Canadian Journal of Forest Research

OPEN ACCESS | Research Article

Mapping old-growth forests using airborne lidar data and satellite images: how do plot size and rarity affect accuracy?

Janne Räty[a], **Mari Myllymäki**[b], **Mikko Peltoniemi**[c], **Aleksi Lehtonen**[d], **and Petteri Packalen**[b]

[a]Natural Resources Institute Finland, Bioeconomy and Environment, Forest Inventory and Planning, Yliopistokatu 6, 80100 Joensuu, Finland; [b]Natural Resources Institute Finland, Bioeconomy and Environment, Forest Inventory and Planning, Latokartanonkaari 9, 00790 Helsinki, Finland; [c]Natural Resources Institute Finland, Bioeconomy and Environment, Ecosystems and Modelling, Latokartanonkaari 9, 00790 Helsinki, Finland; [d]Natural Resources Institute Finland, Bioeconomy and Environment, Carbon Cycle Management, Latokartanonkaari 9, 00790 Helsinki, Finland

Corresponding author: **Janne Räty** (email: janne.raty@luke.fi)

Abstract

Old-growth forests have become rare and fragmented in the boreal biome. Their precise locations are not currently known with sufficient accuracy to support forest conservation and forest management. We studied the mapping of old-growth forests using airborne lidar data and satellite images in the Finnish coniferous forests. We investigated how plot size and the rarity of old-growth forests affect the accuracy of old-growth forest detection. We employed a Gaussian process classifier to distinguish old-growth forests from managed forests. Our field data consisted of 176 old-growth and 1082 managed forest plots. The results showed that an increase in plot size from 20 m × 20 m to 60 m × 60 m improved the performance of the classifier, because the larger plots more likely contain spatial patterns of trees and crown features indicative of forest naturalness. The largest F1-score (0.74) was achieved by data augmentation that generates additional training plots located inside forest boundaries. We also showed that the detection accuracy of old-growth forests decreases as they become rarer in the population. This rarity effect is crucial to understand, because the occurrence of old-growth forests can vary regionally due to different land use pressures. The mapping procedure proposed here can assist in the planning of field-based inventories of old-growth forests.

Key words: airborne laser scanning, nature conservation, satellite imagery, conservation prioritization, data augmentation

Introduction

Terms such as old-growth forests, natural forests, and primary forests are often used interchangeably to refer to forests with a high-level of naturalness and distinct ecological characteristics. Here, we use the term old-growth forest for coniferous-dominated, boreal forests that represent mature and climax phase forests and exhibit a high level of naturalness, and fulfill the following criteria:

(1) Forest and tree characteristics are clearly distinguishable from those of forests used for timber production and reflect long-term ecological processes that have shaped tree species composition, tree size distribution, and spatial patterns of trees.
(2) A visible continuum of dead wood supply exists, especially in fertile soil types.
(3) No significant signs of human activities, that have disturbed natural processes associated with the trees or the ground, are found.

Old-growth forests are important for biodiversity, for the provision of ecosystem services and for the maintenance of habitats for several endangered species. Such forests develop slowly over centuries through natural processes, which emphasizes their role as robust carbon stores (Luyssaert et al. 2008). Moreover, old-growth forests are deemed an irreplaceable part of the boreal biome, which is also recognized in the Biodiversity Strategy 2030 proposed by the European Union (Hermoso et al. 2022). Notwithstanding the crucial role of old-growth forests, official definitions of such forests are manifold, especially with regards to determination of forests that are sufficiently natural to be deemed old-growth forest and so excluded from other forest uses. For example, European Commission defines old-growth forests as "*A forest stand or area consisting of native tree species that have developed, predominantly through natural processes, structures and dynamics normally associated with late-seral developmental phases in primary or undisturbed forests of the same type. Signs of former human activities may be visible, but they are gradually disappearing or too limited to significantly disturb natural processes*" (European Commission 2023). The European Commission's definition reflects the difficulty to determine what is an old-growth forest: the definition rather provides guidelines than quantitative

criteria. Both the difficulty of defining old-growth forests with straightforward criteria and the knowledge of old-growth forest locations hamper the planning of effective forest management and conservation. As such, accurate identification of old-growth forests in the forest landscape would serve to prioritize forest conservation in terms of biodiversity values and connectivity and achieve the target of 10% strictly protected land stated in the European Union's Biodiversity Strategy 2030.

Remote sensing techniques, especially airborne light detection and ranging (lidar), and remotely sensed (RS) data have revolutionized the mapping of forest resources in the 21st century (White et al. 2013; Maltamo and Packalen 2014; Næsset 2014). To date, several countries have established their own campaigns for the collection of airborne lidar data and aerial imagery that are used, among other purposes, for terrain modeling and forest inventories. Satellite missions, such as the Copernicus Sentinel-2, are another source of widely applied RS data in forest inventories. The main benefit of satellite-based imaging compared to airborne missions is the short re-visit interval, which enables continuous monitoring of forests in cloud-free conditions. However, current satellite technology does not allow the characterization of 3D structures with an adequate resolution, which highlights the need for airborne lidar data in the most demanding inventory tasks (Hirschmugl et al. 2023). For example, old-growth forests typically have characteristics (e.g., the spatial pattern of the trees and the height distribution of the trees) that are challenging to detect from medium-resolution satellite images. Instead, the benefit of multispectral satellite images over airborne lidar information is that the former can account for boreal tree species (Kukkonen et al. 2018), and can also help to distinguish between managed and old-growth forests.

Lidar-based forest inventories have evolved in two ways that support forest inventories: individual tree detection (ITD) and an area-based approach (ABA). In the former, the attributes of interest are modelled at the tree-level (Hyyppä and Inkinen 1999), whereas the latter relies on units larger than individual trees, typically field plots, in the modelling of forest attributes (Næsset 1997). Although the ABA principle has matured into an operational standard, the operational applicability of ITD has also improved as the availability of high-density airborne lidar data required for the delineation of individual trees has improved in several countries. However, ITD still exhibits some unsolved bottlenecks that limit its operational application. The major bottlenecks are related to the detection of suppressed trees and the application of tree allometry in the prediction of diameter by tree height (Hyyppä and Inkinen 1999). Although the ITD approach does not permit the detection of all trees in a multilayered forest as such, dominant trees can be detected reasonably well (Vauhkonen et al. 2012), and approaches are currently being developed to deal with suppressed trees (Kansanen et al. 2019; Kostensalo et al. 2023). The detected (and predicted suppressed) trees can be used to extract characteristics, such as the spatial pattern of the dominant trees (Packalen et al. 2013), canopy gaps (Silva et al. 2019), and crown properties (Næsset and Økland 2002). The potential of such airborne lidar-based tree features for the detection of old-growth forests has been reported in recent studies (Sverdrup-Thygeson et al. 2016; Fuhr et al. 2022; Jutras-Perreault et al. 2023).

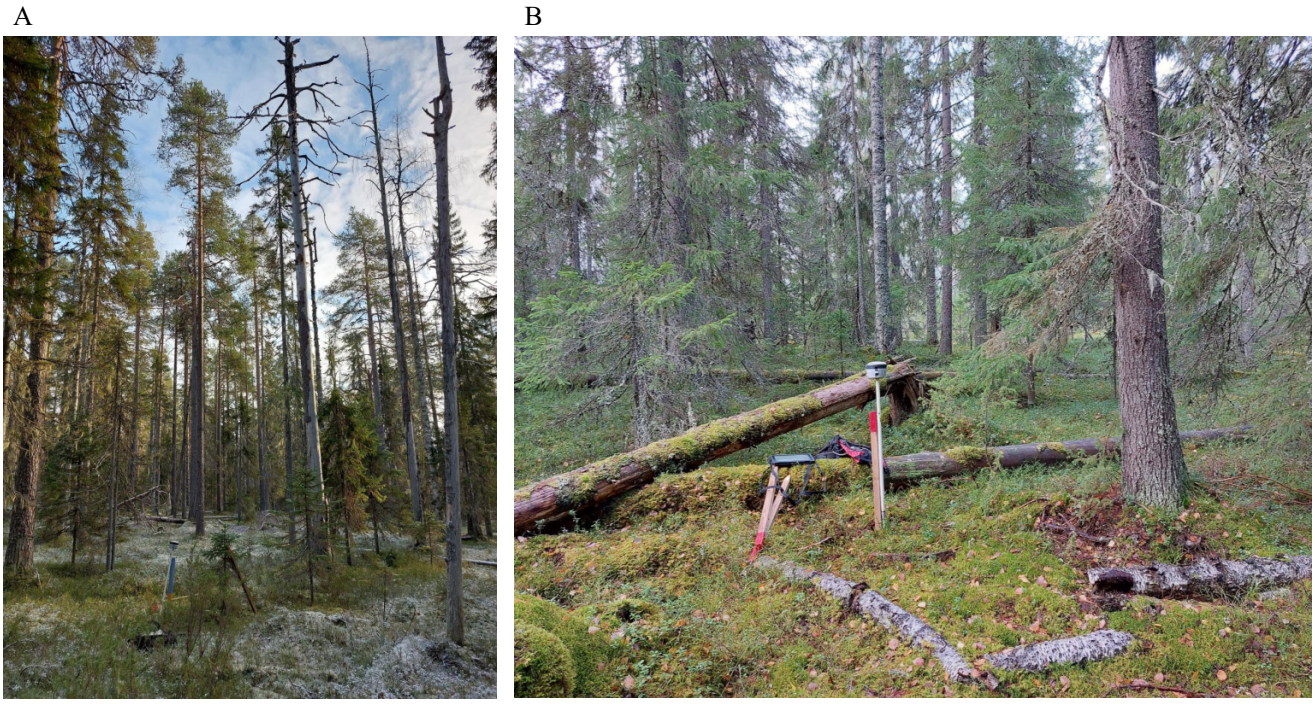
The characteristics of old-growth forests and managed forests have been compared by several authors. For example, Myllymäki et al. (2023) investigated the differences between forests close to their natural state and managed forests using forest attributes extracted from the Finnish National Forest Inventory (NFI) data. They found that the classification accuracy was rather poor when the age of the forest was unknown. Other studies have also investigated the detection of old-growth forests using RS data (Sverdrup-Thygeson et al. 2016; Ørka et al. 2022; Jutras-Perreault et al. 2023; Grabska-Szwagrzyk et al. 2024). However, mapping of old-growth forests remains a challenge although airborne lidar data have provided performance gains compared to the use of satellite data only. Previous studies have also pointed out that a key challenge associated with the mapping of old-growth forest is related to the availability of suitable field data (Jutras-Perreault et al. 2023; Myllymäki et al. 2023). A typical radius of forest inventory plots is approximately 9 m, which may be too small to clearly distinguish ecological characteristics, such as the spatial pattern of the trees, in both old-growth and managed forests. The effect of plot size on the accuracy associated with the classification of old-growth and managed forests using RS data are unclear in the light of the current knowledge.

Old-growth forests have become rare in the boreal region because forests have been used for a wide range of purposes for hundreds of years. In Finland, the rarity of old-growth forests is highly variable regionally, primarily because land use pressure is greater in the densely populated areas of southern Finland. The proportion of protected forest areas also provides good insights into this phenomena: the proportion of protected forest areas is 28% in the northernmost region of the country and 8% in the southernmost region (Niinistö et al. 2023). In general, the rarity of old-growth forests affects mapping accuracy in two ways. First, rarity exacerbates field data collection, because it is laborious to collect a representative sample of the phenomena of interest. Second, the accuracy of the resulting map is subject to the data in which the models are to be tested. Therefore, it is essential to perform accuracy assessments on the predictive models with respect to the realistic proportion of phenomena of interest in a population to avoid reporting unrealistic mapping accuracy values (Toivonen et al. 2024).

This study has the following objectives:

(1) To evaluate the classification accuracy of old-growth and managed forests using plot-level lidar, Sentinel-2 and tree-level lidar features with plot sizes of 20 m × 20 m and 60 m × 60 m.
(2) To improve the classification accuracy by incorporating RS features extracted from areas that surround the field plots. We also studied the aggregation of predictions by moving windows to account for the spatial autocorrelation of the surroundings of the field plots.

Fig. 1. Examples of (A) a pine-dominated old-growth forest and (B) a spruce-dominated old-growth forest in the study area. Photos: Tapio Ylimartimo.



(3) To quantify to what extent mapping accuracy depends on the occurrence of old-growth forests in the area of interest.

Materials and methods

Study area

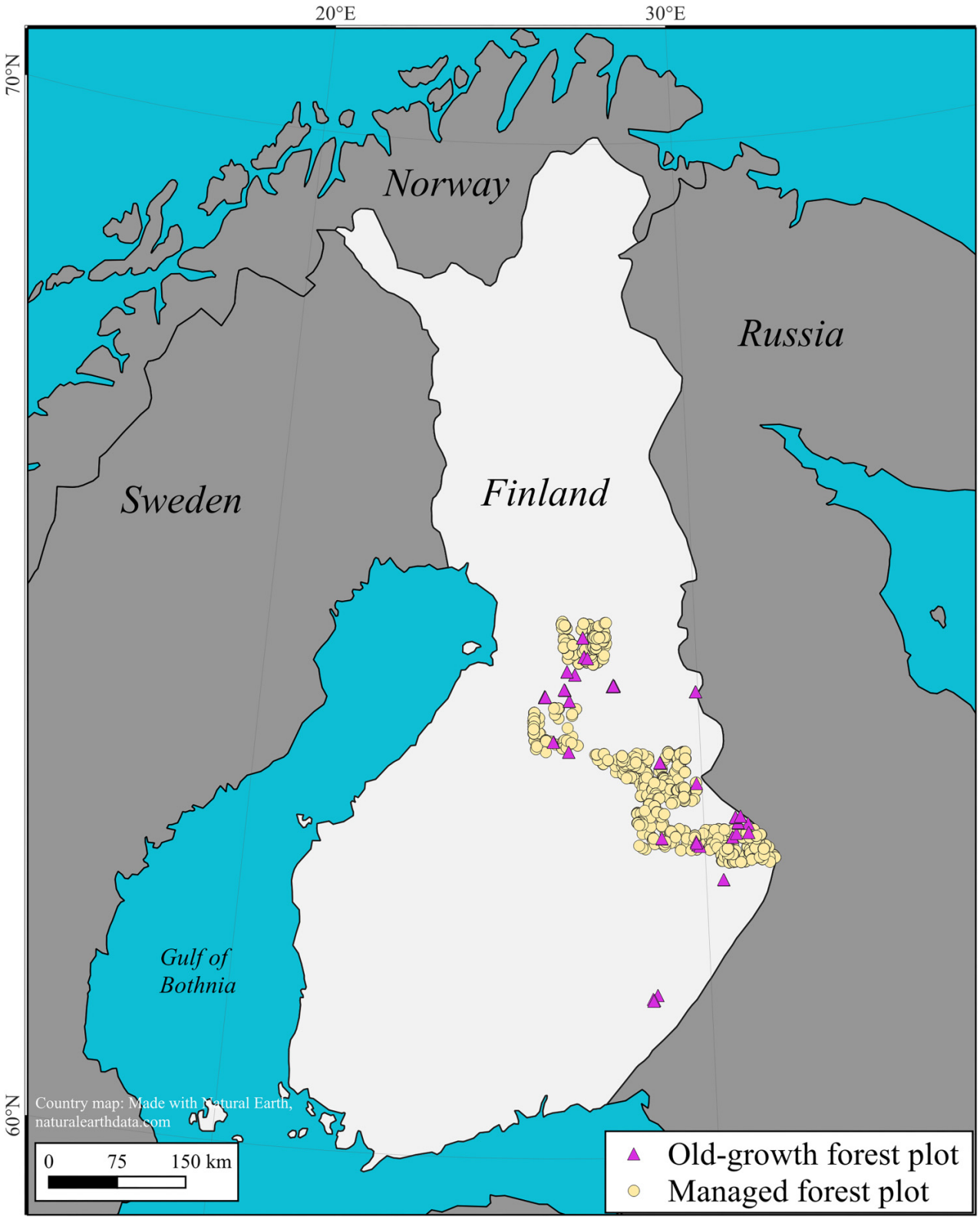
The study area covers parts of North Karelia, North Ostrobothnia, North Savo, South Savo, and Kainuu in Finland. The forests in the study area are dominated by coniferous species, primarily Scots pine (*Pinus sylvestris* [L.]) and Norway spruce (*Picea abies* [L.] Karst.). Broadleaved species, such as birch (*Betula* spp. [L.]) and aspen (*Populus tremula* [L.]), are common as mixtures throughout the area. Aside from anthropogenic disturbances, major disturbance agents include snow, fungi, wind, insects, and moose, which typically cause scattered disturbance events. Old-growth forests are characterized by multilayered canopy structures and a large variety in tree sizes, which is typical when the boreal forest ecosystem develops over the long term in the absence of a non-stand-replacing disturbance. Downed deadwood is abundant in the fertile, spruce-dominated old-growth forests, whereas dead standing trees are common in pine-dominated forest areas. Two representative examples of old-growth forests in the study area are shown in Fig. 1.

Old-growth forest plots

In the 1990s, a set of monitoring plots was established in old-growth forests by Metsäntutkimuslaitos (currently known as Natural Resources Institute Finland (Luke)). The plot locations were selected subjectively by a professional field crew. The field plots were primarily established in existing, state-owned forest conservation areas, i.e., national parks, strict nature reserves, and state-owned protection areas (Isomäki et al. 1998). The conservation status ensures that the forests represent at least their near-natural stage, maintain considerable natural values and have been shaped over centuries without significant silvicultural treatments. The criteria for silvicultural treatments were ensured upon the establishment of each plot by visually inspecting any signs of human intervention. The field plots were circular with a radius that varied from 7.57 to 28.21 m depending on the number of stems in the plot. The field protocol comprised a range of measurements, such as diameter at breast height, height, tree species and tree location relative to plot center. However, tree measurements were considered to be too outdated to be useful in this study. As the response variable in this study was a binary label, old-growth forest plots were labelled as 1.

As the field plot locations were not recorded accurately in the 1990s, it was not possible to establish a reliable link to the RS data. Therefore, accurate locations of the plot centers of 176 old-growth forest plots were determined in September and November 2023 (Fig. 2). Positioning was carried out with a differential global navigation satellite system (dGNSS) device Trimble R2 (Trimble 2022). The plot locations were post-corrected using reference positions provided by the closest Trimble virtual reference station. The locations of the old-growth forest plots were also visually inspected against the most recent available aerial orthophotos (National Land Survey of Finland 2024) to ensure that the field plots contained a concentric 60 m × 60 m cell completely inside

Fig. 2. Locations of the field plots in Finland. Figures were created using QGIS version 3.42.0 (http://qgis.osgeo.org) and assembled from the Natural Earth data (naturalearthdata.com).



the uniform forest areas. The uniform forest areas refer to forest units with similar structural characteristic, cf. forest stands of managed forests. The purpose of this inspection was to ensure that the RS data could be reliably extracted from a larger geographical extent than the original plot extent.

Table 1. The flight and sensor parameters associated with the acquisition campaigns of the airborne lidar (light detection and ranging) data.

Year	Month	Sensor	Flight altitude (m)	Half scan angle (degrees)	Beam divergence ($1/e^2$, mrad)	Pulse frequency (kHz)	Lateral overlap (%)	Pulse density (pts/m^2)
2019	July	Riegl VQ-1560i	1755	20	0.25	1400	22	6.4
2020	June	Riegl VQ-1560II	2100	20	0.25	1338	20	5.1
2021	August	Riegl VQ-1560II	2100	20	0.25	1338	20	5.1
	July	Riegl VQ-780i	1265	20	0.25	1000	21	6.5
	June	Riegl VQ-780II	1356	20	0.25	1200	24	7.0
2022	June	Riegl VQ-1560II-S	2100	20	0.25	1620	20	5.5
	August	Riegl VQ-1560II-S	2100	20	0.25	1620	20	5.5
	July	Riegl VQ-1560II-S	2100	20	0.25	1620	20	5.5
	June	Riegl VQ-1560II-S	2100	20	0.25	1620	20	5.5
2023	August	Riegl VQ-780II-S	1400	24	0.25	1150	21	6.2
	June	Riegl VQ-1560II-S	1800	20	0.25	1400	20	5.5
	June	Riegl VQ-1560II-S	1800	20	0.25	1400	20	5.5

Note: All data acquisitions were carried out under leaf-on conditions.

Managed forest plots

Managed forest plots measured between 2019 and 2023 were retrieved from an openly available database provided by the Finnish Forest Center (Suomen Metsäkeskus 2024). Here, the circular field plots were either 9 m radius or 12.62 m radius depending on the number of stems in the plot. The field plots were accurately positioned with a dGNSS device followed by a post-correction workflow. The field plots were collected for Finnish forest management inventories by inventory blocks that temporally aligned with the collection of the nationwide RS datasets. The locations of the managed forest plots were selected by the Finnish Forest Center. As the original purpose for the field data was to serve as training data in forest management inventories, the sample of field plots is representative of forests that are managed forests and serve as timber production forests, i.e., conservation areas are not of interest. The managed forest plots retrieved from the database were either young, middle-aged or mature forests, and seedling and sapling plots were not used in this study. The locations of the field plots were inspected against orthophotos (similarly to the old-growth forest plots) to select field plots that had a concentric 60 m × 60 m cell completely within the stand boundaries. The set of managed forest plots comprised 1082 plots (Fig. 2) and were marked with a label 0.

Remotely sensed data

In this study, we used airborne lidar data and Sentinel-2 images as RS data. The airborne lidar data belongs to the national data acquisition programs organized by the National Land Survey of Finland (National Land Survey of Finland 2023*a*). The nationwide collection of lidar data are targeted mainly for use in forest management inventories and terrain modelling (National Land Survey of Finland 2023*b*).

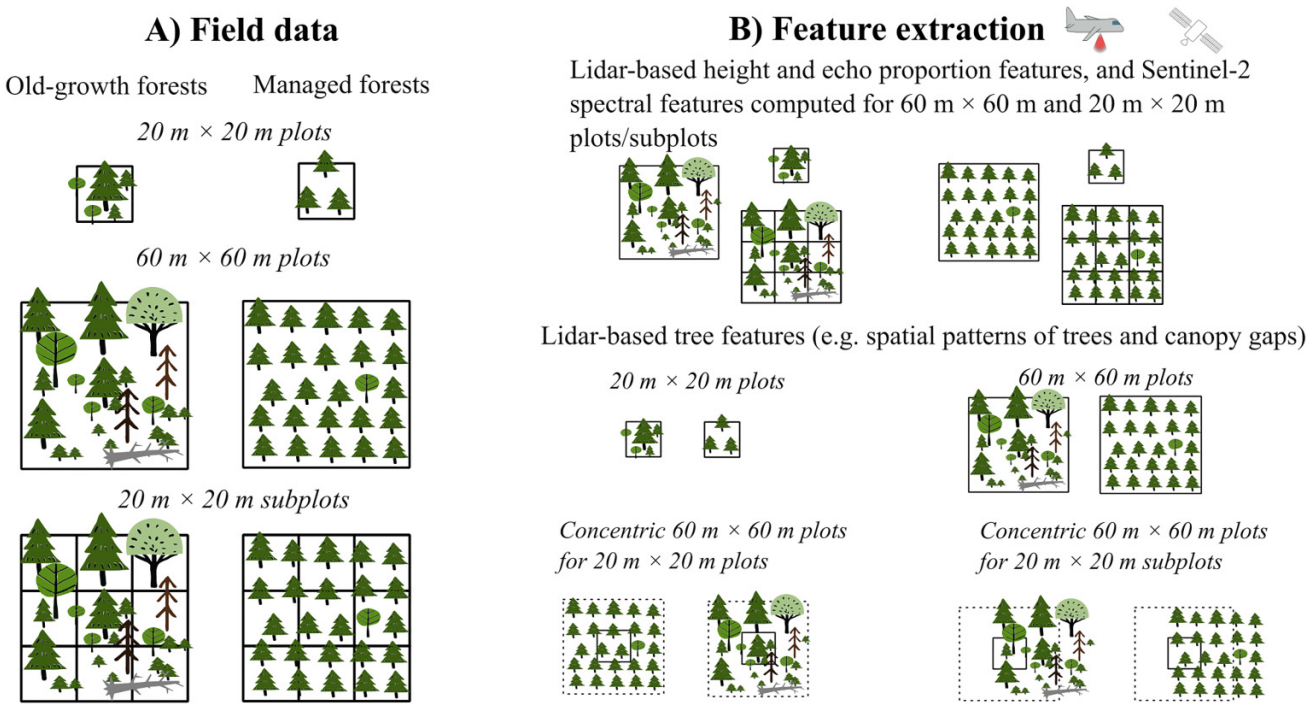
Airborne lidar data were collected between 2019 and 2023 using fixed-wing airplanes. Since the lidar data originated from the nationwide acquisition program, which consists of tens of flight campaigns, the flight parameters and sensor setups differed slightly among the flight campaigns. On average, the geographical area of a flight campaign was 290 000 ha. The study area comprised data from 12 different flight campaigns. Table 1 shows flight parameters associated with the flight campaigns. The echoes were categorized into *last-of-many*, *intermediate*, *first-of-many*, and *only* echoes. The echoes were classified as ground and other returns (Axelsson 2000), and the ground returns were used to generate a digital terrain model using a Delaunay triangulation. The lidar echoes were normalized to the ground-level by subtracting the digital terrain model from the original echo heights.

The Sentinel-2 Level-2 A summer images acquired between 15 June and 15 August 2021 were collected in the Google Earth Engine platform (https://code.earthengine.google.com/). A set of the Sentinel-2 images with an overall cloud coverage up to 15% were selected for further processing, and potential remaining clouds above the threshold of 30% were masked out using the S2_CLOUD_PROBABILITY layers (Skakun et al. 2022). The percentile-based automatic mosaicking method was applied for all the unmasked image areas to compute 40th percentile reflectance values by the Sentinel-2 bands (Pitkänen et al. 2024). The mosaic was resampled to the pixel size of 10 m using the nearest neighbor method. The Sentinel-2 bands of interest were B2 (blue), B3 (green), B4 (red), B5 (red edge), B6 (red edge), B7 (vegetation red edge), B8 (near infrared), B8A (near infrared), B11 (shortwave infrared), and B12 (shortwave infrared).

Preparation of plots and feature extraction

We set the initial plot size of old-growth and managed forest plots to 20 m × 20 m that corresponds well with the size of the original circular field plots. For each plot, we have manually checked against aerial images that a concentric 60 m × 60 m plot represents similar forests as the initial 20 m × 20 m plot. To account for the surroundings of the initial plots as a data augmentation in the model training phase, the 60 m × 60 m plots were split into 20 m × 20 m plots (hereafter

Fig. 3. An illustration of the feature extraction for the field plots.



referred to as subplots). This means that the surroundings of a 20 m × 20 m plot were used to augment the training data to generate additional eight subplots. The preparation of the plot data is illustrated in Fig. 3A.

The feature extraction is illustrated in Fig. 3B. We used the *first* (*first-of-many* and *only*) and *last* (*last-of-many* and *only*) echo categories of the airborne lidar data in the feature extraction. The lidar height and echo proportion features ($lidar_{plot}$) were extracted for the 20 m × 20 m plots and 60 m × 60 m plots without a height cutoff. The lidar data were also used to extract features that characterized individual trees and their spatial patterns. The lidar-based tree ($lidar_{tree}$) features involved statistics associated with the spatial pattern of the trees, crown features, gap features, and neighborhood characteristics derived from ITD or directly from a canopy height model (CHM). In the case of a 20 m × 20 m plot or subplot, $lidar_{tree}$ features were computed using either the actual plot size or the concentric 60 m × 60 m plot to account for the surroundings of the field plots. The extent of $lidar_{tree}$ features is denoted in subscript where necessary, for example $lidar_{tree60}$. Plot features were extracted from the Sentinel-2 mosaic by bands as area-weighted averages. The extracted features and their descriptions and abbreviations are listed in Table 2. A technical description of feature extraction using ITD and CHM follows in the next paragraphs.

We used ITD as a feature extraction tool for the $lidar_{tree}$ features. We followed the CHM-based ITD principle (Hyyppä and Inkinen 1999; Pitkänen et al. 2004) that consisted of four steps (1) generation of the CHM, (2) low-pass filtering of the CHM, (3) detection of local maxima, and (4) segmentation of the trees. The CHM was generated with a cell size of 0.5 m using the pitfree CHM algorithm proposed by Khosravipour et al. (2014). Low-pass filtering was implemented by Gaussian kernel with a sigma value of 0.5 pixels. Local maxima detection was carried out with a local maximum filter with a circular window which radius adapts to the height observed in the CHM cells (Popescu and Wynne 2004). This means that the radius of the filtering circle is larger in mature than young forests. For example, the radius of the window was 1.7 and 3.4 m for CHM heights of 5 and 30 m, respectively. Segmentation followed the region growing algorithm proposed by Dalponte and Coomes (2016). Unrealistically small segments with an area smaller than 1 m^2 were removed. The ITD routines were carried out using the implementations provided in the *lidR* library (Roussel et al. 2020; Roussel and Auty 2022) and the *terra* library (Hijmans 2024) in the R environment (R Core Team 2024).

The CHMs produced in the ITD workflow were used to analyze the gap patterns and forest characteristics. We detected canopy gaps and computed the basic statistics associated with them using the functions of the *ForestGapR* library (Silva et al. 2019). The height threshold for gap detection was set to 5 m, and the minimum and maximum gap sizes were set to 5 and 1000 m^2, respectively. We computed the number of canopy gaps per hectare, and the mean and standard deviation values of CHM heights in canopy gaps. In addition, we computed the mean area of canopy gaps, and the associated standard deviation of canopy gap areas.

Classification

Gaussian processes (GPs) offer a powerful and nonparametric toolbox for a variety of regression and classification problems (Rasmussen and Williams 2006). The core of a GP model is a kernel function that captures the flexible rela-

Table 2. Features extracted from remotely sensed data for field plots.

Feature category	Feature	Description
Lidar height and echo proportion features by *first* and *last* echoes ($lidar_{plot}$)	hmean, hmed, hstd, skew and kurt	Mean, median, standard deviation, skewness, and kurtosis of echo height distribution.
	h5, h10, …, h95	Percentiles (5%, 10%…, 95%) associated with the distribution of echo heights (m).
	d0.5, d2, d5, d10, d15	Proportion of echo heights below a fixed threshold of 0.5, 2, 5, 10, or 15 m (%).
	eprop	Proportion of first/last echoes to all echoes (%).
Lidar-based tree features ($lidar_{tree}$)	itd_n, itd_{mh}, itd_{wmh}, itd_{sdh}	Number of detected trees per hectare, mean height of detected ITD trees, mean height of detected ITD trees weighted by crown area, standard deviation of heights of detected ITD trees.
	itd_{gr15}	Proportion of detected trees taller than 15 m.
	itd_{ls5}	Proportion of detected trees shorter than 5 m.
	$crown_{area}$	Sum of tree segment areas (m^2).
	$crown_{round}$	Average roundness of tree segment computed as $crown\_perimeter^2/(4*\pi*crown_{area})$.
	ann_1, ann_2, ann_3, $sdann_1$, $sdann_2$, $sdann_3$	Average distance to the nearest, second nearest and third nearest neighbor tree, and associated standard deviations. Average distances were divided by the number of detected ITD trees in field plot.
	gap_n, gap_{mh}, gap_{hsd}, gap_{aream}, gap_{areasd}	Number of canopy gaps per hectare, mean CHM height in canopy gaps, standard deviation of CHM heights in canopy gaps, mean area of canopy gaps, standard deviation of canopy gap areas.
Sentinel-2 features by imaging bands (s_2)	wmean	Area-weighted mean of digital number values. Computed by Sentinel-2 bands B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12.

Note: Lidar—light detection and ranging, ITD—individual tree detection, CHM—canopy height model.

tionship between features, i.e., predictor variables, and class labels in the training data. Specifically, GP is based on the Bayesian formalism, which means that predictions can be inferred from the resulting posterior distribution. When using a GP classifier (GPC), an approximate inference for the latent values must be used because the likelihood function is non-Gaussian. The GPC is a so-called soft classifier that produces predictions between 0 and 1.

The Bernoulli likelihood function with a logit link was utilized in the GPC model to model the binary response variable (old-growth, managed). We assumed that observation y_i of each field plot i follows a Bernoulli distribution:

$$y_i \mid p_i \sim \text{Bernoulli}(p_i)$$

where the parameter p_i can be interpreted as a probability of old-growth forest. The Bernoulli parameter $p_i = p(z_i) \epsilon [0, 1]$ was assumed to depend on the normalized features z_i. The probability was modelled by a latent Gaussian process $f = f(z)$ through the logistic sigmoid transformation

$$p(z) = \frac{1}{1 + \exp(-f(z))}, \quad f(z) \sim GP(0, k(z, z'))$$

The Gaussian process f was a priori assumed to have zero mean and the Gaussian kernel function

$$k(z, z') = \exp\left(\frac{\|z - z'\|^2}{2l^2}\right)$$

where $\|z - z'\|$ refers to the Euclidian distance between two points in the normalized feature space and l is the correlation length hyperparameter of the Gaussian kernel that defines the range of the correlation in the GPC model.

The correlation length hyperparameter was tuned by a grid search algorithm as described in Varvia et al. (2021). The grid search evaluated each correlation length candidate using several classification thresholds. The classification threshold refers to the value in [0,1] that was used to convert the GPC predictions into the binary format. The grid search allowed correlation lengths 4.0, 4.5, 5.0, …,10.0, whereas the allowed classification thresholds of the grid search were 0.35, 0.40, …, 0.65. Correlation length values < 4 were excluded due to the risk of overfitting. In the grid search, we utilized a 50-fold cross validation (CV) with a spatial exclusion buffer of 500 m to eliminate spatial autocorrelation between the training and test folds. The computational burden of the hyperparameter optimization was alleviated by using the fully independent training and test conditional GP approximation with 200 inducing points (Quinonero-Candela and Rasmussen 2005). The correlation length that produced the largest F1-score was selected for final GP modeling (for the F1-score, please refer to the Performance assessment section below). The highly correlated predictor variables were omitted with a correlation cut-off value of 0.95 (Kuhn 2008). We fitted the GPC models using the GP implementation outlined in Piironen (2022). Bayesian inference for the latent Gaussian process was carried out using the Laplace approximation.

The GPC model predictions that were produced for the subplots (i.e., 20 m × 20 m subplots) enabled us to carry out a mean-aggregation using a moving window of 9 cells (low-pass filtering). This means that the final prediction for an initial 20 m × 20 m plot was computed as the average value of nine predictions. This was motivated by the fact that the neigh-

boring cells/plots are spatially autocorrelated. The aggregation method smoothed the map and potentially eliminated some mapping artefacts, such as individual grid cells in old-growth forest areas classified as managed. For this reason, we hypothesized that aggregation could improve the classification performance.

Performance assessment

The performance of the GPC model was evaluated by iteratively training and testing in a CV framework. As some of the field plots were closely located to each other, we applied the *dbscan* algorithm (Ester et al. 1996; Hahsler et al. 2019) with a neighborhood radius of 500 m to allocate the field plots into clusters. The use of the leave-cluster-out CV framework ensured that the models were not trained with plots located in the same forest stand as the target plots.

The number of managed forest plots was larger than the number of old-growth forest plots, which caused an unbalanced training dataset. The balancing of training data are important, because many binary classification approaches, including GPC, are sensitive to imbalanced datasets. The class imbalance may cause poor classification performance associated with the minority class of the training data (Yang et al. 2021; Toivonen et al. 2024). To overcome this issue, we employed the synthetic minority over-sampling technique (SMOTE) (Chawla et al. 2002) to balance the data by generating synthetic observations for the minority class, i.e., old-growth forest plots. The SMOTE algorithm searches for k-nearest neighbors for real observations in the feature space to generate synthetic observations based on the neighborhood. The generation of synthetic observations using SMOTE is controlled by the k-value and by the percentage of minority increment. A k value of 5 was used in this study. The percentage of minority increment was iteratively increased from 1 until the class balance was achieved. We applied the SMOTE implementation of the *performanceEstimation* library (Torgo 2014). The SMOTE algorithm was used in the training data in each iteration of the leave-cluster-out CV framework.

The predicted values ranged between 0 and 1 and were transformed into binary values by a classification threshold. Classification accuracy at each classification threshold was measured by an F1-score (eq 1) and its components:

$$\text{(1)} \qquad \text{F1} = \frac{(2{*}\text{recall}{*}\text{precion})}{\text{recall} + \text{precision}}$$

in which $\text{recall} = \frac{\text{TP}}{\text{TP+FN}}$, and $\text{precision} = \frac{\text{TP}}{\text{TP+FP}}$ with TP, FP, and FN denoting the true positives, false positives, and false negatives, respectively.

The performance of the GPC model is subject to the rarity of the old-growth forests in the target population, because the precision component of the F1-score depends on false positives. This was investigated in two ways: (1) The proportion of observed old-growth forests was reduced by the generation of synthetic managed forest plots in the confusion matrix. The frequencies of the confusion matrix associated with the observed managed forests (true negatives and false negatives) were multiplied by a factor that led to the desired proportion of old-growth forests in the confusion matrix. This confusion matrix multiplication assumes that the classification model performs similarly, i.e., the ratio of true and false negatives is fixed, regardless of the number of managed forests in the data; and (2) the CV framework was re-run and 25% of the training data were separated from the model training in each iteration and used for the generation of synthetic managed forest plots using the SMOTE algorithm. In the target data of each CV iteration, the pool of the synthetic managed forest plots and real managed forest plots was randomly subsampled to fulfill desired proportions of old-growth forests for the evaluation of classification accuracy. It should be noted that this approach presumably reduces model performance, as 25% of the training data are omitted in each CV iteration.

Results

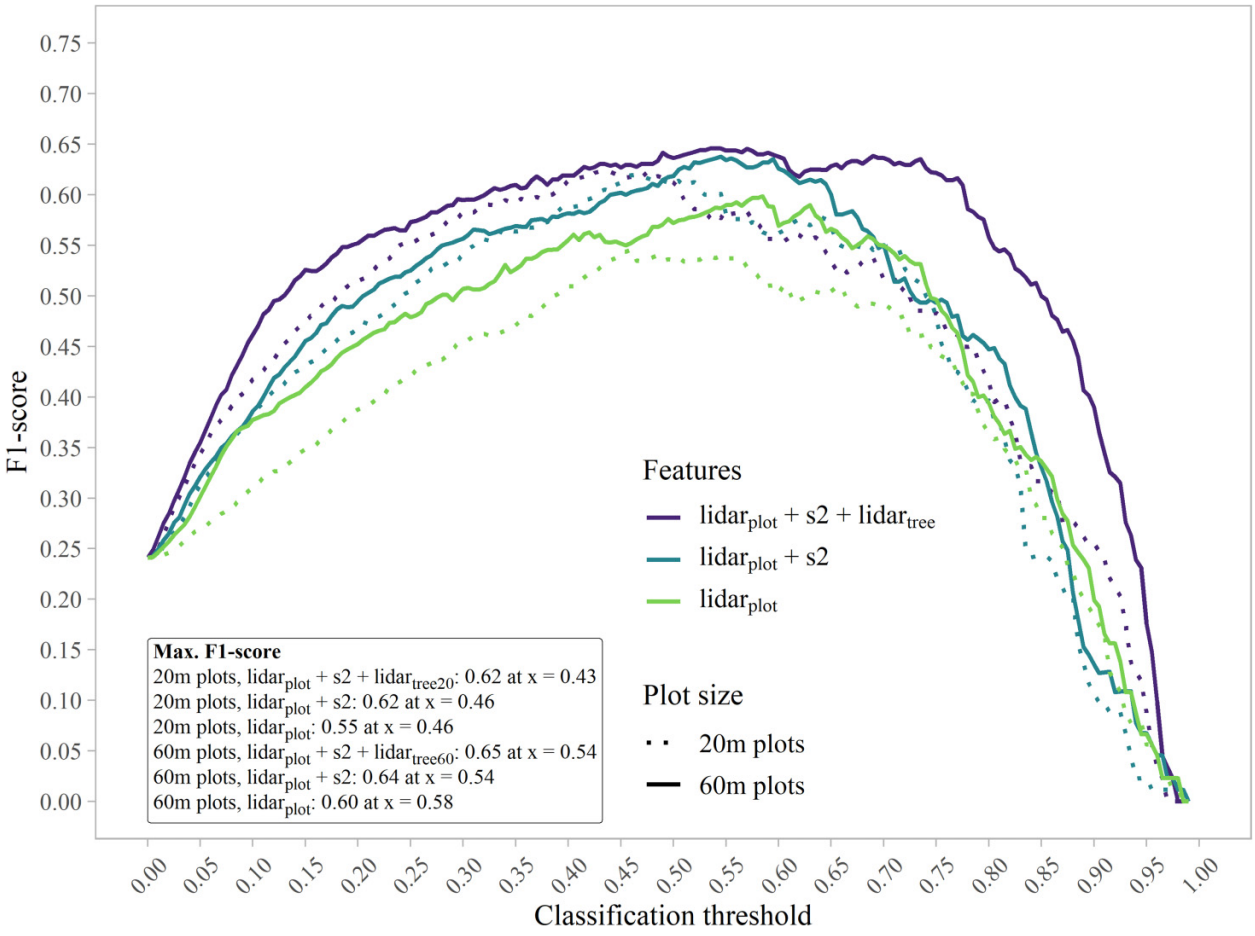
Feature category and plot size

The F1-score values, in terms of classification thresholds by feature categories and plot sizes, are shown in Fig. 3. The use of the 60 m × 60 m plots led to larger F1-scores than the use of the 20 m × 20 m plots. The maximum F1-score achieved with the 60 m × 60 m plots was 0.65, whereas the corresponding score for the 20 m × 20 m plots was 0.62. Further examination of the results revealed that the performance gain was mainly caused by the $\text{lidar}_{\text{tree}}$ features, which did not improve the classification performance in the 20 m × 20 m plots, but improved the performance in the 60 m × 60 m plots. It is worth noting that the shape of the F1-score curves associated with the 60 m × 60 m plots differed from those of the 20 m × 20 m plots (dashed vs. solid lines in Fig. 4). In particular, the use of the $\text{lidar}_{\text{tree}}$ features stretched the right tail of the curve to the right thereby improving the classification performance, compared to the models without the $\text{lidar}_{\text{tree}}$ features, with large classification threshold values. The precision and recall curves associated with all features and both field plot sizes are shown in Fig. 5. The precision and recall curves showed that the F1-scores were maximized at lower classification threshold values with the 20 m × 20 m plots, which means that the user must tolerate lower precision scores when maximizing the F1-score and using smaller field plots.

Surroundings of the 20 m × 20 m plots

A plot size of 60 m × 60 m is not feasible in most RS-based forest inventory applications. To overcome this, we investigated how to take advantage of the surroundings of the target plot/cell in the modeling by either computing the $\text{lidar}_{\text{tree}}$ features from a larger area than the field plot or by the generation of synthetic observations (subplots) in the training phase. The F1-scores by classification thresholds when the surroundings of the field plots were accounted for in the modeling are shown in Fig. 6. The computation of the $\text{lidar}_{\text{tree}}$ features from a concentric 60 m × 60 m area generally improved the classification performance, especially with the classification thresholds > 0.5 and < 0.9. The classification performance improved considerably when the training data of the model were augmented with the additional 20 m

Fig. 4. F1-scores associated with the classification of managed and old-growth forests using 20 m × 20 m and 60 m × 60 m plots with different sets of predictor variables. For feature abbreviations, please refer to Table 2.



× 20 m subplots. Recall that here all the 20 m × 20 m subplots were obtained by splitting the 60 m × 60 m plots into nine subplots. The largest F1-score value of 0.75 was achieved using the 20 m × 20 m subplots in model training and by further computing the final predictions as mean-aggregated values (low-pass filtering) with a moving-window of 9 cells.

Rarity effect

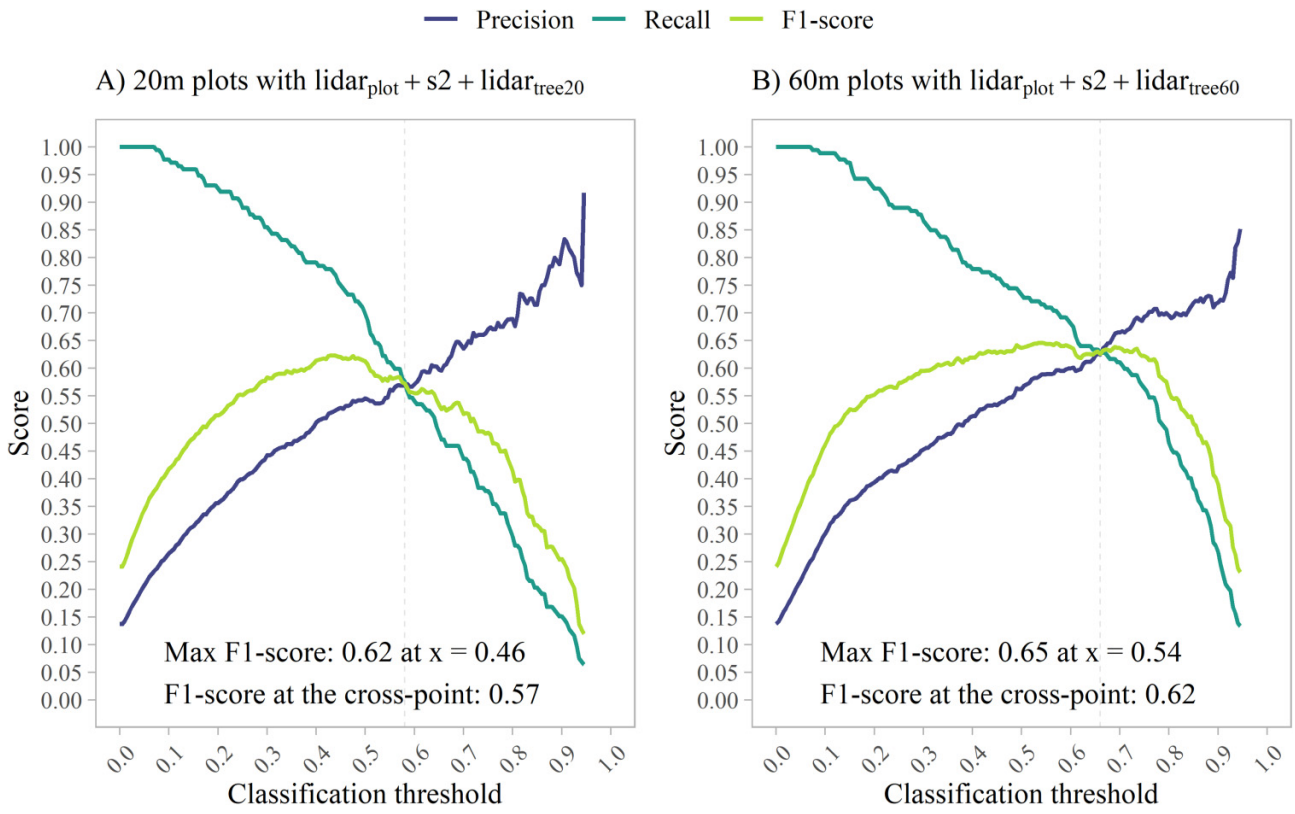
The effect of rarity of old-growth forest plots on the resulting F1-score, precision and recall values is shown in Fig. 7. The rarity effect was investigated in two ways (i.e., confusion matrix multiplication and SMOTE) both of which produced very similar results. The findings indicated that the F1-score decreased when the old-growth forests became rarer in the evaluation data used for the performance assessment. The decrease was due to the deterioration of the precision scores, whereas recall scores remained constant. We showed that the rarity effect on the classification performance can be investigated either by the generation of synthetic majority observations or by the use of a simple matrix multiplication that resulted in very similar curves as the SMOTE technique.

Discussion

We found that the lidar$_{tree}$ features were beneficial in the classification of managed and old-growth forests in the 60 m × 60 m plots but not in the 20 m × 20 m plots. Myllymäki et al. (2023) assessed structural naturalness of forests using attributes available in the Finnish NFI, such as the variation in tree sizes, the spatial pattern of tree locations, and stand ages. While they found forest naturalness was dependent on spatial tree patterns and the variability of tree sizes, the classification performance of managed and old-growth forests was generally low due to overlapping distributions. They also found that stand age better accounted for naturalness than the structural metrics derived from the forest attributes. They used circular field plots with a fixed radius of 9 m, and our findings suggest that greater classification accuracy could be achieved by using structural features computed for larger field plots.

The RS data could potentially provide an additional indicator of forest naturalness alongside traditional field measurements. As an example, 3D RS data enable the analysis of gap characteristics and crown characteristics of individual trees. Jutras-Perreault et al. (2023) computed lidar height features and horizontal features by characterizing the spatial patterns

Fig. 5. Precision, recall and F1-scores associated with the classification of managed and old-growth forests using all predictor variables with (A) the 20 m × 20 m field plots, and (B) the 60 m × 60 m field plots. For feature abbreviations, please refer to Table 2.

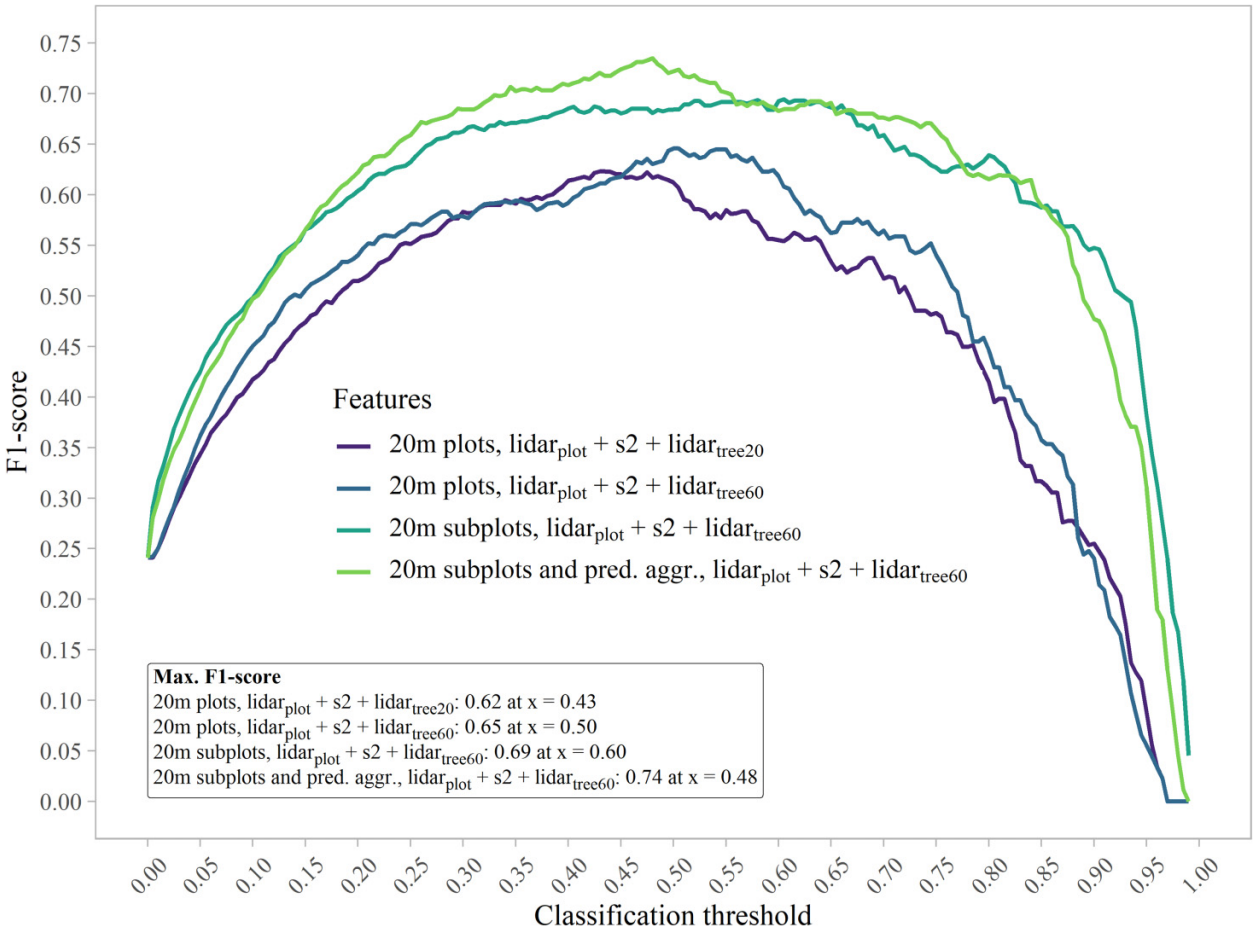


of the trees, as well as gap and crown characteristics to investigate the performance gains that result from the use of the horizontal features in the classification of managed and old-growth forests. However, they did not find improvements in the classification accuracy. It is worth noting that their field plots were similar-sized to the plots used by Myllymäki et al. (2023). Previously, Sverdrup-Thygeson et al. (2016) used circular field plots of 2000 m^2 and reported contrary results, which would indicate that horizontal features can benefit classification performance in managed and old-growth forests. This supports our finding that lidar$_{tree}$ features, i.e., horizontal metrics, can improve the classification accuracy when they are available from a sufficiently large plot size (here 60 m × 60 m). This finding could be linked to the fact that horizontal structure is typically more diverse in an old-growth than in a managed forest due to natural disturbance dynamics that cause canopy gaps in forest. While using small plots or grid cells, these disturbed areas can be misclassified as managed forests, because they are detached from their real-life context. Natural disturbance dynamics should be considered at a larger extent than, for example, a 20 m × 20 m cell. The small plot or cell size may also cause misinterpretations to the other direction: individual cells can be classified as old-growth forests although they are retention patches left outside the silvicultural treatments in managed forest.

The finding on the effect of plot size on classification accuracy motivated us to utilize RS data from a larger area than the extent of the actual field plot. We found that the use of lidar$_{tree}$ features with a larger area improved the classification performance. The computation of features based on an area larger than the actual field plot has some implications for old-growth forest map predictions. Forest edges have a major impact on cell-level predictions: the broader the feature extraction area, the greater the probability that the area overlaps several structurally different forests, such as old-growth forest and managed forest. In our case, the edge-effect was mitigated by computing lidar$_{plot}$ and Sentinel-2 features based on the actual sizes of the field plots. It is also noteworthy that the boreal old-growth forests of a high conservation value typically cover several hectares at a minimum, and thus the anomalies related to the edges of old-growth forest areas are likely minor.

We found that the use of the 20 m × 20 m subplots in model training resulted in the highest classification accuracy, which was further improved by the application of prediction aggregation (low-pass filter) on the map. The main reason for the improvement potentially lies in the fact that data augmentation made the dataset more comprehensive. For example, the original number of old-growth forest plots was 176, which increased nine-fold after surroundings-based data augmentation. The practical limitation of this kind of data augmentation is that it is necessary to ensure (using aerial imagery, for example) that all the subplots are representative of old-growth forests. In addition, it is important to note that the

Fig. 6. F1-scores associated with the classification of managed and old-growth forests using the 20 m × 20 m plot size and the surroundings of the field plots in the modeling procedure. For feature abbreviations, please refer to Table 2.

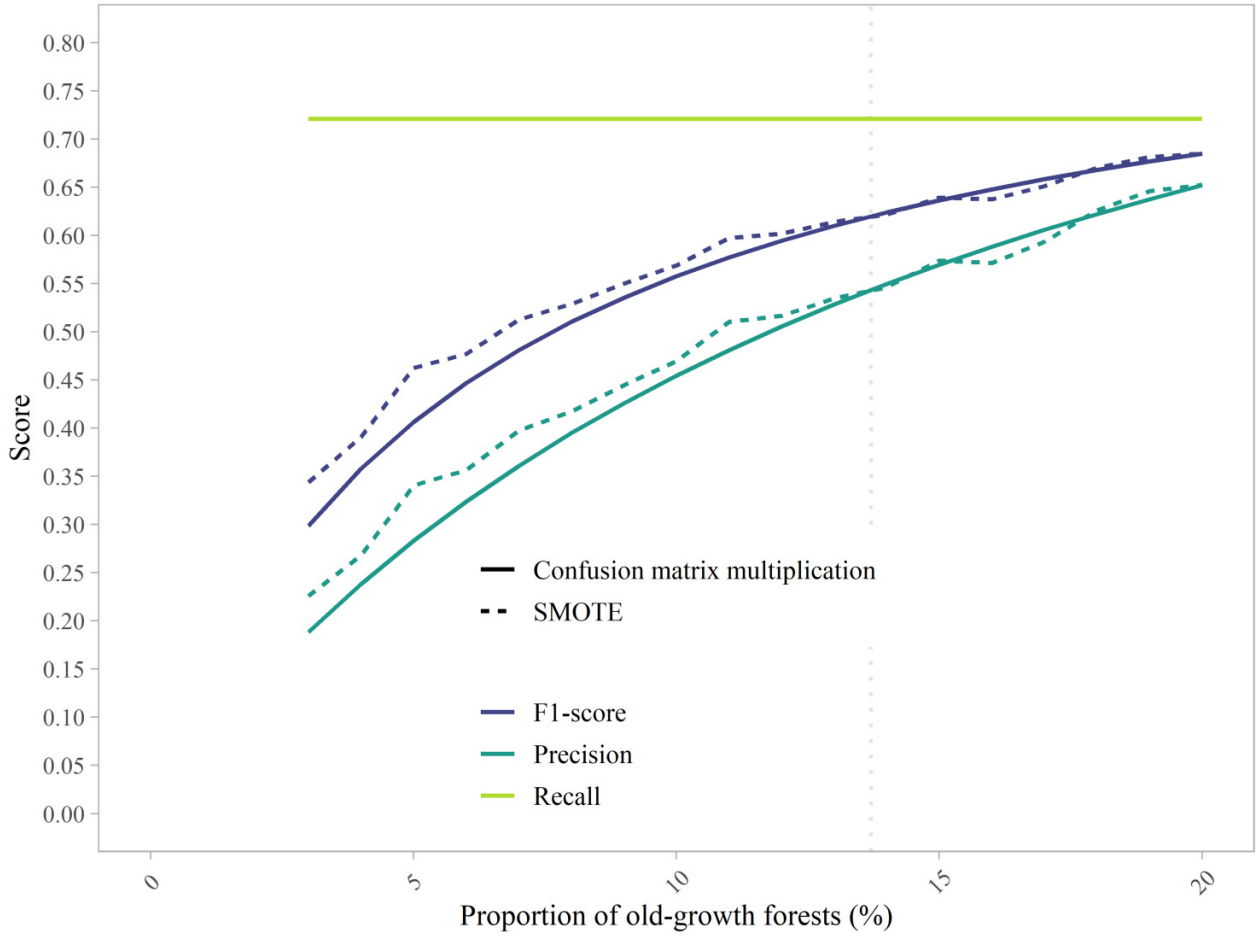


use of subplots is possible only in certain classification problems. For example, the classification of main tree species may be a problem because it is not possible to define a class label for subplots without a field visit. We also reported accuracy gains that resulted from the use of low-pass filtering. This would indicate that the individual cell-level predictions may contain anomalies due to the relatively small cell size, and that low-pass filtering shifts the focus of old-growth forest mapping to larger geographical areas rather than individual cells. For example, the classification model may predict individual cells in a large old-growth forest area as managed forests and vice versa, because the forest structure can be heterogenous both in old-growth forests and managed forest stands.

An important issue related to the mapping of rare phenomena using RS data is the rarity effect (Toivonen et al. 2024). Our findings demonstrated that large area maps of old-growth forests are likely prone to accuracy differences that are subject to the occurrence of the old-growth forests in the mapping region. For this reason, it is necessary to recognize the limitations of RS-based maps and, as far as possible, assess the mapping accuracy in a test dataset that corresponds to the real-life proportions of managed and old-growth forests in the area of interest. The rarity of old-growth forests is unknown but could be approximated, for example, based on the proportion of conservation areas in forest lands (information typically available in NFI).

We showed that 3D RS data sources have potential for mapping of old-growth forests, whereas some practical difficulties may lie in the collection of comprehensive field data with sufficiently large plot sizes from forests of interest that are already rare in the landscape. Here, we proposed that the issue associated with the plot size could be alleviated by the extraction of predictor variables or additional observations from the areas that surround the field plots. The collection of field data comprehensively representing a variety of old-growth forest types in the population is often challenging in practice. In our data, the locations of the old-growth forest plots were in existing conservation forests that have no significant signs of human activity. The existing conservation areas may not cover all types of old-growth forests, but the current conservation status is indicative of high conservation values, potentially including characteristics of old-growth forests. It is important to recognize that the recommended use scenario of an RS-based old-growth forest map is to support the planning of field visits for the inventories of old-growth forests. In

Fig. 7. Effect of the proportion of old-growth forests plots in the evaluation of the classification accuracy. The dotted, vertical line (in gray) indicates the proportion of old-growth forest plots in the field data.



addition, such maps could be useful as auxiliary information to assist sampling and/or inference in sample-based forest inventories, such as NFIs.

There are several possible data sources that could provide further gains in detection accuracy of old-growth forests. Future studies could harness RS data at several time points to ensure that mapping is more robust by identifying the timing of harvest operations in recent decades. Ørka et al. (2022) suggested that time series of satellite images can be used to map silvicultural pressures on forests over decades, which could benefit the detection of old-growth forests. Once high-density ALS data are available for larger areas, the detection of large downed dead wood on the forest floor (Heinaro et al. 2021) could also be used to support the identification of old-growth forest patches that contain large amounts of downed dead wood. Socio-economic features, such as ditches and roads, may also be worth considering in future studies as suggested by Lalechère et al. (2024).

Conclusions

We draw the following conclusions:

- The accuracy associated with the classification of old-growth and managed forests using remotely sensed data was higher with 60 m × 60 m than 20 m × 20 m plots.
- Data augmentation and tree features, such as spatial patterns and canopy gaps, accounting for a surrounding area of the 20 m × 20 m plot extent improved the classification accuracy.
- Accuracy of old-growth forest classification decreased as they became rarer in the forest landscape.

Acknowledgements

We acknowledge Timo Pitkänen for technical support in the processing of the satellite data, and Esko Oksa, Tapio Ylimartimo and Ville Lumberg for their efforts on the field work. This work was supported by Horizon Europe project FORWARDS "The ForestWard observatory to secure resilience of European forests" (nr. 101084481). The work was also supported by the Research Council of Finland through "The Finnish Flagship Programme for the Forest-Human-Machine Interplay—Building Resilience, Redefining Value Networks and Enabling Meaningful Experiences" (UNITE) [337655], the project "Asynchronous datasets in large-area forest inventories by remote sensing" [355267], and the European Union—NextGenerationEU in the Research Council of Finland's project [348154].

Article information

History dates

Received: 29 October 2024
Accepted: 16 April 2025
Accepted manuscript online: 6 May 2025
Version of record online: 9 June 2025

Copyright

Data availability

The data that support the findings of this study are available from the corresponding author upon a reasonable request.

Author information

Author ORCIDs

Janne Räty https://orcid.org/0000-0002-6578-8965
Aleksi Lehtonen https://orcid.org/0000-0003-1388-0388

Author contributions

Conceptualization: JR, MM, MP, AL, PP
Formal analysis: JR
Funding acquisition: MP, AL, PP
Methodology: JR, MM, PP
Project administration: MP, AL, PP
Supervision: MP, AL, PP
Writing – original draft: JR
Writing – review & editing: JR, MM, MP, AL, PP

Competing interests

The authors declare there are no competing interests.

References

Axelsson, P. 2000. DEM generation from laser scanner data using adaptive TIN models. Int. Arch. Photogramm. Remote Sens. **33**: 110–117.

Chawla, NV., Bowyer, KW., Hall, LO., and Kegelmeyer, WP. 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**: 321–357. doi:10.1613/jair.953.

Dalponte, M., and Coomes, DA. 2016. Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. Methods Ecol. Evol. **7**: 1236–1245. doi:10.1111/2041-210X.12575. PMID: 28008347

Ester, M., Kriegel, H-P., Sander, J., and Xu, X. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. *In* Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. AAAI Press, Portland, Oregon. pp. 226–231.

European Commission. 2023. Commission guidelines for defining, mapping, monitoring and strictly protecting EU primary and old-growth forests.

Fuhr, M., Lalechère, E., Monnet, J.-M., and Bergès, L. 2022. Detecting overmature forests with airborne laser scanning (ALS). Remote Sens. Ecol. Conserv. **8**: 731–743. doi:10.1002/rse2.274.

Grabska-Szwagrzyk, E., Jakiel, M., Keeton, W., Kozak, J., Kuemmerle, T., Onoszko, K., et al. 2024. Historical maps improve the identification of forests with potentially high conservation value. Conserv. Lett. e13043. Available from https://conbio.onlinelibrary.wiley.com/doi/abs/10.1111/conl.13043.

Hahsler, M., Piekenbrock, M., and Doran, D. 2019. dbscan: fast density-based clustering with R. J. Stat. Softw. **91**: 1–30. doi:10.18637/jss.v091.i01.

Heinaro, E., Tanhuanpää, T., Yrttimaa, T., Holopainen, M., and Vastaranta, M. 2021. Airborne laser scanning reveals large tree trunks on forest floor. Forest Ecol. Manage. **491**: 119225. doi:10.1016/j.foreco.2021.119225.

Hermoso, V., Carvalho, SB., Giakoumi, S., Goldsborough, D., Katsanevakis, S., Leontiou, S., et al. 2022. The EU Biodiversity Strategy for 2030: opportunities and challenges on the path towards biodiversity recovery. Environ. Sci. Policy **127**: 263–271.

Hijmans, RJ. 2024. terra: Spatial data analysis. Available from https://CRAN.R-project.org/package=terra.

Hirschmugl, M., Sobe, C., Di Filippo, A., Berger, V., Kirchmeir, H., and Vandekerkhove, K. 2023. Review on the possibilities of mapping old-growth temperate forests by remote sensing in Europe. Environ. Model. Assess. **28**: 761–785.

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

Isomäki, A., Niemistö, P., and Varmola, M. 1998. Luonnontilaisten metsien rakenne seurantakoealoilla. Metsäntutkimuslaitoksen tiedonantoja: 75–86.

Jutras-Perreault, M-C., Gobakken, T., Næsset, E., and Ørka, HO. 2023. Detecting the presence of natural forests using airborne laser scanning data. For. Ecosyst. **10**: 100146. doi:10.1016/j.fecs.2023.100146.

Kansanen, K., Vauhkonen, J., Lähivaara, T., Seppänen, A., Maltamo, M., and Mehtätalo, L. 2019. Estimating forest stand density and structure using bayesian individual tree detection, stochastic geometry, and distribution matching. ISPRS J. Photogramm. Remote Sens. **152**: 66–78. doi:10.1016/j.isprsjprs.2019.04.007.

Khosravipour, A., Skidmore, AK., Isenburg, M., Wang, T., and Hussin, YA. 2014. Generating pit-free canopy height models from airborne Lidar. Photogramm. Eng. Remote Sens. (PE&RS), **80**: 863–872.

Kostensalo, J., Mehtätalo, L., Tuominen, S., Packalen, P., and Myllymäki, M. 2023. Recreating structurally realistic tree maps with airborne laser scanning and ground measurements. Remote Sens. Environ. **298**: 113782. doi:10.1016/j.rse.2023.113782.

Kuhn, M. 2008. Building predictive models in R using the caret package. J. Stat. Softw. **28**: 1–26. doi:10.18637/jss.v028.i05.

Kukkonen, M., Korhonen, L., Maltamo, M., Suvanto, A., and Packalen, P. 2018. How much can airborne laser scanning based forest inventory by tree species benefit from auxiliary optical data? Int. J. Appl. Earth Obs. Geoinf. **72**: 91–98. doi:10.1016/j.jag.2018.06.017.

Lalechère, E., Monnet, J-M., Breen, J., and Fuhr, M. 2024. Assessing the potential of remote sensing-based models to predict old-growth forests on large spatiotemporal scales. J. Environ. Manage. **351**: 119865. doi:10.1016/j.jenvman.2023.119865.

Luyssaert, S., Schulze, E-D., Börner, A., Knohl, A., Hessenmöller, D., Law, BE., et al. 2008. Old-growth forests as global carbon sinks. Nature, **455**: 213–215. doi:10.1038/nature07276.

Maltamo, M., and Packalen, P. 2014. Species-specific management inventory in Finland. *In* Forestry applications of airborne laser scanning: concepts and case studies. *Edited by* M Maltamo, E Næsset and J Vauhkonen. Springer, Netherlands, Dordrecht. pp. 241–252. doi:10.1007/978-94-017-8663-8_12 [accessed 22 January 2019].

Myllymäki, M., Tuominen, S., Kuronen, M., Packalen, P., and Kangas, A. 2023. The relationship between forest structure and naturalness in the Finnish National Forest Inventory. Forest. Int. J. For. Res. **97**: 339–348. doi:10.1093/forestry/cpad053.

Næsset, E. 1997. Estimating timber volume of forest stands using airborne laser scanner data. Remote Sens. Environ. **61**: 246–253. doi:10.1016/S0034-4257(97)00041-2.

Næsset, E. 2014. Area-based inventory in Norway—from innovation to an operational reality. *Edited by* M Maltamo, E Næsset and J Vauhkonen. Springer, Netherlands, Dordrecht. 215–240. doi:10.1007/978-94-017-8663-8_11 [accessed 22 January 2019].

Næsset, E., and Økland, T. 2002. Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. Remote Sens. Environ. **79**: 105–115. doi:10.1016/S0034-4257(01)00243-7.

National Land Survey of Finland. 2023a. Laser scanning data 0,5p. Available from https://www.maanmittauslaitos.fi/en/maps-and-spatial-data/datasets-and-interfaces/product-descriptions/laser-scanning-data-05-p.

National Land Survey of Finland. 2023b. Laser scanning data 5p terms of use. Available from https://www.maanmittauslaitos.fi/en/laser-skanning-data/terms-of-use.

National Land Survey of Finland. 2024. NLS orthophotos. Available from https://www.maanmittauslaitos.fi/en/maps-and-spatial-data/datasets-and-interfaces/product-descriptions/orthophotos.

Niinistö, T., Peltola, A., Räty, M., Sauvula-Seppälä, T., Torvelainen, J., Uotila, E., and Vaahtera, E., 2023. Finnish statistical yearbook of forestry. Luonnonvarakeskus (Luke), Helsinki.

Ørka, HO., Jutras-Perreault, M-C., Næsset, E., and Gobakken, T. 2022. A framework for a forest ecological base map—an example from Norway. Ecol. Indic. **136**: 108636. doi:10.1016/j.ecolind.2022.108636.

Packalen, P., Vauhkonen, J., Kallio, E., Peuhkurinen, J., Pitkänen, J., Pippuri, I., et al. 2013. Predicting the spatial pattern of trees by airborne laser scanning. Int. J. Remote Sens. **34**: 5154–5165. doi:10.1080/01431161.2013.787501.

Piironen, J. 2022. gplite: General Purpose Gaussian Process Modelling. Available from https://CRAN.R-project.org/package=gplite.

Pitkänen, J., Maltamo, M., Hyyppä, J., and Yu, X. 2004. Adaptive methods for individual tree detection on airborne laser based., Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **36**: 187–191.

Pitkänen, TP., Balazs, A., and Tuominen, S. 2024. Automatized Sentinel-2 mosaicking for large area forest mapping. Int. J. Appl. Earth Obs. Geoinf. **127**: 103659. doi:10.1016/j.jag.2024.103659.

Popescu, SC., and Wynne, RH. 2004. Seeing the trees in the forest: using lidar and multispectral data fusion with local filtering and variable window size for estimating tree height. Photogramm. Eng. Remote Sens. **70**: 589–604. doi:10.14358/PERS.70.5.589.

Quinonero-Candela, J., and Rasmussen, CE. 2005. A unifying view of sparse approximate Gaussian process regression. J. Mach. Learn. Res. **6**: 1939–1959.

R Core Team. 2024. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Available from https://www.R-project.org/.

Rasmussen, C., and Williams, K. 2006. Gaussian processes for machine learning. The MIT Press.

Roussel, J-R., and Auty, D. 2022. Airborne LiDAR data manipulation and visualization for forestry applications. Available from https://cran.r-project.org/package=lidR.

Roussel, J-R., Auty, D., Coops, NC., Tompalski, P., Goodbody, TR.H., Meador, AS., et al. 2020. lidR: an R package for analysis of Airborne Laser Scanning (ALS) data. Remote Sens. Environ. **251**: 112061. doi:10.1016/j.rse.2020.112061.

Silva, CA., Valbuena, R., Pinagé, ER., Mohan, M., de Almeida, DR.A., North Broadbent, E., et al. 2019. ForestGapR: an R package for forest gap analysis from canopy height models. Methods Ecol. Evol. **10**: 1347–1356. doi:10.1111/2041-210X.13211.

Skakun, S., et al. 2022. Cloud mask intercomparison eXercise (CMIX): an evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2. Remote Sens. Environ. **274**: 112990. doi:10.1016/j.rse.2022.112990.

Suomen Metsäkeskus. 2024. Inventointikoealat. Available from https://www.metsakeskus.fi/sites/default/files/document/tietotuotekuvaus-inventointikoealat.pdf.

Sverdrup-Thygeson, A., Ørka, HO., Gobakken, T., and Næsset, E. 2016. Can airborne laser scanning assist in mapping and monitoring natural forests? Forest Ecol. Manag. **369**: 116–125. doi:10.1016/j.foreco.2016.03.035.

Toivonen, J., Kangas, A., Maltamo, M., Kukkonen, M., and Packalen, P. 2024. Mapping large European aspen (*Populus tremula* L.) in Finland using airborne lidar and image data. Can. J. For. Res. **54**: 762–773. doi:10.1139/cjfr-2023-0271.

Torgo, L. 2014. An infra-structure for performance estimation and experimental comparison of predictive models in R. arXi. doi:10.48550/arXiv.1412.0436.

Trimble. 2022. Timble R2 GNSS receiver. Available from https://trl.trimble.com/docushare/dsweb/Get/Document-787661/022516-200M_R2%20GNSS%20Receiver_DS_USL_0822_LR.pdf [accessed 2 September 2024].

Varvia, P., Räty, J., Korhonen, L., and Packalen, P. 2021. Gaussian process regression for airborne laser scanning based forest inventory: validation and parameter selection. *In* Proceedings of the SilviLaser Conference 2021, Vienna, Austria, 28–30 September 2021.

Vauhkonen, J., Ene, L., Gupta, S., Heinzel, J., Holmgren, J., Pitkänen, J., et al. 2012. Comparative testing of single-tree detection algorithms under different types of forest. Forestry, **85**: 27–40. doi:10.1093/forestry/cpr051.

White, JC., Wulder, MA., Varhola, A., Vastaranta, M., Coops, NC., Cook, BD., et al. 2013. A best practices guide for generating forest inventory attributes from airborne laser scanning data using an area-based approach. For. Chron. **89**: 722–723. doi:10.5558/tfc2013-132.

Yang, L., Heiselman, C., Quirk, JG., and Djurić, PM. 2021. Class-imbalanced classifiers using ensembles of Gaussian processes and gaussian process latent variable models. *In* ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 3775–3779. doi:10.1109/ICASSP39728.2021.9414754.