Forestry *An International Journal of Forest Research*

Institute of Chartered Foresters

*Forestry* 2018; **91**, 121–131, doi:10.1093/forestry/cpx041
Advance Access publication 22 September 2017

# Airborne laser scanning for tree diameter distribution modelling: a comparison of different modelling alternatives in a tropical single-species plantation

**M. Maltamo[1]*, L. Mehtätalo[2], R. Valbuena[3], J. Vauhkonen[4] and P. Packalen[1]**

[1]*University of Eastern Finland, Faculty of Science and Forestry, School of Forest Sciences, PO Box 111, FI-80101 Joensuu, Finland*
[2]*University of Eastern Finland, Faculty of Science and Forestry, School of Computing, PO Box 111, FI-80101 Joensuu, Finland*
[3]*University of Cambridge, Department of Plant Sciences, Forest Ecology and Conservation, Downing Street, CB2 3EA Cambridge, UK*
[4]*Natural Resources Institute Finland (LUKE), Economics and Society, Yliopistokatu 6 (PO Box 68), FI-80101 Joensuu, Finland*

*Corresponding author. E-mail: matti.maltamo@uef.fi

*Received 23 February 2017*

This study examines the potential of airborne laser scanning (ALS) to predict diameter distributions in an even-aged plantation of *Eucalyptus urograndis* in Brazil. The single-species plantation conditions allow different modelling alternatives to be compared without the presence of minor tree species or an understory layer affecting the results. Three modelling alternatives based on the two-parametric Weibull function form; parameter prediction, parameter recovery and distribution matching were tested with a k-nearest neighbour prediction (*k-nn*) method. We also evaluated a parameter prediction alternative, in which the Weibull distribution was predicted using field attributes. The results showed that ALS information can predict diameter distributions with an error margin of slightly more than 10 per cent of the RMSE of the mean of the third power of diameter, and with error index values between 50 and 60. The degree of accuracy was only slightly improved when the Weibull distribution was predicted using field attributes. According to the accuracy metrics, the differences between modelling alternatives were minor but parameter recovery and *k-nn* seem to be the most favourable ALS-based prediction methods. To conclude, the results showed a strong relationship between ALS information and diameter distributions in a tropical single-species plantation and we discuss how these results could be applied in other types of forests.

## Introduction

Diameter at breast height (d.b.h.) is an important tree attribute and is commonly measured in forest inventories, because it can be used to calculate a range of other attributes that cannot be directly measured, such as tree trunk volume and above-ground biomass. The distribution of d.b.h. is a stand-level indicator of forest structure and information on diameter distribution also aids in the calculation of timber assortments. In mixed forest stands, the distribution of d.b.h. is usually computed by tree species. Although d.b.h. can be measured easily and accurately, a full diameter distribution is not usually measured in stand-level management inventories due to the high costs involved. However, field assessed stand attributes can be used as predictors of theoretical distributions, such as the Weibull function (Bailey and Dell, 1973), or the k-nearest neighbour (*k-nn*) prediction (also known as non-parametric tree lists, Maltamo and Kangas, 1998) to estimate the diameter distribution. The two-parameter form of the Weibull distribution has generally been preferred for this task (Gobakken and Næsset, 2005; Siipilehto and Mehtätalo, 2013; Saad *et al.*, 2015). The parameters of the selected distribution, given the known stand attributes, can be estimated using the parameter recovery (Hyink and Moser, 1983) and parameter prediction methods (Rennolls *et al.*, 1985). As parameter recovery is based on the analytical relationships between the distribution parameters and the stand attributes, the number of distribution-related stand attributes, therefore, needs to be equal to the number of parameters in the selected distribution function (Siipilehto and Mehtätalo, 2013). The attributes used in recovery can be moments (mean and quadratic mean) or percentiles (e.g. median) of the distribution. When the equations between the distribution parameters and the stand attributes are solved analytically, they result in parameter estimates that are mathematically compatible with the attributes used in the recovery process. In parameter prediction, the estimated parameters are predicted by regression analysis with stand attributes used as independent variables. In general, parameter recovery is preferred to parameter prediction because of (a) compatible parameter estimates and (b) the analytical relationships between moments (or percentiles) and parameters are stronger than those seen in regression models used for parameter prediction (Hyink and Moser, 1983; Siipilehto and Mehtätalo, 2013).

The emergence of airborne laser scanning (ALS) techniques has had a major impact on stand-level management inventories (White *et al.*, 2013; Vauhkonen *et al.*, 2014). In the Nordic countries, compartment-wise field visits and the visual interpretation of aerial data have been replaced by an area-based approach (ABA) using wall-to-wall coverage of ALS and aerial images and detailed measurements from a limited number of sample plots (Maltamo and Packalen, 2014; Næsset, 2014). Despite these changes, diameter distribution remains one of the most important stand characteristics that cannot be directly produced from ALS measurements and must be predicted independently (Maltamo and Gobakken, 2014). Prediction of diameter distributions by ALS usually consists of a prediction or recovery of the distribution parameters using ALS information instead of stand-level field attributes. Alternatively, it is also possible to predict stand attributes from ALS data using the ABA and to employ them in the prediction of parameter estimates of diameter distributions by applying existing parameter models (Holopainen *et al.*, 2010; Maltamo and Gobakken, 2014).

Earlier studies (e.g. Gobakken and Næsset, 2004, 2005; Maltamo *et al.*, 2006; Thomas *et al.*, 2008; Saad *et al.*, 2015) predicted the Weibull function with ALS using either parameter prediction or recovery. Mehtätalo *et al.* (2007) recovered the parameters of diameter distributions and height–diameter curves simultaneously. Moreover, ALS-based diameter distributions have also been estimated using *k-nn* prediction and Random forest (Packalén and Maltamo, 2008; Shang *et al.*, 2017). In all these papers, tree lists (i.e. tree frequencies by diameter classes) are predicted using field sample plot data.

Several studies have further refined the method, for example, by incorporating multimodal distribution predictions (Bollandsås and Næsset, 2007; Packalén and Maltamo, 2008; Thomas *et al.*, 2008; Magnussen *et al.*, 2013; Magnussen and Renaud, 2016). However, the success of the method development has been varied and problems related to tree species still hamper the prediction of diameter distributions based on ALS. Since the ABA cannot fundamentally separate species, estimations of stand-level d.b.h. distributions are usually carried out for all species together, rather than predicting separate distributions of each species. Yet, species-specific information is needed and such predictions cannot be accurately provided for less abundant tree species, even when optical sensors are involved in the estimation process (Packalén and Maltamo, 2008). In addition, ordinary theoretical probability density functions, such as the Weibull function, can only describe unimodal distributions that occur in managed, even-aged, single-species stands, and a small proportion of minor tree species can cause irregularities or peaks in the resulting distribution. In such cases, the mathematical properties of the Weibull distribution are not met without considering the species separately. Although many of the previous studies have considered the total diameter distributions of multiple species, the potential of ALS data to predict unimodal single-species diameter distributions still requires further research. We hypothesize that the modelling and comparison of ALS-based distribution alternatives in single-species stands could provide valuable information in regard to the different alternatives, which are also applied in more heterogeneous stand structures.

In recent times, fast growing tropical single-species plantations have rapidly attained a prominent role in the forest industry. In both sawnwood and pulpwood plantations, diameter distribution is one of the most important stand attributes (Saramäki, 1992; Nanang, 1998; Mabvurira *et al.*, 2002). Unlike sampling-based field inventories, the application of ALS data provides an opportunity to produce maps of plantation resources. However, to the best of our knowledge, there have been no ALS-based diameter distribution prediction studies in tropical plantations, although other stand attributes, such as volume (Rombouts *et al.*, 2010), dominant height (Packalén *et al.*, 2011b), stand density (Tesfamichael *et al.*, 2009) and site quality (Rombouts *et al.*, 2010) have been considered. The prediction of diameter distributions is more challenging compared to single attributes, such as volume or height, and given their practical importance, there is a need to evaluate their accuracy when they are used with ALS for tropical plantations.

This study examines the potential of ALS to predict diameter distributions in an even-aged *Eucalyptus urograndis* plantation, where the relationship between diameter distribution and ALS information can be analysed without the influence of minor tree species or an understory layer. As modelling alternatives, we compare parameter prediction, parameter recovery and distribution matching, which were based on the Weibull function form. In addition, a non-parametric *k-nn* prediction and a parameter prediction alternative, in which the Weibull distribution is predicted using field attributes available from the plantation stand register database, are also included in the comparison. We evaluate the implementation of different alternatives for ALS-based diameter distribution modelling in tropical plantations and attempt to derive useful conclusions that are valid for modelling diameter distributions in semi-natural forest conditions.

## Materials

### Study area and field data

The study was performed using data for a *E. urograndis* plantation located in Bahia state, Brazil (16°05′S 39°24′W). *Eucalyptus urograndis* is a hybrid between *E. grandis* W. Hill ex Maiden and *E. urophylla* S. T. Blake. In total, 195 circular sample plots (radius 13 m) were established and measured in August and September 2008. Three or four sample plots (depending on the stand size) were placed in 55 randomly chosen forest stands. There were 28 different clones in the 55 stands from which the field data was collected. All of the trees in a stand belonged to the same clone.

Satellite positioning (global positioning systems device: Trimble GPS Pathfinder Pro XRS) was used to determine the position of each plot centre using a real-time differential correction signal obtained from the OmniSTAR satellite (http://www.omnistar.com). The spatial accuracy was not explicitly assessed, but it was assumed to lie in the range of ~1–2 m reported as a standard for the Trimble GPS Pathfinder Pro XRS device. Tree density was fixed to 833 stems per hectare, although damage may have reduced the actual stem density. The d.b.h. of all trees in the plots were recorded and this information was used to calculate the actual number of trees per hectare ($N$, ha$^{-1}$), basal area per hectare ($G$, m$^2$ ha$^{-1}$), and mean d.b.h. ($D$, cm).

The seventh tree in each plot was measured for height ($h$, m). Näslund's (1937) $h$–d.b.h. curve was fitted as a nonlinear mixed-effects model and used to predict heights for the trees without a height measurement. Finally, the dominant height (HD, m) was calculated as the mean height of the 100 thickest trees per hectare, in terms of their d.b.h. Stand age ($T$, years) was obtained from the plantation register database, in which it was recorded with an interval of 1 month. The site index (SI) was predicted using the following form of the Chapman–Richards equation (Clutter *et al.*, 1983):

$$\mathrm{SI} = \mathrm{HD}\left(\frac{1 - \mathrm{e}^{-\beta_1 t_{\mathrm{reference}}}}{1 - \mathrm{e}^{-\beta_1 t_{\mathrm{current}}}}\right)^{\beta_2}, \tag{1}$$

where $t_{\mathrm{reference}}$ is the reference age of 7 years, HD is the current dominant height, $t_{\mathrm{current}}$ is the current age, and $\beta_1$ (0.3341) and $\beta_2$ (1.1442) are known model parameters in the plantation. The main stand characteristics are presented in Table 1.

### Airborne laser scanning data

The ALS data were collected on 16 August 2008, using an Optech ALTM 3100 laser scanning system. The test site was measured from an altitude of ~1200 m above ground level using a field of view of 30 degrees. Pulse repetition frequency was set at 50 000 pulses per second, which resulted in a nominal sampling density of ~1.5 measurements per square metre. The footprint was ~35 cm at ground level.

A digital terrain model (DTM) was generated from the ALS data. First, laser points were classified as ground and non-ground points using the method reported by Axelsson (2000). Then, a raster DTM with a 1 m pixel size was interpolated using ground points and an inverse distance weighting algorithm (Lloyd and Atkinson, 2002). Finally, the raster DTM was subtracted from the ellipsoidal heights of the laser points to normalize the ALS data to the above-ground level (AGL).

## Methods

### Explanatory variables

The laser scanner captured a maximum of four echoes per emitted pulse, categorized as 'first of many', 'last of many', 'only' and 'intermediate' echoes. After preliminary tests, we decided to use only the 'first of many' and 'only' echoes, because the exclusion of 'last of many' and 'intermediate' echoes did not significantly decrease the accuracy of the estimates (Packalén *et al.*, 2011b). This set contains all of the first

**Table 1** Mean, standard deviation (SD) and range of plot characteristics.

| Characteristic[1] | Mean | SD | Minimum | Maximum |
|---|---|---|---|---|
| $G$, m$^2$ha$^{-1}$ | 25.1 | 6.1 | 12.1 | 38.2 |
| $N$, ha$^{-1}$ | 812.9 | 51.7 | 602.7 | 979.4 |
| $D$, cm | 19.4 | 5.6 | 14.0 | 23.9 |
| SI, m | 34.3 | 1.8 | 29.2 | 39.2 |
| $T$, years | 7.3 | 2.6 | 2.5 | 11.9 |

[1] $G$ denotes basal area per hectare, $N$ denotes number of trees per hectare, $D$ denotes mean d.b.h., SI denotes site index and $T$ denotes stand age.

(i.e. surface) echoes, since an 'only' echo may also be considered as a first echo.

In total, 15 height and density metrics were calculated from the combined set of 'first of many' and 'only' echoes using the ABA method (Næsset, 2002). The first step was to calculate height distributions for each sample plot using the normalized ALS data. Height quantiles for 5, 10, 20, …, 80, 90, 95 per cent (*h5*, …, *h95*) of height sums were computed using all echoes, including those from the ground. The corresponding densities (*p5*, …, *p95*) were calculated for the respective quantiles, i.e. $p_{50}$ is the number of laser hits below $h_{50}$ divided by all of the echoes in the plot. In addition, the mean (*hmean*) and standard deviation (*hstd*) of heights were calculated. These metrics formed the set of candidate predictors used for modelling the diameter distributions.

### Methods based on the Weibull distribution

The two-parameter form of the Weibull function was applied to model the diameter distributions from the field reference data. The main advantages of the two-parameter form are the low number of parameters that need to be estimated and the flexibility of the form to describe different shapes of unimodal distributions (Bailey and Dell, 1973; Maltamo and Gobakken, 2014). The probability density function of the two-parameter Weibull distribution for a random variable $x$ (i.e. d.b.h. in our case) is (Dubey, 1967)

$$f(x) = \frac{c}{b}\left(\frac{x}{b}\right)^{c-1}\exp\left[-\left(\frac{x}{b}\right)^{c-1}\right], \quad x \geq 0; \quad b,\ c > 0 \tag{2}$$

where $b$ is a scale parameter and $c$ is a shape parameter. The Weibull parameters of equation (2) were estimated for each field plot by fitting the two-parameter Weibull distribution to the discrete ground reference diameter distributions with 1 cm precision by the maximum likelihood (ML) method (Harter and Moore, 1965). As an alternative, the Weibull parameters were also recovered using the measured stand attributes.

Three approaches were tested to relate the ALS data based predictions with the field reference distributions; parameter prediction (3.2.1), parameter recovery (3.2.2) and histogram matching (3.2.3). In the case of parameter prediction, a field information based alternative was also tested. The distributions were formed according to the original parameter values (ML estimate of equation (2) or the parameter recovery estimate based on actual stand attributes) or predicted parameter values, and transformed to histograms with 1 cm diameter classes for accuracy assessment. Original parameter values were also included to show the best possible diameter distribution fit that can be obtained using the Weibull distribution smoothing.

### Parameter prediction

The ML-estimates of the Weibull parameters were separately modelled using the observed field attributes (number of trees per hectare, $N$; site index, SI; stand age, $T$), and the ALS metrics. SI and $N$ were calculated from the field data, whereas $T$ and clone (used in mixed-effects, see below) were available from the plantation register database. The set of candidate predictor variables included the original values of variables, and the following transformations: the square root, natural logarithm, inverse, and second and third powers. Linear regression models were first constructed using the lm function in R (R Core Team, 2015) to select the independent variables. The function used the Akaike information criterion (AIC) statistic for stepwise selection between the models, allowing for both inclusion and exclusion of the independent variables in the candidate models.

The independent variables had to be statistically significant ($P < 0.05$) to be selected for the model. The number of ALS variables was restricted to two using the AIC. The constructed models were re-fitted using the lme function of the R statistical computing environment (R Core Team, 2015) as mixed-effects models including a random clone-level intercept.

### Parameter recovery

The distribution parameters were recovered by applying the stand attributes $G$, $N$ and $D$. The recovery was implemented separately using the observed and predicted values. In the case of predicted stand attributes, the predictions of the model described in the previous section were used. To illustrate the recovery, consider a forest stand where the density of diameter distribution is $f(x|b,c)$ and the stand density is $N$. The basal area for such a stand is:

$$G = \kappa N E(x^2) \tag{3}$$

where $\kappa = \frac{\pi}{40000}$ and the mean of squared diameter (also known as the quadratic mean diameter) is $E(x^2) = \int_0^\infty x^2 f(x)dx$. Solving this integral for $x$ in the case of the Weibull distribution gives $E(x^2) = b^2\,\Gamma(2/c + 1)$, where $\Gamma$ is the Gamma function. Applying this result to equation (3) gives the first recovery equation.

The second recovery equation matches the observed $D$ with the Weibull mean:

$$D = E(x) = \int_0^\infty x f(x)dx = b\Gamma(1/c + 1). \tag{4}$$

The recovery finds values of $b$ and $c$ that simultaneously fulfil the nonlinear equations (3) and (4). If a solution is found, it is unique. However, no solution is found if $N < G/(\kappa D^2)$, i.e., if $N$ is too small compared to $G$ and $D$. The stand density for a given $D$ and $G$ is minimized when all trees of the stand have diameter $D$, i.e., the diameter distribution is extremely narrow. In a small number of plots, where no solution was found, we replaced the estimated $N$ with the value $N^* = G/(\kappa D^2) - 10$. This replacement leads to a distribution of approximately equal diameters.

The solution to the system of equations is based on a profiling approach, where equation (4) is first solved for $b$ to get $b(c, D) = D/\Gamma(1/c + 1)$.

Applying this solution to equation (3) and rearranging the terms gives:

$$\kappa\frac{D^2}{[\Gamma(1/c + 1)]^2}\,\Gamma(2/c + 1) - \frac{G}{N} = 0. \tag{5}$$

Equation (5) was solved for the unknown parameter $c$ using the Gauss–Newton algorithm. Applying the obtained $c$ to $b(c, D)$ gives the corresponding estimate of $b$. The parameter recovery was implemented in the *recweib* function of the R-package *lmfor* (Mehtätalo, 2015).

### Distribution matching

In the third alternative, the aim was to extract the tree height distributions from the ALS data and to develop a matching function relating them to the d.b.h. distributions. This approach is a combination of two published methods, namely tree row detection (Vauhkonen *et al.*, 2011) and distribution matching (Vauhkonen and Mehtätalo, 2015). The tree height distributions are extracted in two steps, which rely on the availability of tree planting distance from the stand management records, in addition to the ALS data. An algorithm determines the orientation of the rows in which the trees were planted and samples the detected rows for height using a series of windows with sizes that correspond to the planting distance along the row. The full algorithm is described in detail by Vauhkonen *et al.* (2011). To derive the d.b.h. distributions, a quadratic function is fitted to model the percentile-wise transformation between the tree height and d.b.h. for distribution matching (Vauhkonen and Mehtätalo, 2015). Besides using the percentiles of the distributions, we noticed that the fit of the transformation function could be improved with stand-level parameters that localize the function to the actual stand. For this reason the transformation was modelled as

$$d_{pi} = \beta_1 + \beta_2 h90_p^{\beta_3} + \beta_4 h_{pi}^{\beta_5}, \tag{6}$$

where $d_{pi}$ and $h_{pi}$ are the $i$th percentiles ($i = 1, 2, \ldots, 99$) of the Weibull distributions of heights and d.b.h., respectively, *h90* is the 90th height quantile of plot $p$, and $\beta_1$, $\beta_2$, $\beta_3$, $\beta_4$ and $\beta_5$ are the model parameters. The $\beta$ parameters were estimated separately for each plot $p$ by fitting equation (6) to observations of $d_i$, $h_i$, and *h90* of all other plots in the same age class by using the *nls* function of R (R Core Team, 2015). The division of age classes to <4.4, 4.4–6.4, 6.4–8.4, 8.4–10.4, and >10.4 years is the same as that operationally used in the plantation (Vauhkonen *et al.* 2011).

### k-nearest neighbour prediction

Tree lists were estimated using the nearest neighbour prediction (*k-nn*) and canonical correlation analysis to produce a weighting matrix to select the reference observations similar to the object of prediction in terms of the independent variables. The *k-nn* method follows the description by Moeur and Stage (1995), except that the estimates were calculated as weighted averages of the $k$ nearest observations employing weights ($W$) from the inverse of the canonical correlation analysis based distance. This approach was chosen according to earlier experiences of multivariate prediction (see Packalén and Maltamo, 2008; Breidenbach *et al.*, 2010). In this study, the dependent variables were the 5th, 20th, 40th, 50th, 60th, 80th and 95th percentiles of the diameter distribution. The value of $k$ was set to 5 based on iterations varying $k$ from 1 to 10 to minimize the RMSE of the diameter percentiles. The number of ALS metrics used in the iterative variable selection (Gill *et al.*, 1981) was 15. The *k-nn* was implemented using the *yaImpute* package in R (Crookston and Finley, 2008).

The predicted diameter distributions consist of trees observed in the field reference plots. When constructing a diameter distribution for the target plot, the stem frequency represented by each tree of the reference plots is required. The plot level weight $W_{uj}$ of the reference plot $u$ and the target plot $j$ is used for this purpose. The plot level weights of $k$ neighbours sum to one.

Let $T_{uj}$ denote a set of trees $t$ that are predicted from the reference plot $u$ for the target plot $j$, and let $n$ denote the number of trees in $u$:

$$T_{uj} = \{t_{1uj}, t_{2uj}, \ldots, t_{n_u uj}\}. \quad (7)$$

Then the eventual set of trees predicted from the $k$ nearest reference observations for the target plot is

$$T_j = \{W_{1j}T_{1j}, W_{2j}T_{2j}, \ldots, W_{kj}T_{kj}\} \text{ (Packalén and Maltamo, 2008).} \quad (8)$$

### Accuracy assessment

Accuracy assessment was carried out by Leave-One-Out Cross-Validation (LOOCV), in which all plots within a stand were excluded from the training data when predicting for a plot of this particular stand. In the LOOCV model, parameters were repeatedly estimated by ignoring the observations for which the prediction was done. In the case of mixed-effects models, the accuracy of the model prediction calibrated by the clone effect was assessed (Packalén *et al.*, 2011a).

The accuracy assessment was based on relative frequencies between 0 and 1 for the 1 cm wide diameter classes of the formed Weibull distributions or tree lists. The accuracy of estimates was first evaluated in terms of relative root mean squared error (RMSE per cent) and bias% at the plot level, applying the mean of the third power of diameter defined as follows:

$$E(d^3) = \sum_{i=1}^{m} f_i d_i^3 \quad (9)$$

where $m$ is the number of diameter classes, $d$ is the midpoint of a diameter class and $f_i$ are the observed relative frequencies of those classes. This estimate is based on a transformation to third powers of diameters, and it has previously been applied as an approximation of stand volume (Kilkki and Päivinen, 1986). We applied it here since it is based only on the diameter distribution, i.e. it entails no height or volume model error and is measured on the same scale as the actual volume (Vauhkonen *et al.* 2010).

The equations of root mean square error and bias are as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(E(d^3) - \widehat{E(d^3)}\right)^2}{n}} \quad (10)$$

$$\text{BIAS} = \frac{\sum_{i=1}^{n}\left(E(d^3) - \widehat{E(d^3)}\right)}{n} \quad (11)$$

where $n$ is the number of plots, $E(d^3)$ is the observed value for plot $i$, and $\widehat{E(d^3)}$ is the predicted value for plot $i$. Predicted relative frequencies $\hat{f}_i$ of the diameter classes were used to calculate $\widehat{E(d^3)}$ in similar way to $E(d^3)$. Finally, the relative RMSE and biases were calculated by dividing the absolute values (equations (10 and 11)) by the true mean of the variable in question. The level of this relative value is comparable to the corresponding stand volume estimate. The prediction accuracies were studied by analysing the plot-level error of each modelling alternative using repeated measures analysis of variance (ANOVA) (Penner *et al.*, 2013). Predictions from the five modelling alternatives were considered to be the 'repeated measures'. The statistical significance of the differences between biases of different methods was tested using a $t$-test (Maltamo, 1997). Secondly, a validation was performed for unweighted distribution in the 1 cm classes by applying the error index (Reynolds *et al.*, 1988), which has also been applied by Gobakken and Næsset (2004) in an ALS context and is determined as follows:

$$\text{errorindex} = \sum_{i=1}^{m} 100 \left| \frac{f_i - \hat{f}_i}{F} \right| \quad (12)$$

To avoid the effect of other attributes on the evaluation of the diameter distributions, the frequencies were not scaled, i.e. the value of $F$ was 1. The range of error index values was between 0 and 200, with values of 0 and 200 depicting completely identical and completely disjoint distributions, respectively (Reynolds *et al.* 1988).

## Results

Since a LOOCV procedure was applied in all of the constructed models, we do not present exact parameter values. Therefore, only the model forms are presented in Table 2, which includes parameter models for Weibull in the case of both ALS information and field attributes, models for stand attributes in the case of Weibull parameter recovery, and the independent variables of the *k-nn* prediction. For the regression models, there were two (ALS information) or three (field attributes) independent variables (Table 2). For the *k-nn* prediction, there were five iteratively selected independent variables (Table 2). There were both height and density metrics in the ALS information based models, but

**Table 2** Dependent and independent variables of diameter distribution modelling alternatives.

| Estimation method and dependent variables | Independent variables[1] |
|---|---|
| ALS-based parameter prediction | |
| *b* | ln_*h50*, sqrt_*p50* |
| ln_*c* | sqrt_*h50*, sqrt_*h90* |
| Field information based parameter prediction | |
| ln_*b* | *N*, inv_Age, ln_SI |
| ln_*c* | ln_*Age*, SI |
| ALS-based parameter recovery | |
| Basal area | sqrt_*h30*, inv_*p70* |
| Mean diameter | ln_*h50*, ln_*p95* |
| Number of stems | inv_*h50*, inv_*h95* |
| *k-nn* | |
| Diameter percentiles | *hmean, h90, p30, h70, p10* |

[1]sqrt denotes square root, inv denotes inverse, and ln denotes natural logarithm.

height metrics were applied more often. Among the height percentiles, mean value and upper heights were used more often than lower heights. Stand age was not a statistically significant predictor variable in the ALS information based diameter distribution modelling alternatives. In general, Weibull parameter *b* and mean diameter are closely related, and a height percentile of 50 per cent was used as a predictor variable in our models to model both parameters. In the case of field information, stand age and site index were used in the models for Weibull parameters *b* and *c*, but the number of stems was only used for parameter *b*. The relationship between dependent variables and the ALS predictor variables was weakest in the case of models for Weibull parameter *c* and number of stems. In those models, only height metrics were statistically significant predictor variables and a second height metric was added to increase the variation to model predictions. The modelling of parameter *c* has also proven difficult in field information based studies (Maltamo, 1997). The field values of number of stems were close to each other in all plots due to the constant planting density, thus, making their prediction by ALS information a major challenge.

The accuracies of the estimated and predicted distributions are shown in Table 3. In the case of the RMSE% of $E(d^3)$, both the fitted ML-estimates of the Weibull parameters and recovered estimates based on observed stand attributes were below 1 per cent, and showed a very close agreement between actual and smoothed distributions. In contrast, the RMSE% of the estimates based on predictions by ALS or field information was between 10 and 18 per cent. The most accurate prediction was the one based on field information and stand register data followed by parameter recovery of ALS-based estimates. In the latter, 35 of the 195 plots did not converge. In those plots, *N* estimates were then slightly modified to find a converged solution. The most inaccurate alternatives were ALS-based parameter prediction and distribution matching. The results of the repeated measures ANOVA showed statistical differences in accuracy (Wilk's Lambda $P = 0.025$). We also conducted follow-up comparisons on modelling alternatives to determine significant differences. The results showed that the predictions of distribution matching differed significantly from all other alternatives ($P = 0.020$ *vs* ALS-based parameter prediction, $P < 0.0001$ *vs* *k-nn,* $P = 0.0003$ *vs* field information based parameter prediction and $P < 0.0001$ *vs* ALS-based parameter recovery). In addition, ALS-based parameter recovery and ALS-based parameter prediction differed from each other ($P = 0.039$). Finally, ALS-based parameter prediction and distribution matching were the only alternatives that showed a tendency to underestimate the $E(d^3)$. These underestimates were also statistically significant (Table 3).

In the case of error indices, only minor differences between the original fitted and the predicted distribution estimates were seen. The error index values were already over 40 with the original fitted distributions, and between 50 and 60 with the ALS or field information based predicted diameter distribution estimates. In this case, parameter prediction was the most accurate ALS-based alternative, but the differences between the alternatives were small (Table 3).

The examples of diameter distribution predictions in plots with low (~25), medium (~50) and high (~100) values of error index are shown in Figures 1–3. On average, ~4 per cent of predictions resulted in either low or high value categories of error index. The plots were chosen according to error index values of ALS-based parameter prediction alternatives. Thus, the error index values of other alternatives may be considerably different.

In the case of close agreement between estimated or predicted distributions and empirical distributions (Figure 1), i.e. a low error index, the different modelling alternatives showed a high degree of overlap. However, in the case of distribution matching the estimate was peaked. For *k-nn*, the frequencies of larger trees were greater than in the actual distribution. This was also seen in the error index values.

In the average error index example (Figure 2), the figures of the distribution estimates varied according to distribution kurtosis. Nevertheless, the error indices of the most accurate alternatives were close to one another. In contrast, the most inaccurate alternatives were the ALS-based parameter prediction and *k-nn*. The *k-nn* estimate included larger trees than those observed in the actual field distribution (see also Figure 1). Note that although estimates of ALS-based parameter prediction and recovery seemed very similar, the fit of parameter prediction was poor.

In the last example (Figure 3), the distribution estimates of rather high values of error indices were compared. However, high error index values only consider parameter recovery and parameter prediction alternatives whereas distribution estimates seem to be affected by a couple of small trees of 7–8 cm diameter. The error index values were considerably smaller for field information based parameter prediction and distribution matching.

**Table 3** Accuracy metrics of the compared methods.

| Estimation method | RMSE% | bias% | Error index |
|---|---|---|---|
| Maximum likelihood | 0.50 | 0.20 | 43.75 (12.4–78.5) |
| Parameter recovery | 0.90 | 0.50 | 46.30 (11.0–82.7) |
| ALS-based parameter prediction | 14.56 | 2.55 ** | **55.04** (12.3–132.5) |
| Field information based parameter prediction | 10.14 | 0.74 | 50.05 (15.4–131.7) |
| ALS-based parameter recovery | **11.59** | **−0.05** | 60.21 (22.2–125.0) |
| Distribution matching | 18.14 | 5.77 *** | 56.22 (16.1–138.0) |
| *k-nn* | 13.47 | 0.06 | 58.71 (17.5–129.0) |

In the case of error index, range is also given. The most accurate ALS-based prediction method is shown in bold.
Statistical significance of *T*-test.: *** = prob($T$ < t) < 0.001, ** = 0.001 < prob($T < t$) < 0.01, * = 0.01 < prob($T < t$) < 0.05.

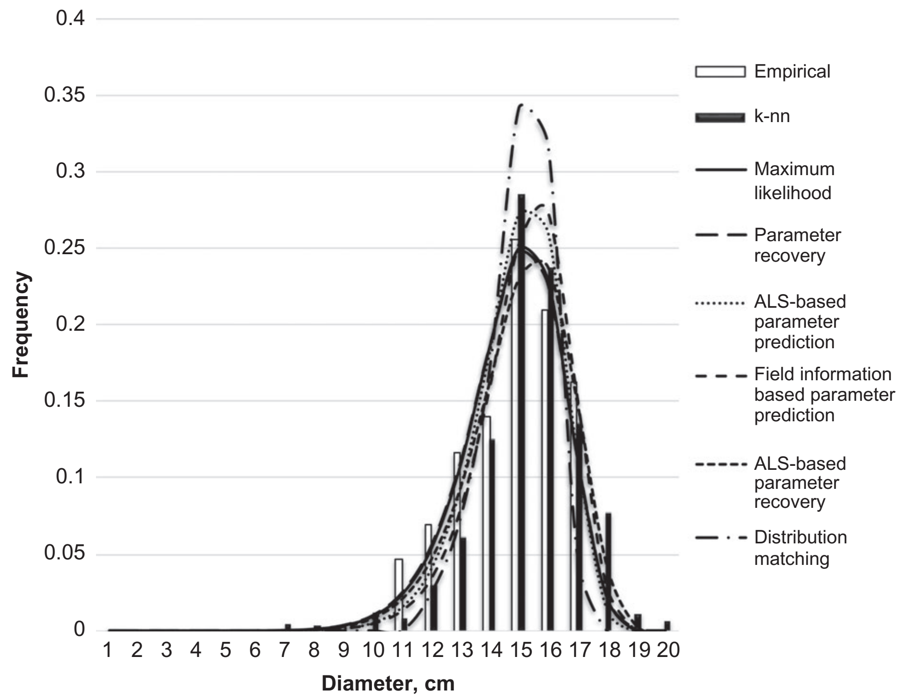


**Figure 1** An example plot with low error index values for the estimated diameter distributions: ML-estimation (20.5), parameter recovery (20.2), ALS-based parameter prediction (26.3), ALS-based parameter recovery (22.3), distribution matching (47.4), field information based parameter prediction (24.0) and *k-nn* (34.8).

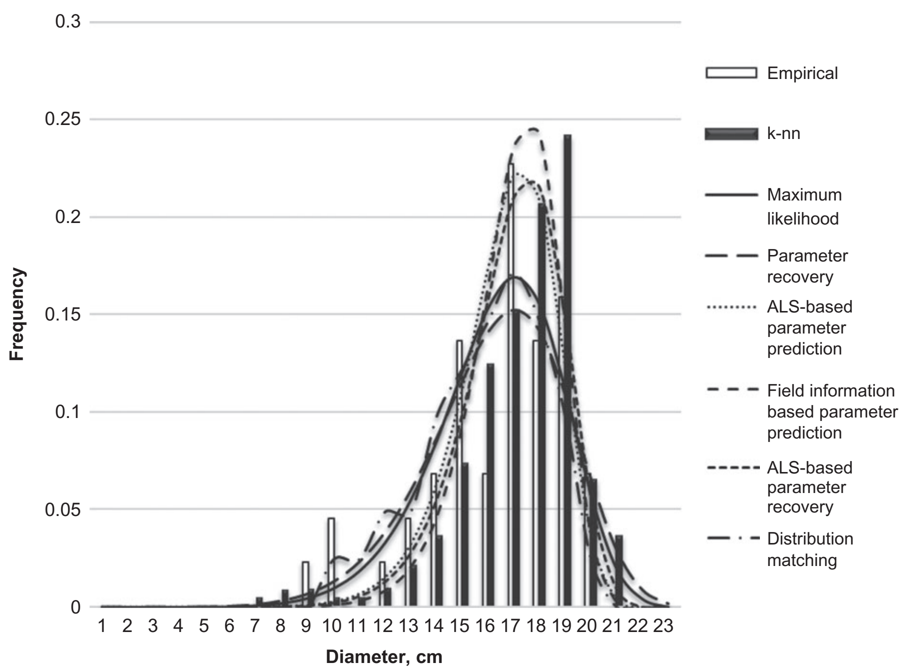


**Figure 2** An example plot with average error index values for the estimated diameter distributions: ML-estimation (37.2), parameter recovery (39.2), ALS-based parameter prediction (55.1), ALS-based parameter recovery (36.7), distribution matching (39.1), field information based parameter prediction (43.3) and *k-nn* (52.7).

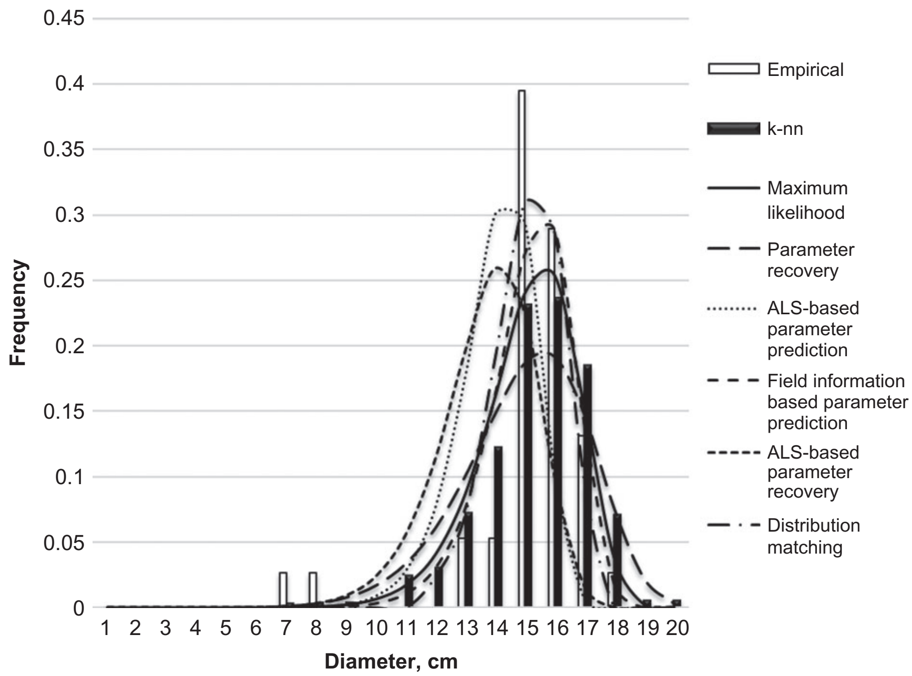


**Figure 3** An example plot with high error index values for the estimated diameter distributions: ML-estimation (48.5), parameter recovery (70.7), ALS-based parameter prediction (97.9), ALS-based parameter recovery (110.5), distribution matching (38.5), field information based parameter prediction (37.4) and *k-nn* (52.2).

## Discussion

This study examined the estimation of diameter distributions with ALS information in the tropical single-species plantation of *E. urograndis* using different modelling alternatives. The results of the study showed that ALS information can predict diameter distributions with a degree of error slightly more than 10 per cent of the RMSE% of $E(d^3)$, and with error index values between 50 and 60.

We compared different statistical methods to predict Weibull distributions with ALS information. Earlier studies based on field data have suggested that parameter recovery yields more accurate and robust estimates than parameter prediction (Hyink and Moser, 1983; Siipilehto and Mehtätalo, 2013). Of course, this is also dependent on the availability of stand attributes for the prediction and recovery and their estimation errors. We confirmed that parameter recovery was an accurate alternative, although the stand attribute predictions were not always compatible with the assumed Weibull function. This incompatibility may have effects on the estimates of stand attributes whenever the parameter recovery is used to predict the diameter distributions of the entire plantation. Otherwise, the effect only impacts on the accuracy of the estimate for *N*, and no effect can be observed on the accuracy metrics. We also found that differences between the parameter recovery and parameter prediction alternatives, with respect to their RMSE% of $E(d^3)$, were statistically significant based on the repeated measures ANOVA. Moreover, the differences between biases were statistically significant in the case of parameter prediction.

Although the performance of the distribution matching approach did not compare favourably with the other Weibull-based methods in terms of all of the accuracy measures, it may be possible to obtain similar levels of accuracy using a lesser number of field plots for training, by employing *a priori* information in the form of tree height distributions (Vauhkonen and Mehtätalo, 2015). Unlike the parameter prediction and recovery approaches, which are based on the use of ALS predictors or stand attributes, distribution matching requires another distribution that is calibrated to match the reference distribution as closely as possible. Even though it may be possible to describe the transformation in detail and use discrete classes, modelling of such a transformation is difficult. To produce a smooth transformation, both the reference and target distribution were based on the Weibull function, but other types of functions may also be considered.

The *k-nn* method produced better estimates than the parameter prediction alternative in terms of the RMSE% of $E(d^3)$. The *k-nn* estimates had higher RMSE% of $E(d^3)$ than the parameter recovery estimates, but the differences were not statistically significant. Unlike other methods based on the continuous Weibull function, the *k-nn* prediction method allows the compilation of distributions with discrete diameter classes. Therefore, the prediction is more complicated, because there are no distribution parameters to predict directly. We wanted to avoid the use of traditional stand attributes, such as the volume or basal area as dependent variables, because they do not describe the shape of the distribution. Instead, we tried to find indicators of the distribution shape and computed the distance metric based on the diameter percentiles. Therefore, it is possible that the *k-nn* estimates could be further improved by employing better optimized variables in the canonical correlation analysis. It should also be noted that only five independent variables were applied in our *k-nn* prediction, which is considerably less than in previous studies (e.g. Maltamo *et al.*, 2009). Aside from the distance metric and the number of neighbours, the amount of available reference data will also have a strong effect on the accuracy of the *k-nn* prediction. Our 195 reference plots is much lower than the number typically used for a corresponding species-specific inventory (Maltamo and Packalen, 2014), but is close to the level used for reference data in total diameter distribution predictions (Maltamo *et al.*, 2009).

The results of the comparison of different ALS-based diameter distribution predictions showed rather small differences in error indices. It was also notable that while RMSE percentage figures were based on the transformation to the third powers of diameter (thus giving more weight to large trees), the error indices were calculated from unweighted distributions. Thus, the highest error index values for parameter recovery may be due to the fact that it produced highly peaked and narrow distribution of estimates (Siipilehto and Mehtätalo, 2013). In our data, the true distributions were usually very narrow, which can lead to a high error index value due to a small estimation error in the mean diameter. On the other hand, parameter prediction methods usually provide more averaged and flat distribution predictions (Maltamo, 1997).

It was also notable that the mean error index values of the fitted distributions (without prediction phase) were already over 40, indicating a rather small difference between the original smoothed and predicted diameter distribution estimates. In studies where the growth of predicted diameter distributions has been simulated, the results have shown that small differences in the original distributions do not lead to larger differences in simulations (Maltamo and Kangas, 1998; Mäkinen *et al.*, 2008). Thus, the slightly larger error index values of parameter recovery may not be a problem if diameter distribution estimate is used for growth prediction.

We also tested field information based diameter distribution prediction. The distribution parameters were predicted by applying true stand attribute values, whereas only assessments of these variables can be used in practical inventories. In the northern boreal zone, the use of true attributes has led to RMSE% values of predicted volume below 10 per cent (Kangas and Maltamo, 2000), whereas when visually assessed stand attributes were compared to the values of a systematic network of circular sample plots within each stand, the corresponding RMSE % values were at least 25 per cent (Haara and Korhonen, 2004). The advantage of remote sensing based approaches over field based approaches is that they provide that same information in a spatially continuous way (Packalén and Maltamo, 2008). In this study, we obtained an RMSE% of $E(d^3)$ of slightly more than 10 per cent for field information based diameter distribution prediction. Our most accurate ALS-based diameter distribution predictions were almost as accurate as the field based predictions and the differences were not statistically significant. The fact that stand age was not chosen in the ALS information based models despite being available as a predictor variable reinforces the premise that ALS data can replace the need for field information. Plantations differ from natural or semi-natural forests in that the stand age is always known in the former. In addition, initial constant planting density, clone type, and some *a priori* information on the site index is known. In this study, we applied the stem density and site index values calculated from the field data. Thus, as this information is only available in plantations, the accuracy values of field information based diameter distribution prediction would be lower when applied to natural forests.

Various studies have applied error indices (Reynolds *et al.*, 1988), statistical tests (Kolmogorov–Smirnoff: Siipilehto, 2000) and accuracy of derived end products, such as stand volume (Siipilehto, 1999) or approximate volume by means of third power of diameter class frequencies (Kilkki and Päivinen, 1986) to validate diameter distribution estimates. We employed the mean of the third power of diameter in order to avoid the effects of other attributes, such as height, and thus it was chosen over the predicted stand volume as an accuracy criterion. In the case of the error index, we tested proportional frequencies between 0 and 1 without scaling them to any stand attribute, which would affect the accuracy of the estimates. This was done in order to focus solely on the relationship between ALS metrics and diameter distributions. Either accuracy metric describes a different aspect of each method; RMSE provides information in regard to the capacity for final volume prediction while the error index indicates the general fit to the diameter distribution. As such, they may result in contradictions, as was the case in this study. However, in regard to statistical significance, our results showed that the most accurate methods were the parameter prediction and *k-nn* and these methods also have the desired property of compatibility between the predicted diameter distributions and measured stand characteristics. In other words, the basal area, mean diameter and number of stems based on the diameter distributions are equal to the values used in the prediction.

Although comparing between different studies is difficult due to inherent variations (e.g. different plot sizes, etc.), it is worth noting that our average error index values were between 50 and 60, whereas Maltamo *et al.* (2009) obtained corresponding values of over 80 in boreal forests in Norway, where the stands typically had more than one species, and were pre-classified according to development stage and dominant tree species. This further emphasizes the strength of the relationship between ALS metrics and diameter distributions, and how ALS information can be used successfully in diameter distribution prediction in plantations. Earlier studies that had predicted stand attributes in the same plantation that we analysed in this study (Packalén *et al.*, 2011a; 2011b; Vauhkonen *et al.*, 2011) reported RMSE% values of volume estimates that varied between 7 and 12 per cent depending on the level of applied calibration (stand, clone). Although this study did not focus specifically on volume estimations, the results obtained for the third power of diameters were, in general, very similar.

The results of this study demonstrated the potential of ALS information to predict diameter distributions in single-species plantations. They revealed only slight differences between the modelling alternatives. Nonetheless, these results can also be discussed in the context of more heterogeneous stand structures. For example, in previous studies carried out in mixed species forests, the *k-nn* method was more accurate than Weibull-based alternatives (Packalén and Maltamo, 2008); while the *k-nn* method can predict multimodal diameter distributions, the Weibull function is less flexible, which hampers its performance in forests with very heterogeneous structures. One way to overcome that weakness is to use a finite mixtures approach (Thomas *et al.*, 2008), although that substantially increases the number of parameters to be estimated. The problem related to parameter recovery is that the system of equations might not converge in all cases. Although this shortcoming was not a major problem in our plantation data, in mixed forests and heterogeneous stands the lack of solutions may be a more severe problem. Thus, in the case of unimodal diameter distributions, both ALS information based parameter recovery and *k-nn* seem to be feasible alternatives for modelling. Previous studies and the findings of this study (e.g. convergence, close RMSE metrics of parameter recovery and *k-nn*) suggest that *k-nn* approaches should be used in more heterogeneous stand structures, especially in the presence of representative reference data and without a need to separate species. The improvement of species specific diameter distribution estimates could be achieved by further refining the separation of species using ALS datasets, for example by integrating leaf off data or applying multispectral ALS data.

## Conclusions

We found a strong relationship between ALS information and diameter distributions in single-species plantations. We found the parameter recovery, in which stand attributes are first predicted by ALS information and distribution parameters are later recovered from those predicted attributes, and the *k-nn* method, which produces non-parametric tree lists, to be the most appropriate alternatives for ALS-based diameter distribution prediction in single-species plantations. This conclusion was reached on the basis of statistically significant differences among methods, the compatibility between predicted diameter distributions and general stand attributes, and the accuracy of the stand volume approximation in terms of $E(d^3)$. The ranking between different modelling alternatives, however, varied depending on which accuracy metrics were considered, and therefore the best choice of method depends on the applied criteria. In more heterogeneous stand structures with several tree species and multimodal diameter distributions, *k-nn* may be the most suitable alternative since Weibull-based methods may be challenged.

## Conflict of interest statement

None declared.

## References


Axelsson, P. 2000 DEM generation from laser scanner data using adaptive TIN models. in *Proceedings of the XIXth ISPRS Conference*, IAPRS, Vol. XXXIII, Amsterdam, The Netherlands, pp. 110–117.

Bailey, R.L. and Dell, T.R. 1973 Quantifying diameter distributions with the Weibull function. *For. Sci.* **19**, 97–104.

Bollandsås, O.M. and Næsset, E. 2007 Estimating percentile-based diameter distributions in uneven-aged Norway spruce stands using airborne laser scanner data. *Scand. J. For. Res.* **22**, 33–47.

Breidenbach, J., Næsset, E., Lien, V., Gobakken, T. and Solberg, S. 2010 Prediction of species-specific forest inventory attributes using a non-parametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. *Remote Sens. Environ.* **114**, 911–924.

Clutter, J.L., Fortson, J.C., Pienaar, L.V., Brister, G.H. and Bailey, R.L. 1983 *Timber Management: A Quantitative Approach*. Wiley, p. 334.

Crookston, N.L. and Finley, A O. 2008 yaImpute: an R package for kNN imputation. *J. Stat. Softw.* **23** (10), 1–16.

Dubey, S.D. 1967 Some percentile estimators for Weibull parameters. *Technometrics* **9**, 119–129.

Gill, P.E., Murray, W. and Wright, M.M. 1981 *Practical Optimization*. Academic Press Inc, p. 401.

Gobakken, T. and Næsset, E. 2004 Estimation of diameter and basal area distributions in coniferous forest by means of airborne laser scanner data. *Scand. J. For. Res.* **19**, 529–542.

Gobakken, T. and Næsset, E. 2005 Weibull and percentile models for LIDAR-based estimation of basal area distribution. *Scand. J. For. Res.* **20**, 490–502.

Haara, A. and Korhonen, K.T. 2004 Kuvioittaisen arvioinnin luotettavuus. *Metsätieteen aikakauskirja* **4/2004**, 489–508. (In Finnish).

Harter, H.L. and Moore, A.H. 1965 Maximum likelihood estimation of parameters of Gamma and Weibull populations from complete and from censored samples. *Technometrics* **4**, 639–643.

Holopainen, M., Vastaranta, M., Rasinmäki, J., Kalliovirta, J., Mäkinen, A., Haapanen, R., *et al.* 2010 Uncertainty in timber assortment predicted from forest inventory data. *Eur. J. For. Res.* **129**, 1131–1142.

Hyink, D.M. and Moser, J.W. 1983 A generalized framework for projecting forest yield and stand structure using diameter distributions. *For. Sci* **29**, 85–95.

Kangas, A. and Maltamo, M. 2000 Performance of percentile based diameter distribution prediction and Weibull method in independent data sets. *Silva Fenn.* **34**, 381–398.

Kilkki, P. and Päivinen, R. 1986 Weibull function in the estimation of the basal-area DBH-distribution. *Silva Fenn.* **20**, 149–156.

Lloyd, C.D. and Atkinson, P.M. 2002 Deriving DSMs from LiDAR data with kriging. *Int. J. Remote Sens.* **23** (12), 2519–2524.

Mabvurira, D., Maltamo, M. and Kangas, A. 2002 Predicting and calibrating diameter distributions of Eucalyptus grandis (Hill) Maiden plantations in Zimbabwe. *New For.* **23**, 207–223.

Magnussen, S., Næsset, E. and Gobakken, T. 2013 Prediction of tree-size distributions and inventory variables from cumulants of canopy height distributions. *Forestry* **86**, 585–593.

Magnussen, S. and 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. *Ann. For. Sci.* **73**, 1089–1098.

Mäkinen, A., Kangas, A., Kalliovirta, J., Rasinmäki, J. and Välimäki, J. 2008 Comparison of treewise and standwise forest simulators by means of quantile regression. *For. Ecol. Manage.* **255**, 2709–2717.

Maltamo, M. 1997 Comparing basal area diameter distributions estimated by tree species and for the entire growing stock in a mixed stand. *Silva Fenn.* **31**, 53–65.

Maltamo, M. and Kangas, A. 1998 Methods based on k-nearest neighbor regression in the estimation of basal area diameter distribution. *Can. J. For. Res.* **28**, 1107–1115.

Maltamo, M., Eerikäinen, K., Packalén, P. and Hyyppä, J. 2006 Estimation of stem volume using laser scanning based canopy height metrics. *Forestry* **79**, 217–229.

Maltamo, M., Næsset, E., Bollandsås, O.-M., Gobakken, T. and Packalén, P. 2009 Non-parametric estimation of diameter distributions by using ALS data. *Scand. J. For. Res.* **24**, 541–553.

Maltamo, M. and Gobakken, T. 2014 Predicting tree diameter distributions. In *Forestry Applications of Airborne Laser Scanning – Concepts and Case Studies*, **Vol. 27**. Maltamo M., Næsset E. and Vauhkonen J. (eds). Springer, pp. 177–191. Managing Forest Ecosystems.

Maltamo, M. and Packalen, P. 2014 Species specific management inventory in Finland. In *Forestry Applications of Airborne Laser Scanning – Concepts and Case Studies*, V**ol. 27**. Maltamo M., Næsset E. and Vauhkonen J. (eds). Springer, pp. 241–252. Managing Forest Ecosystems.

Mehtätalo, L. 2015 lmfor: Functions for Forest Biometrics. *R package version 1.1*.

Mehtätalo, L., Maltamo, M. and Packalén, P. 2007 Recovering plot-specific diameter distribution and height-diameter curve using ALS based stand characteristics. In *Proceedings of ISPRS Workshop Laser Scanning 2007 and Silvilaser 2007*, Rönnholm P., Hyyppä J. and Hyyppä H. (eds.) September 12–14, 2007, Finland. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVI (Part 3/W52), 288–293.

Moeur, M. and Stage, A.R. 1995 Most similar neighbor: an improved sampling inference procedure for natural resource planning. *For. Sci.* **41**, 337–359.

Nanang, D.M. 1998 Suitability of the Normal, Log-normal and Weibull distributions for fitting diameter distributions of neem plantations in Northern Ghana. *For. Ecol. Manage.* **103**, 1–7.

Näslund, M. 1937 Skogsförsökanstaltens gallringsförsök i tallskog. *Meddelanden från Statens Skogsförsöksanstalt* **29**, 169.

Næsset, E. 2002 Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sens. Environ.* **80**, 88–99.

Næsset, E. 2014 Area-based inventory in Norway – from innovation to an operational reality. In *Forestry Applications of Airborne Laser Scanning – Concepts and Case Studies*, **Vol. 27**. Maltamo M., Næsset E. and Vauhkonen J. (eds). Springer, pp. 215–240. Managing Forest Ecosystems.

Packalén, P. and Maltamo, M. 2008 The estimation of species-specific diameter distributions using airborne laser scanning and aerial photographs. *Can. J. For. Res* **38**, 1750–1760.

Packalén, P., Heinonen, T., Vauhkonen, J., Pukkala, T. and Maltamo, M. 2011a Dynamic treatment units in eucalyptus plantation. *For. Sci.* **57**, 416–426.

Packalén, P., Mehtätalo, L. and Maltamo, M. 2011b 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.* **68**, 1085–1092.

Penner, M., Pitt, D.G. and Woods, M.E. 2013 Parametric vs. nonparametric LiDAR models for operational forest inventory in boreal Ontario. *Can. J. Remote Sensing* **39** (5), 426–443.

R Core Team. 2015 R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL https://www.R-project.org/ (accessed on 28 September, 2015).

Rennolls, K., Geary, D.N. and Rollinson, T.J.D. 1985 Characterising diameter distributions by the use of the Weibull distribution. *Forestry* **58** (1), 57–66.

Reynolds, M.R., Jr., Burk, T.E. and Huang, W.C. 1988 Goodness-of-fit tests and model selection procedures for diameter distribution models. *For. Sci.* **34**, 373–399.

Rombouts, J.H., Ferguson, I.S. and Leech, J.W. 2010 Campaign and site effects in LiDAR prediction models for site quality assessment of radiata pine plantations in South Australia. *Int. J. Remote Sens.* **31**, 1155–1173.

Saad, R., Wallerman, J. and Lämås, T. 2015 Estimating stem diameter distributions from airborne laser scanning data and their effects on long term forest management planning. *Scand. J. For. Res.* **30**, 186–196.

Saramäki, J. 1992 A growth and yield prediction model of *Pinus kesiya* (Royle ex Gordon) in Zambia. *Acta For. Fenn.* **230**, 68.

Shang, C., Treitz, P., Caspersen, J. and Jones, T. 2017 Estimating stem diameter distributions in a management context for a tolerant hardwood forest using ALS height and intensity data. *Can. J. Remote Sens.* **43**, 79–94.

Siipilehto, J. 1999 Improving the accuracy of predicted basal-area diameter distribution in advanced stands by determining stem number. *Silva Fenn.* **33** (4), 281–301.

Siipilehto, J. 2000 A comparison of two parameter prediction methods for stand structure in Finland. *Silva Fenn.* **34** (4), 331–349.

Siipilehto, J. and Mehtätalo, L. 2013 Parameter recovery vs. parameter prediction for the Weibull distribution validated for Scots pine stands in Finland. *Silva Fenn.* **47** (4), article id 1057.

Tesfamichael, S.G., Ahmed, F., van Aardt, J. and Blakeway, F. 2009 A semi-variogram approach for estimating stems per hectare in Eucalyptus grandis plantations using discrete-return lidar height data. *For. Ecol. Manage.* **258**, 1188–1199.

Thomas, V., Oliver, R.D., Lim, K. and Woods, M. 2008 Lidar and Weibull modeling of diameter and basal area. *For. Chron.* **84** (6), 866–875.

Vauhkonen, J., Tokola, T., Maltamo, M. and Packalén, P. 2010 Applied 3D texture features in ALS-based forest inventory. *Eur. J. For. Res.* **129**, 803–811.

Vauhkonen, J., Maltamo, M., Næsset, E. and McRoberts, R. 2014 Introduction to forest applications of airborne laser scanning. In *Forestry Applications of Airborne Laser Scanning – Concepts and Case Studies*, **Vol. 27**. Maltamo M., Næsset E. and Vauhkonen J. (eds). Springer, pp. 1–16. Managing Forest Ecosystems.

Vauhkonen, J. and Mehtätalo, L. 2015 Matching remotely sensed and field measured tree size distributions. *Can. J. For. Res.* **45**, 353–363.

Vauhkonen, J., Mehtätalo, L. and Packalén, P. 2011 Combining tree height samples produced by airborne laser scanning and stand management records to estimate plot volume in Eucalyptus plantations. *Can. J. For. Res.* **41**, 1649–1658.

White, J.C., Wulder, M.A., Varhola, A., Vastaranta, M., Coops, N.C., Cook, B. D., *et al.* 2013 A best practices guide for generating forest inventory attributes from airborne laser scanning data using an area-based approach. *For. Chron.* **89**, 6722–723.