Science of Remote Sensing 12 (2025) 100310

Contents lists available at ScienceDirect

Science of Remote Sensing

journal homepage: www.sciencedirect.com/journal/science-of-remote-sensing

Full length article

Transferability of country-wide airborne laser scanning-based models for individual-tree attributes

Valtteri Soininen *, Xiaowei Yu, Matti Hyyppä, Juha Hyyppä

Department of Remote Sensing and Photogrammetry, Finnish Geospatial Research Institute FGI, The National Land Survey of Finland, Vuorimiehentie 5, Espoo, 02150, Finland

ARTICLE INFO

Keywords:
Individual trees
Model transferability
Nationwide laser scanning
Airborne laser scanning
Precision forestry

ABSTRACT

Optimising bioeconomy-related ecosystem services requires more detailed forest information. One option is to go towards individual-tree-based precision forestry. Although many national airborne laser scanning (ALS) programmes can detect individual trees and predict their attributes, training the necessary models remains a challenge. Ideally, each site covered with ALS would have its own reference data, but this requires measuring millions of trees nationwide per every country-level ALS scan. Instead of always collecting site-specific training data, an alternative is to transfer individual tree models from other sites. This approach relies on good model transferability that ensures accurate and realistic estimates. This study tested the transferability of individual tree diameter at breast height (DBH) and stem volume models combining national laser scanning data and the random forest method in Finland. The model that was trained with coordinate information benefitted from training data that were collected within the range of 500 km. The root mean squared error (RMSE) and bias magnitude of the model that was trained without the coordinate information started to increase after 300 km, but the increase could be cancelled by using coordinates as predictor features. Furthermore, when the models were evaluated outside the area for which they were trained, the errors increased at a rate between 0.27–0.28 cm/100 km in RMSE in DBH prediction and 8.08–13.18 dm^3/100 km in stem volume prediction. The same values for bias magnitude were 0.39–0.42 cm/100 km in DBH prediction and 8.32–12.01 dm^3/100 km in stem volume prediction. The increase in training set size slightly slowed the rate. Quick convergence of RMSE was observed in a test in which small amounts of target site data were included in the training data. The same was also observed for bias magnitude, although the results were not as good as with RMSE.

1. Introduction

Forests are central to modern sustainable bioeconomy, offering a wide range of economic, ecological, and social ecosystem services. These include providing timber and biofuel, serving as crucial habitats for biodiversity, offering recreational opportunities, producing food such as berries and mushrooms, and playing a vital role in climate regulation (Kettunen et al., 2012). However, inaccurate data on forest resources and biodiversity can lead to poor management decisions, resulting in significant economic and social repercussions (Haara et al., 2019). Key tree attributes measured in forest inventories typically include stem volume, biomass, height, and species, but also diameter at breast height (DBH) and stem curve. The inventories are based on remote sensing data covering large areas, field plots used for training and testing the models, and predictive models used for deriving the association between the remote sensing data, mainly airborne laser scanning (ALS) data and field plots (Hyyppä et al., 2020a). In Finland, a typical area-based inventory requires about 500 field reference plots for model training (Niemi et al., 2015).

Although area-based forest inventories based on aggregated tree characteristics (Næsset, 2002) are standard in Scandinavia, they do not provide crucial information on the tree size distribution. Historically, adopting an individual tree approach (Hyyppä and Inkinen, 1999) has been deemed too demanding mainly for computational reasons. However, a significant shift is underway. Hyyppä et al. (2024) demonstrated the feasibility of a nationwide individual tree forest inventory, a system now extended to encompass six billion trees in Finland (metsakanta.fi) and most probably one of the world's largest forest inventory databases. This system is built upon Finland's second national laser scanning programme (2020–2025) which provides comprehensive coverage with a density of five points per square metre. The third national programme (starting in 2026) promises an even higher resolution, with around 20 points per square metre. This coincides with the Finnish Forest Centre's

* Corresponding author.
E-mail address: valtteri.soininen@maanmittauslaitos.fi (V. Soininen).

https://doi.org/10.1016/j.srs.2025.100310
Received 8 August 2025; Received in revised form 8 October 2025; Accepted 9 October 2025
Available online 11 October 2025
2666-0172/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

commitment to making both individual tree and area-based forest data accessible to all forest owners. In relation to ALS acquisition, Finland is divided into 137 production areas with the average size of 2550 km^2. Collecting the manual reference data is, however, slow and costly. Therefore, the amount of manual labour should be minimised either by developing fully automatic field inventory techniques or by developing models that are applicable to as large areas as possible. The latter promotes studying the model transferability on a national scale.

While much of the research effort has been focused on the transferability of area-based models and allometric models, limited effort has been put into studying the transferability of individual tree models based on ALS and individual tree approach. Karjalainen et al. (2019) and Korhonen et al. (2019) studied this using the same data between their studies but different methods containing three test sites all located in Eastern Finland. The linear mixed-effect model used by Korhonen et al. (2019) was found to be more robust against the increase in relative root mean squared error (RMSE%) than the *k* nearest neighbour method used by Karjalainen et al. (2019), but less robust against changes in bias. Korhonen et al. (2019) observed that transferring the linear mixed-effect model for Scots pine attributes over a distance of less than 100 km increased the RMSE% by 1.3–6.5 %-points, depending on the attribute and the target site. They also tested how collecting a minimal sample of data from the target plots improved the prediction accuracy, and observed an average decrease of around one %-point in RMSE%.

Past research on individual tree model transferability has reported the growth in error when models are transferred over spatial range in a discrete manner in a limited area with limited data. In this study, the error change rate is studied on a national scale, which greatly expands the study area compared to previous studies. The aim is to support future individual tree data collection campaigns by optimising the number of reference sites needed in large-area inventories with minimum loss of accuracy and by determining the area inside of which the model accuracy does not suffer from transferring. The research questions are as follows. (1) What is the maximum radius for training data beyond which the model prediction capability declines? (2) How does spatial separation between training and prediction areas influence model accuracy? (3) What is the quantitative impact of local calibration data on prediction accuracy within the study area? The random forest method (Breiman, 2001) is used as the regression method and any algorithmic development is excluded from the study.

2. Related studies on prediction and transferability

In the past ten years, model transferability in ecological context has been studied extensively. Model transferability refers to the ability of a predictive model to predict accurately using input data collected from different premises, such as point-in-time, location, or device. Such models are desirable, as they save time, labour, and money (Yates et al., 2018), but often, the prediction ability is impaired when the models are transferred, even though the input data are seemingly similar to the data to which the model was fitted. Drivers of impaired prediction accuracy of transferred models have been recognised, and they can be categorised as originating from modelling choices (van Ewijk et al., 2020; Tompalski et al., 2019; Latifi and Koch, 2012; Sumnall et al., 2024), different sensors (Tompalski et al., 2019; Kotivuori et al., 2016), different ecological conditions (Kangas et al., 2023; Cysneiros et al., 2021; Lines et al., 2012; Schneider et al., 2018), and differences in point-in-time (Domingo et al., 2019; Kangas et al., 2020).

Some models are more resistant to changes in input data than others. Yates et al. (2018) discussed upon several issues related to model transferability. One of the issues is the risk of overfitting to local conditions when using complex models instead of simpler ones. Yet, they mentioned that complex models are more suited for modelling complex phenomena, and the better transferability of simple models does not guarantee correct predictions. Van Ewijk et al. (2020) compared a simple ordinary least squares model with a more complex random forest algorithm when predicting forest attributes over a long distance in Canada. They observed mixed results between the two modelling approaches in their test scenarios, but they mentioned that the random forest algorithm is less robust to site-specific differences in the modelled forest attribute. Tompalski et al. (2019) compared how least squares regression, random forest, and *k* nearest neighbours models transfer over a distance of approximately 350 km in South-West Canada. They discovered that the least squares regression and random forest models transfer better than the *k* nearest neighbours method, but the results depended on the modelled attribute.

Sensors and data collection techniques improve over time. Constant developments in ALS devices include the increase in pulse repetition rate (Yu et al., 2006) and the inclusion of multiple wavelengths and multiple returns (Pfeifer and Briese, 2007; Takhtkeshha et al., 2024). The use of uncrewed aerial vehicle-based sensors has increased the density of point clouds and the accuracy of forest inventories (Hyyppä et al., 2020b). Tompalski et al. (2019) studied how dissimilarities in ALS point clouds affect the model transferability. They found that simpler models are more robust against changes in point cloud density, but more complex models produce more accurate predictions overall. Kotivuori et al. (2016) observed that the different ALS devices have some effect on prediction accuracy.

Different ecological conditions affect the way forests grow (Bonan and Shugart, 1989). Thus, a model developed for a large area, such as a country, may not work optimally if different ecological conditions are not included in the model. Kangas et al. (2023) studied how tree allometry changes as a function of different ecological drivers. They observed that temperature sum is a significant variable in tree volume allometric equations, but soil type (mineral or peatland) is usually not. Lines et al. (2012) examined the effect of various climatic and environmental variables on tree height-DBH and crown diameter-DBH ratios of 26 tree species in Spain. One of their observations was that trees become slenderer in warmer temperatures. This would have immediate consequences on transferability, as tree height is known to be a strong predictor of DBH and stem volume (Yu et al., 2011). Schneider et al. (2018) tested how climate change affects tree stem taper and volume of five tree species in Canada. The model equation for stem taper benefitted from the inclusion of climatic variables, such as 30-year mean annual and seasonal values of temperature, rain, and snowfall.

3. Data and methods

3.1. ALS data and individual tree delineation

The ALS data used in this study originate from the Finland's nationwide laser scanning campaign (National Land Survey of Finland, 2020) which commenced in 2020 with the goal of covering the entire country over a six-year period. For operational purposes, the country was systematically divided into approximately 50 km × 50 km production areas, each scanned independently. 16 production areas were selected from different parts of the country for this study, and they are referred as sites. The sites are plotted in Fig. 1(a). They were laser scanned in the summer of 2022 and 2023 at an altitude ranging from 1265 m to 2100 m above ground level to ensure a point density of five points per square metre on average. That is, the distance between laser points on the ground was on average no more than approximately 0.4 m for all production areas. Different scanners from Riegl (Riegl GmbH, Horn, Austria) and Leica (Leica Geosystems AG, Heerbrugg, Switzerland) were used for the acquisitions. Table 1 presents their characteristics. The resulting point clouds were quality checked and classified into classes representing, for example, ground, vegetation, buildings, and noise before delivery. In this study, only ground and vegetation points

Table 1
Sensor characteristics for the sensors used in ALS surveys. The divergence is at $1/e^2$. NIR = near infrared.

	Riegl VQ-780II	Riegl VQ-780II-S	Riegl VQ-1560II-S	Leica Hyperion2+
Point accuracy /precision	20/20 mm[a]	20/20 mm[a]	20/20 mm[a]	50 × 130 mm/NA[b]
Beam divergence	0.25 mrad	0.23 mrad	0.23 mrad	0.23 mrad
Wavelength	NIR	NIR	NIR	1064 nm
Max. field-of-view/used	60°/40°	60°/40–48°	58°/40°	20–40°/40°
Multipulse	Yes	Yes	Yes	Yes

[a] 1 sd. dev. @ 250 m altitude.
[b] Vertical × horizontal accuracy, 1 sd. dev. @ 1000 m altitude.

Table 2
Treewise features derived from the point clouds. Two models were tested: one with coordinates (*xl* and *yl*) and one without. The rest of the features were shared by both models.

Feature	Description
xl, yl	Tree coordinates in the ETRS-TM35FIN projection
Hmax	Tree height
Hmean, Hstd, Hskew, Hmode	Mean/st. dev./skewness/mode of vertical treewise point distribution
Gele	Ground elevation
cSlope	Mean value of third component of the normal vectors to crown surface triangulations
lowBranch	Height of the lowest branch
crownH, crownA, crownV	Crown height, area and volume
Hp10–95	Six percentiles of vertical treewise point distribution between 10%–95%
D5–9	Five cumulative distribution function values of vertical point distribution between 50%–90% of the height between 2 m above ground and treetop
O, F, L, M	Ratio of only/first/last/intermediate returns to all vegetation returns
min_int, mean_int, std_int	Minimum/mean/st. dev. of point intensity of all points
min25_int, mean25_int, std25_int	Same but below the 25th height percentile
min80_int, mean80_int, std80_int	Same but above the 80th height percentile

were used. These points were then normalised in height by removing the ground elevation from the laser height for further analyses.

Considering the fact that point density of ALS data was relatively low, individual tree detection was carried out with a simple algorithm based on a canopy height model. The algorithm started with the creation of a canopy height model by finding the maximum values of normalised points within a raster image with a resolution of 0.5 m × 0.5 m. The canopy height model was further smoothed by filtering to eliminate the variations in the canopy surface. Then, the tree tops were searched with local maximum filtering followed by a watershed transformation for crown delineation using detected tree tops as control markers. The tree segmentation process is discussed in more detail in Hyyppä et al. (2024). In a previous study by Kaartinen et al. (2012), the method was found to be superior to manual segmentation and many other segmentation methods. Moreover, current deep learning-based segmentation methods require denser point clouds with more than 50–100 points per square metre (Wielgosz et al., 2024; Xiang et al., 2024), which justifies the use of raster-based tree segmentation method developed for sparse point clouds.

Various features were derived from the point clouds bounded by the delineated crown polygons. Table 2 tabulates the features used in model training. Importantly, two types of models were tested: one without coordinate features and one using the coordinates as features. The features involving intensity were scaled in each site by subtracting the mean value and dividing by the standard deviation (z-score normalisation) in order to mitigate dissimilarities between the sensors.

3.2. Reference data

Reference data collection and measurements were carried out in the summer of 2022 and 2023 in the same sites as ALS acquisitions. A small portion of 19 plots were already manually surveyed in 2021, which causes one leap year between the manual survey and the ALS survey in some plots in Alavus and Tervo. The error caused by this is assumed to be negligible. Sample plots were selected for each site representing the range of forest conditions of the site, including mature and young stands, as well as pine-, spruce-, and birch-dominated and mixed stands. Scots pine (*Pinus sylvestris*), Norway spruce (*Picea abies*), downy and silver birch (*Betula pubescens*, *Betula pendula*) are the main tree species in Finland. Field measurements were conducted in all sample plots. For each sample plot, diameter at breast height (DBH) was manually measured for all trees with a DBH greater than 2.5 cm (in some plots, the limit was 3 cm) using callipers.

Tree height for selected trees was measured manually. The height of the rest of the trees was obtained by modelling using different methods and models. In the first method, the height was predicted from the laser scanning data and calibrated by the following model

$$H_{cal} = aH_{laser} + b, \quad (1)$$

where H_{cal} is the calibrated tree height, H_{laser} is the height by laser scanning, and a and b were determined using trees whose height had been measured manually. The model was developed separately for different production areas and tree species. The second method consisted of the Näslund model (see e.g. Mehtätalo et al. (2015)) or the models by Eerikäinen (2009). The species was recorded by visual identification.

The stem volume was calculated using the measured DBH and height values as input to allometric equations that closely resemble the standard Finnish allometric model by Laasasenaho (1982). Lately, Kangas et al. (2023) showed that the parameters of the equations vary throughout Finland. Thus, their two-variable model was evaluated for

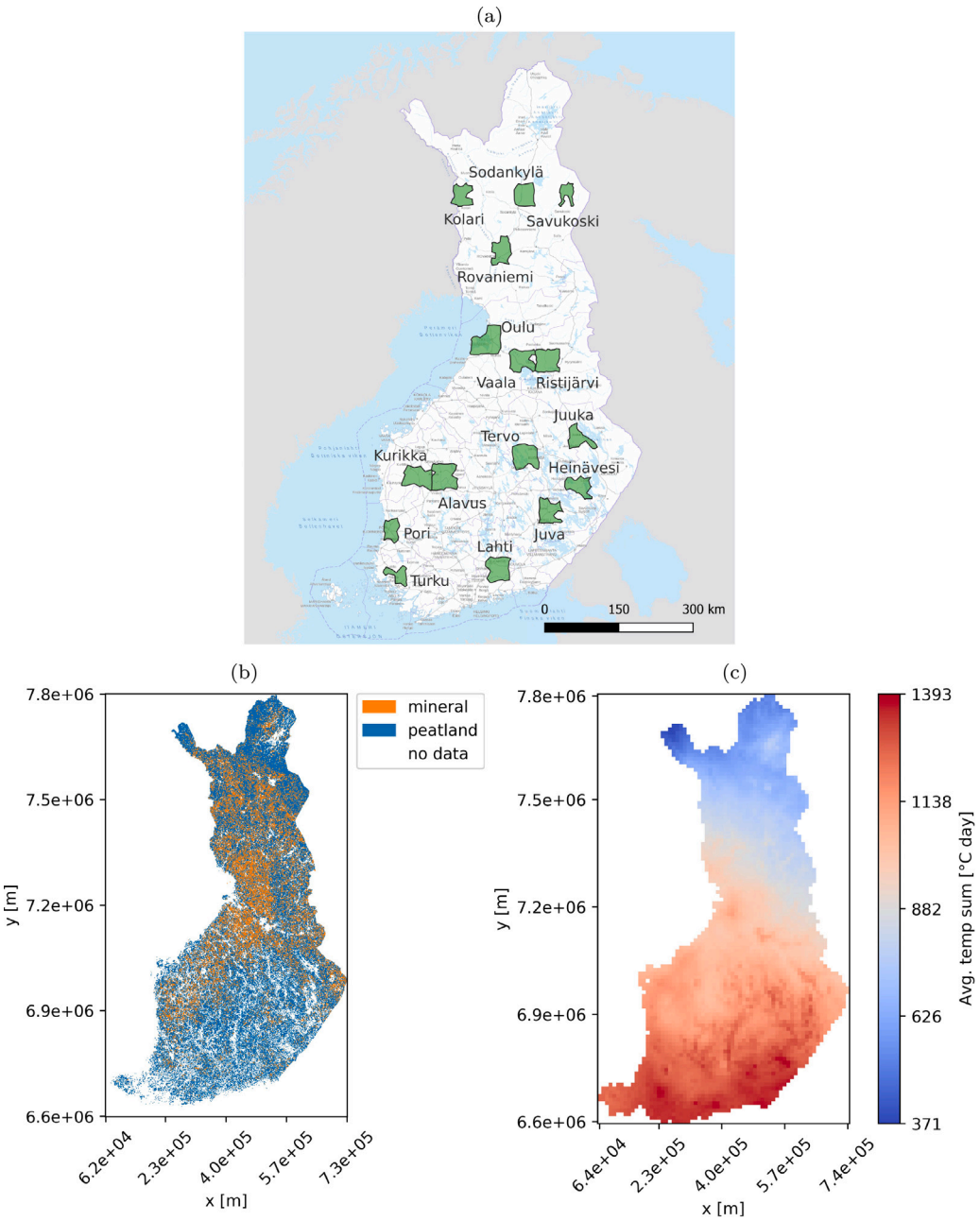

Fig. 1. 1(a): Location of the test sites. The shaded areas mark the distribution of the test plots within the sites. 1(b) and 1(c): Soil types and 30-year average temperature sum (1961–1990). The coordinates are in the ETRS-TM35FIN projection.
Source: The data are obtained from Natural Resources Institute Finland (2024) and Finnish Meteorological Institute (2016).

the reference stem volume to account for the spatial variability of the reference volume values. The model relates the DBH and height measurements to stem volume, but allows spatial dependency through the input variables soil type and 30-year average temperature sum. Only the fixed-effect part of the model was used, and the dataset option was set to *scanned* because that model matched the closest to the collection time of our reference data. Data for the soil type and temperature sum were obtained from Natural Resources Institute Finland (2024) and Finnish Meteorological Institute (2016), respectively, and the models are reported in the study by Kangas et al. (2023).

The improved equations were evaluated for 99.6% of the trees in the dataset, and they resulted in a change with an RMSE of 0.2, 12.4, and 1.3 dm^3 for pine, spruce, and birch, respectively, when compared to the original Laasasenaho equations. The remaining 0.4% of the trees had no soil type data or were of tree species other than pine, spruce, or birch. The stem volume reference for those trees was modelled with the models published in Tomppo et al. (2011), and the volume of trees that could not be associated with the soil type data was predicted with the original equations of Laasasenaho (1982). Tables 3 and 4 present descriptive statistics for each site, and Fig. 1 illustrates the study site locations, temperature sums, and soil types.

3.3. Linking ALS data to reference data

The linking between the detected trees and the field-measured trees was carried out based on the method presented in Yu et al. (2006). The idea is to find the nearest pair of trees between the detected trees and field-measured trees. A pair is determined based on the distance in an n-dimensional feature space within a certain threshold. In this study, the method was applied in a three-dimensional space based on x and y

Table 3
Descriptive statistics for the study sites. Average and standard deviation (in parentheses) reported.

Site	No. plots	No. trees	Height [m]	DBH [cm]	Volume [dm^3]
Sodankylä	192	11 774	12.3 (3.7)	16.2 (6.5)	168.8 (169.1)
Kolari	198	12 251	13.4 (3.8)	18.9 (7.1)	236.6 (228.3)
Savukoski	183	9 612	12.3 (3.6)	17.5 (6.7)	192.8 (204.6)
Rovaniemi	204	11 831	13.8 (3.4)	17.4 (6.0)	197.6 (171.8)
Oulu	204	11 828	16.3 (4.1)	19.9 (7.0)	304.6 (256.5)
Vaala	163	9 667	14.3 (4.2)	16.8 (6.3)	205.5 (210.0)
Ristijärvi	199	11 696	15.0 (3.6)	18.5 (7.1)	258.3 (255.2)
Juuka	168	10 918	19.6 (4.6)	22.5 (7.1)	441.2 (349.5)
Tervo	181	9 942	21.2 (5.4)	22.7 (7.8)	507.3 (432.1)
Alavus	203	12 075	18.9 (5.1)	21.0 (7.2)	393.1 (347.3)
Kurikka	154	10 055	18.9 (4.8)	21.7 (7.3)	409.8 (339.2)
Heinävesi	169	11 203	20.8 (5.1)	22.6 (7.5)	483.8 (387.1)
Juva	173	10 856	20.6 (5.1)	23.1 (7.7)	500.8 (416.0)
Pori	189	14 132	19.4 (4.7)	22.2 (7.4)	434.3 (355.9)
Lahti	148	8 355	21.2 (5.3)	22.8 (7.8)	509.2 (426.8)
Turku	150	9 069	20.2 (4.7)	23.7 (7.5)	503.2 (370.1)
Total	2878	175 264	NA	NA	NA

Table 4
Survey times for each site.

Site	Manual survey time	ALS survey time
Sodankylä	June–October 2023	July 2023
Kolari	June–August 2023	June 2023
Savukoski	July–August 2023	July 2023
Rovaniemi	May–October 2022	June–July 2022
Oulu	May–September 2023	June 2023
Vaala	May–September 2022	June 2022
Ristijärvi	May–September 2023	June 2023
Juuka	May–September 2022	August 2022
Tervo[a]	May and August 2021, May–September 2022	July 2022
Alavus[a]	June–July 2021, May–August 2022	June 2022
Kurikka	May–September 2023	June 2023
Heinävesi	May–August 2023	June 2023
Juva	May–August 2022	June–July 2022
Pori	May–September 2022	June 2022
Lahti	May–August 2022	June–July 2022
Turku	May–August 2023	May–June 2023

[a] 15 plots (847 trees) in Alavus and four plots (265 trees) in Tervo had a leap year between the ALS and manual survey.

coordinates and tree height. If a detected tree was the closest tree to a field-measured tree and vice versa within a distance threshold of 3 m, a successful match was found.

3.4. Random forest regression

The random forest algorithm (Breiman, 2001) trains a forest of predictor trees and outputs the average of the prediction of each tree as the final predicted value. The forest benefits from trees whose prediction errors do not correlate with each other (Ghojogh and Crowley, 2023). Independence is achieved by inputting a bootstrap sample of the training data for each predictor tree and by randomly selecting a number of features at each splitting node. There were three hyperparameters that were tuned in this study. They control the number of features randomly chosen at each node (max_features), the number of predictor trees (n_estimators), and the minimum number of samples in a leaf node (min_samples_leaf). The random forest algorithm was implemented using Python's scikit-learn v. 1.2.2 library.

In order to avoid overfitting to local forest features, the three hyperparameters were optimised with a five-fold cross-validation using 300 training samples per site, which resulted in separate data set with a total of 4800 training samples. The set was discarded after the hyperparameter optimisation. The following hyperparameter values were considered for each response variable and data set with and without coordinates:

- n_estimators: 100, 200, 300, 400
- max_features: *log2* (=5), *sqrt* (=6), 10
- min_samples_leaf: 3, 5, 10.

The cross-validation results showed that the number of predictor trees had very little effect on the cross-validation metric (RMSE). The best hyperparameters were a combination of the same values for max_features and min_samples_leaf and different values for n_estimators. Thus, a set of hyperparameters with the highest-ranking combination of max_features and min_samples_leaf with minimum n_estimators was chosen. This reduced the training time of the model. Table 5 shows the hyperparameter combinations.

Table 5
Cross-validation results for hyperparameter choice.

Resp. variable	n_estimators	max_features	min_samples_leaf	Test RMSE	Train RMSE
DBH[a]	200	10	3	3.54 cm	1.83 cm
DBH[b]	200	10	3	3.49 cm	1.79 cm
Stem volume[a]	100	10	3	127.2 dm^3	66.8 dm^3
Stem volume[b]	100	10	3	124.7 dm^3	65.6 dm^3

[a] Without coordinates.
[b] With coordinates.

3.5. Error metrics

Two error metrics and their relative versions are used to compare the results. RMSE is defined as

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

$$\mathrm{RMSE\%} = \frac{\mathrm{RMSE}}{\frac{1}{n}\sum_{i=1}^{n} y_i} \times 100\%, \tag{3}$$

where $\hat{y}_i$ is the predicted value for DBH or stem volume, and y_i is the corresponding reference value. Bias is defined as the difference between the mean reference value and the mean predicted value

$$\mathrm{bias} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \tag{4}$$

$$\mathrm{relative\ bias} = \frac{\mathrm{bias}}{\frac{1}{n}\sum_{i=1}^{n} y_i} \times 100\%. \tag{5}$$

The magnitude of bias refers to its absolute value.

4. Results

This section presents results for three test scenarios. In the first test, a model for a target site was trained by incrementally adding training sites to the training data set, always choosing the closest first, then the second closest, and so on. This procedure was repeated for each site. The second test was performed at the same time. In it, each trained model was evaluated at the remaining auxiliary target sites that were not in the training set or the original target site of the first test. The procedure for choosing the training sites is illustrated in Fig. 2 and in the video in Appendix B. The third test was run separately. It demonstrated a minimal working setup, where a model was trained with one training site at three different distances, and then some on-site data from the target site were mixed with the training data. Some results are compared with a model that is trained using a certain percentage of on-site data only and referred as the baseline model. In particular, the two-fold cross-validated results using 50% of on-site data for training and 50% for testing are referred as the baseline values.

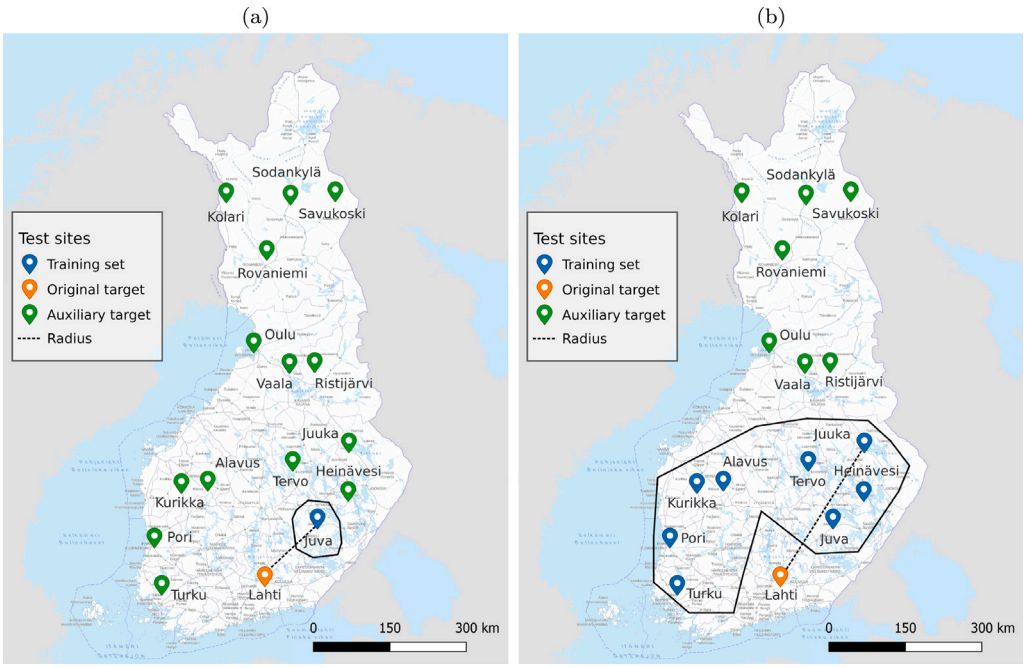

Fig. 2. Illustration how the training sites were chosen in the first and second test. 2(a): First iteration when Lahti is the original target site. 2(b): Eighth iteration. The radius measures the maximum distance between the original target site and the furthermost training site. The distance to auxiliary target sites was calculated as the average distance between each training site and the auxiliary site.

4.1. Target site data not included in the training data

The results for the first test show how the test error changes when no data from the target site were included in the training data, but the closest training sites were incrementally added to the training set. At each increment, a new training site was added to the training set, its distance was recorded, the model was retrained, and then evaluated with the target site. When all 15 training sites were added, the observed minimum error for the current target site was subtracted from each error value in order to standardise the error values between each 16 target sites. Thus, the values in Fig. 3 show the error increase above the minimum error per site in 100 km intervals for both the model that uses coordinates as features and for the model that does not.

The RMSE values of DBH prediction in Fig. 3(a) show that both models improve up to radii between 200–300 km. After that, the model not using coordinates starts to degrade, while the model that uses coordinates stays close to zero increase in RMSE. The minimum is observed between 400–500 km, although the increase in error stays almost constant, and near zero, after 200 km for the model that uses coordinates. In stem volume RMSE in Fig. 3(b), the increase in RMSE behaves similarly to the DBH error values. The two models begin to deviate after 300 km, and the model that uses coordinates obtains its minimum between 400–500 km, although almost as good RMSE is obtained already with radii between 200–300 km.

The decrease in bias magnitude is more pronounced for the model that uses coordinates both in DBH prediction in Fig. 3(c) and in stem volume prediction in Fig. 3(d). However, the decrease is less pronounced compared to the RMSE values. The bias magnitude of the model that does not use coordinates decreases until 200 km in DBH prediction and until 300 km in stem volume prediction. As with the RMSE values, the bias magnitude of the model that uses coordinates stays near the minimum after reaching it around 200–300 km, but the bias magnitude of the model that does not use coordinates increases after reaching its minimum.

As the panels in Fig. 3 show, the interval between 300–400 km is the radius where the RMSE and bias magnitude stop decreasing in practise. Therefore, 400 km was chosen as the optimal distance between the prediction accuracy and the amount of data used for training. Fig. 4 draws the baseline errors and the increase in RMSE% and relative bias magnitude on a map when the models were trained with the 400 km training set radius. The results are from the model that uses coordinates, since at the radius of 400 km, it always produced smaller bias and RMSE values than the model that was trained without coordinates, thus being the better of the two models. On average, the radius was 349 km (min. 240 km, max. 391 km), and it included 9 training sites. Fig. 4(e) illustrates the number of training sites for each original target site.

As the panels show, the increase in RMSE% when not including on-site data is larger with stem volume than with DBH (on average 1.3 %-points versus 0.8 %-points). The horizontal coordinate line 7.1×10^6 divides the RMSE% values into two groups: north to the line, all DBH (stem volume) RMSE% values are $\geq$ 17.2% (36.5%), south to the line the values are $\leq$ 16.2% (31.1%) when the baseline error and the increase in error due to distant training data are summed. The number of training sites correlates negatively with the increase in error. The Spearman correlation between the number of training sites and RMSE% increase is −0.25 with DBH and −0.10 with stem volume.

Figs. 4(c) and 4(d) show that the relative bias magnitude increases significantly when data from the target site are not included in the training data (2.2 %-points in DBH and 3.5 %-points in stem volume prediction). Unlike RMSE%, the increase in the relative bias magnitude is more uniformly distributed. The correlation between the training set size and bias magnitude increase is −0.55 with DBH and −0.31 with stem volume.

4.2. Species and incremental convergence

The error convergence using 10% training data increments was tested using the same radius and training sites as in Fig. 4. In each training site, the data were sampled with 10% increments ranging between 5%–95%. Each increment was randomly sampled three times always retraining the model. The model was trained with coordinates, and its

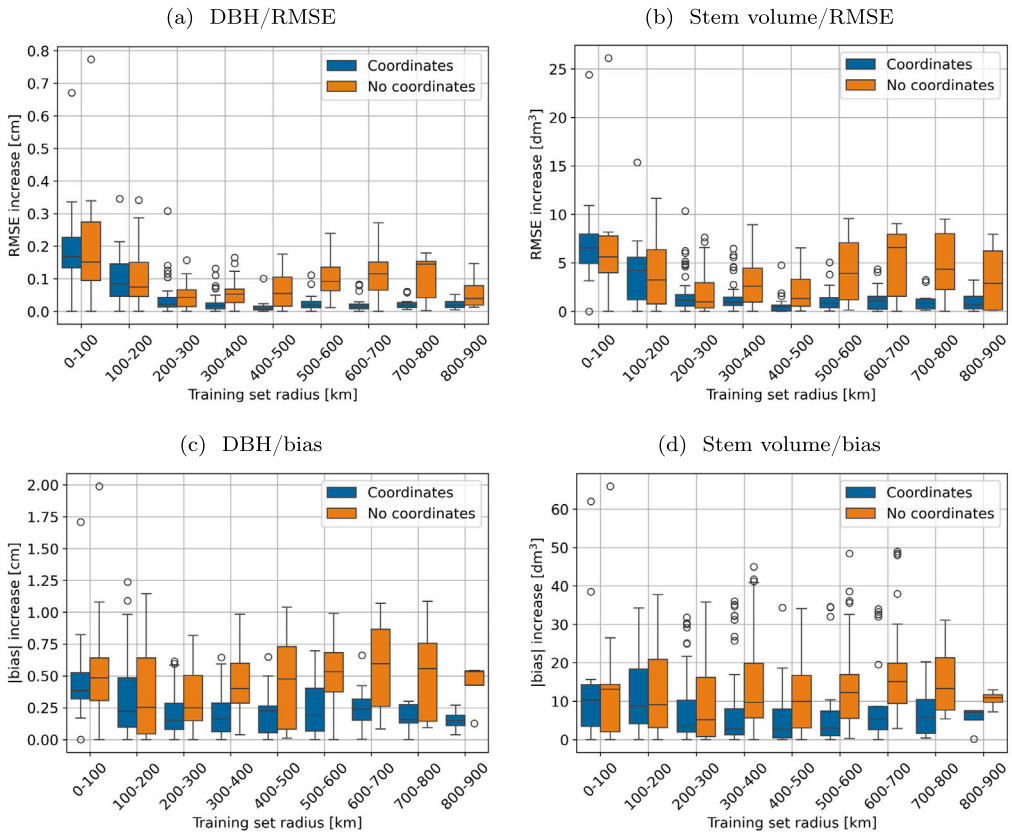

Fig. 3. 3(a) and 3(b): The distribution of RMSE values above the site-wise minimum RMSE as a function of training set radius. 3(c) and 3(d): The same results for bias magnitude.

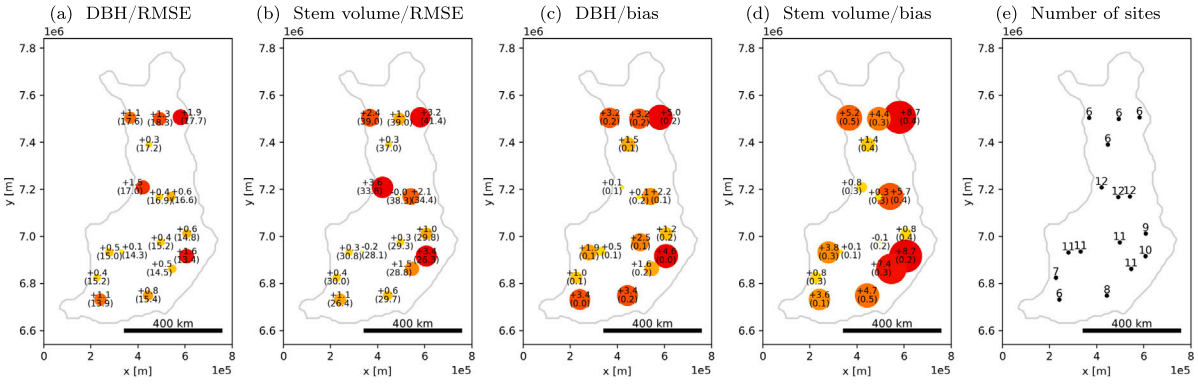

Fig. 4. 4(a) and 4(b): The increase in RMSE% when using data collected outside the target site with the target radius of 400 km. The number in parentheses is the corresponding two-fold cross-validated baseline value (unit in %). The marker is scaled and coloured according to the increase marked above the baseline value (unit in %-points). The coordinates are in the ETRS-TM35FIN projection. 4(c) and 4(d): The same for relative bias magnitude. 4(e): The number of training sites when the target training set radius was set to 400 km.
Source: The borders are adapted from Finnish Meteorological Institute (2016).

prediction error was recorded for pine, spruce, and birch (=99.6% of the data), as well as for the whole data set. The two birch subspecies, silver birch and downy birch, were analysed together. Fig. 5 shows the error convergence.

Fig. 5(a) shows that all species benefit from increasing data in DBH prediction, but the decreasing average RMSE plateaus practically at 35%. Fig. 5(b) shows the same results for stem volume. Pine and spruce benefit substantially more than birch whose range of average improvement is only 3 dm^3. The bias magnitude in DBH and stem volume prediction in Figs. 5(c) and 5(d) show that the average bias magnitude does not considerably improve when more training data are

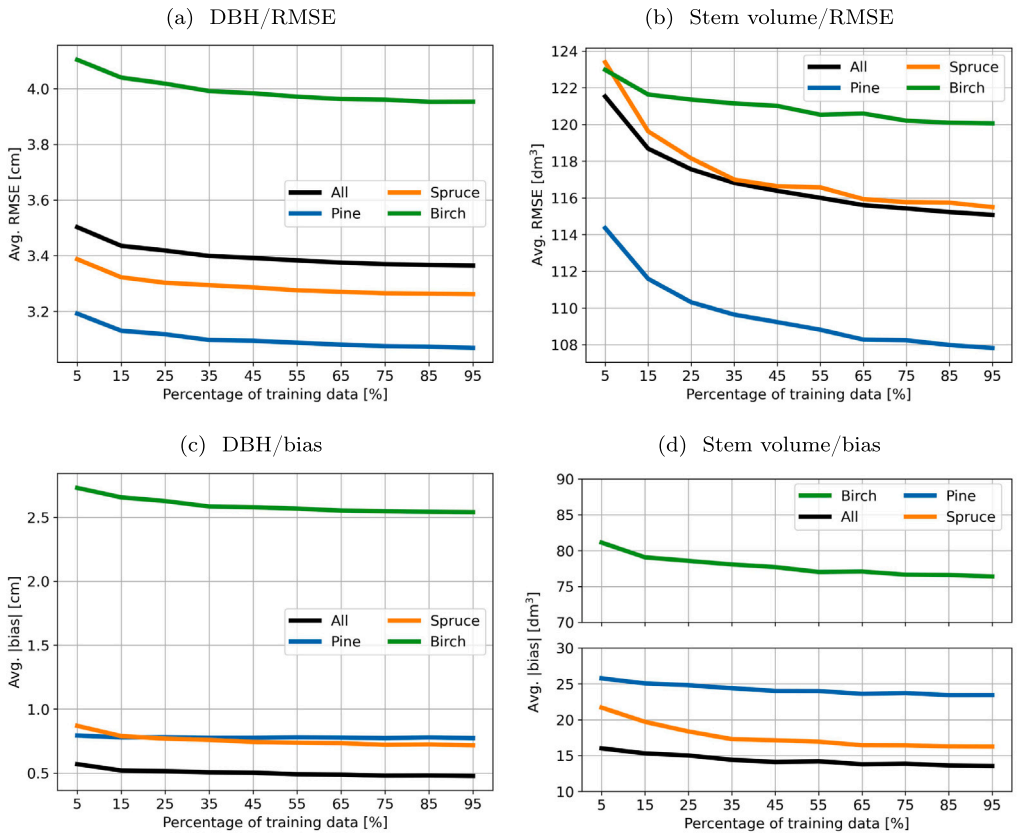

Fig. 5. 5(a) and 5(b): The average RMSE as a function of training data increments. 5(c) and 5(d): The same for bias magnitude. Note that it is possible for the total bias magnitude to be smaller than the combined bias magnitude of pine, spruce, and birch because of removing the negative values when calculating the separate magnitudes of each species.

added. The analysis shows that birch produces the largest RMSE and absolute bias values.

4.3. Robustness of prediction with distant training data

The second test was performed simultaneously while calculating the prediction error increase for each original target site. While incrementally adding the closest training sites of the original target site to the training data set, the prediction error increase was also calculated for each auxiliary target site that was neither the original target site nor in the training data set. The average distance between each site in the training set and the auxiliary test site was calculated, and the minimum error value for each site was subtracted in order to standardise the error increase values between all the test sites. The increase in error was averaged inside a 100 km sliding window between each auxiliary test site. Fig. 6 shows the average error increase for training set sizes of 1, 3, 5, and 7 sites.

The RMSE values in Figs. 6(a) and 6(b) show that the larger the data set, the smaller the error stays with distant training data, although with the most extreme distance and with some mid-range intervals, some smaller data sets produce better results. The use of coordinates gives no significant advantage over the model that does not use coordinates.

With bias magnitude increase in Figs. 6(c) and 6(d), the use of coordinates slightly slows the bias magnitude increase with large data sets of five and seven sites, and with moderate distances of less than 600 km. The effect is more noticeable with DBH bias magnitude compared to stem volume bias magnitude. The small data sets of one and three sites become almost as biased with distances of over 400–500 km.

Regardless of the training set size, RMSE and especially bias magnitude grow significantly as a function of distance, and the size of the set does not effectively protect against the degradation of prediction accuracy. Table 6 collects the slopes of the linear trends fitted for data set sizes of one and seven sites. The fits are plotted in Appendix A.

4.4. Inclusion of on-site calibration data

The third test tested a minimal working setup in which a model was trained with one site at an approximate distance of 200, 400, and 575 km. Then, p percentage of on-site calibration data were mixed with the training data. Each increment in p was resampled, and the model was always retrained. Each increment was repeated five times for each target site, and the average of each run and target site was calculated. The same procedure was repeated for a model that used only on-site data, so that p percentage was used for training and $1 - p$ percentage

Table 6
The mean slope values for the model using coordinates. The bootstrapped 95 % confidence interval is reported in brackets. The unit is for DBH is cm/100 km and dm^3/100 km for stem volume.

	One site	Seven sites
DBH/RMSE	0.28 [0.24, 0.32]	0.27 [0.19, 0.33]
Stem volume/RMSE	13.18 [10.99, 15.73]	8.08 [6.34, 9.63]
DBH/bias magnitude	0.42 [0.36, 0.48]	0.39 [0.29, 0.49]
Stem volume/bias magnitude	12.01 [9.48, 14.68]	8.32 [5.88, 10.70]

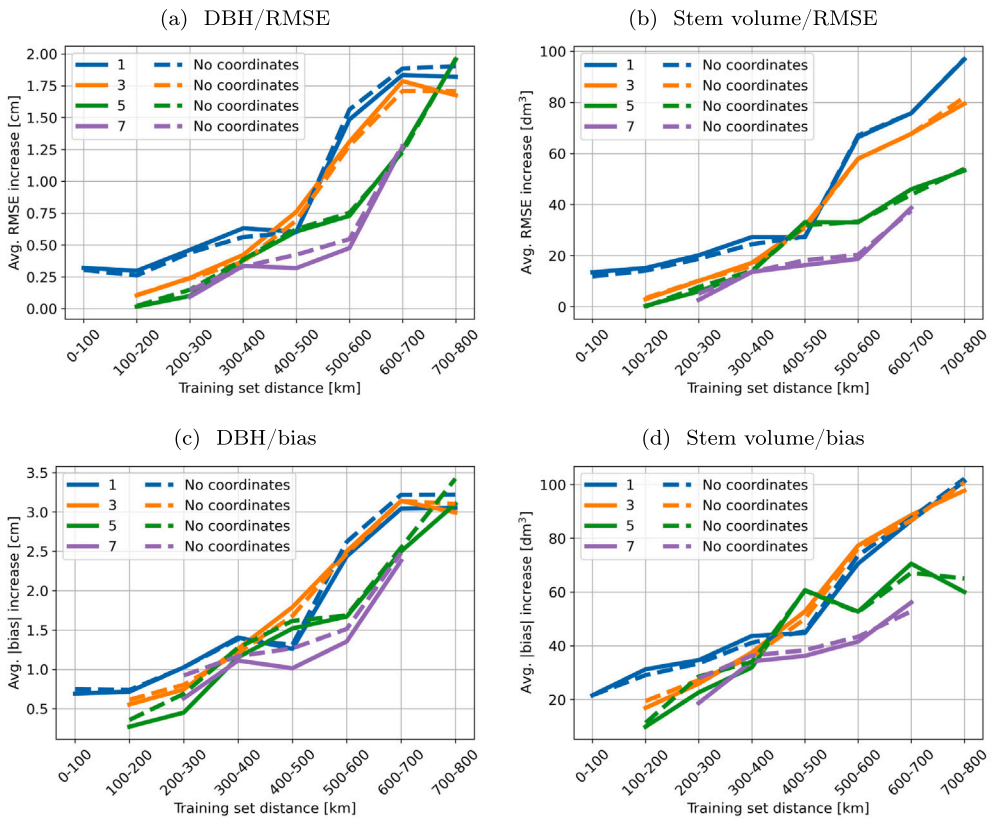

Fig. 6. 6(a) and 6(b): The average increase above site-wise minimum in RMSE as a function of distance. Training sets with 1, 3, 5, and 7 sites tested. 6(c) and 6(d): the same for bias magnitude.

for testing. This model is called the baseline model. Fig. 7 shows the results.

The decrease in RMSE in DBH prediction in Fig. 7(a) shows that when no data from the target site are mixed with the training data, the errors are large. After including only 5% of the on-site data, the models using coordinates are rather indistinguishable from the baseline model. The models that do not use coordinates continue to improve, but never reach the baseline model.

With the stem volume RMSE values in Fig. 7(b), the results show that models trained with a close training site (200 km) slightly outperform the baseline model. When 20% of the on-site data are mixed with the training data, the distant models (>200 km) perform equally with the baseline model. The models that do not use coordinates perform better than with DBH prediction, and even reach the baseline model. Unlike the RMSE values, the bias magnitude never reaches the baseline magnitude both with DBH and stem volume prediction. Interestingly, the models whose average distance between the target site and the training site is 400 km appear to be more biased than those who are trained with training sites whose average distance is 575 km. This might be caused by some exceptionally biased models.

5. Discussion

5.1. The predictions are locally precise, but biased, and not robust far away from the training data area

The first test helps in determining the area inside which the collected training data can be used to train a good model for that specific area. The results imply that a model for individual tree attributes predicts accurately when its training data are collected within the radius of 400 km. The average (minimum and maximum) increase compared to a model that only uses on-site data is 0.8 (0.1–1.9) %-points in RMSE% in DBH prediction and 1.3 (−0.2–3.6) %-points in stem volume prediction. Beyond that radius, the accuracy of the model that uses coordinates as predictors does not substantially improve or degrade. The accuracy of the model that does not use coordinates degrades already after 300 km. This contradicts the findings of Kotivuori et al. (2016) who noted that the use of coordinates did not improve their linear model of area-based volume. However, their model used only two regression features, which might have decreased the impact of coordinates.

Even with the optimal training set radius, the models trained with data collected outside the target site are heavily biased. On average (minimum and maximum), the increase in relative bias magnitude is approximately 2.2 (0.7–5.0) %-points in DBH prediction and 3.5 (−0.1–8.7) %-points in stem volume prediction compared to a value obtained from a two-fold cross-validated baseline model using only on-site data.

Previous studies have reported results where a model trained for one area was evaluated in two nearby areas, such that the setup resembled the first test of this study. In those tests, the amount of data was not varied, and the minimum error was not sought. A comparison shows that the results of this study regarding RMSE% are better than those observed earlier. The results by Karjalainen et al. (2019) suggest an average increase in RMSE% between 2.7–3.1 %-points in DBH prediction and 7.2–8.9 %-points in stem volume prediction when predicting outside the training area. Similarly, Korhonen et al. (2019) reports an increase of 3.8 %-points in RMSE% in DBH prediction. In this study, the average increase in RMSE% was 0.8 %-points in DBH prediction and 1.3 %-points in stem volume prediction. The results of this study were achieved with considerably larger training data set (approximately 65 000–134 000 samples versus less than two thousand), which is most likely the main reason why the transferred predictions are better in this study. For bias magnitude, Karjalainen et al. (2019) obtained an average increase between 1.3–1.4 %-points in DBH prediction and 0.4–4.0 %-points increase in stem volume prediction, while Korhonen

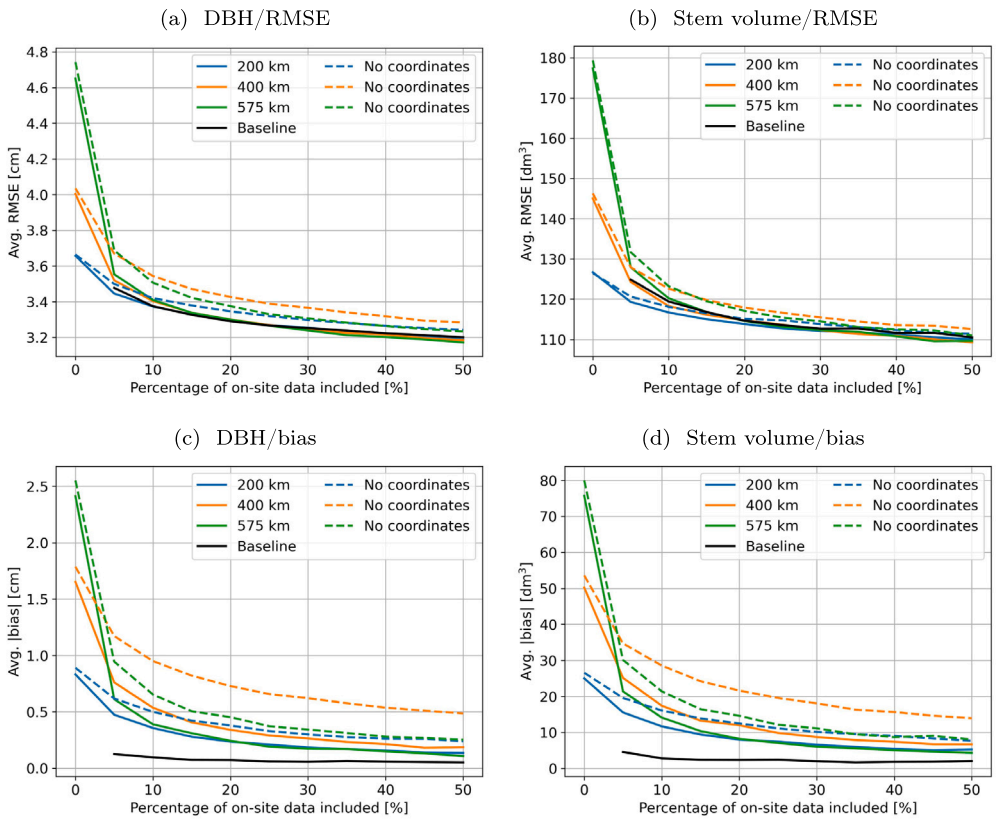

Fig. 7. 7(a) and 7(b): The decrease in average RMSE as a function of mixing proportion p. 7(c) and 7(d): The same for bias magnitude.

et al. (2019) observed an increase of 5.6 %-points in bias magnitude when predicting individual tree DBH. The average values obtained in this study were between those observed earlier, but the variation was large and there were smaller and considerably larger values.

With a radius smaller than 400 km, not enough sites were included in the training set, and the accuracy was suboptimal. In this study, it was not possible to test whether the results had been better if more data were included within a smaller radius. This is because of limited number of sites. In contrast, had there been fewer sites, such that very little data would have been inside the 400 km radius, the limit would probably have stayed the same. This is because the more distant, and thus worse, data would have not included better training samples, as was shown in this study in the first test. The 400 km radius is an optimum between the amount of data that needs to be collected and the area that the collected data cover.

While the first test tested how well the model performs inside the radius induced by the training sites, the second test tested how well the model trained with the same data set could be used outside the area. Unfortunately, the results were not good. The training set size did not effectively protect against the degradation of prediction accuracy caused by the distance between the target site and training sites, as the growth rate was more or less the same. This is observed in Fig. 6 and Table 6. This emphasises that, in order to collect a good training set for individual tree attribute prediction, the data should be collected close to the target sites.

5.2. Model calibration using on-site data rapidly improves the accuracy

While the results of the second test could be used to support systematic nationwide field reference collection, the third test showed that existing data could be augmented with data from the target site in order to improve prediction accuracy. The results were promising, as data collected from a reasonable distance of 400 km or less and augmented with as little as 5% (≈ 550 trees in this study) of on-site data produced results that were comparable or even better than the baseline model. That was, if coordinates could be effectively exploited. The results are in line with past research, as quick convergence of error has been observed when using small amounts of local calibration data in area-based prediction (Latifi and Koch, 2012; Kotivuori et al., 2016), and in individual tree attribute predictions (Korhonen et al., 2019; Sumnall et al., 2024).

However, bias seemed to be much more difficult to improve. As Table 3 shows, the site-wise average values tend to be smaller in north than in south, which is the reason why bias magnitude converged slowly (see also Lappi (2001)). Even with the models that used coordinates, the convergence was slow, and bias magnitude never quite reached the baseline model. Empirical bias reduction could be possible, but this would require studying a model that would take into account the location of the training site and different magnitude of bias per tree species and tree size.

The random forest algorithm is effective in using coordinates as features. If it is assumed that using only the on-site data would minimise the leaf node variance loss function that was used for training the model, then the random forest algorithm should filter out the data that are collected from outside the target site when training the model. This is because the algorithm chooses the optimal splitting feature for reducing the leaf node variance. By assumption, the coordinates are at some point in the tree the optimal features, and they are chosen. Thus, the algorithm prefers the on-site data. This explains why the error difference in Fig. 7 between the baseline model and the models using coordinates as features quickly converged.

5.3. Effects caused by the study setup choices

In this study, each site contained approximately 11 000 trees. Collecting such a number of trees manually for model training data

is laborious. Some insight regarding how well a model trained with limited data collected outside the target area can be obtained from the results that showed how the test error converged with small increments of training data. The results showed that RMSE can be improved more than bias, and approximately 50% of the data used in this study would have been enough for 1% higher error than what was obtained when 95% of the data collected within the radius of 400 km were used for training.

This study did not try to find means for error mitigation when using a transferred model, apart from on-site data collection and the use of coordinates as features. Another study should be conducted to test if some features or information could be used for error mitigation. For example, an empirical bias mitigation model could be tested for birch, which seemed to be the largest error source for the increased RMSE, but especially for bias. The cause of this was not evident from this study.

The results of this study were obtained with the specific characteristics of the Finnish geography and climate conditions. Thus, the results may not generalise directly to other countries. For example, three of the test sites (Sodankylä, Kolari, Savukoski) are completely located above the Arctic Circle, and some parts of the Rovaniemi test site, too. Areas north to the Arctic Circle are characterised by short growing seasons (Fig. 1(c), and e.g. Kersalo and Pirinen (2009)) Also, the proximity of sea to some areas, such as Turku, Pori, and Oulu causes some climatic conditions that are not present in countries without coasts or with vast inland areas.

Kotivuori et al. (2016) showed that ALS scanner has some effect on the transferability results. In this study, the magnitude of this error source was not estimated, but most of the differences caused by the sensor are likely in the features that represent intensity values. Those features were normalised in each site in order to mitigate the differences. However, the results may change if different point cloud characteristics, such as denser data, and methods, such as deep learning applied directly to the point clouds, were used.

The choice of the random forest method for the regression task was based on its reputation for being a good classical machine learning method for individual tree tasks that use hand-crafted features (Malek et al., 2019). Importantly, a random forest model is simple to use, requires only a few hyperparameters, is scalable, and is more robust against overfitting (Ghojogh and Crowley, 2023), which could decrease the transferability. Similar protection against overfitting could be achieved with simpler models but such models may not predict the individual tree attributes as accurately (Malek et al., 2019; Yates et al., 2018). The transferability of more complex, neural network-based methods could be tested with denser point clouds because they benefit from direct use of a point cloud for prediction tasks (Wielgosz et al., 2024; Xiang et al., 2024; Taher et al., 2025). However, the need for very dense point clouds may prevent the use of neural network-based solutions at national scale due to memory space and time complexity issues.

6. Conclusions

This study investigated how the prediction capability of individual tree models changes when their training data are collected from outside the area where the models are applied. Throughout the study, coordinates proved to be beneficial as training data features. It was observed that the models that use coordinates as features reach their RMSE minima when the training data are collected between 400–500 km, but the improvement is minimal after 200–300 km. (answered the research question 1). The minimum was obtained using between 65 000–134 000 (98 000 on average) training samples. The average increase in RMSE% compared to a model that used only on-site data was 0.8 %-points in DBH prediction and 1.3%-points in stem volume prediction. Relative bias magnitude increased more, and the average increase was 2.2 %-points in DBH and 3.5 %-points in stem volume prediction.

Training data set size did not completely protect against the distance-related decline in model accuracy. Larger data sets produced errors that increased at a similar but slightly lower rate compared to the smaller training sets. The increase in RMSE in DBH prediction was between 0.27–0.28 cm/100 km, and in stem volume prediction, the increase was between 8.08–13.18 dm^3/100 km. The same values for bias magnitude were 0.39–0.42 cm/100 km for DBH prediction and 8.32–12.01 dm^3/100 km in stem volume prediction. Thus, the training data set should be collected as close as possible to the target sites. This answered the research question 2.

The final observation was related to the quick convergence of RMSE when using on-site calibration data. When one training site was at a maximum distance of 400 km, only 5% (≈ 550 trees) of the on-site data mixed with the training data were enough for almost indistinguishable or even better RMSE results than from a model that used only on-site data. The result was obtained from a model that was able to use coordinates as predictors. However, bias magnitude never reached the values obtained from a model that used only on-site data. This answered the research question 3.

The results of this study help in planning the reference data collection campaigns at the national level. The results should be repeated in other countries. Furthermore, denser data, more complex regression methods, and species-specific responses may change the results, which leaves opportunities for further research.

CRediT authorship contribution statement

Valtteri Soininen: Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization. **Xiaowei Yu:** Writing – review & editing, Software, Data curation. **Matti Hyyppä:** Writing – review & editing, Software, Data curation. **Juha Hyyppä:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Funding

We acknowledge the Research Council of Finland projects "High-performance computing allowing high-accuracy country-level individual tree carbon sink and biodiversity mapping" (decision number 359203) and "Collecting accurate individual tree information for harvester operation decision making" (decision number 359554) for financial support. The work was done under the Research Council of Finland flagship project "Forest-Human–Machine Interplay – Building Resilience, Redefining Value Networks and Enabling Meaningful Experiences" (decision number 359175) and utilising the research infrastructure of "Measuring Spatiotemporal Changes in Forest Ecosystem" (decision number 346382).

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.

Acknowledgements

Finnish Forestry Centre is acknowledged for collecting the individual tree reference data for the research. Teemu Mielonen, National Land Survey of Finland, is acknowledged for the process of helping to get the national ALS data for the research.

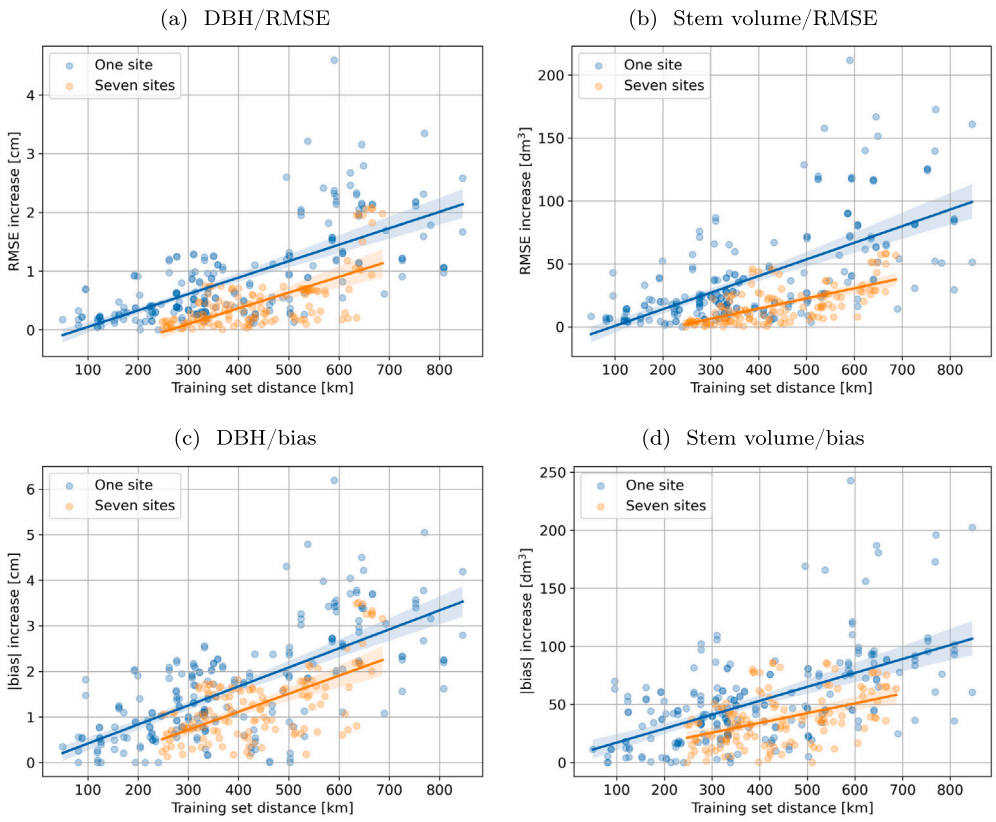

Fig. A.1. The slopes and individual data points for Fig. 6. The shaded area represents the bootstrapped 95% confidence bands.

Appendix A. Slopes

This Appendix presents the individual linear fits and data points for Fig. 6. The results are from a model that uses coordinates.

As the curves in Fig. 6 show, a linear fit is not a perfect fit for the data. Additionally, the data points in Fig. A.1 show heteroskedasticity, such that with long distances of over 500 km, the variance of the points increases, especially when using one training site. This is reflected in the trendline as bootstrap confidence bands that widen at longer distances. However, the 95% intervals do not contain horizontal lines or lines that decrease. Thus, the effect of increasing error is persistent. The linear fit provides an approximate fit that can be used as a guiding value of average prediction degradation rate when making predictions using distant training data.

Appendix B. Supplementary data

The order of choosing the test and training sites can be watched online at https://doi.org/10.1016/j.srs.2025.100310.

Data availability

The data sets for soil and temperature sum are available free of charge in Natural Resources Institute Finland (2024) and Finnish Meteorological Institute (2016). The point clouds used in the study are available in National Land Survey of Finland (2020).

References

Bonan, G.B., Shugart, H.H., 1989. Environmental factors and ecological processes in boreal forests. Annu. Rev. Ecol. Syst. 20, 1–28. http://dx.doi.org/10.1146/annurev.es.20.110189.000245.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (3), 5–32. http://dx.doi.org/10.1023/A:1010933404324.

Cysneiros, V.C., de Souza, F.C., Gaui, T.D., Pelissari, A.L., Orso, G.A., do Amaral Machado, S., de Carvalho, D.C., Silveira-Filho, T.B., 2021. Integrating climate, soil and stand structure into allometric models: An approach of site-effects on tree allometry in Atlantic Forest. Ecol. Indic. 127, e107794. http://dx.doi.org/10.1016/j.ecolind.2021.107794.

Domingo, D., Alonso, R., Lamelas, M.T., Montealegre, A.L., Rodríguez, F., de la Riva, J., 2019. Temporal transferability of pine forest attributes modeling using low-density airborne laser scanning data. Remote. Sens. 11, 261. http://dx.doi.org/10.3390/rs11030261.

Eerikäinen, K., 2009. A multivariate linear mixed-effects model for the generalization of sample tree heights and crown ratios in the finnish national forest inventory. For. Sci. 55 (6), 480–493. http://dx.doi.org/10.1093/forestscience/55.6.480.

van Ewijk, K., Tompalski, P., Treitz, P., Coops, N.C., Woods, M., Pitt, D., 2020. Transferability of ALS-derived forest resource inventory attributes between an eastern and western Canadian boreal forest mixedwood site. Can. J. Remote Sens. 46, 214–236. http://dx.doi.org/10.1080/07038992.2020.1769470.

Finnish Meteorological Institute, 2016. Vuorokauden keskilämpötila, 10 km, 1961–2023, GeoTIFF. Fairdata.fi URL http://urn.fi/urn:nbn:fi:csc-kata00001000000000000661, License: CC BY 4.0 https://creativecommons.org/licenses/by/4.0/..

Ghojogh, B., Crowley, M., 2023. The theory behind overfitting, cross validation, regularization, bagging, and boosting: Tutorial. arXiv:1905.12787, URL https://arxiv.org/abs/1905.12787.

Haara, A., Kangas, A., Tuominen, S., 2019. Economic losses caused by tree species proportions and site type errors in forest management planning. Silva Fenn. 53, 10089. http://dx.doi.org/10.14214/sf.10089.

Hyyppä, J., Inkinen, M., 1999. Detecting and estimating attributes for single trees using laser scanner. Photogramm. J. Finl. 16, 27–42.

Hyyppä, E., Kukko, A., Kaijaluoto, R., White, J.C., Wulder, M.A., Pyörälä, J., Liang, X., Yu, X., Wang, Y., Kaartinen, H., et al., 2020a. Accurate derivation of stem curve and volume using backpack mobile laser scanning. ISPRS J. Photogramm. Remote Sens. 161, 246–262. http://dx.doi.org/10.1016/j.isprsjprs.2020.01.018.

Hyyppä, M., Turppa, T., Hyyti, H., Yu, X., Handolin, H., Kukko, A., Hyyppä, J., Virtanen, J.P., 2024. Concepts towards nation-wide individual tree data and virtual forests. ISPRS Int. J. Geo-Inf. 13, e424. http://dx.doi.org/10.3390/ijgi13120424.

Hyyppä, E., Yu, X., Kaartinen, H., Hakala, T., Kukko, A., Vastaranta, M., Hyyppä, J., 2020b. Comparison of backpack, handheld, under-canopy UAV, and above-canopy UAV laser scanning for field reference data collection in boreal forests. Remote. Sens. 12, 1–31. http://dx.doi.org/10.3390/rs12203327.

Kaartinen, H., Hyyppä, J., Yu, X., Vastaranta, M., Hyyppä, H., Kukko, A., Holopainen, M., Heipke, C., Hirschmugl, M., Morsdorf, F., et al., 2012. An international comparison of individual tree detection and extraction using airborne laser scanning. Remote. Sens. 4 (4), 950–974. http://dx.doi.org/10.3390/rs4040950.

Kangas, A., Henttonen, H., Pitkänen, T., Sarkkola, S., Heikkinen, J., 2020. Re-calibrating stem volume models – is there change in the tree trunk form from the 1970s to the 2010s in Finland? Silva Fenn. 54, 1–23. http://dx.doi.org/10.14214/sf.10269.

Kangas, A., Pitkänen, T., Mehtätalo, L., Heikkinen, J., 2023. Mixed linear and non-linear tree volume models with regional parameters to main tree species in Finland. For.: An Int. J. For. Res. 96, 188–206. http://dx.doi.org/10.1093/forestry/cpac038.

Karjalainen, T., Korhonen, L., Packalen, P., Maltamo, M., 2019. The transferability of airborne laser scanning based tree-level models between different inventory areas. Can. J. For. Res. 49, 228–236. http://dx.doi.org/10.1139/cjfr-2018-0128.

Kersalo, J., Pirinen, P., 2009. Suomen Maakuntien Ilmasto, 2009:8 Yliopistopaino, Helsinki, URL http://hdl.handle.net/10138/15734, [In Finnish].

Kettunen, M., Vihervaara, P., Kinnunen, S., D'Amato, D., Badura, T., Argimon, M., Ten Brink, P., 2012. Socio-Economic Importance of Ecosystem Services in the Nordic Countries - Synthesis in the Context of the Economics of Ecosystems and Biodiversity (TEEB). Nordic Council of Ministers, Copenhagen, http://dx.doi.org/10.6027/TN2012-559.

Korhonen, L., Repola, J., Karjalainen, T., Packalen, P., Maltamo, M., 2019. Transferability and calibration of airborne laser scanning based mixed-effects models to estimate the attributes of sawlog-sized scots pines. Silva Fenn. 53, 10179. http://dx.doi.org/10.14214/sf.10179.

Kotivuori, E., Korhonen, L., Packalen, P., 2016. Nationwide airborne laser scanning based models for volume, biomass and dominant height in Finland. Silva Fenn. 50, 1567. http://dx.doi.org/10.14214/sf.1567.

Laasasenaho, J., 1982. Taper curve and volume function for pine, spruce and birch. Commun. Inst. For. Fenn. 108, 1–74, URL http://urn.fi/URN:ISBN:951-40-0589-9.

Lappi, J., 2001. Forest inventory of small areas combining the calibration estimator and a spatial model. Can. J. For. Res. 31, 1551–1560. http://dx.doi.org/10.1139/cjfr-31-9-1551.

Latifi, H., Koch, B., 2012. Evaluation of most similar neighbour and random forest methods for imputing forest inventory variables using data from target and auxiliary stands. Int. J. Remote. Sens. 33, 6668–6694. http://dx.doi.org/10.1080/01431161.2012.693969.

Lines, E.R., Zavala, M.A., Purves, D.W., Coomes, D.A., 2012. Predictable changes in aboveground allometry of trees along gradients of temperature, aridity and competition. Glob. Ecol. Biogeogr. 21, 1017–1028. http://dx.doi.org/10.1111/j.1466-8238.2011.00746.x.

Malek, S., Miglietta, F., Gobakken, T., Næsset, E., Gianelle, D., Dalponte, M., 2019. Prediction of stem diameter and biomass at individual tree crown level with advanced machine learning techniques. IForest 12 (3), 323–329. http://dx.doi.org/10.3832/ifor2980-012.

Mehtätalo, L., de Miguel, S., Gregoire, T.G., 2015. Modeling height-diameter curves for prediction. Can. J. For. Res. 45 (7), 826–837. http://dx.doi.org/10.1139/cjfr-2015-0054.

Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sens. Environ. 80 (1), 88–99. http://dx.doi.org/10.1016/S0034-4257(01)00290-5.

National Land Survey of Finland, 2020. Laser scanning data 5 p. URL https://www.maanmittauslaitos.fi/en/maps-and-spatial-data/datasets-and-interfaces/product-descriptions/laser-scanning-data-5-p, License: https://www.maanmittauslaitos.fi/en/laser-skanning-data/terms-of-use..

Natural Resources Institute Finland, 2024. Monilähteisen valtakunnan metsien inventoinnin (MVMI) kartta-aineisto 2021: Kasvupaikan päätyyppi / Site main class. Fairdata.fi URL http://urn.fi/urn:nbn:fi:att:f8f2ff42-6ab6-45a6-a822-b8d62e48afe8, License: CC BY 4.0 https://creativecommons.org/licenses/by/4.0/..

Niemi, M., Vastaranta, M., Peuhkurinen, J., Holopainen, M., 2015. Forest inventory attribute prediction using airborne laser scanning in low-productive forestry-drained boreal peatlands. Silva Fenn. 49 (2), 1218. http://dx.doi.org/10.14214/sf.1218.

Pfeifer, N., Briese, C., 2007. Laser scanning–principles and applications. In: Geosiberia 2007 - International Exhibition and Scientific Congress. EAGE Publications BV, Houten, the Netherlands, pp. cp–59–00077. http://dx.doi.org/10.3997/2214-4609.201403279.

Schneider, R., Franceschini, T., Fortin, M., Saucier, J.P., 2018. Climate-induced changes in the stem form of 5 North American tree species. Forest Ecol. Manag. 427, 446–455. http://dx.doi.org/10.1016/j.foreco.2017.12.026.

Sumnall, M.J., Carter, D.R., Albaugh, T.J., Platt, E., Host, T., Cook, R.L., Campoe, O.C., Rubilar, R.A., 2024. Evaluating the transferability of airborne laser scanning derived stem size prediction models for Pinus taeda L. stem size estimation to two different locations and acquisition specifications. Int. J. Remote Sens. 45, 5267–5294. http://dx.doi.org/10.1080/01431161.2024.2370499.

Taher, J., Hyyppä, E., Hyyppä, M., Salolahti, K., Yu, X., Matikainen, L., Kukko, A., Lehtomäki, M., Kaartinen, H., Thurachen, S., Litkey, P., Luoma, V., Holopainen, M., Kong, G., Fan, H., Rönnholm, P., Polvivaara, A., Junttila, S., Vastaranta, M., Puliti, S., Astrup, R., Kostensalo, J., Myllymäki, M., Kulicki, M., Stereńczak, K., Pires, R.d., Valbuena, R., Carbonell-Rivera, J.P., Torralba, J., Chen, Y.-C., Winiwarter, L., Hollaus, M., Mandlburger, G., Takhtkeshha, N., Remondino, F., Lisiewicz, M., omiej Kraszewski, B., Liang, X., Chen, J., Ahokas, E., Karila, K., Vezeteu, E., Manninen, P., Näsi, R., Hyyti, H., Pyykkönen, S., Hu, P., Hyyppä, J., 2025. Multispectral airborne laser scanning for tree species classification: a benchmark of machine learning and deep learning algorithms. arXiv:2504.14337, URL https://arxiv.org/abs/2504.14337.

Takhtkeshha, N., Mandlburger, G., Remondino, F., Hyyppä, J., 2024. Multispectral light detection and ranging technology and applications: A review. Sensors 24, http://dx.doi.org/10.3390/s24051669.

Tompalski, P., White, J., Coops, N., Wulder, M., 2019. Demonstrating the transferability of forest inventory attribute models derived using airborne laser scanning data. Remote. Sens. Environ. 227, 110–124. http://dx.doi.org/10.1016/j.rse.2019.04.006.

Tomppo, E., Heikkinen, J., Henttonen, H.M., Ihalainen, A., Katila, M., Mäkelä, H., Tuomainen, T., Vainikainen, N., 2011. Designing and Conducting a Forest Inventory - case: 9th National Forest Inventory of Finland. Springer Dordrecht, http://dx.doi.org/10.1007/978-94-007-1652-0.

Wielgosz, M., Puliti, S., Xiang, B., Schindler, K., Astrup, R., 2024. SegmentAnyTree: A sensor and platform agnostic deep learning model for tree segmentation using laser scanning data. Remote Sens. Environ. 313, e114367. http://dx.doi.org/10.1016/j.rse.2024.114367.

Xiang, B., Wielgosz, M., Kontogianni, T., Peters, T., Puliti, S., Astrup, R., 2024. Automated forest inventory: Analysis of high-density airborne LiDAR point clouds with 3D deep learning. Remote Sens. Environ. 305, e114078. http://dx.doi.org/10.1016/j.rse.2024.114078.

Yates, K.L., Bouchet, P.J., Caley, M.J., Mengersen, K., Randin, C.F., Parnell, S., Fielding, A.H., Bamford, A.J., Ban, S., Barbosa, A.M., Dormann, C.F., Elith, J., Embling, C.B., Ervin, G.N., Fisher, R., Gould, S., Graf, R.F., Gregr, E.J., Halpin, P.N., Heikkinen, R.K., Heinänen, S., Jones, A.R., Krishnakumar, P.K., Lauria, V., Lozano-Montes, H., Mannocci, L., Mellin, C., Mesgaran, M.B., Moreno-Amat, E., Mormede, S., Novaczek, E., Oppel, S., Crespo, G.O., Peterson, A.T., Rapacciuolo, G., Roberts, J.J., Ross, R.E., Scales, K.L., Schoeman, D., Snelgrove, P., Sundblad, G., Thuiller, W., Torres, L.G., Verbruggen, H., Wang, L., Wenger, S., Whittingham, M.J., Zharikov, Y., Zurell, D., Sequeira, A.M., 2018. Outstanding challenges in the transferability of ecological models. Trends Ecol. Evol. 33, 790–802. http://dx.doi.org/10.1016/j.tree.2018.08.001.

Yu, X., Hyyppä, J., Kukko, A., Maltamo, M., Kaartinen, H., 2006. Change detection techniques for canopy height growth measurements using airborne laser scanner data. Photogramm. Eng. Remote. Sens. 72 (12), 1339–1348. http://dx.doi.org/10.14358/PERS.72.12.1339.

Yu, X., Hyyppä, J., Vastaranta, M., Holopainen, M., Viitala, R., 2011. Predicting individual tree attributes from airborne laser point clouds based on the random forests technique. ISPRS J. Photogramm. Remote Sens. 66, 28–37. http://dx.doi.org/10.1016/j.isprsjprs.2010.08.003.