This is an electronic reprint of the original article.
This reprint *may differ* from the original in pagination and typographic detail.

Author(s):   Matti Maltamo, Hermanni Kinnunen, Annika Kangas and Lauri Korhonen

Title: Predicting stand age in managed forests using National Forest Inventory field data and airborne laser scanning

Year: 2020

Version: Publisher's version

Please cite the original version:

RESEARCH

Predicting stand age in managed forests using National Forest Inventory field data and airborne laser scanning

Matti Maltamo[1*], Hermanni Kinnunen[1], Annika Kangas[2] and Lauri Korhonen[1]

Abstract

Background: The aim of this study was to construct a nationwide stand age model by using National Forest Inventory (NFI) data and nationwide airborne laser scanning (ALS) data. In plantation forestry, age is usually known. While this is not the case in boreal managed forests, age is still seldom predicted in forest management inventories. Measuring age accurately in situ is also very laborious. On the other hand, tree age is one of the accurately measured sample tree attributes in NFI field data. Many countries also have a nationwide coverage of airborne laser scanning (ALS) data. In this study, we merged these data sources and constructed a nationwide, area-based model for stand age.

Results: While constructing the model, we omitted old forests from the data, since the correlation between ALS height metrics and stand age diminished at stands with age > 100 years. Additionally, the effect of growth conditions was considerable, so we also utilized different geographical and NFI variables such as site fertility and soil type in the modeling. The resultant nationwide model for the stand age of managed forests yielded a root mean square error (RMSE) of about 14 years. The model could be improved further by additional forest structure variables, but such information may not be available in practice.

Conclusions: The results showed that the prediction of stand age by ALS, geographical and NFI information was challenging, but still possible with moderate success. This study is an example of the joint use of NFI and nationwide ALS data and re-use of NFI data in research.

Keywords: Forest stock age, LiDAR, NFI, Nationwide model, Growth conditions

Background

Stand age is a highly important attribute in even-aged forestry. It is required in the characterization of the development stage of a stand, as a predictor variable in some growth and yield models, and in site indexing with dominant height/age models (e.g. Eerikäinen et al. 2002; Kneeshaw and Gauthier 2003; Racine et al. 2014). In plantation forestry with short rotation times, stand age is usually known exactly (Packalén et al. 2011). However, in managed boreal forests, age is usually not known due to the long rotation times, natural regeneration of minor

tree species, and inadequate content of historical stand register data. In forest management inventories the estimation of stand age is laborious, since visual assessments are usually not accurate, and exact measurement requires costly and slow boring of a tree or even several trees per stand (Koivuniemi and Korhonen 2006). In addition, boring is usually not allowed without forest owner's permission. Due to these difficulties in tree age measurement, different approaches, methods and models that do not rely on stand age information have been developed for forest planning systems (Hynynen et al. 2002).

In some early references, tree age has been modelled as a function of other forest attributes such as tree diameter and site fertility (Korhonen 1987; Vähäsaari 1988). In those studies, tree age estimates were needed in the

* Correspondence: matti.maltamo@uef.fi
[1]University of Eastern Finland, School of Forest Sciences, P.O. Box 111, 80101 Joensuu, Finland
Full list of author information is available at the end of the article

modelling of taper curve and timber sortiments, respectively. Kalliovirta and Tokola (2005) modelled the age of individual sample trees based on height and maximum crown diameter with root mean square error (RMSE) of 9.2%–12.8%. These models were intended for use with high-resolution remotely sensed data. In later studies, stand age has been modelled directly as a function of airborne laser scanning (ALS) information. Maltamo et al. (2009b) included stand age as one of the predicted attributes in species specific multivariate forest attribute prediction by *k*-nearest neighbor (see Maltamo and Packalen 2014) and obtained an RMSE of about 20 years. Straub and Koch (2011) achieved a similar level of stand age RMSE using regression analysis in a study concentrating on estimation of bioenergy potential. Racine et al. (2014) studied different ALS based variables, including site attributes (elevation, slope, aspect, solar radiation, wetness index, catchment area, and flow path length) in stand age prediction in research forest area in Canada, obtaining an RMSE of about 10 years. Kinnunen (2018) constructed a local ALS model for stand age including also site fertility indices as predictor variables. The resultant RMSE was less than 10 years. Age has also been predicted based on a time series of remotely sensed data (Vastaranta et al. 2015), optical remote sensing information and inversed yield models (Frate et al. 2015) or indirectly by predicting stand development stages (Falkowski et al. 2009; Weber and Boss 2009; Kane et al. 2010).

Recently, remotely sensed data have been applied to obtain nationwide predictions of forest attributes. For example, Nilsson et al. (2017) utilized National Forest Inventory (NFI) field data and ALS data collected systematically over the country by the Swedish National Land Survey to obtain a nationwide forest attribute map for Sweden. Astrup et al. (2019) constructed a similar map product for Norway, employing photogrammetric point clouds instead of ALS data. Other studies merged data from several ALS inventory projects to construct large area prediction models for forest attributes such as volume, biomass and dominant height (Gopalakrishnan et al. 2015; Kotivuori et al. 2016, 2018). Tree age is one of the accurately measured sample tree attributes in NFI field data, but it has not been included so far in large scale ALS modelling studies.

The aim of this study was to construct a nationwide stand age model by a joint use of NFI and ALS data. We also utilize and analyze the use of different NFI and geographical variables in the prediction of stand age.

Materials and methods

National Forest Inventory field data

We used field sample plot data from the 11th NFI of Finland (2009–2013). Tree measurements were based on truncated angle count sample plots, where the maximum inclusion radius was 12.5 m. We only used plots that were fully located within one stand. Here the forest attributes calculated from tree measurements on a plot mimic a stand level. The stand age was estimated or each sample plot based on age core measurements taken from sample trees at breast height (VMI11 Maastotyöohje 2013). Every seventh sampled tree was measured as a sample tree. There are different definitions for stand age (Chirici et al. 2011) but since our tree measurements are based on angle count samples, this automatically led to the use of a basal area weighted stand age. Stand age is also commonly defined as the mean age at breast height weighted by the basal area (Sharma et al. 2011). Biological age could have been obtained by adding site and species-specific correction factors, but in this study, we focused on the breast height age. Some basic information concerning NFI plots applied in the study are given in Table 1.

We also used some categorical NFI variables, such as soil type, site fertility, land use, and forest structure, which are defined at stand level. The soil type was either mineral soil or peat land. Site fertility classes were estimated based on forest floor vegetation (Cajander 1926). The following classes were applied: Oxalis–Maianthemum (site1), Oxalis–Myrtillus (site2), Myrtillus (site3), Vaccinum (site4), Calluna (site5) and Cladina (including rocky and sandy sites) (site6). Land use class included productive and low productive forest lands. Forest structure was described by classes: natural stage, almost natural stage, or managed forest.

Airborne laser scanning data and selection of study areas

In Finland, ALS data are collected systematically for digital terrain mapping and operational ALS-based forest management inventories. The selection of the inventory areas was based on the publicly available ALS data acquisitions by the Finnish National Land Survey. The acquisitions are made in both leaf-on and leaf-off conditions, but we used only areas with leaf-on data. The data have low pulse densities, typically 0.5–1.5 per m^2, which is sufficient for the application of area-based forest inventories.

We selected randomly 1–2 inventory areas from each forest vegetation zone (Fig. 1, Table 2). Finland is divided into zones of vegetation based on similar soil,

Table 1 Characteristics of the 715 modeling plots

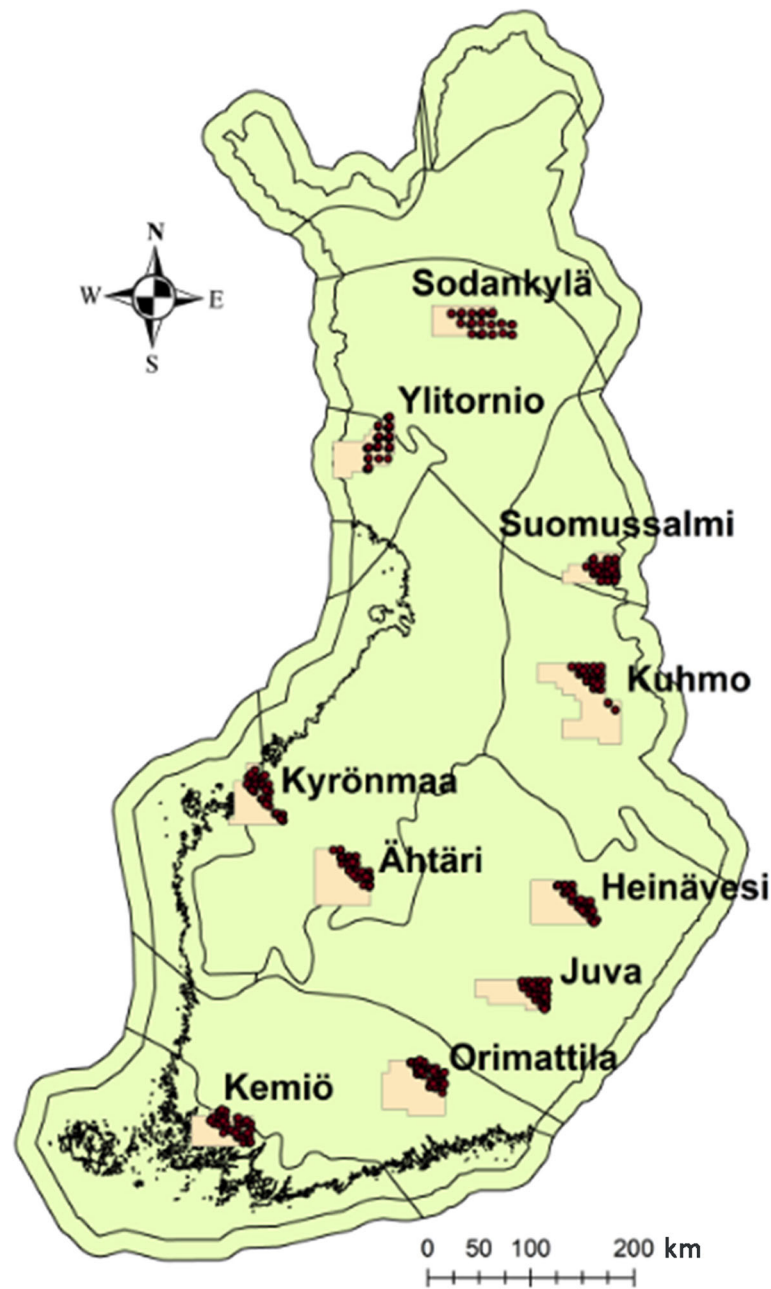|	Mean	Standard deviation	Minimum	Maximum
Degree days (C)	1058	173.9	707	1344
Elevation (m)	141	70.1	1	306
Basal area (m^2·ha^{-1})	16.4	8.5	0	49
Mean height (m)	13.1	5.7	2	30.9
Mean diameter (cm)	14.9	7.7	0	41
Stand age (years)	44.7	21.6	10	100

Fig. 1 A map of the forest vegetation zones in Finland, and the locations of the ALS based forest inventory areas and NFI field plots employed in this study

topographic and climate conditions (Kalliola 1973). These data were available from the website of the Finnish Environmental Institute (Forest vegetation zones/© Finnish Environment Institute (SYKE) 2015). Two ALS data acquisition areas were chosen from the largest vegetation zones, and one from the others. The southernmost and northernmost vegetation zones were omitted from the analyses, because the structure of forests deviated considerably from the rest of the country. Thus altogether ten areas were chosen (Table 2, Fig. 1).

For each area, 100 NFI plots were selected from its north-eastern part based on the sum of their Cartesian coordinates. The plots were not selected randomly, because we wanted to avoid downloading and processing unnecessarily large amounts of ALS data.

The ALS data sets were downloaded in laz-format from a public database (National Land Survey of Finland 2018) and processed using the LASTools software (Isenburg 2018). The lasclip tool was first employed to intersect the ALS data within a 12.5-m radius around each NFI plot.

Table 2 Summary of the airborne laser scanning data available at each site

Area	Year	Scanner	Point density (m^{-2})	Height accuracy (m)	Flight altitude (m)
Ähtäri	2013	Optech ALTM GEMINI	0.7	0.15	1730
Heinävesi	2017	Riegl LM-7800	1.5	0.15	1870
Juva	2016	Leica ALS 60	0.88	0.15	2400
Kemiö	2017	Leica ALS 60	0.7	0.15	2000
Kuhmo	2014	Leica ALS 60	1.33	0.15	2150
Kyrönmaa	2016	Leica ALS 80 HP	1.55	0.15	2000
Orimattila	2016	Leica ALS 60	0.5	0.15	2500
Sodankylä	2015	Optech ALTM_Gemini	0.7	0.15	1890
Suomussalmi	2017	Riegl LM-7800	1.5	0.15	1700
Ylitornio	2015	Optech ALTM_Gemini	0.8	0.15	1600

Orthometric heights of the original ALS data sets were normalized into above ground heights by employing the lasground tool, which utilized an existing echo classification available in the laz files. Finally, a set of ALS-based canopy height distribution metrics was computed for each plot. These included the mean, maximum, and standard deviation of heights, height percentiles, canopy density percentiles or "bincentiles", and a canopy cover metric. Each variable was computed separately using 1) single and first-of many echoes, and 2) using single and last-of-many echoes (Næsset 2002). A height threshold of 2.0 m was used in the computation of the canopy cover metric, while other metrics were computed without a height threshold. The XY-coordinates and elevations of the plots, as well as the scanner type, were also saved for further use in the modelling.

Final modelling datasets and variables

The temporal differences between field and ALS data acquisitions were corrected plot-by-plot by adding the respective time difference to the stand age. Some silvicultural actions had also been conducted between the data collections, which could have considerable effects on relationship between the stand age and ALS metrics. In order to avoid such problems, all plots where the difference between the field-based forest stock height and the 90th canopy height percentile (p90) (see Table 3) was smaller than − 5 m or greater than 7 m were removed from the data.

Furthermore, we observed that there was no correlation between ALS metrics and stand age in old stands. Since most of the forests in Finland are managed with rotation ages < 100 years, plots with stand age > 100 years are mainly found at conservation areas in the northern part of the country. The reason for reduced correlation between ALS metrics and stand age in old-growth forests is obvious: tree age will increase linearly also when its height growth has already ceased. The greatest stand age values in our original data were > 300 years. Thus,

we decided to restrict our analysis to managed forests where the greatest stand age at breast height was 100 years. We also removed the youngest (stand age below 10 years) seedling stand sample plots. After these modifications our data consisted of 715 plots (Table 1). A majority of the plots (91.6%) were located at productive forest land, and at mineral soils (69%). Correspondingly, the proportions of site fertility classes were 3.4%, 17.8%,

Table 3 Candidate predictor variable. The ALS height distribution variables are supplemented with a prefix f or l, which means that the variable was computed from first or last echoes, respectively

Type	Variable	Abbreviation
ALS-A	height percentiles	p05, p10, …, p90, p95
	canopy density percentiles	b05, b10, …, b90, b95
	canopy cover	cov
	mean height maximum height standard deviation of heights	avg max stdev
	scanner type	Leica60, Leica80, Optech, Riegl
ALS-B	inventory area	Ähtäri, Heinävesi, Juva, Kemiö, Kuhmo, Kyrönmaa, Orimattila, Sodankylä, Suomussalmi
Geographic	major region	other, Lapland
	latitude	
	longitude	
	elevation	
	degree days	dd
	slope	
NFI-A	site fertility class	site1, … site5
	land use class	productive, low
	soil type	mineral, peat
NFI-B	forest structure	natural, almost natural, managed
	mean diameter	meanD

39.2%, 30.1%, 8.3% and 1.4% for Oxalis–Maianthemum, Oxalis–Myrtillus, Myrtillus, Vaccinum, Calluna and Cladina types, respectively.

All candidate predictor variables are presented in Table 3. In addition to ALS height distribution variables (ALS-A), we also included geographic variables and NFI variables as candidate predictors in the modelling. Geographic, ALS-A and some of the NFI variables (NFI-A in Table 3) would typically be available in a practical modelling case. Scan area ID (ALS-B) is applicable only for the considered inventory areas. Some of the NFI variables (NFI-B in Table 3) are only defined for the NFI field plots and they cannot be assumed to known practical applications databases. Their role in the current study is exploratory, i.e. they are used to describe the relationship between age and stand structure.

The ALS-A variables include the area based canopy height distribution metrics described in the previous section, and a categorical variable describing ALS sensor manufacturer. Geographic variables included plot latitude and longitude, elevation and slope computed from ALS-based digital terrain models, and the number of degree days that is a simplified description of climate (obtained in raster format from the Natural Resources Institute of Finland). Another geographic variable was the major region, which was obtained by separating the province of Lapland (Sodankylä and Ylitornio) from the other parts of the country (the rest of the inventory areas). The NFI-A variables included soil type, site fertility, and land use class. The NFI-B variables included mean diameter of the NFI plot and a categorical forest structure variable.

Statistical analysis

The nationwide stand age model was constructed using linear regression analysis with the NFI based age as the dependent variable. We constructed two separate models for stand age. First model included the ALS-A metrics, geographic variables, and the NFI-A variables. The second model included also the ALS-B and NFI-B variables. The role of the second model was to test a scenario where more factors affecting stand age prediction are known. Various interactions between ALS metrics and class level information were also tested in both models. All categorical variables were included as dummy variables. Different modifications (natural logarithm, square root, inverse, second power) of both independent and dependent variables were also tested.

Linear regression models were constructed using the lm function in R (R Core Team 2015). Due to the heteroscedasticity we experimented with transformations to dependent variable. The independent variables were selected by means of the stepwise function (step) which applies the AIC statistic to include or exclude non-significant predictors from the models. Both inclusion and exclusion of the independent variables in the candidate models were allowed. The independent variables had to be statistically significant ($p < 0.05$) to be selected for the model.

Accuracy assessment was performed through a cross validation procedure in which either each plot (leave-one-plot-out cross validation) or all plots within an inventory area (leave-one-inventory area-out cross validation) were excluded from the training data, and their values were predicted with a model refitted with the remaining plots. The residuals of the models were also analyzed to evaluate the heteroscedasticity.

The accuracy of the cross validated predictions was evaluated in terms of root mean squared error:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(t-\hat{t}\right)^2}{n}} \qquad (1)$$

where n is the number of plots, t is the observed stand age for plot i, and $\hat{t}$ is the predicted stand age for plot i. Subsequently, the RMSE% was calculated by dividing the RMSE by the predicted mean of the stand age.

Results

The relationship between ALS metrics and stand age was not very strong in the whole data (Fig. 2). It can be seen from Fig. 2 that the relationships are clearly different at mineral soils and peat lands. Thus categorical site variables are required in any realistic stand age model.

Due to the heteroscedasticity the response variable was a logarithmic transformation of stand age in both models. The first model included altogether five ALS metrics, dummy variables for scanner manufacturer Leica and mineral soils, as well as interactions between ALS metric l_max and large geographical area other, mineral soil, and low productivity forests (Table 4). Additionally, there were interactions between ALS variable l_avg and site fertility classes. The residual standard error of the model was 0.29 (14 years when back-transformed). The R^2 value was 0.71.

The logarithmic residuals of the first model (Table 4) are presented in Fig. 3. It can be seen that the residuals appear homogeneous, although the model overestimated the age of some younger stands. On the other hand, back-transformed stand age estimates (Fig. 4) showed that there were rather large variations in the estimates for greater stand ages. The plot level cross-validated RMSE of the first model was 32.9% and inventory area level cross-validated RMSE was 33.9%. Thus, the increase in RMSE% was modest, since the relative RMSE of the original model was 31.5%.
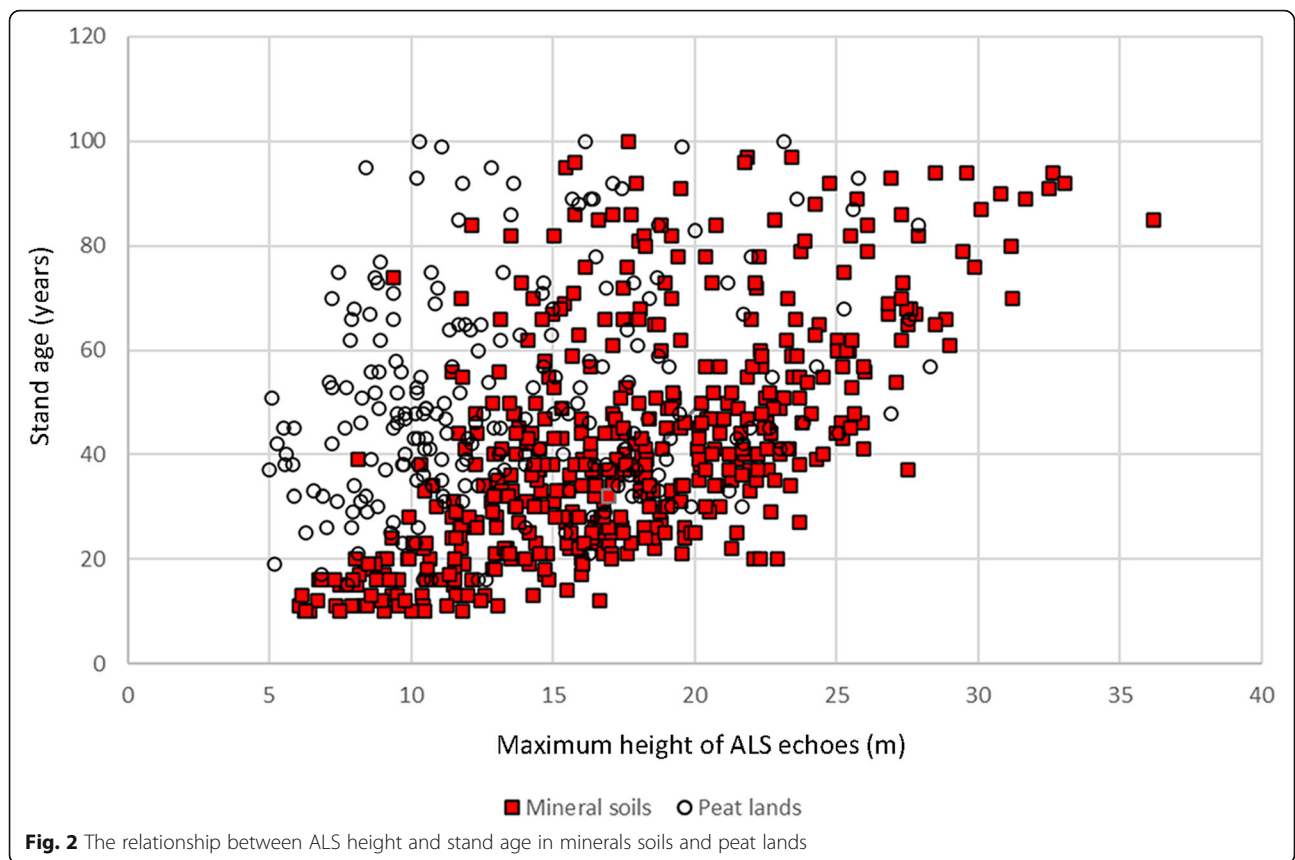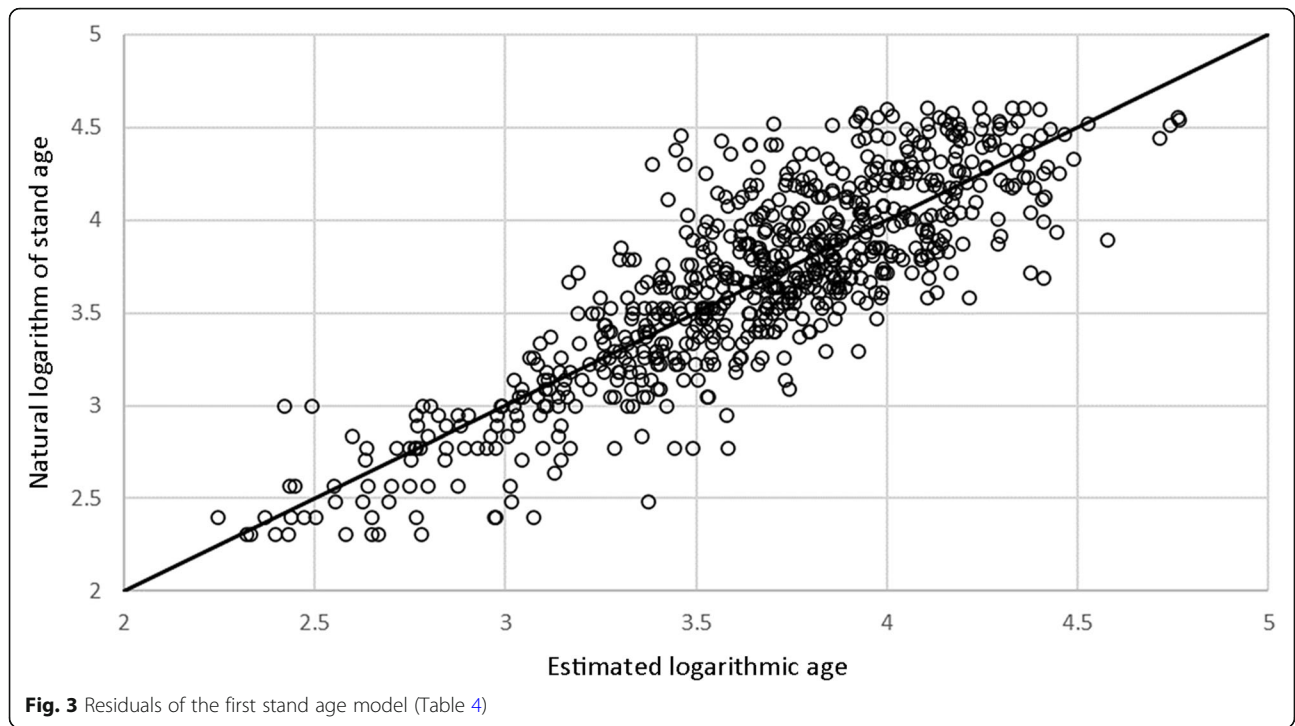
Fig. 2 The relationship between ALS height and stand age in minerals soils and peat lands

Table 4 Independent variables of the first stand age model. Sqrt refers to square root

	Estimate	Standard error	*t*-value	*P*-value
Intercept	2.331	0.133	17.570	< 2e−16
f_p90	0.058	0.008	7.741	3.45e−14
f_b05	0.030	0.005	5.566	3.71e−08
f_b10	−0.023	0.005	−4.253	2.40e−05
l_avg	0.426	0.066	6.483	1.70e−10
sqrt (l_max)	0.226	0.047	4.763	2.32e−06
Leica60	−0.193	0.025	−7.719	4.07e−14
mineral	−1.244	0.141	−8.803	< 2e−16
sqrt (l_max × other)	−0.129	0.008	−16.353	< 2e−16
sqrt (l_max_ × mineral)	0.240	0.037	6.550	1.12e−10
sqrt (l_max_ × low)	0.115	0.018	6.430	2.36e−10
l_avg × site1	−0.412	0.066	−6.247	7.29e−10
l_avg × site2	−0.431	0.065	−6.633	6.59e−11
l_avg × site3	−0.404	0.065	−6.230	8.04e−10
l_avg × site4	−0.361	0.065	−5.546	4.15e−08
l_avg × site5	−0.321	0.069	−4.643	4.11e−06

The second model included also the estimated forest structure classification and mean diameter as candidate predictors (Table 5). There were two ALS metrics and altogether 12 interactions between ALS metrics and geographic or NFI variables, including two inventory areas, land use class low productive forests, latitude, and field-measured mean diameter. Categorical variables for mineral soils and almost natural forest structure were also included. Compared to the first model, the residual standard error decreased to 0.24 (12.2 years when back-transformed), and the R^2 increased to 0.80. The residuals of the model showed a slightly better fit than the first model. Corresponding cross-validation showed 28.7% RMSE for plot level and 31.2% RMSE for inventory area level, while the original RMSE was 27.9%. (Figs. 5 and 6).

Discussion

Stand age is rarely predicted by remote sensing. In boreal forests comprehensive field data acquisitions of accurate stand age are time consuming and expensive. However, the use of NFI sample plot data offers possibilities for such studies for large areas. Accurate forest stock description is a crucial demand for NFIs, and this kind of information can be used with wall-to-wall ALS data to obtain large area predictions.

Fig. 3 Residuals of the first stand age model (Table 4)

Our aim was to construct a nationwide stand age model based on a fusion of NFI plots and ALS metrics. However, when constructing the model, we noticed that old forests could not be included in the model, since the correlation between ALS height metrics and stand age diminished after a threshold age. Thus we restricted our

model to managed stands where the stand age at breast height was at maximum 100 years. Additionally, the NFI data consisted of a comprehensive selection of forests, including e.g. mineral soils and peat lands where growth conditions are very different, and similar sized tree stocks can have large age variations. Also the rather low
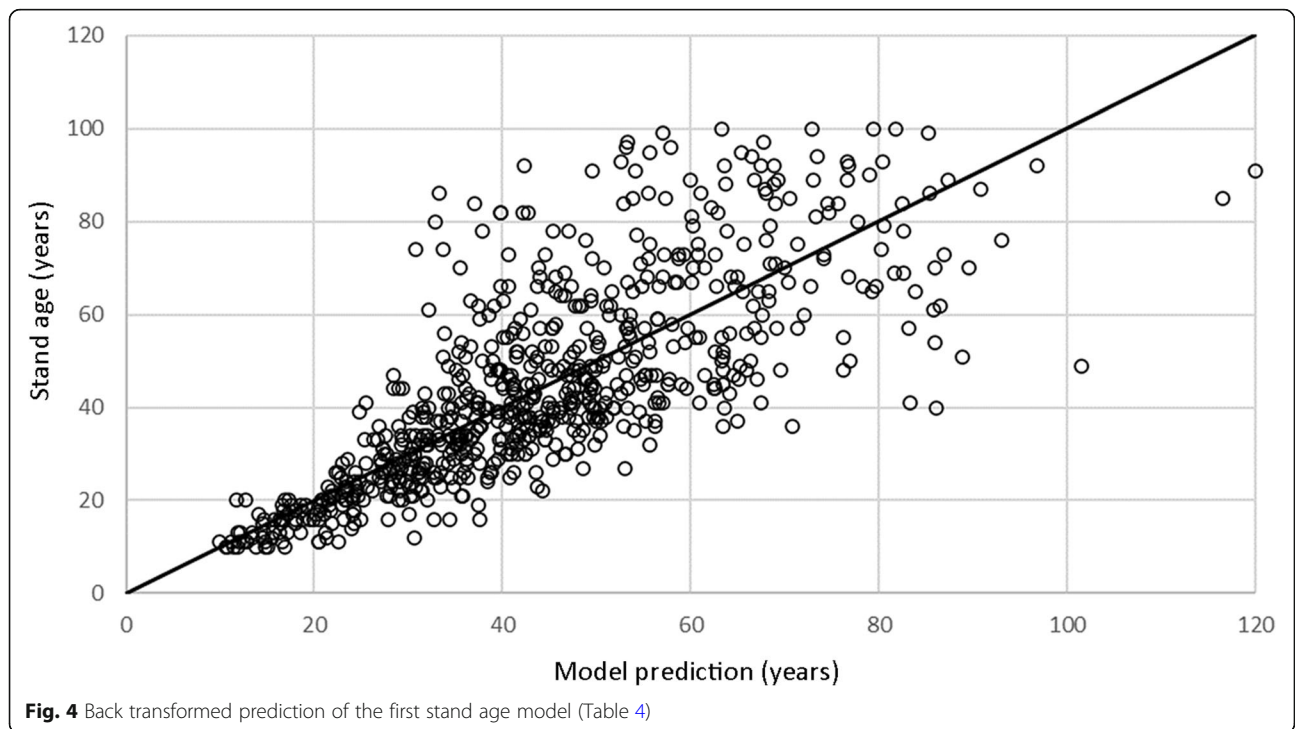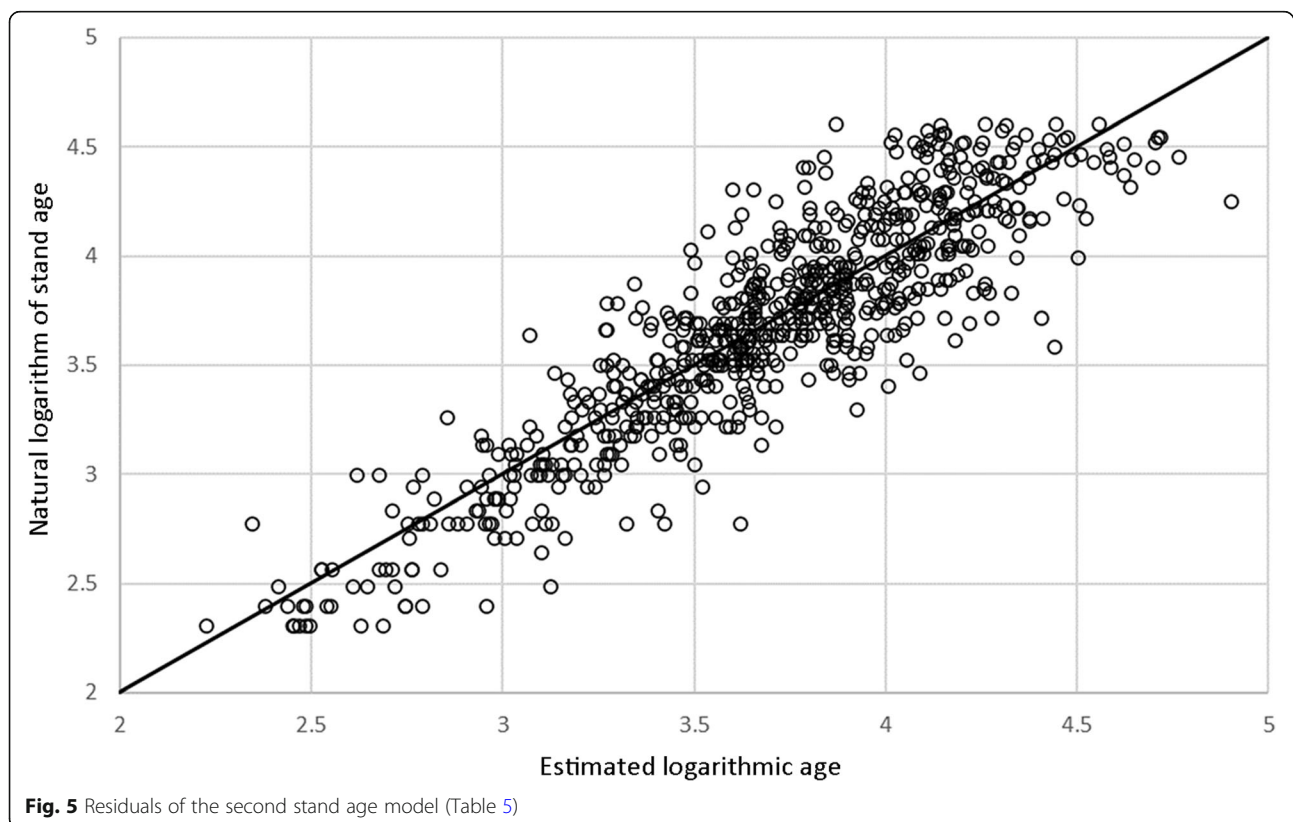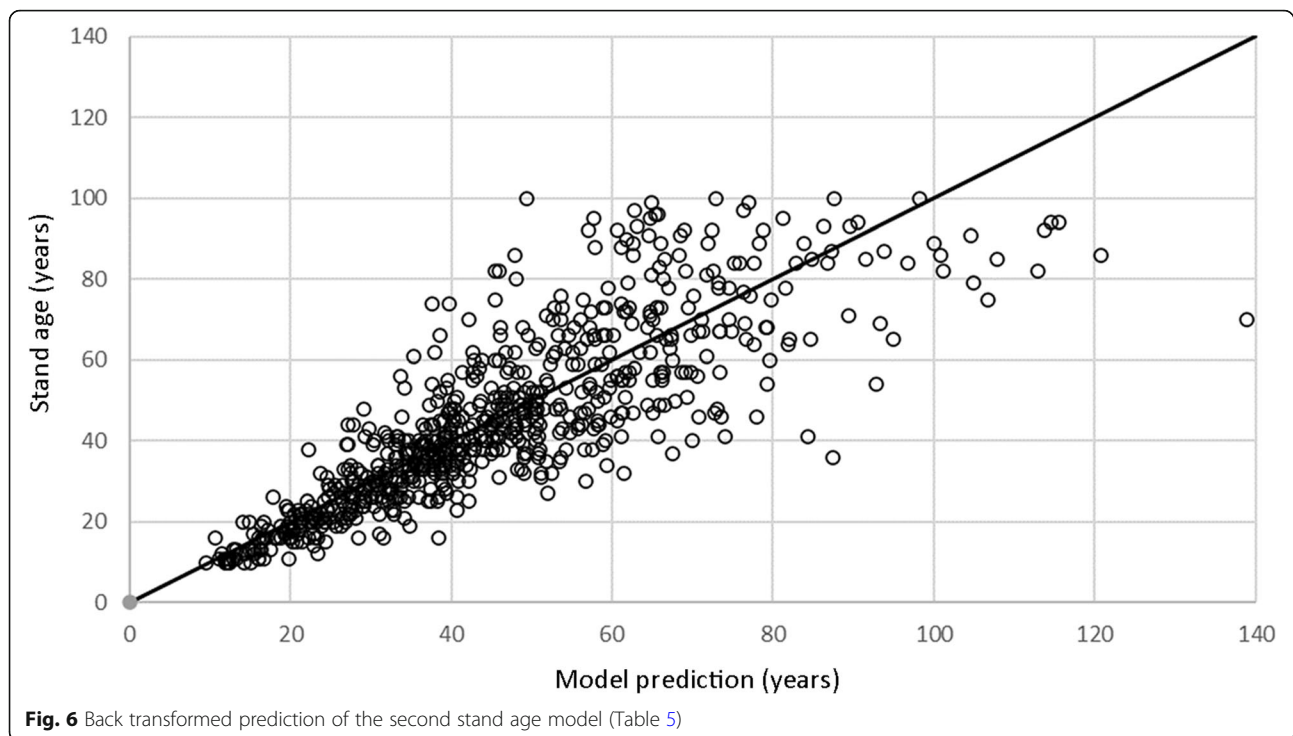

Fig. 4 Back transformed prediction of the first stand age model (Table 4)

Table 5 Independent variables of the second stand age model

	Estimate	Standard error	*t*-value	*P*-value
Intercept	2.444	0.076	32.091	< 2e−16
f_b05	0.008	0.001	9.898	< 2e−16
l_b20	0.004	0.001	4.456	9.70e−06
mineral	−0.977	0.109	−8.937	< 2e−16
sqrt (l_max × north)	0.306	0.044	6.963	7.68e−12
sqrt (l_max × mineral)	0.170	0.028	6.124	1.52e−09
sqrt (l_max × low)	0.303	0.019	16.228	< 2e−16
l_max × site1	−0.018	0.003	−6.218	8.67e−10
l_max × site2	−0.019	0.002	−11.508	< 2e−16
l_max × site3	−0.009	0.001	−6.400	2.85e−10
f_b95 × Lapland	−0.012	0.002	−7.083	3.45e−12
l_avg × lat	1.37e−08	1.10e−09	12.460	< 2e−16
f_b05 × other	−0.006	0.001	−7.843	1.65e−14
l_b70 × meanD	0.0007	2.94e−05	22.907	< 2e−16
almost natural	0.222	0.038	5.787	1.08e−08
sqrt (l_max × Juva)	−0.035	0.007	−4.836	1.63e−06
f_b05 × Kuhmo	−0.004	0.001	−5.334	1.30e−07

number of age measurements in a NFI plot may increase the uncertainty of age prediction. This resulted in difficulties when predicting stand age with ALS metrics, because ALS data and geographic variables alone cannot describe this variation. Therefore, we included several NFI variables into the variable selection.

The resultant model (Table 4) included ALS height percentiles describing the height of top layers of the canopy, and density metrics computed for the bottom layers of the canopy (i.e. the fraction of echoes above the crown base height). The categorical variable separating peatlands and mineral soils was significant in the model, and was also included as an interaction term with ALS metrics. Additionally, interactions of ALS metrics and site fertility classes were significant predictors. Also large scale geographical areas, low productivity forests and one laser scanner manufacturer were significant categorical predictors. Information on mineral soils and site fertility also behaved logically in the model, although the interpretation of a multiplicative model is not unambiguous. In applications it can be expected that all this information is available. For example, site fertility has been comprehensively assessed in stand management during previous decades (Koivuniemi and Korhonen 2006), while the soil type and land use information can be found from thematic maps.

Fig. 5 Residuals of the second stand age model (Table 5)

Fig. 6 Back transformed prediction of the second stand age model (Table 5)

We had also other types of information available. However, variables describing elevation, slope, degree days or geographical coordinates were not statistically significant in the model. While there were several geographic variables available, only the large scale geographical region was included in the model. Also only one categorical variable related to the laser scanner was included in the model. This agrees with previous studies that have indicated that data from different laser scanners can be combined in the models (Kotivuori et al. 2018). Regarding location information, the results from previous large-scale studies vary. For instance, Maltamo et al. (2016) used elevation, latitude and degree days in their biomass prediction study in Norway, where the geographic variations, for example in elevation, are considerably larger.

The observed RMSE of about 14 years is in between of those obtained in earlier studies with considerably smaller study areas (Maltamo et al. 2009b; Racine et al. 2014). Direct comparison with earlier studies is problematic since those studies have applied only remotely sensed predictor variables. In addition, Maltamo et al. (2009b) had stand age as one of the independent variables in a multivariate forest attribute model, but did not pay any specific attention to it when selecting predictor variables. On the other hand, Racine et al. (2004) presented a thorough analysis of the effects of different ALS based variables on age prediction, but the study was conducted on a rather small area. The same is true also for the study of Kinnunen (2018), while the current

study aimed to construct a large-scale age prediction model.

The residuals plots of our model were satisfactory but not perfect for young and old observations. Our second model (Table 5) demonstrated that an inclusion of mean diameter improved both RMSEs and residual plots. This means that the tree diameter distribution contains additional age information (RMSE% improved 13%) that is lacking from the ALS metrics. This can also be seen from the inaccuracies observed in ALS based diameter prediction studies (Maltamo et al. 2009a, 2009b; Sumnall et al. 2016). Similar to age, the tree diameter also increases after the tree height growth has ceased.

Our study indicated that stand age can be predicted with ALS metrics with moderate success. At least the accuracy was better than in a previous study in Finland (Maltamo et al. 2009b). The relationship between age and ALS metrics is not as strong as in the case of volume or biomass (e.g. Næsset 2002; Lim et al. 2003). Particularly the effect of growth conditions on stand age - ALS height relationship is considerable. Our approach was similar to studies concerning ALS-based estimation of volume and biomass: a total amount of the attribute is predicted by variables describing the height and density of the forest canopy. In the case of age, the use of multitemporal lidar data might improve the predictions, since it would allow the use of metrics describing height changes. This would also lead to the use of age predictions in site indexing (Noordermeer et al. 2018).

Conclusion

Prediction of stand age by ALS and class level field information is challenging but still possible with moderate success. Especially the role of information describing growth conditions of the forest stock is important. Here the national level model for age of the managed forests led to an RMSE of about 14 years. The usability of such age predictions varies according to applications. It may be inadequate for site indexing, but enough for growth and yield prediction. This study is an example of the joint use of NFI and National Land Survey ALS data and especially of re-use of NFI data in research.

Abbreviations
ALS: Airborne laser scanning; cov: Canopy cover; b05, b10, …, b90, b95: Canopy density percentiles; dd: Degree days; f: First echoes; p05, p10, …, p90, p95: Height percentiles; l: Last echoes; max: Maximum height; meanD: Mean diameter; avg.: Mean height; NFI: National Forest Inventory; RMSE: Root mean square error; sqrt: Square root; stdev: Standard deviation of heights

Authors' contributions
Matti Maltamo planned the study, participated to modelling and wrote the manuscript. Hermanni Kinnunen did the data processing, calculation and modelling. Lauri Korhonen participated to planning of the study, processing of the data and modelling. He also wrote and commented the manuscript. Annika Kangas participated to planning of the study and modelling. She also commented the manuscript. The authors read and approved the final manuscript.

Availability of data and materials
All airborne laser scanning data were made publicly available by the national land survey of Finland. Some of the data were not directly available in the public file service but required an access to a separate directory user interface. See https://www.maanmittauslaitos.fi/en/maps-and-spatial-data/expert-users/topographic-data-and-how-acquire-it for more information. National Forest Inventory data can be requested from Annika Kangas (annika.kangas@luke.fi), Natural Resources Institute Finland.

Ethics approval and consent to participate
This research is not considered to involve ethical problems.

Consent for publication
Not applicable.

Competing interests
There are no competing interests.

Author details
[1]University of Eastern Finland, School of Forest Sciences, P.O. Box 111, 80101 Joensuu, Finland. [2]Natural Resources Institute Finland, Bioeconomy and Environment, P.O. Box 68, 80100 Joensuu, Finland.

References
Astrup R, Rahlf J, Bjørkelo K, Debella-Gilo M, Gjertsen AK, Breidenbach J (2019) Forest information at multiple scales: development, evaluation and application of the Norwegian forest resources map SR16. Scand J Forest Res 34:6

Cajander AK (1926) The theory of forests types. Acta For Fenn 29(3):7193

Chirici G, Winter S, McRoberts RE (2011) National Forest Inventories: contributions to forest biodiversity assessments. Springer, Dordrecht, Heidelberg, London, New York

Eerikäinen K, Mabvurira D, Nshubemuki L, Saramäki J (2002) A calibrateable site index model for *Pinus kesiya* plantations in southeastern Africa. Can J For Res 32:1916–1928

Falkowski MJ, Evans JS, Martinuzzi S, Gessler PE, Hudak AT (2009) Characterizing forest succession with lidar data: an evaluation for the inland northwest, USA. Remote Sens Environ 113:946–956

Finnish Environment Institute (SYKE) (2015) Finnish environment institute (SYKE) forest vegetation zones http://www.syke.fi/en-US/Open_information/Spatial_datasets. Accessed 13 Nov 2018

Frate L, Carranza M, Garfì V, Febbraro M, Tonti D, Marchetti M, Ottaviano M, Santopuoli G, Ghirici G (2015) Spatially explicit estimation of forest age by integrating remotely sensed data and inverse yield modeling techniques. iForest 9:63–71

Gopalakrishnan R, Thomas VA, Coulston JW, Wynne RH (2015) Prediction of canopy heights over a large region using heterogeneous lidar datasets: efficacy and challenges. Remote Sens 7:11036–11060

Hynynen J, Ojansuu R, Hökkä H, Siipilehto J, Salminen H, Haapala P (2002) Models for predicting stand development in MELA system. Metsäntutkimuslaitoksen Tiedoantoja, the Finnish Forest research institute, research papers 835. Vantaa Research Center, Vantaa, p 116

Isenburg M (2018) LAStools - efficient LiDAR processing software (version 27092018, academic) http://rapidlasso.com/LAStools. Accessed 23 Nov 2019

Kalliola R (1973) Suomen kasvimaantiede. Weiner Söderström Osakeyhtiön kirjapaino, Porvoo

Kalliovirta J, Tokola T (2005) Functions for estimating stem diameter and tree age using tree height, crown width and existing stand database information. Silva Fenn 39:227–248

Kane VR, Bakker JD, McGaughey RJ, Lutz JA, Gersonde RF, Franklin JF (2010) Examining conifer canopy structural complexity across forest ages and elevations with LiDAR data. Can J For Res 40:774–787

Kinnunen VH (2018) Puuston iän ennustaminen laserkeilaustunnuksilla. Bachelor thesis, University of Eastern Finland

Kneeshaw D, Gauthier S (2003) Old growth in the boreal forest: a dynamic perspective at the stand and landscape level. Environ Rev 11(Sup 1): S99–S114

Koivuniemi J, Korhonen KT (2006) Inventory by compartments. In: Kangas A, Maltamo M (eds) Forest inventory. Methodology and applications. Managing forest ecosystems, vol 10. Springer, Dordrecht, pp 271–278

Korhonen KT (1987) Kuusen runkokäyrämalli lineaarisella sekamallimenetelmällä. Master thesis, University of Joensuu

Kotivuori E, Korhonen L, Packalen P (2016) Nationwide airborne laser scanning based models for volume, biomass and dominant height in Finland. Silva Fenn 50(4):1567. https://doi.org/10.14124/sf.1567

Kotivuori E, Maltamo M, Korhonen L, Packalen P (2018) Calibration of nationwide airborne laser scanning based stem volume models. Remote Sens Environ 210:179–192

Lim K, Treitz P, Baldwin K, Morrison I, Green J (2003) Lidar remote sensing of biophysical properties of tolerant northern hardwood forests. Can J Remote Sens 29(5):658–678

Maltamo M, Bollandsås OM, Gobakken T, Næsset E (2016) Large-scale prediction of aboveground biomass in heterogeneous mountain forests by means of airborne laser scanning. Can J For Res 46:1138–1144

Maltamo M, Næsset E, Bollandsås OM, Gobakken T, Packalén P (2009a) Non-parametric estimation of diameter distributions by using ALS data. Scand J Forest Res 24:541–553

Maltamo M, Packalen P (2014) Species specific management inventory in Finland. In: Maltamo M, Naesset E, Vauhkonen J (eds) Forestry applications of airborne laser scanning –concepts and case studies. Managing Forest ecosystems, vol 27. Springer, Dordrecht, pp 241–252

Maltamo M, Packalén P, Suvanto A, Korhonen KT, Mehtätalo L, Hyvönen P (2009b) Combining ALS and NFI training data for forest management planning -a case study in Kuortane, Western Finland. Eur J For Res 128:305–317

Næsset E (2002) Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sens Environ 80:88–99

National Land Survey of Finland (2018) Stereo model classified airborne laser scanning data from the National Land Survey of Finland topographic

database 10/2018 https://tiedostopalvelu.maanmittauslaitos.fi/tp/kartta?lang=
    en. Accessed 23 Nov 2019

Nilsson M, Nordkvist K, Jonzén J, Lindgren N, Axensten P, Wallerman J, Egberth M,
    Larsson S, Nilsson L, Eriksson J, Olsson H (2017) A nationwide forest attribute
    map of Sweden predicted using airborne laser scanning data and field data
    from the national forest inventory. Remote Sens Environ 194:447–454

Noordermeer L, Bollandsås OM, Gobakken T, Næsset E (2018) Direct and indirect
    site index determination for Norway spruce and scots pine using bitemporal
    airborne laser scanner data. For Ecol Manag 428:104–114

Packalén P, Heinonen T, Pukkala T, Vauhkonen J, Maltamo M (2011) Dynamic
    treatment units in eucalyptus plantation. For Sci 57:416–426

R Core Team (2015) R: a language and environment for statistical computing. R
    Foundation for Statistical Computing, Vienna https://www.R-project.org/.
    Accessed 23 Nov 2019

Racine EB, Coops NC, St-Onge B, Bégin J (2014) Estimating forest stand age from
    LiDAR-derived predictors and nearest neighbor imputation. For Sci 60:128–136

Sharma RP, Brunner A, Eid T, Øyen B-H (2011) Modelling dominant height
    growth from National Forest Inventory individual tree data with short time
    series and large age errors. For Ecol Manag 262:2162–2175

Straub C, Koch B (2011) Enhancement of bioenergy estimations within forests
    using airborne laser scanning and multispectral line scanner data. Biomass
    Bioenergy 35:3561–3574

Sumnall M, Hill RA, Hinsley SA (2016) Comparison of small-footprint discrete
    return and full waveform airborne lidar data for estimating multiple forest
    variables. Remote Sens Environ 173:214–223

Vähäsaari H (1988) Puutavaralajirakenteen arvioiminen eri mittausmenetelmillä.
    Master thesis, University of Joensuu

Vastaranta M, Niemi M, Wulder M, White J, Nurminen K, Litkey P, Honkavaara E,
    Holopainen M, Hyyppä J (2015) Forest stand age classification using time
    series of photogrammetrically derived digital surface models. Scand J Forest
    Res 31:194–205

VMI11 Maastotyöohje (2013) Metsäntutkimuslaitos

Weber TC, Boss DE (2009) Use of LiDAR and supplemental data to estimate forest
    maturity in Charles County, MD, USA. For Ecol Manag 258:2068–2075