Forest Ecology and Management 547 (2023) 121395

Contents lists available at ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco

Mapping site index in coniferous forests using bi-temporal airborne laser scanning data and field data from the Swedish national forest inventory

Alex Appiah Mensah *, Jonas Jonzén, Kenneth Nyström, Jörgen Wallerman, Mats Nilsson

Department of Forest Resource Management, Swedish University of Agricultural Sciences, SE-901 83, Umeå, Sweden

ARTICLE INFO

Keywords:
Regression analysis
Forest site productivity
National forest inventory
Airborne laser scanning
Scots pine
Norway spruce

ABSTRACT

Recent advancements in remote sensing of forests have demonstrated the capabilities of three-dimensional data acquired by airborne laser scanning (ALS) and, consequently, have become an integral part of enhanced forest inventories in Northern Europe. In Sweden, the first national laser scanning revolutionised forest management planning through low-cost production of large-scale and spatially explicit maps of forest attributes such as basal area, volume, and biomass, compared to the earlier practice based on field survey data. A second scanning at the national level was launched in 2019, and it provides conditions for the estimation of height growth and site index. Accurate and up-to-date information about site productivity is relevant for planning silvicultural treatments and for the prognosis of forest status and development over time. In this study, we explored the potential of bi-temporal ALS data and other auxiliary information to predict and map site productivity by site index according to site properties (SIS) of Norway spruce (*Picea abies* (L.) Karst) and Scots pine (*Pinus sylvestris* L.) in even-aged stands in Sweden. We linked ground survey data of SIS from more than 11,500 plots of the Swedish National Forest Inventory (NFI) to bi-temporal ALS data to predict and map site index using an area-based method and two regression modelling strategies: (1) a multiple linear regression (MLR) model with an ordinary least-squares parameter estimation method, and (2) a non-parametric random forests (RF) model optimised for hyper-parameter tuning. For model development, permanent plots were used, whereas the validation was done on the temporary plots of the Swedish NFI and an independent stand-level dataset. Species-specific models were developed, and the root mean square error (RMSE) metric was used to quantify the residual variability around model predictions. For both species, the MLR model gave precise and accurate estimates of SIS. The RMSE for SIS predictions was in the range of 1.96 – 2.11 m, and the relative RMSE was less than 10 % (7.68 – 9.49 %) of the reference mean value. Final predictors of site index include metrics of 90th percentile height and annual increment in the 95th percentile height, altitude, distance to coast, and soil moisture. Country-wide maps of SIS and the corresponding pixel-level prediction errors at a spatial resolution of 12.5 m grid cells were produced for the two species. Independent validations show the site index maps are suitable for use in operational forest management planning in Sweden.

1. Introduction

Sweden's forested land area is about 28 million hectares, from which 23.4 million hectares have the potential to produce at least one cubic meter of wood per hectare per year (i.e. productive forestland). On productive forestlands, 50 % of the area is privately owned, and the other half is shared between large forest companies and the state. Wood production (mainly timber and pulpwood) is the most dominant objective (Lindahl et al., 2017). Therefore, finding strategies to improve total wood production is central to most silvicultural interventions.

When optimising total wood volume production, large efforts are being made to individually plan silvicultural treatments for each forest stand. Access to fundamental data such as stem volume, mean tree height, and stem diameter for each stand is necessary to support this planning process. These stand attributes have been traditionally obtained from field surveys. However, accurate estimates of these variables can directly or indirectly be made using modern remote sensing techniques (Maltamo et al., 2014). Of current mapping technologies,

* Corresponding author.
E-mail addresses: alex.appiah.mensah@slu.se (A. Appiah Mensah), jonas.jonzen@slu.se (J. Jonzén), Kenneth.Nystrom@slu.se (K. Nyström), Jorgen.Wallerman@slu.se (J. Wallerman), Mats.Nilsson@slu.se (M. Nilsson).

https://doi.org/10.1016/j.foreco.2023.121395
Received 7 July 2023; Received in revised form 25 August 2023; Accepted 29 August 2023
Available online 8 September 2023

airborne laser scanning (ALS) features prominently due to its superior ability to resolve three-dimensional (3D) vegetation structures (Maltamo et al., 2014). Since its advent in a pilot study of a coastal Scots pine stand (Nilsson, 1996), ALS has been seen as a breakthrough for forest remote sensing in Sweden. Sweden's first national laser scanning began in 2009 and ended in 2016. Compared to the earlier practice based on field survey data, the first scan revolutionised forest management planning by producing large-scale and spatially explicit maps of forest attributes. The maps are open for public use and highly appreciated by Sweden forest owners (Nilsson et al., 2017). The Swedish National Mapping Agency began a second nationwide scanning in 2019 and is expected to end by 2024. The second scan means new possibilities, for example, for the estimation of height growth and site index.

Accurate and up-to-date information about site productivity is vital when deciding upon assumed efficient silvicultural treatments or when making a prognosis of forest status and development over time. Forest site productivity is defined here as the potential volume production that can be realised at a certain site with a given species, genotype and management (Skovsgaard and Vanclay, 2008). Site productivity decides the value of forestland, the choice of tree species and spacing, and the timing and priority of silvicultural operations. Site productivity is also used as a legislative boundary between productive and unproductive forestlands. National statistics about forest resources are reported differently based on this boundary (Nilsson, 2021).

Site index is by far the most common expression of forest site productivity. In the Swedish forest site classification system, site index based on height development curves (denoted as SIH) at a predetermined age (e.g., 100 years for *Pinus sylvestris* – Scots pine and *Picea abies* – Norway spruce) is the most preferred method in homogenous stands (Hägglund, 1981; Mensah et al., 2022). The curves show the development of mean heights of the dominant (top) trees (e.g., the 100 largest trees in diameter per hectare) over age. At sites where the dominant height cannot be measured (e.g., after clear-cut or thinning from above), it is possible to assess the site index using ground vegetation and site properties (the resulting site index is denoted as SIS), although this is a less accurate alternative to SIH (Hägglund and Lundmark, 1977). SIS is also species-specific, age-independent, and expresses the expected height at a reference age of 100 years in even-aged stands. The application of SIS in Sweden is manifold. Due to poor age records (especially on the permanent plots) and difficulties in determining top-height trees, the SIS has been the basis for reporting by the Swedish National Forest Inventory (NFI) since 2003. SIS is also an important predictor in the growth models for projecting young stand height development in the forestry-planning package, Heureka (Wikström et al., 2011). Mensah et al. (2023) used SIS to study growth trends, and Ekö et al. (2008) to compare growth differences of the major tree species in Swedish forests. Most private forest owners and large forest companies use SIS in their forest management plans. Nevertheless, a major limitation of SIS is its availability only limited to sample plots and field registers of forest companies that are not publicly available. To stimulate activities in the forest sector, increased availability of up-to-date and accurate spatial information on SIS is needed. Such readily available map data may be significant for local decision-making (Ulvdal et al., 2023) and large-scale operational forest management planning (Wilhelmsson, 2023).

In recent years, 3D data acquired by ALS have been useful for site index prediction and mapping. Essentially, the use of bi-temporal ALS data to determine site index has gained popularity due to two main reasons: (1) the ability to detect and exclude disturbed forest areas improves site index determination (e.g., Moan et al., 2023) and (2) because site index models are also height growth models, the addition of a growth rate parameter allows for both reliable predictions of the maximum potential height and inference of stand development (e.g., Sharma et al., 2011; Riofrío et al., 2023). For a forested country like Sweden, where there is a strong latitudinal gradient in growth (i.e. higher yield capacity in the south than in the north) and variation in forest management intensity (i.e. shorter rotation length in the south than in the north), information from repeated ALS campaigns is expected to improve the determination of site index in even-aged coniferous stands.

So far, bi-temporal ALS-based site index estimation has been generally approached in two main ways, where site index is predicted from either individual tree segmentation from ALS data (e.g., Kandare et al., 2017; Solberg et al., 2019) or employing an area-based approach (e.g., Persson and Fransson, 2016; Socha et al., 2017; Noordermeer et al., 2018). Kandare et al. (2017) predicted the site index in a boreal forest site in Norway with an accuracy (relative RMSE) of 27 % by fusing ALS and hyperspectral data through individual tree crown delineation. A novel method of age-independent site index estimation using repeated single-tree ALS data was demonstrated with high accuracy by Solberg et al. (2019) in a spruce-dominated area in southern Norway. From an area-based approach, Noordermeer et al. (2018) also determined the site index of Scots pine (RMSE = 1.08 m) and Norway spruce (RMSE = 1.78 m) dominated stands in southeastern Norway with bi-temporal ALS data. Persson and Fransson (2016) predicted the site index of a hemi-boreal forest in Remningstorp – Sweden, using data from two airborne laser scans (RMSE = 2.3 m). Socha et al. (2017) successfully estimated site-specific growth trajectories of Norway spruce stands from a short time series of repeated ALS data in Poland. Clearly, the above studies show that there is a potential in linking field-measured site indices to the time series of ALS data in Swedish forests.

One potential area for improvement of site index determination is model development. For a given species, location, growth resources, and management, it is well established that site productivity shows significant site-dependent variations and that it is very difficult to design the relations between the predictors and their effects appropriate for a specific site (Hägglund and Lundmark, 1977). For example, the geographical location (e.g., latitude and altitude) of a forest stand may affect the site index via its influence on the length of the vegetation period (Langlet, 1936). Further, the deficiency in precipitation in areas close to the coast also means that the site index maximum does not always occur at sea level. In such areas, the site index culminates at an altitude of 75 – 200 m, especially on dry and mesic sites (Lundmark, 1974). Thus, the complexity of growth-site factors potentially affects the distribution of the expected site index as well as the structure and form of the site index model. In many studies of site index determination, site index is linked to remotely sensed data either through parametric statistical models (e.g., general and generalised linear models), semi-parametric models (e.g., generalised additive models), or non-parametric models (e.g., random forests and boosted regression trees) (e.g., Nothdurft et al., 2012; Noordermeer et al., 2018; Antón-Fernández et al., 2023). Non-parametric models (for instance, machine learning models) have gained some popularity due to their ability to automatically learn from data, explain patterns and describe complex nonlinear relationships without making explicit assumptions about the error distribution, how the predictor variables relate to each other, and the response variable (Penner et al., 2013; Watt et al., 2021). However, a major drawback of these models is that erroneous predictions can be expected when the models are applied outside the range of the training data (Sabatia and Burkhart, 2014). On the other hand, parametric models are still relevant because they account for the distribution of the response variable (Næsset et al., 2005). Owing to the biological limitations of the site index (Antón-Fernández et al., 2016), a logical parametric model would probably do better than a non-parametric machine learning model when applied outside the range of the training data (Zhao et al., 2018).

This current study aims to provide cost-efficient site index information for forest management using probability field samples and available remote sensing data in Sweden. Specifically, we aimed to (1) examine the suitability of bi-temporal ALS data and other remotely sensed data for SIS prediction, (2) compare the performance of parametric and non-parametric models for SIS prediction, and (3) produce national-level

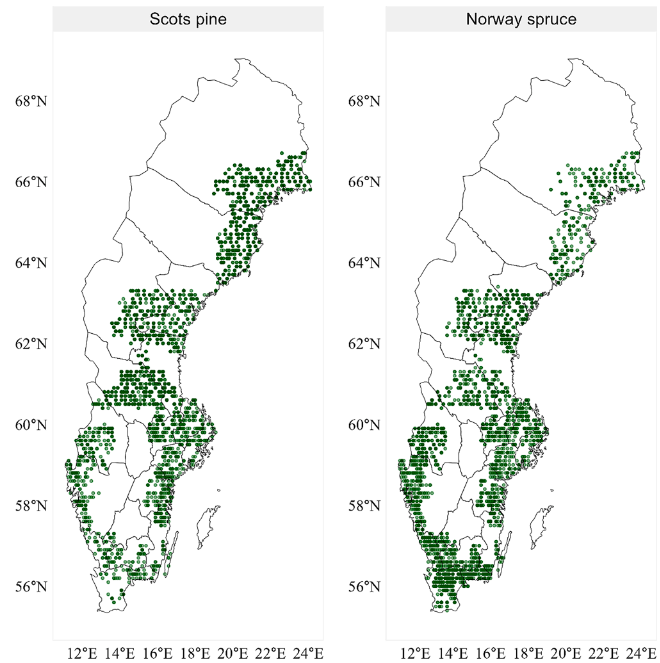

Fig. 1. Locations of the NFI temporary sample plots (green points) where there are repeated ALS scanning in the years 2011 to 2021. Note that the permanent NFI plots are not shown here in order not to disclose their locations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

maps of SIS and associated prediction errors. To address these objectives, we used observations of SIS from the temporary and permanent plots of the Swedish NFI, bi-temporal ALS data from the National Land Survey, and other available auxiliary data describing site conditions (e.g., soil moisture, distance to coast and altitude above sea level). We explored two regression modelling strategies (general linear regression with ordinary least squares for parameter estimation and a random forests regression) to link SIS and the remotely sensed predictor variables. The validity of the functions was checked by comparison with data not used for constructing the functions. The best model was used to produce wall-to-wall fine-resolution (12.5 m × 12.5 m) SIS maps for the two most dominant tree species, Scots pine and Norway spruce, in productive forestlands in Sweden.

2. Materials and methods

2.1. Swedish NFI: design, tree, stand and site variables

Knowledge of the Swedish forests – their area, volume, composition, etc. – has been obtained through forest inventories. For the whole of Sweden, inventories have been performed since 1923. The Swedish NFI is a probability sample with a statistical design characterised as systematic cluster sampling with partial replacement of units (Ranneby et al., 1987). Over the years, the survey methods have been changed several times and adapted to new conditions (Fridman et al., 2014). Between 1953, when the system with tracts (i.e. square clusters of sample plots) was introduced, and 1982, all the tracts were temporary. Since 1983, the NFI has been performed with an interpenetrating system where a set of temporary and permanent tracts are annually measured. Since 2017, the temporary sample plots have been designed based on a balanced sampling method (Grafström et al., 2017). The permanent plots are revisited every five years, and each plot has an area of 314 m^2 (radius of 10 m). The temporary plots are surveyed once, and each plot has an area of 154 m^2 (radius of 7 m). The position of the plot centres is determined with GPS receivers (Garmin GPSMAP 64), which have a horizontal accuracy of about 5 m (Nilsson et al., 2017).

On all plots, a large number of variables concerning the administration (e.g. land ownership), site (e.g. latitude, altitude, and site productivity), tree (e.g., total height, diameter, species, etc.), and stand (e.g. relative density, the basal area measured by relascope sampling, tree species composition, and the type and type of forest management activities) are measured. The biometric measurements of tree and site variables are done on the 7 and 10-m radius plots respectively for the temporary and permanent plots. The stand variables are measured on a 20 m radius in both plot types.

2.1.1. SIS determination on NFI plots

On a 10-m radius of temporary and permanent plots, site productivity is determined by the SIS-method, i.e., Site Index according to Site properties (SIS, m) using species-specific functions. The functions are based on primary site variables such as climate (average temperature sum within the growing season – degree days when air temperatures exceed + 5 °C), inclination and direction of the slope, soil depth, texture, moisture, humidity index, and field vegetation. The SIS function was developed earlier by (Hägglund and Lundmark, 1977), using site index predicted from height development curves (SIH) as the response variable and the above site variables as the predictor variables in a regression analysis. Controlled assessments of SIS on the same plots by independent field crews have shown a measurement error of about 2 m and no bias (Fridman et al. 2019). Sites are classified as Scots pine sites or Norway spruce sites according to dominant species (i.e., species with more than 65 % of the total basal area on a plot) or the species with the

Table 1
Descriptive statistics for NFI plot-level data separated by species and plot type.

	Permanent plots				Temporary plots			
Variable, units	Mean	Median	Min.	Max.	Mean	Median	Min.	Max.
Scots pine								
SIS, m	21	21	11	28	21	21	11	28
Basal area, m^2/ha	16	15	0	64	16	15	0	60
Mean height, m	15	15	5	31	14	14	5	32
Volume, m^3/ha	122	108	0	796	118	99	0	758
Yield capacity, m^3/ha/yr	5	5	1	8	5	5	1	8
Number of plots, n	2242				3740			
Norway spruce								
SIS, m	27	28	11	39	26	28	11	40
Basal area, m^2/ha	15	14	0	70	16	14	0	84
Mean height, m	17	17	5	34	16	16	5	38
Volume, m^3/ha	132	99	0	963	135	92	0	1200
Yield capacity, m^3/ha/yr	8	8	2	16	8	8	2	17
Number of plots, n	2077				3453			

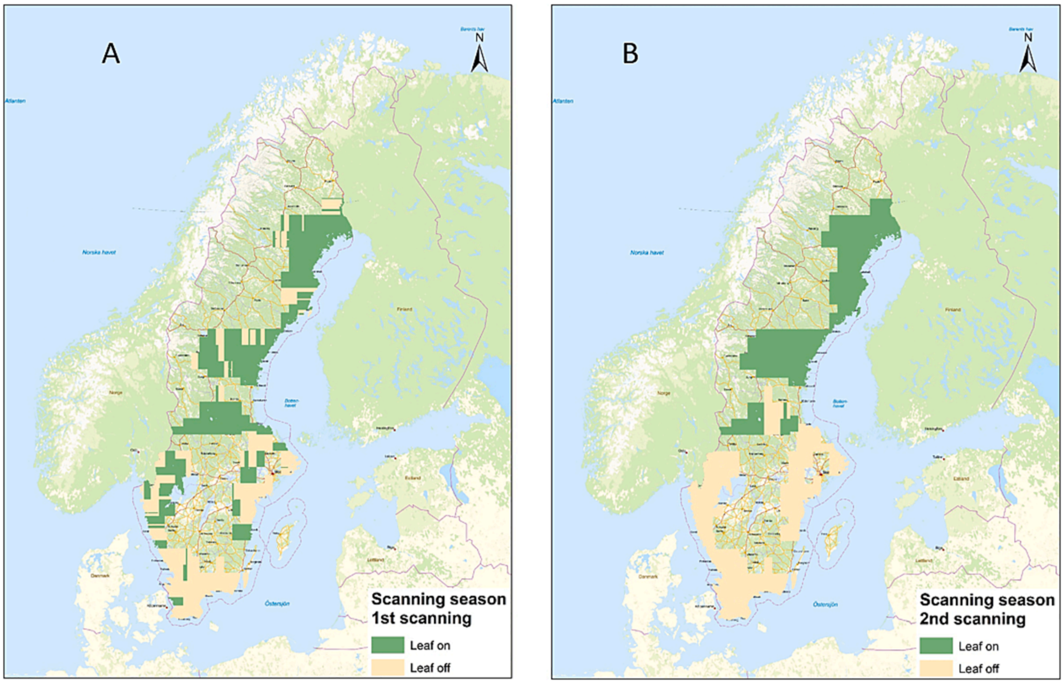

Fig. 2. The ALS data used in the study. Maps show areas with repeated ALS scanning (A – first scanning and B – second scanning) and the scanning season (leaf-on and leaf-off). Copyright Lantmäteriet Topographic Map.

highest SIS value.

In this study, plot-level observations of SIS were considered as the response variable in the regression analysis. We used field data from the years 2011 to 2020 (Fig. 1), corresponding to areas with repeated airborne laser scanning data. Descriptive statistics about the NFI field data used in the study are provided in Table 1. The distribution of SIS in the temporary and permanent NFI sample plots is shown in Fig. A1 (appendix).

2.2. Laser scanning data

The campaign for a new national digital elevation model (DEM) in Sweden began in 2009 when the National Mapping Agency launched the first airborne laser scanning. The first nationwide ALS was completed in 2019, with more than 90 % of the forestland scanned in 2016. The scanning campaign was organised in blocks of sizes 25 km × 50 km. The scanning altitude ranged from 1700 m to 2300 m, and a pulse density of 0.5 – 1 pulses/m^2. The ALS data were acquired in different seasons (i.e. leaf-on and leaf-off) using 13 different scanners from Leica (72.3 % of the scanned blocks), Optech (24.9 %), and Reigl and Trimble (2.8 %). Details of the first national scanning can be found in Nilsson et al. (2017). The second campaign began in 2019 and is planned to end in 2024. It is carried out in the same way as the first scanning but with a slightly higher pulse density (1.0 – 1.51 pulses/m^2) using Leica scanners. Presently, about 70 % of the productive forestland in Sweden has been scanned twice (Fig. 2).

The new national DEM (2 m × 2 m) produced by the National Mapping Agency was used as the ground reference when computing ALS height returns above ground (Nilsson et al. 2017). For each NFI field plot and map unit, laser metrics were derived using the Fusion software (McGaughey, 2015) following the area-based method (Næsset, 2002). The metrics were calculated for 12.5 m × 12.5 m grid cells and comprised the mean, mode, standard deviation, interquartile distance, and percentiles from the height distribution of laser returns between 1.5 m and 50 m height above ground. The metrics were computed independently and separately for the two scanning campaigns.

In addition to the standard metrics, we also computed metrics of change in height. Since the national laser scan occurs at different times of the year, affecting the length of the vegetation period, we normalised the change in height with the number of growing seasons between scans to produce estimates of annual height growth (hereafter referred to as height increment). The number of growing seasons between scans was adjusted for the start and end of the growing season based on the average values (of the start and end of the growing season) in the last ten years in Sweden (SMHI, 2018). The annual height increments were derived for the 80th, 90th, 95th, and 99th percentile heights as the quotient of the height difference between the second and first scanning and the number of growing seasons between scans.

2.3. Other auxiliary data.

To account for the variations in growth and site conditions from the north to the south of Sweden, environmental variables describing the location (e.g., latitude), altitude (meters above sea level, m a.s.l), distance to the nearest coast (km) and the degree of moisture in soils (expressed as a percentage, %) were obtained. Altitude was derived from the new national DEM. Distance to the coast was computed as the distance from the centre of a pixel or the field plot to the nearest coastline. The coastline was a vector file also provided by the National Mapping Agency. For soil moisture, we extracted the information from the available soil moisture maps with a spatial resolution of 2 m across the entire Swedish forest landscape (Ågren et al., 2021). The map data shows the degree of soil wetness, with values expressed in a probability index scale of 0 – 100 % (i.e., dry to wet). The maps were produced using machine learning and data from ALS-derived terrain indices and field plots of the Swedish NFI. These variables were selected because they are readily available. Before the statistical modelling, the spatial resolution of each raster was resampled to match that of the ALS metrics (i.e. 12.5 m).

2.3. Statistical models

The area-based approach (Næsset, 2002) was considered for the model calibration. In the first stage, observed site indices in georeferenced field plots were regressed on the ALS metrics available for each plot unit to develop a predictive model. In the second stage, the models were applied over tessellations of individual grid cells available to predict and generate a countrywide wall-to-wall map of the site index.

Before the calibration of models, we optimised the selection of candidate predictor variables by testing different ALS metrics and a large number of possible combinations following the method proposed by Ekström and Nilsson (2021). Afterwards, less strongly intercorrelated variables were selected (i.e. $\rho \leq 0.5$) for the regression modelling. The final set of predictors were ALS-derived 90th percentile height at second scanning, the annual increment in the 95th percentile height, location, altitude, distance to the coast and soil moisture. The correlation levels (Pearson's) among the final predictors are shown in Fig. A2 (appendix).

For the treatment of outliers, we first used bivariate scatterplots (not presented here) to examine the relationship between laser-returned 95th percentile height at the second scanning and the field-measured basal area weighted mean height. Observations with differences of ±5 m in mean height were removed (about 5 % of the original data) since they were considered to be affected by silvicultural operations such as thinning and clearcutting.

The models were calibrated on the permanent sample plots of the Swedish NFI. In the regression analysis, each sample plot was treated as an independent observation without any consideration of potential spatial autocorrelation. This approach was motivated by the following reasons (see Ranneby et al., 1987): (1) the stratification of the country into five inventory regions to obtain the efficient orientation of the tracts (i.e., tracts are oriented at 45° to account for the elliptical correlation in the north–south and east–west directions), (2) the tracts have a distance between them of 5 km in the south and 15 km in the north and (3) the distance between sample plots in a tract is 300 m in the south and 600 m in the north to have low spatial autocorrelation.

2.3.1. Model development

The predictor variables were broadly grouped into the site, growth resource availability, and tree size. For given values of the predictor variables, we examined the model relating the expected mean site index (i.e., mean SIS) as a function of the predictor variables as:

$$\mathrm{E(SIS)} = \mathrm{f(site, resource, size)} \tag{1}$$

Latitude (Lat) was used together with altitude (Alt) and distance to the nearest coastline $(\mathrm{Dist_{coast}})$ to describe the geographical location of a plot and thus represent essential features of temperature and light climate on the plot (Lundmark, 1974). Due to the supply of nutrients and oxygen to the soil, which follows the surface or subsurface water flow, we used soil moisture $(\mathrm{Soil_{wet}})$ (Ågren et al., 2021) as a proxy for the positive effect of mobile water on forest growth (Troedsson, 1965).

The calculation of SIS does not necessarily require information about dominant height and age but rather it is computed based on combined geocentric and phytocentric indicators (*see section 2.1.1*). This implies that the resulting site index is density-dependent and may vary with stand development over time (e.g., due to the different light requirements of understory vegetation). In addition, volume production varies within a given site index due to the differences in the carrying capacity of stands (Mensah et al., 2022). To capture the effects of density and variation in growth on predictions of SIS, it is imperative to include metrics of size and increment. Both $\mathrm{H_{p90}}$ and $\mathrm{H_{p95}}$ metrics correlate strongly with many important attributes (e.g., Lorey's mean height, diameter, basal area and stem volume) in Sweden (Nilsson et al., 2017). Therefore, in Equation 1, the size and increment variables were defined by the mean height $(\mathrm{H_{p90.t2}})$ which is density-dependent and the annual height increment rate $(\Delta\mathrm{H_{p95}})$ which is strongly correlated with volume production. These variables can be regarded as direct biological indicators of site productivity (Assmann, 1970).

We expanded Equation 1, where the site index was modelled using two regression approaches: parametric general linear and a random forests model. The models were fitted separately for Norway spruce and Scots pine.

2.3.2. Linear regression model

For the linear model, we assumed that the expected mean SIS was linear in the parameters of the function f (Eq. 1), describing the relationship between the mean site index and the remotely sensed predictor variables. To decrease the influence of unknown and complex growth-site interactions, we chose a general multiple linear regression (hereafter referred to as the MLR model) with an additive error structure (Eq. 1.1). This model was a contrast to the model by Hägglund and Lundmark (1977) which assumed that the effects of different site factors work together in a multiplicative way and therefore used a logarithmic site index as the dependent variable.

$$\mathrm{E(SIS)} = \alpha_0 + \beta \bullet (\mathrm{site}) + \gamma \bullet (\mathrm{resource}) + \omega \bullet (\mathrm{size}) \tag{1.1}$$

where α_0 is the intercept, β is the vector of coefficients for the physical site variables, γ is the vector of coefficients for the growth resource variable, and ω is the vector of coefficients for the tree size variables. The explicit forms of the various predictor variable groups were defined as:

$$\beta \bullet (\mathrm{site}) = \beta_1 \bullet (\mathrm{Alt}) + \beta_2 \bullet (\mathrm{Lat}) + \beta_3 \bullet (\mathrm{Dist_{coast}}) \tag{1.2}$$

$$\gamma \bullet (\mathrm{resource}) = \gamma_1 \bullet (\mathrm{Soil_{wet}}) + \gamma_2 \bullet (\mathrm{Soil_{wet}})^2 \tag{1.3}$$

$$\omega \bullet (\mathrm{size}) = \omega_1 \bullet (\mathrm{H_{p90.t2}}) + \omega_2 \bullet (\Delta\mathrm{H_{p95}}) \tag{1.4}$$

Parameters of the MLR models were estimated by the ordinary least squares method, which minimised the sum of squares for the error. The MLR models were fitted respectively with the "lm" function in the R Statistical Environment (R Core Team, 2022), and the significance of the model parameters was tested at a 5 % alpha level. To determine whether each predictor variable introduced to the model contributes to significantly improving the quality of fit of the model to the data, an F-test was used to compare the sum of square errors for the model with and without the variable of interest. Preliminary fits of the models to the data revealed that the individual random model errors were homoscedastic and that the assumption of additive errors seemed appropriate.

2.3.3. Random forests regression model

We fitted the random forests regression model (hereafter referred to as the RF model) to examine the relationship between the mean site index and the remotely sensed predictor variables (Eq. 1). The RF model is based on an ensemble tree method where a set of independent and less correlated decision trees are generated and aggregated to reduce the variance of predictions (Breiman, 2001). Further, the RF model is non-parametric; thus, no explicit assumptions about how the predictor variables relate to each other and the response variable is needed. RF requires two parameters to be set: (1) "mtry", the number of predictor variables executing the data splitting at each node and (2) "ntree", the total number of decision trees to be grown in the model run. To optimise the RF model, a two-stage hyperparameter tuning of mtry and ntree was carried out. First, a train control was set up using a resampling method with repeated cross-validation where the number of folds and repeats was set to 10 and 3, respectively. Second, a grid search where one to seven predictor variables were tested for splitting at each tree node, and the total number of trees grown in the model run was set to 500. We used the measure of the Gini decrease in node impurity to evaluate the importance of predictor variables in the RF model. The RF model was fitted using the "randomForest" function in the R Statistical Environment (R Core Team, 2022).

2.4. Model validation and predictive site index maps

To evaluate the predictive performances of the fitted MLR (Eq. 1.1) and RF models, validation was done on two independent data sets: (1) temporary plots of the Swedish NFI and (2) stand-level data from the state-owned forest company Sveaskog. The stand-level data was

Table 2
Goodness-of-fit statistics of the species-specific SIS models based on the NFI data. Calibration and validation refer to the permanent and temporary NFI sample plots.

Model	Metric	Scots pine Calibration	Scots pine Validation	Norway spruce Calibration	Norway spruce Validation
MLR	RMSE (m)	1.958	2.019	2.072	2.105
	$RMSE_{rel}$ (%)	9.209	9.492	7.683	7.983
	R^2_{adj} (%)	63.471	60.775	87.409	87.641
	RE	1.03		1.02	
RF	RMSE (m)	0.846	2.184	0.920	2.112
	$RMSE_{rel}$ (%)	3.982	10.271	3.412	8.011
	R^2_{adj} (%)	94.399	59.636	97.775	87.651
	RE	2.58		2.29	

collected for another management purpose by the company and consisted of 311 stands (84 stands for Norway spruce and 227 stands for Scots pine) of average size 20 ha (range 0.6 – 159 ha). The stands were subjectively surveyed using on average eight circular plots (of radius 8 m) per stand. Fig. A3 (appendix) shows the distribution of the validation stands. Model diagnostics involved the evaluation of predicted versus observed SIS and the distribution of the individual model errors on the plot-level predictions of SIS and other site characteristics.

The accuracy of the final models was compared based on the following performance estimators (Eqs. 2 – 6): adjusted coefficient of determination ($R^2{}_{adj}$, %), root mean square error (RMSE, m), relative root mean square error ($RMSE_{rel}$, %); mean deviation (MD, m) as a proxy for bias; and relative mean deviation (MD_{rel}, %).

$$R^2{}_{adj} = 1 - \frac{(n-1)\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{(n-k)\sum_{i=1}^{n}(\hat{y}_i - \overline{y}_i)^2} \tag{2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n-k}} \tag{3}$$

$$RMSE_{rel} = 100\,x\,\frac{RMSE}{\overline{y}} \tag{4}$$

$$MD = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i) \tag{5}$$

$$MD_{rel} = 100\,x\,\frac{MD}{\overline{y}} \tag{6}$$

where $\hat{y}_i$, y_i, and $\overline{y}_i$ are the predicted, measured, and average values of the response variable, respectively; *n* represents the total number of observations used for model fitting; and *k* denotes the number of model parameters.

In addition to the above performance estimators, we also computed (Eq. 7) an index of relative efficiency (RE) to determine the stability of the predictive performance of the fitted models. Since the SIS of a given species is determined by the same field protocol for both the temporary and permanent NFI sample plots and the distribution of observed SIS is similar in the two plot types (Fig. A1, see appendix), lower values (RE less than and close to 1) of the ratio of the residual variability around model validation (i.e., on temporary plots) to residual variability around model calibration (i.e., on permanent plots) reflect better and a more stable predictive model and otherwise for an over-fitted model (i.e. RE values greater than 1).

$$RE = \frac{RMSE_{validation}}{RMSE_{calibration}} \tag{7}$$

where $RMSE_{validation}$ is the root mean square statistic of model validation (i.e. on temporary NFI plots), and $RMSE_{calibration}$ is the root mean square statistic of model calibration (i.e. on permanent NFI plots).

Predictive maps of site index for Scots pine and Norway spruce were developed based on the best-obtained model (i.e. the model with high statistical precision and low prediction error). The model was then applied to tessellated grid cells of the predictor variables to generate wall-to-wall maps of predicted SIS at a spatial resolution of 12.5 m × 12.5 m.

2.4.1. Uncertainty estimation

Since the site index predictions are model-based and conditioned on the modelling data set (Saarela et al., 2020; Kangas et al., 2023), it is important to quantify the amount of uncertainty in the site index maps. The estimated uncertainty may serve both as a measure of model accuracy and reflect the reliability of the estimated site maps for resource planning at different decision levels (Ulvdal et al., 2023). Further, statistical models are highly uncertain outside the range of the calibration data. This means future predictions of SIS at the areas presently not covered by the two ALS scans may be subjected to prediction errors. Generally, uncertainties in model predictions are influenced by model misspecification, values of the predictor variables, model form and parameter estimation, and the residual variability around model predictions (Nyström and Ståhl, 2001; McRoberts and Westfall, 2016). We focused on the latter two for the estimation of pixel-level prediction errors, as the site index maps are essentially predictions with locations (Kangas et al., 2023).

In the case of the parametric MLR model, we utilised the error estimates available from the model fit to the observed data. Specifically, the analytical form of the prediction error variance for each pixel prediction of SIS is a combination of the estimation error of model parameters and variance of the residual errors under the assumptions of model prediction unbiasedness and homoscedasticity of model residuals (Eq. (5) in Saarela et al., 2020) as:

$$\widehat{RMSE}_{SIS_i} = \sqrt{X_i \bullet \widehat{Cov}(\hat{\beta}) \bullet X_i{}^T + \hat{V}(\varepsilon_i)} \tag{8}$$

where $\widehat{RMSE}_{SIS_i}$ is the estimated standard error for the predicted SIS in pixel i ($\widehat{SIS}_i$), X_i is a (k + 1) length vector of partial derivatives (obtained by the delta method) of the MLR model concerning the estimated parameters ($\hat{\beta}$s), $\widehat{Cov}(\hat{\beta})$ is the estimated covariance matrix of the $\hat{\beta}$s, and $\hat{V}(\varepsilon_i)$ is the variance of the residual errors. The corresponding relative uncertainty ($\widehat{RMSE}_{SIS_i-rel}$) was then calculated as follows:

$$\widehat{RMSE}_{SIS_i-rel} = \left(\frac{\widehat{RMSE}_{SIS_i}}{\widehat{SIS}_i}\right) \bullet 100 \tag{8.1}$$

In the case of the non-parametric RF model, bootstrapping was used to assess pixel-wise uncertainties. We used the pairs (i.e. non-parametric) bootstrap resampling method to estimate the standard errors for the predicted SIS following Esteban et al. (2019). The following steps were carried out: (1) For each value of the predictor variable in pixel i, we generated 1000 bootstrap replications. (2) We applied the RF model where the number of decision trees was set to 500. (3) For each decision tree, a new bootstrap sample of size 1000 was randomly selected with replacement to predict SIS. (4) The standard deviation of the 500 estimates of the mean predicted SIS, one for each of the default 500 RF trees, was then used as a proxy for the standard error of prediction in pixel i. We selected 1000 bootstrap runs based on the criterion that larger values enhance the stability of the bootstrap estimate of the standard error over replications (McRoberts and Westfall, 2016), as well as to reduce computational cost.

3. Results

3.1. Models

Table 2 summarises the predictive performance of the two fitted site index models – the general multiple linear regression model (MLR) and

Table 3
Coefficients of the MLR site index model (Eq. 1.1). SE is the standard error, and SE_{rel} is the standard error relative to the estimate.

Parameter	Variable/Group	Estimate	SE	SE_{rel} (%)	P-value
		Scots pine			
$\hat{\alpha}_0$	Intercept	45.87	1.0120	2.206	< 0.0001
$\hat{\beta}_1$	Lat/Site	-3.877×10^{-6}	1.427×10^{-7}	−3.681	< 0.0001
$\hat{\beta}_2$	Alt/Site	−0.0042	3.825×10^{-4}	−9.217	< 0.0001
$\hat{\beta}_3$	$Dist_{coast}$/Site	−0.0012	5.168×10^{-4}	−44.939	0.0305
$\hat{\omega}_1$	$H_{p90.t2}$/Size	0.0079	3.547×10^{-4}	4.507	< 0.0001
$\hat{\omega}_2$	ΔH_{p95}/Size	11.66	0.4896	4.199	< 0.0001
$\hat{\gamma}_1$	$Soil_{wet}$/Resource	0.0649	0.0049	7.683	< 0.0001
$\hat{\gamma}_2$	$(Soil_{wet})^2$/Resource	−0.0010	5.659×10^{-5}	−5.619	< 0.0001
		Norway spruce			
$\hat{\alpha}_0$	Intercept	115.4	1.1210	0.971	< 0.0001
$\hat{\beta}_1$	Lat/Site	-1.326×10^{-5}	1.685×10^{-7}	−1.271	< 0.0001
$\hat{\beta}_2$	Alt/Site	−0.0105	4.906×10^{-4}	−4.695	< 0.0001
$\hat{\beta}_3$	$Dist_{coast}$/Site	9.319×10^{-4}	5.976×10^{-5}	6.413	0.119
$\hat{\omega}_1$	$H_{p90.t2}$/Size	0.0036	3.083×10^{-4}	8.65	< 0.0001
$\hat{\omega}_2$	ΔH_{p95}/Size	4.9340	0.4011	8.129	< 0.0001
$\hat{\gamma}_1$	$Soil_{wet}$/Resource	0.0096	0.0054	55.95	0.074
$\hat{\gamma}_2$	$(Soil_{wet})^2$/Resource	-3.678×10^{-4}	5.978×10^{-5}	−16.253	< 0.0001

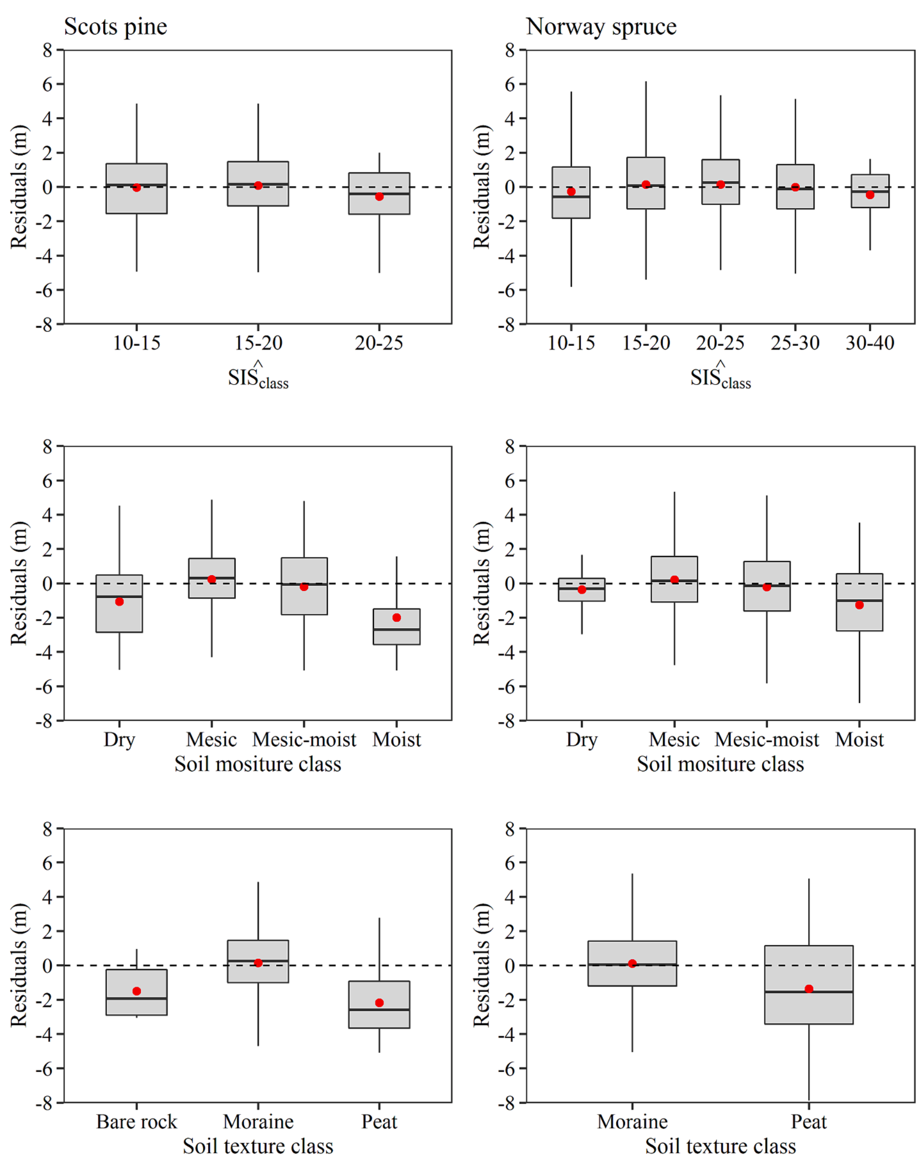

Fig. 3. Residuals diagnostics of the fitted MLR site index models.

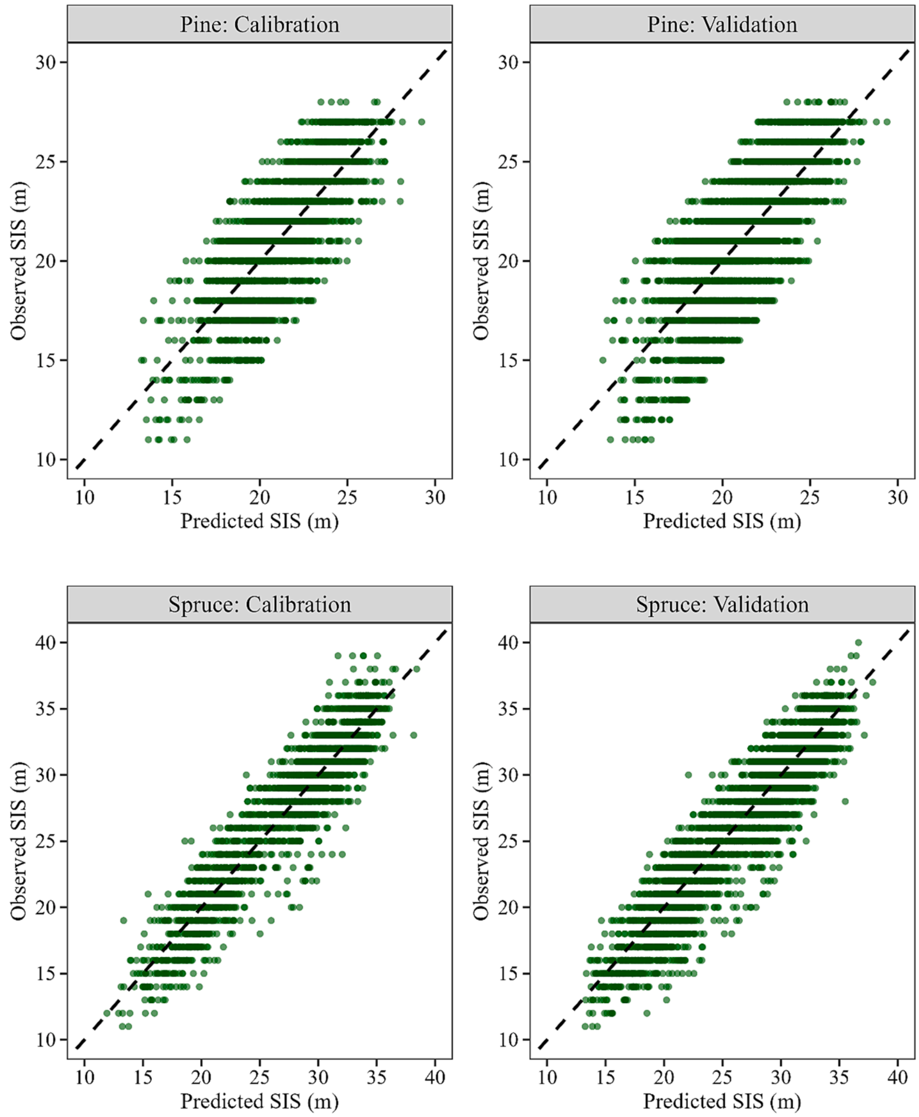

Fig. 4. Observed (y-axis) vs. predicted (x-axis) SIS (from the MLR model) on the calibration (permanent NFI plots) and validation (temporary NFI plots) data for Scots pine and Norway spruce-dominated stands.

the random forests regression model (RF). Scots pine and Norway spruce's site index prediction accuracy was higher for the MLR models than for the RF models. Further, the relative efficiency (RE) values obtained under the MLR models were close to 1 (1.03 for Scots pine and 1.02 for Norway spruce), while the corresponding values obtained under the RF models were larger than two folds (2.58 for pine and 2.29 for spruce). Thus, the MLR models reflected an accurate and more stable predictive model of SIS than the RF models, which showed overfitting (Table 2 and Fig. A4). Hence, the RF models were excluded from further analysis.

The full MLR site index model (Eq. 1.1) for the two species was explicitly determined as follows:

$$SIS = \alpha_0 + \beta_1 \bullet (\mathrm{Lat}) + \beta_2 \bullet (\mathrm{Alt}) + \beta_3 \bullet (\mathrm{Dist_{coast}}) + \omega_1 \bullet \left(\mathrm{H_{p90.t2}}\right)^2 + \omega_2 \bullet \left(\Delta\mathrm{H_{p95}}\right)^2 + \gamma_1 \bullet (\mathrm{Soil_{wet}}) + \gamma_2 \bullet (\mathrm{Soil_{wet}})^2 \quad 9$$

An F-test was used to determine whether reduced models (i.e. models fitted with only variables of a predictor group, Eqs. 1.2 – 1.4) performed significantly better than the full model (Eq. 1.1). The obtained results show that the full model was significantly better than the reduced models and that each predictor variable introduced to the model contributes to statistically significantly improve the quality of fit of the model to the data (Tables A.1 and A.2, appendix). The total variance explained (R^2_{adj}) by the full model was 60 % for pine and 87 % for spruce. Further, the average model errors (RMSEs) ranged from 1.96 m to 2.11 m, and the corresponding relative errors were less than 10 % (7.68 – 9.49 %) of the reference mean values (Table 2).

Coefficients for the effects of site (Eq. 1.2), resource (Eq. 1.3) and size (Eq. 1.4) in the final model are given in Table 3 by species. All estimated parameters, except for parameters γ_1 and β_3 in the Norway spruce model, were significant (p less than 0.05). The size coefficients (ω_1 and ω_2) were positive for both species, suggesting increasing SIS as average height and growth increase. The β_i coefficients for the geographic effects of latitude, altitude, and distance to the coast were all significant and negative for pine, but the distance to the coast was non-significant for Norway spruce. The significance of parameter γ_2 ($\mathrm{Soil_{wet}}$) indicates that the effects of mobile water on-site productivity can be best described in a quadratic form of soil moisture.

Variable definitions: Lat is latitude; Alt is Altitude; $\mathrm{Dist_{coast}}$ is the distance to coast; $\mathrm{H_{p90.t2}}$ is 90th percentile height at second scanning; $\Delta\mathrm{H_{p95}}$ is the annual increment in the 95th percentile height, and $\mathrm{Soil_{wet}}$ is soil moisture.

One statistical prerequisite for the general multiple linear regression (MLR) analysis was that the variance of the residuals should be constant.

Table 4
Goodness-of-fit statistics for different classes of predicted SIS. Values in parentheses correspond to the validation data (temporary NFI plots), and *n* refers to the number of observations per class.

Species	Metric	Predicted SIS (m) 10 – 15	15 – 20	20 – 25	25 – 30	30 – 40
Scots pine	RMSE (m)	2.083	1.916	1.695		
		(2.233)	(1.945)	(1.509)		
	$RMSE_{rel}$ (%)	11.357	8.565	6.682		
		(12.137)	(8.733)	(5.858)		
	MD (m)	−0.047	0.081	−0.559		
		(-0.048)	(0.116)	(-0.155)		
	MD_{rel} (%)	−0.254	0.363	−2.201		
		(-0.259)	(0.520)	(-0.602)		
	n	721	1377	144		
		(1197)	(2303)	(240)		
Norway spruce	RMSE (m)	2.044	2.069	2.217	1.972	1.734
		(2.043)	(2.422)	(2.171)	(1.831)	(2.287)
	$RMSE_{rel}$ (%)	11.594	9.203	7.874	6.112	4.932
		(11.868)	(11.021)	(7.798)	(5.742)	(6.596)
	MD (m)	−0.279	0.139	0.119	−0.042	−0.463
		(-0.602)	(-0.279)	(-0.235)	(-0.392)	(-1.005)
	MD_{rel} (%)	−1.583	0.618	0.421	−0.129	−1.316
		(-3.497)	(-1.271)	(-0.845)	(-1.228)	(-2.898)
	n	304	428	604	703	38
		(539)	(729)	(1000)	(1134)	(51)

Table 5
Goodness-of-fit statistics of the species-specific MLR site index models based on independent stand-level data and *n* is the number of stands used in the validation.

Metric	Scots pine	Norway spruce
RMSE (m)	1.80	1.56
$RMSE_{rel}$ (%)	9.98	6.20
MD (m)	0.7	0.2
MD_{rel} (%)	3.7	0.8
Mean of observed SIS (m)	18	25
n	227	84

Diagnostics of the MLR models showed that the variance of residuals was fairly constant over the predicted site index, but slight trends in the residuals were observed when validated against soil moisture and textural class from the Swedish NFI (Fig. 3). When evaluated against stand density (i.e., basal area) and productivity (i.e., stem volume), no obvious trends were observed in the residuals (Fig. A5). Further evaluations (not presented) of the model residuals showed no trends for the leaf-on and off-scanning campaigns. The observed versus predicted SIS corresponded well in both the calibration and validation data for the two species (Fig. 4), and the accuracy generally increased with larger values of predicted SIS (Table 4).

Further, to determine the reliability of the fitted MLR models, additional validation was made using independent data from stands managed by the state-owned forest company Sveaskog. About 84 and 227 stands, respectively, for Norway spruce and Scots pine were used for the validation. The results from the validation are summarised in Table 5 below. For Scots pine, the accuracy (RMSE) was 1.8 m (relative RMSE = 10 %) and a bias of less than 1 m (MD = 0.7 m and relative MD = 3.7 %). For Norway spruce, the model prediction accuracy (RMSE) was 1.56 m (relative RMSE = 6.2 %) and no bias (MD = 0.7 m and relative MD = 3.7 %).

3.2. Mapping and uncertainty estimation

The species-specific fitted MLR site index models were applied to tessellated grid cells of the predictor variables to predict the SIS for each pixel unit and the corresponding prediction errors at a spatial resolution of 12.5 m × 12.5 m. Given that the variance of the residuals from the fitted MLR models was unbiased and homogenous over predictions of SIS (Fig. 3), we applied the estimators in Equations (8) and (8.1) to compute the standard prediction errors in each pixel. We utilised the combined information from the estimation error of the model parameters (i.e., the estimated covariance matrix and derivative of the model concerning the parameters) and the residual error variance (given in Table 2). The estimated covariance matrix of the model parameters is given in Table A3 (appendix). The map of SIS predictions for Norway spruce-dominated stands is shown in Fig. 5. Maps at full resolution for an area of about 100 km^2 around Sundsvall detail the predicted site index and the corresponding relative errors at the pixel level. Largely, the relative errors of prediction decreased with larger values of predicted SIS for both species (Fig. 6).

4. Discussion

4.1. Data and models

The present study explored the potential of ALS data for site productivity estimation by site index in even-aged dominant stands of Scots pine and Norway spruce in Sweden. Our approach utilised data from field observations by the Swedish NFI, the national ALS data, and other remotely sensed data describing site conditions. For many large area projects, such as making predictions and mapping forest attributes, ALS data has played a significant role in enhancing operational forest management planning and research (Maltamo et al., 2014). Mostly, ALS data are trained with field observations from NFIs through the area-based approach to generate wall-to-wall predictions of forest attributes such as mean height, basal area, stem volume, and biomass. Such large-scale mapping has been demonstrated with satisfactory accuracy in several countries (Nord-Larsen and Schumacher, 2012; Nilsson et al., 2017).

So far, the use of ALS data for estimation and mapping of site index has been successful in operational forest management inventories (e.g., Tompalski et al., 2015; Socha et al., 2017; Noordemeer et al., 2018) and in many large area projects (e.g., Guerra-Hernández et al., 2021; Antón-Fernández et al., 2023). In the above studies, site index according to height development curves (i.e., from height–age observations, SIH) was predicted and used as a basis for characterising stand productivity. Strictly, this means that the stand itself is used as a biological marker of site productivity. However, such a test is valid only if the condition of the stand is such that the site index estimated is not unduly influenced by stand history and meets the standard assumptions of site index regarding age, stocking, and species composition (Elfving and Nyström, 1996). Thus, SIH models are species-specific and may be restricted to areas

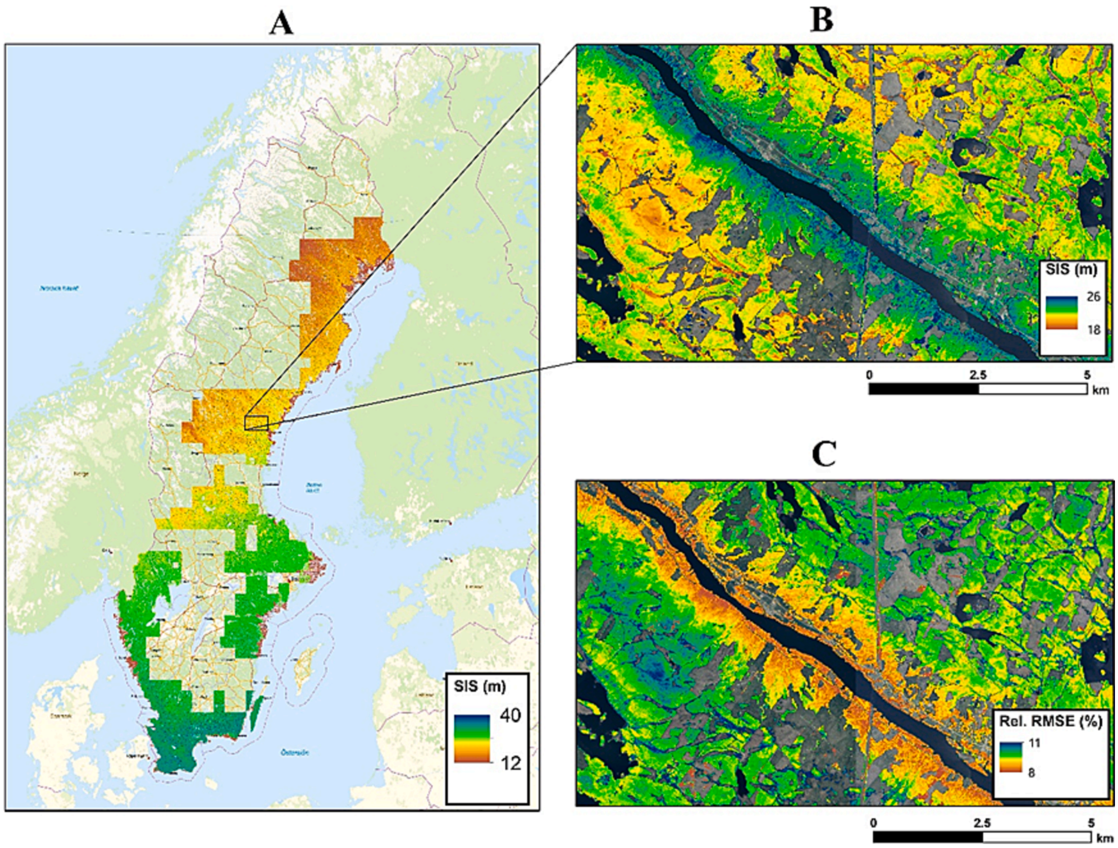

Fig. 5. (A) Map of predicted site index (SIS) for Norway spruce for areas where there are bi-temporal ALS data. (B) Fine-resolution maps of predicted SIS based on Equation (9) and (C) corresponding relative standard errors of SIS predictions for the area around Sundsvall. Copyright Lantmäteriet Topographic Map.

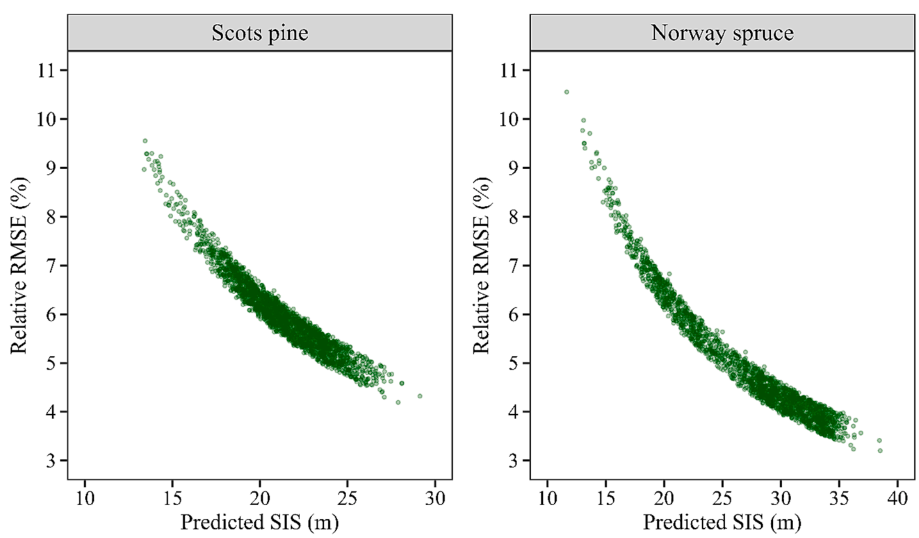

Fig. 6. Predicted site index (SIS) versus relative RMSE for Scots pine and Norway spruce.

where age and dominant (top) height information is known with less uncertainty. However, these variables (age and top height) are difficult to determine in NFIs, and their use for site quality assessment is questionable because of the initial suppression of advanced reproduction, especially for tolerant tree species like Norway spruce (Peng, 2000). Furthermore, the SIH method has poor precision in young stands and, of course, does not apply to areas without forests (Hägglund and Lundmark, 1977). These restrictions mean that SIH can only be used in 30 – 40 % of the total Swedish forest area. In the Swedish forest site classification system, site index estimation using site properties (i.e., SIS) has been used to overcome most of the above restrictions, though the site index is predicted to have an accuracy considerably lower (~4 m) than when using SIH in stands suitable for the two methods (Hägglund and Lundmark, 1977).

The above restrictions make it challenging to compare the accuracy of our results to those reported in other studies. However, we compare and discuss the results obtained in this study with the literature in the context of the RMSE, which is a commonly reported uncertainty statistic in forest attribute predictions by remote sensing (Persson and Ståhl, 2020).

The accuracy obtained under the general multiple linear regression (MLR) site index model (Eq. 2.6) was higher than that of the non-parametric RF model (Table 2 and Fig. A4). This finding may be due to (1) the better generality and logical behaviour of parametric linear models compared to RF models (Zhao et al., 2018) and (2) the less complex structure of even-aged coniferous forests (e.g., homogeneity in stem diameter, height, and species) (Sabatia and Burkhart, 2014). Other studies (e.g., Aertsen et al., 2010, Nothdurft et al., 2012) have found generalised additive models (GAM) and generalised linear models (GLM) to be robust for site index determination. We emphasize that in the earlier investigations of the present study, a GLM model (with both Gaussian and square root link functions) was tested for site index prediction. The accuracy obtained for the GLM models (not reported in the current study) was similar to that of the MLR models. As GAM is by default an extension of GLMs (Hastie and Tibshirani, 1990), we expect a GAM model (e.g., with a Gaussian link function) to yield similar results as those obtained for the MLR models. We postulate that a parametric linear model that adequately describes the population of interest may reasonably extrapolate beyond the range of the calibration data (Penner et al., 2013). Similar to previous reports, a well-formulated linear model, even with a relatively small sample size, can result in accurate and unbiased predictions of attributes from LiDAR (Næsset et al., 2005; Bontemps and Bouriaud, 2014; Tompalski et al., 2021).

The prediction accuracy obtained was similar to that reported for SIH prediction using ALS data (Noordemeer et al., 2018; Antón-Fernández et al., 2023). The reliability of the developed MLR-SIS functions was further strengthened by the evaluations of independent data not used for constructing the functions. Similar accuracies to that of the calibration data (i.e. permanent NFI sample plots) were obtained for the temporary plots of the Swedish NFI and stand-level data from the State-owned forest company Sveaskog (Table 5). The conclusion from the evaluations suggests that the functions are sound and reliable and can be used for site index determination in Swedish forests.

4.2. Sources of uncertainty in site index predictions

Because our model was linear, and the accuracy of the prediction had been studied mainly using the residual sum of squares, we used the adjusted multiple correlation coefficients (R^2_{adj}) as a measure of explained variation. The R^2_{adj} values obtained were 63 % for pine and 88 % for spruce, suggesting that a large part of the variance in SIS can be explained by the remotely sensed predictor variables. The estimated coefficients of the predictor variables had both biological and logical meanings (Table 3). For example, the size variables (Eq. 1.4) described by mean height $\left(H_{p90.t2}\right)$ and annual height increment $\left(\Delta H_{p95}\right)$ were all significant and positive. This suggests that for a given species and height, an increase in height growth may correspond to a larger site index (Assmann, 1970). The ALS height metrics H_{p90} and H_{p95} strongly correlate with stand basal area and volume (Fig. A6) and are used in forest attribute prediction models in Sweden (Nilsson et al., 2017). As the computation of the site index using the SIS-method does not require information about the dominant or top height, the use of an average height generally implies that SIS is sensitive to stand density. However, the plots of residuals from the MLR models against field-measured stand basal area (on NFI plots) indicated no obvious trends (Fig. A5). This probably suggests that the density effect on SIS may be captured in the models by the H_{p90} metric describing average stand height. Similarly, the model residuals did not show any apparent trends over standing volume (Fig. A5), suggesting that the inclusion of an increment parameter (i.e. height growth rate - ΔH_{p95}) adequately described the differences in volume production among sites along the latitudinal gradient (north to south) of Sweden. Those variables describing site characteristics (Eq. 1.2) were negative, and their effects were greater for Norway spruce than for Scots pine.

The predictor variables and the functional form of the final MLR models imply that the site index predictions are biologically and statistically sound. However, it is important to highlight that there are more factors influencing site index than those included in our models. For example, we found an overestimation of the site index in peat and soils with bare rocks (Fig. 3). In the MLR models, we used only ALS data and other available remotely sensed data as predictors. Currently, spatial information on soil properties is lacking in Sweden. We acknowledge that if there exists any geographical information on the soil type with sufficient accuracy, including such variables may increase the accuracy of the final models. Also, those predictor variables included in the final model (Eq. 9) present additional sources of errors that might affect the residual sum of squares in the regression analysis. Here, these errors are not evaluated wholly in the present study, but we discuss their sources and potential effects on prediction accuracy and recommendations for further studies.

The use of multiple data sources for site index prediction presents many sources of error, which may propagate in unknown ways (Nyström and Ståhl, 2001). Generally, uncertainty in model predictions is influenced by (1) errors in the dependent variable, (2) model misspecification, (3) values of the predictor variables, (4) model parameter estimates and (5) the residual variability around model predictions (Nyström and Ståhl, 2001; McRoberts and Westfall, 2016). In this study, the uncertainty analysis (Eq. (8) and Figs. 5 and 6) mainly involved the latter two sources of errors, which might represent a conservative estimate of the total uncertainty in the predicted SIS.

Model misspecification results from the lack of appropriate data for model calibration (McRoberts and Westfall, 2016) and the form or structure of the model fit (Næsset et al., 2005). However, we assume these to have negligible influence since the calibration data were obtained from representative ground plot measurements. Generally, the Swedish NFI is a probabilistic survey design with well-spread samples and covers large gradients of growth and climatic conditions. This guarantees an unbiased sample of the Swedish forest tree populations (Fridman et al., 2014). Thus, it is likely that the full range of variation in forest site productivity was captured during the calibration of the models.

Errors in the values of the predictor variables may stem from, for example, the geolocation of sample plots, soil moisture map and ALS data. Geolocation errors between the field plot centres and ALS data can be small due to the use of enhanced GPS receivers (Garmin GPSMAP 64), which have a horizontal accuracy of about 5 m during the field inventory (Nilsson et al., 2017). At the level of a large area inventory such as the Swedish NFI, such an accuracy of 5 m in geolocation can be highly substantial. An earlier evaluation by Nilsson et al. (2017) showed that at this accuracy level, good correspondence can be observed between the NFI plots and ALS data in Sweden. The soil moisture map has poor accuracy near roads, at sites on coarse sediments and in areas with steep local topography (Ågren et al., 2021). When compared with soil moisture information from the NFI sample plots, we observed trends in the model residuals, with large errors in the predicted SIS at dry and wet sites (Fig. 3). Probably, this effect can be minimised by examining the effect of the slope in and around the pixels together with the moisture index.

Similarly, we used bi-temporal ALS data for the prediction of the site index, and the possibility of error introduction in the models cannot be ruled out. Such errors may be influenced by the scanner brands and scanner configurations in the two ALS campaigns, as well as the size of the grid cell for height computation. In the first scan, ALS data were acquired using 13 different scanners from Leica (72.3 % of the scanned blocks), Optech (24.9 %), and Reigl and Trimble (2.8 %). Previous studies showed that the sensor effect is minimised largely by doing block modelling using only reference data scanned with the same scanner brand, acquired during the same season (leaf-on or leaf-off), and from a limited geographic area (Nilsson et al., 2017). We recommend such an approach to future studies on SIS modelling where separate model parameters can be estimated for the different scanner brands. Though SIS is determined on a radius of 10 m for both permanent and temporary plots, the determination of height from ALS data and subsequent predictions of site index were done on 12.5 × 12.5 m pixels. This was to ensure that the resolution of the SIS map data is compatible with other derived forest data products at the national scale. Again, from previous research (Nilsson et al. 2017), it was seen that the gain in precision (relative RMSE, %) of height estimation by increasing the grid-cell size was small, i.e., 7.4 % in 12.5 × 12.5 m to 7.3 % for 20 × 20 m grid-cell sizes. Thus, the chosen 12.5 m cell size can be regarded as optimal for SIS prediction in Swedish forests dominated by pine and spruce.

4.3. Application of the site index maps

So far, the models developed are species-specific SIS functions. From these models, generic SIS maps were produced separately for Scots pine and Norway spruce. At the application, it is expected that users choose the most suitable map according to the given tree species information, as tree species information derived from ALS data at the national scale is presently unavailable in Sweden.

An important part of this study was to make the map data available to authorities, forest owners, forest companies, and the general public. Already, map data (at a spatial resolution of 12.5 m) on other forest attributes, such as basal area, mean height, standing volume, and biomass, are available for use in Sweden (Nilsson et al., 2017). The addition of the SIS maps may provide additional benefits. For example, in forest planning, it is seen that 3D information obtained from ALS data is imperative for modelling ecosystem processes using dynamic treatment units optimised at the level of grid cells. Spatial information from SIS may improve the cell-level optimisations rather than the current approach of using a single value obtained from stand registers (Wilhelmsson, 2023).

The accompanying estimates of uncertainty in the cell-level predictions of SIS may also be valuable for large forest companies when handling uncertainties in forest information during the hierarchical planning process (Ulvdal et al., 2023). Predictions of SIS can facilitate site-adapted forest management on the one hand and, on the other hand, can enhance efficient sampling strategy by providing extra information on site productivity when selecting representative field samples during the design phase (Grafström et al., 2014).

An important question is how the product can be updated in the future. It is not clear how often the nationwide ALS campaigns may be carried out in the future. Possible options might be to use RADAR and digital photogrammetry data. Already, Persson and Fransson (2016) showed that TanDEM-X data could be used to predict site index with satisfactory accuracy (RMSE = 4.4 m and relative RMSE = 12.1 %) in Sweden. Bohlin et al. (2017) also reported that point clouds derived from the national image program could be used for large-area forest attribute mapping. The above data sources may be combined with older ALS data within the framework of data assimilation to concurrently improve the predictions of and update the SIS map product (Lindgren et al., 2022).

5. Conclusion

This study evaluated the potential of bi-temporal ALS data and other remotely sensed data for site productivity determination by site index according to site factors (SIS) in even-aged coniferous stands (Scots pine and Norway spruce) on productive forestlands in Sweden. This was approached by utilising field observations from the Swedish NFI, bi-temporal ALS data from the National Land Survey, and other auxiliary data describing site conditions (e.g., soil moisture, distance to coast and altitude above sea level). Two techniques of regression analysis were evaluated for site index prediction, a general multiple linear regression (MLR) and a non-parametric random forests (RF) model.

The results obtained showed that site index in even-aged coniferous stands could be determined by a sufficiently reasonable degree of accuracy (1.96 m for Scots pine and 2.11 m for Norway spruce) using bi-temporal ALS data and the MLR model. Model validation on independent data (i.e., data not used for model construction) gave similar accuracy as those obtained during model fitting. Final predictors of site index include metrics of 90th percentile height (at time two) and annual increment in the 95th percentile height, altitude, distance to coast, and soil moisture. Evaluations of the final models revealed that the residual variations were homoscedastic for the predictions of site index as well as for wider ranges of density (stand basal area) and productivity (stem volume).

Further, a map of site index predictions and the corresponding pixel-level prediction errors was produced at a spatial resolution of 12.5 m separately for stands of Scots pine and Norway spruce. The SIS maps have a potentially high value for operational forest management planning. The predictor variables may also be beneficial for the estimation of age and yield capacity in Swedish forests.

CRediT authorship contribution statement

Alex Appiah Mensah: Conceptualization, Methodology, Data curation, Formal analysis, Software, Writing – original draft, Writing – review & editing. **Jonas Jonzén:** Methodology, Data curation, Software, Validation, Writing – original draft, Writing – review & editing. **Kenneth Nyström:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Jörgen Wallerman:** Writing – original draft, Writing – review & editing. **Mats Nilsson:** Conceptualization, Methodology, Data curation, Writing – original draft, Writing – review & editing, Project administration, Funding acquisition, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We thank the two anonymous reviewers for the constructive comments that greatly improved the earlier version of the manuscript. This study was financially supported by MISTRA Digital Forest. Field data was provided by the Swedish National Forest Inventory and the State-owned forest company Sveaskog.

Appendix

Table A.1
F-test for comparing reduced (Eqs. 1.2 – 1.4) and full (Eq. 1.1) models of site index fitted by multiple linear regression (MLR) for Scots pine.

*Predictor group	Model	DF_{res}	RSS	DF_{diff}	SSQ_{diff}	F	Pr(>F)
Site	M_{red}	2238	6298.9				
	M_{full}	2234	3482	4	2816.9	451.8	< 0.001
Resource	M_{red}	2239	8162.7				
	M_{full}	2234	3482	5	4680.7	600.6	< 0.001
Size	M_{red}	2239	5634.1				
	M_{full}	2234	3482	5	2152.1	276.2	< 0.001

Variable definitions: M_{red} is reduced model (i.e. contains only variables for a specific predictor group), M_{full} denotes the full model (i.e. contains all predictor variables of the reduced model and other predictors), DF_{res} is residual degrees of freedom, *RSS* is the residual sum of squares, DF_{diff} is the degrees of freedom associated with the difference in DF_{res} for the reduced and full models, SSQ_{diff} is the sum of squares for the difference between *RSS* of the full and reduced models, and *Pr(>F)* denotes the significance of the calculated F-statistic at an alpha level of 0.05.
* See Eqs. (1.2) – (1.4) for definitions of the predictor groups.

Table A.2
F-test for comparing reduced (Eqs. 1.2 – 1.4) and full (Eq. 1.1) models of site index fitted by multiple linear regression (MLR) for Norway spruce.

*Predictor group	Model	DF_{res}	RSS	DF_{diff}	SSQ_{diff}	F	Pr(>F)
Site	M_{red}	2073	4348.2				
	M_{full}	2069	3384.8	4	963.4	147.2	< 0.001
Resource	M_{red}	2074	22925.5				
	M_{full}	2069	3384.8	5	19,541	2388.9	< 0.001
Size	M_{red}	2074	20135.2				
	M_{full}	2069	3384.8	5	16,720	2047.8	< 0.001

Variable definitions: M_{red} is reduced model (i.e. contains only variables for a specific predictor group), M_{full} denotes the full model (i.e. contains all predictor variables of the reduced model and other predictors), DF_{res} is residual degrees of freedom, *RSS* is the residual sum of squares, DF_{diff} is the degrees of freedom associated with the difference in DF_{res} for the reduced and full models, SSQ_{diff} is the sum of squares for the difference between *RSS* of the full and reduced models, and *Pr(>F)* denotes the significance of the calculated F-statistic at an alpha level of 0.05.
* See Eqs. (1.2) – (1.4) for definitions of the predictor groups.

Table A.3
Estimated covariance matrix of the MLR site index model parameters.

	Scots pine							
	$\widehat{\alpha}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\omega}_1$	$\widehat{\omega}_2$	$\widehat{\gamma}_1$	$\widehat{\gamma}_2$
$\widehat{\alpha}_0$	1.024							
$\widehat{\beta}_1$	$-1.43\text{x}10^{-7}$	$2.04\text{x}10^{-14}$						
$\widehat{\beta}_2$	$4.23\text{x}10^{-5}$	$-1.03\text{x}10^{-11}$	$1.46\text{x}10^{-7}$					
$\widehat{\beta}_3$	$-5.67\text{x}10^{-5}$	$8.99\text{x}10^{-12}$	$-9.72\text{x}10^{-8}$	$2.67\text{x}10^{-7}$				
$\widehat{\omega}_1$	$-1.7\text{x}10^{-4}$	$1.87\text{x}10^{-11}$	$2.68\text{x}10^{-8}$	$-1.27\text{x}10^{-8}$	$1.26\text{x}10^{-7}$			
$\widehat{\omega}_2$	-0.132	$1.39\text{x}10^{-8}$	$-1.09\text{x}10^{-5}$	$2.52\text{x}10^{-6}$	$6.81\text{x}10^{-5}$	0.239		
$\widehat{\gamma}_1$	$6.92\text{x}10^{-4}$	$-1.25\text{x}10^{-10}$	$8.06\text{x}10^{-8}$	$-1.35\text{x}10^{-7}$	$-1.08\text{x}10^{-7}$	$-3.1\text{x}10^{-4}$	$2.49\text{x}10^{-5}$	
$\widehat{\gamma}_2$	$-9.73\text{x}10^{-6}$	$1.55\text{x}10^{-12}$	$-4.98\text{x}10^{-10}$	$1.02\text{x}10^{-9}$	$2.36\text{x}10^{-9}$	$5.22\text{x}10^{-6}$	$-2.71\text{x}10^{-7}$	$3.2\text{x}10^{-9}$
	Norway spruce							
	$\widehat{\alpha}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\omega}_1$	$\widehat{\omega}_2$	$\widehat{\gamma}_1$	$\widehat{\gamma}_2$
$\widehat{\alpha}_0$	1.257							
$\widehat{\beta}_1$	$1.87\text{x}10^{-7}$	$2.84\text{x}10^{-14}$						
$\widehat{\beta}_2$	$1.5\text{x}10^{-4}$	$-2.88\text{x}10^{-11}$	$2.41\text{x}10^{-7}$					
$\widehat{\beta}_3$	$-3.18\text{x}10^{-5}$	$5.59\text{x}10^{-12}$	$-1.30\text{x}10^{-7}$	$3.57\text{x}10^{-7}$				
$\widehat{\omega}_1$	$-1.1\text{x}10^{-4}$	$1.12\text{x}10^{-11}$	$2.44\text{x}10^{-8}$	$-5.78\text{x}10^{-9}$	$9.51\text{x}10^{-8}$			
$\widehat{\omega}_2$	-0.124	$1.35\text{x}10^{-8}$	$1.43\text{x}10^{-5}$	$-9.34\text{x}10^{-6}$	$4.91\text{x}10^{-5}$	0.161		
$\widehat{\gamma}_1$	$4.2\text{x}10^{-4}$	$-1.01\text{x}10^{-10}$	$8.57\text{x}10^{-8}$	$-1.14\text{x}10^{-7}$	$-5.78\text{x}10^{-7}$	$-9.34\text{x}10^{-6}$	$2.91\text{x}10^{-5}$	
$\widehat{\gamma}_2$	$-3.12\text{x}10^{-6}$	$7.54\text{x}10^{-13}$	$-7.33\text{x}10^{-10}$	$3.85\text{x}10^{-10}$	$1.35\text{x}10^{-9}$	$2.59\text{x}10^{-6}$	$-3.11\text{x}10^{-7}$	$3.57\text{x}10^{-9}$

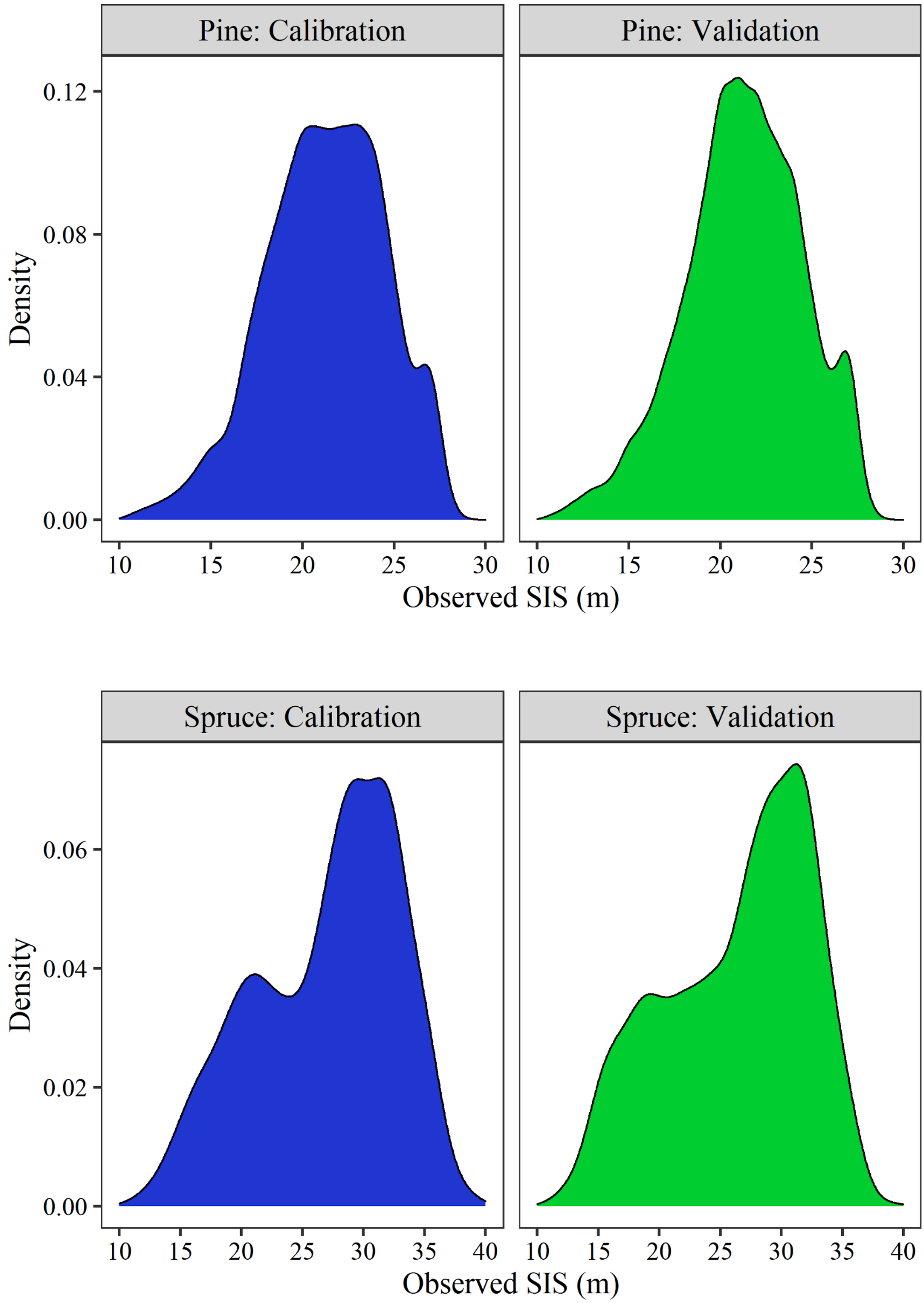

Fig. A1. Distribution of observed SIS on the NFI sample plots. Calibration refers to permanent sample plots, whereas validation denotes temporary sample plots.

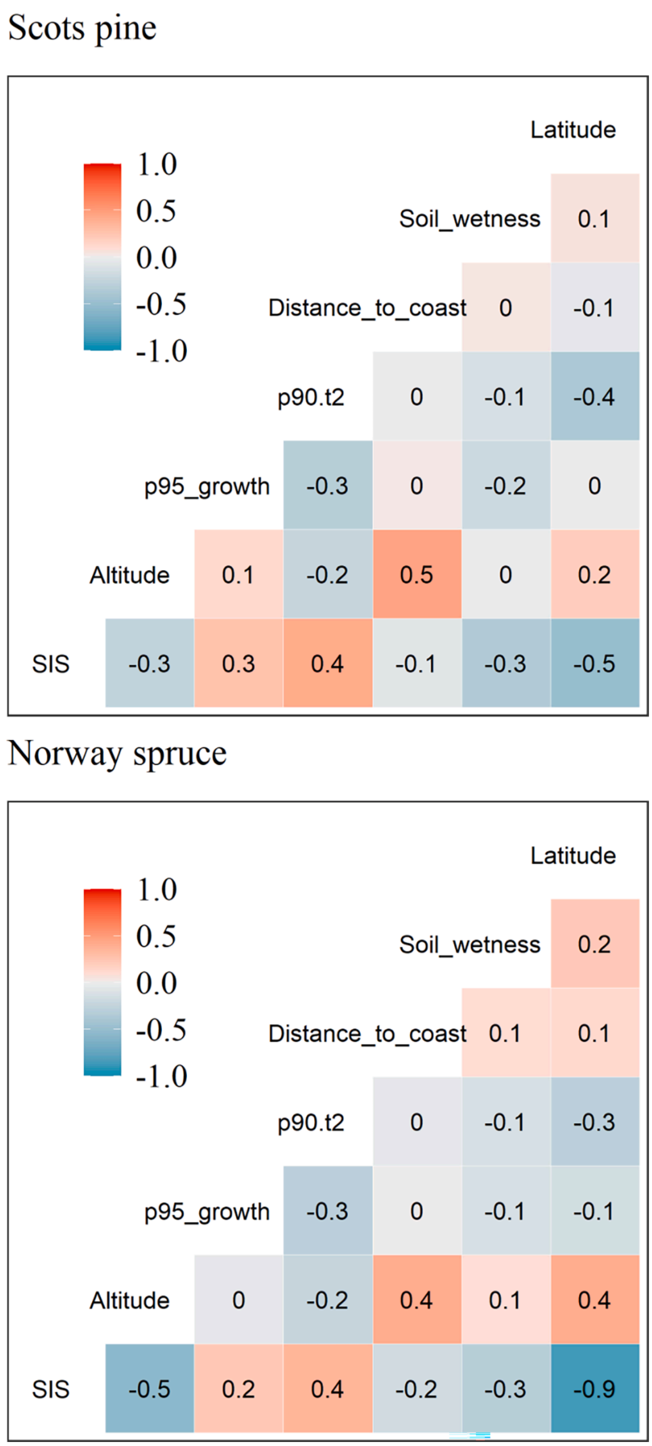

Fig. A2. Pearson's correlation among the final predictors of SIS.

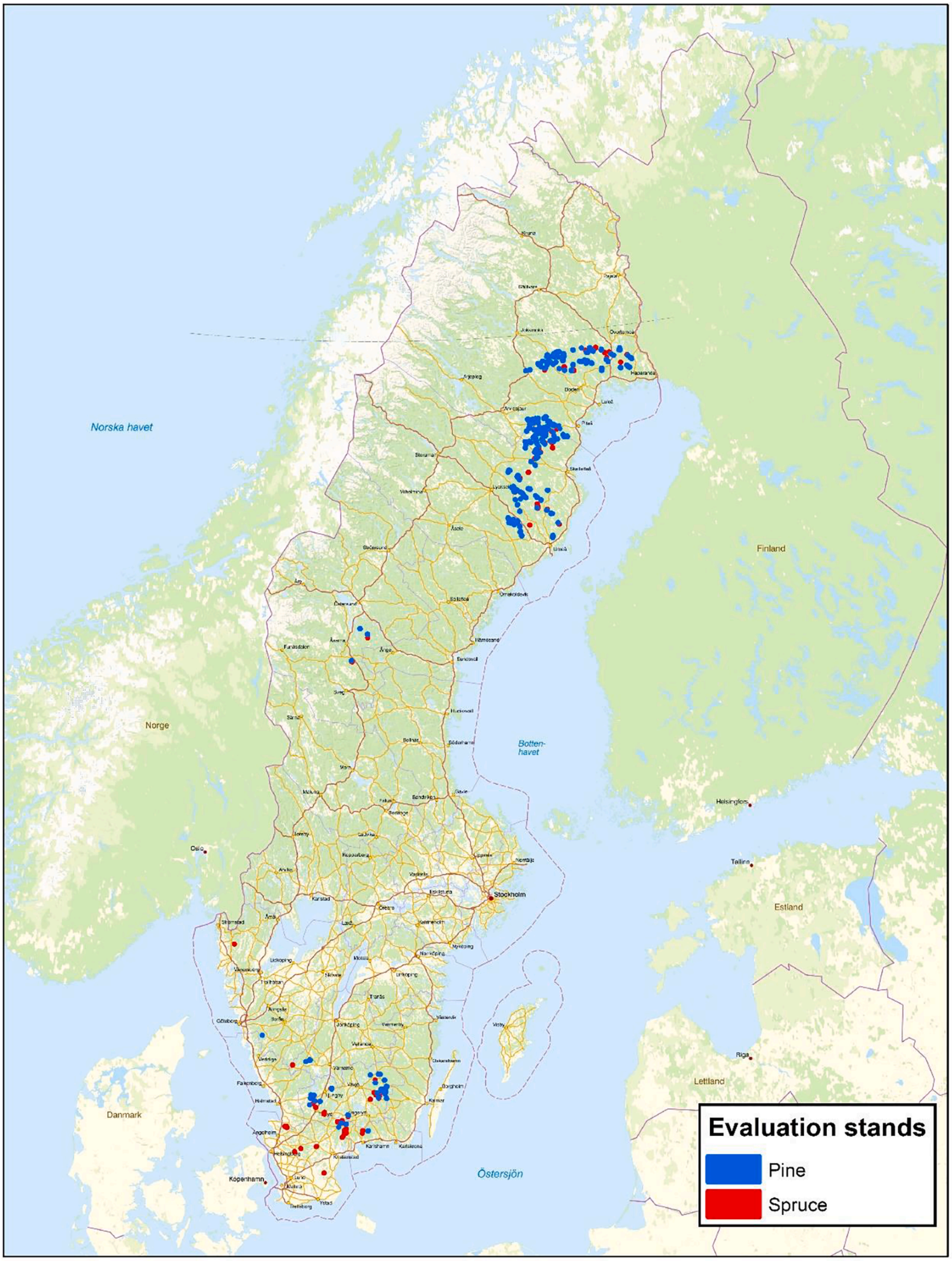

Fig. A3. Location of the forest stands (managed by Sveaskog) used for validation. Copyright Lantmäteriet Topographic Map.

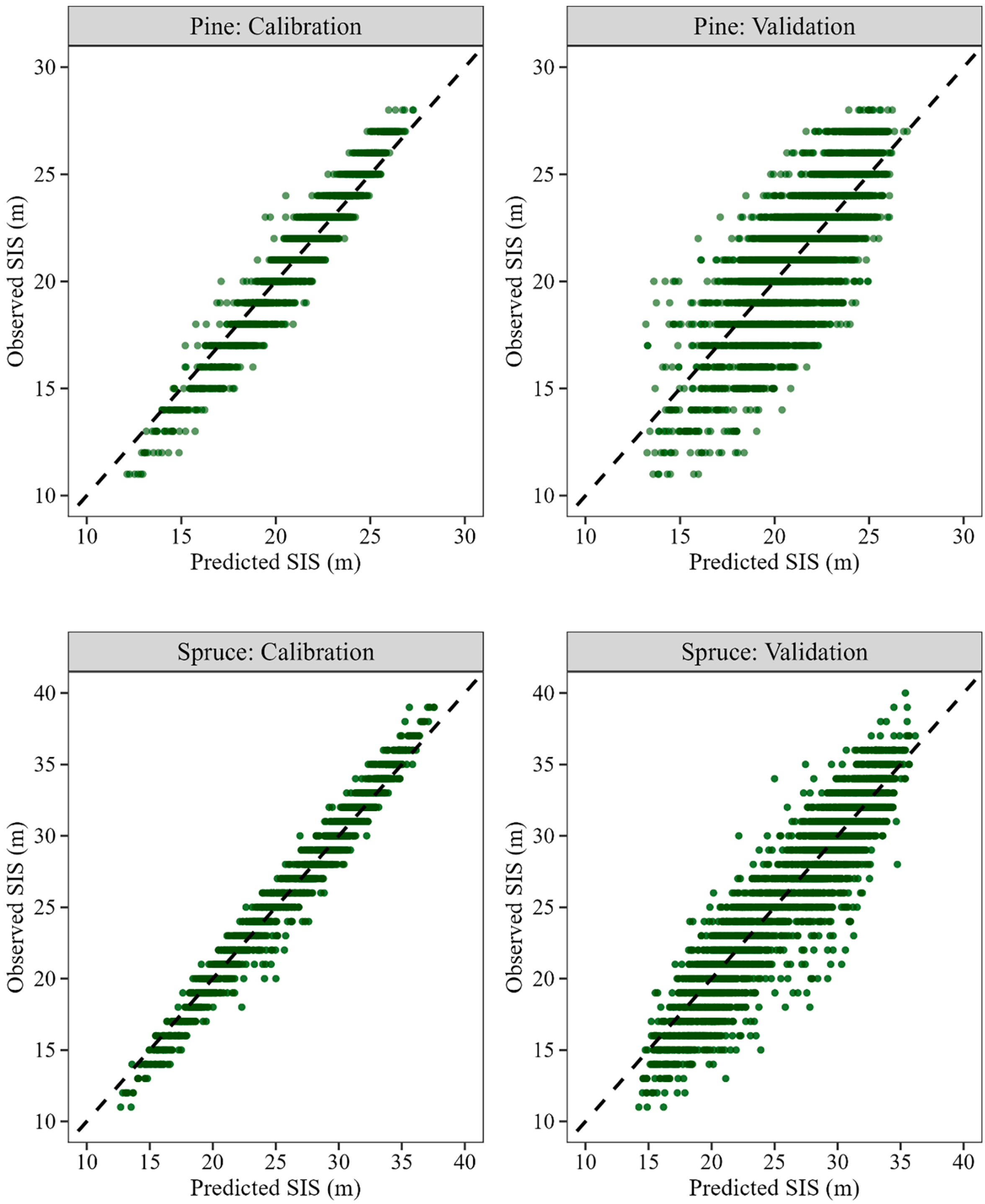

Fig. A4. Observed vs. predicted SIS (from the RF model) on the calibration (permanent plots) and validation (temporary plots) dataset for Scots pine and Norway spruce dominated stands.

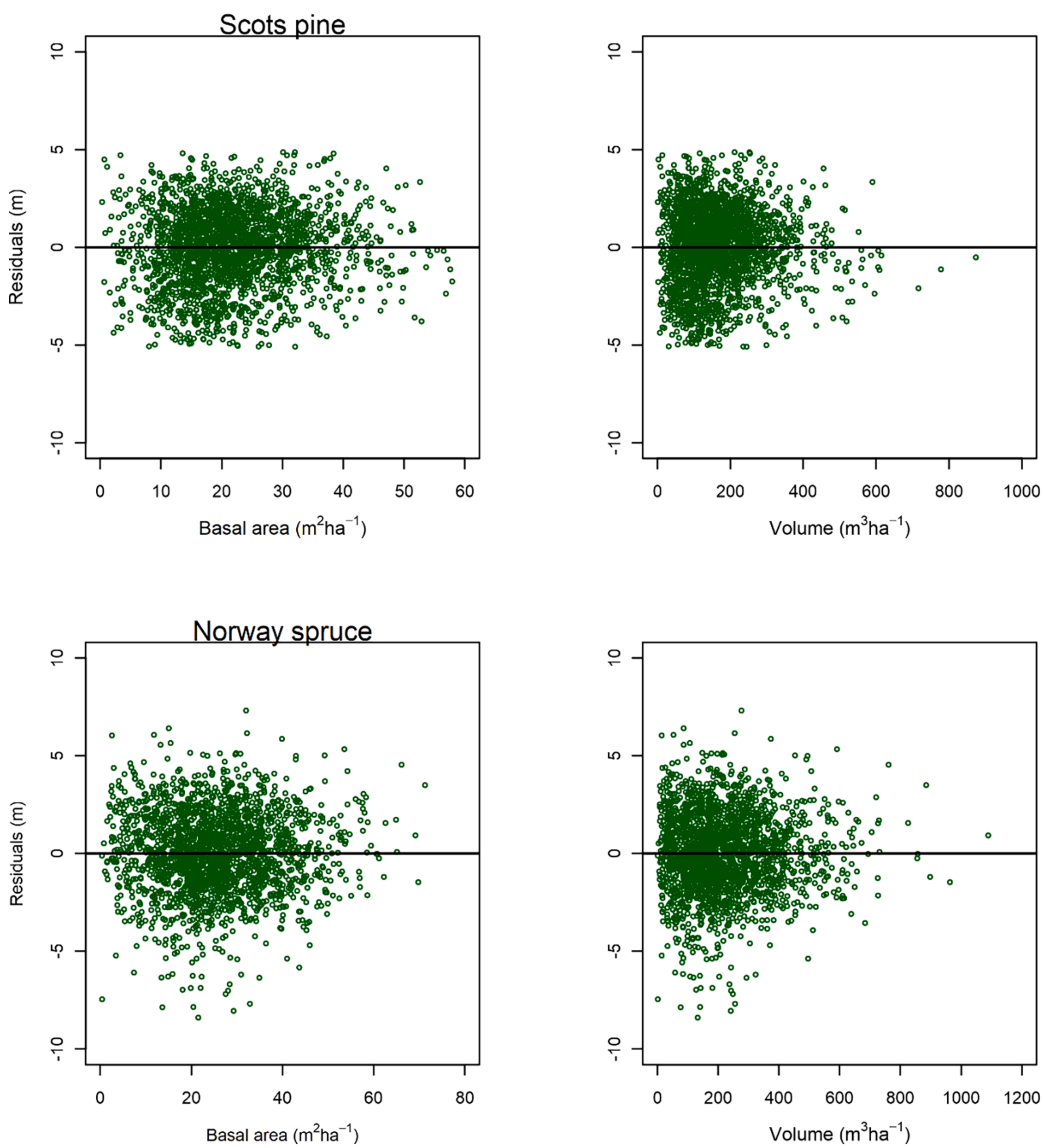

Fig. A5. Residuals vs. ground-measured forest stand attributes.

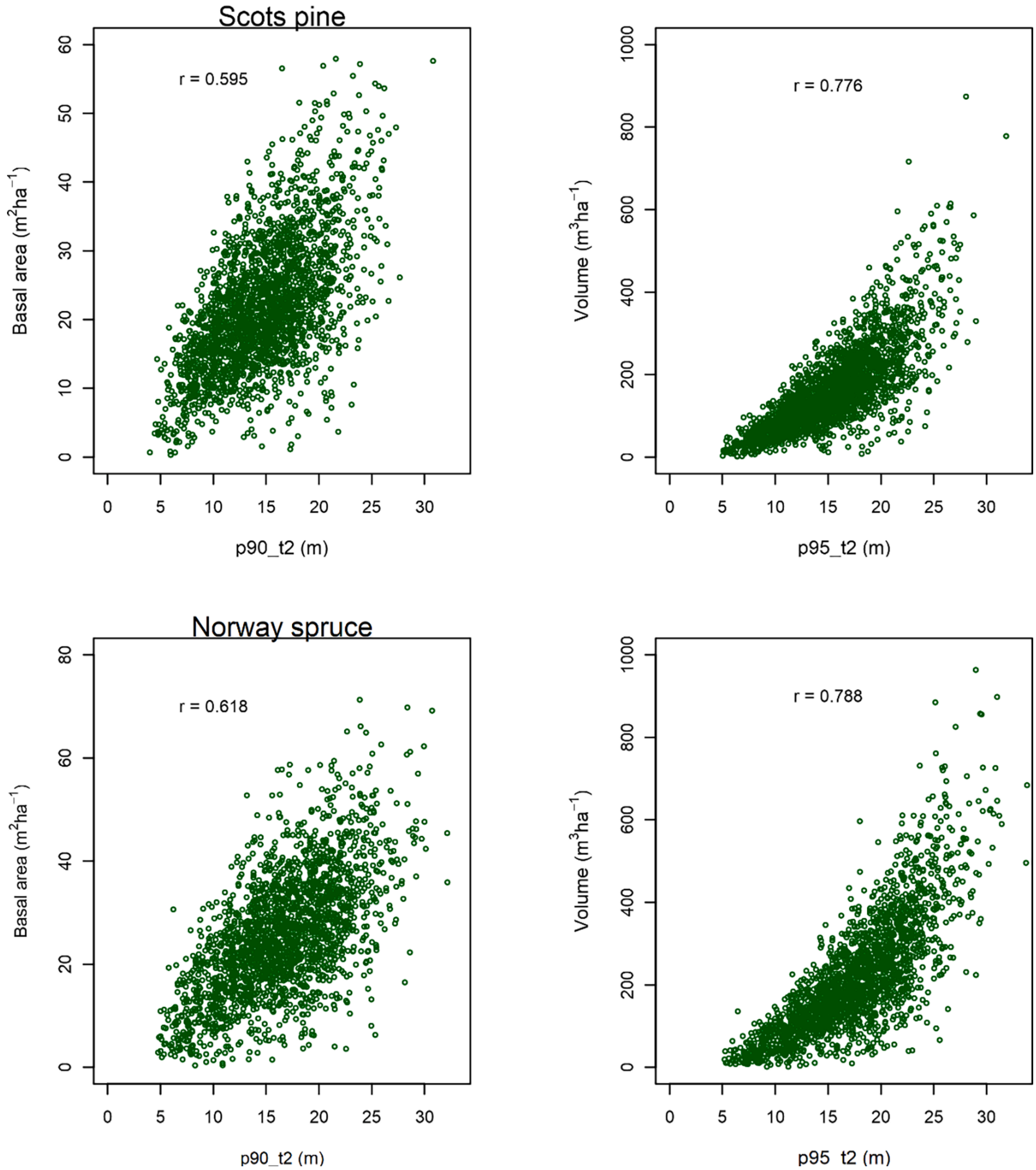

Fig. A6. Relationship between ground-measured stand attributes (basal area and volume) and ALS metrics of p90 and p95 at second scanning. The correlation value shows Spearman's rank correlation coefficient (r).

References

Aertsen, W., Kint, V., Van Orshoven, J., Özkan, K.K., Muys, B., 2010. Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. Ecol. Model. 221, 1119–1130. https://doi.org/10.1016/j.ecolmodel.2010.01.007.

Ågren, A.M., Larson, J., Paul, S.S., Laudon, H., Lidberg, W., 2021. Use of multiple LIDAR-derived digital terrain indices and machine learning for high-resolution national-scale soil moisture mapping of the Swedish forest landscape. Geoderma 404, 115280. https://doi.org/10.1016/j.geoderma.2021.115280.

Antón-Fernández, C., Mola-Yudego, B., Dalsgaard, L., Astrup, R., 2016. Climate-sensitive site index models for Norway. Can. J. For. Res. 46, 794–803. https://doi.org/10.1139/cjfr-2015-0155.

Antón-Fernández, C., Hauglin, M., Breidenbach, J., Astrup, R., 2023. Building a high-resolution site index map using boosted regression trees: the Norwegian case. Can. J. For. Res. 53 (6), 416–429.

Assmann, E., 1970. The principles of forest yield study. Pergamon Press, Oxford, New York, p. 506.

Bohlin, J., Bohlin, I., Jonzén, J., Nilsson, M., 2017. Mapping forest attributes using data from stereophotogrammetry of aerial images and field data from the national forest inventory. Silva Fennica 51.

Bontemps, J.-D., Bouriaud, O., 2014. Predictive approaches to forest site productivity: recent trends, challenges and future perspectives. Forest. Int. J. Forest Res. 87, 109–128. https://doi.org/10.1093/forestry/cpt034.

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Ekö, P.-M., Johansson, U., Petersson, N., Bergqvist, J., Elfving, B., Frisk, J., 2008. Current growth differences of Norway spruce (Picea abies), Scots pine (Pinus sylvestris) and birch (Betula pendula and Betula pubescens) in different regions in Sweden. Scand. J. Forest Res. 23, 307–318. https://doi.org/10.1080/02827580802249126.

Ekström, M., Nilsson, M., 2021. A Comparison of Model-Assisted Estimators, With and Without Data-Driven Transformations of Auxiliary Variables, With Application to Forest Inventory. Fronti. Forests Global Change 4.

Elfving, B., Nyström, K., 1996. The yield capacity of planted *Picea abies* in northern Sweden. Scand. J. Forest Res. 11, 38–49. https://doi.org/10.1080/02827589609382910.

Esteban, J., McRoberts, R.E., Fernández-Landa, A., Tomé, J.L., Næsset, E., 2019. Estimating Forest Volume and Biomass and Their Changes Using Random Forests and Remotely Sensed Data. Remote Sens. (Basel) 11, 1944. https://doi.org/10.3390/rs11161944.

Fridman, J., Wulff, S., Dahlgren, J., 2019. Resultat från Nyckelord: kontrolltaxering av Riksskogstaxeringens datainsamling 2012-2016. Umeå: Sveriges lantbruksuniversitet. (Arbetsrapport/Sveriges lantbruksuniversitet, Institutionen for skoglig resurshushållning, 500). [In Swedish with English summary]. Retrieved from https://pub.epsilon.slu.se/.

Fridman, J., Holm, S., Nilsson, M., Nilsson, P., Ringvall, A.H., Stahl, G., 2014. Adapting National Forest Inventories to changing requirements - the case of the Swedish National Forest Inventory at the turn of the 20th century. Silva. Fenn. 48, 1095. https://doi.org/10.14214/sf.1095.

Grafström, A., Saarela, S., Ene, L.T., 2014. Efficient sampling strategies for forest inventories by spreading the sample in auxiliary space. Can. J. For. Res. 44, 1156–1164. https://doi.org/10.1139/cjfr-2014-0202.

Grafström, A., Zhao, X., Nylander, M., Petersson, H., 2017. A new sampling strategy for forest inventories applied to the temporary clusters of the Swedish National Forest inventory. Can. J. For. Res. 47, 1161–1167. https://doi.org/10.1139/cjfr-2017-0095.

Guerra-Hernández, J., Arellano-Pérez, S., González-Ferreiro, E., Pascual, A., Sandoval Altelarrea, V., Ruiz-González, A.D., Álvarez-González, J.G., 2021. Developing a site index model for P. Pinaster stands in NW Spain by combining bi-temporal ALS data and environmental data. For. Ecol. Manage. 481, 118690 https://doi.org/10.1016/j.foreco.2020.118690.

Hägglund, B., Lundmark, J.E., 1977. Site index estimation by means of site properties. Scots pine and Norway spruce in Sweden, Studia Forestalia Suecica, p. 38.

Hägglund, B., 1981. Evaluation of forest site productivity. Presented at the For Abstr, pp. 515–527.

Hastie, T.J., Tibshirani, R.J., 1990. Generalised Additive Models. Chapman and Hall, London.

Kandare, K., Ørka, H.O., Dalponte, M., Næsset, E., Gobakken, T., 2017. Individual tree crown approach for predicting site index in boreal forests using airborne laser scanning and hyperspectral data. Int. J. Appl. Earth Obs. Geoinf. 60, 72–82. https://doi.org/10.1016/j.jag.2017.04.008.

Kangas, A., Myllymäki, M., Mehtätalo, L., 2023. Understanding uncertainty in forest resources maps. *Silva Fennica*. https://doi.org/10.14214/sf.22026.

Langlet, O., 1936. Studier över tallens fysiologiska variabilitet och dess samband med klimatet. Medd. från statens skogsförsöksanstalt, vol. 29, no 16. (In Swedish).

Lindahl, K.B., Sténs, A., Sandström, C., Johansson, J., Lidskog, R., Ranius, T., Roberge, J.-M., 2017. The Swedish forestry model: More of everything? Forest Policy and Economics, Alternative Pathways to Sustainability? Comparing Forest Governance Mod. 77, 44–55. https://doi.org/10.1016/j.forpol.2015.10.012.

Lindgren, N., Olsson, H., Nyström, K., Nyström, M., Ståhl, G., 2022. Data Assimilation of Growing Stock Volume Using a Sequence of Remote Sensing Data from Different Sensors. Can. J. Remote. Sens. 48, 127–143. https://doi.org/10.1080/07038992.2021.1988542.

Lundmark, J.-E., 1974. Use of site properties for assessing site index in stands of Scots pine and Norway spruce. Inst. för växtekologi och marklära, skogshögskolan. Rapporter uppsatser 33, 243–253. In Swedish.

Maltamo, M., Næsset, E., Vauhkonen, J. (Eds.), 2014. Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies, Managing Forest Ecosystems. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-94-017-8663-8.

McGaughey, R., 2015. FUSION/LDV: software for LIDAR data analysis and visualization. v. 3.3. Washington: USDA/Forest Service. http://forsys.cfr.washington.edu/fusion/.

McRoberts, R.E., Westfall, J.A., 2016. Propagating uncertainty through individual tree volume model predictions to large-area volume estimates. Ann. For. Sci. 73, 625–633. https://doi.org/10.1007/s13595-015-0473-x.

Mensah, A.A., Holmström, E., Nyström, K., Nilsson, U., 2022. Modelling potential yield capacity in conifers using Swedish long-term experiments. Forest Ecology and Management 512, 120162. Doi: 10.1016/j.foreco.2022.120162. Doi: 10.1093/forestscience/36.1.1.

Mensah, A.A., Petersson, H., Dahlgren, J., Elfving, B., 2023. Taller and slenderer trees in Swedish forests according to data from the National Forest Inventory. For. Ecol. Manage. 527, 120605 https://doi.org/10.1016/j.foreco.2022.120605.

Moan, M. Å., Noordermeer, L., White, J. C., Coops, N. C., Bollandsås, O. M., 2023. Detecting and excluding disturbed forest areas improves site index determination using bi-temporal airborne laser scanner data. *Forestry: Int. J. Forest Res.*, cpad025. https://doi.org/10.1093/forestry/cpad025.

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. https://doi.org/10.1016/S0034-4257(01)00290-5.

Næsset, E., Bollandsås, O.M., Gobakken, T., 2005. Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data. Remote Sens. Environ. 94, 541–553. https://doi.org/10.1016/j.rse.2004.11.010.

Nilsson, M., 1996. Estimation of tree heights and stand volume using an airborne lidar system. Remote Sens. Environ. 56 (1), 1–7. https://doi.org/10.1016/0034-4257(95)00224-3.

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. https://doi.org/10.1016/j.rse.2016.10.022.

Nilsson, P., 2021. Skogsdata (Official Forest Statistics of Sweden): aktuella uppgifter om de svenska skogarna från Riksskogstaxeringen. Tema: Den döda veden. Institutionen för skoglig resurshushållning, Umeå. Swedish University of Agricultural Sciences. (In Swedish with English summary).

Noordermeer, L., Bollandsås, O.M., 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. Manage. 428, 104–114. https://doi.org/10.1016/j.foreco.2018.06.041.

Nord-Larsen, T., Schumacher, J., 2012. Estimation of forest resources from a country wide laser scanning survey and national forest inventory data. Remote Sens. Environ. 119, 148–157. https://doi.org/10.1016/j.rse.2011.12.022.

Nothdurft, A., Wolf, T., Ringeler, A., Böhner, J., Saborowski, J., 2012. Spatio-temporal prediction of site index based on forest inventories and climate change scenarios. For. Ecol. Manage. 279, 97–111. https://doi.org/10.1016/j.foreco.2012.05.018.

Nyström, K., Ståhl, G., 2001. Forecasting probability distributions of forest yield allowing for a Bayesian approach to management planning. Silva Fennica 35, 185–201.

Peng, C., 2000. Growth and yield models for uneven-aged stands: past, present and future. For. Ecol. Manage. 132, 259–279. https://doi.org/10.1016/S0378-1127(99)00229-7.

Penner, M., Pitt, D.G., Woods, M.E., 2013. Parametric vs. nonparametric LiDAR models for operational forest inventory in boreal Ontario. Can. J. Remote. Sens. 39, 426–443. https://doi.org/10.5589/m13-049.

Persson, H.J., Fransson, J.E.S., 2016. Estimating Site Index From Short-Term TanDEM-X Canopy Height Models. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9, 3598–3606. https://doi.org/10.1109/JSTARS.2016.2563158.

Persson, H.J., Ståhl, G., 2020. Characterizing Uncertainty in Forest Remote Sensing Studies. Remote Sens. (Basel) 12, 505. https://doi.org/10.3390/rs12030505.

R Core Team., 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/.

Ranneby, B., Cruse, T., Hägglund, B., Jonasson, H., Swärd, J., 1987. Designing a new national forest survey for Sweden. Studia Forestalia Suecica Report No. 177, 1–29. https://pub.epsilon.slu.se/4634/.

Riofrío, J., White, J.C., Tompalski, P., Coops, N.C., Wulder, M.A., 2023. Modelling height growth of temperate mixed wood forests using an age-independent approach and multi-temporal airborne laser scanning data. For. Ecol. Manage. 543, 121137 https://doi.org/10.1016/j.foreco.2023.121137.

Saarela, S., Wästlund, A., Holmström, E., Mensah, A.A., Holm, S., Nilsson, M., Fridman, J., Ståhl, G., 2020. Mapping aboveground biomass and its prediction uncertainty using LiDAR and field data, accounting for tree-level allometric and LiDAR model errors. For. Ecosyst. 7, 43. https://doi.org/10.1186/s40663-020-00245-0.

Sabatia, C.O., Burkhart, H.E., 2014. Predicting site index of plantation loblolly pine from biophysical variables. For. Ecol. Manage. 326, 142–156. https://doi.org/10.1016/j.foreco.2014.04.019.

Sharma, R.P., 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. Manage. 262 (12), 2162–2175. https://doi.org/10.1016/j.foreco.2011.07.037.

Skovsgaard, J.P., Vanclay, J.K., 2008. Forest site productivity: a review of the evolution of dendrometric concepts for even-aged stands. Forestry 81, 13–31. https://doi.org/10.1093/forestry/cpm041.

SMHI. Swedish Meteorological and Hydrological Institute, 2018. *Data*. http://www.smhi.se/klimatdata.

Socha, J., Pierzchalski, M., Bałazy, R., Ciesielski, M., 2017. Modelling top height growth and site index using repeated laser scanning data. For. Ecol. Manage. 406, 307–317. https://doi.org/10.1016/j.foreco.2017.09.039.

Solberg, S., Kvaalen, H., Puliti, S., 2019. Age-independent site index mapping with repeated single-tree airborne laser scanning. Scand. J. For. Res. 34 (8), 763–770. https://doi.org/10.1080/02827581.2019.1616814.

Tompalski, P., Coops, N.C., White, J.C., Wulder, M.A., 2015. Augmenting Site Index Estimation with Airborne Laser Scanning Data. For. Sci. 61, 861–873. https://doi.org/10.5849/forsci.14-175.

Tompalski, P., Coops, N.C., White, J.C., Goodbody, T.R.H., Hennigar, C.R., Wulder, M.A., Socha, J., Woods, M.E., 2021. Estimating Changes in Forest Attributes and Enhancing Growth Projections: A Review of Existing Approaches and Future Directions Using Airborne 3D Point Cloud Data. Curr Forestry Rep 7, 1–24. https://doi.org/10.1007/s40725-021-00135-w.

Troedsson, T., 1965. The hydrology of forest soils. Kungl Skogs- och Lantbruksakademiens Tidskrift 104, 142–154.

Ulvdal, P., Öhman, K., Eriksson, L.O., Wästerlund, D.S., Lämås, T., 2023. Handling uncertainties in forest information: the hierarchical forest planning process and its use of information at large forest companies. Forestry: An International Journal of Forest Research 96, 62–75. https://doi.org/10.1093/forestry/cpac028.

Watt, M.S., Palmer, D.J., Leonardo, E.M.C., Bombrun, M., 2021. Use of advanced modelling methods to estimate radiata pine productivity indices. For. Ecol. Manage. 479, 118557 https://doi.org/10.1016/j.foreco.2020.118557.

Wikström, P., Edenius, L., Elfving, B., Eriksson, L.O., Lämås, T., Sonesson, J., Öhman, K., Wallerman, J., Waller, C., Klintebäck, F., 2011. The Heureka forestry decision support system: an overview. Math. Comput. For. Nat. -Res. Sci. 3, 87–94.

Wilhelmsson, P., 2023. Forest planning utilizing high spatial resolution data. Acta Universitatis Agriculturae Sueciae. https://doi.org/10.54612/a.4h25q0pofl.

Zhao, K., Suarez, J.C., Garcia, M., Hu, T., Wang, C., Londo, A., 2018. Utility of multitemporal lidar for forest and carbon monitoring: Tree growth, biomass dynamics, and carbon flux. Remote Sens. Environ. 204, 883–897. https://doi.org/10.1016/j.rse.2017.09.007.