OXFORD

Institute of Chartered Foresters

Forestry: An International Journal of Forest Research, 2025, 1–14

https://doi.org/10.1093/forestry/cpaf083

Original Article

Combining tree-boosting and mixed effects models improves the performance of remote-sensing based forest age predictions

Janne Toivonen [1,2,*], Annika Kangas [1], Timo P. Pitkänen [3], Mari Myllymäki [3], Matti Maltamo [2], Mikko Kukkonen [1], Petteri Packalen [3]

[1]Bioeconomy and environment, Natural Resources Institute Finland (Luke), Yliopistokatu 6, P.O. Box 111, FI-80100 Joensuu, North Karelia, Finland
[2]School of Forest Sciences, University of Eastern Finland (UEF), Yliopistokatu 7, P.O. Box 111, FI-80101 Joensuu, North Karelia, Finland
[3]Bioeconomy and environment, Natural Resources Institute Finland (Luke), Latokartanonkaari 9, P.O. Box 2, FI-00791 Helsinki, Uusimaa, Finland
*Corresponding author. Natural Resources Institute Finland (Luke), Yliopistokatu 6, P.O. Box 111, FI-80100 Joensuu, Finland. E-mail: janne.toivonen@luke.fi

Abstract

Forest age is an important attribute from the perspectives of forest management and biodiversity, but prediction with remote sensing data is difficult. In this study, we evaluated the performance of airborne laser scanning (ALS) and Sentinel-2 data in plot-level forest age predictions. In addition, we accounted for site conditions in the modelling by utilizing categorical variables, such as the main site type of the forest plot. Categorical variables were derived from field data but were available for the entire landscape. We compared two prediction methods: linear mixed effects (LME) modelling and tree-boosted mixed effects (GPBoost) modelling. Our field data contained 870 National Forest Inventory plots in northern Finland with ages that ranged from 0 to 300 years. Some plots contained seedling and retention trees (hereafter hold-over tree) left from the previous generation, which make the age prediction of these plots a major challenge. To mitigate this, we tested an alternative strategy that included a prior classification step to identify hold-over plots. Overall, three age modelling strategies were tested (1) without categorical variables, (2) with categorical variables, and (3) with both categorical variables and hold-over plot classification. Our results showed that GPBoost was superior to LME in each tested scenario, and the addition of categorical variables led to a clear decrease in the prediction error. When categorical variables were added as random components, the relative root mean square error (RMSE) values for LME improved from 46.2% to 40.2% and from 41.7% to 38.5% for GPBoost. The best performing modelling strategy included hold-over plot classification before age modelling, which yielded RMSE values of ∼38.2% and 36.3% for LME and GPBoost, respectively. Compared to earlier research, our approach exhibited better prediction performance for older forests (≥150 years old) which in turn enables better identification of old-growth forests.

Keywords: airborne laser scanning (ALS); biodiversity; forest structure; linear mixed effects model; GPBoost

Introduction

Forest age is an important attribute for sustainable forest management and planning (Franklin *et al.* 2018, Rogers *et al.* 2022) and for the assessment of forest biodiversity and habitats (Costanza *et al.* 1998, Pan *et al.* 2011). It is also relevant for various forest attribute models such as site index (Eerikäinen *et al.* 2002, Racine *et al.* 2014) and regional or national level reporting (Gillis *et al.* 2005). Moreover, forest age is reported to be a more important indicator of forest naturalness than tree size distribution parameters and species composition (Myllymäki *et al.* 2023). According to Eid (2000), stand age and site index are also the most valuable variables for correct harvest decisions. Age is usually known in even-aged plantation forests (Packalen *et al.* 2011), but not in boreal managed forests. In Finland, for example, the long rotation period of the main tree species and the natural regeneration of minor tree species, such as aspen (*Populus tremula* L.), during the rotation period are reasons why forest age information is not always available (e.g. Maltamo *et al.* 2020).

Forest age is also key for the identification of old-growth forests, although there is no single age threshold. In Finland, for example, the definition of old-growth forest varies by region and by site type but typically refers to forests which are at least 120–150 years old (Kouki *et al.* 2018). In addition to the actual age, characteristics of old-growth forests include multi-layered canopies, a heterogenous spatial pattern of canopy gaps, a substantial variation in tree sizes and age structure, and a large amount and variety of dead wood (Ziaco *et al.* 2012, FAO 2022, European Commission 2023). Moreover, old-growth forest is commonly defined as a forest that consists of native tree species, that has no obvious signs of human activity and exhibits no significant disturbances in its ecological processes (da Silva *et al.* 2019, Barredo *et al.* 2021, FAO 2022). Approximately 34% of global forests are old-growth, and they are often fragmented and small in extent (FAO 2022). Old-growth forests have considerable provisioning potential with regard to ecosystem services, they store significant amounts of carbon and they provide habitats for various endemic and endangered

Handling editor: Dr. Ewa Grabska
Received 20 December 2024; **revised** 23 September 2025; **accepted** 14 November 2025

species (Eckelt et al. 2018, Hyvärinen et al. 2019). In Finland, for example, >34% of endangered forest species live in old-growth forests (Hyvärinen et al. 2019).

Age definition of single trees is usually straightforward, but definition of age at the plot- or stand-level, which is often the main interest, is more complicated. For example, in mixed forests, age often varies between tree species, and many old forests have canopy gaps caused by wind or other disturbances that lead to a substantial variation in tree age in those gaps and adjacent areas. The presence of single overstorey trees and vertically stratified canopies also make it difficult to accurately define plot- or stand-level age. Therefore, the definition may differ depending on silvicultural or ecological viewpoints.

The age of a tree is typically measured in the field through coring of the tree stem and the counting of tree rings. This field-derived age can be used as reference data in age modelling with remote sensing data. However, coring is a laborious process and, typically, multiple trees must be cored in a plot or stand. In many cases, coring is not allowed without the permission of the landowner (Koivuniemi and Korhonen 2006). An alternative approach in estimating forest age is visual aerial photo interpretation (Gillis et al. 2005). However, this can be problematic because it is a subjective process that relies on the experience and skill of the interpreter. Notwithstanding, visual photo interpretation and other remote sensing-based approaches have great advantages as they can cover large spatial areas, which allows for the generation of wall-to-wall maps, and they also reduce logistical and other costs.

Previous research has shown that forest age can be modelled with reasonable accuracy using remote sensing data. For this purpose, spectral data (Jensen et al. 1999, Gillis et al. 2005, Dye et al. 2012), airborne laser scanning (ALS) data (Maltamo et al. 2009, Racine et al. 2014, Wylie et al. 2019, Maltamo et al. 2020) or a combination of these (Straub and Koch 2011, Schumacher et al. 2020) have been tested. In mature forests, with increasing height the canopy is altered as gaps are formed due to death of single trees. These forests have been shown to exhibit a distinct reflectance in near-infrared (NIR) and short-wave infrared (SWIR) regions (Rautiainen et al. 2018, Hallik et al. 2019), which is why spectral data could be effective data in the mapping of forest age. Moreover, ALS data provide 3D structural information of the forest and so, could assist in the detection of altered canopy structures in ageing forests. These differences are potentially visible in ALS data. In previous research, age has commonly been modelled using linear regression models between a reference age as dependent variable and remote sensing features as independent variables, and stand-level has been the most utilized assessment scale (e.g. Maltman et al. 2023). Other previously tested approaches to map and predict forest age with remote sensing have included time series analysis (Vastaranta et al. 2016), change detection (post-disturbance) (Pan et al. 2011), indirect prediction through site index (Kandare et al. 2017) or successional stage (Falkowski et al. 2009), and forest biomass (e.g. Zhang et al. 2014).

The earliest studies related to remote sensing-based prediction of forest age often utilized Landsat TM satellite data. Cohen and Spies (1992) tested spectral and texture features of satellite data to predict forest age in a coniferous-dominated national park in the USA. Landsat TM satellite data were also utilized by Jensen et al. (1999) who compared traditional statistical approaches and artificial neural networks for age prediction in coniferous stands in Brazil. More recently, 3D data from ALS has been tested for predicting forest age. For example, Racine et al. (2014) and Wylie et al. (2019) utilized ALS data with the k-nearest neighbour (k-NN) approach to predict forest stand age in boreal forests in Canada. Both studies reported that ALS data offered extremely useful information for age predictions. Also, the joint use of ALS and multispectral aerial image data was tested by Straub and Koch (2011) in Germany.

Forest age prediction with remote sensing data has also been studied in Nordic countries. One of the earliest was a study by Reese et al. (2003) who predicted stand age using Landsat 7 satellite data and Swedish National Forest Inventory (NFI) data with the k-NN approach, whilst ALS data were utilized by Maltamo et al. (2009) to predict stand age in Finland with the k-most similar neighbour (k-MSN). To our knowledge, this was the only study conducted in Finland that predicted species-specific forest ages with remote sensing data. More recently, Maltamo et al. (2020) constructed nationwide ALS models to estimate forest stand age using Finnish NFI data. Schumacher et al. (2020) have also utilized ALS and Sentinel-2 data to map forest ages in Norway at both the plot- and stand-levels.

Earlier studies have pointed out that age predictions for forests older than ~100 years is a major challenge (Wylie et al. 2019, Maltamo et al. 2020). For example, Wylie et al. (2019) reported that the stand age-height relationship weakened after 120 years in boreal forests in Canada, which was why they opted for a separate model for plots older than 120 years. Maltamo et al. (2020) also reported that the relationship between stand age and ALS height metrics diminished after 100 years. In their study, they decided to omit stands with an age >100 years from their dataset. In Norway, Ørka et al. (2022) classified forests as natural and non-natural based on ALS-predicted age. They utilized 140- and 160-year age thresholds for natural forests, and classified forests as younger or older than the specific age threshold but with poor results. This suggests that existing approaches for estimating forest age from remote sensing still suffer from some drawbacks, with one key-challenge being the correct identification of old-growth forests.

In this study, we predicted forest age using ALS and Sentinel-2 data combined with a range of categorical variables available for the entire landscape. These categorical variables have previously been reported to be important in the modelling of forest age (Maltamo et al. 2020). For ground-truthing, we utilized the age of Finnish NFI plots located in northern Finland. Different combinations of predictor variables served as inputs to linear mixed effects (LME) models and tree-boosted mixed effects (GPBoost) models. A side-objective of our study was to test whether the identification of 'hold-over' plots prior to the prediction of actual age improves model performance. Here, 'hold-over' plots refer to plots that contained either *seedlings* or *retention trees* that were left unharvested on the site to promote natural regeneration or biodiversity.

Materials and methods

Study area

The study area was located in the boreal zone in northern Finland (Fig. 1). It was comprised of 12 ALS data blocks that were included in the national data acquisition programme (see *Remote sensing data* Section). In that programme, Finland has been divided into *blocks* that are approximately similar in the area covered. Aerial images and ALS data from ~40 blocks per year are produced in the programme. The total area covered by our 12 blocks was ~24 000 km^2. The main tree species in the study area were Scots Pine (*Pinus sylvestris* L.) and Norway spruce (*Picea abies* (L.) H. Karst). The most common deciduous tree species was downy birch (*Betula pubescens* Ehrh.). Other deciduous species, such as silver birch

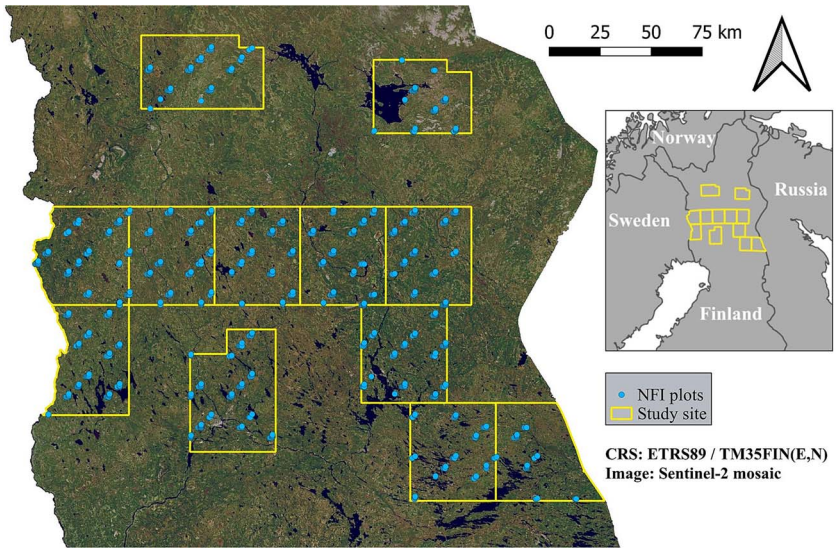

Figure 1. Locations of the 12 airborne laser scanning (ALS) data blocks and Finnish National Forest Inventory (NFI) field plots used in this study. Figures were created using QGIS 3.40.7 and assembled from the following data sources: Clipped Sentinel-2 mosaic (Copernicus browser) and clipped map of northern Europe (Esri world countries generalized).

(*Betula pendula* Roth) and European aspen, were in the minority. Most forests in the region are mixed, i.e. there are multiple tree species in a stand. Terrain height at the plot locations (blue dots in Fig. 1) ranged from 80 m to 420 m above sea level with a mean terrain height of 207 m above sea level.

Field data

The NFI plots were measured between 2019 and 2022. The number of trees per plot varied between 0 and 32. The sampling design of the Finnish NFI is explained in Korhonen *et al.* (2021) and all details related to Finnish NFI field measurements are provided in a field manual (Luke 2021). In brief, NFI plots are concentric with two radii employed depending on diameter at breast height (DBH) values within the plot: trees with DBH values >95 mm were included within a radius of 9 m radius and trees with DBH values 45–94 mm were included within a 4 m radius. Tree species and DBH were recorded and measured for all included trees. Tree height was measured for sample trees (two sample trees per plot, on average) and the heights of the other trees were predicted with a mixed effect model proposed by Myllymäki (unpublished report). The height model is presented in detail in Supplementary Material Table S3.

Age was determined for each tree stratum, defined based on tree species and canopy layer, and also—in the seedlings—the origin. This means that age was determined for all tree species and canopy layer mixtures (i.e. stratum) in all plots other than those that contained seedlings. For the latter, the origin of the seedling (regeneration type) was used as the third factor separating the stratum. For the main tree species in the dominant stratum, age was always defined by coring. The number of trees to be cored was dependent on the plot. For an even-aged plot, one tree was deemed sufficient, whilst for older plots and plots with an uneven-age structure, more than one tree was cored. The age for other strata of a plot was estimated visually. Age for these plots was calculated as the basal area weighted mean age of the trees. For the seedlings (mean height < 1.3 m), age was calculated from the number of branch whorls. In the larger seedlings (mean height > 1.3 m), age was calculated as the mean age of the trees. This age was used as the *observed* plot age in our predictions. Plot ages were updated to match the year of ALS data acquisition. Due to the systematic sampling design, NFI plots can be located either completely inside a forest stand or in between multiple forest stands of different ages. We only used NFI plots that were completely within one forest stand because plot age is not unambiguous if a plot was located in several stands.

One objective of our study was to identify plots that contained hold-over tree(s) (hereafter hold-over plots). Hold-over plot identification was needed as plot age is defined by the age of the dominant stratum. On hold-over plots, the dominant stratum was composed of seedling trees, and the hold-over trees may hamper plot age prediction when ALS data is used. A plot was considered as a hold-over plot if it had at least one tree that belonged to the hold-over stratum and the recorded age of the plot was ≤50 years. In this study, plots other than hold-over plots are referred to as 'other plots'.

In total, there were 870 plots across the 12 ALS data blocks. From these, 50 plots were identified as hold-over plots (5.7% of all plots). Plot age ranged between 0 and 300 years and the most common age class was 70–80 years (Fig. 2). The mean and median plot ages were 87 and 79 years, respectively. The number of plots ≥150 years old was 83.

The extent to which basal area weighted mean height and DBH vary by plot age is illustrated in Fig. 3 and is indicative of why age prediction is such a challenging task when remote sensing data are used. For example, basal area weighted mean height varied considerably within the 100-year old plots: the smallest values were ∼4 m and the largest ∼20 m (Fig. 3A). Correspondingly, mean height may be the same in the 50- and 250-year old plots. Mean diameter (Fig. 3B) showed a similar but weaker trend. Given the similarities in mean height between plots of different ages, it was clear that we also needed predictor variables other than canopy height to ensure accurate age predictions.

Remote sensing data

In this study, we used ALS data and Sentinel-2 satellite imagery. The former are part of the national data collection campaign (KALLIO) organized by the National Land Survey of Finland. The

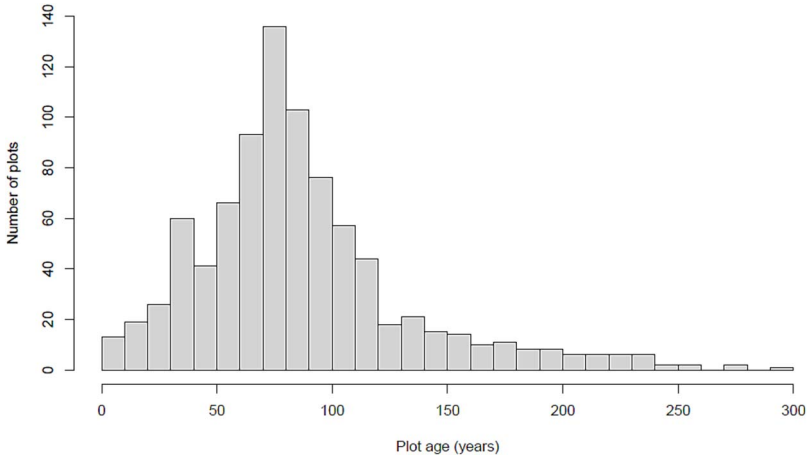

Figure 2. Histogram for the age distribution of plots used in this study. Each bar equates to a 10-year age class.

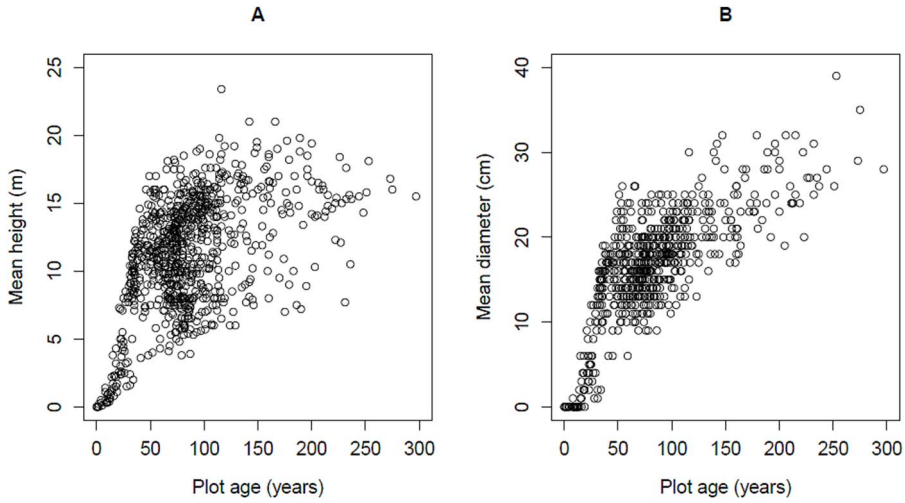

Figure 3. Scatter plots of plot ages with respect to field measured basal area (A) weighted mean height, and basal area (B) weighted mean diameter.

Table 1. Metadata of the airborne laser scanning (ALS) datasets used in this study.

Laser scanners	RIEGL VQ-780i RIEGL VQ-780 II RIEGL VQ-780 II-S RIEGL VQ-1560 II RIEGL VQ-1560 II-S
Wavelength	1064 nm
Flying altitude	855–2100 m a.g.l.
Flying speed	140–160 knots
Scanning frequency	131–230 Hz
Pulse repetition frequency	700–1620 kHz
Point density	5.1–9.5 p/m^2
Maximum scanning angle	20–23°
Side overlap	20–24%

ALS data were acquired during leaf-on conditions during the summers of 2020, 2021 and 2022, depending on the ALS block. Details of the ALS data are shown in Table 1. The original echo heights were normalized to above ground level (a.g.l.) using a digital terrain model interpolated from ground echoes. All processing of ALS data was carried out in the PALUS remote processing platform provided by the scientific computation center 'CSC–IT Center for Science' (CSC 2025).

The Sentinel-2 mosaic was generated in the Google Earth Engine (GEE) using the percentile-based method described in Pitkänen *et al.* (2024). The mosaic included all the bands, where the 20 m pixels were resampled to 10 m pixel size using the nearest neighbour method. The mosaicking process started with selection of all the summertime (between June 15 and August 15) Level-2A images between 2020 and 2022 that had a maximum overall cloud cover of 25%. Then, these initial images were masked to remove the majority of the remaining clouds by using a threshold of 30%, based on the S2_CLOUD_PROBABILITY layers (Zupanc 2017). The final output was calculated using the band-wise 40th reflectance percentile values, derived from all the unmasked pixels. The 40th reflectance percentile was selected as it has been shown to be relatively powerful in avoiding cloud shreds and shadows in the final mosaic (Pitkänen *et al.* 2024). The resulting mosaic was also visually checked and verified to confirm that no clouds or shadows were present.

Plot level-metrics

The ALS and Sentinel-2 metrics and categorical variables were calculated at the plot-level. The ALS metrics were calculated separately for the *first* (*f*) and *last* (l) echo categories using the *lascanopy* tool of LAStools (LAStools 2024). Height metrics included height percentiles (*p20, p40,* …), height bincentiles (*b05, b10,*

b20, ...), maximum (max), mean (*mean*), standard deviation (*sd*), skewness (*ske*) and kurtosis (*kur*) for height, and vertical complexity index. Height bincentiles (a.k.a. deciles) were calculated as the proportion of all echoes below a specific percentage of the maximum height (e.g. *b20* is the cumulative fraction of echoes up to the 20% of height). Vertical complexity indices were calculated for 1 m (*vc1*) and 2 m (*vc2*) sized bins as proposed by van Ewijk *et al.* (2011). Density metrics *d0–5*, *d5–10*, and *d10–15* were calculated as proportions of echoes between specific heights. The first digit of each density metric denotes the starting value (m) and the second denotes the end value (m) of the interval. Canopy cover (*cov*) was calculated as the number of *first* echoes above the cover cutoff (2 m) divided by the number of all *first* echoes. Similarly, canopy density was calculated as the number of all echoes above the cover cutoff divided by the number of all echoes. The ALS metrics also included two terrain height metrics: the terrain height at the plot location above sea level (*TH*) and the relative terrain height calculated as the terrain height at the plot location divided by the mean terrain height in the 3 km radius around the plot (*rTH*).

The Sentinel-2 metrics contained the mean values for each band (e.g. *B2*), the ratios between the bands (e.g. *B2/B11*) and the Normalized Difference Vegetation Index (*NDVI*). Categorical variables (*ALS block*, *WoodProdRestr,* and *MainType*) were derived for the plots from the NFI data and were available in raster or vector format for the whole of Finland. The ALS block signified the inventory areas of this study (i.e. 12-level factor). The WoodProdRestr variable describes wood production restrictions of a forest and has three levels: 'no restrictions', 'protected forests' and 'protected mires'. The 'protected forests' category mostly includes forested plots located in national parks, old-growth forest reserves and other nature reserves. MainType describes the main type of forest plot and has three levels: 'mineral soil', 'drained peatland', and 'undrained peatland'.

Prediction scenarios and methods

We compared age predictions between the LME and GPBoost models for three scenarios (SC1, SC2, and SC3). These approaches to age prediction represented two of the main categories of models used to predict forest parameters: statistical models (LME) and machine learning models (GPBoost). Both statistical models and machine learning models are built upon a statistical framework. In general, the objective of machine learning is to obtain repeatable and accurate predictions for the target variable. In contrast, statistical modelling is more about model interpretability and examination of the relationships between the variables. Whilst machine learning models can capture more complex nonlinear relationships, they tend to be sensitive to overfitting, which is usually not an issue for the structurally simpler statistical models. In this study, the age predictions with LME are perceived as the benchmark for GPBoost.

The workflow of age predictions is presented in Fig. 4. In the first scenario (SC1), plot age was predicted using only plot metrics calculated from remote sensing data. In the second scenario (SC2), the categorical variables were added to the configuration of SC1. In the third scenario (SC3), the hold-over plot classification prior to age prediction was added to the configuration of SC2. The proposed scenario-wise age prediction was made to highlight the effect of categorical variables (SC1 vs. SC2), and whether the detection of hold-over plots would decrease the prediction error (SC2 vs. SC3). Age models were fitted separately for the hold-over and other plots in SC3. Age predictions for hold-over and other plots differed in the structure of the random component of the models. For other plots, we treated categorical variables as random effects, whereas in the models for the hold-over plot (SC3 only) we did not include them in the models at all as the hold-over plots only contained a small number of observations in most categories. The categorical variables of interest are typically modelled as fixed effects, but we decided to model them as random effects, so that in cases where the information would not be available, the model would still be usable. Also, as ALS blocks can be seen as a sample of all ALS blocks, their treatment as random effects (as other categorical variables) ensures that the model behaves more logically. Analyses in this study were implemented in the R environment (R Core Team 2024).

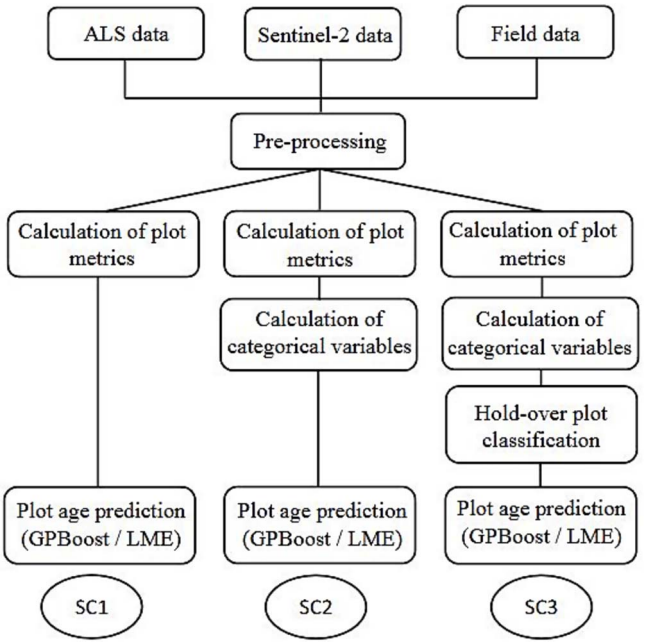

Figure 4. Workflow for age predictions on National Forest Inventory (NFI) plots across the three scenarios: SC1, SC2, and SC3.

The initial set of predictor variables consisted of 42 ALS and 56 Sentinel-2 metrics. Prior to the fitting of the age models, we removed predictor variables until the correlations between the remaining metrics were <0.9. In practice, the correlation matrix was formed, upper triangle and diagonal were set to zero, and all rows with values >0.9 were removed. From the remaining metrics, we selected 10 metrics that exhibited the strongest correlation with plot age. This was done within each 10-fold cross validation scheme as described in *Validation and feature importance* Section. The resulting 10 metrics in each fold were then included in the age models in SC1, SC2, and SC3.

Linear mixed effects model

As the basis for the LME model, we used the *lmer* function from the lme4 package (Bates *et al.* 2015). Furthermore, a selection of the maximal model that can still converge was carried out using the *buildmer* function from the buildmer package (Voeten 2021). First, we formulated the 'maximum model' with *lmer*, which contains all the predictors that were used in the corresponding GPBoost model. In SC1, we only used fixed effects (i.e. ALS and Sentinel-2 predictor variables) in the model. In SC2 and SC3, categorical variables were included in the mixed effects model by adding random intercepts for the categorical variables. Categorical variables were treated as crossed random effects. After model formulation, *buildmer* finds the maximal feasible model by starting with an 'empty model' and adding terms to this model until the model

can no longer converge. New terms were added in order of their contribution to the Akaike Information Criterion (AIC). After the maximal feasible model was found, stepwise elimination was applied. The elimination was carried out with a default direction (backward stepwise elimination) and AIC was used as the criterion for term elimination. In SC2 and SC3, ALS block, MainType and WoodProdRestr were included in the random part of the models. This means that the 'empty' *lmer* model was not entirely empty but contained the pre-defined variables in the random component of the model. Finally, the best model of *buildmer* was fitted with *lmer*.

GPBoost

Tree-boosted mixed effects (GPBoost) is a relatively new software library used to combine tree boosting with the Gaussian process and grouped random effects models (Sigrist *et al.* 2021, Sigrist 2022). To predict age, we used tree-boosted mixed effects, and the Gaussian process part of the library was not used. Here, boosting relaxes the linearity assumptions in mixed effects models, whilst the mixed effects, in turn, allow relaxation of the assumption of independent data in boosting. Our model was of the following form (Sigrist 2022)

$$y = F(X) + Zb + \varepsilon \quad (1)$$

where b is the parameter vector related to the grouped random effects that model categorical variables, and ε is the vector of independent normally distributed error terms. In our model, the response variable y is a vector that contains the plot ages in the n plots, X is the matrix that contains the values of the 10 chosen plot metrics of ALS and Sentinel-2 data for all plots, and Z is a matrix that contains the necessary information related to the grouping structure provided by the categorical variables. In the LME models, it would be assumed that $F(X) = X^T\beta$, where β is the vector of coefficients; however, here $F(X)$ represents the fixed effects given by a regression tree (Breiman *et al.* 1984), see details in Sigrist (2022). The group effects related to the groups of a categorical variable were assumed to be independent and identically normally distributed. The model was trained using the *gpboost* function, including optimization of tuning parameters with a grid search (function *gpb.grid.search.tune.parameters* of the gpboost package) using a principle of nested cross-validation (see *Validation and feature importance* Section). The input set of values for hyperparameter tuning are presented in Supplementary Material (Table S1). The chosen hyperparameter values are thresholds, which are based on experiments with our datasets.

Logistic regression

Scenario 3 (SC3) included the hold-over plot classification, which was implemented with logistic regression using the *glm* function from the R package 'stats' (R Core Team 2024). For the stepwise model selection, we used the *stepAIC* function from the MASS package. The candidate set of predictor variables was the same as for age prediction, although categorical variables were included as the fixed effects in the model formulation.

Validation and feature importance

The results were validated using a 10-fold cross validation procedure. That is, we split the data into 10 subsamples, i.e. folds. Then, each plot fold (~10%) was, in turn, used as the test data, and the remainder of the plots (~90%) were used as training data. For each fold, we sampled plots from each of the 12 inventory areas without replacement. The same folds were used for both prediction methods (LME and GPBoost), logistic regression and for all scenarios. Inside of training folds, we further divided plots into actual training data and validation data: hyperparameter optimization of GPBoost was validated using the validation data.

To minimize the effect of randomness due to random selection of observations into the 10 folds in cross-validation, we repeated the validation process 50 times. This means that age was predicted 50 times for each plot and the mean of these 50 values was used as the predicted value. We used relative RMSE (Eq. 2) and relative bias (Eq. 3) as measures of prediction accuracy:

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

$$\%\mathrm{BIAS} = 100 \times \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)}{\sum_{i=1}^{n} y_i} \quad (3)$$

where $\hat{y}_i$ is the predicted age on plot i, y_i is the observed age on field plot i, $\bar{y}$ is the mean of observed ages on plot i, and n is the number of plots.

Finally, we analyzed the most important predictors of the hold-over classification and the GPBoost models of SC3. We used the importance measure *Gain* for GPBoost. It indicates the relative contribution of the predictor variable to the model: the higher the Gain value compared to other predictors, the more important it is for the model (Sigrist *et al.* 2021). A Gain value of 1 denotes high variable importance and Gain value of 0 denotes low variable importance, respectively. The presented Gain values are the mean values over the 50 repetitions. The metric importance for the hold-over plot classification was the number of times that the predictor variable was chosen in the logistic regression models divided by the total number of logistic regression models. The total number of logistic regression models equals 500 (10 folds × 50 repetitions = 500 models).

Generating wall-to-wall map

Wall-to-wall mapping of forest age was also tested for the area of interest (AOI) (Fig. 8). The AOI was located in the Huttutunturi area (near Pyhä-Luosto National Park), and it contained both managed forests and forests located in the national park. Wall-to-wall mapping was based on the GPBoost model of the SC3 approach. First, a logistic model for hold-over plot classification was fitted and predictions were applied for each 16 m × 16 m grid cell in the AOI. Next, separate age models were fitted for the hold-over and other plots, based on predicted classes in the hold-over plot classification. For the production of the final age map, a subset of the Corine Land Cover 2018 dataset was utilized to mask out water bodies and non-forested areas (SYKE 2018).

Results

Predicted vs. observed age scatter plots of our three scenarios are presented in Fig. 5. The first scenario (SC1) yielded the largest RMSE values for both methods (Fig. 5). The RMSE value associated with the LME model was over 4 percentage points larger than the value associated with GPBoost model. The LME model produced smaller bias than the GPBoost model although the bias for both models were marginal. Both methods clearly failed to predict the ages of plots that had an observed value close to zero. Overall, the predictions of GPBoost (Fig. 5) fitted better to the 1:1 line than the predictions of LME (Fig. 5). The range of predictions was clearly smaller in the LME model compared to the GPBoost model: the age of young plots was clearly overpredicted in LME and the age of old plots was underpredicted.

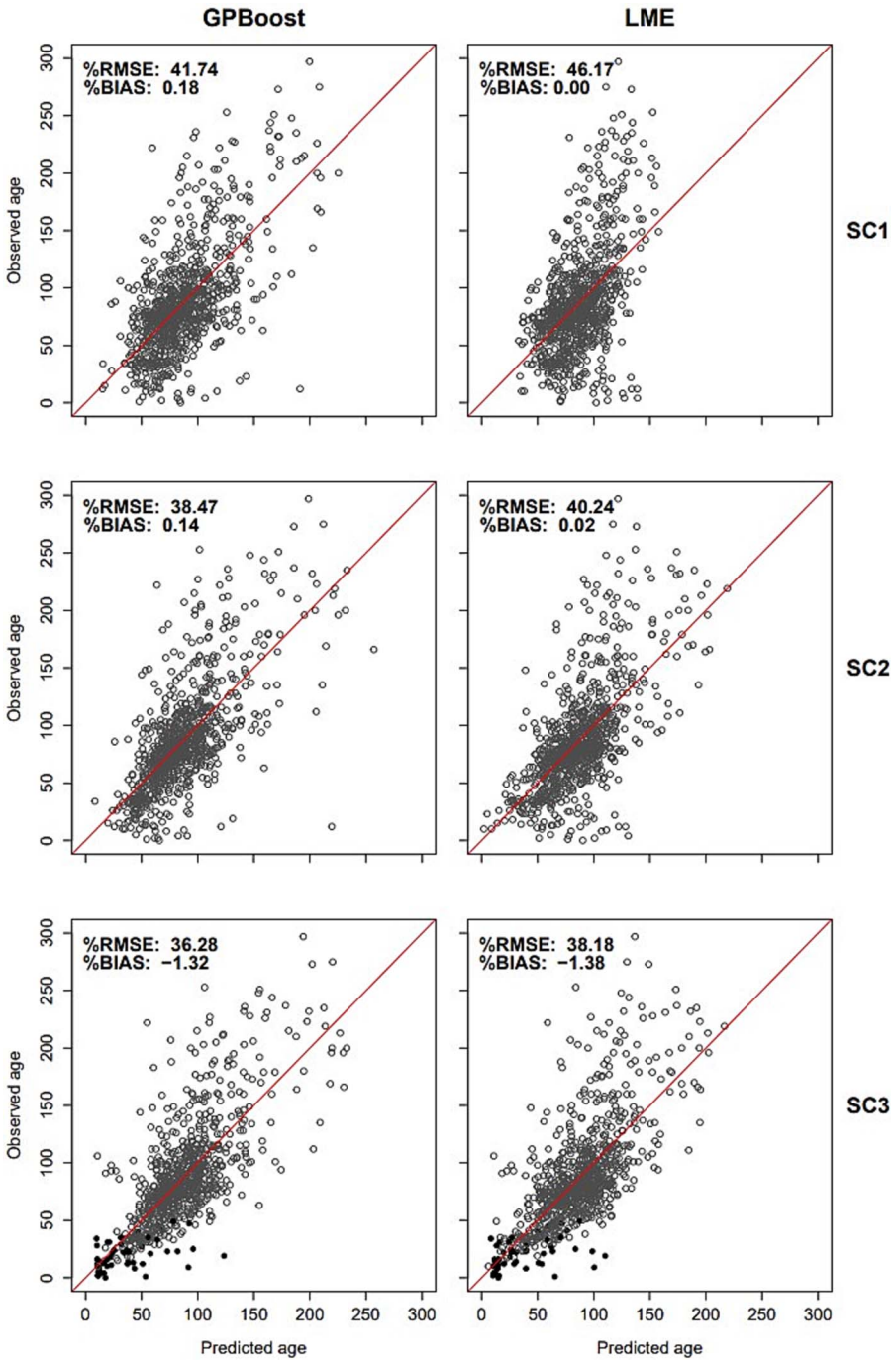

Figure 5. Predicted vs. observed age for the linear mixed effects (LME) model and tree-boosted mixed effects (GPBoost) model in the first scenario (SC1), second scenario (SC2), and third scenario (SC3). Observed hold-over plots are highlighted in black in the SC3 scatter plots.

In SC2, both methods yielded improved accuracies compared to SC1 (Fig. 5). The RMSE value associated with LME decreased more than the RMSE value associated with GPBoost (decrease in 5.93 percentage points vs. 3.27 percentage points) compared to SC1. With regard to RMSE, GPBoost was clearly the better alternative of the two methods. The LME model produced smaller bias than the GPBoost model, although bias in both cases was again minimal when compared to RMSE. As in SC1, both methods failed to accurately predict the age of plots that had an observed value close to zero, and this tendency was stronger in LME. However, in some instances, ages close to zero were better predicted with LME than with GPBoost. For both methods, the predicted values in SC2 fitted better to the 1:1 line than in SC1. The LME predictions had a clearly wider range in SC2 than in SC1 and, therefore, overprediction of young ages and underprediction of old ages were notably reduced.

The third scenario (SC3) utilized the hold-over plot classification prior to age prediction to separate age predictions for hold-over and other plots. Logistic regression models for hold-over plot classification provided an overall accuracy of 94.7% and a *kappa* coefficient of 0.58. The error matrix for the hold-over plot classification over 50 repetitions is provided in Supplementary Material (Table S2).

In SC3, both methods performed better than in SC2 (Fig. 5). The RMSE value was reduced by 2.06 percentage points in LME and by 2.19 percentage points in GPBoost compared to SC2. Bias

Table 2. Root mean square error (RMSE) and bias values for plots with observed ages ≥150 years. Values are calculated for each scenario and for each modelling approach, respectively.

	%RMSE	%BIAS
SC1		
LME	44.58	−41.19
GPBoost	37.11	−31.27
SC2		
LME	39.47	−32.34
GPBoost	36.33	−28.59
SC3		
LME	38.92	−31.78
GPBoost	35.78	−27.85

were clearly larger in this alternative but the addition of hold-over plot classification clearly resulted in better prediction of plots with observed ages close to zero compared to SC1 and SC2. Poor age predictions for the remaining single hold-over plots close to zero can mostly be accounted for by the fact that hold-over classification failed in those plots. It should be noted that both methods used the same classification for hold-over and other plots. Predictions in plots with observed ages close to zero were slightly improved compared to SC2, and in general, the predictions more closely followed the 1:1 line.

We further investigated age prediction accuracy for plots with observed ages ≥150 years (Table 2). In each scenario, the GPBoost model yielded lower RMSE and bias values for these plots. For both approaches, the overall trend in the accuracy statistics was similar to those computed from the complete age range (see Fig. 5).

The random effect of the SC3 models for 'other plots' is shown in Fig. 6. The effect of ALS block was minor compared to the other categorical variables. The strongest random effect was in the 'protected forests' category, which increased the predicted age by 30 years in the LME model and by 40 years in GPBoost model. The second strongest random effect was observed in the 'undrained peatlands' category, which increased the predicted age by 20 years in the LME model and by 25 years in the GPBoost model. For the 'drained peatlands' category, the increase in age was 5 years in the LME model and 10 years in the GPBoost model. The strongest negative random effect was in the 'mineral soil' category, which decreased the predicted age by 25 years in the LME model and by 10 years in the GPBoost model. For plots with 'no wood production restrictions', the predicted age decreased by 25 years in the LME model and by ∼5 years in GPBoost model. The random effect for the 'protected mires' category for the LME model decreased predicted ages by 5 years, although it increased the predicted ages in the GPBoost model by 20 years. This is likely a consequence of the different prediction levels in the fixed parts of the two models.

Standard deviations associated with the random effects were similar in LME and GPBoost (Table 3). The greatest difference was observed between residual errors: ∼31 years for LME and 17 years for GPBoost. Overall, WoodProdRestr exhibited the largest variation between groups (∼30 years). The variation in MainType random effects was ∼20 years, whilst ALS block exhibited the smallest variation between groups and was close to zero.

The 10 most important metrics for age prediction in the hold-over and other plots, and hold-over plot classification are presented in Fig. 7. These correspond to the GPBoost models in SC3

Table 3. Estimates of random effect standard deviations (in years) from linear mixed effects (LME) models and tree-boosted mixed effects (GPBoost) models in scenario 3 (SC3).

Random effect	LME Std. Dev.	GPBoost Std. Dev.
WoodProdRestr	28.34	27.68
MainType	21.07	18.37
ALS block	1.76	2.26
Residual	31.43	17.01

reported above. Overall, ALS metrics were found to be the most important predictors for age prediction (Fig. 7A and B). For the hold-over plot classification (Fig. 7C), all categorical variables and Sentinel-2 metrics were amongst the most important predictors. The ALS metrics related to maximum height and density were very important for age prediction in both hold-over and other plots (Fig. 7A and B). The greatest difference between hold-over and other plots was that the most important Sentinel-2 metric was ranked as second for hold-over plots, whereas for the other plots, the most important Sentinel-2 metric was only the fifth most important. Also, the most important Sentinel-2 metric for hold-over plots (B3) was not ranked in the 10 most important metrics for other plots. For both hold-over plots and other plots, half of the 10 most important metrics were Sentinel-2 and ALS metrics, respectively. The ratio of some Sentinel-2 bands were more often amongst the most important metrics than the value of the Sentinel-2 band *per se*.

Wall-to-wall map of forest age in the Huttutunturi area highlighted differences in the forest age between managed forests and forests located in conservation area (Fig. 8). White areas in the aerial image are rocky tops of the Huttutunturi fell that were masked out by the forest mask. In the lower left corner of the aerial image there is a dark area which represents a lake (Huttujärvi) that separates managed forest area on its left-hand-side from the natural park on its right-hand-side. In the lower right corner of the map, there is also a clear distinction in the predicted ages between managed forest and national park forest. Forest areas in the national park in this map (Fig. 8) are, on average, older than the forests in managed forest region.

Discussion

We predicted plot ages over a large area in northern Finland using a combination of ALS data, Sentinel-2 images and categorical variables that were available for the entire landscape. We compared two prediction methods using the same features as predictor variables: LME modelling and the GPBoost algorithm. Our results showed that predictions by GPBoost were better than those by LME in each tested scenario and that the categorical variables included as random effects were important predictor variables. To our knowledge, this is the first study that has used tree-boosted mixed effects (GPBoost) modelling in the field of remote sensing of forests.

In this study, we formulated LME and GPBoost models by adding random effects to the intercept of the models, although we also tested multiple structures of random effects, including random slopes and nested random effects. The purpose of using random effects in our models was to improve prediction accuracy. It is assumed that the categorical variables in the random part are always available, however, if this is not the case, the specific

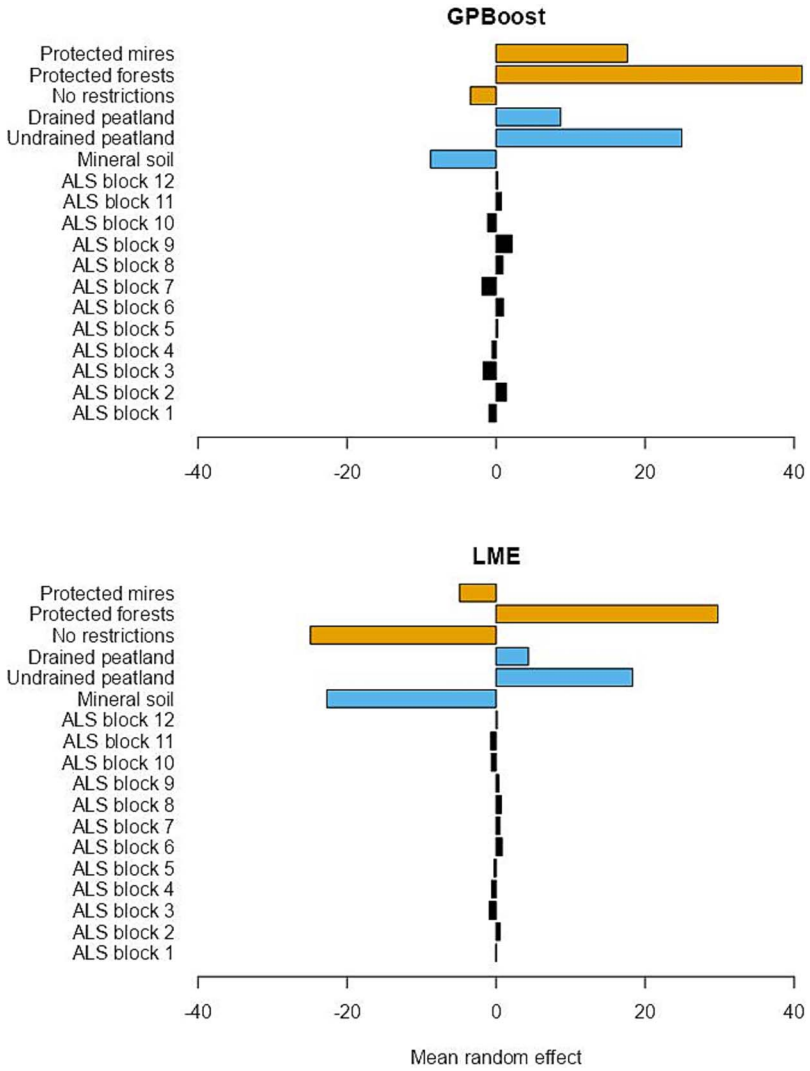

Figure 6. Mean random effects by categorical variables in the 'other plots' models of scenario 3 (SC3). Effects are presented separately for the linear mixed effects (LME) model and tree-boosted mixed effects (GPBoost) model. Categorical variables are presented in the following colours: WoodProdRestr (levels 0,1,2) as orange, MainType (levels 0,1,2) as blue, and airborne laser scanning (ALS) block (levels 1 to12) as black.

variable from the random component can be dropped in the prediction phase.

The addition of categorical variables was most notable for LME: the range of predictions increased from 33 to 158 years in SC1 to 2–219 years in SC2, and the RMSE value decreased by 5.93 percentage points. The same phenomenon has also been reported in previous studies. For example, Maltamo *et al.* (2020) reported that the categorical variable that separated peatlands and mineral soils was significant in their age model. The stratification into peatlands and mineral soils has also been used in the multi-source Finnish NFI (Mäkisara *et al.* 2022). This is because trees on mineral soils grow faster than trees on peatlands and are, on average, younger when they reach a specific height.

During the construction of the age models, we noticed that plots which contained hold-over trees exhibited greater prediction errors, so we included a hold-over plot classification prior to the fitting of the age models in SC3. One potential advantage of this setup is that the age model can be used even when the actual hold-over plots are not known. Prior classification of hold-over and other plots, and the fitting of separate age models for these two plot classes was beneficial as predictions for plots with observed ages close to zero clearly improved. However, the observed age of single hold-over plots were incorrectly predicted, which can partly be explained by the failed classification of hold-over plots (Fig. 5). One reason for these misclassifications is that hold-over plots exhibited a very wide diversity of forest structures, although the proportion of hold-over plots was only 5.7% of all plots.

In SC3, the predictions yielded slightly increased bias values compared to SC1 and SC2, which was associated with the division of age modelling between hold-over plots and other plots. Further inspection revealed that the predictions for other plots in SC3 yielded small positive bias (~1%) for both LME and GPBoost models, whilst the age predictions for hold-over plots had very high negative bias (approximately −60%). This explains the overall increased bias observed in SC3. One reason for such a high bias may rise from the small dataset of hold-over plots in which the variation in forest structures and spectral responses for plots with similar ages was relatively higher than in the larger dataset of other plots.

In Fig. 5, there is a small cluster of six plots in SC3 that had observed ages near 100 years and predicted ages between 0–30 years. Further inspection of these plots revealed that all have recently been harvested and seed trees were left from the previous

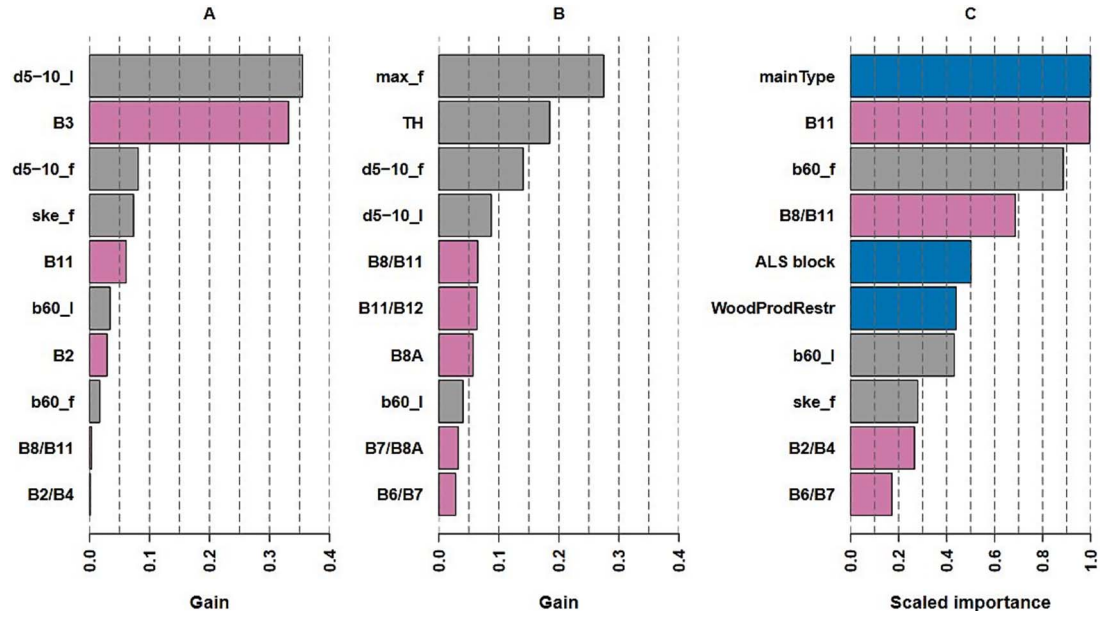

Figure 7. Importance (gain) for the 10 best predictor variables in tree-boosted mixed effects (GPBoost) models in scenario 3 (SC3) for hold-over plots (A) and other plots (B). Plot C signifies the importance of the 10 best metrics for the hold-over plot classification with logistic regression. Grey, pink, and blue signify airborne laser scanning (ALS), Sentinel-2, and categorical variables, respectively. Categorical variables were not considered in sub-figures A and B, as they were not included in the model for case A and were included in the random part of the model in case B. All abbreviations for the predictor variables are defined in the *Plot level-metrics* Section.

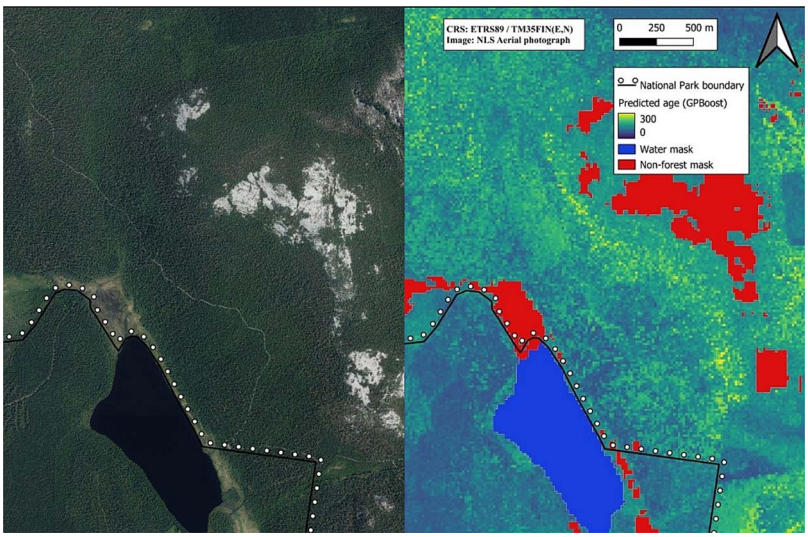

Figure 8. Aerial image of the Huttutunturi region (left) and wall-to-wall map of forest age predicted with GPBoost utilizing approach of SC3 (right). Blue areas and red areas in the age map represent water bodies and non-forested areas, respectively. Figures were created using QGIS 3.40.7 and assembled from the following data sources: Clipped aerial image (National Land Survey of Finland) and Corine land cover 2018 (SYKE 2018).

generation. The ages of these plots were obtained based on the ages of the seed trees as the seedlings had not yet emerged. The ALS data and Sentinel-2 image indicated that these plots were significantly younger than the observed value. This was because there were few ALS hits above-ground and most of the spectral response came from the ground areas with scant vegetation.

Further investigation of the age prediction accuracy of plots with observed ages ≥150 years (Table 2) showed mostly similar trends as with the whole dataset (Fig. 5). The reduction in RMSE values for the ≥150-year-old plots from SC1 to SC3 was not as large as for the whole dataset: for example, with GPBoost, the reduction in RMSE values for the whole dataset was 5.46 percentage points and was 1.33 percentage points for the ≥150-year-old plots. Similarly, the reduction in RMSE values for LME was larger than for GPBoost, but not as large as in the whole dataset (7.99 percentage points vs. 5.66 percentage points). This means that most of the accrued prediction accuracy for LME was due to increased accuracy in the older plots (≥150 years old). However, for GPBoost, most of the gained prediction accuracy originated from increased accuracy of the ages of the younger plots.

Group-level means for random effects indicate that ALS block did not have a significant effect compared to the other random effects (Fig. 6). This implies that various ALS acquisition setups and ALS data processing phases had only minor effects in age modelling. In practice, this would suggest that the ALS data

acquired with different sensors was usable and did not affect the age predictions to a large extent. Despite this, the premise of this study was that the age model is never transferred to an ALS block not existing in this data. However, the inclusion of ALS block as a random variable is justified as ALS data were acquired block-by-block at different times within the growing season, and with a range of ALS sensor types and data acquisition configurations.

The mean random effects for the WoodProdRestr variable were largest for plots located in protected forests. The large positive random effect would indicate that plots in protected forests are ∼30 (Fig. 6, LME) or 40 (Fig. 6, GPBoost) years older than the expected value. The absolute mean random effects for the MainType variable were largest for the undrained peatlands (Fig. 6). Plots on undrained peatlands were ∼20 years older than the expected value as they have low productivity compared to drained peatlands, if their 3D structure and spectral responses are assumed to be similar. The negative random effect observed for plots on mineral soils can be similarly explained: mineral soil plots are the most productive of the MainType levels and are, on average, younger than other plots.

A direct comparison to earlier studies that have assessed the use of remote sensing in predicting forest age is not straightforward for a number of reasons. For example, ALS and image data used may vary considerably between studies, the predictor variables computed from the data may also differ, and the target forests are not the same (e.g. forest type and age range). Despite these discrepancies, there are a number of published studies to which we can compare our results. For instance, in the study by Schumacher *et al.* (2020), ages ranged from 3 to 270 years for spruce-dominated plots, which is similar to the range in our dataset. They reported a mean RMSE value of 25.4% for their site index specific models. Prior information of site index was critical in their age models, as the remote sensing-based classification of plot/stand site index clearly weakened the performance of the age models. In Canada, Racine *et al.* (2014) reported a RMSE value of 19% for their age models that used ALS data. The age range in their field data was clearly narrower than in our study: between 11 and 94 years. Maltamo *et al.* (2009) reported RMSE values for spruce-, pine-, and deciduous-dominated plots, which were higher than in our study: 88%, 51%, and 102%, respectively. This is a logical outcome since the prediction of age by tree species in mixed stands is a much more difficult task than the prediction of the age of the total growing stock. More recently, Maltamo *et al.* (2020) predicted the plot-level age of trees by ignoring tree species. In their age model, they included common ALS metrics, a set of variables computed from NFI data (e.g. a dummy variable for mineral soils) and geographical variables as predictor variables. They reported a plot-level RMSE value of 32.9% for their model. The RMSE values reported in our study are similar, although the data used in Maltamo *et al.* (2020) had a narrower age range than our data. For Straub and Koch (2011), the ages of the field plots ranged from 0 to 153 years. They reported a RMSE value of 29%, which is less than in this study, and may be explained by the fact that their study was conducted over a relatively small area (9.2 km^2). Ørka *et al.* (2022) reported that age-based classification of forests as natural or non-natural (140 and 160 year thresholds) resulted in kappa values <0.05, which is indicative of very poor performance. Aside from Maltamo *et al.* (2020), Schumacher *et al.* (2020) and Ørka *et al.* (2022), the studies mentioned above were conducted on substantially smaller areas than our study. Also, the age range across the plots was clearly wider in Schumacher *et al.* (2020) than in the other studies. Given the range of ages in the field data and the extent of our study area, our results correspond well with the earlier research.

The most important metrics for age models in our study are in agreement with those reported in earlier studies. The relationship between forest age and height is widely recognized, which is why ALS height information has often been reported as an important age predictor variable (e.g. Racine *et al.* 2014, Vastaranta *et al.* 2016). The importance of the high order percentiles (e.g. 90^{th} and 95^{th}) and maximum vegetation height have been reported often in earlier studies (Straub and Koch 2011, Maltamo *et al.* 2020, Schumacher *et al.* 2020). This is in line with our study as we reported that maximum vegetation height was the most important metric for plots other than the hold-over plots (Fig. 7). The ALS metric d5–10 was one of the most important metrics in our age models as it fundamentally describes the complexity of a canopy. We did not find a previous study that used this specific ALS metric in age models. Other ALS metrics that describe the complexity of the canopy (e.g. canopy cover, canopy closure) were reported to be important for age models in some studies (Straub and Koch 2011, Wylie *et al.* 2019, Schumacher *et al.* 2020). Terrain height, a metric that describes the elevation of a plot in a national height system, was the second most important metric in this study and its importance has also been reported in earlier research (Racine *et al.* 2014, Wylie *et al.* 2019, Schumacher *et al.* 2020). The importance of terrain height is rational as trees at higher altitudes have limited growth potential compared to trees at lower altitudes.

For Sentinel-2 metrics, our results on metric importance showed good agreement with the earlier research. As in our study, the importance of B8A and B11 were reported to be the most important Sentinel-2 bands by Schumacher *et al.* (2020). They reported that the Sentinel-2 metrics contributed less to the model than the ALS metrics, which also agrees with our observations. Similarly, on a smaller scale, Rautiainen *et al.* (2018) reported that the age of a pine needle has influence on reflectance in NIR (equivalent to B8 and B8A in Sentinel-2) and SWIR (equivalent to B11 in Sentinel-2) regions.

Most of the hold-over plots were not located on undrained or drained peatlands. This explains why MainType was found to be the most important metric for hold-over plot classification. Another important metric was B11 and its division with B8, also known as the moisture stress index (Hunt and Rock 1989). The B11 surface reflectance was greater on hold-over plots than on other plots. Also, B8/B11 was an important metric, as it exhibited lower values on hold-over plots compared to other plots (caused by the greater B11 values). The 60^{th} bincentile of ALS heights was an important metric in the classification as well: average values were significantly greater on hold-over plots than on other plots (92% vs. 80%). A greater value for this bincentile means that most of the echoes are clearly located lower than the dominant canopy, which would indicate the presence of hold-over trees.

The most important metrics of the hold-over plot age models differed slightly from the metrics for the other plots. The absence of ALS maximum height can mostly be explained by the fact that age in the hold-over plots was determined by seedlings and, therefore, maximum height, which represents hold-over trees, was not an important metric. Another difference between the most important metrics between the hold-over and other plot age models were the Sentinel-2 metrics B2 and B3, which were not amongst the 10 most important metrics in the other plot age models, although B3 was reported to be the second most important metric for hold-over plot age models. Further inspection of the hold-over plots showed a strong negative and linear age-B3

relationship, which suggests that younger plots tend to have larger B3 values than older plots.

In the past, prediction of forest age using optical remote sensing data has been considered a major challenge due to the limited spectral link to the attribute of interest, especially as forests get older (Cohen and Spies 1992, Maltman *et al.* 2023).

Similarly, forest age predictions that have used ALS data have been reported to be unsatisfactory due to the weakened height-age relationship after ~100 years (Wylie *et al.* 2019, Maltamo *et al.* 2020). For example, Wylie *et al.* (2019) found that there was a clear decrease in the relationship between the 50th and 90th ALS percentiles of vegetation height and mean tree age at ~120 years. With regard to the results presented in our study, the predictions for the older plots exhibit more variation than those for the younger plots (Fig. 5, SC3, GPBoost). Most of the variation occurs after 150 years of the observed plot age (Fig. 5, SC3, GPBoost). However, our predictions are generally in-line up to 250 years of the observed plot age. The issue with the diminished ALS height-age relationship is due to the reduced height growth of trees during later successional stages where tree age can vary significantly for similar tree heights because tree diameter continues to increase for a longer period than tree height. The specific age at which height growth ceases is mostly dependent on tree species and site characteristics.

The accuracy of the age predictions in this study could be improved if the ages of the hold-over plots were defined differently in the field. In our dataset, hold-over plot ages were defined from a forest management perspective. If the ages of these plots were defined from a biodiversity perspective (i.e. based on hold-over trees), age prediction using ALS data would yield improved prediction accuracies.

Discussion of old-growth forests is an active topic in the biodiversity arena. For example, the EU Biodiversity Strategy for 2030 is targeted at the protection of all remaining EU primary and old-growth forests (European Commission 2021). Old-growth forests are important habitats for several species and, therefore, age is a highly relevant attribute from a biodiversity point of view. Currently, detailed information on the current status and extent of old-growth forests is sparse (Hirschmugl *et al.* 2023), and accurate inventories and mapping of these ecosystems, particularly in countries with high forest-cover, require substantial resources. Remote sensing is one tool that can bridge this gap and offers a range of data that may assist in the mapping of old-growth forests and their properties (Sabatini *et al.* 2020).

Conclusion

In this study, tree-boosted mixed effects modelling (GPBoost) provided better prediction performance of plot-level age than linear mixed-effects modelling (LME) in each tested scenario. The categorical variables that describe site conditions (e.g. mineral soil, undrained/drained peatland and forest use restrictions) were found to be highly beneficial. Their inclusion as random-effects clearly decreased prediction errors and increased the range of predictions, which enabled better identification of old-growth forests. Metrics computed from ALS data were found to be more useful than Sentinel-2 metrics, which is in line with previous studies. The separate modelling of age in plots that contained hold-over trees improved prediction accuracies, especially when the observed age of the plot was close to zero.

Tree-boosted mixed effects modelling enables efficient use of indicator variables and hierarchical setups. There are many applications in the field of forest remote sensing where this type of approach may be beneficial. We consider that the error rate obtained in this study is sufficiently low to be useful in some real-world applications, and that a comparable error rate was not obtained earlier for Nordic countries.

Acknowledgement

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

Author contributions

Janne Toivonen (Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing—original draft, Writing—review & editing), Annika Susanna Kangas (Conceptualization, Funding acquisition, Supervision, Validation, Writing—review & editing), Timo P. Pitkänen (Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Software, Supervision, Writing—review & editing), Mari Myllymäki (Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Validation, Writing—review & editing), Matti Maltamo (Conceptualization, Funding acquisition, Supervision, Validation, Writing—review & editing), Mikko Kukkonen (Conceptualization, Funding acquisition, Methodology, Software, Supervision, Writing—review & editing), and Petteri Packalen (Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Resources, Software, Supervision, Validation, Writing—original draft, Writing—review & editing)

Supplementary data

Supplementary data are available at *Forestry* online.

Conflict of interest: The authors declare there are no competing interests.

Funding

This work was supported by Kone Foundation [202 201 370 to J.T.]; Research Council of Finland [337 127 to M. Maltamo, 337 655 to A.K., 355 267 to P.P.]; European Union—NextGenerationEU instrument [352 782 to P.P, M.K. and T.P.P.]; and European Union Horizon Europe (HORIZON) Research & Innovation programme [101 056 907 to M. Myllymäki].

Data availability

ALS data are available at: https://asiointi.maanmittauslaitos.fi/karttapaikka/tiedostopalvelu. However, free data are 0.5 pulses/m^2 and the original data used in this paper are 5.0 pulses/m^2. Sentinel-2 images (Level-2A) are freely available at Copernicus Browser: https://dataspace.copernicus.eu/browser. The Sentinel-2 mosaic used in this study is available at request from the authors of this paper. Finnish NFI plot data are not publicly available.

References

Barredo J, Brailescu C, Teller A. *et al. Mapping and Assessment of Primary and Old-Growth Forests in Europe, EUR 30661 EN*. Luxembourg. ISBN

978-92-76-34230-4: Publications Office of the European Union, 2021. https://data.europa.eu/doi/10.2760/797591.

Bates D, Mächler M, Bolker B. *et al.* Fitting linear mixed-effects models using lme4. *J Stat Softw* 2015;**67**:1–48 https://doi.org/10.18637/jss.v067.i01.

Breiman L, Friedman J, Olshen RA. *et al. Classification and regression trees*. 1st edition. New York, USA: Chapman and Hall/CRC, 1984. https://doi.org/10.1201/9781315139470.

Cohen WB, Spies TA. Estimating structural attributes of Douglas-fir/western hemlock forest stands from landsat and SPOT imagery. *Remote Sens Environ* 1992;**41**:1–17. https://doi.org/10.1016/0034-4257(92)90056-P.

Costanza R, d'Arge R, de Groot R. *et al.* The value of the world's ecosystem services and natural capital. *Ecol Econ* 1998;**25**:3–15. https://doi.org/10.1016/S0921-8009(98)00020-2.

CSC, IT Center for Science. https://csc.fi/en/ (assessed on 26 March 2025).

Dye M, Mutanga O, Ismail R. Combining spectral and textural remote sensing variables using random forests: predicting the age of Pinus patula forests in KwaZulu-Natal, South Africa. *J Spatial Sci* 2012;**57**:193–211. https://doi.org/10.1080/14498596.2012.733620.

Eckelt A, Müller J, Bense U. *et al.* "Primeval forest relict beetles" of Central Europe: a set of 168 umbrella species for the protection of primeval forest remnants. *J Insect Conservation* 2018;**22**:15–28. https://doi.org/10.1007/s10841-017-0028-6.

Eerikäinen K, Mabvurira D, Nshubemuki L. *et al.* A calibrateable site index model for Pinus kesiya plantations in southeastern Africa. *Can J For Res* 2002;**32**:1916–28. https://doi.org/10.1139/x02-106.

Eid T. Use of uncertain inventory data in forestry scenario models and consequential incorrect harvest decisions. *Silva Fennica (Helsinki, Finland: 1967)* 2000;**34**:89–100. https://doi.org/10.14214/sf.633.

European Commission: Directorate-general for environment. In: *EU Biodiversity Strategy for 2030: Bringing Nature Back into our Lives*. Luxembourg: Publications Office of the European Union, 2021. https://data.europa.eu/doi/10.2779/677548.

European Commission: Directorate-general for environment. In: *Commission Guidelines for Defining, Mapping, Monitoring and Strictly Protecting EU Primary and Old-Growth Forests*. Luxembourg: Publications Office of the European Union, 2023. https://data.europa.eu/doi/10.2779/481811.

van Ewijk KY, Treitz PM, Scott NA. Characterizing Forest succession in Central Ontario using lidar-derived indices. *Photogrammetric Eng Remote Sensing* 2011;**77**:261–9 https://doi.org/10.14358/PERS.77.3.261.

Falkowski MJ, Evans JS, Martinuzzi S. *et al.* Characterizing forest succession with lidar data: an evaluation for the inland northwest, USA. *Remote Sens Environ* 2009;**113**:946–56. https://doi.org/10.1016/j.rse.2009.01.003.

FAO. The state of the World's forests 2022 – Forest pathways for green recovery and building inclusive, resilient and sustainable economies. *Rome*. 2022;166. https://doi.org/10.4060/cb9360en.

Franklin JF, Johnson KN, Johnson DL. *Ecological Forest Management*. Long Grove, Illinois: Waveland Press Inc, 2018.

Gillis MD, Omule AY, Brierley T. Monitoring Canada's forests: the National Forest Inventory. *Forestry Chronicle* 2005;**81**:214–21. https://doi.org/10.5558/tfc81214-2.

Hallik L, Kuusk A, Lang M. *et al.* Reflectance properties of hemiboreal mixed forest canopies with focus on red edge and near infrared spectral regions. *Remote Sensing*. 2019;**11**:1717. https://doi.org/10.3390/rs11141717.

Hirschmugl M, Sobe C, Di Filippo A. *et al.* Review on the possibilities of mapping old-growth temperate forests by remote sensing in Europe. *Environ Modeling Assessment* 2023;**28**:761–85. https://doi.org/10.1007/s10666-023-09897-y.

Hunt ER, Rock BN. Detection of changes in leaf water content using near and middle-infrared reflectances. *Remote Sens Environ* 1989;**30**:43–54.

Hyvärinen E, Juslén A, Kemppainen E. *et al. The 2019 Red List of Finnish Species*. Helsinki: Ympäristöministeriö & Suomen ympäristökeskus, 2019, 704.

Jensen JR, Qiu F, Ji M. Predictive modelling of coniferous forest age using statistical and artificial neural network approaches applied to remote sensor data. *Int J Remote Sensing* 1999;**20**:2805–22. https://doi.org/10.1080/014311699211804.

Kandare K, Ørka HO, Dalponte M. *et al.* Individual tree crown approach for predicting site index in boreal forests using airborne laser scanning and hyperspectral data. *Int J Appl Earth Obser Geoinform* 2017;**60**:72–82. https://doi.org/10.1016/j.jag.2017.04.008.

Koivuniemi J, Korhonen KT. Inventory by compartments. In: Kangas A, Maltamo M, (eds.). *Forest Inventory. Methodology and Applications. Managing Forest Ecosystems*, Vol. **10**. Dordrecht: Springer, 2006, 271–8.

Korhonen KT, Ahola A, Heikkinen J. *et al.* Forests of Finland 2014–2018 and their development 1921–2018. *Silva Fennica. Helsinki: The Finnish Society of Forest Science* 2021;**55**:49. https://doi.org/10.14214/sf.10662.

Kouki J, Junninen K, Mäkelä K. *et al.* Metsät. In: Kontula T, Raunio A, (eds.) *Suomen Luontotyyppien Uhanalaisuus 2018. Luontotyyppien Punainen Kirja – Osa 1: Tulokset Ja Arvioinnin Perusteet. Helsinki: Suomen ympäristökeskus & ympäristöministeriö* 2018;171–201. http://urn.fi/URN:ISBN:978-952-11-4816-3.

LAStools. *Efficient LiDAR Processing Software* (v2.0.1). http://rapidlasso.com/LAStools (5 February 2024, date last assessed).

Luke - Luonnonvarakeskus (Natural Resources Institute Finland). *VMI13 Maastotyön Ohje* 2021;**2021**:113.

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;73.

Maltamo M, Kinnunen H, Kangas A. *et al.* Predicting stand age in managed forests using National Forest Inventory field data and airborne laser scanning. *Forest Ecosyst* 2020;**7**:1–11. https://doi.org/10.1186/s40663-020-00254-z.

Maltamo M, Packalén P, Suvanto A. *et al.* Combining ALS and NFI training data for forest management planning: a case study in Kuortane, western Finland. *Eur J For Res* 2009;**128**:305–17. https://doi.org/10.1007/s10342-009-0266-6.

Maltman JC, Hermosilla T, Wulder MA. *et al.* Estimating and mapping forest age across Canada's forested ecosystems. *Remote Sens Environ* 2023;**290**:113529. https://doi.org/10.1016/j.rse.2023.113529.

Myllymäki M, Tuominen S, Kuronen M. *et al.* The relationship between forest structure and naturalness in the Finnish national forest inventory. *Forestry* 2023;**97**:339–48. https://doi.org/10.1093/forestry/cpad053.

Ørka HO, Jutras-Perreault M-C, Næsset E. *et al.* A framework for a forest ecological base map – an example from Norway. *Ecol Indic* 2022;**136**:108636. https://doi.org/10.1016/j.ecolind.2022.108636.

Packalen P, Heinonen T, Pukkala T. *et al.* Dynamic treatment units in eucalyptus plantation. *Forest Sci* 2011;**57**:416–26.

Pan Y, Chen JM, Birdsey R. *et al.* Age structure and disturbance legacy of north american forests. *Biogeosciences* 2011;**8**:715–32. https://doi.org/10.5194/bg-8-715-2011.

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

R Core Team. R: *A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2024. Available online: https://www.R-project.org/.

Racine EB, Coops NC, St-Onge B. *et al.* Estimating forest stand age from LiDAR-derived predictors and nearest neighbor imputation. *Forest Sci* 2014;**60**:128–36. https://doi.org/10.5849/forsci.12-088.

Rautiainen M, Lukeš P, Homolová L. *et al.* Spectral properties of coniferous forests: A review of in situ and laboratory measurements. *Remote Sensing* 2018;**10**:207. https://doi.org/10.3390/rs10020207.

Reese H, Nilsson M, Pahlen T. *et al.* Countrywide estimates of forest variables using satellite data and field data from the national forest inventory. *Ambio* 2003;**32**:542–8. https://doi.org/10.1639/0044-7447(2003)032[0542:CEOFVU]2.0.CO;2.

Rogers BM, Mackey B, Shestakova TA. *et al.* Using ecosystem integrity to maximize climate mitigation and minimize risk in international forest policy. *Front Forests Global Change* 2022;**5**:30. https://doi.org/10.3389/ffgc.2022.929281.

Sabatini FM, Bluhm H, Kun Z. *et al.* European primary Forest database (EPFD) (v2.0bioRxiv). 2020. https://doi.org/10.6084/m9.figshare.13194095 (3 December 2025, date last assessed).

Schumacher J, Hauglin M, Astrup R. *et al.* Mapping forest age using National Forest Inventory, airborne laser scanning, and Sentinel-2 data. *Forest Ecosyst* 2020;**7**:1–14. https://doi.org/10.1186/s40663-020-00274-9.

Sigrist F. Gaussian process boosting. *J Mach Learn Res* 2022;**23**:1–46.

Sigrist F, Gyger T, Kuendig F. GPBoost: combining tree-boosting with gaussian process and mixed effects models. R package (version 1.3.0). 2021. https://github.com/fabsig/GPBoost (3 December 2025, date last assessed).

da Silva LP, Heleno RH, Costa JM. *et al.* Natural woodlands hold more diverse, abundant, and unique biota than novel anthropogenic forests: a multi-group assessment. *Eur J For Res* 2019;**138**:461–72. https://doi.org/10.1007/s10342-019-01183-5.

Straub C, Koch B. Enhancement of bioenergy estimations within forests using airborne laser scanning and multispectral line scanner data. *Biomass Bioenergy* 2011;**35**:3561–74. https://doi.org/10.1016/j.biombioe.2011.05.017.

SYKE – Suomen Ympäristökeskus (Finnish Environment Institute). *Corine Land Cover 2018*. Available online: https://ckan.ymparisto.fi/dataset/corine-maanpeite-2018 (3 December 2025, date last assessed).

Vastaranta M, Niemi M, Wulder MA. *et al.* Forest stand age classification using time series of photogrammetrically derived digital surface models. *Scandinavian J Forest Res* 2016;**31**:194–205. https://doi.org/10.1080/02827581.2015.1060256.

Voeten CC. Using 'buildmer' to automatically find & compare maximal (mixed) models. *CRAN* 2021. Available online: https://cran.r-project.org/web/packages/buildmer/vignettes/buildmer.html (3 December 2025, date last assessed).

Wylie RRM, Woods ME, Dech JP. Estimating stand age from airborne laser scanning data to improve models of black spruce wood density in the boreal Forest of Ontario. *Remote Sens* 2019;**11**:2022. https://doi.org/10.3390/rs11172022.

Zhang C, Ju W, Chen JM. *et al.* Mapping forest stand age in China using remotely sensed forest height and observation data. *J Geophys Res Biogeo* 2014;**119**:1163–79. https://doi.org/10.1002/2013JG002515.

Ziaco E, Di Filippo A, Alessandrini A. *et al.* Old-growth attributes in a network of Apennines (Italy) beech forests: disentangling the role of past human interferences and biogeoclimate. *Plant Biosystems* 2012;**146**:153–66. https://doi.org/10.1080/11263504.2011.650729.

Zupanc A. Improving Cloud Detection with Machine Learning. *Medium* 2017. Available online: https://medium.com/sentinel-hub/improving-cloud-detection-with-machine-learning-c09dc5d7cf13 (20 February 2024, date last assessed).

Forestry: An International Journal of Forest Research, 2025, 1–14
https://doi.org/10.1093/forestry/cpaf083
Original Article