Author(s): Mehtätalo, Lauri; Yazigi, Adil; Kansanen, Kasper; Packalen, Petteri; Lähivaara, Timo; Maltamo, Matti; Myllymäki, Mari; Penttinen, Antti

Title: Estimation of forest stand characteristics using individual tree detection, stochastic geometry and a sequential spatial point process model

Year: 2022

Version: Published version

Please cite the original version:

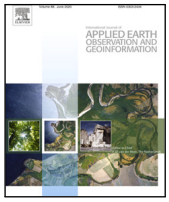
Mehtätalo, L., Yazigi, A., Kansanen, K., Packalen, P., Lähivaara, T., Maltamo, M., Myllymäki, M., & Penttinen, A. (2022). Estimation of forest stand characteristics using individual tree detection, stochastic geometry and a sequential spatial point process model. International Journal of Applied Earth Observation and Geoinformation, 112, Article 102920. https://doi.org/10.1016/j.jag.2022.102920

International Journal of Applied Earth Observations and Geoinformation 112 (2022) 102920

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Estimation of forest stand characteristics using individual tree detection, stochastic geometry and a sequential spatial point process model

Lauri Mehtätalo [a,*], Adil Yazigi [b], Kasper Kansanen [b], Petteri Packalen [a], Timo Lähivaara [c], Matti Maltamo [d], Mari Myllymäki [e], Antti Penttinen [f]

[a] Natural Resources Institute Finland (Luke), Yliopistokatu 6, 80100 Joensuu, Finland
[b] School of Computing, University of Eastern Finland, Postal Box 111, 80101 Joensuu, Finland
[c] Department of Applied Physics, University of Eastern Finland, Postal Box 1627, Kuopio, Finland
[d] School of Forest Sciences, University of Eastern Finland, Postal Box 111, 80101 Joensuu, Finland
[e] Natural Resources Institute Finland (Luke), Latokartanonkaari 9, 00790 Helsinki, Finland
[f] University of Jyväskylä, PO Box 35, FI-40014, Finland

ARTICLE INFO

Keywords:
Forest inventory
Airborne Laser Scanning
Horvitz–Thompson-like estimator
Stand density
Tree height

ABSTRACT

Airborne Laser Scanning (ALS) results in point-wise measurements of canopy height, which can further be used for Individual Tree Detection (ITD). However, ITD cannot find all trees because small trees can hide below larger tree crowns. Here we discuss methods where the plot totals and means of tree-level characteristics are estimated in such context. The starting point is a previously presented Horvitz–Thompson-like (HT-like) estimator, where the detectability is based on the larger tree crowns and a tuning parameter α that models the detection condition. We propose a new method which is based on modeling the spatial pattern of hidden tree locations using a sequential spatial point process model, with a tuning parameter θ. We also explore whether the variability of the tuning parameters α and θ can be predicted using ALS features to improve the predictions. The accuracy of stand density, dominant height and mean height is used as comparison criteria in a cross-validation procedure. The HT-like estimator with empirically estimated tuning parameter α performed the best. The overall performance of the new method was comparable. The new method was computationally less demanding, which makes it attractive for practical use.

1. Introduction

Airborne Laser Scanning (ALS) results in point-wise measurements of canopy height, which allow detection of individual tree crowns using individual tree detection (ITD). The detected trees can be ordered according to their shortest distance to the sensor, and trees may hide behind earlier trees; we call larger trees "earlier" in this context. Here we focus on the estimation of totals and means of all forest trees, including those that are hidden below larger tree canopies. The missed trees have a direct impact on the total estimates, such as stand density (N, trees per ha). Also the means (e.g. mean height) may be biased since small trees are missed more easily than large trees (Vauhkonen et al., 2010).

Mehtätalo (2006) proposed a Horvitz–Thompson-like (HT-like) estimator for the estimation stand density based on ITD. He used a sequential construction and assumed that a three is detected if the crown center is not covered by larger tree crowns. If tree locations follow complete spatial randomness, the probability to detect tree i can be computed using union of detected crown segments. The method is an application of stochastic geometry (Chiu et al., 2013), which analyzes sets generated by marked point processes (Illian et al., 2007). Kansanen et al. (2016, 2019, 2022) generalized the method by assuming that a tree is detectable if it does not hit the set A_i, which is obtained by subtracting/adding a buffer of width $|\alpha|r_i$ from/to the union of larger crown segments. The tuning parameter α ($-1 \leq \alpha \leq 1$) needs to be known beforehand, e.g. based on the properties of the ITD algorithm or using empirical training data.

The constructions described above assume that the density of trees of a given size is similar in the visible and hidden parts of the sample plot. However, the hidden parts have worse growing conditions e.g. in terms of sunlight availability. In addition, foresters may have manipulated the point pattern of tree locations e.g. through thinnings.

* Corresponding author.
E-mail addresses: lauri.mehtatalo@luke.fi (L. Mehtätalo), adil.yazigi@uef.fi (A. Yazigi), kasperkansanen@gmail.com (K. Kansanen), petteri.packalen@luke.fi (P. Packalen), timo.lahivaara@uef.fi (T. Lähivaara), matti.maltamo@uef.fi (M. Maltamo), mari.myllymaki@luke.fi (M. Myllymäki), antti.k.penttinen@jyu.fi (A. Penttinen).

https://doi.org/10.1016/j.jag.2022.102920
Received 31 March 2022; Received in revised form 12 July 2022; Accepted 12 July 2022
Available online 29 July 2022

Table 1
Mean, standard deviation, minimum and maximum of stand density (N), quadratic mean diameter (QMD) and basal area (BA) in Liperi. The full data contains 110 field plots.

Attribute	mean	sd	min	max
N, stems ha^{-1}	1085.99	625.18	211.11	3900
QMD, cm	18.48	5.92	8.23	37.18
BA, m^2 ha^{-1}	23.87	7.80	7.84	46.19
Mean height, m	15.86	4.30	8.94	26.61
Dominant height, m	22.12	4.98	12.83	34.28

Empirical estimation of α takes this problem into account at least partially, but a better solution might be explicit modeling of the local stand density in the hidden parts. An intuitive model for this purpose is the finite sequential spatial point process model (SSPP), which was recently proposed for forest data by Yazigi et al. (2021). In SSPP, the points (tree locations) within an observation window (sample plot) are ordered from the earliest (largest) to the latest (smallest), and the earlier points have an effect on the locations of the latter points. They assumed that the order in terms of tree age can be sufficiently well approximated by the order in terms of tree diameter. The first location of the sequence is generated from the homogeneous Poisson process. For the latter trees, homogeneous Poisson process is also used for generating a proposal location. If the proposal is within distance R from any of the earlier locations, the proposal is accepted with probability θ. Otherwise it is accepted with probability $1-\theta$.

The SSPP model of Yazigi et al. (2021) allows the generation of different spatial point patterns depending on the values of the parameter θ. For $\theta = 1/2$, the SSPP model is a homogeneous Poisson process. Values $\theta < 1/2$ and $\theta > 1/2$ correspond to clustered and regular spatial point patterns respectively. If the parameter R is selected so that it corresponds to the detection condition in the HT-like estimator, the ratio $\frac{\theta}{1-\theta}$ describes the ratio of stand densities in the hidden and visible parts of the forest, leading to a straightforward extension of Kansanen et al. (2016). An additional possible extension is modeling the parameter θ of Yazigi et al. (2021) or/and the parameter α of Kansanen et al. (2016) empirically using predictors based on the ALS data and justified by point process theory (Häbel et al., 2021).

In this paper, we extend the HT-like estimator of Kansanen et al. (2016) so that it takes into account the interaction of trees through the SSPP model of Yazigi et al. (2021), and compare the extended estimator with the previously published one. We will also evaluate whether the variability of the tuning parameters can be explained by predictors derived from the ALS point cloud to improve the accuracy of the extended HT-like estimator. The evaluation with empirical data is based on a leave-one-out cross-validation approach, using the accuracy of stand density, dominant height and mean height as performance criteria. This paper extends the analysis of Mehtätalo et al. (2021) by (1) generalizing the estimator for plot total and mean, (2) exploring the possibility to explain the tuning parameters using laser features, (3) evaluating based on cross-validation and (4) presenting a more detailed evaluation using dominant and mean heights of the plot.

2. Material

We use 110 fixed-area plots of size 30 by 30 m from North Carelia, Finland (Table 1). Scots pine (Pinus sylvestris L.) is the dominant species on 39%, Norway spruce (Picea abies [L.] Karst.) on 43%, and birch species (Betula pendula Roth and B. Pubescens Ehrh.) on 18% of the plots. The trees were measured for location, diameter (DBH), and height on the field. ALS data were acquired using an Optech Titan instrument on July 2–10, 2016 using the scanning altitude of 850 m above ground level, the scanning half angle of 20 degrees, the pulse repetition frequency of 250 kHz and the sampling density of 4.8 pulses per m^2. We used only the 1064 nm channel in this study.

3. Methods

3.1. Individual tree detection

In the applied Bayesian framework-based ITD algorithm, individual trees are modeled as rotationally symmetric objects. For each tree the *maximum a posteriori* (MAP) estimates of their location, crown radius and height, the lower limit of the living crown, and crown shape parameter are determined. For a more detailed discussion about the ITD algorithm, we refer to Lähivaara et al. (2014) and Kansanen et al. (2022).

The trees found by ITD are those that are visible to the scanner. As we discussed in the introduction, some trees may be hidden below other tree canopies, which means that plot totals (of stand density, standing volume or biomass, for example) are underestimated. Furthermore, because the hidden trees are on average smaller than the visible trees, the means of tree-specific variables will be overestimated. To adjust the estimates based on ITD for these problems, we use a previously published HT-like estimator and a new variant of it, which are explained in the next subsections.

3.2. The basic estimator

We assume that ALS is used to detect the projections of tree crowns to the ground level C_i, where i is an index for tree. The n observed trees within a sample plot window W are ordered according to the distance to the sensor, which in practice means ordering in terms of height in decreasing order. Furthermore, assuming that crown area is a monotonic function of tree height, this corresponds to ordering in terms of $|C_i|$ from largest to smallest. Now, small (later) trees may hide behind larger (earlier) trees in the ordered sequence. An intuitive starting point to approximate the number of hidden trees is based on the assumptions that (i) a tree is detected if the center point is visible to the scanner and (ii) the density (trees per ha) in the hidden parts of W is similar to that in the visible part of W for each detected tree. Then the detectability for tree i is the proportion of the total area not covered by larger tree crowns (Mehtätalo, 2006; Kansanen et al., 2016, 2019)

$$\pi_i = 1 - \frac{\left|W \cap \bigcup_{j=1}^{i-1} C_j\right|}{|W|}.$$

We are interested in the plot total of the tree-level characteristic m_i, $\tau = \sum_{i=1}^{N} m_i$. For estimation, we use the Horvitz–Thompson-like estimator

$$\hat{\tau}_m = \sum_{i=1}^{n} \frac{m_i}{\pi_i}, \tag{1}$$

where the inclusion probability π_i is called *detectability* (Thompson, 2012; Kansanen et al., 2019). It is straightforward to implement estimator (1) using the ordered set of observed tree crowns. Notice that using $m_i = 1$ for all trees gives the estimator of the total number of trees, which we denote by $\hat{\tau}_1$. For the plot-level mean of the tree-level characteristic m, such as the mean height, we use the estimator (Thompson, 2012, p. 69)

$$\hat{\mu}_m = \frac{\hat{\tau}_m}{\hat{\tau}_1}. \tag{2}$$

Writing the estimator (1) as

$$\hat{\tau}_m = \sum_{i=1}^{n} m_i + \sum_{i=1}^{n} \left(\frac{1}{\pi_i} - 1\right) m_i \tag{3}$$

shows that it is the sum of the total of observed trees and an estimated total of hidden trees. Because π_i is smaller than or equal to 1, the multipliers $\left(\frac{1}{\pi_i} - 1\right)$ are non-negative. Therefore, the estimator adjusts the total of detected trees upwards, which is intuitive if the two above-mentioned assumptions on the detection condition and spatial pattern

of tree locations in the hidden parts of the plot are met, and no other nuisance effects are present. However, the spatial pattern of the small trees is affected by nearby larger trees. Furthermore, the ITD algorithm may miss also visible trees or may find trees that do not exist. The next subsections present adjustments for these shortcomings.

3.3. Adjusting for the detection condition

Kansanen et al. (2016, 2022) proposed an adjustment that allowed parameter α to control the detection condition. To estimate the detectability of tree i, a buffer of width αr_i is removed (a negative α means that a buffer is added) from the union of larger tree crowns, and the detectability is defined as the probability that the center point of tree i is within this buffered set. An intuitive range for α is the interval $[-1, 1]$; using $\alpha = 1$ means that the tree will be detected if any part of the crown disk is outside the union of larger tree crowns, whereas $\alpha = -1$ means that the tree crown should not overlap with larger tree crowns at all to make the tree detectable. However, there is no technical reason to restrict the value of α. Especially, if it is estimated empirically, it implicitly describes the density of the hidden trees below the larger ones, and no upper or lower bounds for α are necessary.

Formally, the detectabilities are defined as

$$\pi_i = \begin{cases} 1 - \frac{|W \cap [\bigcup_{j=1}^{i-1} C_j \ominus B(\alpha r_i)]|}{|W|}, & \alpha > 0 \\ 1 - \frac{|W \cap [\bigcup_{j=1}^{i-1} C_j \oplus B(|\alpha| r_i)]|}{|W|}, & \alpha < 0 \end{cases}, \tag{4}$$

where $\ominus$ is erosion and $\oplus$ is dilation and $B(r)$ is a disk of radius r (Chiu et al., 2013). To apply formula (4), parameter α needs to be known. In applications, a value based on the known properties of the applied tree detection algorithm could be used based on the restrictions of inter-tree distances used in the applied ITD algorithm. It would also be possible to remove such small, detected trees that are not detectable according to the applied detection condition. Such an approach might make the data more closely follow the model assumptions, which might improve the final estimation result. Alternatively, the parameter can be estimated empirically, as we do here.

3.4. Adjusting for the spatial pattern in the hidden parts of the plot

If the value of α is determined using the ITD algorithm, the estimator (1) with detectability (4) assumes that the density in the hidden parts (λ_h, trees per ha) is similar to the density in the visible part (λ_v, trees per ha) for each tree separately. That assumption may be unrealistic, therefore it might be better to allow $\lambda_h \neq \lambda_v$. However, we might assume that the ratio of densities in the hidden and visible parts is constant for all trees of a sample plot, regardless of tree size. An intuitive way to parameterize such assumption is to use parameter θ, $(0 \leq \theta < 1)$, such that

$$\frac{\lambda_h}{\lambda_v} = \frac{\theta}{1-\theta}.$$

Value $\theta = 0.5$ means that the density in the hidden part is similar to that in visible part, $\theta < 0.5$ means that the density in hidden parts is smaller than in the visible part (small trees avoid large trees), and $\theta > 0.5$ means that density in the hidden parts is larger than in the visible parts (large trees attract small trees). The HT-like estimator (3) can then be adjusted as follows:

$$\hat{\tau}_m = \sum_{i=1}^{n} m_i + \frac{\theta}{1-\theta} \sum_{i=1}^{n} \left(\frac{1}{\pi_i} - 1 \right) m_i \,. \tag{5}$$

For convenience, we rewrite the formula in the form that mimics the Horvitz–Thompson-like estimator

$$\hat{\tau}_m = \sum_{i=1}^{n} \frac{m_i}{\pi_i^*}, \tag{6}$$

where

$$\pi_i^* = \frac{1}{1 + \frac{\theta}{1-\theta}\left(\frac{1}{\pi_i} - 1\right)} = \frac{\pi_i - \theta\pi_i}{\theta + \pi_i - 2\pi_i\theta}.$$

However, π_i^* cannot be interpreted as an inclusion probability or detectability, because θ is a parameter of the spatial point process model of the tree population. To employ Eq. (6), parameter θ needs to be known. Notice that using $\theta = 0.5$ leads to estimator (3) and $\theta = 0$ gives $\hat{\tau}_m = \sum m_i$.

3.5. Estimating α and θ for a forest plot using field data

In this section, we present how the values of α and θ needed in estimators (3) and (6) can be estimated for sample plots using field-measured trees. These estimates are then used as true values in the subsequent analysis. Especially, we aim at predicting them using remote sensing data to enable practical application of estimators (1) and (5).

3.5.1. The N-matching method to estimate α or θ

Estimation of α or θ can be based on matching the field-measured values for the number of trees with the HT-estimate. For every plot where the field-measured tree count N_f is larger than the number of detected trees n, it is possible to find such value of α, conditional on a predefined value of θ, that $\hat{\tau}_1 = N_f$; the estimated tree count $\hat{\tau}_1$ is based on Eqs. (3) and (4). Corresponding solution can also be found for θ conditional on a predefined value of α.

The left-hand side of equation $N_f - \hat{\tau}_1(\alpha|\theta) = 0$ is a monotonic function of α. Therefore, it has a unique solution, which can be found e.g. by using a step-halving algorithm. We used implementation `lmfor::updown` in R (Mehtätalo and Kansanen, 2022). For θ, the exact solution to equation $N_f - \hat{\tau}_1(\theta|\alpha) = 0$ is

$$\hat{\theta} = \frac{N_f - n}{\sum_{i=1}^{n} \frac{1}{\pi_i} + N_f - 2n}.$$

Notice that N-matching utilizes only the field-measured tree counts, whereas the SSPP-method described below utilizes also the field-measured tree locations.

For plots where the number of detected trees is greater or equal to the number of field-measured trees, we considered $\alpha = \frac{R_{max}}{R_{min}}$ as the N-matching estimate. It is the smallest value of α that gives the minimum distance $\left|N_f - \hat{\tau}_1(\alpha|\theta)\right|$ and $\hat{\tau}_1 = n$ when trees are isolated. For θ, we used $\theta = 0$ as the N-matching estimate of θ when $n \geq N_f$, which always gives $\hat{\tau}_1 = n$. An alternative is to consider N-matching estimate to be undefined in those plots.

3.5.2. The SSPP-method to estimate θ

An alternative to N-matching for estimating θ for a forest plot is an adjusted version of the SSPP model of Yazigi et al. (2021). The model is based on ordering trees from largest to smallest. The model assumes that the location of the first tree of the sequence is generated by a homogeneous spatial Poisson process. For the later trees, a proposal location is generated using the Poisson process. The proposal is accepted with probability θ if the model is within distance R (interaction radius) from any of the earlier trees and with probability $1 - \theta$ otherwise. The model of Yazigi et al. (2021) has parameters R and θ, which are common to all trees of the plot. The parameters were estimated using maximum likelihood. Here we use the predicted crown radius as a predefined tree-specific value of R, which was predicted using tree DBH. Therefore, only parameter θ_i was estimated when fitting the SSPP model for forest plot i.

For prediction of crown radius, a data set of ITD-based crown radii and field-measured tree DBH was constructed by manually matching the ITD data with field-measurements. Only sure matches were included in the data set. Thereafter, a mixed-effect model

$$R_{ij} = \beta_0 + \beta_1 D_{ij} + \beta_2 (D_{ij} - \bar{D}_i) + a_i + b_i D_{ij} + e_{ij} \,, \tag{7}$$

Table 2
The ALS features calculated on the basis of marginal distribution of return heights and spatial information from thresholded canopy height models (TCHM).

Feature	Symbol	x-values
Height features		
Quantile of first return heights	p_x	50, 80, 95 (%)
Proportion of first returns above a threshold	d_x	2, 10, 15 (m)
Spatial features		
Number of patches	no_x	20, 40, 60, 80 (%)
Average size of patches in number of pixels	means_x	20, 40, 60, 80 (%)
Standard deviation of size of patches	sdS_x	20, 40, 60, 80 (%)
Average number of same pixel type in a 4-neighborhood	SP_x	20, 40, 60, 80 (%)
Euler number for TCHMs	E_x	20, 40, 60, 80 (%)
Integrated deviation of F-function from theoretical reference	IF_x	20, 40, 60, 80 (%)
KL-type divergence of F-function from theoretical reference	DF_x	20, 40, 60, 80 (%)
Pairwise KL-type divergence of F-functions between TCHMs	F_x1 × 2	20, 40, 60, 80 (%)

was fitted, where R_{ij} is the maximum crown radius based on ITD and D_{ij} is the field-measured DBH of tree i on plot j; $\bar{D}_i$ is the mean of field-measured diameters and a_i and b_i are random zero-mean bivariate normal random effects on plot i. Term e_{ij} is residual error with variance $\sigma^2\left|\tilde{R}_{ij}\right|^{2\delta}$ (Mehtätalo and Lappi, 2020). The plot-level predictions of this model were used to predict crown radii for all field-measured trees. Thereafter, an estimate of parameter θ_i was obtained for each plot by fitting the SSPP model of Yazigi et al. (2021) separately for each plot, using maximum likelihood, conditioning on the tree-specific estimates of crown radii as the interaction radii.

3.6. Predicting α and θ without field data

A simple way for estimating α and θ without field-data is to use the mean of the plot-specific 'true' values in a training data set. However, better performance might be obtained by regressing them on conventionally used features of the ALS return heights or such features that tell about the spatial pattern of trees. Häbel et al. (2021) proposed a set of such metrics, based on the theory of spatial point patterns and stochastic geometry (Table 2). We explored if they improve the prediction of α and θ compared to using the mean of training data.

For each evaluated plot-level parameter estimate (α and θ based on N-matching and θ based on SSPP), we used stepwise model selection, as implemented in R-function `step` (R. Core Team, 2021). We used BIC as the model selection criterion instead of AIC, because AIC sometimes led to models with more than ten predictors, which was regarded as a too large number of predictors in the data of 110 plots (Harrell, 2001). However, for some of the characteristics, one or few plots had a very different value of the characteristics than the others, and those characteristics were not used in the analysis to avoid problems caused by overly influential observations. For parameter α, a linear model of form

$$\alpha_i = \boldsymbol{x}_i'\boldsymbol{\beta} + e_i \tag{8}$$

was used and fitted using OLS. For θ, a nonlinear model of form

$$\theta_i = \frac{1}{1+\exp\left(-\boldsymbol{x}_i'\boldsymbol{\beta}\right)} + e_i \tag{9}$$

was used to restrict the value of predicted θ to range $(0, 1)$; estimation was based on nonlinear OLS. The variable selection was first done using a linear model of form (8) for $\mathrm{logit}(\theta)$.

3.7. Analysis

3.7.1. The evaluated methods

To evaluate the estimator (6) using different values of α and θ, we start with the detected tree crowns. The following methods were thereafter used for determining α and θ in the estimator.

1. $\alpha = 0$, $\theta = 0$ (Using detected trees only)
2. $\alpha = 0$, $\theta = 0.5$ (Mehtätalo, 2006)
3. $\alpha = 0$, θ is the mean of the plot-level values based on the SSPP model.
4. $\alpha = 0$, θ is the mean of the plot-level values based on N-matching.
5. $\alpha = 0$, θ is predicted using ALS metrics. The estimates based on SSPP-model are used when fitting model for θ.
6. $\alpha = 0$, θ is predicted using ALS metrics. The estimates based on N-matching are used when fitting model for θ.
7. α is the mean of the plot-level values based on N-matching, $\theta = 0.5$
8. α is predicted using ALS metrics, $\theta = 0.5$
9. α is the mean of the plot-level values based on N-matching. The plot-level estimates of θ where further found by N-matching. The mean of these estimates was used in prediction.
10. Similar to method (9) but θ was based on a regression model of the plot-level estimates.

3.7.2. Cross-validation procedures

The evaluation for methods (1–8) was based on leave-one-out cross-validation, where each plot, in turn, was used for evaluation and all steps of training, including the stepwise variable selection of the predictive models, were carried out without the evaluation plot in question.

For methods 9 and 10, a two-step cross-validation approach was needed because the training was needed for both parameters in separate steps. To conduct the cross-validation for plot i, the training data of step 1 included all plots except for plot i. A further leave-one-out procedure was conducted within this training set to estimate θ for each plot j, $j = 1, \ldots, n-1$ using N-matching. When estimating θ for plot j, the α parameter was estimated as the mean of the $n-2$ training plots, i.e. in the training data of step 2 that excluded plots i and j. In method 9, the mean of the obtained $n-1$ estimates of θ in the training data was then used as a prediction for plot i. In method 10, the predictions of θ were based on a nonlinear regression model of form (9) fitted to the $n-1$ estimates of θ.

3.7.3. Comparison criteria

The methods were evaluated by estimating their ability to predict the stand density (trees per ha), dominant height (meters), and mean tree height (meters). Dominant height was determined as the mean of 100 tallest trees per hectare, i.e., the mean of 9 tallest trees on the 30 m by 30 m sample plot.

It is well-known that ALS-based estimates of tree height are underestimates because of penetration of laser pulses to the crown and missed treetops (Gaveau and Hill, 2003). For example, in the data set of matched field-measured and ITD-trees (see Section 3.5.2), the mean difference was 0.626 m. However, the matching data was biased towards larger trees, and the absolute difference seemed to increase with increasing height. Therefore, the bias was eliminated from the estimated heights using the model

$$h_{ij}^{(field)} = \beta_0 + \beta_1 h_{ij}^{(ALS)} + a_i + b_i h_{ij}^{(ALS)} + e_{ij},$$

Table 3
The predictors that remained in the predictive models of α_{Nmatch}, θ_{Nmatch} and θ_{SSPP} (Eqs. (8) and (9)) after stepwise variable selection in the full model fitting data, the model fitting statistics of the models and the summary (minimum–median–maximum) of the number of predictors in the 110 models fitted to the leave-one-out cross-validation data sets.

Response	Predictors	Residual standard error	R^2	# of predictors in LOO
$\widehat{\theta}_{Nmatch}$	p_95, d_10, IF_0.2	0.117	0.217	2 - 3 - 10
$\widehat{\theta}_{SSPP}$	no_0.8, DF_0.2	0.111	0.139	2 - 2 - 7
$\widehat{\alpha}_{Nmatch}$	p_50, p_80, d_10, DF_0.2, F_84, F_82	0.236	0.297	4 - 6 - 9

Table 4
The cross-validation results about the accuracy of stand density for all plots on the left and for plots with highest field-measured stand density on the right ($N > 114$, 37 plots). The figure ME/se in parentheses is the ratio between mean error and the standard error of mean.

			All plots				Plots with highest N			
			ME		RMSE		ME		RMSE	
Method	α	θ	1/ha (ME/se)	%	1/ha	%	1/ha (ME/se)	%	1/ha	%
1	0	0	−387.3 (−11.1)	−35.5	532.4	48.8	−755.9 (−12.8)	−42.1	834.9	46.5
2	0	0.5	621.7 (10.4)	57.0	881.5	80.9	1045.1 (8.0)	58.2	1307.2	72.8
3	0	$\tilde{\theta}_{SSPP}$	38.1 (1.6)	3.5	246.1	22.6	2.6 (0.0)	0.1	318.8	17.7
4	0	$\tilde{\theta}_{Nmatch}$	7.9 (0.4)	0.7	**236.9**	**21.7**	**−51.0 (−1.0)**	**−2.8**	**309.3**	**17.2**
5	0	$\check{\theta}_{SSPP}$	80.7 (2.8)	7.4	310.3	28.5	105.7 (1.5)	5.9	438.2	24.4
6	0	$\check{\theta}_{Nmatch}$	−10.5 (−0.4)	−1.0	263.4	24.2	−85.6 (−1.4)	−4.8	386.1	21.5
7	$\tilde{\alpha}_{Nmatch}$	0.5	−38.1 (−1.8)	−3.5	**219.5**	**20.1**	**−159.0 (−3.8)**	**−8.9**	**298.5**	**16.6**
8	$\check{\alpha}_{Nmatch}$	0.5	−27.2 (−1.0)	−2.5	274.4	25.2	−116.2 (−1.8)	−6.5	414.8	23.1
9	$\tilde{\alpha}_{Nmatch}$	$\tilde{\theta}_{Nmatch}$	−101.1 (−4.6)	−9.3	250.8	23.0	−267.1 (−6.3)	−14.9	368.2	20.5
10	$\check{\alpha}_{Nmatch}$	$\check{\theta}_{Nmatch}$	−19.9 (−0.6)	−1.8	320.9	29.4	−15.3 (−0.2)	−0.9	503.1	28.0

which was fitted to the same data set that was used for fitting the model shown in Eq. (7). Here $h_{ij}^{(field)}$ and $h_{ij}^{(ALS)}$ are the field-measured and ALS-based tree height in meters, β_0 and β_1 are fixed regression coefficients, a_i and b_i random plot effects and e_{ij} the residual error, based on the standard default assumptions of mixed-effects models (Mehtätalo and Lappi, 2020). Also crown radius was evaluated as a candidate predictor but it did not improve the fit. The parameter estimates of the model were $\widehat{\beta}_0 = -0.0614$, $\widehat{\beta}_1 = 1.039$, $\mathrm{var}(a_i) = 0.360^2$, $\mathrm{var}(b_i) = 0.0275^2$, $\mathrm{cor}(a_i, b_i) = -0.946$ and $\mathrm{var}(e_{ij}) = 0.953^2$. The plot-level predictions from this model were used as the heights of detected individual trees when computing the mean and dominant heights based on them.

The mean error (ME, empirical bias), RMSE, and standard deviation were used to measure the performance. Relative RMSE and ME were also used for the stand density, where RMSE and ME were divided by the mean of the field-measured value.

4. Results

4.1. Estimation and prediction of α and θ

Among plots where field-measured tree count was higher than the number of detected trees in the ALS data, the plot-specific estimates of θ based on N-matching varied from 0.018 to 0.627, with the mean of 0.281 and median 0.286 (Fig. 1). The estimates based on $SSPP$ varied from 0.039 to 0.550, with a mean of 0.301. The plot-specific estimates of α varied between −0.26 and 1.24, with a mean of 0.400. The estimates of $\widehat{\alpha}_{Nmatch}$ and $\widehat{\theta}_{Nmatch}$ had very strong nonlinear dependence, whereas the dependencies between $\widehat{\theta}_{SSPP}$ and the estimates based on N-matching were much weaker. Including also the nine plots where field-measured tree count was equal or smaller than the number of detected trees changed the results dramatically, especially in the case of α. Therefore, and based on some initial analysis, we defined the solution of N-matching as undefined for those plots. In practice it means that these plots were not included in the training data sets in the cross-validation, but each of them was used as an evaluation plot to avoid biased results.

Even the best performing predictors of α and θ explained only a small portion of the total variability (Table 3). In cross-validation, the full models reported in Table 3 were not used but the variable selection was done separately for each training set. The set of predictors varied a lot among the training sets, as illustrated by the summary of the number of parameters in them summarized in the last column of Table 3.

4.2. Performance of the adjusted HT like estimator

The Horvitz–Thompson-like estimator, where the estimates of α or θ are based on the training data, performed well compared to the naive method 1, where only detected trees are used. This happened regardless of the method used in estimating α and θ. There was a clear improvement in mean error, standard deviation and RMSE for all three applied stand characteristics (Tables 4 and 5) regardless of the method used in determining the parameters α and θ.

The best method to determine the combination of α and θ for estimation of stand density was the method 7, where $\theta = 0.5$ and α are determined as the mean of training plots (Table 4). It was clearly better than the second-best method 4, where $\alpha = 0$ and θ is the mean of $\widehat{\theta}_{Nmatch}$ of the training plots. However, method 7 shows signs of systematic underestimation of stand density for plots with the highest true density (Fig. 2, top; Table 4, right). Even though it is the best method in terms of RMSE also among those plots, the almost 9% underestimation can hardly be explained by sampling errors ($|ME|/se(ME) \approx 3.8$. In contrast, method 4 where θ is the mean of plot-specific estimates of θ based on N-matching has both low ME and RMSE among those plots. It was also the second-best method in the full data set.

In terms of dominant height and mean height, the method 2 where the values $\alpha = 0$ and $\theta = 0.5$ do not utilize the training data at all are the best (Table 5). However, that method is clearly the worst for estimating stand density and cannot be suggested for practical use. Among methods 3–10, where α and θ are estimated using the training data, the best method in terms of RMSE of dominant height is method 6 where $\alpha = 0$ and θ is based on a regression model on $\widehat{\theta}_{Nmatch}$. In terms of RMSE in mean height, methods 7 and 8 where $\theta = 0.5$ and α is estimated either as the mean of training plot or using a regression model are the best and have very similar performance. However, the differences between methods 3–8 in the accuracy in dominant and mean height are small and may be much affected by random variability. The value of $|ME/se|$ is clearly above 2 for all methods for dominant and mean height, which indicates that the underestimation of mean and dominant heights is significant; being about 0.20 m for dominant height for methods 3–10 and about 1.1 m for mean tree height.

Fig. 3 gives a closer look on the height estimation. It shows the distributions of tree height based on field-measurements and methods 1, 2, and 4. The HT-like estimator adjusts only the weights of detected trees. Therefore, the (discrete) support of the adjusted height distribution is the same as that of the detected trees. ALS-based ITD seldom

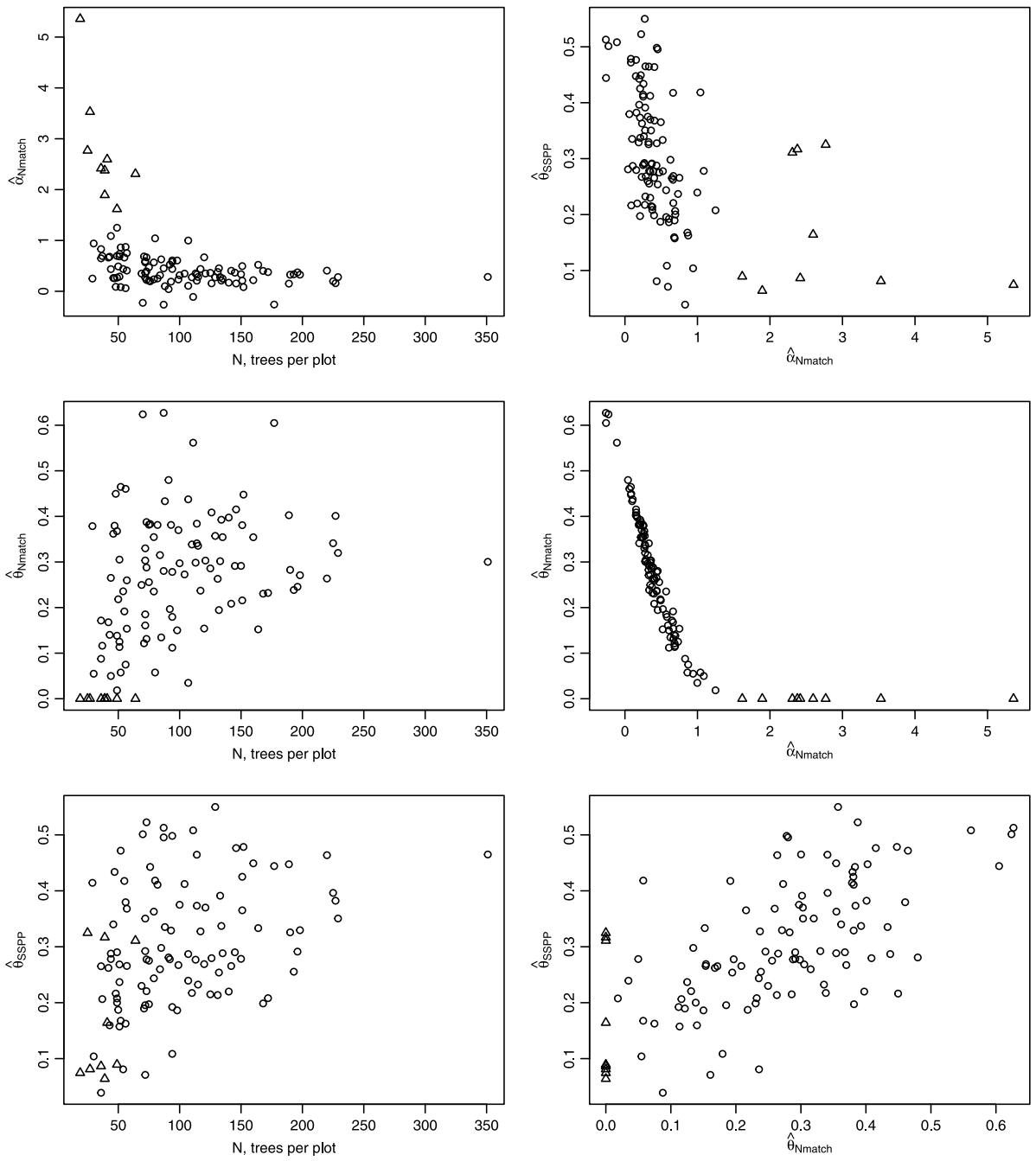


Fig. 1. The plot-specific estimates of α and θ -parameters based on N-matching and θ based on SSPP on field-measured tree count and on each other. Triangle is used as symbols for those plots where field-measured tree count was smaller or equal to the detected tree count.

Table 5
The cross-validation results about the accuracy of dominant height and mean tree height.

			Dominant height					Mean tree height				
			ME		RMSE		SD	ME		RMSE		SD
Method	α	θ	m (ME/se)	%	m	%	m	m (ME/se)	%	m	%	m
1	0	0	−0.31 (−6.54)	0.59	−1.42	2.68	0.50	1.84 (8.80)	11.63	2.86	18.04	2.20
2	0	0.5	**−0.10 (−2.32)**	**0.44**	**−0.43**	**2.00**	**0.43**	**0.66 (3.72)**	**4.16**	**1.97**	**12.40**	**1.86**
3	0	$\tilde{\theta}_{SSPP}$	−0.20 (−4.70)	0.50	−0.92	2.25	0.46	1.08 (5.82)	6.79	2.21	13.96	1.94
4	0	$\tilde{\theta}_{Nmatch}$	−0.21 (−4.84)	0.50	−0.96	2.27	0.46	1.11 (5.97)	7.01	2.24	14.11	1.95
5	0	$\hat{\theta}_{SSPP}$	−0.20 (−4.56)	0.49	**−0.89**	**2.22**	**0.45**	1.05 (5.67)	6.60	2.19	13.84	1.94
6	0	$\hat{\theta}_{Nmatch}$	−0.21 (−4.88)	0.49	**−0.93**	**2.20**	**0.44**	1.13 (6.07)	7.12	2.25	14.16	1.95
7	$\tilde{\alpha}_{Nmatch}$	0.5	−0.21 (−4.90)	0.51	−0.97	2.29	0.46	1.05 (5.82)	6.62	**2.15**	**13.58**	**1.89**
8	$\tilde{\alpha}_{Nmatch}$	0.5	−0.21 (−4.96)	0.49	−0.96	2.23	0.45	1.05 (5.85)	6.64	**2.15**	**13.58**	**1.89**
9	$\tilde{\alpha}_{Nmatch}$	$\tilde{\theta}_{Nmatch}$	−0.23 (−5.20)	0.52	−1.05	2.34	0.47	1.15 (6.27)	7.23	2.23	14.04	1.92
10	$\tilde{\alpha}_{Nmatch}$	$\hat{\theta}_{Nmatch}$	−0.23 (−5.10)	0.51	−1.02	2.33	0.46	1.13 (6.26)	7.13	2.20	13.86	1.89

finds trees that are shorter than the shortest trees in the field data, but can easily miss some of the smallest size classes, as illustrated by plots 2, 31, 36, and 94. This leads to a general overestimation of individual tree height. However, the adjusted height distribution based on method 7 ($HT(\bar{\alpha}, 0.5)$) looks otherwise very similar to the observed height distribution in many of these plots. Especially, the right tail of the estimated height distribution is nicely estimated, which explains the low RMSE of dominant height, and much smaller bias than for mean tree height. For method $HT(0, 0.5)$, the number of stems has a severe overestimation, especially for the classes of shortest detected trees.

5. Discussion

This study extended the Horvitz–Thompson-like estimator proposed by Mehtätalo (2006) and further developed by Kansanen et al. (2016, 2019, 2022) to take into account the departure from the assumption of complete spatial randomness in the tree population and evaluated the methods with empirical data. The new adjusted HT-like estimator (6) includes two parameters: α that models the detection condition and θ that models how large trees affect the occurrence of smaller trees within the area of their crown projection.

As already demonstrated in the previous papers, the adjustment of the set of trees found by ITD based on the detection condition is highly useful. If only the detected trees were used, the stand density would be underestimated by 35%. In addition, empirical estimation of the parameter of the estimator is suggested: assuming that a tree is detected if center point is visible ($\alpha = 0$) and using Poisson process for the hidden trees ($\theta = 0.5$), the stand density was overestimated by 57%. The best-performing methods were those where either α or θ was estimated as the mean of training plots and the other parameter was given the above-mentioned default value. Those methods led also to much better results in terms of estimation of mean height and dominant height than using ITD-trees only. Especially, the distribution of tree heights was much more realistic than one would get by using only the detected trees or using adjustment based on the default values. However, dominant height was underestimated by approximately 0.2 m and mean height overestimated by approximately one meter. These estimates are already corrected for the well-known underestimation based on penetration of laser pulse to tree crowns and missing treetops (Gaveau and Hill, 2003). The overestimation of tree height is caused by a higher detection probability for large trees. This bias was clearly decreased but not completely removed by the proposed estimator.

Including separate parameters for modeling the detectability condition and the spatial pattern of tree locations make conceptually much sense. However, our results showed that explicit modeling of the spatial point pattern through the sequential spatial model of Yazigi et al. (2021) provided only minor improvements compared to the method of Kansanen et al. (2016). A natural explanation is that, as discussed already in Kansanen et al. (2022), empirical estimation of parameter α implicitly takes into account also the departure from the complete spatial randomness. Our results showed that this implicit modeling of the spatial point pattern seems to be so efficient that a well-formulated spatial model did not provide clear improvement in overall accuracy, even though it led to better estimation of stand density for those sample plots where the true stand density was the largest. However, it is not sure whether this is a property of the methods that generalizes to other data sets.

The adjusted model was based on an assumption that the ratio of densities in the hidden and visible parts of the forest is constant for all trees of a sample plot, regardless of tree size. This is much more realistic assumption than that of the previous methods, which assumed that the densities in the hidden and visible parts are equal, but can still be criticized e.g. because the trees in the hidden parts may have larger probability to be removed in thinnings or due to mortality, compared to the trees in visible parts. The model might be adjusted for this problem by allowing parameter θ to depend on the tree size. However, the benefit of such extension might be minor, in a similar way as the effect of switching from a model that assumes similar density to hidden and visible parts was found to be minor in this study.

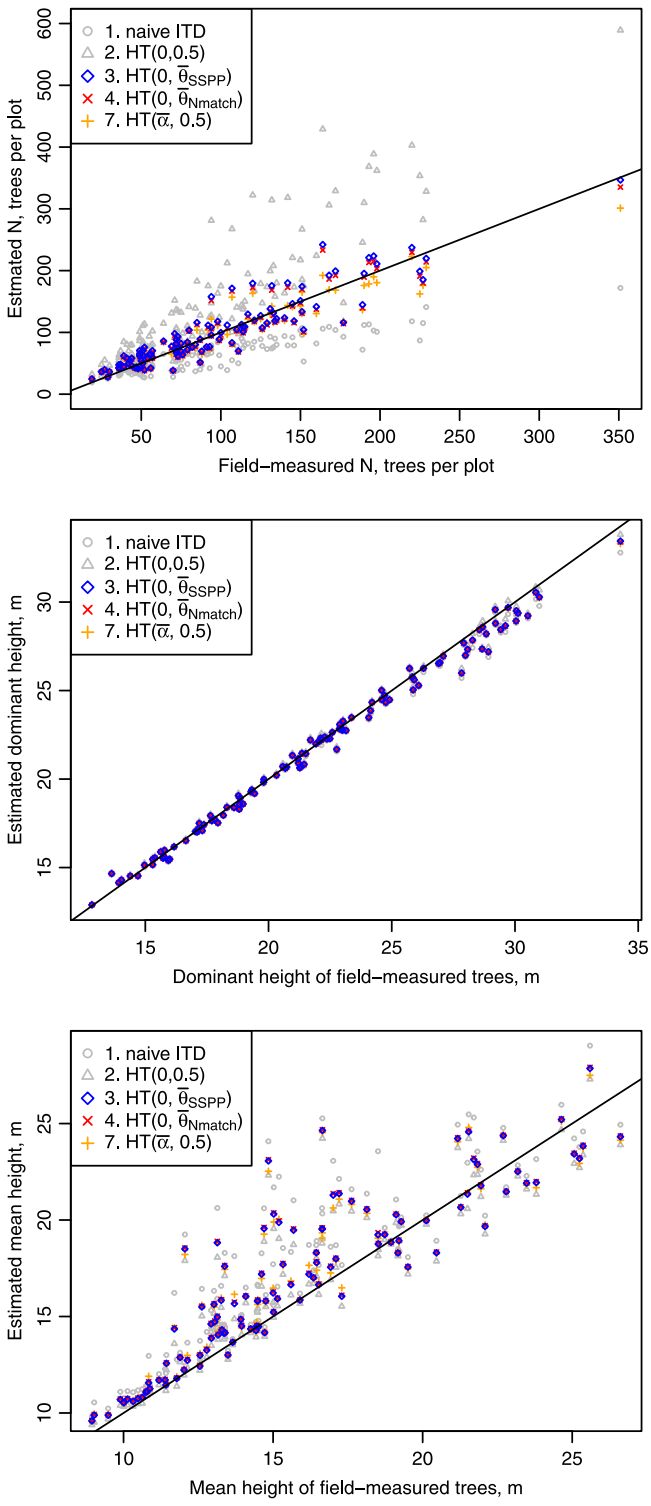


Fig. 2. The estimates of number of stems, dominant height and mean height on field-measured values using methods 1 (naive ITD), 2 (naive HT) and the three best-performing methods to determine α and θ for the Horvitz–Thompson like estimator.

Another important difference between methods where $\alpha = 0$ and θ is estimated vs. the methods where $\theta = 0.5$ and α is estimated is in computing time. When $\alpha \neq 0$, we need to dilate or erode a union of

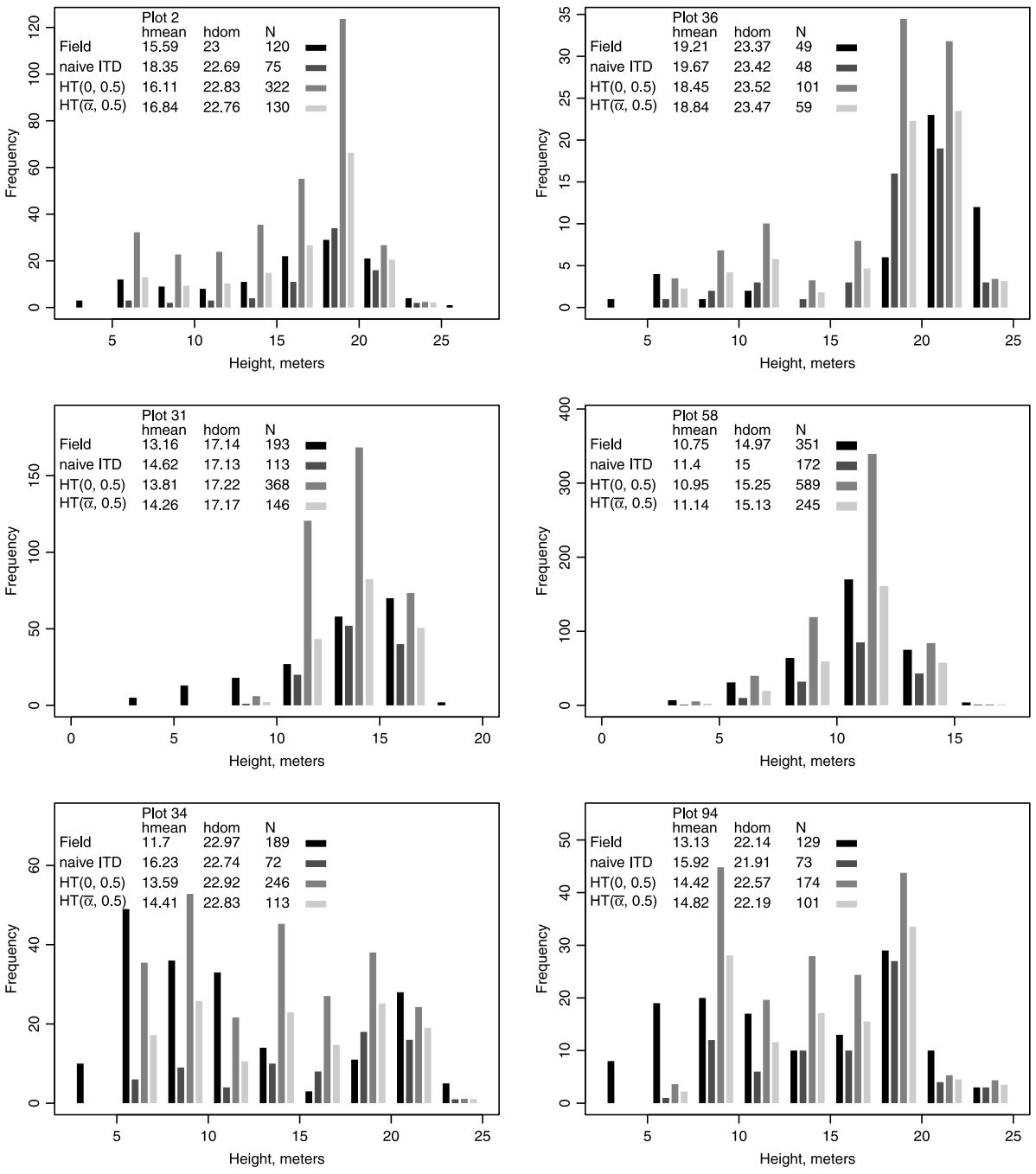


Fig. 3. The true height distributions of six selected sample plots and estimated distributions based on methods 1, 2 and 7, and the mean height (hmean), dominant height (hdom) and number of stems (N) in each of the distributions.

crown segments n times in a plot with n detected trees, which can take lot of time especially if the estimation is done on stands with thousands of trees. The computational burden for estimators where $\alpha = 0$ is much lower.

The approach where both parameters α and θ were estimated empirically did not lead to improvements compared to methods where only either of these parameters was fixed to a justified value and the other was estimated. A natural explanation to this behavior is the strong dependence of the estimates of α and θ when N-matching was used (Fig. 1). These parameters are so strongly dependent that using both of them leads to an overparameterized model. Or in other words, formulating a justified model for the process of canopy visibility and estimating the related parameter α using N-matching leads to very similar results than modeling spatial point pattern using a well-justified model and estimating the related parameter θ using N-matching.

Fig. 1 shows that the optimal values of α and θ vary much between sample plots. Therefore, we expected that a model that could predict these parameters even slightly better compared to the mean of the estimates of the training plots would perform better in prediction of stand characteristics. However, even though the spatially justified predictors of Häbel et al. (2021) were able to predict some variability in the estimates of α and θ, the improved accuracy did not pay back in improvements in the accuracy of stand density, mean and dominant height.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research was financially supported by the Academy of Finland through (1) the Finnish Centre of Excellence of Inverse Modeling and Imaging, (2) the flagship program "Forest-Human–Machine Interplay - Building Resilience, Redefining Value Networks and Enabling Meaningful Experiences (UNITE, decision number 337655)", and (3) research projects 295100, 310073, 321761, 327211 and 351525.

References

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

Gaveau, D.L., Hill, R.A., 2003. Quantifying canopy height underestimation by laser pulse penetration in small-footprint airborne laser scanning data. Can. J. Remote Sens. 29 (5), 650–657. http://dx.doi.org/10.5589/m03-023.

Häbel, H., Balazs, A., Myllymäki, M., 2021. Spatial analysis of airborne laser scanning point clouds for predicting forest structure. Math. Comput. Forestry Nat.-Resour. Sci. (MCFNS) 13 (1).

Harrell, Jr., F., 2001. Regression Modeling Strategies with Applications To Linear Models, Logistic Regression, and Survival Analysis. Springer, New York, USA, 562 p..

Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., 2007. Statistical Analysis and Modelling of Spatial Point Patterns. John Wiley & Sons, Ltd, ISBN: 9780470725160, http://dx.doi.org/10.1002/9780470725160.ch1.

Kansanen, K., Packalen, P., Lähivaara, T., Seppänen, A., Vauhkonen, J., Maltamo, M., Mehtätalo, L., 2022. Refining and evaluating a horvitz-thompson -like stand density estimator in individual tree detection based on airborne laser scanning. Can. J. Forest Res. 52 (4), 527–538. http://dx.doi.org/10.1139/cjfr-2021-0123.

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

Kansanen, K., Vauhkonen, J., Lähivaara, T., Seppänen, A., Maltamo, M., Mehtätalo, L., 2019. Estimating forest stand density and structure using Bayesian individual tree detection, stochastic geometry, and distribution matching. ISPRS J. Photogramm. Remote Sens. (ISSN: 0924-2716) 152, 66–78. http://dx.doi.org/10.1016/j.isprsjprs.2019.04.007.

Lähivaara, T., Seppänen, A., Kaipio, J., Vauhkonen, J., Korhonen, L., Tokola, T., Maltamo, M., 2014. Bayeslan approach to tree detection based on airborne laser scanning data. IEEE Trans. Geosci. Remote Sens. 52 (5), 2690–2699.

Mehtätalo, L., 2006. Eliminating the effect of overlapping crowns from aerial inventory estimates. Can. J. Forest Res. 36 (7), 1649–1660.

Mehtätalo, L., Kansanen, K., 2022. Lmfor: Functions for forest biometrics. R package version 1.6.

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

Mehtätalo, L., Yazigi, A., Kansanen, K., Packalen, P., Lähivaara, T., Maltamo, M., Myllymäki, M., Penttinen, A., 2021. Using stochastic geometry and sequential spatial point process model for estimation of stand density based on ALS-itd. In: Hollaus, M., Pfeifer, N. (Eds.), Proceedings of the SilviLaser Conference 2021. pp. 114–116. http://dx.doi.org/10.34726/wim.1934.

R. Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/.

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

Vauhkonen, J., Korpela, I., Maltamo, M., Tokola, T., 2010. Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. Remote Sens. Environ. (ISSN: 0034-4257) 114 (6), 1263–1276. http://dx.doi.org/10.1016/j.rse.2010.01.016.

Yazigi, A., Penttinen, A., Ylitalo, A.-K., Maltamo, M., Packalen, P., Mehtätalo, L., 2021. Modeling forest tree data using sequential spatial point processes. J. Agric. Biol. Environ. Stat. 27, 88 – 108. http://dx.doi.org/10.1007/s13253-021-00470-2.