Refining and evaluating a Horvitz-Thompson-like stand density estimator in individual tree detection based on airborne laser scanning

Kansanen, Kasper

2022-04

Kansanen , K , Packalen , P , Lahivaara , T , Seppanen , A , Vauhkonen , J , Maltamo , M & Mehtatalo , L 2022 , ' Refining and evaluating a Horvitz-Thompson-like stand density estimator in individual tree detection based on airborne laser scanning ' , Canadian Journal of Forest Research , vol. 52 , no. 4 , pp. 527-538 . https://doi.org/10.1139/cjfr-2021-0123

http://hdl.handle.net/10138/349610
https://doi.org/10.1139/cjfr-2021-0123

acceptedVersion

Refining and evaluating a Horvitz–Thompson-like stand density estimator in individual tree detection based on airborne laser scanning

Kasper Kansanen, Petteri Packalen, Timo Lähivaara, Aku Seppänen, Jari Vauhkonen, Matti Maltamo, and Lauri Mehtätalo

Abstract: Horvitz–Thompson-like stand density estimation is a method for estimating the stand density from tree crown objects extracted from airborne laser scanning data through individual tree detection. The estimator is based on stochastic geometry and mathematical morphology of the (planar) set formed by the detected tree crowns. This set is used to approximate the detection probabilities of trees. These probabilities are then used to calculate the estimate. The method includes a tuning parameter, which needs to be known to apply the method. We present a refinement of the method to allow more general detection conditions than the previous papers and present and discuss the methods for estimating the tuning parameter of the estimator using a functional k-nearest neighbors method. We test the model fitting and prediction in two spatially separate data sets and examine the plot-level accuracy of estimation. The estimator produced a 13% lower RMSE than the benchmark method in an external validation data set. We also analyze the effects of similarity and dissimilarity of training and validation data to the results.

Key words: segmentation, stochastic geometry, detectability, nearest neighbor, forest inventory.

Kasper Kansanen. School of Computing, University of Eastern Finland, Postal Box 111, 80101 Joensuu, Finland.
Petteri Packalen. School of Forest Sciences, University of Eastern Finland, Postal Box 111, 80101 Joensuu, Finland. petteri.packalen@uef.fi
Timo Lähivaara. Department of Applied Physics, University of Eastern Finland, Postal Box 1627, Kuopio, Finland. timo.lahivaara@uef.fi
Aku Seppänen. Department of Applied Physics, University of Eastern Finland, Postal Box 1627, Kuopio, Finland. aku.seppanen@uef.fi
Jari Vauhkonen. Department of Forest Sciences, University of Helsinki, P.O. Box 27, 00014 Helsinki, Finland, jari.vauhkonen@helsinki.fi
Matti Maltamo. School of Forest Sciences, University of Eastern Finland, Postal Box 111, 80101 Joensuu, Finland. matti.maltamo@uef.fi
Lauri Mehtätalo.[1] School of Computing, University of Eastern Finland, Postal Box 111, 80101 Joensuu, Finland and Natural Resources Institute Finland (Luke), Bioeconomy and Environment Unit, Yliopistokatu 6, 80101 Joensuu, Finland. lauri.mehtatalo@luke.fi

[1]Corresponding author (e-mail: lauri.mehtatalo@luke.fi).

DOI: 10.1139/Z*xx-xxx* Published by NRC Research Press

1. Introduction

Volume and biomass related attributes are usually of primary interest in remote sensing based forest inventories. If these quantities are estimated at tree level, scaling them to area-level requires information about stand density, which is most commonly measured using basal area or number of stems. The number of stems can be used directly by multiplying the mean tree volume or biomass by the estimated number of trees (if the joint behaviour of estimation errors of the two variables is ignored). Scaling using basal area is also possible, but it requires additional information about the diameter distribution in the form of quadratic mean diameter (Mehtätalo and Lappi, 2020, Chapter 11).

Two commonly used approaches for estimating stand density and other stand characteristics using remote sensing data, especially airborne laser scanning (ALS) data, are the area based approach (ABA) and individual tree detection (ITD) (Vauhkonen et al., 2014). In ABA, stand density can be estimated using parametric or nonparametric regression techniques. The resulting predictions typically have relative root mean square errors (RMSE%) of 20-30% at the plot level (Næsset, 2007; Packalén and Maltamo, 2007; Maltamo et al., 2014), which are rather high values compared to other total stand attributes predicted by ABA. The reason for the rather low accuracy is mainly the high variability of stand density with respect to remotely sensed predictors, which depends on how the existence of small trees is considered in the data (e.g., minimum diameter limit applied in the field measurements). In ITD, individual tree crowns are algorithmically detected from the data and stand density is directly estimated as number of detected trees per area unit. A major limitation of ITD is that not all trees can be correctly detected, which is especially true in forests with high stand density, vertically multilayered stand structures and tree groups with interlaced crowns (Vauhkonen et al., 2012). Typically, detection rates detectrate (number of detected trees / number of field-measured trees) of ITD under boreal managed forests vary between 40-90% (Lähivaara et al., 2014). The detection rate could further be divided into omission and commission error rates, which can be estimated by linking detected trees to a reference tree map.

newreferences Many kind of attempts have been done to correct the ITD for the bias of stand density. Maltamo et al. (2004) combined tree size distributions predicted by both ABA and ITD such that small trees were predicted by the ABA and the ITD was used to calibrate the large tree fraction. Ene et al. (2012) used stand densities predicted by ABA as *a priori* estimates to guide the ITD process. Flewelling (2008) and Melville et al. (2015) revisited sampling theory in an attempt to improve the precision of remotely sensed tree counts for plantation forests. Hou et al. (2016) first derived species-specific diameter distributions from both ABA and ITD and then calibrated the large tree fractions of the ABA-derived distribution by the corresponding fractions of the ITD. Breidenbach et al. (2010) imputed crown segments with a summation of field reference attributes instead of treating them as single trees, thereby reducing the area-level bias and also taking the commission errors into account. Strub and Osborne (2021) modeled tree counts within delineated segments by using a zero-deflated one-inflated Poisson model. Drone data has also been used to detect tree trunks within crown segments to improve the detection of understory trees (e.g. Kukkonen et al., 2021). Mehtätalo (2006) and Kansanen et al. (2016) presented methods based on stochastic geometry for estimating the true stand density based on tree crown objects detected by ITD and Kansanen et al. (2019) presented how the method could be used as a part of an ALS inventory chain.

This paper is continuation the work of Kansanen et al. (2016), where two different estimators were proposed. We concentrate on the more promising one: the Horvitz–Thompson-like (HT-like) estimator where the inclusion probability, called *detectability*, is estimated based on the ordered sequence of detected tree crown segments. We implicitly think that trees with larger crown segments are also taller. Therefore, their shortest distance to the measurement device is smaller, and the target-tree sized trees can be hidden below their canopies. Computing the detectability of a target starts with the union of the crown segments of trees that are larger than the target tree. If the target tree were located in that union, it would not be detected because the tree is covered by larger tree crowns. Therefore, it is justified to

define detectability as the ratio of the uncovered area and the total area of the sample plot. However, we need to define more specifically when a tree is located within the union. Mehtätalo (2006) required that the center point of the target tree is within the larger tree crowns, whereas Kansanen et al. (2016) introduced a tuning parameter α ($0 \leq \alpha \leq 1$) to model the condition for detectability so that $\alpha = 0$ corresponds to the condition of Mehtätalo (2006) and $\alpha = 1$ requires that the target tree crown is completely within the larger tree crowns to be hidden. comment4a Technically, a buffer is removed from the union of larger crown segments and the detectability is based on the area of that set; a mathematical formulation will be given later. Similar concept has also been applied to terrestrial inventory to take into account the hiding of trees behind other trees in terrestrial laser scanning (Kansanen et al., 2021). Also in that case, the order of trees is based on the shortest distance to the measurement device, but it is not related to tree size.

htlike The estimator was called the Horvitz-Thompson type estimator by Kansanen et al. (2016). The change in terminology follows from the use of similar estimators and naming conventions in distance sampling (see e.g. Buckland et al., 2004), where HT-like estimators are used in estimation of animal populations. The difference between these estimators and the standard Horvitz–Thompson estimator (Horvitz and Thompson, 1952) is that the inclusion probabilities in the HT estimator are known and fixed, whereas in HT-like estimators the detection probabilities have to be estimated, probably by using a model. Therefore, HT-estimator is design-based whereas the HT-like estimator is model-based. justdet The term detectability is also used in distance sampling and in estimation of animal populations in general (Thompson, 2012).

The tuning parameter α of Kansanen et al. (2016) is intended to take into account the differences of ITD algorithms in their ability to detect trees. The problem of selecting a value for the tuning parameter was not studied. comment1a If the tuning parameter α is estimated empirically using observed field data of training plots, the method can implicitly take into account other reasons for errors than hiding below larger canopies. For example, if the density of trees of a certain size class differs between the observable (uncovered) and hidden (covered) areas, the optimal value of α could simultaneously take into account both the detectability condition and interaction of tree locations. Especially, the ratio of hidden and visible parts is an unbiased estimator of detectability only if tree locations follow the complete spatial randomness. If the spatial pattern of trees is clustered, the hidden parts have higher stand density than the visible parts, and taking into account this implies smaller values of α than suggested by the detection conditon alone. Under regular pattern, the situation is opposite. Therefore, allowing a wider range to the tuning parameter might be justified. It is also justified to assume that the optimal tuning parameter α varies among plots. Furthermore, one could expect that sample plots with similar size distribution of detected trees could also have similar values of α. Therefore, a functional k-nearest neighbors (k-NN) methodology (see e.g. Ferraty and Vieu, 2006) might be a good choice for the problem at hand. In functional k-NN, the data is represented by functions (here empirical cumulative distribution functions of detected trees) and the similarity between data points is determined as distances between those functions. The use of functional k-NN in remote sensing inventory has been previously studied for example by Peuhkurinen et al. (2008).

In this paper, we extend the method of Kansanen et al. (2016) to allow also negative tuning parameters, which corresponds to not detecting a tree even when the center point is not covered by larger crowns. Furthermore, we evaluate the functional k-NN method for estimation of the tuning parameter. The tuning parameter is estimated using a training data set and the performance is evaluated in a spatially separate validation data set to evaluate the performance in the absence of local calibration field plots from the target area. An area-based method is used as a benchmark.

2. Material

There are two data sets corresponding to two spatially separated forest areas used in this study, Kiihtelysvaara and Liperi. Both of the study areas are typical boreal managed forest areas in Eastern

Table 1. Information relating to the data gathering in the two data sets used in this study.

	Kiihtelysvaara (training)	**Liperi (validation)**
location	62° 31′ N, 30° 10′ E	62° 25′ N, 29° 10′ E
time of ALS	June 26, 2009	July 2-10, 2016
scanning height, m	720	850
half angle, degrees	26	20
pulse repetition frequency, kHz	125	250
sampling density, pulses/m^2	11.9	14.0
time of field measurements	May-June 2010	May-August 2017
number of field plots	79	111

Table 2. Mean, standard deviation, minimum and maximum of stand density (N), quadratic mean diameter (QMD) and basal area (BA) in Kiihtelysvaara. The full data contains 79 field plots, of which 36 have ITD data.

Attribute	**n**	**mean**	**sd**	**min**	**max**
N, stems $\cdot$ ha^{-1}	79	1082.90	505.40	464	2850
	36	1046.55	438.45	466.67	2175
QMD, cm	79	17.91	3.78	10.98	29.04
	36	18.17	3.98	10.98	29.04
BA, m^2 $\cdot$ ha^{-1}	79	24.45	6.21	13.61	39.96
	36	24.69	6.31	15.31	39.96

Finland. In Kiihtelysvaara, Scots pine (Pinus sylvestris L.) represents 73% of the volume, Norway spruce (Picea abies [L.] Karst.) 16% of the volume and deciduous trees altogether about 11% of the volume, whereas in Liperi the corresponding proportions are 39%, 43% and 18%, respectively. The Kiihtelysvaara data are used as training data and the Liperi data as validation data.

The airborne laser scanning (ALS) data for Kiihtelysvaara were collected using an Optech ALTM Gemini laser scanning system, whereas in Liperi an Optech Titan was used. Optech Titan is a multi-spectral airborne LiDAR system that provides measurements from three different wavelengths, however, in this study we used only channel 2 (1064 nm). There is a side overlap of 55% in the data acquisition in both areas, which means that each location was covered from two flight lines in order to increase the probability that trees have ALS hits each side.

In both areas, field plots were placed subjectively in attempt to represent the species and size variation over the area. In Kiihtelysvaara, plot size varies between 20×20 m^2, 25×25 m^2 and 30×30 m^2, whereas in Liperi all of the plots are 30×30 m^2. In both areas, location, diameter at breast height (DBH) and height of a tree were measured and species was registered. In Kiihtelysvaara, trees were chosen under the criterion of either DBH $\geq$ 5 cm or height $\geq$ 4 m, whereas in Liperi DBH $\geq$5 cm was the criterion for recording their information. Information relating to ALS data collection and field measurements is presented in Table 1. Central plot-level attributes for Kiihtelysvaara are presented in Table 2 and for Liperi in Table 3.

As field data collection in Kiihtelysvaara and Liperi was performed using different criteria, we had to make the data sets similar. Hence, all trees with DBH $<$ 5 cm were removed from the Kiihtelysvaara field data. We assume that these small trees would not be detected by the ITD.

Table 3. Mean, standard deviation, minimum and maximum of stand density (N), quadratic mean diameter (QMD) and basal area (BA) in the validation data set (111 field plots).

Attribute	mean	sd	min	max
N, stems $\cdot$ ha^{-1}	1085.99	625.18	211.11	3900
QMD, cm	18.48	5.92	8.23	37.18
BA, m$^2 \cdot$ ha^{-1}	23.87	7.80	7.84	46.19

3. Methodology

3.1. Individual tree detection

The computational method used for ITD in this paper was presented by Lähivaara et al. (2014). It is based on a 3D template matching type approach, where simplified, rotationally symmetric crown shape models are fitted to ALS data. For each tree, the crown dimensions and shape are represented by four parameters: the crown radius, crown height, lower limit of the living crown and a crown shape parameter. In addition, the positions, i.e., horizontal coordinates of the tree crown center points are considered as unknown parameters. The tree positions and crown size and shape parameters of all trees are estimated simultaneously from the 3D ALS point cloud data by an iterative algorithm, which is based on Bayesian inference. More specifically, the *maximum a posteriori* (MAP) estimates for the parameters are computed.

The Bayesian inference used for the 3D-ITD is based on the *prior* model of the unknown parameters and the *likelihood* model for the measurements. Both models are constructed as in Lähivaara et al. (2014). For the unknown parameters, a Gaussian prior is used, written by combining data from field measurements with allometric tree shape models given by Muinonen (1995). bayesdist In addition, the prior model includes a constraint: center points of trees cannot be overlapped by the crowns of other trees. A Gaussian approximation is written also for the likelihood. We note that by using non-Gaussian likelihood model, the tree detection rates as well as height estimates could potentially be improved (Luostari et al., 2018). However, because the use of non-Gaussian model also leads to high increase of computational burden, and consequently, long computation times, the approach proposed by Luostari et al. (2018) is not taken in the present work.

comment2a The ITD method needs training data for the Gaussian prior. Hence, 43 field plots of the training data were used for this purpose. This leaves us the remaining 36 training plots to be used as training data for our stand density estimator. Figure 1 shows an example of detected tree crown segments and field measurements.

3.2. HT-like stand density estimation

Kansanen et al. (2016) presented a HT-like stand density estimator based on stochastic geometry and mathematical morphology (Chiu et al., 2013). We interpret the forest as a realisation of a germ-grain model of discs $\Xi = \bigcup B(x_i, R_i)$ and consider it in some area of interest $W \subset \mathbb{R}^2$. W is some field plot in our studies. The x_i are locations of crown center points, distributed as a homogeneous point process of intensity N (the stand density). Closed discs B with random radii R_i correspond to the projections of tree crowns to the ground; hereafter we will refer by word crowndef"crown" to projections of crowns to the ground level. ITD produces estimates of the tree locations and crown radii from which we derive $\hat{\Xi}$ (the union of detected crowns). Model formulation assumes only omission errors, and therefore $\hat{\Xi}$ is a thinned version of Ξ. To combat possible edge effects, the detection should be done in a larger window W^+ that contains W, so that the effect of those crowns whose centers are outside the sample plot but still cover parts of W on detectability would also be taken into account.

Let us assume that n trees with crown radii r_i, $i = 1, \ldots, n$ and center points in W have been

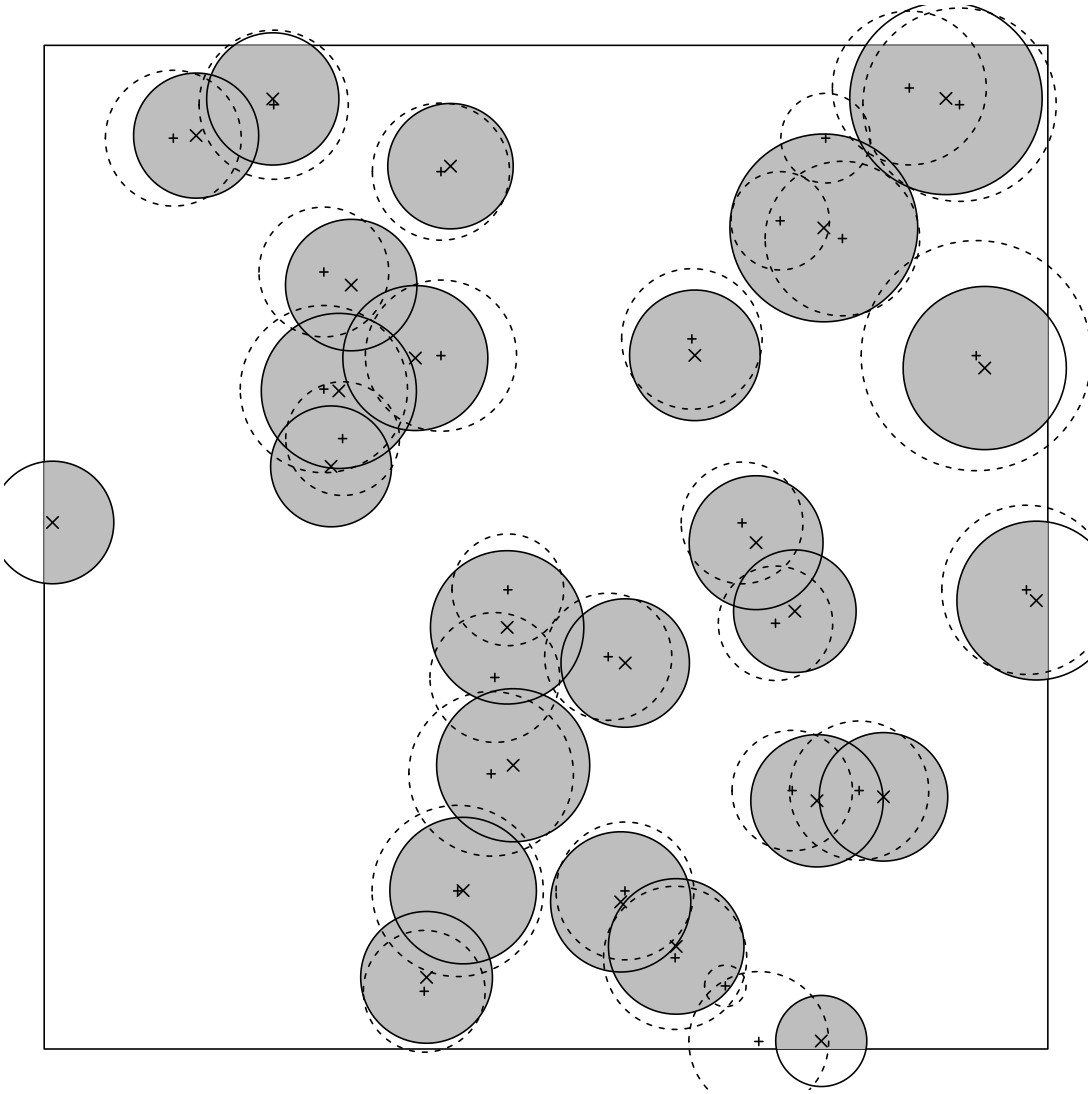


Fig. 1. An example of detected and field-measured trees in one sample plot of our data, with 25 detected and 28 field-measured trees. The gray discs show the detected tree crowns and crosses (x) their center points. The field-measured stem locations (at breast height) are shown by +; the dashed circle around each of them is poroprtional to tree DBH.

detected. Then the HT-like stand density estimator is

$$\hat{N} = \frac{10000}{|W|} \sum_{i=1}^{n} \frac{1}{p(r_i)}, \tag{1}$$

where the first term is just scaling $\hat{N}$ to $\mathrm{stems} \cdot \mathrm{ha}^{-1}$, and $p(r_i)$ is the estimated inclusion probability, *the detectability*, of a tree with crown radius r_i. As we already discussed in the introduction, the estimator is model-based if the detectability is estimated with a model, whereas the classical Horvitz-Thompson estimator with fixed, known inclusion probabilities (Horvitz and Thompson, 1952)

is design-based. The classical HT-estimator is unbiased, whereas the bias and variance properties of HT-like estimator depends on how well the detectabilities have been estimated.

comment4b Estimation of detectability is based on a sequential construction, where trees are ordered according to the crown size (i.e., the area of the projection of tree crown onto the ground level) from largest to smallest. We assume detectability 1 for the largest tree in W. For the other trees, we start by defining the union of tree crowns for the larger trees, which is further modified by removing or adding a buffer of width $|\alpha|r$ to determine where a tree with radius r would not be detectable (Fig. 2). Here $-1 \leq \alpha \leq 1$ is a tuning parameter which needs to be estimated using a training data; the negative value means that a buffer is added to the set of larger crowns and positive value means that it is removed. The detectability is then the probability that a uniformly distributed random point would not hit this set. This is written formally as

$$p_\alpha(r) = \begin{cases} 1 - \frac{|W \cap [\hat{\Xi}_{R>r} \ominus B(o,\alpha r)]|}{|W|}, & \alpha > 0 \\ 1 - \frac{|W \cap \hat{\Xi}_{R>r}|}{|W|}, & \alpha = 0 \\ 1 - \frac{|W \cap [\hat{\Xi}_{R>r} \oplus B(o,|\alpha| r)]|}{|W|}, & \alpha < 0 \end{cases}, \tag{2}$$

where r is the crown radius, $\hat{\Xi}_{R>r}$ is a subset of the detected germ-grain model formed by crowns with larger radii than r, $B(o,r)$ is an origin-centered closed disc of radius r, and $|.|$ is an area operator. dilation The operator $A \ominus B(o,\rho)$ is the erosion (removing a buffer of width ρ) and $A \oplus B(o,\rho)$ is the dilation (adding a buffer of width ρ) of set A.

The parameter $\alpha \in [-1,1]$ can be thought of as controlling the proportion of radius that should be covered by the larger trees for non-detection of the target tree. Positive parameter values correspond to situations where the center point of a tree should be inside the set formed by larger tree crowns to be hidden. Negative values correspond to situations where the center point of a crown can be outside the crowns of the larger trees, but part of the tree canopy is still within the larger tree canopies and the three is hidden. For example, $\alpha = 1$ assumes that trees are hidden only if their crowns are fully covered by larger ones. When $\alpha = -1$, a tree will be detected only if its crown does not overlap with larger tree crowns at all. When $\alpha = 0$, we assume that the tree it is not detected if the center point of the crown is covered by larger tree crowns but is detected otherwise. The case with erosion was considered in our earlier works on aerial inventory (Kansanen et al., 2016, 2019) whereas dilation has been previously used only in the very different situation of terrestrial laser scanning (Kansanen et al., 2021). To make the connection between the estimator and the parameter α clear, we write our estimator as a function of α, $N(\alpha)$.

comment1b Although we have justified the model and parameter α through the omission errors caused by overtopping larger tree crowns, the estimation of α through matching the field-measured stand density with $\widehat{N}(\alpha)$ will be affected by other factors, such as the spatial pattern of tree locations and commission error rate. For example, if the true detection condition for the applied ITD algorithm corresponds to the case $\alpha = 0$ but the trees are located regularly so that all trees are detectable, the estimate would be $\widehat{\alpha} = 1$. On the other hand, if the detection condition corresponds to $\alpha = 1$ but the tree locations are clustered, the fitted value of α would be less than 1 to compensate. Especially, our definition of detectability implies an assumption of complete spatial randomness for the hidden trees, but empirical estimation of α will compensate for the deviations from complete spatial randomness. Furthermore, it can compensate for commission error rates as well, commission even though the estimated number of stems cannot be smaller than the number of detected trees. Notice also that no model assumptions are made about the point process of the observed trees and the size distribution of trees, because the estimation is done conditional on the observed tree locations and sizes. In general, our approach is based on rather weak model assumptions.

move1 The optimal value of α for each plot of the training data was estimated by determining such value of α that minimizes the difference between observed stand density and the HT-like estimate. If

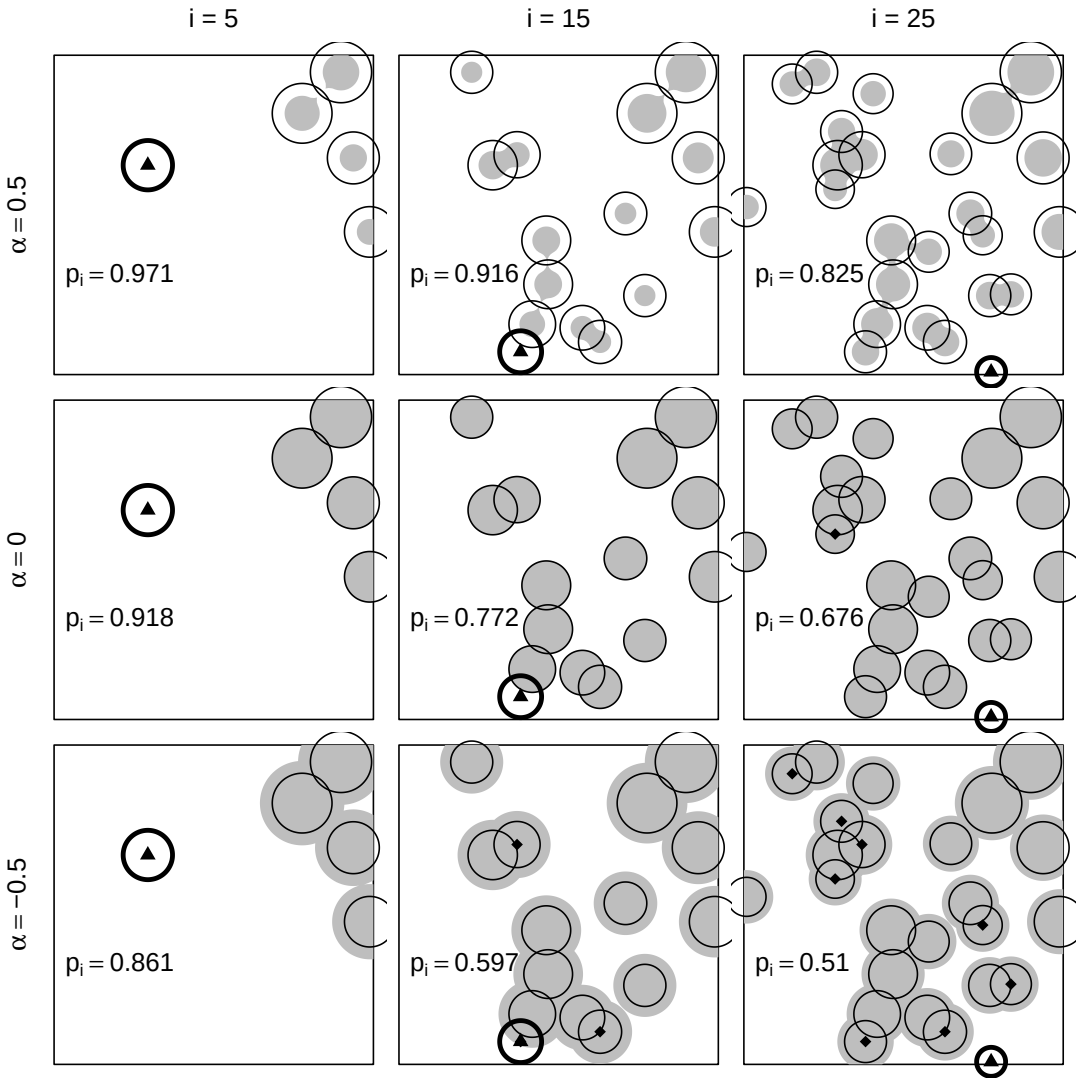


Fig. 2. Illustration of the computation of detectability for the 5^{th} (left) 15^{th} (middle) and 25^{th} (right) largest tree in the sample plot of Figure 1 of when α is 0.5 (top) 0 (middle) and -0.5 (bottom). The thin circles show the detected crowns for trees that are larger than the tree i, shown using thick circle and marked by (▲). A buffer of width $|\alpha|r_i$ is either removed from or added to the union of observed larger tree crowns (thin circles) to get the shaded area where the center point of tree i should be located to make it undetected. Detectability for tree i, p_i, is the relative proportion of the non-shaded area of the plot area. Trees marked by (⧫) are hidden under each of the detection conditions.

the number of detected trees is smaller or equal to the true stand density and the tuning parameter α is not restricted to range $(-1, 1)$, the optimal value leads to an exact match. If the observed density of detected trees is higher than the true density, then an optimal match cannot be found because the HT-like estimate cannot give a lower estimate than the density of detected trees.

3.2.1. Choosing α with functional k-NN

We assume that plots that have similar distributions of detected crown radii should also have similar values of α. Therefore a functional k-nearest neighbors (k-NN) method (Ferraty and Vieu, 2006) is used to find plots similar to the validation plots from the training data set and estimate α. The term "functional" means that we are assessing the similarity via distance between plot-specific functions in the target plot and in each plots of the training set. The function whose similarity is analyzed is the cumulative size distributions of observed trees .

Let $i \in I$ and $j \in J$ be indices of plots in training and validation sets, respectively. For every j, calculate the pairwise distances d_{ij} between the empirical cumulative distribution functions of crown radii, where i goes over the training plots. Here we use the Kolmogorov-Smirnov statistic (also called L^∞), as the distance metric

$$d_{ij} = \max_{r \in [0,\infty)} |F_i(r) - F_j(r)|, \tag{3}$$

which is a widely used similarity measure between two distributions. cdfIn our case, F_i and F_j are cumulative distribution functions of crown radii for two sample plots. For every validation plot j, we choose from among the training plots a total of k neighbours with the smallest distances d_{ij} to form a set K. The estimate α_j is determined by minimizing the root-mean-squared error of stand density in this set:

$$\widehat{\alpha}_j = \min_{\alpha \in [-1,1]} \sqrt{\sum_{i \in K} \frac{(N_i - \hat{N}_i(\alpha))^2}{k}}. \tag{4}$$

This minimizing procedure differs from the standard k-NN way of calculating an (inverse distance weighted) average of the variable of interest from the nearest neighbors and using that as the predicted value for new data. However, it should be noted that for $k = 1$ these procedures produce the same results, and for $k > 1$, the α_j based on (4) can also be considered as a weighted average of the α values of the neighbors, where the weights are related to the plot-level estimation errors. Conceptually, our approach first finds a training set of size k that corresponds well to a validation plot (i.e. all plots of the training set are sufficiently close to the target plot in terms of distance 3) and then determines the optimal α for that set.

As an alternative to the plot-level selection of α, we also consider choosing one fixed α for the whole validation set. This is done by taking the median of all estimates of α_j estimated by the k-NN. median Median was used instead of mean because it is more robust to exceptionally large or small values. This approach can be seen as a model averaging strategy that could improve prediction accuracy, especially if outliers are present among the estimated values of α_j.

3.3. The benchmark method

We use the area based approach (ABA) as a benchmark method. A linear regression model is fitted between the plot-level stand density and the explanatory variables derived from the ALS data. The models were fitted in the training data set (79 plots). All combinations of two and three explanatory variables were fitted and the model with the smallest root-mean-squared error in leave-one-out cross-validation test was chosen. The potential explanatory variables were the mean and standard deviation sd of ALS return heights, vegrat proportion of hits above the height of 2 meters, and the percentiles q_k and corresponding proportional densities p_k of ALS-based canopy height distribution, $k = 5, 10, \ldots, 95$. In addition, several transformations of these variables were considered: squared, square root, logarithmic and inverse. Models with larger number of explanatory variables were not considered to avoid overfitting. fitcriteriaThe model was selected purely based on a good fit empirically. The models chosen

for stand density N was

$$\widehat{\log N} = \hat{\beta}_0 + \hat{\beta}_1\sqrt{q_5} + \hat{\beta}_2 p_{50}^2 + \hat{\beta}_3\frac{1}{sd}. \quad (5)$$

To diminish the bias introduced by the logarithmic transformation in the stand density model a bias correction based on normal errors was introduced to the estimate biascorr (Mehtätalo and Lappi, 2020, Section 10.2):

$$\widehat{N} = \exp(\widehat{\log N}) \times \exp(0.5\hat{\sigma}^2)\,, \quad (6)$$

where $\hat{\sigma}^2$ is the estimated variance of residuals in (5).

3.4. Evaluation

Root-mean-squared errors,

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}, \quad (7)$$

means of errors

$$ME = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)}{n}, \quad (8)$$

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

Our training data set is quite small. Therefore, the results on the performance of the methods may partially be explained by lack of sufficiently good neighbours for some of the evaluation plots. To further analyze the effect of this, we explore the effect of similarity (4) between the training and validation data sets to the estimation errors. Especially, the k-NN method picks from among the training plots the k nearest neighbours for each plot of the validation data. These neighbors are the nearest ones that we have, but not necessary near enough, and a larger training data might lead to much better predictive performance. Therefore, we explore the effect of distance to the neighbours to the accuracy of our estimates of N as follows.

1. Order the validation plots based on the distance of the kth neighbor (the farthest among the k nearest ones) to get a list where the validation plots are ordered in terms of the quality (tightness) of the applied neighbourhood (worst is the last).

2. Compute the evaluation statistics in the current validation data set.

3. Drop the last plot from the list (the one with the worst neighbourhood) and repeat step 2 until there is no plots in the set.

4. Report the accuracy for each applied size of validation data set.

More formally, we have a chain of sets $J_{111} \supset J_{110} \supset \ldots \supset J_j$ where set J_n contains n plots and J_{n-1} is obtained from it by removing the plot that has the largest distance to its kth neighbor.

We also examine the estimation errors in subsets determined by dominant species, which is the species that has the largest share of the total plot volume. In the evaluation data set, 50 (45%), 42 (38%) and 19 (17%) plots are dominated by pine, spruce and deciduous trees, respectively. In the training data set, the corresponding numbers are 61 (77%), 12 (15%) and 6 (8%) among the 79 plots

available for the ABA estimator and 28 (78%), 6 (17%) and 2 (5%) among the 36 plots available for the HT-like estimator comment2b (recall that 41 plots were used for training the ITD algorithm).

All calculations were done using R (R Core Team, 2017). The HT-like estimator was implemented in function `HTest` of package `lmfor` (Mehtätalo, 2019), which further utilizes functions of `spatstat` (Baddeley et al., 2015).

4. Results

The number of neighbors k in k-NN needs to be chosen before prediction. This choice can be done with leave-one-out cross-validation in the training data, using the minimum RMSE as the criterion. This value for the training data set is $k = 7$. We concentrate on the estimates produced by this number of neighbors.

For the full validation data set, the lowest RMSE (24.1%) was produced by the HT-like estimator using a common fixed value of clari tuning parameter α (Eq. (2)) for all plots (Table 4). It produced a 51% improvement when compared to ITD, a 22% improvement when compared to the estimator with plot-level α, and a 13% improvement when compared to ABA. ABA produced lower RMSE than the HT-like estimator with plot-level α and has also the lowest ME value. The estimates produced by ITD were accurate in low-density plots but underestimates on denser plots (Figures 3 and 4). The HT-like estimator moves the point scatter closer to the identity line (Fig. 3) and zero line (Fig. 4) but seems still underestimate stand density on the plots with highest densities. ABA does not show similar underestimation but shows larger variance of estimation errors especially for high true densities. The common estimate of α for all plots of the validation data set was 0.495, which implies erosion with buffer $0.495r_i$. The plot-specific estimates of α, based on the functional k-NN ranged within $[0.395, 0.879]$.

move2 To analyze the potential of the method, it is interesting to look also at the performance of the method using optimal values of α. The optimal values of α for each validation plot lead to RMSE% of 1.0 and ME% of 0.3 and the estimates of α ranged within $[-0.263, 1.000]$. The errors are always overestimates and caused by few plots where the ITD algorithm found so many trees that $\alpha = 1$ led to an overestimate. An optimal common α for all validation plots is 0.3548, which leads to RMSE% 19.4 and ME% 0.2. These figures imply that using a common α for all plots, the RMSE% cannot be below 19.4, and the estimate with lowest RMSE% would also be (practically) unbiased. If each plot is allowed to have its own α, the theoretical lower bound for RMSE% is 1 %.

The quality of the applied 7-plot neighbourhood has a clear effect on the estimation accuracy. For example, among the 20 validation plots with the best neighbourhood, the RMSE of the most accurate estimator was approximately 15%, which is very low compared to the value 24% in the whole data set (Figure 5 and Table 4). In the reduced sets of validation plots, the HT-like estimator with fixed α shows lowest RMSE. The second best is the HT-like estimator with plot-level alpha if 40 plots with the best neighbourhood are considered, but ABA when plots with less tight neighbourhood are also included in validation. The common estimates of α for the 40 and 20 plots with best neighbourhood were $\widehat{\alpha} = 0.551$ and $\widehat{\alpha} = 0.497$, respectively, and the plot-specific estimates ranged within $[0.395, 0.762]$ and $[0.395, 0.617]$. The optimal common values of α were $\alpha = 0.387$ and $\alpha = 0.421$, leading to RMSE% of 12.87 and 12.39, respectively, and not meningful bias.

Both HT-like estimators perform better than ABA among plots dominated by Scots pine and Norway spruce (Table 5), but worse among the plots dominated by deciduous trees. Also here, a fixed α leads to lower RMSE and absolute ME than a plot-level α, and in all cases ABA has lowest absolute ME.

Table 4. The results of stand density estimation in the validation data set when sequentially smaller subsets that are closer to the training data are used for validation. ITD: estimator based on the number of detected trees, ABA: the area based approach estimator, HT: the HT-like estimator, where 7 nearest neighbors have been either used to choose α for every plot (plot-level α), or the median of the estimated α is used for every plot (fixed α). The column "n" indicates the number of validation plots used to calculate the errors.

n	method	RMSE	RMSE%	ME	ME%
111	ITD	530.0	48.8	-384.5	-35.4
	ABA	301.6	27.8	-8.5	-0.8
	HT, plot-level α	335.3	30.9	-151.2	-13.9
	HT, fixed α	262.2	24.1	-119.7	-11.0
40	ITD	559.9	40.7	-479.4	-34.8
	ABA	313.4	22.8	-19.6	-1.4
	HT, plot-level α	278.7	20.3	-169.4	-12.3
	HT, fixed α	264.4	19.2	-188.7	-13.7
20	ITD	535.7	40.4	-451.1	-34.0
	ABA	310.9	23.5	-31.4	-2.4
	HT plot-level α	208.2	15.7	-78.0	-5.9
	HT, fixed α	191.6	14.5	-94.1	-7.1

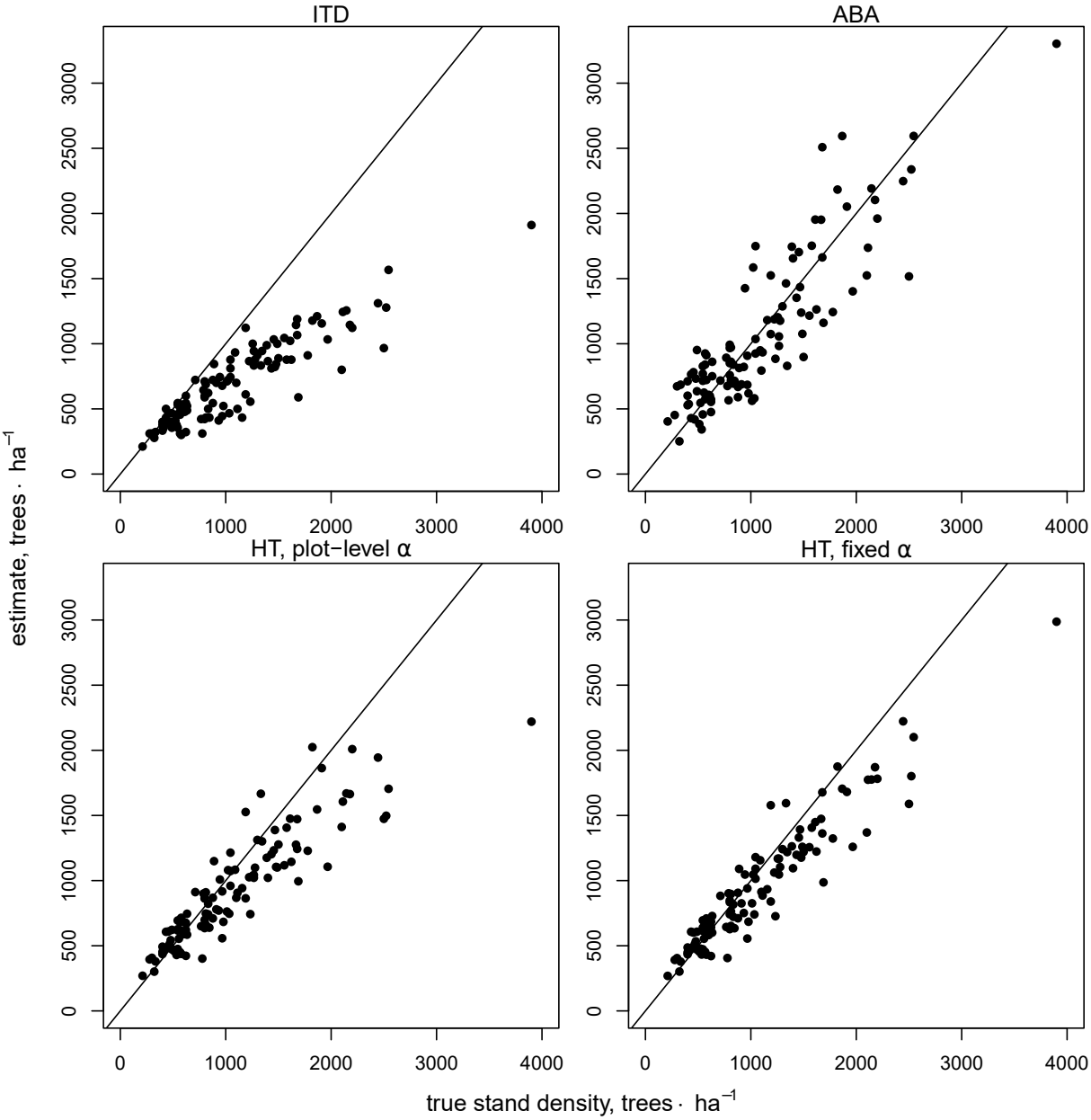
Table 5. The results of stand density estimation in the validation data set when plots are divided into subsets based on their dominant species. ITD: estimator based on the number of detected trees, ABA: the area-based approach estimator, HT: the Horvitz–Thompson-like estimator with 7 neighbors (plot-level α), or using median α over all training plots (fixed α).

dominant	method	RMSE	RMSE%	ME	ME%
pine	ITD	506.3	46.4	-377.1	-34.6
	ABA	315.4	28.9	-77.6	-7.1
	HT, plot-level α	304.6	27.9	-145.2	-13.3
	HT, fixed α	264.9	24.3	-126.9	-11.6
spruce	ITD	494.4	47.6	-350.8	-33.8
	ABA	305.9	29.5	87.8	8.5
	HT, plot-level α	296.3	28.5	-134.4	-12.9
	HT, fixed α	222.1	21.4	-94.5	-9.1
deciduous	ITD	652.9	55.3	-478.4	-40.5
	ABA	250.5	21.2	-39.2	-3.3
	HT plot-level α	467.2	39.6	-203.9	-17.3
	HT, fixed α	328.7	27.8	-156.6	-13.3

5. Discussion

This study explored the problem of determining stand density based on an individual tree detection algorithm that utilizes aerial laser scanning data. We generalized the HT-like estimator of Kansanen et al. (2016) to a situation where a tree can remain undetected even when the center point of the tree is not covered by a crown of a larger tree. We also explored for the first time how the method performs in an external validation data and how the tuning parameter should be determined in such situation. Two strategies were presented for the estimation of the tuning parameter: using a common fixed tuning parameter obtained as the median of plot-specific optimal values in the training data, and a value specified separately for each plot by using functional k-NN method. The results were compared to an

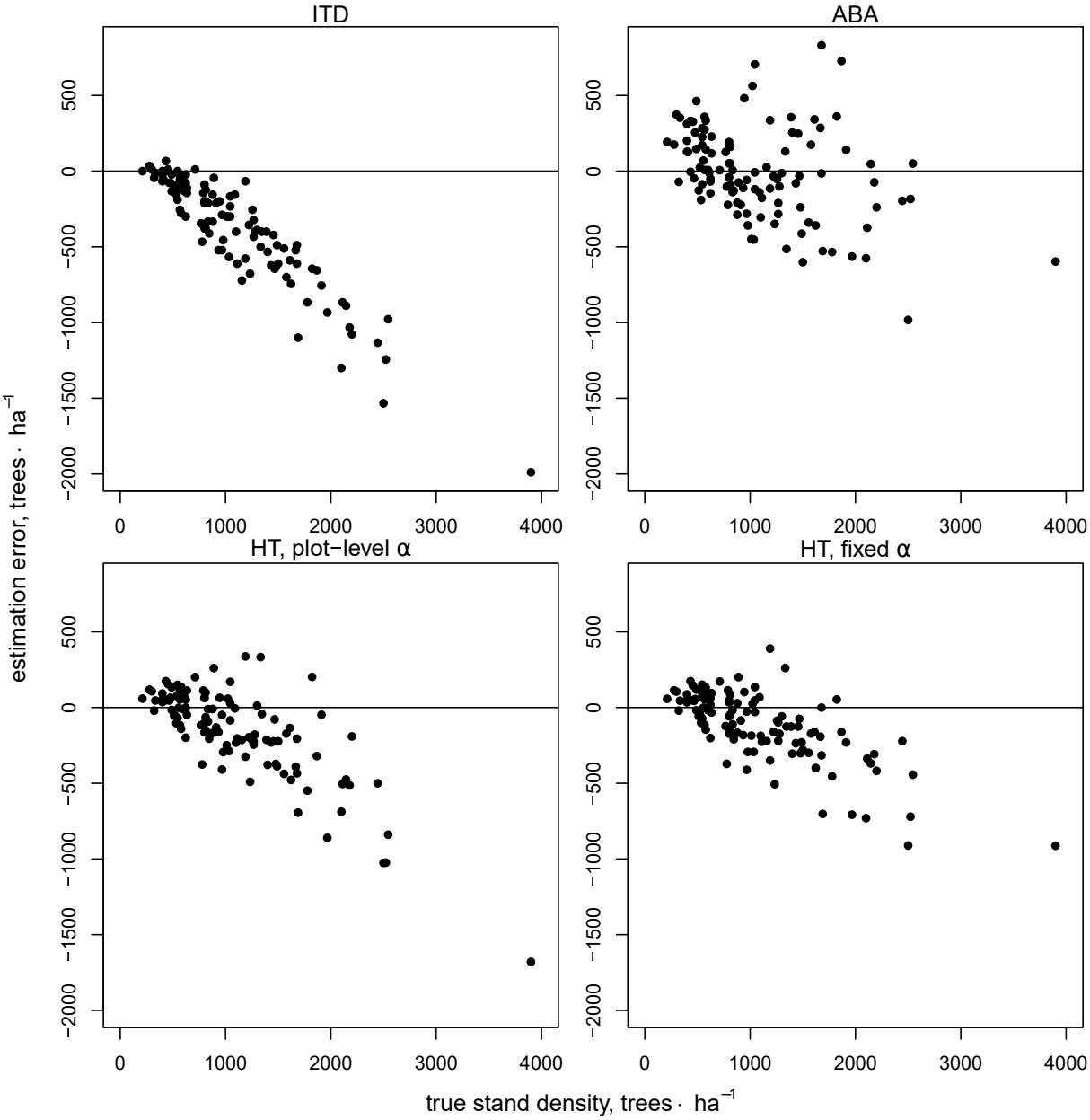
Fig. 3. The estimated stand densities vs. the true stand densities.



area-based approach. Furthermore, the effect of the quality of the neighbourhood in the results was analyzed in a novel way that might be generalized to other studies as well.

The HT-like stand density estimator performed better than the area-based method and the individual tree detection approach without correction for hiding trees. The best way to find the tuning parameter α was to use a the median over the optimal plot-level values of α. That method produced lower RMSE in the full validation data than the benchmark method and the HT-like estimator with plot-level α. Surprisingly, predicting α separately for each plot by using the functional k-NN did not perform as

Fig. 4. The estimation errors (estimated - observed stand density) vs. the true stand densities.



well as we expected: it provided noticeably larger RMSE than the best method, and also higher RMSE than the area-based method in the full validation data set. comment2c However, when the validation data set was restricted to 40 or less plots with the best neighbourhood, the accuracy of the HT-like estimator was much better than in the full data, and both approaches for selecting α performed rather similarly, as demonstrated in Figures 5 and Table 4. These results indicate that our training data of 36 plots could not provide a good neighbourhood of 7 plots for all validation plots and with larger training data set that fills the function space where the distances between plots are calculated well, then the

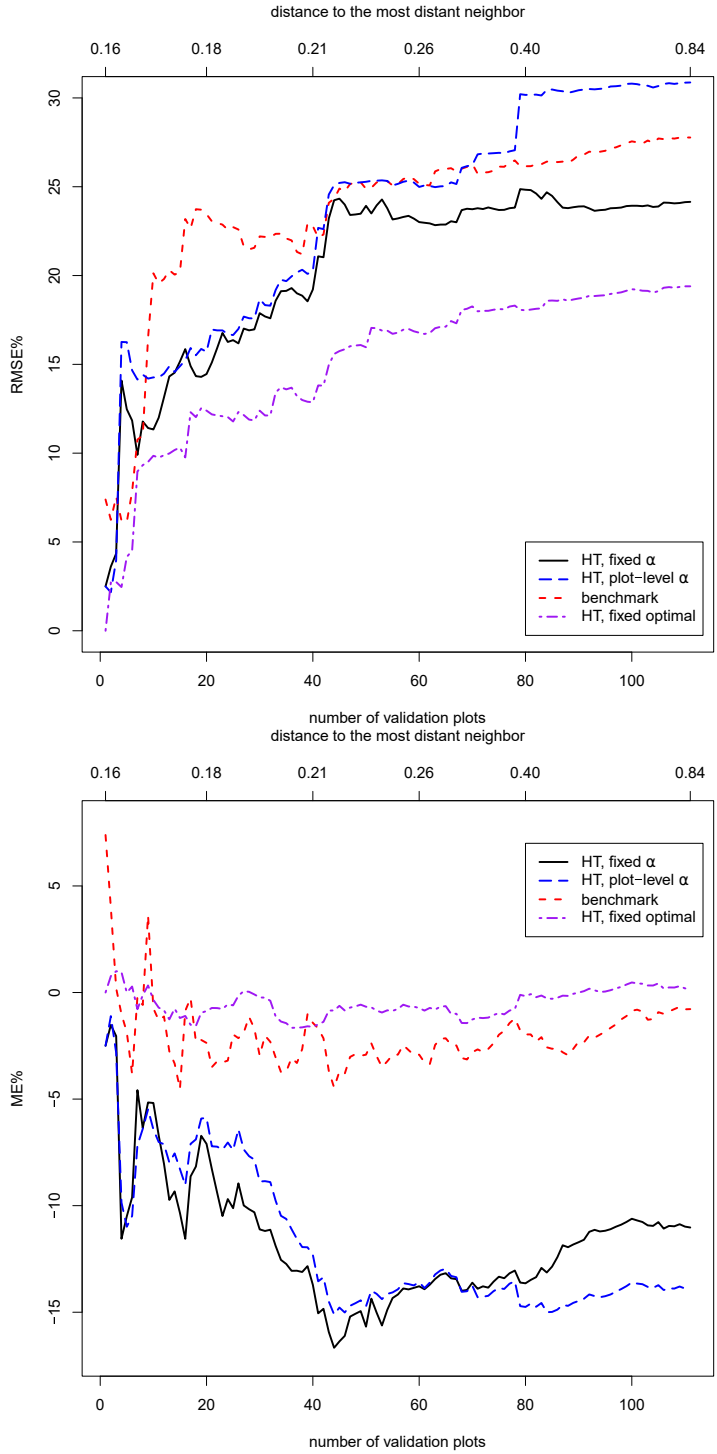


Fig. 5. The relative root-mean-squared error (top) and relative mean error (bottom)) of the HT-like estimator as a function of number of validation plots, with 7 nearest neighbors used to choose α. Also results using the benchmark method and using an optimal fixed alpha for all plots are shown.

HT-like estimator could lead to relative RMSE of 15% for the number of stems.

If optimal values of α could be used on each plot separately, the number of stems could be estimated without meaningful errors. If an optimal common valuer of α were used for all plots, practically unbiased estimates of stand density would be obtained and the RMSE% would be approximately 20 in the full data and 15 in the reduced data sets with 40 or 20 plots with best neighbourhood. Comparing these values to the RMSE% and bias based on estimates of α from the training data shows that the estimates based on the training data do not reach optimal performance, especially with the plot-level α. It seems that the difference in the size distributions of detected trees does not include such information of plot-specific properties (such as the spatial pattern of tree locations) that would lead to significantly better choices of α than a common estimate for all plots.

The HT-like estimator produced better RMSE values than ABA in plots dominated by coniferous trees, both with plot-level and fixed α (Table 5), and the deciduous plots are a major source of estimation errors for the HT-like estimator. comment1d An explanation may be that the deciduous tree crowns may have such an irregular shape that the ITD algorithm does not perform as well. In addition, the detection conditions applied by our method may not be as well justified for deciduous tree crowns. Also the small sample size of deciduous plots may have an effects especially because the training data of ABA had more deciduous plots.

A functional k-NN method based on seven neighbors was used in choosing the plot-level value of α. The distance between the empirical cumulative distribution functions of detected crown radii was used as the measure of similarity between a target plot and its neighbors. Also other measures were considered, such as empirical cumulative distribution functions of detected tree heights and laser return heights, but neither of these options produced noticeably better results. As most of the statistics used to study the behaviour of random sets are functional – for example the spherical contact distribution function and set covariance function used in Diggle (1981) – their use in the functional k-NN method is possible. In addition to the K-S metric (L^∞) other L^p metrics

$$d_{ij} = \sqrt[p]{\int (F_i(x) - F_j(x))^p dx} \tag{9}$$

could be used to measure the similarity. L^2 metric was considered, but it did not produce better results than L^∞. A *global* choice of $k = 7$ Ferraty and Vieu (2006) was used. Also *local* choice was evaluated, where for every validation plot the closest neighbor from the training set is selected. Then the optimal k for estimating stand density for that closest neighbor would be chosen as the number of neighbors used to estimate the stand density in the validation plot. This approach attempts to take advantage of the structure of the function space used for k-NN. However, the local choice of k did not lead to considerable improvement in results, and the chosen values of k did not correspond well to the optimal values of k for the validation plots. However, the local choice of k could work better with a larger training data.

One might question the necessity of the HT-like estimation step in our k-NN scheme as it is possible to do also a more traditional k-NN estimation (Ferraty and Vieu, 2006): after k closest neighbors have been chosen, a weighted mean of the stand densities in those neighbors could be calculated and used as an estimate. The weights are usually the distances to the neighbors manipulated by some kernel function, for example a quadratic kernel, to give larger weights for the closer neighbors and smaller weights for the more distant ones. This procedure was tested. Again, the global number of neighbors ($k = 7$) was chosen via leave-one-out cross-validation in the training data. Arranging the results again to the chain of subsets induced by the distances of the 7th neighbors, RMSE% values in all of the subsets were over 100, indicating that the HT-step was necessary.

Our training and validation sets are different in several aspects. Even though the means of stand density, quadratic mean diameter and basal area are quite similar, there is more variation and larger

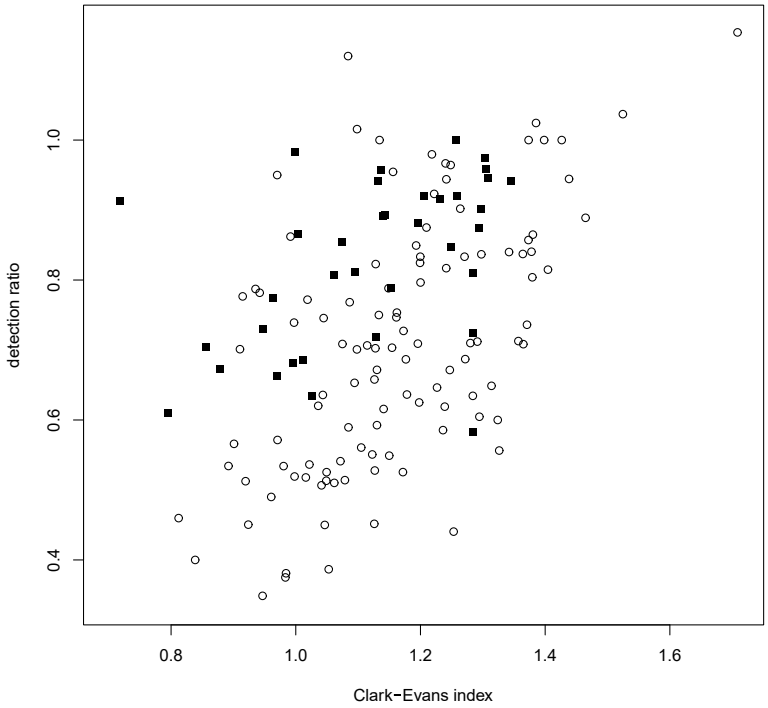
range for these values in the validation data (Tables 2 and 3). The spatial scale of the data sets is also different: the area of the convex hull containing all 79 plots of the training data set is roughly 550 ha, whereas in the validation data set the area is roughly 44500 ha. This difference in scales could lead to differences in the scale of spatial variation. From the ITD standpoint, the plot-level detection ratios also differ: in the training data set, these range from 58% to 100% of trees detected, whereas in the validation data set the range is from 35% to 115% (notice that these ratios are just simple ratios of numbers of trees given by ITD vs. the number of field measured trees). Optimal plot-level values of α are connected to these detection ratios: the fewer trees that have been detected, the smaller α needs to be to minimize the estimation error. The correlations between the detection ratios and optimal values of α are 0.87 and 0.84 in training and validation data sets, respectively. The values of α reach in both data sets the upper bound 1, whereas the lower bound is 0.16 in the training data set and -0.26 in the validation data set. If the validation set is chosen from the full validation set in such a way that the detection ratios fall to the range that is present in the training data, leading to a validation set of 72 plots, and α is estimated by minimizing RMSE over the whole training data, then the HT-like estimator achieves a RMSE% 18.8, which is much lower than compared to RMSE% 28.9 in the full validation data set of 111 plots. For comparison, ABA produces RMSE% 28.9 in this smaller validation set with 72 plots. It seems that if the training and validation sets have similar detection ratios the HT-like estimator produces good results even without k-NN. Of course, the detection ratio for a new data set meant for estimation is not known and hence this is not a viable strategy for choosing a similar training set in a prediction situation.

varianceThe variance of Horvitz-Thompson estimator is known (Horvitz and Thompson, 1952) and one might therefore consider application of it also here in a similar way as it was applied in the TLS context by Kansanen et al. (2021). However, in the TLS context, trees remain unobserved if they are located behind each other. With approximately circular tree stems and high-density TLS device, one can rather accurately derive the plot geometry based on observed tree stem diameters and locations. Therefore, the estimation errors are mainly caused by sampling error, which can be well approximated by the general formulas for HT-estimator under complete spatial randomness. In the aerial case, additional uncertainty is caused (at least) by irregular tree size and empirical estimation of α. Analysis and derivation of the variance estimators is therefore rather complex and was left as a future work.

bayesalphaThe model formulation was justified by crown overlap. In our case, the prior distribution of the ITD algorithm was formulated so that tree crown is not allowed to overlap with the center points of other trees, which is consistent to choosing $\alpha = 0$. However, as we already mentioned in Section 3.2, the spatial structure of forest and commission and omission errors also affect the optimal value of α. Furthermore, the order of trees in terms of height is not exactly same as the order in terms of crown size, which causes further errors in estimation. Therefore, we suggest empirical estimation of α instead of a value based on the inter-tree distance restrictions implemented in the ITD algorithm.

To analyze the effect of spatial pattern on the results, Clark–Evans indices (Clark and Evans, 1954) were calculated using the field-measured tree locations to assess their spatial structure. These indices were also used to test the plots for complete spatial randomness (CSR) at significance level 0.05. Both data sets exhibit positive correlation between the detection ratios and the index values, 0.46 for the training data and 0.57 for the validation data set. Hence, larger portion of trees has been detected in plots where the trees are in a regular pattern than in plots where they are clustered. Correlations between the index values and plot-level optimal α values were similar, 0.41 and 0.56. As the HT-like estimator has an inbuilt assumption of CSR, one might assume that fitting α in CSR training plots and predicting in CSR validation plots would produce better results. However, the optimal α over the CSR training plots is 0.544, whereas in CSR validation plots it is 0.282. This difference results in severe underestimation and RMSE% 32.4. The effect of spatial structure – at least given by the Clark–Evans index – on detection seems to be different in the two data sets. This can also be seen in Fig. 6. Although

Fig. 6. The detection ratios of ITD and Clark–Evans indices of training (■) and validation (○).



the scatter plots in both data sets have a similar shape, on average the detection ratios in the training data set are higher than in the validation data set for similar index values.

comment1c We restricted α within the range -1 to 1, which is justified if we assume that detection and non-detection of a tree is related to the overlap of the tree crown with the crowns of larger trees. However, on might expect that very small trees are hard to detect by ITD even if they are not in contact with the large tree crowns. This could justify a very small (negative) value of α for small trees to model the very small detectability, or using a constant detectability that does not depend on mathematical morphology of the larger tree crowns at all. On the other hand, if we let α to grow without limit, the HT-like estimator would equal the estimate given by ITD. This would make it possible for the HT-like estimator to give the best estimate in cases where ITD has detected all of the trees in a plot. These expansions would lead us from interpreting α as a parameter to model the detectability due crown overlap to just as a general tuning parameter based on a of a geometric transformation. The tuning parameter α could also be a function of crown size.

Another way of modifying the detectability without changing the interpretation of α would be adding other parameters to the model. For example, $p_\alpha(r)$ could be replaced in the estimator with $\theta p_\alpha(r)$ and bounding this value to the interval $(0, 1]$ to preserve its nature as a probability. One natural way of modelling θ would be to connect it to the spatial structure of the forest: for an area where the trees are CSR, the value of θ would be 1, for clustered patterns it would be less than 1, and for regular patterns greater than 1. The magnitude of the deviance from CSR would affect the value of θ. Estimation of θ could be based on a sequential spatial point process models, such as the one introduced in Penttinen and Ylitalo (2016) and applied to forestry by Yazigi et al. (2021). Modelling θ this way would remove the effect of spatial structure from the values of α. Of course, in addition to the detection conditions related to trees covering each other, α would still contain variation induced by other possible factors, such as scanning angles and other data gathering parameters. As discussed above and shown

in Fig. 6, different data sets can also differ on the way that the spatial structure influences detection. This way of modelling θ would also require estimating a statistic describing the spatial structure, such as the Clark–Evans index, from the remote sensing data.

6. Acknowledgements

This study was funded by the Academy of Finland (projects 295489, 250215, 310072, 323484, 310073 and 337655) and the strategic funding of the University of Eastern Finland.

References

Baddeley, A., E. Rubak, and R. Turner. 2015. Spatial Point Patterns: Methodology and Applications with R. Chapman and Hall/CRC Press, London. URL http://www.crcpress.com/Spatial-Point-Patterns-Methodology-and-Applications-with-R/Baddeley-Rubak-Turner/9781482210200/.

Breidenbach, J., E. Næsset, V. Lien, T. Gobakken, and S. Solberg. 2010. Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. Remote Sensing of Environment 114(4):911–924. URL https://www.sciencedirect.com/science/article/pii/S0034425709003599.

Buckland, S., D. Anderson, K. Burnham, J. Laake, D. Borchers, and L. Thomas ((eds.)) 2004. Advanced distance sampling. Oxford University Press.

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

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

Diggle, P.J. 1981. Binary mosaics and the spatial pattern of heather. Biometrics 37(3):531–539. URL http://www.jstor.org/stable/2530566.

Ene, L., E. Næsset, and T. Gobakken. 2012. Single tree detection in heterogenous boreal forests using airborne laser scanning and area-based stem number estimates. International journal of remote sensing 33(16):5171–5193.

Ferraty, F., and P. Vieu. 2006. Nonparametric Functional Data Analysis: Theory and Practice (Springer Series in Statistics). Springer-Verlag, Berlin, Heidelberg.

Flewelling, J. 2008. Probability models for individually segmented tree crown images in a sampling context. P. 284–294, *in* Proceedings of SilviLaser 2008, 8th International Conference on LiDAR Applications in Forest Assessment and Inventory, Hill, R., J. Rosette, and J. SuÃ¡rez, (eds.).

Horvitz, D.G., and D.J. Thompson. 1952. A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association 47(260):663–685.

Hou, Z., Q. Xu, J. Vauhkonen, M. Maltamo, and T. Tokola. 2016. Species-specific combination and calibration between area-based and tree-based diameter distributions using airborne laser scanning. Canadian Journal of Forest Research 46(6):753–765. URL https://doi.org/10.1139/cjfr-2016-0032.

Kansanen, K., P. Packalen, M. Maltamo, and L. Mehtätalo. 2021. Horvitz-Thompson-like estimation with distance-based detection probabilities for circular plot sampling of forests. Biometrics 77(2):715-728. URL https://doi.org/10.1111/biom.13312.

Kansanen, K., J. Vauhkonen, T. Lähivaara, and L. Mehtätalo. 2016. Stand density estimators based on individual tree detection and stochastic geometry. Canadian Journal of Forest Research 46(11):1359–1366. URL http://dx.doi.org/10.1139/cjfr-2016-0181.

Kansanen, K., J. Vauhkonen, T. Lähivaara, A. Seppänen, M. Maltamo, and L. Mehtätalo. 2019. Estimating forest stand density and structure using Bayesian individual tree detection, stochastic geometry, and distribution matching. ISPRS Journal of Photogrammetry and Remote Sensing 152:66–78. URL https://www.sciencedirect.com/science/article/pii/S0924271619301030.

Kukkonen, M., M. Maltamo, L. Korhonen, and P. Packalen. 2021. Fusion of crown and trunk detections from airborne uas based laser scanning for small area forest inventories. International Journal of Applied Earth Observation and Geoinformation 100:102327. URL https://www.sciencedirect.com/science/article/pii/S0303243421000349.

Lähivaara, T., A. Seppänen, J. Kaipio, J. Vauhkonen, L. Korhonen, T. Tokola, and M. Maltamo. 2014. Bayesian approach to tree detection based on airborne laser scanning data. IEEE Transactions on Geoscience and Remote Sensing 52(5):2690–2699.

Luostari, T., T. Lähivaara, P. Packalen, and A. Seppänen. 2018. Bayesian approach to single-tree detection in airborne laser scanning – use of training data for prior and likelihood modeling. Journal of Physics: Conference Series 1047:012008.

Maltamo, M., K. Eerikäinen, J. Pitkänen, J. Hyyppä, and M. Vehmas. 2004. Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. Remote Sensing of Environment 90(3):319–330. URL https://www.sciencedirect.com/science/article/pii/S0034425704000136.

Maltamo, M., E. Næsset, and J. Vauhkonen ((eds.)) 2014. Forestry applications of airborne laser scanning - Concepts and case studies. Springer, Dorcdrecht.

Mehtätalo, L. 2006. Eliminating the effect of overlapping crowns from aerial inventory estimates. Canadian Journal of Forest Research 36(7):1649–1660.

Mehtätalo, L. 2019. lmfor: Functions for Forest Biometrics. URL https://CRAN.R-project.org/package=lmfor. R package version 1.3.

Mehtätalo, L., and J. Lappi. 2020. Biometry for Forestry and Environmental Data: with Examples in R. 1st edition. Chapman & Hall / CRC Press, Boca Raton. 411 p.

Melville, G.J., A.H. Welsh, and C. Stone. 2015. Improving the efficiency and precision of tree counts in pine plantations using airborne lidar data and flexible-radius plots: Model-based and design-based approaches. Journal of Agricultural, Biological, and Environmental Statistics 20(2):229–257. URL https://doi.org/10.1007/s13253-015-0205-6.

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

Næsset, E. 2007. Airborne laser scanning as a method in operational forest inventory: Status of accuracy assessments accomplished in scandinavia. Scandinavian Journal of Forest Research 22(5):433–442. URL https://doi.org/10.1080/02827580701672147.

Packalén, P., and M. Maltamo. 2007. The k-msn method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. Remote Sensing of Environment 109(3):328–341. URL https://www.sciencedirect.com/science/article/pii/S0034425707000387.

Penttinen, A., and A.K. Ylitalo. 2016. Deducing self-interaction in eye movement data using sequential spatial point processes. Spatial Statistics 17:1–21. URL https://www.sciencedirect.com/science/article/pii/S2211675316300021.

Peuhkurinen, J., M. Maltamo, and J. Malinen. 2008. Estimating species-specific diameter distributions and saw log recoveries of boreal forests from airborne laser scanning data and aerial photographs: a distribution-based approach. Silva Fennica 42(4):625–641.

R Core Team. 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Strub, M.R., and N. Osborne. 2021. Correcting tree count bias for objects segmented from lidar point clouds. Mathematical and Computational Forestry & Natural-Resource Sciences (MCFNS) 13(1):29–35.

Thompson, S.K. 2012. Sampling. 3rd edition. Wiley, New York. 446 p.

Vauhkonen, J., L. Ene, S. Gupta, J. Heinzel, J. Holmgren, J. Pitkänen, S. Solberg, Y. Wang, H. Weinacker, K.M. Hauglin, V. Lien, P. Packalén, T. Gobakken, B. Koch, E. Næsset, T. Tokola, and M. Maltamo. 2011. Comparative testing of single-tree detection algorithms under different types of forest. Forestry: An International Journal of Forest Research 85(1):27–40. URL https://doi.org/10.1093/forestry/cpr051.

Vauhkonen, J., M. Maltamo, R. McRoberts, and E. Næsset, 2014. Introduction to forestry applications of airborne laser scanning. *in* Forestry applications of airborne laser scanning: concepts and case studies, *Managing Forest Ecosystems*, volume 27, Maltamo, M., E. NÃ¦sset, and J. Vauhkonen, (eds.), P. 1–16. Springer.

Yazigi, A., A. Penttinen, A.K. Ylitalo, M. Maltamo, P. Packalen, and L. Mehtätalo. 2021. Modeling Forest Tree Data Using Sequential Spatial Point Processes. Journal of Agricultural, Biological end Environmental statistics (Early view) URL https://doi.org/10.1007/s13253-021-00470-2.