Forestry: An International Journal of Forest Research, 2026, 99(4), cpag071
https://doi.org/10.1093/forestry/cpag071
Published: 08 September 2026
Original Article

Institute of Chartered Foresters OXFORD

Predicting forest inventory variables with remote sensing data: the role of number and locality of the training plots

Timo P. Pitkänen [1,*], Annika Kangas [2], Mari Myllymäki [1,3], Petteri Packalen [1]

[1] Bioeconomy and Environment, Natural Resources Institute Finland (Luke), Latokartanonkaari 9, 00790 Helsinki, Finland
[2] Bioeconomy and Environment, Natural Resources Institute Finland (Luke), Yliopistokatu 6, 80100 Joensuu, Finland
[3] Department of Mathematics and Statistics, Faculty of Science, University of Helsinki, P.O. Box 68 (Pietari Kalmin katu 5), 00014 University of Helsinki, Finland
*Corresponding author. E-mail: timo.p.pitkanen@luke.fi.

Abstract

Field measurements of national forest inventories (NFIs) and auxiliary data derived from remote sensing can be used to model forest inventory variables. Because NFI observations typically cover large areas, training data for modeling is often limited by the number and proximity of the available plots. Determining suitable thresholds for these, however, is complicated and depends on various factors. In this study, we examine how the number and locality of the NFI plots affect the modeling accuracy. With locality, we refer to the spatial proximity of the training plots to the target location, which is often considered a proxy of their representativeness. We used field measurements from the Finnish NFI, collected mainly from managed, conifer-dominated boreal forests, and modeled growing stock volumes and trees' dominant heights with Sentinel-2 and airborne laser scanning (ALS) data. We then compared prediction errors derived from varying numbers of training plots and modeling attempts where we purposely excluded the nearby plots within a specific radius. As a special case of training plots' representativeness, we also assessed if models estimated using ALS features were transferable over the production area borders, i.e. scanned with similar quality standards but at different times and using different scanners. Our results indicated that a set of 100 training plots could be regarded as the minimum preferred count regardless of the auxiliary data. Locality affected modeling bias, which increased when the exclusion radius was extended to 50 km or beyond. This bias connected particularly to models' abilities for adapting to local conditions which resulted in regional under- and overestimates, but these effects were only mildly observable through global statistics over all the applied validation data. Differences between ALS production areas were generally small, but they varied by inventory variable and were more pronounced for features derived from first echoes than from last or all echoes.

Keywords wall-to-wall forest maps, random forest modeling, optical satellite data, airborne laser scanning, prediction accuracy

Introduction

National forest inventories (NFIs), which rely on measuring field plots according to a probability sampling design, have an essential role in assessing large-scale forest resources using design-based inference. In recent decades, however, applications have become more diverse, creating demand for timely estimates which focus on relatively small spatial units (McRoberts et al. 2010). As the availability of accurate and up-to-date remote sensing data sources has simultaneously increased, model-based inference that applies remote sensing data as auxiliary information has attracted more attention (McRoberts 2010, Ståhl et al. 2016, Fassnacht et al. 2024). In this context, estimated models and derived maps rely on correlation between field observations and the remotely sensed features. This makes it possible to predict inventory variables for smaller areas than with design-based methods and to assess forest resources at the level of single forest stands.

Remote sensing auxiliary data sources applied in these models may be broadly categorized into 2D and 3D sources depending on their principal information content. 2D data refers primarily to aerial and satellite images, which represent measurements of reflected electromagnetic radiation at different wavelengths. Aerial images have higher resolution than most satellite images, but fewer spectral bands and less frequent coverage. Their low flying altitude and wide viewing angle can also cause uneven lighting and image distortions (Tuominen and Pekkarinen 2004, Melin et al. 2017). High-resolution satellites offer an alternative to aerial images, but for large-area mapping, mid-resolution images (i.e. with 10–30 m pixel size) are often preferred as they provide a good balance between spatial resolution, coverage, and availability. Data from Sentinel-2 and Landsat-8/9 satellites are widely used because the images are freely available and frequently acquired, making them suitable for example for assessing and monitoring changes in growing stock volume or biomass. (Kangas et al. 2018, Esteban et al. 2019, Fassnacht et al. 2021, Tian et al. 2023). Optical satellite images are, however, sensitive to atmospheric defects and their coarse pixel size does not necessarily correspond well to the field

Handling editor: Dr. Nicholas Coops
Received: 11 March 2026. **Revised:** 13 July 2026. **Accepted:** 2 August 2026

plots, weakening the correlation between reflectance values and field measurements (Saarela et al. 2016, Tuominen et al. 2017, Kangas et al. 2018).

Compared to 2D data sources, 3D auxiliary data provides primarily information on vegetation heights. Common data in this category are point clouds derived from airborne laser scanning (ALS), which are operationally acquired in many countries and can describe forest structure in detail over a diverse range of forest types (White et al. 2016). A common approach is to select ALS points within field plots, summarize them into height and density metrics, and use these features to predict inventory attributes (Næsset 2004, White et al. 2016, Coops et al. 2022). Although these metrics are generally robust and can be applied over large areas, combining data from different campaigns may reduce consistency due to differences in acquisition time, season, point densities, flight parameters, and scanner characteristics (Gobakken and Næsset 2008, Keränen et al. 2016, Maltamo et al. 2025). Digital aerial photogrammetry can also provide 3D data, but it captures only the canopy surface, requires a terrain model for height normalization, and is generally less accurate than ALS. (Goodbody et al. 2019, Noordermeer et al. 2019). 3D-derived features are also often combined with 2D spectral data, which can improve the results, especially for species-specific inventory variables (Fassnacht et al. 2014, Kukkonen et al. 2018, Schumacher et al. 2020).

Apart from the information content of the applied remote sensing data per se, several other factors affect the accuracy of model-based estimates. First, field observations and auxiliary data should have good spatial and temporal alignment, particularly in heterogeneous and dynamic landscapes. Besides differences between field plot sizes and remote sensing pixels, errors in plot positioning can further weaken the results, particularly if the field plot locations have been measured using low-accuracy GNSS equipment (Persson and Ståhl 2020, Saarela et al. 2016). Because simultaneous acquisition of all the data is virtually impossible in large-scale forest inventories, some temporal mismatch needs to be accepted (McRoberts et al. 2018, Knott et al. 2023). The acceptable time gap between field observations and auxiliary data is often based on expert judgment and depends highly on the dynamism of the target landscape. For example, in the context of European managed forests, the lag should preferably be within a few years and not exceed about five years, and changed plots due to known management activities or disturbances should be excluded (Waser et al. 2017, Hawryło et al. 2020, Blickensdörfer et al. 2024).

Another important matter in model-based inference is the representativeness of the training data (Ståhl et al. 2016, Hau et al. 2018, Lister et al. 2020). When large field datasets, such as NFI plots, are used for small-area estimation or wall-to-wall mapping, models are typically built using a subset of close-by plots. This limits training data to plots that, due to spatial autocorrelation, are expected to have forest and environmental conditions most like the target area (Nilsson et al. 2017, Mäkisara et al. 2022, Miettinen et al. 2025). NFI plots however tend to be sparsely located because they are designed to provide unbiased estimates over large areas while avoiding redundant measurements (Henttonen and Kangas 2015, Räty et al. 2019). This creates a challenge in selecting enough training plots while keeping them close to the target area. Modeling performance is also connected to the feature (i.e. covariate) space of the auxiliary data, which should be well covered by the field plots to avoid extrapolation and decrease the modeling bias (Junttila et al. 2013, Wadoux et al. 2019). Auxiliary datasets are also assumed to be consistent across all plots, but this may not hold when multiple datasets are combined and they differ, for example, in atmospheric conditions or laser scanning parameters.

As discussed above, selecting training plots or remote sensing variables is not a straightforward task, but these choices strongly affect model-based inference. To quantify the prediction accuracy and bias over various modeling scenarios, we present a systematic study that utilizes commonly used auxiliary data sources and examines the influence of the number and locality of the training plots. With locality, we refer to the spatial proximity of the training plots to the target location, which has not been commonly studied in detail. We evaluate the effect of locality on training data representativeness by constructing models with similar numbers of plots but different distances to the target location. As field data, we use NFI measurements from boreal forests in Finland, focusing on plot volume and dominant height. Regarding the auxiliary data, we apply ALS data and optical Sentinel-2 images, both alone and combined. Modeling is based on the Random Forest (RF) method and by creating individual models for each respective plot. Our approach is purely experimental rather than an operationally applicable framework due to high computational load but it allows to assess the influence of actual distances between the estimation and prediction locations. With this setting, we aim to provide practical guidance and identify key challenges in building accurate and unbiased models from NFI plots and remote sensing data.

Data and methods

Study area

The study area covers nearly all of Finland, approximately 900 × 550 km, excluding areas without available ALS data (see Remote sensing materials section) (Fig. 1). The northernmost Finland (above 68° N latitude) and Åland Islands in the southwest were also left out, the former due to sparser NFI sampling and the latter because of isolation from the otherwise continuous NFI plot coverage. Most of the study area is characterized by extensive forests and numerous lakes. Topography is relatively flat with gently undulating hills, elevations ranging from generally less than 100 m in most southern and western parts to over 200 m in the eastern and northern parts. The highest hills are in the north and reach elevations of over 500 m, but very few of the applied NFI plots are located above 350 m elevation. Regarding soil types, mineral soils (mainly glacial till) and organic peatlands with bedrock outcrops are common while clay soils dominate part of the coastal zone. Forests belong to the boreal zone and are mostly managed for timber production. The most common tree species are Scots pine (*Pinus sylvestris L.)* and Norway spruce (*Picea abies (L.) H. Karst*) and, to a lesser extent, deciduous trees consisting mainly of two birch species (*Betula pendula* and *Betula pubescens*). Forest characteristics change gradually between the milder maritime climate in the southwest and colder, more continental conditions in the east and north, but there are no clear borderlines or abrupt changes in environmental conditions.

Field data

Field data used in the study were derived from the Finnish NFI measurements between years 2019 and 2023. Individual NFI plots were organized as clusters, consisting of 8–11 plots at 250–450 m distances depending on the sampling subregion and time of their establishment, and clusters were located several kilometers apart (for further details, see Korhonen et al. 2024). At each plot, the field crew

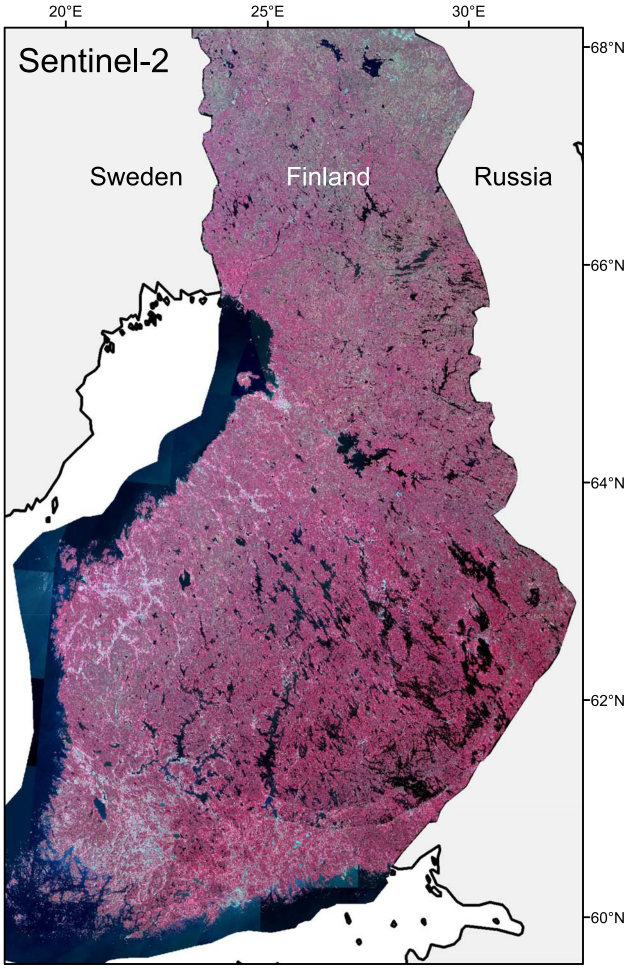

Figure 1 Study area and the data used. More details are provided in the text.

used a handheld GNSS receiver to navigate as close as possible to the predetermined plot location, but the final plot center coordinates were always measured using a survey-grade GNSS device with post-processing procedures. On each individual plot located on forestry land, the locations, diameters at breast height (DBH), tree species, and several other variables were measured from trees within a radius (r) of 9.00 m. These tally trees were selected based on their diameter so that trees with DBH $\geq$ 95 mm were measured from the whole plot, trees with 45 $\leq$ DBH < 95 mm within r = 4.00 m, and trees with DBH < 45 mm using an angle count plot. In total, the number of NFI plots, as limited by the study area, temporal range, coverage of all the auxiliary data, and location on forestry land, was 30 098, divided into 4720 clusters. Of these plots, 64.4% were located on mineral soil and 35.6% on peatland. The number of tallied trees ranged from 0–81 with a median value of 16 trees.

As target variables in our experiment, we focused on plot-level standing wood volumes and dominant height of the trees, the latter being defined as the mean height of 100 thickest trees (by DBH) per hectare. Tree and plot-level volumes were calculated as documented in Mäkisara et al. (2022). Total plot volumes ranged between 0 and 1128 m^3/ha (mean 115 m^3/ha), which were further divided into volumes of coniferous trees (0–1056 m^3/ha, mean value 92 m^3/ha) and deciduous trees (0–637 m^3/ha, mean value 24 m^3/ha), respectively. In total, 10.1% of all plots were treeless with no measurable wood volume, mostly due to recent clearcuts. As the r = 9.00 m plot covers roughly 1/40 of a hectare, dominant heights were calculated as the mean height of the three thickest trees, of which the third one was included using half weight. Tree heights were derived from manual field measurements for sample trees (i.e. a subset of tally trees with extended measurements), and from values modeled according to Myllymäki (2016) as described in Toivonen et al. (2026, Supplement S3) for other tally trees. Dominant heights varied between 0 and 39.3 m with a mean value of 14.6 m. Figure 2 shows maps of all the targeted inventory variables, interpolated over all the selected NFI plots (including also those not covered by the ALS data) at a 1 km cell size. While the value ranges vary depending on the variable, all of them have relatively similar spatial patterns with the largest values in the south and the smallest in the north.

Remote sensing materials

Remote sensing materials consisted of Sentinel-2 Level-2A images and ALS data. Regarding Sentinel-2 images, a cloudless mosaic covering the whole study area was processed using Google Earth Engine (GEE). The mosaic was targeted nominally at year 2021, but due to frequent cloudiness and the short span of the growing season, input images were selected between 1 June and 31 August from years 2020–2022. Initial image tiles were selected with a cloud percentage up to 50%

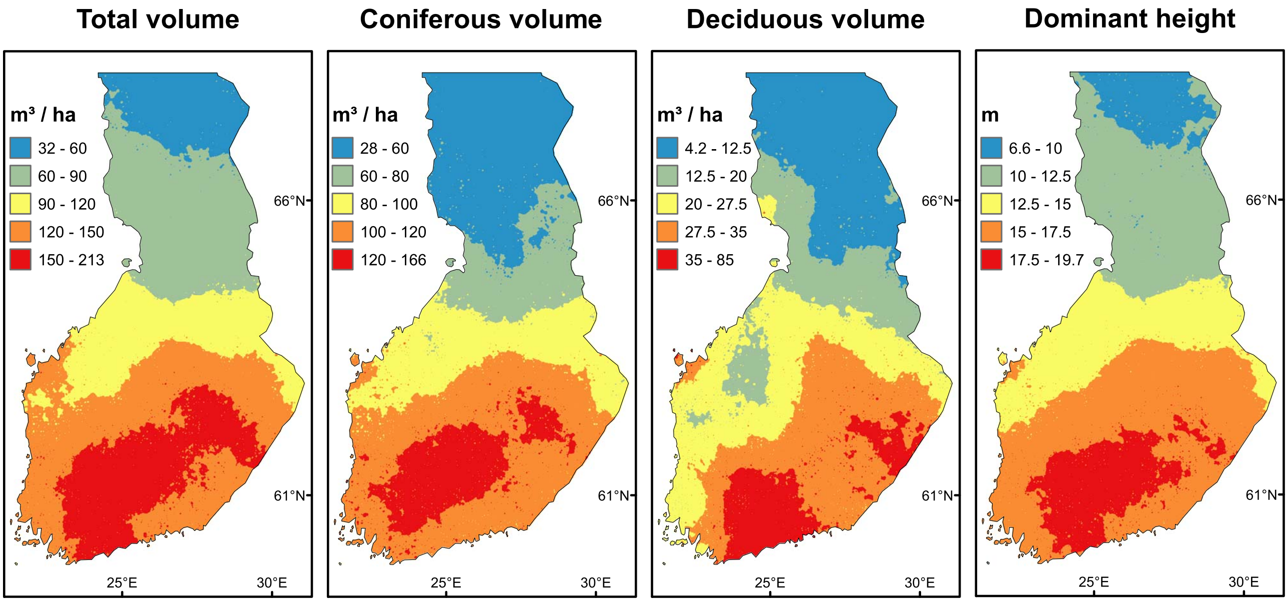

Figure 2 General patterns of the focused inventory variables.

according to Sentinel-2 metadata, their pixels were further evaluated using CloudScore+ quality layers (Pasquarella et al. 2023), and the final mosaic was calculated using band-wise median reflectance values over all the approved pixels. An advantage of CloudScore+ layers compared to many other alternatives is the sensitivity of quality scores not only for clouds and haze but also for shadows.

All Sentinel-2 bands with initial resolutions of 10 or 20 m were included in the mosaic (B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12), and the 20 m bands were resampled to 10 m pixel size using nearest neighbor method. In addition to the initial band-wise reflectance values, generic normalized difference indices (ND) of all the unique two-band combinations were calculated according to Eq. 1:

$$ND_{a,b} = (\rho_b - \rho_a) / (\rho_b + \rho_a) \quad (1)$$

where ρa,b are reflectance values of bands a and b. These indices have been found useful in vegetation applications and for reducing atmospheric or other sources of disturbances (Stagakis et al. 2010, Verrelst et al. 2015). All the single band and ND values were then assigned to plot-wise features as their mean values, weighted by the pixel proportions within the field plot extents (Table 1).

ALS data were acquired from the Finnish national laser scanning program, targeted at covering almost the whole country at a nominal density of 5 pulses/m^2 within a 6-year period starting from 2020 (NLS 2025). Scanning had been divided into designated production areas each year as indicated in Fig. 1, covering 2000–3000 km^2 each. For this study, all the available data from summertime acquisitions between 2020 and 2023 were selected, totaling 74 production areas scanned by six different contractors. Of these, 45 areas were acquired using Riegl VQ-780 i/ii/ii-S scanners, 21 using Riegl VQ-1560 i/ii/ii-S scanners, and eight using Leica CityMapper-2 L or TerrainMapper-2 scanners, respectively. Realized pulse densities ranged generally between 5.1 and 7.5 pulses/m^2 with one exceptional production area reaching 9.5 pulses/m^2.

Processing of the ALS data included extracting all the echoes within the plot limits and using them to calculate features indicated in Table 1, which had been selected based on earlier studies conducted in boreal forests (Gobakken and Næsset 2008, Packalen et al. 2019, de Lera et al. 2020). The primary focus was on structural characteristics, and intensity features were not included, mainly due to their inconsistencies across scanners. Calculated features included statistics related to the overall distribution of the normalized echo heights as well as their percentiles and proportions below the selected threshold values. Features were calculated for both the first and the last echoes (including single echoes in both categories) within $r = 9.00$ m, partially using only echoes at or above 2 m height to exclude low vegetation (denoted in Table 1 as '2 m+').

Experimental design

To address our objectives—namely, to evaluate how the number and locality of training plots affect prediction accuracy—we built separate prediction models for each NFI plot (Fig. 3a). That is, a set of spatially closest neighboring plots was used to build a model for the response variable, which was then applied to predict the respective value of the target plot. Finally, all the individual plots' prediction errors were combined for assessing overall accuracy. Each of the following cases and their combinations were tested in modeling:

- using features derived from only Sentinel-2 data (2D), only ALS data (3D), or both Sentinel-2 and ALS data (2D + 3D), which are later referred to as S2, ALS and S2 + ALS models;
- selecting 10, 25, 50, 100, or 500 neighboring plots to train the model; and
- applying exclusion radii of 0 km (i.e. using all the available neighbor plots), 10 km, 50 km, or 100 km around the target plot for selecting the neighbors.

Further, we assessed modeling across ALS production area boundaries as a special case of model transferability, where locality effects are mainly linked to the similarity of the auxiliary data and may change abruptly at dataset borders. For this purpose, all pairs of neighboring production areas that shared a long common border but had been scanned by a different contractor and/or scanner brand/model

Table 1 Summary of features extracted from the remote sensing materials.

Source	Description	Abbreviation	Feature count
Sentinel-2	Reflectance values of all the selected bands, and normalized difference indices of all their unique pairwise combinations.	*Ba* for reflectance of the band *a*, *NDa/b* for normalized difference of bands a and b	55
ALS data	Features derived from plot-wise points, calculated separately for the first and the last echoes: • general statistics of mean and maximum heights, coefficient of variation and standard deviation (2 m+); • percentile heights 10, 20, ..., 90 (2 m+); • proportions of echoes below the upper limit of the first four bins, when the height range between 2 m ... 95th percentile is divided into five bins of equal heights; and • proportions of echoes below fixed heights of 2 m, 5 m, 10 m, and 15 m.	*meanF|L*, *maxF|L cvF|L*, and *sdF|L* for general statistics; *percXF|L* for percentiles; *binXF|L* for bin proportions; and *fixXF|L* for fixed height proportions, where *F* and *L* indicate whether the feature derives from the first or the last echoes and *X* denotes the applied percentile, bin number, or fixed height.	46

ALS features described with ‘2 m+’ indicate that echoes below the height of 2 m were discarded when computing the metrics.

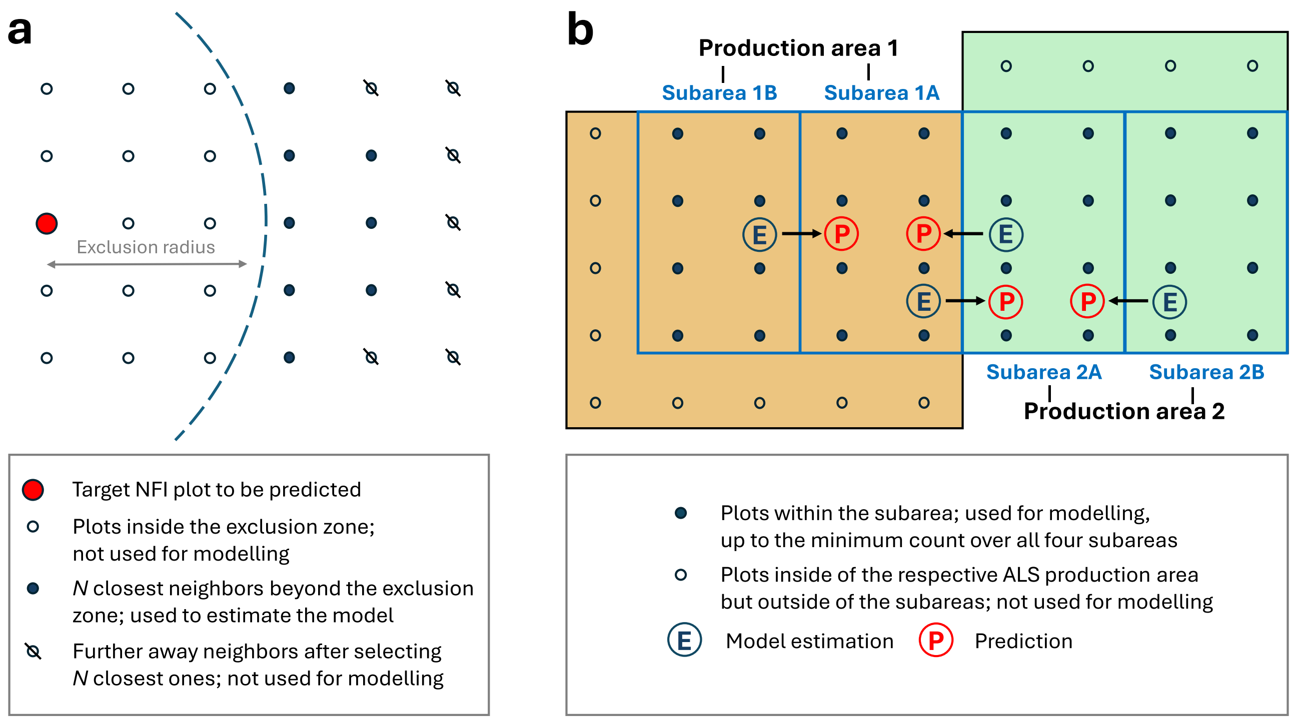

Figure 3 Schematic figures of assessing modeling performance using individual plots (a) and for comparing potential differences of two adjacent ALS production areas (b).

were identified ($n = 18$; see Fig. 1). For each pair of production areas (1 and 2; Fig. 3b), we delineated similarly sized subareas (A and B) on both sides of the shared border. Plots were selected from each subarea, and sample sizes were equalized by randomly removing plots from the larger set. Then, plots from each subarea were used to build a model, which was then applied to plots in the neighboring subarea (i.e. using the same model for all the target plots). Modeling was based solely on ALS features, using features derived from all, first, and last echoes separately to assess their influence on prediction performance. Finally, prediction errors from all target plots were pooled and compared between models built with ALS data from the same production area and those built with data from the neighboring area, i.e. with equal numbers of plots at approximately similar distances.

Modeling and accuracy assessment

Before modeling, plots with observed or recorded management activities during the investigated time span were excluded to avoid mismatches between field measurements and remote sensing data. As NFI data included observations on forest management as well as their approximate timing, plots which had recorded cutting on or after year 2020 (i.e. the earliest auxiliary data year) were first removed. To iden-

Table 2 Features selected for different models.

Auxiliary data	Inventory variable	Selected feature set
S2	Total volume	B5, ND2/3, ND5/12, ND3/4, B6, ND3/5, ND6/7, ND4/5, ND7/8A, ND7/8
	Coniferous volume	B11, B8, B2, ND3/4, ND2/8, ND6/7, ND2/5, ND4/11, ND2/11
	Deciduous volume	ND5/8, B4, ND5/12, ND3/4, ND6/8A, ND4/12, ND2/3, ND3/11, B8A, ND7/8A, ND8/8A
	Dominant height	B5, ND2/4, ND2/5, ND5/8, B8, ND3/5, ND5/12, ND7/8A, ND11/12, ND6/8A, ND7/8
ALS, all echoes	Total volume	fix10F, fix15L, cvF, sdF, bin4L, meanF, bin1F, perc10L
	Coniferous volume	fix15L, fix10F, perc90L, cvF, bin4L, perc10L
	Deciduous volume	fix5L, perc30F, bin4F, cvF, cvL, fix15F, perc10L
	Dominant height	fix10F, fix15L, maxL, perc10L, fix5L
ALS, only first echoes	Total volume	fix10F, cvF, meanF, sdF, fix2F
	Coniferous volume	fix10F, perc30F
	Deciduous volume	fix5F, perc70F, perc10F, fix15F, cvF
	Dominant height	fix10F, perc90F, cvF, bin2F
ALS, only last echoes	Total volume	fix10L, perc90L, perc10L
	Coniferous volume	fix15L, perc20L, perc80L, bin4L
	Deciduous volume	fix5L, perc10L, perc50L, cvL, fix15L, sdL
	Dominant height	maxL, fix15L, fix5L
S2 + ALS	Total volume	fix10F, fix15L, perc20L, meanF, fix5L, B12, ND6/7, B8, ND3/4, ND2/12, ND2/8, ND4/11
	Coniferous volume	fix15L, fix10F, B8, perc10L, meanF, B11, ND4/11, ND2/8, bin4L, ND5/11, ND8/8A, ND7/8
	Deciduous volume	ND3/8, fix5L, B5, ND6/7, ND4/11, B8, fix10F, perc30F, sdL, ND3/4, ND8/8A, ND7/8, ND6/8A, ND7/8A
	Dominant height	fix10F, fix15L, maxL, B4, bin2F, ND2/4

For feature abbreviations, see Table 1. ALS features derived from only first or only last echoes were used solely for comparing the adjacent ALS production areas.

tify harvesting between field measurements and later remote sensing acquisitions, we used forest use declarations that forest owners are required to submit prior to any planned or otherwise required (e.g. after insect damage) management action according to the Forest Act (Ministry of Agriculture and Forestry 2014). A declaration does not obligate the forest owner, and it has a validity period of three years, but in most cases the planned action is committed shortly after the declaration. If a declaration was found at a plot location and it dated earliest one year before the NFI measurement (given the potential lag for implementation), and if any remote sensing data had been acquired on or after the measurement year, the plot was excluded due to its high potential for changes. After these operations, the number of plots dropped from 30 098 to 24 437 and the number of distinct clusters from 4720 to 4660. Since the removed plots were from many development classes, mean values of the inventory variables did not significantly change, but the proportion of zero-volume plots dropped to 6.3% as several recent clearcuts were excluded. For the comparison of adjacent ALS production areas, the number of plots per subarea ranged from 69 to 258, with a median of 170.

Modeling of the inventory variables was based on the RF algorithm, implemented using R statistical software and *randomForest* package using its default parameters (Liaw and Wiener 2002, R Core Team 2022). RF is considered one of the best-performing prediction methods with a high degree of robustness (Esteban et al. 2019, Cosenza et al. 2020, Omoniyi and Sims 2024). To account for the stochastic nature of RF modeling, each model was trained five times and the predicted values were averaged to produce the final prediction. Altogether, separate RF models were trained for 24 437 individual plots, using combinations of four inventory variables, three feature sets, five training plot counts, four exclusion radii, and five repetitions of each estimation, totaling almost 30 million RF models to complete the assessments. Analyzing the adjacent 18 ALS production area pairs with four models between the subareas, three ALS-derived feature sets, and five repetitions of each model, further 1080 RF models were needed.

Because of the large number of RF models and remote sensing features, we applied rigorous variable selection to reduce the number of predictors for each combination of auxiliary data and inventory variable. First, we applied variable selection using random forest (VSURF) algorithm, implemented using the *VSURF* package (Genuer et al. 2015), which has been found to be an effective method for feature selection in remote sensing applications (Durante et al. 2019, Speiser et al. 2019, Fassnacht et al. 2021). And second, further filtering was based on Pearson's correlations between the value pairs of features. The maximum allowed correlation was set to 0.8. Starting with the most highly correlated feature pair, the feature with a lower VSURF importance score was removed, and this process was repeated until all the pairwise correlations were below the threshold. Although simple to apply, Pearson's correlation may overlook complementary nonlinear relationships among predictors. This can lead to dropping possibly useful predictors, which may compact the feature space and make it less expressive to detecting subtle information. However, we assumed this procedure to preserve the most essential features for comparing the models while keeping the computational load manageable. Selected auxiliary variables are presented in Table 2.

The performance and errors of the predicted values were assessed using relative root mean square error (RMSE; Eq. 2) and bias (Eq. 3):

$$RMSE\% = 100 \times \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}}{\bar{y}} \tag{2}$$

$$bias\% = 100 \times \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)}{\bar{y}} \tag{3}$$

where $\hat{y}_i$ is the predicted value, y_i is the measured value of NFI plot i over all the n respective plots, and $\bar{y}$ is the mean value of the targeted inventory variable over all the NFI measurements. For comparing the adjacent ALS production areas, significance of the absolute prediction error differences (i.e. not concerning any potential bias) between the two modeling strategies was also assessed using paired t-tests.

Additionally, prediction bias was spatially interpolated across the study area to examine whether over- or underestimates were spatially aggregated. This was done at a level of NFI clusters, i.e. by averaging coordinates and prediction bias over all the cluster-wise plots and excluding clusters with less than three plots. This approach aimed at indicating regional tendencies of bias rather than emphasizing any single outlier plots. Interpolation was based on inverse distance weighting (IDW) algorithm with 100 km maximum distance, 1 km cell size and 1 as the power parameter, producing a smooth surface without excessive influence from individual clusters.

Results

Errors of individual plots

RMSEs (Fig. 4) clearly depended on the auxiliary data and the number of plots used for model training. Errors were invariably the largest with 10 training plots but dropped rapidly when increasing the number of plots to 50–100. Increasing plot count from 100 to 500 benefited mainly species-specific variables (coniferous and deciduous volumes) and particularly S2 + ALS models. Regarding the differences in auxiliary data, errors of the structural variables (total volume and dominant height) depended mainly on whether ALS data were included in the model. With ALS features included, RMSE was nearly halved compared to using only Sentinel-2 data, regardless of whether Sentinel-2 features were also included. This difference remained similar across all training plot counts. For coniferous and deciduous volumes, however, S2 + ALS models performed significantly better than either data source alone. When predicting deciduous volumes, S2 model also slightly outperformed ALS model with the exception of the smallest plot counts. For the exclusion zones, differences in RMSE for total volumes and dominant heights were small, although using nearby plots consistently produced slightly smaller error. For coniferous and deciduous volumes, however, RMSE differed by about 5–10 percentage points between models with no exclusion zone and those with the largest zone (*i.e.* 100 km distance). Wider exclusion of the nearest plots affected ALS-inclusive models more than Sentinel-2, particularly when applying ALS data alone.

Regarding the prediction bias (Fig. 5), the S2 model predictions tended to overestimate the measured values, while the predictions of the ALS and S2 + ALS models were considerably less biased. Biases in total volume and dominant height predictions were especially small for the ALS and S2 + ALS models, particularly with plot counts of 100–500. The bias of coniferous volume predictions resembled the total volumes with the exception that the ALS model had a slight positive bias throughout the whole range of the plot counts. For the deciduous volume, however, the biases were generally the smallest with only 10 training plots. When increasing the number of plots, all the predictions were overestimates regardless of the used auxiliary data or the applied exclusion radius. Deciduous volume predictions were the most sensitive to excluding nearby plots, showing the largest bias difference between exclusion distances of 0 and 100 km. In general, however, widening the exclusion zone did not necessarily lead to an increased bias, and particularly with the S2 models the smallest biases were often not associated with zero exclusion (i.e. when including all the nearest neighbors).

Errors associated with ALS production areas

RMSE differences were evident between models trained with data from the same versus the adjacent ALS production area (Table 3). Training data from the same production area consistently outperformed the adjacent area, regardless of the inventory variable or the used echoes (first, last, or all), and all the differences were statistically significant ($P < .001$). First echoes showed the largest RMSE differences between production areas on average, except for coniferous and deciduous volumes, which had the largest differences with last and all echoes, respectively. Bias, however, was often smaller when models were trained with data from the adjacent production area. This was the case particularly for coniferous volumes, which tended to be slightly overestimated by models trained within the same area. For deciduous volumes, however, using the adjacent production area for training led to considerably larger bias if first or all echoes were used. In addition, the largest positive or negative errors were mostly located in northern Finland for models estimated using the same production area, whereas no clear geographic pattern was observed for models trained with data from the adjacent area.

Spatially interpolated bias

Interpolated maps indicating bias over the whole study area with all the applied combinations of the inventory variables, auxiliary features (S2, ALS, or S2 + ALS), numbers of training plots, and exclusion distances are presented in the supplementary materials. The spatial patterns in the interpolated bias maps generally differed from those of the measured inventory variables (cf. Fig. 2), although some similarities existed. For example, for coniferous and total volumes, using few training plots together with large exclusion distances generally resulted in underestimation in high average volumes, which shifted to overestimation toward lower-volume areas. While this likely reflects the lack of nearby training data, some areas remained more challenging than others irrespective of the measured values, particularly south-western and south-eastern parts of the country. Although these difficulties partly reflect spatial gaps in nearby training data and may be affected by larger proportions of deciduous trees, not all the regions near the map edges showed unusually high bias.

Part of the bias maps' tendencies resembled the calculated bias curves (Fig. 5), such as S2 models' tendency for overestimation. But similar bias at the level of the whole data could also derive from very different regional characteristics. For example, when the S2 models and 500 training plots were used for predicting coniferous volumes, bias according to Fig. 5 did not depend virtually at all on the applied exclusion zone. The interpolated bias maps, however, showed that expanding the exclusion zone also polarized regional differences in bias, creating areas of higher positive and negative biases that largely canceled each other out when deriving the overall bias. While moderate influence from this could be seen in the RMSEs (Fig. 4), the interpolated maps identified several regions where these effects were particularly strong.

The two extremes in the interpolated bias maps were dominant heights and deciduous trees' volumes. Regarding the dominant heights, the interpolated bias was low over the whole study area, almost regardless of the exclusion zone, particularly for ALS and S2 + ALS models. The predicted deciduous volumes, however, were predominantly divided into regions of substantial positive or negative bias, which appeared to depend more on the exclusion zone than the number of plots. The ALS models without Sentinel-2 predictors showed particularly strong bias polarization when exclusion distances of 50–100 km were used. Noteworthy also is that, when comparing, for example, predictions of deciduous trees' volumes with 500 training plots and a 50 km exclusion zone, the

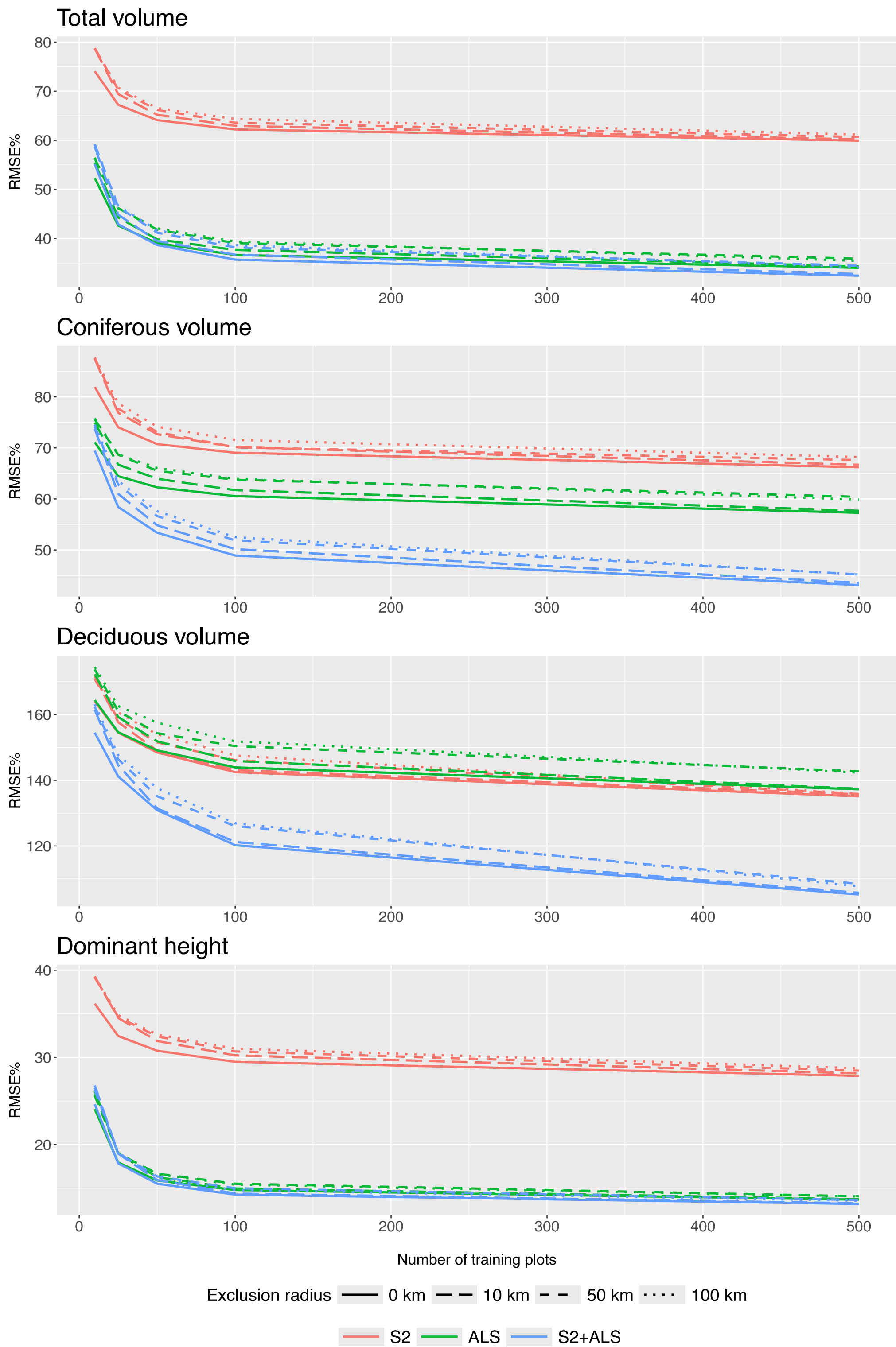

Figure 4 RMSEs of individual NFI plots (n = 24 437) as derived from all predictor combinations.

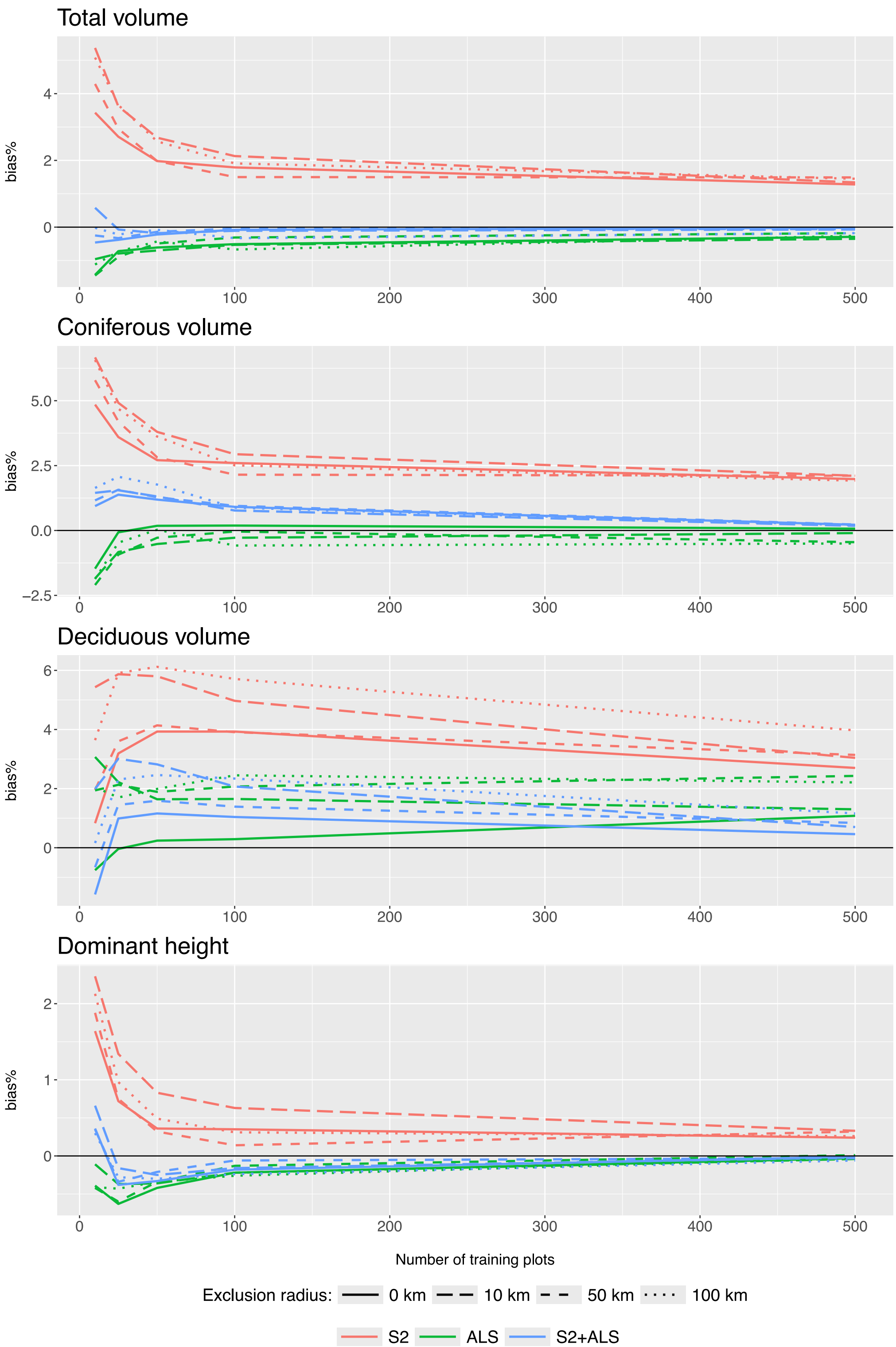

Figure 5 Biases of individual NFI plots (n = 24 437) as derived from all predictor combinations.

Table 3 Prediction errors of models estimated using first (F), last (L), or all (FL) ALS echoes derived from the same (s) or the adjacent (a) ALS production area.

	RMSE%						bias%					
	Fs	Fa	Ls	La	FLs	FLa	Fs	Fa	Ls	La	FLs	FLa
Total volume	40.82	45.32	41.22	43.12	38.57	40.20	0.04	−0.21	0.61	−0.47	0.68	0.47
Coniferous volume	74.95	75.93	67.30	69.43	65.55	67.31	1.35	−0.86	1.48	0.82	1.64	0.67
Deciduous volume	165.17	175.53	166.65	171.15	158.26	169.36	0.40	5.74	0.55	0.74	0.73	4.79
Dominant height	15.88	16.65	15.29	15.96	15.22	15.75	0.31	0.11	0.12	0.00	0.23	-0.08

overall bias (Fig. 5) was higher with the S2 than the ALS model. But according to the bias maps, the S2 model had much lower bias in many regions—although with tendency to positive values, which explains its higher overall bias.

Discussion

NFI field observations, combined with remote sensing data and model-based inference, provide an effective way to predict inventory variables across different spatial scales. The estimates, however, are not necessarily unbiased, and their accuracy depends on various factors as presented in our study. In NFIs, many field plots are usually available, but their size, shape, and spacing are normally designed to make fieldwork cost-efficient and to provide accurate estimates of core variables such as forest area or total volume of the growing stock at national or regional levels (Tomppo et al. 2008). Although NFI data can be used to predict any measured variable and produce estimates for small areas, one key modeling challenge is deciding how large an area should be used to select reference plots for modeling. These reference measurements should resemble the target location relatively well but still capture enough variation from the auxiliary data.

Based on our RMSE results, using 100 training plots appears to provide a good balance between model accuracy and keeping the training data geographically local. This applied particularly to the tested structural inventory variables—total volume and dominant height—which are known to have a relatively strong link to ALS information. For those inventory variables, 3D data generally outperforms 2D spectral features and allows accurate prediction across the full value range, once training data contains enough variation (Junttila et al. 2013, Tuominen et al. 2017). However, selecting another prediction method instead of RF could alter these results. For example, a smaller but well-selected training set might be enough for K-nearest neighbor algorithm with a small value of *k*, although this often increases noise. In contrast, parametric methods such as linear regression could perform most consistently when predictions require extrapolation beyond the training data.

Our results agree with earlier studies, such as Kallio et al. (2010), who estimated tree species and total volumes using ALS and aerial images, and Ruotsalainen et al. (2019), who predicted stand attributes with ALS data to support forest management decisions. In both studies modeling errors dropped rapidly until reaching 100–150 training plots, and gradual improvement continued up to 400–500 plots for models targeting diverse inventory variables. Stereńczak et al. (2018) used ALS data up to 900 training plots of different sizes to predict growing stock volumes and found that increasing the number of plots beyond 300–400 did not improve accuracy regardless of the plot size. On the other hand, Junttila et al. (2013) suggested that as few as forty well-selected training plots may be enough for modeling above-ground biomass with ALS data. However, this approach relied on a dense network of nearby plots that can capture local forest variation well but does not reflect typical NFI sampling. Regarding different modeling methods, Balazs et al. (2026) compared RF, k-nearest neighbor, and multi-layer perceptron for predicting above-ground biomasses and tree volumes using 100–3000 training plots. They found that particularly RF method produced relatively accurate and unbiased predictions even with the smallest number of training plots.

In our study, RMSEs of coniferous and deciduous volumes continued to decrease beyond 100 training plots, especially when 2D and 3D auxiliary data were used together. The contribution of spectral features seemed to relate particularly to the near-infrared band B8 (see Table 2), which improves the separation of tree species groups (Kukkonen et al. 2018, Shang et al. 2019). The added value provided by S2 was the highest for deciduous trees, which in the study area are mostly found as minor components in mixed stands with relatively low contribution to the total volume. For them, more training plots were likely needed to provide a representative set of training data that captured sufficient variation in factors such as stand age and forest health (McRoberts 2010). Additional challenges may arise from Sentinel-2 reflectance values at fixed-sized pixel sizes, which can miss fine-scale within-plot variation due to mixed signals, especially near to sharp boundaries such as managed stand edges (Tomppo et al. 2008).

The exclusion zones had only a small influence on RMSEs. Coniferous and deciduous volumes were affected more than total volumes or dominant heights, with the largest differences between the 10 km and 50 km exclusion radii. Differences may result from local similarities in forest types, management, soil characteristics or mesoclimatic conditions. They can also reflect differences in the consistency of the auxiliary data which potentially varies depending on the acquisition distance. However, the effects of excluding nearby plots could be much larger in ecosystems with greater variation in species composition, structure, or environmental conditions than in relatively homogeneous managed boreal forests.

Prediction bias, as shown in Fig. 5, was generally low and changed little with the number of training plots or the applied exclusion zone, especially when ALS data were included in the model. This is typical for models trained using remote sensing data, which often overestimate the lowest and underestimate the highest values but still provide accurate mean estimates and low overall bias (Avitabile et al. 2015, Kukkonen et al. 2021; Ståhl et al. 2024). However, models based only on Sentinel-2 data were more prone to positive bias regardless of the inventory variable, number of training plots, or the exclusion distance used. While reflectance saturation is often a major problem in forests with high growing stock volumes and typically causes underestimation rather than overestimation (Frampton et al. 2013, Wai et al. 2022), managed boreal forests contain many young, sparse stands and clearcuts, which sets the key challenge to predicting low volumes correctly (Astola et al. 2021, Kukkonen et al. 2021). Capturing this appears more difficult with Sentinel-2 reflectance values that are tied to fixed pixels than with ALS data, which provide direct 3D measurements of forest structure.

The challenge of predicting predominantly low values applied particularly to deciduous volumes in our target population. While approximately 10% of the plots had no measured trees of any species due to recent harvesting operations, the proportion of plots with no deciduous trees was almost 40%. In this case, the large RMSE values likely resulted from difficulties in separating the often small deciduous components from conifer-dominated stands, while the positive bias was mainly caused by the tendency of the RF algorithm to pull predictions toward the mean. These characteristics should preferably be addressed already in the sampling design if the applied framework allows it (Goodbody et al. 2023) or mitigated by strategies such as extreme gradient boosting, or models based on Tweedie error distribution with a log link function (Li et al. 2020, Omoniyi and Sims 2024, Moreno-Fernández et al. 2026).

In our modeling, however, it was somewhat surprising that the smallest numbers of training plots performed comparatively well to predict deciduous trees' volumes. We assume this was due to two reasons: first, both high and low volumes of deciduous trees are somewhat clustered in our study area, which is likely to enhance the prediction capabilities of models based on few nearby plots. And second, when only a few training plots and large exclusion zones were used, the training data may by chance have included enough plots without deciduous trees, helping the RF model predict true zero values more accurately for non-deciduous target plots.

As expected, ALS-based models trained using data from the same production area outperformed those derived from the adjacent area. However, the differences in RMSEs were smaller than anticipated. Additionally, we expected that differences between the areas would mainly result from laser beams' ability to penetrate the canopy—thus leading to larger error differences for last echoes than for first echoes—but we found the results to disagree with this assumption. This suggests that first returns may exhibit higher degree of variation between the scanning campaigns than last echoes, which can hinder model transferability.

Previous studies to discuss or compare our findings are few. Nilsson et al. (2017) investigated the effects of ALS production areas in Sweden when predicting NFI-derived inventory variables. They concluded that the scanner brand influenced the results, but an updated configuration of the same scanner might also lead to similar consequences. These differences may relate, for example, to pulse repetition rate, which can be adjusted in most scanners, but which simultaneously affects the energy of individual laser pulses. This may alter pulses' capabilities for penetrating the canopy even when other acquisition settings remain unchanged (Brede et al. 2022, Chasmer et al. 2006, Næsset 2005). In this context, deciduous trees with their sparser crowns would be likely to show greater variation than denser conifers, which agrees with our results. Laser beams' backscattering properties or patterns may also change due to external conditions, such as phenology or water content of the leaves and needles (Ceccato et al. 2001, Gaulton et al. 2013, Shcherbacheva et al. 2024). Tompalski et al. (2021), on the other hand, suggested that area-based ALS models can still be used across different ALS campaigns if the vertical distribution of points remains similar. However, this assumes that there are no systematic differences in the recorded signals. Although our data covers only limited observations from conifer-dominated boreal forests, the results suggest that, somewhat counterintuitively, using last echoes for modeling tree volumes and heights may improve feature consistency when combining ALS data from multiple campaigns.

The interpolated bias maps provide useful insights into how regional patterns are affected by the number and locality of the training plots, which are often obscured by summary statistics. Some over- or underestimations may be explained by spatial gaps in the training data or differences in ALS features between production areas. They may also be linked to anomalies in forest characteristics caused by factors such as soil fertility, land use history, or climatic differences e.g. due to large water bodies. Regardless of the cause, prediction bias increased particularly when only a small number of training plots were available, and the exclusion zone was 50 km or more.

The bias, however, did not increase evenly over the whole study area, but rather got polarized. That is, the area affected by large under- or overestimates increased, while simultaneously the overall bias (i.e. calculated over all the predicted target plots) remained largely unchanged. Similar polarization effect was noted by Balazs et al. (2026), who concluded that excluding nearby plots increased bias particularly in locations characterized by strong biomass gradients. In other words, limitations that prevent the use of nearby training plots can substantially affect local and regional estimates, even if global accuracy measures suggest little impact. This highlights the importance of representative training data in modeling, for which the number and proximity of training plots can serve as useful indicators.

Conclusions

Accurate model-based prediction of forest inventory variables requires relevant and representative auxiliary data. Our results support earlier findings that 3D point cloud features are more effective than 2D optical satellite data. However, combining both data sources may be beneficial when the target variable is influenced not only by forest structure but also by spectral characteristics. Based on our results, 100 field plots can be considered as the minimum training set size, but more plots may be needed when the target variable is rare in the population. Although our study does not suggest a specific maximum distance for selecting training plots, the results show that proximity matters even in relatively homogeneous conditions. While our findings suggest that plot representativeness starts to gradually decrease beyond 10 km, these effects may be stronger in more diverse and complex ecosystems, or near to substantial ecotones.

Our findings also underline that unrepresentative training data can easily increase prediction bias. However, under- and overestimates may not be evenly distributed across the target area but tend to cluster spatially, reflecting local patterns that the model failed to capture. This increases the risk of making incorrect interpretations on local or regional scales, and the effects may not be visible in overall performance statistics. These results highlight the importance of considering locality when assessing the representativeness of training data.

Acknowledgements

The authors wish to acknowledge CSC—IT Center for Science, Finland, for computational resources.

Author contributions

Timo P Pitkänen (Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Writing—original draft, Writing—review & editing), Annika Susanna Kangas (Conceptualization, Writing—review & editing), Mari Myllymäki (Conceptualization, Writing—review & editing), Petteri Packalen (Conceptualization, Funding acquisition, Methodology, Supervision, Writing—review & editing).

Supplementary material

Supplementary material is available at *Forestry: An International Journal Of Forest Research* online.

Conflicts of interest

None declared.

Funding

This work was supported by the Research Council of Finland through the projects 'Is climate smart forestry a utopia if the preferences of landowners are not considered? (UTOPIA)' (352782), 'Asynchronous datasets in large-area forest inventories by remote sensing (ASYNC)' (355267), and the Research Council of Finland's flagship 'Forest-Human-Machine Interplay—Building Resilience, Redefining Value Networks and Enabling Meaningful Experiences' (UNITE) (359174).

Data availability

The generated Sentinel-2 mosaic can be shared on reasonable request to the corresponding author. The applied ALS data requires licensing from the National Land Survey of Finland and cannot be directly shared. Data on the used NFI measurements or their locations are not publicly available and cannot be shared.

References

Astola H, Seitsonen L, Halme E. *et al.* Deep neural networks with transfer learning for forest variable estimation using Sentinel-2 imagery in boreal forest *Remote Sens*. 2021;**13**:2392. https://doi.org/10.3390/rs13122392.

Avitabile V, Herold M, Heuvelink GBM. *et al.* An integrated pan-tropical biomass map using multiple reference datasets *Glob Chang Biol*. 2015;**22**:1406–20. https://doi.org/10.1111/gcb.13139.

Balazs A, Miettinen J, Nilsson M. *et al.* Large-scale forest resource mapping with spatial gaps in the training data: comparison of different modeling approaches *Int J Appl Earth Obs Geoinf*. 2026;**146**:105104. https://doi.org/10.1016/j.jag.2026.105104.

Blickensdörfer L, Oehmichen K, Pflugmacher D. *et al.* National tree species mapping using Sentinel-1/2 time series and German National Forest Inventory data *Remote Sens Environ*. 2024;**304**:114069. https://doi.org/10.1016/j.rse.2024.114069.

Brede B, Bartholomeus HM, Barbier N. *et al.* Peering through the thicket: effects of UAV LiDAR scanner settings and flight planning on canopy volume discovery *Int J Appl Earth Obs Geoinf*. 2022;**114**:103056. https://doi.org/10.1016/j.jag.2022.103056.

Ceccato P, Flasse S, Tarantola S. *et al.* Detecting vegetation leaf water content using reflectance in the optical domain *Remote Sens Environ*. 2001;**77**:22–33. https://doi.org/10.1016/S0034-4257(01)00191-2.

Chasmer L, Hopkinson C, Smith B. *et al.* Examining the influence of changing laser pulse repetition frequencies on conifer forest canopy returns *Photogram Eng Remote Sens*. 2006;**72**:1359–67. https://doi.org/10.14358/PERS.72.12.1359.

Coops NC, Tompalski P, Goodbody TRH. *et al.* Framework for near real-time forest inventory using multi source remote sensing data *Forestry*. 2022;**96**:1–19. https://doi.org/10.1093/forestry/cpac015.

Cosenza DN, Korhonen L, Maltamo M. *et al.* Comparison of linear regression, k-nearest neighbour and random forest methods in airborne laser-scanning-based prediction of growing stock *Forestry*. 2020;**94**:311–23. https://doi.org/10.1093/forestry/cpaa034.

Durante P, Martin-Alcón S, Gil-Tena A. *et al.* Improving aboveground forest biomass maps: from high-resolution to national scale *Remote Sens*. 2019;**11**:795. https://doi.org/10.3390/rs11070795.

Esteban J, McRoberts RE, Fernández-Landa A. *et al.* Estimating Forest volume and biomass and their changes using random forests and remotely sensed data *Remote Sens*. 2019;**11**:1944. https://doi.org/10.3390/rs11161944.

Fassnacht FE, Hartig F, Latifi H. *et al.* Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass *Remote Sens Environ*. 2014;**154**:102–14. https://doi.org/10.1016/j.rse.2014.07.028.

Fassnacht FE, Poblete-Olivares J, Rivero L. *et al.* Using Sentinel-2 and canopy height models to derive a landscape-level biomass map covering multiple vegetation types *Int J Appl Earth Obs Geoinf*. 2021;**94**:102236. https://doi.org/10.1016/j.jag.2020.102236.

Fassnacht FE, White JC, Wulder MA. *et al.* Remote sensing in forestry: current challenges, considerations and directions *Forestry*. 2024;**97**:11–37. https://doi.org/10.1093/forestry/cpad024.

Frampton WJ, Dash J, Watmough G. *et al.* Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2013;**82**:83–92. https://doi.org/10.1016/j.isprsjprs.2013.04.007.

Gaulton R, Danson FM, Ramirez FA. *et al.* The potential of dual-wavelength laser scanning for estimating vegetation moisture content *Remote Sens Environ*. 2013;**132**:32–9. https://doi.org/10.1016/j.rse.2013.01.001.

Genuer R, Poggi J-M, Tuleau-Malot C. VSURF: an R package for variable selection using random forests *The R Journal*. 2015;**7**:19–33. https://doi.org/10.32614/RJ-2015-018.

Gobakken T, Næsset E. Assessing effects of laser point density, ground sampling intensity, and field sample plot size on biophysical stand properties derived from airborne laser scanner data *Can J For Res*. 2008;**38**:1095–109. https://doi.org/10.1139/X07-219.

Goodbody TRH, Coops NC, Queinnec M. *et al.* sgsR: a structurally guided sampling toolbox for LiDAR-based forest inventories *Forestry*. 2023;**96**:411–24. https://doi.org/10.1093/forestry/cpac055.

Goodbody TRH, Coops NC, White JC. Digital aerial photogrammetry for updating area-based forest inventories: a review of opportunities, challenges, and future directions *Cur For Rep*. 2019;**5**:55–75. https://doi.org/10.1007/s40725-019-00087-2.

Hau Z, McRoberts RE, Ståhl G. *et al.* How much can natural resource inventory benefit from finer resolution auxiliary data? *Remote Sens Environ*. 2018;**209**:31–40. https://doi.org/10.1016/j.rse.2018.02.039.

Hawryło P, Francini S, Chirici G. *et al.* The use of remotely sensed data and polish NFI plots for prediction of growing stock volume using different predictive methods *Remote Sens*. 2020;**12**:3331. https://doi.org/10.3390/rs12203331.

Henttonen HM, Kangas A. Optimal plot design in a multipurpose forest inventory *For Ecosyst*. 2015;**2**:31. https://doi.org/10.1186/s40663-015-0055-2.

Junttila V, Finley AO, Bradford JB. *et al.* Strategies for minimizing sample size for use in airborne LiDAR-based forest inventory *For Ecol Manage*. 2013;**292**:75–85. https://doi.org/10.1016/j.foreco.2012.12.019.

Kallio E, Maltamo M, Packalen P. Effect of sampling intensity on the accuracy of species-specific volume estimates derived with aerial data: A case study on five privately owned forest holdings. In: *Proceedings of the 10th International Conference on LiDAR Applications for Assessing Forest Ecosystems*. Freiburg, Germany, 2010.

Kangas A, Astrup R, Breidenbach J. *et al.* Remote sensing and forest inventories in Nordic countries – roadmap for the future *Scand J For Res*. 2018;**33**:397–412. https://doi.org/10.1080/02827581.2017.1416666.

Keränen J, Maltamo M, Packalen P. Effect of flying altitude, scanning angle and scanning mode on the accuracy of ALS based forest inventory *Int J Appl Earth Obs Geoinf*. 2016;**52**:349–60. https://doi.org/10.1016/j.jag.2016.07.005.

Knott JA, Liknes GC, Giebink CL. *et al.* Effects of outliers on remote sensing-assisted forest biomass estimation: a case study from the United States national forest inventory *Methods Ecol Evol*. 2023;**14**:1587–602. https://doi.org/10.1111/2041-210X.14084.

Korhonen K, Räty M, Haakana H. *et al.* Forests of Finland 2019–2023 and their development 1921–2023 *Silva Fennica*. 2024;**58**:24045. https://doi.org/10.14214/sf.24045.

Kukkonen M, Korhonen L, Maltamo M. *et al.* How much can airborne laser scanning based forest inventory by tree species benefit from auxiliary optical data? *Int J Appl Earth Obs Geoinf*. 2018;**72**:91–8. https://doi.org/10.1016/j.jag.2018.06.017.

Kukkonen M, Kotivuori E, Maltamo M. *et al.* Volumes by tree species can be predicted using photogrammetric UAS data, Sentinel-2 images and prior field measurements *Silva Fennica*. 2021;**55**:10360. https://doi.org/10.14214/sf.10360.

de Lera GA, Gobakken T, Ørka HO. *et al.* Reuse of field data in ALS-assisted forest inventory *Silva Fenn*. 2020;**54**:10272. https://doi.org/10.14214/sf.10272.

Li Y, Li M, Li C. *et al.* Forest aboveground biomass estimation using Landsat 8 and sentinel-1A data with machine learning algorithms *Sci Rep*. 2020;**10**:9952. https://doi.org/10.1038/s41598-020-67024-3.

Liaw A, Wiener M. Classification and regression by randomForest *R News*. 2002;**2/3**:18–22.

Lister AJ, Andersen H, Frescino T. *et al.* Use of remote sensing data to improve the efficiency of national forest inventories: a case study from the United States National Forest Inventory *Forests*. 2020;**11**:1364. https://doi.org/10.3390/f11121364.

Mäkisara K, Katila M, Peräsaari J. The multi-source National Forest Inventory of Finland – methods and results 2017 and 2019. In: *Natural Resources and Bioeconomy Studies 90/2022*. Helsinki: Natural Resources Institute Finland, 2022.

Maltamo M, Packalen P, Laukkanen L. *et al.* The transferability and cross-use of airborne laser scanning-based leaf-off and leaf-on biomass models *Eur J Remote Sens*. 2025;**58**:2542870. https://doi.org/10.1080/22797254.2025.2542870.

McRoberts RE. Probability- and model-based approaches to inference for proportion forest using satellite imagery as ancillary data *Remote Sens Environ*. 2010;**114**:1017–25. https://doi.org/10.1016/j.rse.2009.12.013.

McRoberts RE, Chen Q, Gormanson DD. *et al.* The shelf-life of airborne laser scanning data for enhancing forest inventory inferences *Remote Sens Environ*. 2018;**206**:254–9. https://doi.org/10.1016/j.rse.2017.12.017.

McRoberts RE, Tomppo EO, Næsset E. Advances and emerging issues in national forest inventories *Scand J For Res*. 2010;**25**:368–81. https://doi.org/10.1080/02827581.2010.496739.

Melin M, Korhonen L, Kukkonen M. *et al.* Assessing the performance of aerial image point cloud and spectral metrics in predicting boreal forest canopy cover *ISPRS J Photogramm Remote Sens*. 2017;**129**:77–85. https://doi.org/10.1016/j.isprsjprs.2017.04.018.

Miettinen J, Breidenbach J, Adame P. *et al.* Pan-European forest maps produced with a combination of earth observation data and national forest inventory plots *Data Brief*. 2025;**60**:111613. https://doi.org/10.1016/j.dib.2025.111613.

Ministry of Agriculture and Forestry. *Forest Act, 1093/1996; Amendments up to 567/2014 Included*, 2014 Available online: https://finlex.fi/api/media/statute-foreign-language-translation/216237/mainPdf/main.pdf (accessed 2 June 2025).

Moreno-Fernández D, Adame P, Breidenbach J. *et al.* Cross-country patterns of compositional and structural forest diversity across European regions *Eur J For Res*. 2026;**145**:74. https://doi.org/10.1007/s10342-026-01909-2.

Myllymäki M. *On Generalization of Sample Tree Characteristics in the Finnish National Forest Inventory Using Linear Mixed Effects Models*. Helsinki: Natural Resources Institute Finland, 2016.

Næsset E. Practical large-scale forest stand inventory using a small-footprint airborne scanning laser *Scand J For Res*. 2004;**19**:164–79. https://doi.org/10.1080/02827580310019257.

Næsset E. Assessing sensor effects and effects of leaf-off and leaf-on canopy conditions on biophysical stand properties derived from small-footprint airborne laser data. *Remote Sensing of Environment*. 2005;**98**:356–70. https://doi.org/10.1016/j.rse.2005.07.012.

Nilsson M, Nordkvist K, Janzén J. *et al.* A nationwide forest attribute map of Sweden predicted using airborne laser scanning data and field data from the National Forest Inventory *Remote Sens Environ*. 2017;**194**:447–54. https://doi.org/10.1016/j.rse.2016.10.022.

NLS. Laser scanning data 5p. In: *National Land Survey of Finland*, 2025 Available online: https://www.maanmittauslaitos.fi/en/maps-and-spatial-data/datasets-and-interfaces/product-descriptions/laser-scanning-data-5-p (accessed 2 June 2025).

Noordermeer L, Bollandsås OM, Ørka HO. *et al.* Comparing the accuracies of forest attributes predicted from airborne laser scanning and digital aerial photogrammetry in operational forest inventories. *Remote Sensing of Environment*. 2019;**226**:26–37. https://doi.org/10.1016/j.rse.2019.03.027.

Omoniyi TO, Sims A. Enhancing the precision of forest growing stock volume in the Estonian National Forest Inventory with different predictive techniques and remote sensing data *Remote Sens*. 2024;**16**:3794. https://doi.org/10.3390/rs16203794.

Packalen P, Strunk J, Packalen T. *et al.* Resolution dependence in an area-based approach to forest inventory with airborne laser scanning *Remote Sens Environ*. 2019;**224**:192–201. https://doi.org/10.1016/j.rse.2019.01.022.

Pasquarella VJ, Brown CF, Czerwinski W. *et al.* Comprehensive quality assessment of optical satellite imagery using weakly supervised video learning. In (éditeur), *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* 2023. https://doi.org/10.1109/cvprw59228.2023.00206.

Persson HJ, Ståhl G. *et al.* Characterizing Uncertainty in Forest Remote Sensing Studies. *Remote Sensing*. 2020;**12**:505. https://doi.org/10.3390/rs12030505.

R Core Team. R: A language and environment for statistical computing. In: *R Foundation for Statistical Computing*. Vienna, Austria, 2022. Available online: https://www.R-project.org/ (accessed 2 June 2025).

Räty M, Heikkinen J, Korhonen KT. *et al.* Effect of cluster configuration and auxiliary variables on the efficiency of local pivotal method for national forest inventory *Scand J For Res*. 2019;**34**:607–16. https://doi.org/10.1080/02827581.2019.1662938.

Ruotsalainen R, Pukkala T, Kangas A. *et al.* The effects of sample plot selection strategy and the number of sample plots on inoptimality losses in forest management planning based on airborne laser scanning data *Can J For Res*. 2019;**49**:1135–46. https://doi.org/10.1139/cjfr-2018-0345.

Saarela S, Schnell S, Tuominen S. *et al.* Effects of positional errors in model-assisted and model-based estimation of growing stock volume *Remote Sens Environ*. 2016;**172**:101–8. https://doi.org/10.1016/j.rse.2015.11.002.

Schumacher J, Hauglin M, Astrup R. *et al.* Mapping forest age using National Forest Inventory, airborne laser scanning, and Sentinel-2 data *For Ecosyst*. 2020;**7**:60. https://doi.org/10.1186/s40663-020-00274-9.

Shang C, Treitz P, Caspersen J. *et al.* Estimation of forest structural and compositional variables using ALS data and multi-seasonal satellite imagery *Int J Appl Earth Obs Geoinf*. 2019;**78**:360–71. https://doi.org/10.1016/j.jag.2018.10.002.

Shcherbacheva A, Campos MB, Wang Y. *et al.* A study of annual tree-wise LiDAR intensity patterns of boreal species observed using a hyper-temporal laser scanning time series *Remote Sens Environ*. 2024;**305**:114083. https://doi.org/10.1016/j.rse.2024.114083.

Speiser JL, Miller ME, Tooze J. *et al.* A comparison of random forest variable selection methods for classification prediction modeling *Expert Syst Appl*. 2019;**134**:93–101. https://doi.org/10.1016/j.eswa.2019.05.028.

Stagakis S, Markos N, Sykioti O. *et al.* Monitoring canopy biophysical and biochemical parameters in ecosystem scale using satellite hyperspectral imagery: an application on a Phlomis fruticosa Mediterranean ecosystem using multiangular CHRIS/PROBA observations *Remote Sens Environ*. 2010;**114**:977–94. https://doi.org/10.1016/j.rse.2009.12.006.

Ståhl G, Gobakken T, Saarela S. *et al.* Why ecosystem characteristics predicted from remotely sensed data are unbiased and biased at the same time – and how this affects applications. *Forest Ecosystems*. 2024;**11**:100164. https://doi.org/10.1016/j.fecs.2023.100164.

Ståhl G, Saarela S, Schnell S. *et al.* Use of models in large-area forest surveys: comparing model-assisted, model-based and hybrid estimation *For Ecos*. 2016;**3**:5. https://doi.org/10.1186/s40663-016-0064-9.

Stereńczak K, Lisańczuk M, Parkitna K. *et al.* The influence of number and size of sample plots on modelling growing stock volume based on airborne laser scanning *Drewno*. 2018;**61**:5–22. https://doi.org/10.12841/wood.1644-3985.D11.04.

Tian L, Wu X, Tao Y. *et al.* Review of remote sensing-based methods for forest aboveground biomass estimation: progress, challenges, and prospects *Forests*. 2023;**14**:1086. https://doi.org/10.3390/f14061086.

Toivonen J, Kangas A, Pitkänen TP. *et al.* Combining tree-boosting and mixed effects models improves the performance of remote-sensing based forest age predictions *Forestry*. 2026;**99**:cpaf083. https://doi.org/10.1093/forestry/cpaf083.

Tompalski P, Coops NC, White JC. *et al.* Estimating changes in forest attributes and enhancing growth projections: a review of existing approaches and future directions using airborne 3D point cloud data *Curr For Rep*. 2021;**7**:1–24. https://doi.org/10.1007/s40725-021-00135-w.

Tomppo E, Olsson H, Ståhl G. *et al.* Combining national forest inventory field plots and remote sensing data for forest databases *Remote Sens Environ*. 2008;**112**:1982–99. https://doi.org/10.1016/j.rse.2007.03.032.

Tuominen S, Pekkarinen A. Local radiometric correction of digital aerial photographs for multi source forest inventory *Remote Sens Environ*. 2004;**89**:72–82. https://doi.org/10.1016/j.rse.2003.10.005.

Tuominen S, Pitkänen T, Balazs A. *et al.* Improving Finnish Multi-Source National Forest Inventory by 3D aerial imaging *Silva Fenn*. 2017;**51**:7743. https://doi.org/10.14214/sf.7743.

Verrelst J, Rivera JB, Veroustraete F. *et al.* Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods – A comparison *ISPRS J Photogram Remote Sens*. 2015;**108**:260–72. https://doi.org/10.1016/j.isprsjprs.2015.04.013.

Wadoux AMJ-C, Brus DJ, Heuvelink GBM. Sampling design optimization for soil mapping with random forest *Geoderma*. 2019;**355**:113913. https://doi.org/10.1016/j.geoderma.2019.113913.

Wai P, Su H, Li M. Estimating aboveground biomass of two different Forest types in Myanmar from Sentinel-2 data with machine learning and geostatistical algorithms *Remote Sens*. 2022;**14**:2146. https://doi.org/10.3390/rs14092146.

Waser LT, Ginzler C, Rehush N. Wall-to-wall tree type mapping from countrywide airborne remote sensing surveys *Remote Sens*. 2017;**9**:766. https://doi.org/10.3390/rs9080766.

White J, Coops NC, Wulder MA. *et al.* Remote sensing technologies for enhancing forest inventories: a review *Can J Remote Sens*. 2016;**42**:619–41. https://doi.org/10.1080/07038992.2016.1207484.

Forestry: An International Journal of Forest Research, 2026, **99**, cpag071
https://doi.org/10.1093/forestry/cpag071
Original Article