according to the needs of the tree detection algorithm (Section 3.1.2): only plots that were lying below the flight lines were chosen to the validation set. The central plot-level attributes for the 36 plots used in this study, and the 43 plots used as training data by the tree detection algorithm, are presented in Table 1.

3. Methodology

As motivated in the Introduction, we attempt to improve the distribution matching method of Vauhkonen and Mehtätalo (2015). The method can be broken down to three separate steps and presented as a sequence "ITD + Correction + Matching", i.e. the full method requires (1) an ITD algorithm to detect and segment treetops (Section 3.1); (2) a method to model the tree crown radius distribution and correct it for the missing small trees (Section 3.2); and (3) a method to transform the crown radii distribution to tree diameter distribution (Section 3.3). Fig. 1 is a schematic diagram of the sequence.

The original method of Vauhkonen and Mehtätalo (2015) is considered as a benchmark and described as a sequence of 2D-ITD + Boolean + Polynomial. To assess the effects of each component on the accuracies of the diameter distributions, we consider three alternative model sequences that are obtained by modifying the parts of the benchmark sequence one by one, as reasoned below:

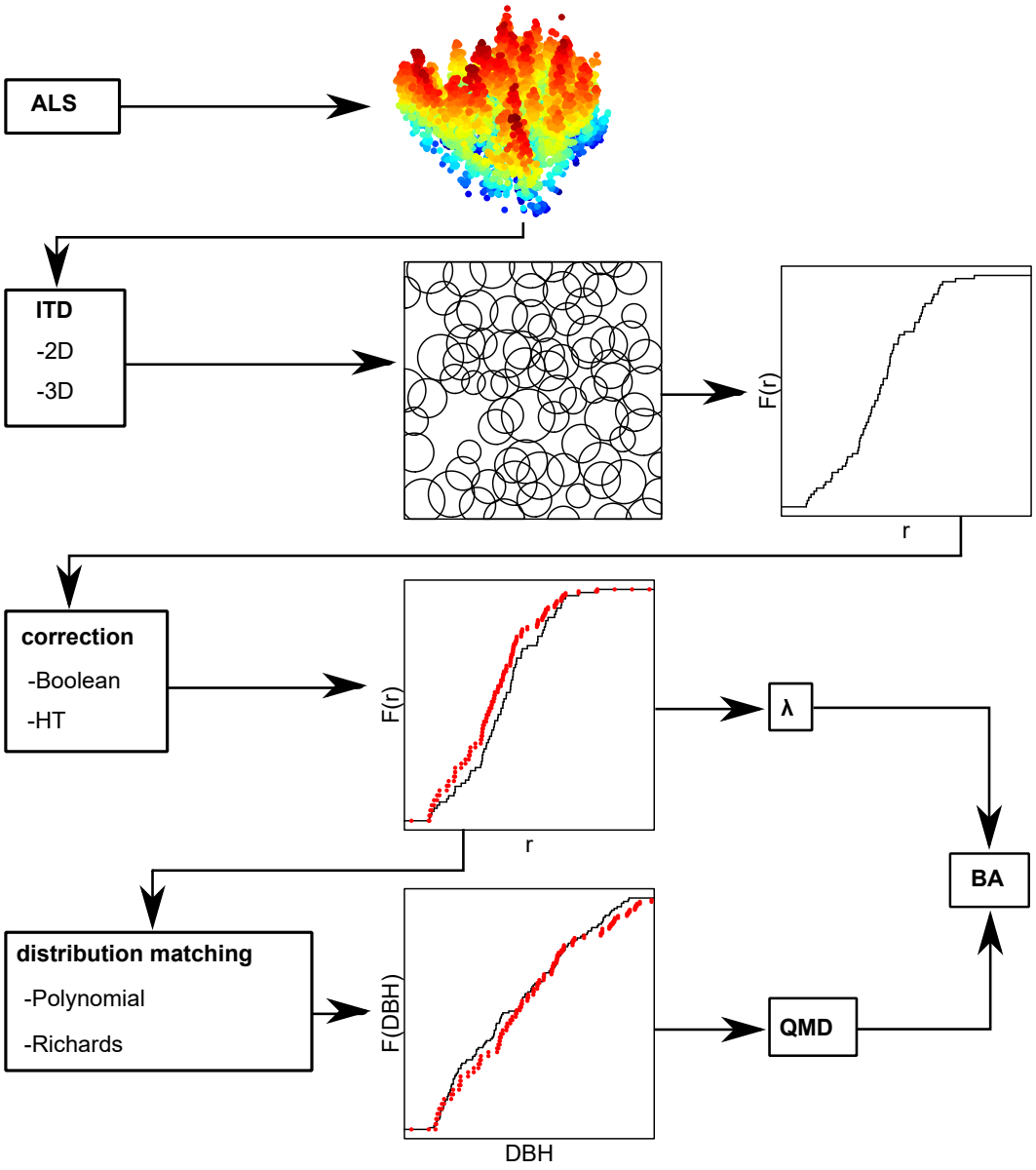

Figure 1: Schematic diagram of the modeling chain. Airborne laser scanning data are first interpreted by an individual tree detection algorithm that produces tree objects and crown radius (r) distributions. These distributions are corrected to compensate for tree detection errors, which produces corrected crown radius distributions (illustrated by red line) and estimates of stand density λ. The corrected crown radius distributions are matched to distributions of DBH, producing a transformation function used to predict the latter from the former. The evaluation is based on the estimated stand density λ, quadratic mean diameter (QMD) of the predicted DBH distribution, and basal area estimated using both λ and QMD.

1. **3D-ITD** + Boolean + Polynomial: Conventional 2D-ITD method based on image analysis of interpolated canopy height surfaces (Section 3.1.1) is replaced by an improved ITD algorithm that uses a priori knowledge on tree crown shapes and operates in 3D space (Section 3.1.2). Expected improvements are due to being able to detect more trees, but also because the initial crown radius distribution obtained using rotationally symmetric tree crown approximations may be more compatible with the Correction step.
2. 3D-ITD + **HT** + Polynomial: The correction based on assuming a Boolean model with complete spatial randomness of locations and independent identically distributed crown radii (Section 3.2.1) is replaced by a reformulated, Horvitz-Thompson type (HT) estimator (Section 3.2.2). Expected improvements are due to more realistic modeling of the proportion of small trees with fewer assumptions on the spatial patterns.
3. 3D-ITD + HT + **Richards**: Distribution matching function with a polynomial model form (Section 3.3.1) is replaced by a nonlinear function form, also known as the Richards' curve (Section 3.3.2). Expected improvements are due to monotonically increasing function form that better fits the data.

3.1. Individual tree detection

The main task for the ITD in our method chain is to obtain the initial crown radius distribution, which could be possible based on a number of different approaches. Since the benchmark ITD method (Vauhkonen and Mehtätalo, 2015) was based on image analysis of canopy surface height models interpolated from the point data, it is reasoned to test an approach with different fundamentals to assess the importance of ITD in the model sequence. Thus, although ITD methods similar to the benchmark are often referred to as "2.5D" because of including height, the abbreviations for our methods are selected to emphasize a main difference between the methods to operate either with raster images (2D-ITD) or vector data examined in 3D point space (3D-ITD). As mentioned in

the Introduction, both the approaches were compared for estimating the stem number in the presently studied area by Lähivaara et al. (2014).

3.1.1. 2D-ITD

The 2D-ITD method (Pitkänen et al., 2004) carries out adaptive low-pass filtering aiming to produce a single local height maximum for each tree top, using Gaussian scale parameters that were subjectively defined for different tree height classes as explained by Packalén et al. (2013). Segments are created around the local maxima of the height-filtered canopy surface model using watershed segmentation to delineate the tree crowns (Pitkänen, 2005). The drainage direction following segmentation algorithm delineates tree crowns as regions bounded by other segments and the background, determined as pixels with height < 2 m. The crown dimensions are therefore obtained solely based on image analysis of eight-neighborhoods of the pixels in the interpolated canopy surfaces. The unfiltered surface model pixels with highest value within the segments were considered as tree locations and the maximum diameter in four cardinal directions passing the crown location was taken as the crown diameter.

3.1.2. 3D-ITD

In this ITD method, single tree crowns are modeled by parametric, rotationally symmetric surfaces; the parameters defining the dimensions of each crown are: crown radius, the crown height, the lower limit of the living crown, and the crown shape parameter. These parameters, and the horizontal coordinates of tree crown center points are estimated based on ALS data. The estimation problem is written in the Bayesian framework of inverse problems (Kaipio and Somersalo, 2005) – the advantage of this approach over, e.g., ordinary least squares fitting or maximum likelihood estimation is that it allows for utilizing *a priori* information on the tree shapes in the ALS based estimates. As a Bayesian estimate for the model unknowns – the positions and crown shape parameters of each tree – we consider the maximum a posterior (MAP) estimate which is computed by a Newton-based optimization method.

As in Lähivaara et al. (2014), the likelihood model is based on an approximation of additive, mutually independent Gaussian noise in the ALS measurements, and all the model unknowns are modeled as Gaussian random variables on the basis of a training set consisting of field measurements from 43 plots together and allometric models for tree shapes by Muinonen (1995). The ITD is applied to a total of 36 plots that were different from plots in the training set.

3.2. Stand density and crown radii distribution corrections

The two correction methods discussed have a common basis in stochastic geometry (Chiu et al., 2013). The forest is interpreted as a realisation of a germ-grain model of discs $\Xi = \bigcup B(x_i, R_i)$ in some area of interest $W \subset \mathbb{R}^2$. Here, x_i are locations of crown center points, distributed as a homogeneous point process of intensity λ (the stand density). The surface areas under tree crowns are modeled as closed discs B with random radii R_i. From the output of the ITD (estimates of the tree locations and crown shapes), we derive $\hat{\Xi}$, the collection of patches on the ground surface covered by the crowns. A standard germ-grain model is the Boolean model, where the disc radii are independently and identically distributed and the disc center points are distributed as a Poisson process. This means that the number of points in an arbitrary planar set is Poisson distributed with parameter that depends on the area of the planar set and the intensity λ. The locations of the points are completely independent of each other.

3.2.1. Boolean detectability

Under the Boolean model assumption, the tree density can be written as

$$\lambda = -\frac{\ln(1 - cc)}{\pi E[R^2]} \tag{1}$$

measured in trees $\cdot$ ha^{-1}, where cc is the relative canopy cover and $E[R^2]$ the expected value of the squared crown radius (Mehtätalo, 2006). Additional assumption of a tree being detectable only when its location is not covered by

the crowns of the larger trees leads to the probability to be detected p, or the detectability:

$$p(r) = \exp\left(-\lambda\pi \int_r^\infty t^2 f(t)dt\right),$$

where f is the probability density function of the crown radii. The density function of the detected tree crown radii can then be written as

$$f_D(r) = \frac{p(r)f(r)}{\int_0^\infty p(t)f(t)dt}$$

and used to estimate the parameters of f through maximum likelihood. The fitted distribution f is then used to calculate $E[R^2]$ to be used in Equation (1). Vauhkonen and Mehtätalo (2015) assumed f to be a Weibull density.

3.2.2. Horvitz-Thompson type detectability

Kansanen et al. (2016) presented a Horvitz-Thompson type stand density estimator. Let us consider each detected crown radius r_i^* as a representative of a size class. The total number of trees in a size class r_i^* is calculated by scaling the detected number of trees, which we assume to be one, by the detectability p:

$$\hat{N}_i = \frac{1}{p(r_i^*)}.$$

If n trees have been detected, the stand density estimator is formed by summing the size class specific tree numbers and scaling with the area of W in hectares:

$$\hat{\lambda} = \frac{\sum_{i=1}^n \hat{N}_i}{|W|}.$$

Detectability for a certain size class is estimated through the probability of a uniformly distributed random point hitting a set formed by the crowns of larger trees in such a way that its crown is suitably covered. It can be written as

$$p_\alpha(r) = 1 - \frac{|W \cap (\hat{\Xi}_{R>r} \ominus B(o, \alpha r))|}{|W|},$$

where r is the crown radius, $\hat{\Xi}_{R>r}$ is a subset of the detected Boolean model formed by discs larger in radii than r, $B(o,r)$ is an origin-centred closed disc of radius r, $|.|$ is an area operator and $\ominus$ a Minkowski-subtraction or erosion,

$$\hat{\Xi}_{R>r} \ominus B(o,r) = \{x \in \hat{\Xi}_{R>r} : B(x,r) \subset \hat{\Xi}_{R>r}\}.$$

The parameter $\alpha \in [0,1]$ controls the proportion of radius that should be covered by the larger trees for non-detection. For example, $\alpha = 1$ corresponds to a situation where trees are not detectable only if their crowns are fully covered by larger ones, whereas $\alpha = 0$ corresponds to a situation where a tree is detectable if the center point of the crown is not covered by a larger tree. Because the optimal value of α likely depends on the ITD algorithm used, the quality of ALS data and properties of the forest, it was determined based on earlier tests in the Kiihtelysvaara data described in Kansanen et al. (2016). The buffer size was fixed to $\alpha = 0.4$, which yielded best results in a cross-validation experiment that further solidified the position of the Horvitz-Thompson type estimator as the best method tested and showed that the estimator is rather robust to the choice of α.

The size class specific tree numbers $\hat{N}_i$ can be used to nonparametrically estimate the distribution of crown radii. This is done by using the tree numbers as weights in an empirical distribution function:

$$F(r) = \frac{\sum_{r_i^* \leq r} \hat{N}_i}{\hat{\lambda}|W|}. \tag{2}$$

We need the percentiles of this distribution, which are calculated through the inverse of the empirical distribution function. All of the calculations related to the weighted empirical distribution functions were done with Hmisc package of R (Harrell Jr et al., 2016).

3.3. Distribution matching

We wish to find a monotonically increasing transformation from the corrected distribution of ITD crown radii to the distribution of field-measured

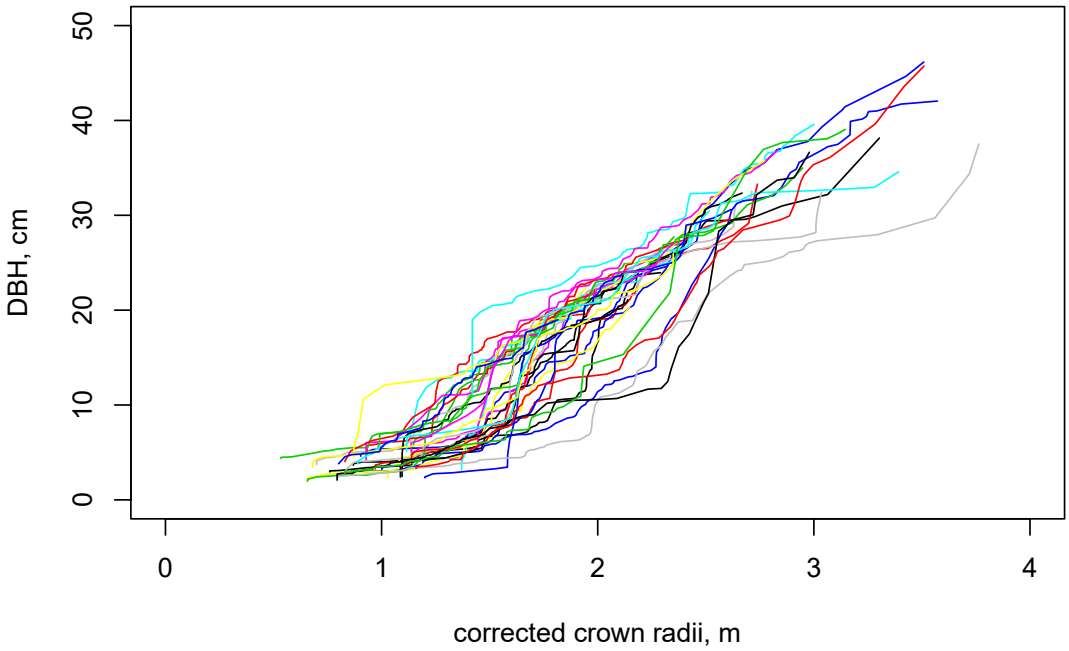

Figure 2: The percentiles of the distribution of diameters at breast height as a function of the percentiles of corrected crown radii distribution in the test data, corrected with the method of (Kansanen et al., 2016). Each line represents one field plot.

stem diameters. For random variables X and Y having cumulative distribution functions F_X and F_Y and $Y = g(X)$ where g is monotonically increasing it can be shown that

$$F_X(x) = P\{X \leq x\} = P\{g^{-1}(Y) \leq x\} = P\{Y \leq g(x)\} = F_Y(g(x)) = t,$$

which leads to formulas $F_Y^{-1}(t) = g(x)$ and $F_X^{-1}(t) = x$. Hence, given t-percentiles of two distributions connected by some unknown transformation g this transformation can be estimated by using the percentiles as data points.

Let d_{ij} be the jth percentile ($j = 1, 2, \ldots, 99$) of the diameter distribution on plot i, and let r_{ij} be the corresponding percentile of the corrected crown radii distribution. We define the transformation g as a mixed-effects model

(Lindstrom and Bates, 1990):

$$d_{ij} = g(r_{ij}, \boldsymbol{\phi}_i) + \varepsilon_{ij}, \tag{3}$$

where the parameter vector $\boldsymbol{\phi}_i$ consists of fixed effects $\boldsymbol{\beta}$ common to all data, possible plot-specific covariates $\boldsymbol{x}_i$, and plot-specific random effects $\boldsymbol{b}_i \sim \mathrm{N}(\boldsymbol{0}, \sigma^2\boldsymbol{D})$ that are independent for all $i \neq k$, that is, from plot to plot. The covariance matrix $\sigma^2\boldsymbol{D}$ is unknown and has to be estimated. In addition the residual errors $\varepsilon_{ij} \sim \mathrm{N}(0, \delta^2)$ are assumed to be independent for all $ij \neq kl$ with an unknown variance δ^2.

When a mixed-effects model is fitted, the predicted values of random effects $\tilde{\boldsymbol{b}}_i$ are only available for plots with observations of the response variable d. Hence, only the expected value (zero) of $\boldsymbol{b}_i$ can be used for plot-specific predictions. However, if a plot-specific covariate explained the between-plot variation, the predicted values of random effects could possibly be replaced by such covariate(s) to mimic the between-plot differences described by the random effects. Our motivation to add plot-specific covariates to the model was in particular to replace the random effects in prediction situations as reasoned above.

Several different covariates were tested for inclusion in the model to make plot-specific predictions. The potential covariates included the mean and standard deviation of ALS return heights, the 5th, 10th, 15th, ..., 95th percentiles and corresponding proportional densities of the ALS-based canopy height distribution computed according to Korhonen et al. (2008), and also stand density estimates, canopy coverage estimates derived from ITD, means, variances and the 5th, 10th, 15th, ..., 95th percentiles of the (non-corrected) ITD detected tree height distribution. The details of the model fitting and covariate choosing procedure are discussed in the following sections. The model fitting was done with the nlme package of R (Pinheiro et al., 2016).

3.3.1. Polynomial model

Vauhkonen and Mehtätalo (2015) assumed g in Equation (3) having a quadratic

polynomial form:

$$g(r_{ij}) = \beta_1 r_{ij} + \beta_2 r_{ij}^2 + b_{1i} r_{ij} + b_{2i} r_{ij}^2, \quad (4)$$

where β_1 and β_2 are fixed effects, b_{1i} and b_{2i} are the plot-specific random effects. Equation (4) is used only with its predicted values of random effects, in other words, without added covariates. When adding these variables, the transformation is first fitted in a simplified form

$$g(r_{ij}) = \beta_1 r_{ij} + \beta_2 r_{ij}^2 + b_{1i} r_{ij}$$

to avoid overfitting. Similar to Vauhkonen and Mehtätalo (2015), we predict the values of the random effect using a linear regression model with one plot-specific covariate. The most suitable covariate for the model was identified as the covariate x_i having the highest absolute correlation with predicted b_{1i}. It is added to the model:

$$g(r_{ij}) = \beta_1 r_{ij} + \beta_2 r_{ij}^2 + (\beta_3 x_i + b_{1i}) r_{ij},$$

and the model is fitted again. When predicting stem diameters with the model, the random effects are set to their expected value, zero, because they are not known in a prediction situation.

3.3.2. Model with Richards' curve

The quadratic transformation is not necessarily monotonically increasing. This flaw can be corrected by using a nonlinear transformation function, for example the generalized logistic function, also known as Richards' curve:

$$g(r_{ij}, \boldsymbol{\phi}_i, v) = \frac{K_i}{(1 + \exp(Q_i - B_i r_{ij}))^{\frac{1}{v}}},$$

where the parameters are divided to $\boldsymbol{\phi}_i = [Q_i, B_i, K_i]^T$ and v to emphasize v as a purely fixed effect. The model was chosen by visual inspection of the Kiihtelysvaara data (Fig. 2). The data seems to support the logistic curve, having variation between plots in the sigmoidal center points, growth rates

and maximum values, governed by the plot-specific parameters Q_i, B_i and K_i, respectively. Possible asymmetric behaviour around the sigmoidal center points is taken into account with parameter v. Although preliminary analysis of the data by fitting separate models to plots showed variation also in v, we were not able to include it as a plot-specific parameter due to convergence problems in model fitting.

The plot-specific parameters were first modeled as $\boldsymbol{\phi}_i = \boldsymbol{\beta} + \boldsymbol{b}_i$. The variables $\boldsymbol{x}_i$ with the highest absolute correlations with $\boldsymbol{b}_i$ (separately for each parameter) were added to the model, giving $\boldsymbol{\phi}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{x}_i + \boldsymbol{b}_i$ where $\boldsymbol{\beta}_1 \boldsymbol{x}_i$ is an element-wise multiplication, and the model was fitted again. When predicting stem diameters with the model, the random effects were set to their expected value, zero.

The model fitting procedure requires starting values for the fixed effects. Preliminary values were chosen as described in Fekedulegn et al. (1999), and refined by minimizing residual squared error of the Richards' curve without any random effects. These same values were used for $\boldsymbol{\beta}_0$ when fitting the model with covariates, while $\boldsymbol{\beta}_1$ were set to zero.

3.3.3. The estimated tree diameter distribution

Let us mark the random variables related to crown radii and diameter at breast height as R and DBH, respectively. To formulate $F_{DBH}(d)$, one has to consider the probability

$$F_{DBH}(d) = P\{DBH \leq d\} = P\{g(R) \leq d\}.$$

The inequality $g(r) \leq d$ needs to be solved to produce probabilities regarding R, hence performing a change of variable in the cumulative distribution function of R. The distribution function resulting from a Weibull distribution of crown radii and a quadratic transformation is presented in the Appendix. When using nonparametric distributions, the cumulative distribution function for diameters at breast height in plot i is simply

$$\hat{F}_i(d) = \frac{\sum_{g(r^*_{ij}) \leq d} \hat{N}_{ij}}{\hat{\lambda}_i |W_i|},$$

where summation goes over the index j. It is essentially a weighted empirical distribution function calculated from the detected crown radii transformed to diameters with the transformation g weighted by the corresponding sizes of the radius classes. Notice the similarity to the corrected cumulative distribution in Equation (2).

3.4. Performance measures

We use quadratic mean diameter (QMD) measured in cm and basal area (BA) measured in $\mathrm{m}^2 \cdot \mathrm{ha}^{-1}$ as measures of model performance. The true value for QMD in plot i is calculated as

$$QMD_i^{true} = \sqrt{\frac{\sum_{j=1}^{n_i} (d^*_{ij})^2}{n_i}},$$

where n_i is the number of trees in the plot and d^*_{ij} is the observed diameter at breast height of tree j. The true value for BA is calculated as

$$BA_i^{true} = \lambda_i \cdot QMD_i^{true} \frac{\pi}{40000}.$$

We estimate QMD as

$$\widehat{QMD}_i = \sqrt{\frac{\sum_j \hat{N}_{ij} (g(r^*_{ij}))^2}{\sum_j \hat{N}_{ij}}},$$

where index j goes over the detected tree crown radii, when using the nonparametric models and as

$$\widehat{QMD}_i = \sqrt{E[d^2]} = \sqrt{\int_{-\infty}^{\infty} (g(r))^2 f(r) dr}$$

when using the methods with the quadratic transformation g and probability density function f for the crown radii. The estimated BA is calculated as

$$\widehat{BA}_i = \hat{\lambda}_i \cdot \widehat{QMD}_i \frac{\pi}{40000}.$$

It should be noted that the estimate of BA depends on both the estimates of QMD and tree density.

Root mean squared errors,

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

means of errors

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

and their normalized variants (RMSE%, ME%) calculated by dividing the error with the mean of true values and multiplied by 100 are used as goodness-of-fit measures. In the formulas y_i is the true value of plot-level statistic, $\hat{y}_i$ the estimate and n the number of plots.

To compare the fitting of the estimated diameter distributions we also calculate L^2 distances induced by the well known L^2 norm (Rudin, 1987, Chap. 3), defined as

$$||F(d) - \hat{F}(d)||_2 = \sqrt{\int_{-\infty}^{\infty}(F(t) - \hat{F}(t))^2 dt},$$

where F is the true empirical cumulative distribution function and $\hat{F}$ is the estimated cumulative distribution function. This integral is approximated numerically by the R function integrate.

The Clark-Evans aggregation index (Clark and Evans, 1954) with the edge-effect correction of Donnelly (1978) was calculated for every plot to assess the effect of spatial distribution of locations on the estimates and their errors. Index values close to one suggest complete spatial randomness, whereas values > 1 suggest ordering and values < 1 suggest clustering.

In addition to considering the performance measures calculated from fitted distributions, leave-one-out (LOO) cross-validation experiments were performed to assess the predictive capabilities of the models. In LOO the n plots are divided into $n - 1$ plots where the model is fitted and the one plot where these

fitted models are used for predicting. This is done n times, leading to a prediction for every plot. In every prediction case the whole distribution matching procedure is performed: in the $n-1$ plots the model is first fitted with random effects, the best covariate explaining the variation in the predicted values of random effects is added to the model and the model is fitted again, and the prediction is performed, without random effects, which are not available during prediction.

4. Results

Vauhkonen and Mehtätalo (2015) considered only pine-dominated plots, defined as plots with $> 95\%$ of the basal area consisting of Scots pine. A precursory analysis comparing pine-dominated plots to those dominated by the other species indicated that random effects were differently distributed in these two subsets of data. This resulted to selecting different covariates for plots dominated either by pine or other species. Hence, we evaluated the predictions separately for full data and pure pine plots.

4.1. Stand density estimation

Results of stand density estimation are presented in Table 2. The results related to 3D-ITD without correction and with both correction methods in the full data have been previously presented in Kansanen et al. (2016). In full data, the RMSE of stand density is the highest when 2D-ITD is used without corrections. Switching to 3D-ITD provides a reduction to it. The correction methods further reduce the RMSE for both ITD methods. The HT correction provides substantially lower RMSE than the Boolean correction. The reduction in RMSE going from the worst results to the best results is 69%. All the corrections also shift ME considerably towards zero.

In the pure pine plots, both of the ITD methods have lower values of RMSE and ME closer to zero than in the full data. When the Boolean correction is used with 2D-ITD, the RMSE is higher than with no correction. With 3D-ITD the Boolean and HT corrections again produce lower RMSE values than

n	ITD	Correction	RMSE	RMSE%	ME	ME%
36	2D-ITD	-	718.5	59.0	-564.4	-46.3
	2D-ITD	Boolean	541.8	44.5	-27.2	-2.2
	3D-ITD	-	486.8	39.9	-380.1	-31.2
	3D-ITD	Boolean	303.1	24.9	-21.2	-1.7
	3D-ITD	HT	221.6	18.2	-39.5	-3.2
18	2D-ITD	-	500.0	44.6	-384.3	-34.3
	2D-ITD	Boolean	574.9	51.3	3.7	0.3
	3D-ITD	-	302.8	27.0	-232.6	-20.7
	3D-ITD	Boolean	280.1	25.0	103.3	9.2
	3D-ITD	HT	177.0	15.8	73.1	6.5

Table 2: Errors of stand density estimates (stems $\cdot$ ha^{-1}) used in predicting basal areas. The column "n" specifies whether the full 36 field plots or the 18 plots with $> 95\%$ pine were used. Column "ITD" specifies whether the original algorithm by Pitkänen or the algorithm by Lähivaara et al. was used. The column "Correction" specifies the type of stand density estimator used, see Section 3.2.

using no corrections. Contrary to the full data, all of the correction methods produce positive ME values, indicating overestimation. The result of HT could be improved by using a slightly higher value of α.

4.2. Distribution matching

We present results of distribution matching relating to QMD, BA and L^2 distances in three different cases: (1) in the modelling data using predicted values of random effects, (2) in the modelling data with added ALS or ITD covariates that try to explain the variation in the predicted values of random effects and leaving predicted values of random effects out (i.e. giving them their expected value 0), and (3) leave-one-out cross-validation (LOO), again with added covariates and no random effects, which are not available for the plot where the prediction is performed. The first case illustrates the potential of the model if the variation in the shape of the model curve from plot to plot could be estimated optimally, and tells mostly about model fit. The second case

shows the model performance when the optimal values of random effects can not be utilized (i.e., prediction), but tells still about model fit. The third case illustrates the model performance in a practical prediction situation.

4.2.1. All plots

When predicted random effects are used in distribution matching, progressively better error values are achieved when modifying the benchmark model (2D-ITD + Boolean + Polynomial) by changing the ITD algorithm, correction method and distribution matching model function, especially with regards to BA (Table 3, rows 1-4). The largest improvements are caused by changing the correction method from Boolean to HT, which is explained by the improved estimates of stand density (Table 2). 3D-ITD + HT + Richards produces the smallest RMSE for QMD and BA.

When covariates are included in the models, and the resulting models are used without predicted random effects, all of the modified models still have lower RMSE of QMD than the benchmark but they do not differ from each other very much. The RMSE values for BA follow the same order as the stand density estimates used in calculating them. 3D-ITD + HT + Richards has clearly the highest ME of both QMD and BA in this case.

Leave-one-out cross-validation results in 3D-ITD + Boolean + Polynomial having the lowest RMSE for QMD and BA, and 3D-ITD + HT + Richards having the highest RMSE for QMD and second highest for BA (Table 3, rows 9-12). Although only the model that differs from the benchmark by different ITD achieves a slightly lower RMSE for QMD, all of the modified models achieve lower RMSE of BA than the benchmark in this case. It should be noted that due to problems in model convergence with the Richards' curve when a certain plot was removed, an assumption of diagonal random-effect variance-covariance matrix $\boldsymbol{D}$ had to be made during leave-one-out cross-validation for the prediction in that plot.

When predicted random effects are used, 3D-ITD + HT + Richards achieves the smallest mean and maximum L^2 distances, as well as lowest variation in the

		QMD		BA	
case	**model**	**RMSE%**	**ME%**	**RMSE%**	**ME%**
fit, RE	2D-ITD + Boolean + Polynomial	1.4	-0.7	45.2	2.7
	3D-ITD + Boolean + Polynomial	1.5	-0.8	24.4	2.1
	3D-ITD + HT + Polynomial	1.2	1.0	17.1	2.0
	3D-ITD + HT + Richards	0.7	0.5	17.0	0.9
fit, no RE	2D-ITD + Boolean + Polynomial	13.9	0.1	34.9	1.0
	3D-ITD + Boolean + Polynomial	12.2	-1.3	26.3	1.6
	3D-ITD + HT + Polynomial	12.7	0.5	25.6	3.0
	3D-ITD + HT + Richards	13.3	5.4	23.6	11.0
LOO	2D-ITD + Boolean + Polynomial	14.7	0.2	34.9	1.0
	3D-ITD + Boolean + Polynomial	14.5	-1.8	30.2	1.0
	3D-ITD + HT + Polynomial	15.0	-0.1	30.2	2.2
	3D-ITD + HT + Richards	19.0	7.3	33.6	17.0

Table 3: Normalized root mean square errors and means of errors for quadratic mean diameter and basal area for several model fittings and predictions in the full 36 field plots. The column "case" specifies if the results are calculated in the modeling data with the predicted values of random effects (fit, RE), or if the random effects have been explained by covariates derived from ALS or ITD (fit, no RE), or if the results come from leave-one-out cross-validation (LOO).

distances (Table 4, rows 1-4). These results combined with the performance of the model when predicting QMD indicate that the applied Richards' model is sufficiently flexible and can well model the various forms of the plot-specific relationship between stem diameter and crown radius. When the predicted values of random effects are explained by covariates, the means of L^2 distances for 3D-ITD + HT + Richards and 3D-ITD + Boolean + Polynomial are the lowest and very close to each other (Table 4, rows 5-8). However, the larger standard deviation and maximum value of the former indicate very large errors in the distribution fitting for some plots and very small for others. In leave-one-out cross-validation, the mean, standard deviation and maximum value of L^2 distances for 3D-ITD + HT + Richards are higher than for the other methods (Table 4, rows 9-12). 3D-ITD + Boolean + Polynomial has the best performance, producing lowest mean and maximum distance. Examples of the fitted cumulative distribution functions are shown in Fig. 3.

There were differences in the selected covariates between the different methods when the predicted values of random effects were replaced with covariates. For the benchmark, the 5th ITD height quantile was selected, whereas for the two methods with 3D-ITD and polynomial model curve the variance of ITD heights was selected. For the Richards' curve, the covariate with highest absolute correlations for K and B was the 95th ALS quantile and for Q the 95th proportional density value. The leave-one-out cross-validation selected the same covariates as above for the benchmark, the 3D-ITD + Polynomial methods and B of Richards' curve every time. Covariates for K and Q were mostly as above, but in some cases a few different covariates had the highest absolute correlations.

4.2.2. Pure pine plots

For pure pine plots, the results with the mixed-effects models without any added covariates are similar to each other for all models when it comes to QMD, although RMSE of 3D-ITD + HT + Polynomial is surprisingly high compared to the other methods (Table 5, rows 1-4). 3D-ITD + HT + Richards attains the lowest RMSE for BA, but also exhibits high ME, as do 3D-ITD + Boolean

case	**model**	**mean**	**sd**	**min**	**max**
fit, RE	2D-ITD + Boolean + Polynomial	0.30	0.14	0.11	0.61
	3D-ITD + Boolean + Polynomial	0.29	0.14	0.10	0.61
	3D-ITD + HT + Polynomial	0.31	0.11	0.17	0.55
	3D-ITD + HT + Richards	0.21	0.07	0.11	0.38
fit, no RE	2D-ITD + Boolean + Polynomial	0.60	0.28	0.16	1.31
	3D-ITD + Boolean + Polynomial	0.56	0.21	0.16	1.18
	3D-ITD + HT + Polynomial	0.58	0.20	0.29	1.22
	3D-ITD + HT + Richards	0.53	0.25	0.20	1.53
LOO	2D-ITD + Boolean + Polynomial	0.62	0.30	0.16	1.37
	3D-ITD + Boolean + Polynomial	0.59	0.24	0.16	1.24
	3D-ITD + HT + Polynomial	0.62	0.23	0.30	1.28
	3D-ITD + HT + Richards	0.68	0.35	0.22	1.64

Table 4: Mean, standard deviation, minimum and maximum of L^2 distances between true and estimated cumulative distribution functions in the full 36 field plots. The column "case" specifies if the results are calculated in the modeling data with the predicted values of random effects (fit, RE), or if the random effects have been explained by covariates derived from ALS or ITD (fit, no RE), or if the results come from leave-one-out cross-validation (LOO).

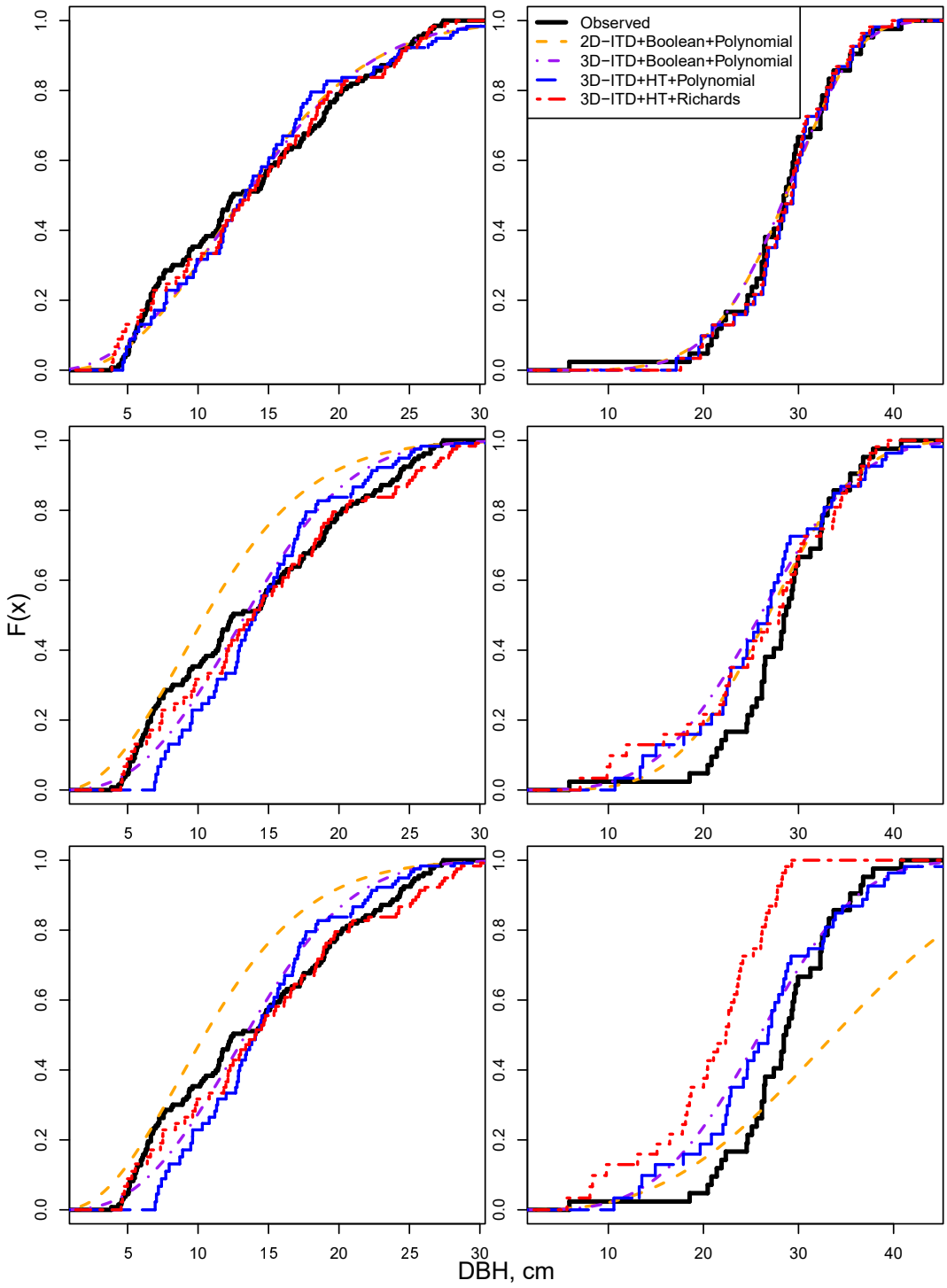

Figure 3: Examples of fitted cumulative distribution functions when all of the 36 plots have been used. At left, the plot where 3D-ITD + HT + Richards achieves the best fit in leave-one-out cross-validation, and at right, the worst fit. Goodness-of-fit measured through L^2 distance. Top panels: fits with the predicted random effects. Middle panels: fits with the random effects explained by covariates derived from ALS and ITD. Bottom panels: the leave-one-out cross-validation fits.

		QMD		BA	
case	**model**	**RMSE%**	**ME%**	**RMSE%**	**ME%**
fit, RE	2D-ITD + Boolean + Polynomial	0.6	-0.4	35.0	1.8
	3D-ITD + Boolean + Polynomial	0.6	-0.4	20.8	11.8
	3D-ITD + HT + Polynomial	1.2	1.0	15.7	10.2
	3D-ITD + HT + Richards	0.7	0.6	14.9	9.3
fit, no RE	2D-ITD + Boolean + Polynomial	11.7	0.2	14.2	-1.7
	3D-ITD + Boolean + Polynomial	4.6	-0.6	23.7	12.3
	3D-ITD + HT + Polynomial	5.4	0.8	22.1	11.7
	3D-ITD + HT + Richards	4.1	-0.2	12.1	7.1
LOO	2D-ITD + Boolean + Polynomial	15.6	0.3	15.0	-1.6
	3D-ITD + Boolean + Polynomial	4.9	-0.5	24.0	12.5
	3D-ITD + HT + Polynomial	9.8	2.1	26.2	14.1
	3D-ITD + HT + Richards	5.4	-0.5	13.4	6.1

Table 5: Normalized root mean square errors and means of errors for quadratic mean diameter and basal area for several model fittings and predictions in the 18 plots with $> 95\%$ pine. The column "case" specifies if the results are calculated in the modeling data with the predicted values of random effects (fit, RE), or if the random effects have been explained by covariates derived from ALS or ITD (fit, no RE), or if the results come from leave-one-out cross-validation (LOO).

+ Polynomial and 3D-ITD + HT + Polynomial, too. The explanation may be the large mean error of stand density among pure pine plots (Table 2).

When adding covariates to the models, 3D-ITD + HT + Richards attains the lowest values for RMSE of QMD and BA, although the benchmark has the best ME values (Table 5, rows 5-8). In leave-one-out cross-validation 3D-ITD + HT + Richards has the second lowest RMSE of QMD, 3D-ITD + Boolean + Polynomial having the lowest, and the lowest RMSE of BA (Table 5, rows 9-12). In this case all of the modified models produce RMSE values of QMD lower than the benchmark, but only the model with the Richards' curve achieves lower RMSE of BA than the benchmark.

When the predicted values of random effects are used, 3D-ITD + HT +

case	model	mean	sd	min	max
fit, RE	2D-ITD + Boolean + Polynomial	0.27	0.15	0.12	0.61
	3D-ITD + Boolean + Polynomial	0.26	0.14	0.10	0.61
	3D-ITD + HT + Polynomial	0.32	0.12	0.18	0.51
	3D-ITD + HT + Richards	0.20	0.06	0.11	0.34
fit, no RE	2D-ITD + Boolean + Polynomial	0.52	0.24	0.21	1.10
	3D-ITD + Boolean + Polynomial	0.38	0.17	0.14	0.74
	3D-ITD + HT + Polynomial	0.44	0.16	0.25	0.77
	3D-ITD + HT + Richards	0.37	0.13	0.21	0.61
LOO	2D-ITD + Boolean + Polynomial	0.62	0.32	0.23	1.22
	3D-ITD + Boolean + Polynomial	0.39	0.18	0.15	0.76
	3D-ITD + HT + Polynomial	0.50	0.17	0.26	0.85
	3D-ITD + HT + Richards	0.42	0.17	0.24	0.80

Table 6: Mean, standard deviation, minimum and maximum of L^2 distances between true and estimated cumulative distribution functions in the 18 plots with $> 95\%$ pine.The column "case" specifies if the results are calculated in the modeling data with the predicted values of random effects (fit, RE), or if the random effects have been explained by covariates derived from ALS or ITD (fit, no RE), or if the results come from leave-one-out cross-validation (LOO).

Richards exhibits the lowest mean, standard deviation and maximum value of L^2 distances (Table 6, rows 1-4). When covariates are added to the models and the predicted values of random effects are not used, the situation is the same (Table 6, rows 5-8). In leave-one-out cross-validation, all of the modified models achieve better statistics for L^2 distances than the benchmark, 3D-ITD + Boolean + Polynomial achieving best values (Table 6, rows 9-12).

The chosen covariates for the benchmark and the two 3D-ITD + Polynomial models were the same as with the full data, the 5th ITD height quantile and the variance of ITD heights, respectively. For 3D-ITD + HT + Richards, the best covariates for K, Q and B were variance of the ITD heights, the 95th proportional density and 5th proportional density, respectively. The leave-one-

out cross-validation chose the same covariates as above for the benchmark and 3D-ITD + Boolean + Polynomial every time. The other methods had more variation in the chosen covariates.

5. Discussion

We have presented a methodology of matching crown radii distributions extracted from airborne laser scanning data through individual tree detection to distributions of diameters at breast height. The methodology is based on distribution matching, as described in Vauhkonen and Mehtätalo (2015). Unlike previously, no distributional assumptions on tree diameters or crown radii were made, and a new nonlinear monotonic transformation was used. A new ITD algorithm of Lähivaara et al. (2014) and correction method of Kansanen et al. (2016) were used.

The methodological choices adopted here generally improved the distribution matching compared to the benchmark (2D-ITD + Boolean + Polynomial; Vauhkonen and Mehtätalo, 2015). Reduced RMSEs for QMD and BA were achieved in almost all of the tested cases with a modified model. The benchmark did, however, achieve ME values closer to zero than the other models in 4 out of the 6 test cases for both QMD and BA. High ME values were obtained especially when using the Richards' curve with random effects explained by ALS covariates, whereas the polynomial matching function always yielded either a better or not markedly worse result than the benchmark method.

Changing the 2D-ITD algorithm to 3D-ITD leads to clear improvements to the performance. Even if the correction and matching methods were not changed, i.e. using 3D-ITD + Boolean + Polynomial, the lowest RMSE of QMD was obtained in 3 cases, two of which are the leave-one-out cross-validation experiments, and lowest RMSE of BA in the leave-one-out cross-validation experiment with the full data. The method also produces lowest mean of L^2 distances in both leave-one-out cases, indicating best performance in predicting DBH distributions, on average.

Changing the correction method based on Boolean model to the HT corresponds to replacing the Weibull distributions of crown radii and DBH with nonparametric distributions and the correction method of Mehtätalo (2006) with the Horvitz-Thompson type correction of Kansanen et al. (2016). The benefit of this choice can be assessed by comparing the performance of 3D-ITD + Boolean + Polynomial with 3D-ITD + HT + Polynomial. In most cases, the change results in lower RMSE of BA. Exceptions are the cross-validation experiments. With full data the errors between the methods are very close, but in the pine plots, 3D-ITD + HT + Polynomial almost doubled the RMSE of 3D-ITD + Boolean + Polynomial in the cross-validation experiment.

Distribution matching using the predicted random effects indicates that the Richards' function is able to describe the variability in the transformations from crown radius to tree diameter. However, this variability is not well explained by the covariates. Especially, in the cross-validation in all data the other models beat the most modified method 3D-ITD + HT + Richards. The simpler models are more robust, signified by the same covariates being chosen every time, whereas the Richards' curve is more sensitive, and the chosen covariates do not accurately represent the transformation in the target plot. The Richards' model might also be overfitted in this data and a larger data set would produce better results. However, the data of all 36 plots is quite heterogeneous, and the performance of 3D-ITD + HT + Richards was the best when the cross-validation was restricted to the more homogeneous pure pine plot data. Better results could be explained by smaller variability in stand density, quadratic mean diameter and basal area (see Table 1), change in the accuracy of the stand density estimators (see Table 2), or the more homogeneous forest structure.

The higher ME% values for BA in the pure pine plot data mostly result from higher ME% values for stand density in this data. Especially, most of the pure pine plots have regular spatial pattern, whereas the whole data includes more random and clustered plots (Fig. 4). The lower RMSE values of the stand density estimator related to the new methodology also result in good RMSE values for BA. Curiously, although the benchmark model has the highest RMSE

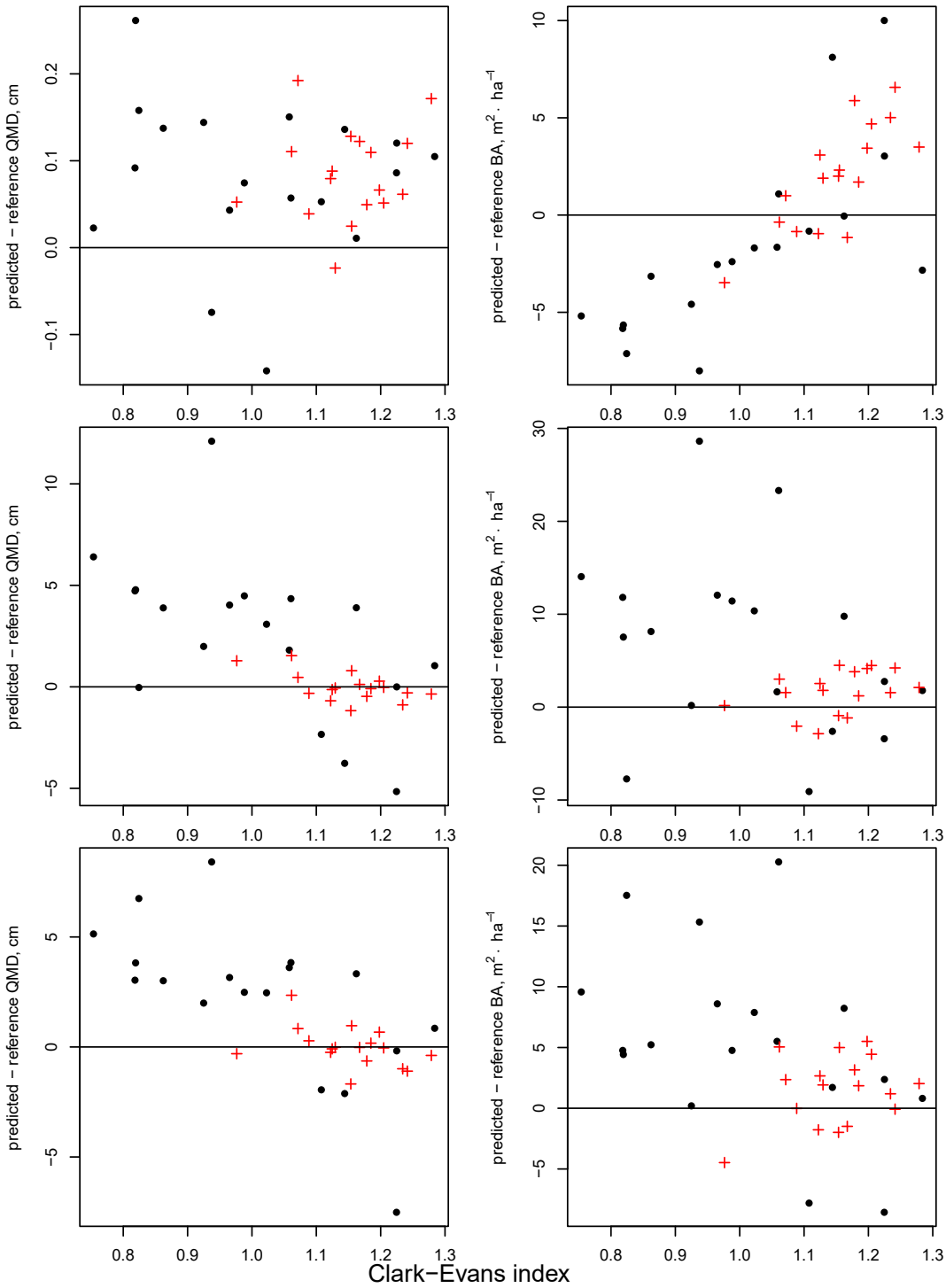

Figure 4: The estimation errors of quadratic mean diameter and basal area for 3D-ITD + HT + Richards as functions of the Clark-Evans aggregation index. Top panels: fits with the predicted random effects. Middle panels: fits with the random effects explained by covariates derived from ALS and ITD. Bottom panels: the leave-one-out cross-validation fits. Pure pine plots represented by +.

for QMD and the stand density estimator connected to it has the highest RMSE, it attains RMSE of BA very close to that of 3D-ITD + HT + Richards. The errors somehow cancel each other out.

Fig. 4 indicates that the underlying spatial distribution of trees, measured through the Clark-Evans index, does not have an impact on the estimation of QMD when 3D-ITD + HT + Richards is used with the predicted random effects. The estimation errors of BA with the same model do exhibit more underestimation in the clustered plots and overestimation in the regular plots, which can be attributed to the similar behaviour of the stand density estimator (see Kansanen et al., 2016). However, when the random effects have been explained with ALS and ITD covariates QMD is underestimated in the clustered plots and overestimated in the regular plots. This combined with the opposite nature of the stand density estimator might explain why the estimation errors of BA are not impacted by the value of Clark-Evans index.

In several cases, the covariates based on ITD detected tree heights were chosen as the best covariates in all models, especially the benchmark methods. It would seem that utilizing the height information is useful for distribution matching. Maltamo et al. (2018) tested distribution matching from tree height to diameter distribution in a pulpwood plantation, where tree planting pattern was known and no compensation for undetected trees was needed. However, for semi-natural forests that method would require a correction method for the tree height distribution based on the ITD detected tree heights. Thus, using the corrected tree height distribution to predict the stem diameter distribution instead of the crown radii distribution would be the next step, and possibly more fruitful, as there are several models connecting the tree heights to stem diameters.

Compared to earlier studies (Vauhkonen and Mehtätalo, 2015; Maltamo et al., 2018), this study produced important findings with respect to applying distribution matching in practice. According to our results, the method does not need to be restricted to forests with known spatial pattern or species. For example, applying the method only in forests that met all the assumptions stated by

Vauhkonen and Mehtätalo (2015) would have narrowed its application down to a small number of forests. After the modifications introduced above, the method was not specifically sensitive to the properties of the target forests, for which reason it can be potentially applied in a wall-to-wall manner for entire inventory areas similar to other methods. As discussed earlier (Vauhkonen and Mehtätalo, 2015; Maltamo et al., 2018), distribution matching can complement both ITD and area-based methods in diameter distribution predictions: especially, due to employing ALS-observed distributions as a priori information, the further matching with the diameter distribution can potentially be based on a lesser number of field measurements than with alternative methods. On the other hand, some data are needed to calibrate the ITD method and fit the matching function. The requirements for these data should be more carefully studied in the future, and for now, the results presented above apply to cases where local training data are available from forest plots that are highly similar to target forests of predictions.

6. Conclusions

It is possible to improve the results of diameter distribution estimation methodology of Vauhkonen and Mehtätalo (2015) while abandoning distributional assumptions. Especially, the use of improved ITD algorithm (Lähivaara et al., 2014), nonparametric distributions, and Horvitz-Thompson type correction (Kansanen et al., 2016) improve the results. Nonlinear transformation via a Richards' curve is flexible enough for diameter distribution estimation due to the good fitting results when the random effects are included in the model. It is useful for prediction when the population from which the field plots have been sampled is homogeneous. In this case the random effects can be modeled by using statistics derived from ALS return heights and ITD as covariates relatively well. In the case where the field plots are not homogeneous, a simpler quadratic transformation can still produce good results when compared to the benchmark method. To avoid bias in a real prediction situation, where also random effects need to be

modeled, the safest choice is to use the polynomial transformation with 3D-ITD and either of the tested correction types for the undetected trees.

The newly formulated model with the Richards' curve provided the best predictions in the situation where the random effects of the Richards' model were known. This result shows that the model we formulated describes the modeled process of non-detection and crown diameter – tree-diameter relationship well. Unfortunately, the variation in the crown diameter – stem diameter relationship was not very well explained by the ALS and ITD covariates in the rather heterogeneous full data set. This may, however, partially result from overfitting as the number of plots is rather limited compared to the number of model parameters, and the best model varied quite a lot among the cross-validation replicates. The results were, however, promising when the analysis was restricted to pure pine plots. A larger dataset should be used to further validate the method in the future.

7. Acknowledgements

This study was funded by Research Funds of the University of Helsinki, Academy of Finland (projects 295489, 250215 and 310073), the strategic funding of the University of Eastern Finland, and Finnish Cultural Foundation, North Karelia Regional fund.

Appendix A

The cumulative distribution function of the Weibull distribution is

$$F(r) = \begin{cases} 1 - \exp\left[-\left(\frac{r}{\gamma}\right)^k\right], & r \geq 0 \\ 0, & r < 0 \end{cases},$$

where γ is the scale and k the shape parameter. In the following, let us assume that these are the parameters that have been estimated for the distribution of crown radii. Let us write the estimated quadratic transformation for

simplicity as $g(r) = \beta_1 r + \beta_2 r^2$. To formulate $F_{DBH}(d)$, one has to consider the probability $P\{DBH \leq d\} = P\{g(R) \leq d\}$ and solve the inequality $g(r) \leq d$ to produce probabilities regarding R, hence performing a change of variable in the cumulative distribution function of R. This inequality has differing solutions dependent on the values of β_1 and β_2. Let us write

$$D_- = \frac{-\beta_1 - \sqrt{\beta_1^2 + 4\beta_2 d}}{2\beta_2 \gamma}$$

and

$$D_+ = \frac{-\beta_1 + \sqrt{\beta_1^2 + 4\beta_2 d}}{2\beta_2 \gamma}.$$

659 If $\beta_1 > 0$ and $\beta_2 = 0$,

$$F_{DBH}(d) = 1 - \exp\left[-\left(\frac{d}{\gamma\beta_1}\right)^k\right].$$

660 If $\beta_1 \geq 0$ and $\beta_2 > 0$,

$$F_{DBH}(d) = 1 - \exp\left[-D_+^k\right].$$

661 If $\beta_1 < 0$ and $\beta_2 > 0$,

$$F_{DBH}(d) = \begin{cases} 0, & d \leq -\frac{\beta_1^2}{4\beta_2} \\ \exp\left[-D_-^k\right] - \exp\left[-D_+^k\right], & -\frac{\beta_1^2}{4\beta_2} < d \leq 0 \\ 1 - \exp\left[-D_+^k\right], & d > 0 \end{cases}.$$

662 If $\beta_1 > 0$ and $\beta_2 < 0$,

$$F_{DBH}(d) = \begin{cases} \exp\left[-D_-^k\right], & d \leq 0 \\ 1 + \exp\left[-D_-^k\right] - \exp\left[-D_+^k\right], & 0 < d < -\frac{\beta_1^2}{4\beta_2} \\ 1, & d \geq -\frac{\beta_1^2}{4\beta_2} \end{cases}.$$

References

References

Arias-Rodil, M., Diéguez-Aranda, U., Álvarez-González, J.G., Pérez-Cruzado, C., Castedo-Dorado, F., González-Ferreiro, E., 2018. Modeling diameter distributions in radiata pine plantations in spain with existing countrywide lidar data. Annals of Forest Science 75, 36. URL: https://doi.org/10.1007/s13595-018-0712-z, doi:10.1007/s13595-018-0712-z.

Chiu, S.N., Stoyan, D., Kendall, W.S., Mecke, J., 2013. Stochastic Geometry and Its Applications. 3rd ed., Wiley, New York.

Clark, P.J., Evans, F.C., 1954. Distance to nearest neighbor as a measure of spatial relationships in populations. Ecology 35, 445–453. doi:10.2307/1931034.

Donnelly, K., 1978. Simulations to determine the variance and edge-effect of total nearest neighbour distance, in: Hodder, I. (Ed.), Simulation studies in archaeology. Cambridge University Press, pp. 91–95.

Duncanson, L., Cook, B., Hurtt, G., Dubayah, R., 2014. An efficient, multi-layered crown delineation algorithm for mapping individual tree structure across multiple ecosystems. Remote Sensing of Environment 154, 378 – 386. URL: http://www.sciencedirect.com/science/article/pii/S0034425714000984, doi:https://doi.org/10.1016/j.rse.2013.07.044.

Fekedulegn, D., Mac Siurtain, M.P., Colbert, J.J., 1999. Parameter estimation of nonlinear growth models in forestry. Silva Fennica 33, 327–336.

Gobakken, T., Næsset, E., 2004. Estimation of diameter and basal area distributions in coniferous forest by means of airborne laser scanner data. Scandinavian Journal of Forest Research 19, 529–542. URL: http://dx.doi.org/10.1080/02827580410019454, doi:10.1080/02827580410019454, arXiv:http://dx.doi.org/10.1080/02827580410019454.

Gomes, M.F., Maillard, P., Deng, H., 2018. Individual tree crown detection in sub-meter satellite imagery using marked point processes and a geometrical-optical model. Remote Sensing of Environment 211, 184 – 195. URL: http://www.sciencedirect.com/science/article/pii/S0034425718301470, doi:https://doi.org/10.1016/j.rse.2018.04.002.

Gonzalez, R., Woods, R.E., 2008. Digital image processing. 3rd ed., Prentice Hall, U.S.A.

Harrell Jr, F.E., with contributions from Charles Dupont, many others., 2016. Hmisc: Harrell Miscellaneous. URL: https://CRAN.R-project.org/package=Hmisc. r package version 3.17-4.

Kaipio, J., Somersalo, E., 2005. Statistical and Computational Inverse Problems. Springer-Verlag.

Kansanen, K., Vauhkonen, J., Lähivaara, T., Mehtätalo, L., 2016. Stand density estimators based on individual tree detection and stochastic geometry. Canadian Journal of Forest Research 46, 1359–1366. URL: http://dx.doi.org/10.1139/cjfr-2016-0181, doi:10.1139/cjfr-2016-0181, arXiv:http://dx.doi.org/10.1139/cjfr-2016-0181.

Korhonen, L., Peuhkurinen, J., Malinen, J., Suvanto, A., Maltamo, M., Packalen, P., Kangas, J., 2008. The use of airborne laser scanning to estimate sawlog volumes. Forestry 81, 499–510. URL: http://forestry.oxfordjournals.org/content/81/4/499.abstract, doi:10.1093/forestry/cpn018, arXiv:http://forestry.oxfordjournals.org/content/81/4/499.full

Lähivaara, T., Seppänen, A., Kaipio, J., Vauhkonen, J., Korhonen, L., Tokola, T., Maltamo, M., 2014. Bayesian approach to tree detection based on airborne laser scanning data. IEEE Transactions on Geoscience and Remote Sensing 52, 2690–2699.

Lamb, S.M., MacLean, D.A., Hennigar, C.R., Pitt, D.G., 2017. Imputing tree lists for new brunswick spruce plantations through nearest-neighbor matching of airborne laser scan and inventory plot data. Canadian Journal of Remote Sensing 43, 269–285. URL: https://doi.org/10.1080/07038992.2017.1324288, doi:10.1080/07038992.2017.1324288, arXiv:https://doi.org/10.1080/07038992.2017.1324288.

Li, W., Guo, Q., Jakubowski, M.K., Kelly, M., 2012. A new method for segmenting individual trees from the lidar point cloud. Photogrammetric Engineering & Remote Sensing 78, 75–84. URL: https://www.ingentaconnect.com/content/asprs/pers/2012/00000078/00000001/art00006, doi:doi:10.14358/PERS.78.1.75.

Lindberg, E., Eysn, L., Hollaus, M., Holmgren, J., Pfeifer, N., 2014. Delineation of tree crowns and tree species classification from full-waveform airborne laser scanning data using 3-d ellipsoidal clustering. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7, 3174–3181. doi:10.1109/JSTARS.2014.2331276.

Lindstrom, M.J., Bates, D.M., 1990. Nonlinear mixed effects models for repeated measures data. Biometrics 46, 673–687.

Lu, X., Guo, Q., Li, W., Flanagan, J., 2014. A bottom-up approach to segment individual deciduous trees using leaf-off lidar point cloud data. ISPRS Journal of Photogrammetry and Remote Sensing 94, 1 – 12. URL: http://www.sciencedirect.com/science/article/pii/S0924271614000860, doi:https://doi.org/10.1016/j.isprsjprs.2014.03.014.

Magnussen, S., Renaud, J.P., 2016. Multidimensional scaling of first-return airborne laser echoes for prediction and model-assisted estimation of a distribution of tree stem diameters. Annals of Forest Science 73, 1089–1098. URL: http://dx.doi.org/10.1007/s13595-016-0581-2, doi:10.1007/s13595-016-0581-2.

Maltamo, M., Mehtätalo, L., Valbuena, R., Vauhkonen, J., Packalen, P., 2018. Airborne laser scanning for tree diameter distribution modelling: a comparison of different modelling alternatives in a tropical single-species plantation. Forestry: An International Journal of Forest Research 91, 121–131. URL: http://dx.doi.org/10.1093/forestry/cpx041, doi:10.1093/forestry/cpx041.

Maltamo, M., Packalen, P., 2014. Species-specific management inventory in finland, in: Maltamo, M., Næsset, E., Vauhkonen, J. (Eds.), Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies, Springer Netherlands, Dordrecht. pp. 241–252. URL: https://doi.org/10.1007/978-94-017-8663-8_12, doi:10.1007/978-94-017-8663-8_12.

Mehtätalo, L., 2006. Eliminating the effect of overlapping crowns from aerial inventory estimates. Canadian Journal of Forest Research 36, 1649–1660.

Mehtätalo, L., Maltamo, M., Packalén, P., 2007. Recovering plot-specific diameter distribution and height-diameter curve using als-based stand characteristics. IAPRSS 36/ Part 3/W52, 288–293.

Muinonen, E., 1995. Metsikön heijastussuhteen ennustaminen geometrisella latvustomallilla. Licenciate thesis, University of Joensuu. In Finnish.

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

Packalén, P., Vauhkonen, J., Kallio, E., Peuhkurinen, J., Pitkänen, J., Pippuri, I., Strunk, J., Maltamo, M., 2013. Predicting the spatial pattern of trees with airborne laser scanning. International Journal of Remote Sensing 34, 5154–5165.

Persson, A., Holmgren, J., Söderman, U., 2002. Detecting and measuring indi-

vidual trees using airborne laser scanner. Photogrammetric Engireening and Remote Sensing 68, 925–932.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team, 2016. nlme: Linear and Nonlinear Mixed Effects Models. URL: http://CRAN.R-project.org/package=nlme. r package version 3.1-128.

Pitkänen, J., 2005. A multi-scale method for segmentation of trees in aerial images, in: Hobbelstad, K. (Ed.), Proceedings of the SNS Meeting at Sjusjøen — Forest Inventory and Planning in Nordic Countries, Norway, 6–8 September 2004, Norwegian Institute of Land Inventory. pp. 207–216.

Pitkänen, J., Maltamo, M., Hyyppä, J., Yu, X., 2004. Adaptive methods for individual tree detection on airborne laser based canopy height model, in: Theis, M., Koch, B., Spiecker, H., Weinacker, H. (Eds.), Proceedings of ISPRS working group VIII/2: Laser-Scanners for Forest and Landscape Assessment, University of Freiburg, Freiburg, Germany. pp. 187–191.

Rana, P., Vauhkonen, J., Junttila, V., Hou, Z., Gautam, B., Cawkwell, F., Tokola, T., 2017. Large tree diameter distribution modelling using sparse airborne laser scanning data in a subtropical forest in nepal. ISPRS Journal of Photogrammetry and Remote Sensing 134, 86 – 95. URL: http://www.sciencedirect.com/science/article/pii/S0924271617303386, doi:https://doi.org/10.1016/j.isprsjprs.2017.10.018.

Reitberger, J., Schnörr, C., Krzystek, P., Stilla, U., 2009. 3d segmentation of single trees exploiting full waveform lidar data. ISPRS Journal of Photogrammetry and Remote Sensing 64, 561 – 574. URL: http://www.sciencedirect.com/science/article/pii/S0924271609000495, doi:https://doi.org/10.1016/j.isprsjprs.2009.04.002.

Rudin, W., 1987. Real and Complex Analysis, 3rd Ed. McGraw-Hill, Inc., New York, NY, USA.

Shang, C., Treitz, P., Caspersen, J., Jones, T., 2017. Estimating stem diameter distributions in a management context for a tolerant hardwood forest using als height and intensity data. Canadian Journal of Remote Sensing 43, 79–94. URL: http://dx.doi.org/10.1080/07038992.2017.1263152, doi:10.1080/07038992.2017.1263152, arXiv:http://dx.doi.org/10.1080/07038992.2017.1263152.

Spriggs, R.A., Coomes, D.A., Jones, T.A., Caspersen, J.P., Vanderwel, M.C., 2017. An alternative approach to using lidar remote sensing data to predict stem diameter distributions across a temperate forest landscape. Remote Sensing 9. URL: http://www.mdpi.com/2072-4292/9/9/944, doi:10.3390/rs9090944.

Thomas, V., Oliver, R.D., Lim, K., Woods, M., 2008. Lidar and weibull modeling of diameter and basal area. The Forestry Chronicle 84, 866–875. URL: http://dx.doi.org/10.5558/tfc84866-6, doi:10.5558/tfc84866-6, arXiv:http://dx.doi.org/10.5558/tfc84866-6.

Vauhkonen, J., Mehtätalo, L., 2015. Matching remotely sensed and field measured tree size distributions. Canadian Journal of Forest Research 45, 353–363.

Vega, C., Hamrouni, A., Mokhtari, S.E., Morel, J., Bock, J., Renaud, J.P., Bouvier, M., Durrieu, S., 2014. Ptrees: A point-based approach to forest tree extraction from lidar data. International Journal of Applied Earth Observation and Geoinformation 33, 98 – 108. URL: http://www.sciencedirect.com/science/article/pii/S0303243414001135, doi:https://doi.org/10.1016/j.jag.2014.05.001.

Zhou, J., Proisy, C., Descombes, X., Le Maire, G., Nouvellon, Y., Stape, J.L., Viennois, G., Zerubia, J., Couteron, P., 2013. Mapping local density of young eucalyptus plantations by individual tree detection in high spatial resolution satellite images. Forest ecology and management 301, 129–141.