Canadian Journal of Forest Research

OPEN ACCESS | Article

Building a high-resolution site index map using boosted regression trees: the Norwegian case

Clara Antón-Fernández, Marius Hauglin, Johannes Breidenbach, and Rasmus Astrup

Division of Forestry and Forest Resources, Norwegian Institute of Bioeconomy Research, P.O. Box 115, NO-1431 Ås, Norway

Corresponding author: **Clara Antón-Fernández** (email: clara.anton.fernandez@nibio.no)

Abstract

Accurate estimation of site productivity is essential for forest projections and scenario modelling. We present and evaluate models to predict site index (SI) and whether a site is productive (potential total stem volume production $\geq$ 1 $m^3 \cdot ha^{-1} \cdot year^{-1}$) in a wall-to-wall high-resolution (16 m $\times$ 16 m) SI map for Norway. We investigate whether remotely sensed data improve predictions. We also study the advantages and disadvantages of using boosted regression trees (BRT), a machine-learning algorithm, to create high-accuracy SI maps. We use climatic and topographical data, soil parent material, a land resource map, and depth to water, together with Sentinel-2 satellite images and airborne laser scanning metrics, as predictor variables. We use the SI observed at more than 10 000 National Forest Inventory (NFI) sample plots throughout Norway to fit BRT models and validate the models using 5822 independent temporary plots from the NFI. We benchmark our results against SI estimates from forest monitoring inventories. We find that the SI from BRT has root mean squared error (RMSE) ranging from 2.3 m (hardwoods) to 3.6 m (spruce) when tested against independent validation data from the NFI temporary plots. These RMSEs are similar or marginally better than an evaluation of SI estimates from operational forest management plans where SI normally stems from manual photo interpretation.

Key words: machine learning, BRT, ALS, Norway spruce, Scots pine

Introduction

Site productivity is a key variable in most forest management decisions such as species selection, stand density, or number and timing of thinnings. Forest site productivity is the production that can be realized at a certain site with a given genotype and a specified management regime (Skovsgaard and Vanclay 2008). Site productivity is, therefore, the combined effect of climate, soil quality, and management on the growth of a fixed genotype on a site. The most widely used indicator for site productivity is site index (SI), which is defined as the average height of the dominant trees of a site at a certain age. Note that SI is a proxy for site productivity, and expresses the potential height growth, whereas the productivity is also affected by additional factors, which SI is assumed to be independent of, such as density. SI as a proxy for productivity has many challenges in cases such as uneven-aged stands or mixed-species stands (Berrill and O'Hara 2014) but is still the most widely used measure of productivity in operational forest management in Fennoscandia and is used in the vast majority of operational forest management planning systems.

Operational forest management is usually based on information from forest management inventories (FMIs), which are spatially explicit inventories carried out for management of individual forest properties. Although SI from FMIs fulfill the need of the individual forest properties, they cannot be used to do planning for larger areas, e.g., municipality or region, since (1) FMIs are not available for all forest area (Lindgren et al. 2021), (2) they are not useful for distinguishing between unproductive and productive areas, since FMIs are only available for productive forest, (3) they cannot be used to estimate productivity of nonforested areas, (4) they estimate SI only for the current dominant species, and thus offer no information about SI of other currently or potentially coexisting species, which is crucial information for projections and scenario modelling. To overcome these drawbacks, a wall-to-wall map of productivity is needed.

There are several alternative methodologies to produce a wall-to-wall productivity map. Productivity could be estimated using canopy height growth estimated from bitemporal airborne laser scanning (ALS) data, or digital aerial photogrammetry (DAP) and existing SI curves (see Tompalski et al. (2021) for a review of ALS and DAP for estimating forest attributes, including SI). However, this approach shares most of the drawbacks described above: bitemporal ALS/DAP is not available for all forest areas, cannot be used to estimate productivity of nonforested areas, and it can only be used to estimate SI for the current dominant species. It also has the drawback of being hard to use if there has been some disturbance (harvest or mortality of dominant trees) in the period in between the laser acquisitions. Alternatively, SI can be estimated using proxy variables such as climate

(Weiskittel et al. 2011; González-Rodríguez and Diéguez-Aranda 2021), soil characteristics (Sabatia and Burkhart 2014; Antón-Fernández et al. 2016; Barrio-Anta et al. 2020), topographic variables (Huang et al. 2017), satellite data, and/or lithology (Watt et al. 2010; Rahimzadeh-Bajgiran et al. 2020). Since these proxy variables are usually available for large areas and not limited to productive or managed forest or by the species currently growing in the area, the SI derived from them can potentially overcome the drawbacks outlined above for the current FMIs.

While the use of biophysical variables and remotely sensed data for productivity mapping has a relatively long history (e.g., Latta et al. 2009), and bitemporal ALS data have been used to predict and map SI in operational forest inventories (Noordermeer et al. 2020), using ALS-derived variables, such as the height of the 95th percentile of the point cloud, to produce SI maps for large regions has not been, to our knowledge, yet tested. Height ALS-derived variables could potentially help improve the SI predictions (Tompalski et al. 2015), particularly in stands with high SI, since mature high-SI stands will reach heights that are unlikely to be reached by a low-SI stand, or in complex, heterogeneous forest stands (Tompalski et al. 2015). However, this could pose a problem in recently felled areas, since tree heights would be low but SI could be high, especially because stands with high SI are more likely to be felled (shorter rotations) than stands with low SI. This could lead to larger uncertainties in the estimated SI in recently felled areas and it could also lead to underprediction of SI. The same potential risk is applicable to other remotely sensed data, since the lack of vegetation after felling could bias the estimation of SI in those areas.

SI models based on biophysical variables have been developed using multiple statistical and machine-learning techniques (e.g., Monserud et al. 2006; Watt et al. 2021), but recently machine-learning algorithms are gaining popularity due to their promising results in predictive performance (e.g., Schratz et al. 2019; Barrio-Anta et al. 2020; Watt et al. 2021). Watt et al. (2021) compared several parametric, machine learning (ML), and geospatial models to map radiata pine (*Pinus radiata* D. Don.) across New Zealand, and found that the ML algorithms clearly outperformed both parametric and geospatial models. Sabatia and Burkhart (2014) evaluated random forest models and nonlinear least squares regression models for predicting SI in loblolly pine (*Pinus taeda* L.) plantations from biophysical variables and concluded that the ML models had better fit and prediction statistics than the parametric models but exhibited the potential to give illogical predictions under extrapolation. Sabatia and Burkhart (2014) also found that SI predictions from both modelling approaches exhibited a regression towards the mean SI. Aertsen et al. (2011) compared several statistical, ML, and hybrid techniques for modelling SI in two contrasting ecoregions and concluded that boosted regression trees (BRT), an ML algorithm, was one of the most effective techniques. BRT method shares the advantages of tree-based methods and has high predictive performance (Elith et al. 2008). BRT has several advantages over traditional regression techniques such as multiple linear regression or generalized additive models: (1) it is insensitive to outliers and multicollinearity, (2) it can fit complex nonlinear relationships and is able to handle discontinuities, such as ecological thresholds (Aertsen et al. 2010, 2011), (3) it can automatically handle the interactive effects between predictors, (4) it can handle missing data, (5) it can handle different types of predictor variables, and (6) it does not require prior data transformation (Elith et al. 2008). BRT shares with other ML techniques the disadvantage of being harder to interpret than traditional statistical models, and potentially more difficult to extrapolate (Sabatia and Burkhart 2014).

The aim of this work was to create a wall-to-wall high-resolution (16 m × 16 m) SI map for the forest area in Norway. Our research objectives were (1) to examine the usefulness of remotely sensed data, (2) to investigate whether including such variables causes bias in recently felled areas, and (3) to investigate the performance, advantages, and disadvantages of BRT as a tool to create high-accuracy SI maps. The models are constructed using data from the Norwegian National Forest Inventory (NFI).

Materials and methods

Data

The study area covers the forest area of Norway, which is approximately 37% of the Norwegian land area. Climatic conditions range between a mean annual temperature of −2 °C and about 6 °C, and a mean annual precipitation from about 400 mm up to 1300 mm. The permanent NFI plots used here cover the whole country from latitude 58.1°N up to 71.1°N and longitude 4.6°E up to 31°E (Breidenbach et al. 2020). There are three main dominant species in Norway: Norway spruce (*Picea abies* (L.) Karst.), Scots pine (*Pinus sylvestris* L.), and birch (*Betula* spp.). Although the largest fraction of the forest area is dominated by broadleaves (42%), mainly birch, Norway spruce (42%), and Scots pine (30%) form the majority of biomass in Norway (Breidenbach et al. 2020).

In Norway, SI is defined as the average height of the 100 thickest trees per hectare at the reference age of 40 years (Tveite and Braastad, 1981). SI data used to fit the models came from the permanent plots of the Norwegian NFI (Breidenbach et al. 2020) measured between 2013 and 2017 (Table 1). The NFI plots are circular with a size of 250 m^2 and are systematically distributed on a 3 km × 3 km (Easting × Northing) grid in lowlands, a 3 km × 9 km grid in the mountains excluding Finnmark, and a 9 km × 9 km grid in Finnmark. Larger relative spacing is used in the mountains and in Finnmark due to the lower proportions of forests there (Breidenbach et al. 2020). Tree measurements are performed within the 250 m^2 plots, on which the diameter at breast height (DBH), species, status, and position are recorded for all trees equal or larger than 5 cm in DBH. On a 1000 m^2 area surrounding each circular plot landscape and stand-level characteristics such as SI, stand age, soil type, and land-use category are assessed. The stand age is the biological age, and it is determined from increment cores from one or two representative trees outside the 250 m^2 plots and within the 1000 m^2 circle. Note that this does not imply that the plots are even-aged. Trees are selected to represent a typical dominant tree,

Table 1. Number of plots in each of the SI classes and for unproductive stands (unprod.) for the three datasets used.

	Permanent NFI				Temporary NFI				Temporary NFI plots in FMI stands		
SI (m)	Unprod.	Spruce	Pine	Hardwoods	Unprod.	Spruce	Pine	Hardwoods	Spruce	Pine	Hardwoods
6		229	383	409		45	215	271	12	36	9
8		522	976	928		104	515	653	20	337	8
11		659	700	549		169	428	578	71	348	83
14		773	368	230		238	252	273	195	192	183
17		564	150	137		234	75	97	251	35	80
20		270	29	64		194	14	52	205	3	19
23		120	2	14		146		5	99	1	5
26		11				29	1		6		1
Total	2492	3148	2608	2331	1234	1159	1500	1929	859	952	388

and should be one of the 10 thickest trees within the 1000^2 but should not be an outlier. Increment cores taken at breast height are used to estimate tree age at breast height by counting tree rings. Tree rings are usually measured in the field; however, if the tree rings are difficult to measure, they are taken back to the lab. Stand age is, in forests that consist of either one or more than two layers, the basal-area weighted age of all trees. In two-layered forests, age is the basal-area weighted age of all trees in the overstory. In uneven-aged or multilayered forests, SI and stand age are usually estimated using two trees: one of the largest and oldest trees within the 1000 m^2, and one of the younger trees that is still growing relatively well, for example one that is about to reach or has recently reached maturity age. SI is determined using Tveite and Braastad (1981) curves, which use stand age at breast height and top height to determine SI. SI estimation, particularly for uneven-aged and multilayered stands, is relatively subjective, since the selection of the representative trees and the estimation of stand age as basal-area weighted age are subjectively done.

During the 6th NFI (1986–1993), SI was assessed for the main species by selecting at least one tree outside the 250 m^2 plots and within the 1000 m^2 ha circle of the NFI. The heights of those trees were recorded, and their corresponding ages were estimated based on increment cores. The resulting data were used to estimate the SI of the stand represented by the sample plot according to Tveite and Braastad (1981), which places SI in eight classes: 6, 8, 11, 14, 17, 20, 23, and 26 m. We will call this SI estimated at the plot "observed SI", although it is estimated from the sampled trees. SI is reevaluated for the main species of the plot at each cycle of the Norwegian NFI. The current spatial distribution of productive/unproductive plots and SI by species are available in the Supplementary material (Figs. S1 and S2).

A total of 10 579 NFI permanent plots were used, 8087 of which were in productive forest and 2492 in unproductive forest plots. SI is only estimated for productive forest, that is, for forest with a potential for minimum annual volume increase of 1 m^3 per hectare, and only for the main species. Following the convention of forest management in Norway, tree species are classified into three groups: spruce, pine, and broadleaves (Tomter et al. 2010). The main species is defined as the one with the largest proportion of basal area. In young stands, it is defined as the species with the largest proportion of basal area after precommercial thinning. All broadleaves use the SI curves for birch.

We had two independent datasets to benchmark the SI models (Table 1; Fig. 1). We used 5822 temporary NFI plots measured during the period 2008–2018 as a validation dataset. SI at the temporary plots is assessed following the same criteria as is followed for permanent NFI plots. We included only temporary and permanent NFI plots that fell entirely within a single stand. Temporary plots are measured by county every 15 years. Since we used the last available 10 years, some counties are not represented in our dataset. As an additional benchmark, we used the SI from the FMIs (SI-FMI), which are used for operational forest management, to assess the quality of the SI estimations. SI for FMIs in Norway is currently typically derived through photo interpretation and existing stand age information, existing local SI maps, or a combination of the two (Bollandsås et al. 2019). FMIs do not have the same coverage as the temporary plots, and only 2856 of the 5822 temporary NFI plots fell into a stand with an FMI. We used only those 2856 temporary NFI plots when benchmarking our results against SI-FMI.

In our analysis, we included variables that have been previously found useful to predict SI and that could correlate to SI, including climatic, geographic, soil parent material, and topographic variables, but also a land resource map (Table 2). We restricted the variables to those that were available for all of Norway, except the depth to water (DTW), which is only available for parts of the country determined by the availability of a detailed terrain model at the time of creation of the DTW dataset. We also tested ALS-derived variables such as crown cover and height, and Sentinel-2 data. Table 2 describes the variables used, their sources, and resolution. Table S1 presents further details on the variables.

The Norwegian national land resource map (Ahlstrøm et al. 2014), AR5, includes land information on soil conditions, vegetation, site productivity (low, medium, high, very high), and tree species groups. AR5 is based on extensive field inventories and aerial image interpretation carried out in the 1960s and 1970s, before the first permanent NFI was established in 1986 (Breidenbach et al. 2020). AR5 uses a qualitative classification of site productivity, while forest planning is done based on the quantitative SI classification used by the NFI; it

Fig. 1. Approximate location of permanent (left) and temporary (right) NFI plots. Temporary NFI plots within stands with an FMI are in blue. Base map image for Norway from Kartverket (https://www.kartverket.no).

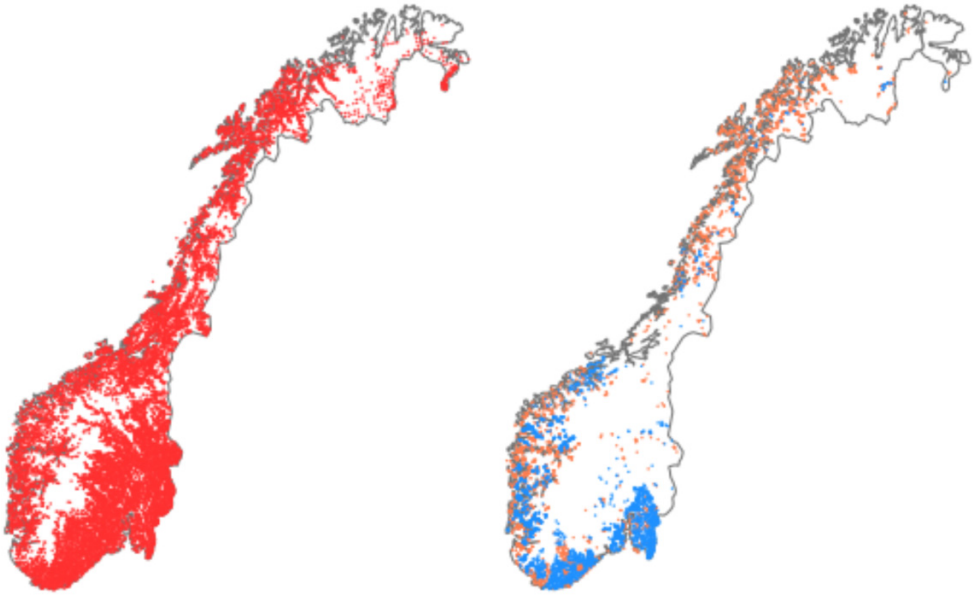


Table 2. Description of the explanatory variables, their sources, and resolution.

Variable type	Description	Source	Resolution
Climatic	30-year downscaled monthly average temperature and precipitation	Norwegian Meteorological Institute (Lussana et al. 2019)	1 km
Land resource map AR5	Land information on soil conditions, vegetation, site productivity, and tree species groups	Ahlstrøm et al. (2014)	1:5000
Parent material	Soil parent material	Geological survey of Norway (Norges Geologiske Undersøkelse 2021)	1:20.000, 1:50.000 and 250.000, and parts of Finnmark 1:500.000
Topographical variables	Slope, altitude	Derived from the 10 m resolution national digital terrain model provided by the Norwegian Mapping Authority	10 m
Depth to water (DTW)	Vertical distance to groundwater table. DTW takes into account the terrain slope, but not soils or parent material	NIBIO DTW map	1 m
Geographical	Distance to coast	Based on a 100 m resolution map produced by The Norwegian Water Resources and Energy Directorate	100 m
	Latitude, longitude		16 m
Sentinel-2	Bands B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12, and the Normalized Difference Vegetation Index (NDVI) based on bands B4 and B8	Breidenbach et al. (2021)	B2, B3, B4, and B8 10 m. B5, B6, B7, B8A, B11 and B12 20 m
ALS	A national ALS campaign: metrics derived from the heights of the first return ALS echoes (mean and percentiles), and the spatial distribution of echoes (crown cover)	Base information of the forest resource map SR16 (Hauglin et al. 2021)	16 m

is based on 50 years old data, and does not cover all forest land in Norway (e.g., drained mires).

Slope and altitude were extracted from a 10 m resolution national digital terrain model provided by the Norwegian Mapping Authority. Distance to coast is based on a 100 m resolution map produced by the Norwegian Water Resources and Energy Directorate. Temperature and precipitation were extracted from the SeNorge_2018 dataset produced by the Norwegian Meteorological Institute (Lussana et al. 2019). All meteorological data were available at 1 km × 1 km spatial

resolution and daily time steps, which were aggregated to monthly means. Within each 1 km × 1 km pixel, the altitude range can be large, particularly in mountainous areas. To account for differences in temperature within each 1 km × 1 km pixel, we corrected temperature with altitude by adjusting with 0.65 °C per 100 m altitude difference between the 1 km × 1 km grid and the plot altitude (Skaugen et al. 2003).

A combination of bands in the visible and near-infrared, NIR, (B2–B8A) and short-wave infrared, SWIR, (B11 and B12) spectrum have been found useful to estimate canopy biophysical variables in a boreal forest (Majasalmi and Rautiainen 2016) and elsewhere (e.g., Wang et al. 2018; Rahimzadeh-Bajgiran et al. 2020). Therefore, we used data from the bands B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12 from a level 2A mosaic of Sentinel-2 satellite images (Table 2). Based on bands B4 and B8, we also calculated the Normalized Difference Vegetation Index (NDVI). The mosaic consisted of imagery acquired under leaf-on conditions during July 2018 and was obtained using the Sentinel-2 Global Mosaic web service (Breidenbach et al. 2021). Sentinel-2 delivers 13 spectral bands ranging from 10 to 60 m pixel size. Its blue (B2), green (B3), red (B4), and near-infrared (B8) bands have a 10 m resolution. Its red-edge (B5), NIR (B6, B7, and B8A) and SWIR (B11 and B12) have a ground sampling distance of 20 m.

ALS metrics were calculated from point cloud data acquired in a national scanning campaign from 2010 to 2019 by the Norwegian Mapping Authority. The first return echoes were used, and six metrics were calculated from the above-ground heights of these echoes: the mean and the heights corresponding to the 25th, 50th, 75th, 90th, and 95th percentiles (Hauglin et al. 2021). We also used two metrics describing crown cover at 5 and 10 m above ground. These were derived by calculating the proportion of subvolumes with and without presence of laser echoes. The calculations of crown cover are further described by Schumacher et al. (2022). The ALS data were acquired over a period of 10 years, but no measures were taken to harmonize these data. The main reason was that there is no straightforward way to account for the varying acquisition time of metrics such as crown cover.

All variables were scaled to the 16 m × 16 m grid of the Norwegian forest resources map SR16 (Astrup et al. 2019; Hauglin et al. 2021). SR16 is based on empirical models fit using NFI sample plots with an area of 250 m^2. A cell edge length of 16 m was chosen because it is the closest integer to create a squared cell that resembles the size of an NFI plot (250 m^2).

Modelling SI and productivity

SI is defined as the average height of dominant trees of a certain species at a certain age. SI is, as such, a continuous variable. However, in Norway, SI is assessed according to categories defined by intervals. For example, a stand is classified as SI 8 if the height of the dominant trees at age 40 is between 6.5 and 9.5 m. The estimated SI is then an interval variable, an ordinal variable where the difference between two values is meaningful. Therefore, we decided to model SI as an ordinal variable.

The problem of modelling an ordinal variable can be transformed into a series of binary classification problems. Following the approach suggested by Frank and Hall (2001), we transform a k-class ordinal regression problem to a k-1 binary classification problem. In terms of SI classes

$$
\begin{aligned}
&\Pr(\mathrm{SI}=6) = 1 - \Pr(\mathrm{SI}>6)\\
&\Pr(\mathrm{SI}=8) = \Pr(\mathrm{SI}>6) - \Pr(\mathrm{SI}>8)\\
&\Pr(\mathrm{SI}=11) = \Pr(\mathrm{SI}>8) - \Pr(\mathrm{SI}>11)\\
&\ldots\\
&\Pr(\mathrm{SI}=20) = \Pr(\mathrm{SI}>17) - \Pr(\mathrm{SI}>20)\\
&\Pr(\mathrm{SI}=23) = \Pr(\mathrm{SI}>20)
\end{aligned}
\tag{1}
$$

We fitted models for Pr(SI > 6), Pr(SI > 8),..., Pr(SI > 20) for the three main species in Norway: Norway spruce, Scots pine, and birch. We used these models to calculate the probabilities that a given site has SI classes 6, 8,..., 23, that is, Pr(SI = 6), Pr(SI = 8),..., Pr(SI = 23) following eq. 1. See Fig. 2 for a schema of the general approach we took to model SI. Not enough data were available to fit models for SI > 20 for pine and birch.

Applying the above approach results in a predicted probability for each SI class and species at each prediction unit (p.u.). A p.u. is a temporary NFI plot in the case of validation or a pixel in the case of mapping.

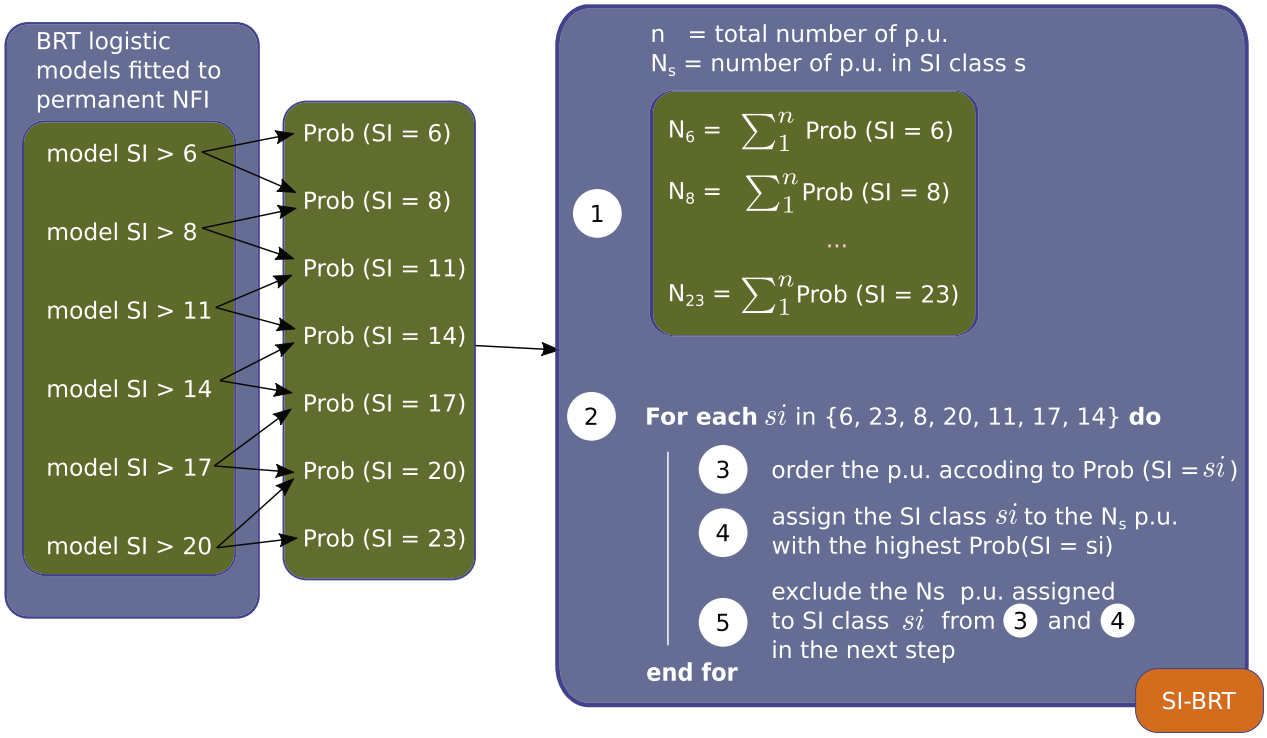
These probabilities can be used to assign an SI class to each p.u. for each of the three species modelled. One way to do this would be to calculate the mean predicted value ($\widehat{y}_i$) for observation i as the expected value of a discrete uniform distribution (Casella and Berger 2002, p. 86)

$$
\widehat{y}_i = \sum_{k=1}^{K} \mathrm{SI}_k \times \widehat{\Pr}(\mathrm{SI}_K|\boldsymbol{x}_i) \tag{2}
$$

where $\widehat{\Pr}(\mathrm{SI}_K|\boldsymbol{x}_i)$ is the probability of observing SI_k as predicted by the BRT models (eq. 1).

However, this results in underrepresentation of SI classes at the extremes of the range, mostly SI classes 6, 23, and 20, and overrepresentation of the SI classes at the center of the range. This regression to the mean has been reported previously (Sabatia and Burkhart 2014). To overcome this problem, we took a different approach, which we outline in Fig. 2. (1) We calculated the expected frequencies of each SI class by adding the probabilities for all p.u. for each SI class, that is, N_s in Fig. 2 where s = (6, 8,..., 23). (2) Then, we selected an SI class s, (3) we ordered the p.u. according to their probability of belonging to that SI class, and (4) we assigned the SI class s to the N_s p.u. with the highest probability of belonging to that class. (5) We repeat the process for all SI classes excluding from step (4) the p.u. that already has an SI class assigned. The order in which we selected SI classes in step (2) affected our results. Therefore, we tested several options, including starting by the lowest SI, the highest SI, or by the extremes (lowest and highest first). The option that resulted in the best performance measures in the fitting dataset was the extremes option, where we select for step (2) first the

Fig. 2. Workflow to calculate SI from the binary models. *s* indicates an SI class; p.u. are prediction units.



lowest SI class, then the highest, then the second lowest, and so on. We refer to the resulting SI as SI-BRT.

To produce an SI map for the forest area in Norway, we fitted a model to predict whether a site is productive or unproductive (model of productive site), and, on the productive sites, we fitted models for SI (SI models). We fitted the model of productive site and the SI models twice, once with ALS-derived and remotely sensed variables (full models) and once without these variables (simple models).

We modelled whether a site is productive or not with a logistic regression BRT model that models the probability that a plot falls within a productive forest, $y = 1$ conditional on plot covariates $\boldsymbol{X}$, i.e., $P(y = 1|\boldsymbol{X})$.

Modelling approach

The objective of this work was to produce a high-resolution (16 m × 16 m) SI Map for Norway and test whether the variables derived from remotely sensed data improved the SI estimations. Since our focus is on prediction, we selected an ML algorithm because of their generally higher prediction power, particularly when large datasets are available (Elith et al. 2008; Aertsen et al. 2011; Watt et al. 2021). Among the ML algorithms, we selected BRT, an ensemble method for fitting statistical models that can accommodate any type of variable (continuous, categorical, also missing and nonindependent data) and can deal with highly correlated independent variables. BRT identifies important predictor variables and enable complex functions and their interactions to be modelled without making assumptions about the type of functions or interactions (Elith et al. 2008). Model fitting and validation were performed using the R Package dismo (Hijmans et al. 2020) version 1.3-5, which uses gbm package (Greenwell et al. 2020) in R version 4.1.3 (R Core Team 2020).

BRT is an ensemble method, which draws on techniques from both statistical and ML traditions. It combines the advantages of regression trees (models that relate a response to their predictors by recursive binary splits) and boosting (a method for combining many simple models to improve predictive performance). Elith et al. (2008) offers a comprehensive and accessible introduction to BRT, while Ridgeway (2007) provides mathematical details.

BRT offers several options to help with the interpretability of the results. The main one is the relative importance. The relative importance is a measure of the influence of input variables. It is calculated by averaging the number of times a variable is selected for splitting, weighted by the squared improvement to the BRT model, as a result of each split, and it is then scaled so the values sum up to 100 (Friedman 2001).

Hyperparameter tuning

BRT uses a sequential model fitting approach that allows regression trees to be added until the data are completely overfitted. Then, it uses regularization to avoid the reduction in generality derived from overfitting. Regularization methods (Elith et al. 2008; Hastie et al. 2009) balance model fit and predictive performance. In BRT, regularization is achieved through shrinking that is applied to each new tree as it is fitted. Analytically, BRT regularization involves jointly optimizing the number of regression trees (nt), learning rate (lr), and tree complexity (tc). tc—the number of nodes in a tree—controls the maximum level of interactions that can be quantified. The lr, also known as the shrinkage parameter,

determines the contribution of each tree as it is added to the model. Decreasing (slowing) lr increases the nt required, and in general a smaller lr (and larger nt) are preferable, conditional on the number of observations and time available for computation. Ridgeway (2007) recommends selecting lr as small as possible while still able to fit the model in a reasonable amount of time. Ridgeway (2007) proposes values of lr between 0.01 and 0.001. A smaller lr will always give improved predictive performance (Ridgeway 2007), but at a computational cost, both in terms of storage and CPU time. In BRT, stochasticity is controlled through a "bag fraction" that specifies the proportion of data to be selected at each step. The optimal nt is automatically calculated in the dismo package.

The dismo package uses cross-validation as described in Hastie et al. (2009) to first determine the optimal nt, then fits a final model to all the data. The first step in the cross-validation is to divide the dataset randomly into n subsets (we chose $n = 10$), then n different training sets each comprising a unique combination of $n - 1$ subsets for fitting and one testing subset are defined. Then, for each training set a BRT model is fitted and their predictive performance measured on their respective testing subset. The mean predictive performance is used to select the optimal nt. The code and a more detailed description of the cross-validation method are available in Elith et al. (2008).

To select the rest of the hyperparameters, we tested for all models all combinations of the following learning rates, tree complexities, and bag fractions: learning rate (0.01, 0.005, 0.001), tree complexity (1, 2, 5, 10), and bag fraction (0.5, 0.6, 0.75). From these combinations, we selected the models with the lowest deviance, a measure of loss of predictive performance (Elith et al. 2008).

Variable selection

The dataset used to fit the models included 54 variables, some of which contributed more to the model than others. We used the methodology suggested in Elith et al. (2008), which uses methods analogous to backward selection in regression, to simplify the predictor set in our models, although it is generally not necessary to drop unimportant variables in BRT (Elith et al. 2008). The number of predictors dropped ranged from 0 for spruce SI 8 and birch SI 6, and 49 for pine SI 17 for the final models.

Evaluation of the models and SI-BRT performance

Individual models' predictive performance was evaluated in terms of discrimination. Discrimination addresses how well the model can distinguish between, for example, a site with an SI $>$ 8 and an SI $\leq$ 8. We evaluated goodness of fit with calibration measures. Calibration is the extent to which the predictive probabilities agree with the observed frequencies. The model's predictive discrimination was evaluated using the probability of concordance, c or c-index (Harrell et al. 1982, 1984), also known as the area under a receiver operating characteristic curve (Hanley and McNeil 1982). The c-index takes values from 0.5 to 1, where a value of 0.5 indicates random predictions, and a value of 1 indicates perfect separation. A c-index larger than 0.8 indicates a model with excellent discrimination (Hosmer and Lemeshow 2000). We used the calibration slope (Cox 1958) to evaluate the miscalibration. Ideally, for the model being evaluated, the calibration equation would have a slope of 1 and an intercept of zero (Miller et al. 1991). Lack of calibration was tested with a χ^2 test with two degrees of freedom (H_0: intercept $= 0$, slope $= 1$).

We evaluated the accuracy of SI-BRT on the validation dataset both graphically and numerically. As a summary measure for accuracy, we estimated the root mean squared error (RMSE; eq. 3), and the relative RMSE (rRMSE; eq. 3) for the productive forest in the validation dataset, and the mean absolute error (MAE; eq. 3). The RMSE combines bias and variance. As a measure of systematic tendency, we estimated bias as the average difference between the SI in the validation dataset and the predictions of the models for the main species of the plot (eq. 3).

We calculated RMSE, rRMSE, MAE, and bias as follows:

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}} \\ \text{rRMSE} &= \frac{\text{RMSE}}{\bar{O}} \\ \text{MAE} &= \frac{\sum_{i=1}^{n}|P_i - O_i|}{n} \\ \text{Bias} &= \bar{O} - \bar{P} \end{aligned} \tag{3}$$

where P are predicted values, O observed values, n number of observations, and $\bar{O}$ and $\bar{P}$ are the mean observed values and the mean predicted values, respectively.

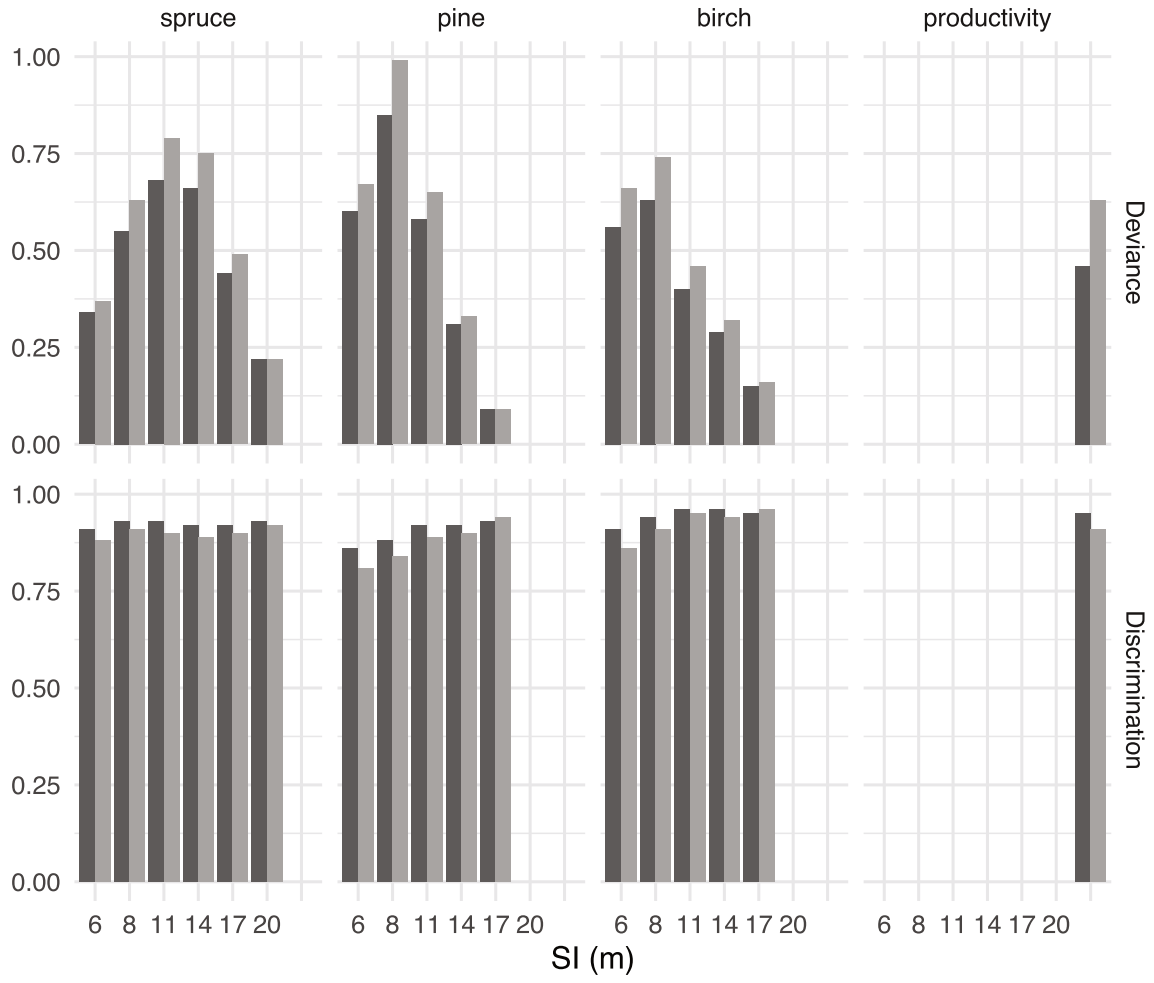
Results

Full model and simple model performance

In this section, we compare the performance of the full and simple models. Hyperparameters for the models are presented in Tables S2 and S3. Mean discrimination (c-index) from the 10-fold cross-validation used to select nt and deviance are presented in Fig. 3. All calibration tests had a p value of 0.29 or higher indicating that there was no lack of goodness of fit. All models had a discrimination of 0.81 or above. Full models had an equal or higher discrimination and equal or lower deviance than the simple models, for both the model of productive site and the SI models. In general, pine models had the lowest discrimination. For all species, the model for SI $>$ 6 had worse discrimination than the rest.

In terms of RMSE, the full models had consistently, in both the fitting and the validation datasets, better results than the simple models (Fig. 4; Table 3), while in terms of bias the simple model outperforms the full model for spruce and pine in the validation dataset. Full models correctly classified 86.8% of the plots, while the simple models classified 79.4% of the plots correctly. Full models incorrectly classified 11.1% of the plots as productive and 3.1 % as unproductive, while simple models incorrectly classified 14.6% as productive and 6.1% as unproductive. For the productivity full model, in the validation dataset 85.9% of the plots were correctly classified, 3.1%

Fig. 3. Comparison of performance in the fitting dataset for the simple (light grey) and full (dark grey) models in terms of discrimination and deviance.



were classified as productive when where actually unproductive, and 11.1% were classified as unproductive while on productive sites.

The rRMSE when predicting for the fitting dataset, that is, the NFI permanent dataset (Table 3), ranges from 11.6% for birch to 17.4% for pine, while the bias ranges from 0.01 for pine and spruce to −0.02 for birch. When we compare the SI estimates of our models with the observed SI in the NFI temporary validation dataset, SI-BRT estimates have higher rRMSE and larger bias for all species (Table 3) than for the fitting dataset. With respect to bias when predicting in the validation dataset, both simple and full models tended to underpredict spruce and pine. However, both rRMSE and bias for the subset of the fitting dataset that are in municipalities where there is at least one plot in the validation dataset have similar values as in the validation dataset (Table 5).

Recently harvested and nonforested areas with small or zero height and little crown cover could potentially have biased SI estimations. We calculated RMSE for the simple, full models, and FMI for the validation dataset plots that had the 90th percentile (h90) less than 2 m and crown cover at 5 m (cc5) less than 10%. The results presented in Table 4 show that the full models did not perform systematically worse or better than the simple models, and that both simple and full models performed similarly or worse than the FMI SI estimates. For spruce, the best rRMSE is given by the full models (21.0%) followed by the simple models and the SI from the FMIs (21.39%). For pine and birch, the best results are given by the FMI (25.7% and 23.4%, respectively) followed by the full models for pine (30.7%), and by the simple models for birch (31.0%).

The relative importance of the predictor variables for all the models is presented in Figs. S3–S5 for full models and in Figs. S6–S8 for the simple models. In general, the variables with relative high importance in the simple models remained so in the full models, albeit with lower relative importance. The variables with the highest relative importance were AR5 site productivity (AR5 for. prod.), and temperatures during the growing season (April–July). For the productivity models, AR5 site productivity (AR5 for. prod.), and crown cover at 5 m (full models), cc5, were the variables with the highest relative importance. For the spruce models, the variables

Fig. 4. Root mean square error (RMSE) and mean absolute error (MAE) for SI from the forest management inventories (SI-FMI) and for the predicted SI-BRT from the final (full) models for a subset of the NFI temporary dataset where FMI data were available.

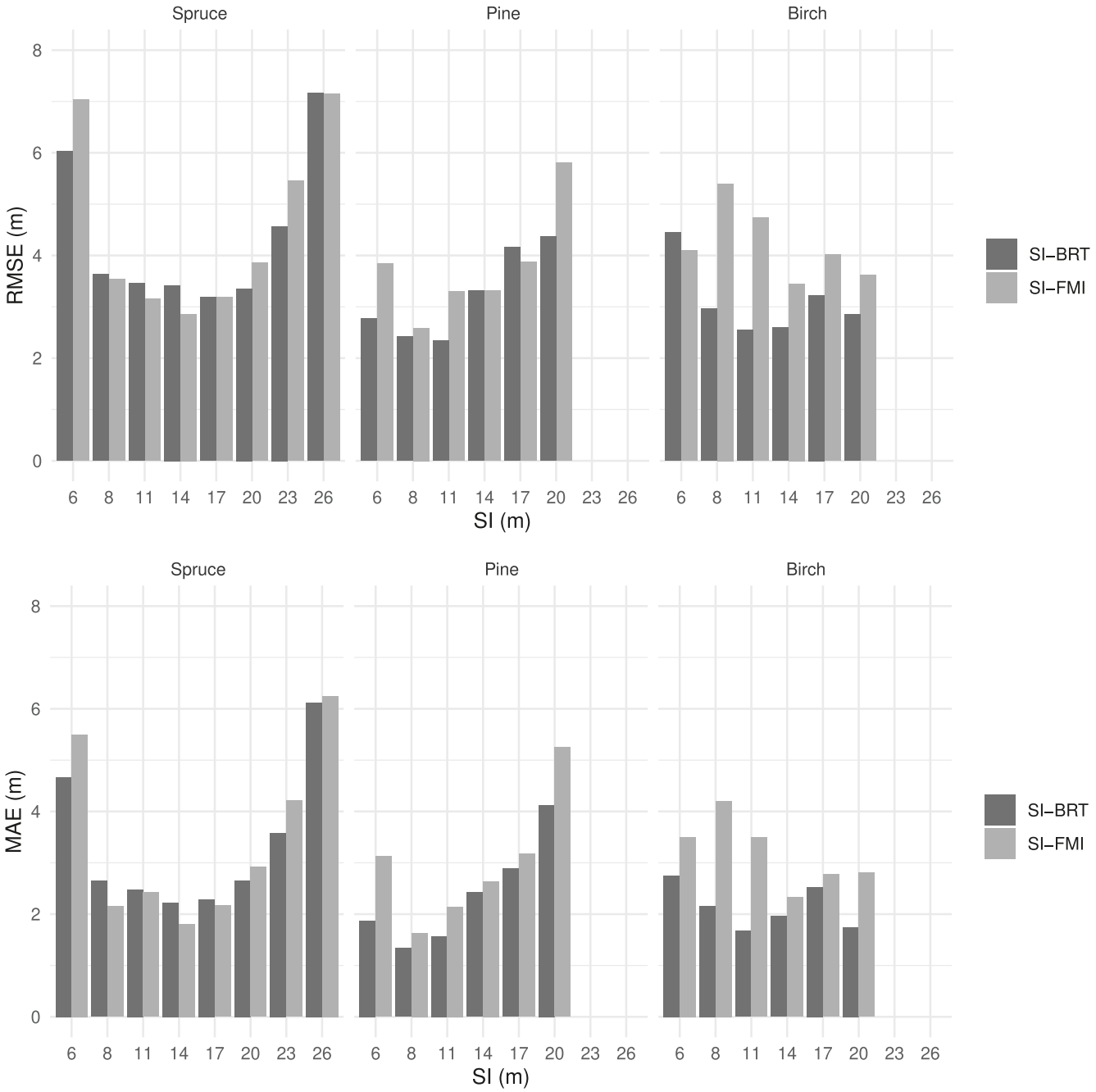


Table 3. Performance measures for SI-BRT for the NFI permanent and temporary datasets.

		NFI permanent			NFI temporary		
	Species	RMSE (m)	rRMSE (%)	Bias (m)	RMSE (m)	rRMSE (%)	Bias (m)
Full models	Spruce	1.94	14.65	0.01	3.59	22.77	0.33
	Pine	1.74	17.37	0.01	2.78	27.43	0.25
	Birch	1.14	11.55	−0.02	2.30	22.37	0.04
Simple models	Spruce	2.72	20.59	0.02	4.04	25.63	0.21
	Pine	2.36	23.58	0.00	3.27	32.19	0.12
	Birch	1.83	18.54	0.00	2.51	24.43	−0.07

Table 4. Performance measures for SI-BRT for the subset of the validation dataset where the 90th percentile (h90) is less than 2 m and crown cover at 5 m (cc5) less than 10% by stand composition for the FMI, full models, and simple models.

	FMI		Full models		Simple models		
Species	rRMSE	Bias	rRMSE	Bias	RMSE	Bias	Number of plots
Spruce	21.39	0.55	21.00	0.18	21.39	0.36	33
Pine	25.72	−0.18	30.71	1.09	33.88	−0.45	11
Birch	23.44	−2.38	32.20	−1.69	30.95	−2.15	13

Table 5. Performance measures for SI-BRT for the NFI permanent plots (fitting dataset) that are in municipalities that have at least one plot in the validation dataset.

Species	rRMSE (m)	Bias (m)
Spruce	22.61	0.34
Pine	26.94	0.26
Birch	22.04	−0.05

with higher contribution were AR5 site productivity (AR5 for. prod.), and temperature in April and May. For pine, the variables with higher contribution were AR5 site productivity (AR5 for. prod.), the ALS heights, S2-NDVI, S2-B12 (only for full models), and temperature in May and July. For birch, AR5 site productivity, temperature in May, June, and July, and for the full models height of the 95th percentile, longitude, and crown cover at 10 m (cc10) were the variables with the highest relative importance. We found that the most useful Sentinel-2-derived variable was NDVI (S2_NDVI).

Not all models had interactions (tc of 1 in Tables S2 and S3). The two models for the higher SI for spruce (SI classes 17 and 20) and the models for SI classes 6 and 14 for pine had no interactions in both the full and the simple models. The most important interactions included, for all models with interactions, the variable with the highest relative importance.

We selected the full models as the final models because they had a better performance on both validation and model fitting. The remainder of the results are focused on these final models.

Final models

If we group the variables according to the variable types in Table 2, the group with the highest relative importance, calculated as the average of the sum of all relative importance of the variables within each group for each model, was ALS for pine (28.1%), land resource map AR5 for spruce (31.1%) and the model of productive site (40.8%), while for birch it was the climatic variables group (40.2%). For all the species, the four groups with the highest relative importance were land resource map AR5, climatic variables, ALS, and Sentinel-2 variables. The sum of the relative importance of these four variables add up to at least 82.0% (birch) of the relative importance for all species, and up to 92.5% (spruce).

The frequency distribution of SI classes for the NFI temporary validation dataset (Fig. 5) is closely followed by SI-BRT. For spruce, the SI-BRT overestimates the number of plots in SI 14 and underestimates the number of plots in SI 11 and >20.

With respect to stand composition, the highest rRMSE and MAE occurred in spruce dominated stands, while the lowest rRMSE and MAE occurred in birch dominated stands (Table S4). With respect to maturity class (Table S5), the highest rRMSE is found in young forest (27.5%), while the lowest rRMSE is found in older production forest (23.2 %).

For spruce and pine, the models can discriminate well between low and high SI, but within low/high SI classes they cannot discriminate accurately to which SI class a plot belongs (Fig. S9). For birch, the models can distinguish better among SI classes, at least for the lower ones.

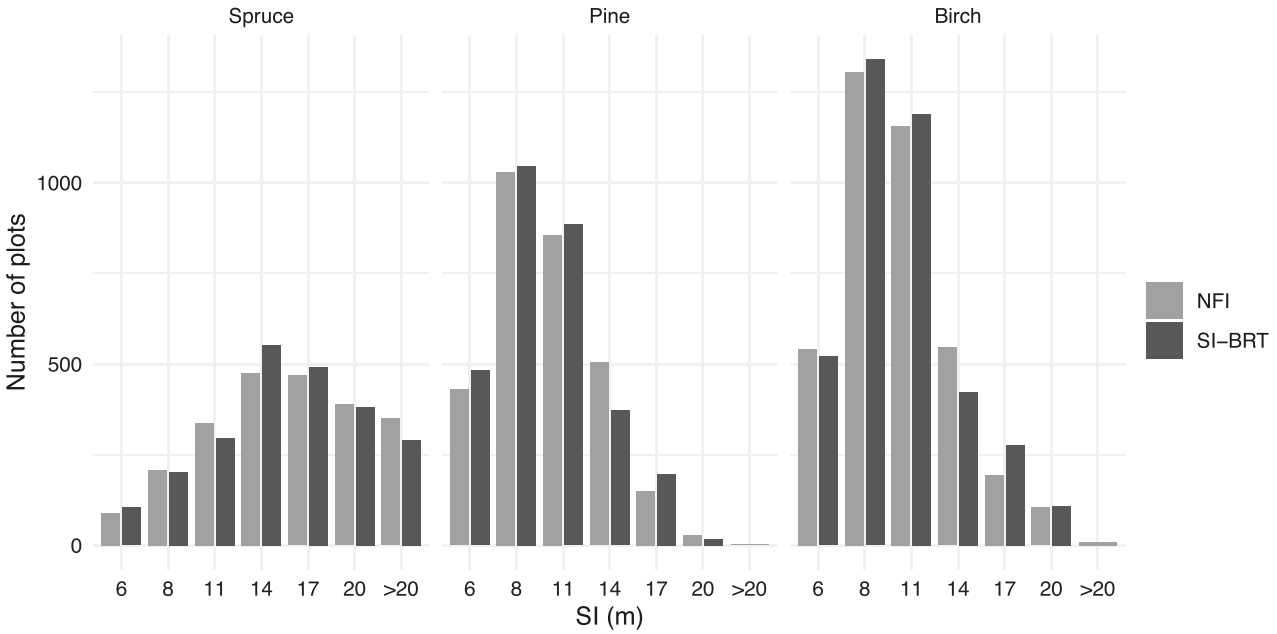
In most instances, SI-BRT predictions outperform the SI from FMIs, SI-FMI (Fig. 4). Only in SI classes 8, 11, and 14 for spruce, SI-FMI outperformed SI-BRT in terms of RMSE and MAE. The SI predicted by the BRT models has lower RMSE (SI-BRT 3.45 m, SI-FMI 3.82 m) and lower bias (SI-BRT 0.15 m, SI-FMI −0.18 m) than the SI-FMI. While the SI-FMI tends to underpredict SI from the temporary NFI plots for spruce (bias of 1.06 m), it tends to overpredict in pine and birch dominated stands (bias of −0.33 and −1.81 m, respectively). SI-BRT tends to underpredict SI for the same plots (bias of 0.51 m for spruce, 0.12 for pine), and overpredict for birch (−0.35).

The validation dataset does not cover the same area as the fitting dataset as some counties are missing due to their inventories being older than 10 years. We recalculated RMSE and bias for the fitting dataset for the municipalities where validation data were available (6048 plots). Results (Table 5) show that for these areas, the rRMSE values for the fitting dataset are significantly higher than for the full fitting dataset, and close to the rRMSE and bias for the full validation dataset.

Discussion

We present two sets of models that can be used to produce a high-resolution SI map for Norway, one with remotely sensed data and one without. We fitted the models using BRT, an ML technique with high prediction performance. The response variables, SI and whether a site is productive or not, were observed at 10 579 Norwegian NFI plots. We used available wall-to-wall variables, including climatic, topographic, geographical variables, and remotely sensed data from the Copernicus Sentinel-2 mission and ALS.

Fig. 5. Histogram of observed (NFI) and estimated SI (SI-BRT) using the full models for the NFI temporary validation dataset.



The resulting SI-BRT and model of productive site can be used to produce a wall-to-wall SI map for Norway. They can be used to estimate productivity for all forest area (1). The model of productive site can distinguish between unproductive and productive sites (2). They can be used to estimate productivity of nonforested areas (3). They give SI estimates for the three main species in Norway (4). However, since the models use data from the current dominant species, such as the ALS-derived variables and the Sentinel-2 variables, the estimates for the other two species that are not currently the dominant ones will likely have less accuracy than the estimates for the current dominant species. For the nondominant species, it is possible that the simple models (with no predictors related to current forest structure) outperform the full models (with predictors related to current forest structure), but we do not have any data to test whether this is the case or in general evaluate the predictions for the nondominant species.

Adding remotely sensed data improved the SI models and the model of productive site and did not result in biased estimates in recently felled plots. However, it also did not improve the estimates in plots with high SI. The largest improvement was in the model of productive site, where one of the variables with the highest relative importance in the full models was the ALS-derived variable cc5. When SI-BRT is used to estimate productivity of nonforested areas, remotely sensed data would be similar to that of recently felled areas. Since SI-BRT is not biased in the case of recently felled areas, we expect no bias in nonforested areas, although we expect higher uncertainty.

The selected algorithm to fit the models, BRT, was easy to use, and fitting the models was relatively straightforward. Since BRT can fit complex nonlinear relationships, and it is insensitive to multicollinearity, there was no need to transform the variables, or to worry about collinearity. Therefore, we were able to include as potential candidates a high number of variables, some of which were highly correlated, such as temperature for all months. ML techniques, such as BRT, are becoming widespread across the quantitative sciences for prediction and forecasting, but there are many known methodological pitfalls, being data leakage the predominant one, which have led to exaggerated claims about their performance (Greener et al. 2022; Kapoor and Narayanan 2022). We present here the three arguments suggested by Kapoor and Narayanan (2022) determining that our scientific results that use ML methods do not suffer from data leakage. The first argument being that our separation of training (fitting) and testing (validation) data is clean. Our fitting and validation dataset came from different datasets that were imported, prepared, and used in separate scripts and files. The datasets did not have any duplicated data point. The second argument is that each feature in the model is legitimate, that is, if the models have access to features (variables), that should not be legitimately available for use in the modelling exercise. We used for both model fitting and model validation the same wall-to-wall maps of all variables, which are the ones that will be used in the production of the SI map for Norway; therefore, all variables included in the model are legitimate. The third argument is that the test (validation) set is drawn from the distribution of interest. Our fitting and validation datasets are independent and were collected independently in separate campaigns, but following the same procedures and definitions. However, in our validation datasets, some of the counties were missing because they were measured over 10 years ago and we deemed the data to be too old to be included, especially because SI in Norway is changing (Antón-Fernández et al. 2016). This could have led to optimistic total RMSE since these counties have a larger representation of spruce-dominated forest, which has the largest rRMSE. On the other side, the counties missing have lower SI

variability compared with the full country, and we argue that this is the main reason behind the higher rRMSE in the validation dataset. Given that the rRMSE and bias for the municipalities included in the validation dataset show similar results for both fitting and validation datasets (Table 5), we believe that if we had had temporary NFI data covering the whole country our validation results would look similar to those reported in Table 5 for the fitting dataset.

Quantification and communication of uncertainty for individual predictions is important to improve decision-making, but it is often neglected in ML models (Kompa et al. 2021). The R package we used to fit the models and make predictions for the validation dataset does not have built-in mechanisms to assess uncertainty, but we can use the probabilities obtained for each SI class as an estimate of the uncertainty in our predictions. The final SI map, which will be available at http://kilden.nibio.no, will include the predicted SI class (SI-BRT) and the probabilities for all SI classes from eq. 1.

Our final models perform similarly as or better than other published SI models for Norway that require in situ information. For example, Sharma et al. (2012) models had RMSE of 2.04 (rRMSE 14.9%) and 1.95 (rRMSE 18.9%) m for spruce and pine for their fitting dataset (15.2% for spruce and 17.2% for pine in our models) using multiple linear regression with climatic variables, year of stand origin, vegetation type, soil depth, and topographical and geographical variables. They also used Norwegian NFI data but included variables that are only available at the NFI plots, such as soil depth and vegetation type. This impedes mapping, i.e., the wall-to-wall application of the model. Antón-Fernández et al. (2016) also used climatic variables, soil depth, and vegetation group to create climate-sensitive SI models for Norway, with RMSE of 2.91 m for spruce, 2.47 m for pine, and 2.49 m for deciduous. Antón-Fernández et al. (2016) also included soil depth and vegetation type, which are only available at the NFI plots. We included in our model soil parent material, which had a low relative importance for all models. If detailed maps with information on soil depth and soil nutrient content were available for Norway, they would very likely improve our estimates.

There are several studies outside of Norway that fit SI models for Norway spruce and/or Scots pine, but results are difficult to compare since their datasets are much smaller and usually more homogeneous, and they use a different base age for SI. For example, in the Western Carpathians, the best model from Socha (2008) for Norway spruce has an RMSE of 2.39 m (base age 100 years) and an rRMSE of 7.2% using topographic features and soil parent material variables in a dataset of 347 plots. Also, for Norway spruce in the Western Carpathians, Bošeľa et al. (2013) had for their best model a residual standard error of 3.46 m (base age 100 years), and a relative residual standard error of 12.4%, using climate and soil parameters for 201 NFI plots. In France, Seynave et al. (2005) using climatic variables and soil characteristics such as soil depth, nutritional variables, or topographic concavity estimated on the field had an RMSE of 3.83 m for Norway spruce for 1270 pure even-aged stands, and 4.16 for a larger (2087 plots) dataset including uneven-aged and mixed stands. For Scots pine, the best model from Aertsen et al. (2012), who used variables estimated on the field, such as soil, forest floor, and ecosystem predictors, has an RMSE of 2.41 m (base age 100 years) and an rRMSE of 10% using a dataset of 55 plots.

The number of plots in SI class 26 in the fitting dataset (NFI permanent plots) was too small to fit a model for SI $>$ 23 for any of the species, and thus our models can only predict the probability of SI $>$ 20 for spruce and $>$17 for pine and birch. A potential solution for spruce would be to use the combined dataset of permanent and temporary NFI plots to fit a model for SI $>$ 23.

When we first attempted to model SI as a continuous variable, the lowest and highest SI classes were underpredicted, particularly SI class 6. This was likely due to the difficulty distinguishing between SI classes 6 and 8, and the fact that SI is a truncated variable, that is, when top height is below 5 m (lower limit of SI class 6), SI is not recorded and the stand is classified as unproductive. This regression to the mean SI was also observed by Sabatia and Burkhart (2014) when modelling SI for loblolly pine. Modelling SI as an ordinal variable allowed us to avoid this underestimation of SI 6 and other underrepresented SI classes.

The high-resolution SI map derived from the BRT models presented here outperforms for almost any combination of SI and species the SI estimates from the FMIs. It has also the advantage over FMIs in that the SI-BRT is available for all the forest areas in Norway at no extra cost, and that the resolution of the SI map derived from our models has a substand resolution. High-resolution productivity maps, like the one presented here, can be used to do forest planning and management at the substand level, enabling precision silviculture.

The SI map resulting from the models presented here can be coupled with the available climate-sensitive SI models (Antón-Fernández et al. 2016) to produce wall-to-wall SI projections for Norway under different climate change scenarios.

Conclusions

The models presented here have similar or better performance than similar models for Norway that use locally measured variables. The exclusive use of variables that are available nationwide, and not only at plot sites, allows mapping of SI and productivity (whether a plot/pixel is productive). We tested and discussed the advantages and disadvantages of using BRT and remotely sensed data compared with traditional regression approaches. We concluded that BRT possesses many interesting advantages and has high predictive performance, and that using remotely sensed data improved the models, albeit mostly for the productivity model and the low SI models, and did not seem to be problematic in recently felled stands. The selected methodology allowed for a distribution of the SI classes that matched the predicted probabilities of the classes, avoiding the underrepresentation of the lowest and highest classes.

Article information

History dates

Received: 6 August 2022
Accepted: 13 January 2023

Accepted manuscript online: 30 January 2023
Version of record online: 15 March 2023

Data availability
The data material used in this study can be made available upon request to the author. An exception applies for the distribution of the exact NFI plot coordinates. This restriction is due to (1) the role of NFI plots as the main source for representative national forest statistics and (2) their use as a data source for greenhouse gas reporting. With reference to the latter cause, it is referred to the "Good Practice Guidelines" given by the Intergovernmental Panel on Climate Change (IPCC), which recommends the signatory nations to make efforts to maintain the secrecy of NFI plot locations in cases where there are no independent validation data sources available for the purpose of greenhouse gas reporting.

Author information

Author ORCIDs
Clara Antón-Fernández https://orcid.org/0000-0001-5545-3320
Johannes Breidenbach https://orcid.org/0000-0002-3137-7236

Author notes
Johannes Breidenbach served as an Associate Editor at the time of manuscript review and acceptance and did not handle peer review and editorial decisions regarding this manuscript.

Author contributions
Conceptualization: RA
Data curation: MH
Formal analysis: CA-F
Funding acquisition: RA
Methodology: CA-F
Resources: MH
Software: CA-F
Validation: CA-F
Visualization: CA-F
Writing – original draft: CA-F
Writing – review & editing: CA-F, MH, JB, RA

Competing interests
The authors declare that there are no competing interests.

Funding information
This research was supported by The Research Council of Norway (grant number 301922—"Northern Forest").

Supplementary material
Supplementary data are available with the article at https://doi.org/10.1139/cjfr-2022-0198.

References
Aertsen, W., Kint, V., van Orshoven, J., Özkan, K., and Muys, B. 2010. Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. Ecol. Modell. **221**(8): 1119–1130. doi:10.1016/j.ecolmodel.2010.01.007.

Aertsen, W., Kint, V., Van Orshoven, J., and Muys, B. 2011. Evaluation of modelling techniques for forest site productivity prediction in contrasting ecoregions using stochastic multicriteria acceptability analysis (SMAA). Environ. Modell. Software, **26**(7): 929–937. doi:10.1016/j.envsoft.2011.01.003.

Aertsen, W., Kint, V., Vos, B.D., Deckers, J., Orshoven, J.V., and Muys, B. 2012. Predicting forest site productivity in temperate lowland from forest floor, soil and litterfall characteristics using boosted regression trees. Plant Soil, **354**(1–2): 157–172. doi:10.1007/s11104-011-1052-z.

Ahlstrøm, A., Bjørkelo, K., and Frydenlund, J. 2014. AR5 klassifikasjonssystem: klassifikasjon av arealressurser. Technical report, Norsk institutt for Skog og landskap. Available from https://nibio.brage.unit.no/nibio-xmlui/handle/11250/2440173.

Antón-Fernández, C., Mola-Yudego, B., Dalsgaard, L., and Astrup, R. 2016. Climate-sensitive site index models for Norway. Can. J. For. Res. **46**(6): 794–803. doi:10.1139/cjfr-2015-0155.

Astrup, R., Rahlf, J., Bjørkelo, K., Debella-Gilo, M., Gjertsen, A.K., and Breidenbach, J. 2019. Forest information at multiple scales: development, evaluation and application of the Norwegian forest resources map SR16. Scand. J. For. Res. **34**(6): 484–496. doi:10.1080/02827581.2019.1588989.

Barrio-Anta, M., Castedo-Dorado, F., Cámara-Obregón, A., and López-Sánchez, C.A. 2020. Predicting current and future suitable habitat and productivity for Atlantic populations of maritime pine (Pinus pinaster Aiton) in Spain. Ann. For. Sci. **77**(2): 41. doi:10.1007/s13595-020-00941-5.

Berrill, J.P., and O'Hara, K.L. 2014. Estimating site productivity in irregular stand structures by indexing the basal area or volume increment of the dominant species. Can. J. For. Res. **44**(1): 92–100. doi:10.1139/cjfr-2013-0230.

Bollandsås, O.M., Ørka, H.O., Dalponte, M., Gobakken., T., and Næsset, E. 2019. Modelling site index in forest stands using airborne hyperspectral imagery and bi-temporal laser scanner data. 10.3390/rs11091020 Remote Sens. **11**(9): 1020.

Bošeľa, M., Máliš, F., Kulla, L., Šebeň, V., and Deckmyn, G. 2013. Ecologically based height growth model and derived raster maps of Norway spruce site index in the Western Carpathians. Eur. J. For. Res. 1–15.10.1007/s10342-013-0708-z

Breidenbach, J., Granhus, A., Hylen, G., Eriksen, R., and Astrup, R. 2020. A century of National Forest Inventory in Norway: informing past, present, and future decisions. For. Ecosyst. **7**: 46.10.1186/s40663-021-00315-x

Breidenbach, J., Waser, L.T., Debella-Gilo, M., Schumacher, J., Rahlf, J., Hauglin, M., et al., 2021. National mapping and estimation of forest area by dominant tree species using Sentinel-2 data. Can. J. For. Res. **51**(3): 365–379. doi:10.1139/cjfr-2020-0170.

Casella, G., and Berger, R.L. 2002. Statistical inference. Cengage Learning.

Cox, D.R. 1958. Two further applications of a model for binary regression. Biometrika, **45**: 562–565. doi:10.1093/biomet/45.3-4.562.

Elith, J., Leathwick, J.R., and Hastie, T. 2008. A working guide to boosted regression trees. J. Anim. Ecol. **77**(4): 802–813. doi:10.1111/j.1365-2656.2008.01390.x.

Frank, E., and Hall, M. 2001. A simple approach to ordinal classification. *In* Machine learning: ECML 2001. Vol. 2167. *Edited by* G. Goos, J. Hartmanis, J. van Leeuwen, L. De Raedt and P. Flach. Springer, Berlin, Heidelberg. pp. 145–156.

Friedman, J.H. 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. **29**(5): 1189–1232. doi:10.1214/aos/1013203451.

González-Rodríguez, M.A., and Diéguez-Aranda, U. 2021. Rule-based vs parametric approaches for developing climate-sensitive site index

models: a case study for Scots pine stands in northwestern Spain. Ann. For. Sci. **78**(1): 23. doi:10.1007/s13595-021-01047-2.

Greener, J.G., Kandathil, S.M., Moffat, L., and Jones, D.T. 2022. A guide to machine learning for biologists. Nat. Rev. Mol. Cell Biol. **23**(1): 40–55. doi:10.1038/s41580-021-00407-0.

Greenwell, B., Boehmke, B., Cunningham, J., and Developers, G. 2020. gbm: Generalized Boosted Regression Models. R package version 2.1.8. Available from https://CRAN.R-project.org/package=gbm.

Hanley, J., and McNeil, B. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, **143**: 29–36. doi:10.1148/radiology.143.1.7063747.

Harrell, F.E., Califf, R.M., Pryor, D.B., Lee, K.L., and Rosati, R.A. 1982. Evaluating the yield of medical tests. J. Am. Med. Assoc. **247**(18): 2543–2546. doi:10.1001/jama.1982.03320430047030.

Harrell, F.E., Lee, K.L., Califf, R.M., Pryor, D.B., and Rosati, R.A. 1984. Regression modelling strategies for improved prognostic prediction. Stat. Med. **3**(2): 143–152. doi:10.1002/sim.4780030207.

Hastie, T., Tibshirani, R., and Friedman, J., 2009. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. Springer.

Hauglin, M., Rahlf, J., Schumacher, J., Astrup, R., and Breidenbach, J. 2021. Large scale mapping of forest attributes using heterogeneous sets of airborne laser scanning and National Forest Inventory data. For. Ecosyst. **8**(1): 65. doi:10.1186/s40663-021-00338-4.

Hijmans, R.J., Phillips, S., Leathwich, J., and Elith, J. 2020. dismo: Species Distribution Modeling. Available from https://cran.r-project.org/web/packages/dismo/dismo.pdf.

Hosmer, D.W., and Lemeshow, S. 2000. Applied logistic regression. 2nd ed. John Wiley and Sons, New York, NY. doi:10.1002/0471722146.

Huang, S., Ramirez, C., Conway, S., Kennedy, K., Kohler, T., and Liu, J. 2017. Mapping site index and volume increment from forest inventory, Landsat, and ecological variables in Tahoe National Forest, California, USA. Can. J. For. Res. **47**(1): 113–124. doi:10.1139/cjfr-2016-0209.

Kapoor, S., and Narayanan, A. 2022. Leakage and the reproducibility crisis in ML-based science. Available from https://arxiv.org/abs/2207.07048. doi:10.48550/ARXIV.2207.07048.

Kompa, B., Snoek, J., and Beam, A.L. 2021. Second opinion needed: communicating uncertainty in medical machine learning. NPJ Digit. Med. **4**(1): 1–6. doi:10.1038/s41746-020-00367-3.

Latta, G., Temesgen, H., and Barrett, T.M. 2009. Mapping and imputing potential productivity of Pacific Northwest forests using climate variables. Can. J. For. Res. **39**(6): 1197–1207. doi:10.1139/X09-046.

Lindgren, N., Wåstlund, A., Bohlin, I., Nyström, K., Nilsson, M., and Olsson, H. 2021. Updating of forest stand data by using recent digital photogrammetry in combination with older airborne laser scanning data. Scand. J. For. Res. **36**(5): 401–407. doi:10.1080/02827581.2021.1936153.

Lussana, C., Tveito, O.E., Dobler, A., and Tunheim, K. 2019. seNorge_2018, daily precipitation, and temperature datasets over Norway. Earth Syst. Sci. Data, **11**(4): 1531–1551. doi:10.5194/essd-11-1531-2019.

Majasalmi, T., and Rautiainen, M. 2016. The potential of Sentinel-2 data for estimating biophysical variables in a boreal forest: a simulation study. Remote Sens. Lett. **7**(5): 427–436. doi:10.1080/2150704X.2016.1149251.

Miller, M.E., Hui, S.L., and Tierney, W.M. 1991. Validation techniques for logistic regression models. Stat. Med. **10**: 1213–1226. doi:10.1002/sim.4780100805.

Monserud, R.A., Huang, S., and Yang, Y. 2006. Predicting lodgepole pine site index from climatic parameters in Alberta. For. Chron. **82**(4): 562–571. doi:10.5558/tfc82562-4.

Noordermeer, L., Gobakken, T., Næsset, E., and Bollandsås, O.M. 2020. Predicting and mapping site index in operational forest inventories using bitemporal airborne laser scanner data. For. Ecol. Manage. **457**: 117768. doi:10.1016/j.foreco.2019.117768.

Norges Geologiske Undersøkelse. 2021. Løsmasser N50 N250. Available from https://kartkatalog.geonorge.no/metadata/loesmasser/3de4ddf6-d6b8-4398-8222-f5c47791a757.

Rahimzadeh-Bajgiran, P., Hennigar, C., Weiskittel, A., and Lamb, S. 2020. Forest potential productivity mapping by linking remote-sensing-derived metrics to site variables. Remote Sens. **12**(12): 2056. doi:10.3390/rs12122056.

R Core Team. 2020. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ridgeway, G. 2007. Generalized boosted models: a guide to the gbm package. Update 1.

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

Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., and Brenning, A. 2019. Performance evaluation and hyperparameter tuning of statistical and machine-learning models using spatial data. Ecol. Modell. **406**: 109–120. doi:10.1016/j.ecolmodel.2019.06.002.

Schumacher, J., Toft, H., McLean, J.P., Hauglin, M., Astrup, R., and Breidenbach, J. 2022. The utility of forest attribute maps for automated Avalanche Terrain Exposure Scale (ATES) modelling. Scand. J. For. Res. **37**(4). doi:10.1080/02827581.2022.2096921.

Seynave, I., Gégout, J.C., Hervé, J.C., Dhôte, J.F., Drapier, J., Bruno, E., and Dumé, G. 2005. Picea abies site index prediction by environmental factors and understorey vegetation: a two-scale approach based on survey databases. Can. J. For. Res. **35**: 1669–1678. doi:10.1139/x05-088.

Sharma, R.P., Brunner, A., and Eid, T. 2012. Site index prediction from site and climate variables for Norway spruce and Scots pine in Norway. Scand. J. For. Res. **27**(7): 619–636. doi:10.1080/02827581.2012.685749.

Skaugen, T., Hanssen-Bauer, I., and Førland, E.J. 2003. Adjustment of dynamically downscaled temperature and precipitation data in Norway. Technical Report 20/02 KLIMA, Norwegian Meteorological Institute, Oslo, Norway.

Skovsgaard, J.P., and Vanclay, J.K. 2008. Forest site productivity: a review of the evolution of dendrometric concepts for even-aged stands. Forestry, **81**(1): 13. doi:10.1093/forestry/cpm041.

Socha, J. 2008. Effect of topography and geology on the site index of Picea abies in the West Carpathian, Poland. Scand. J. For. Res. **23**(3): 203–213. doi:10.1080/02827580802037901.

Tompalski, P., Coops, N.C., White, J.C., and Wulder, M.A. 2015. Augmenting site index estimation with airborne laser scanning data. For. Sci. **61**(5): 861–873. doi:10.5849/forsci.14-175.

Tompalski, P., Coops, N.C., White, J.C., Goodbody, T.R., Hennigar, C.R., Wulder, M.A., et al., 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. For. Rep. **7**(1): 1–24. doi:10.1007/s40495-021-00250-z.

Tomter, S.M., Hylen, G., and Nilsen, J.E. 2010. Norway. *In* National forest inventories. *Edited by* E. Tomppo, T. Gschwantner, M. Lawrence and R.E. McRoberts. Springer, Netherlands, Dordrecht. pp. 411–424.

Tveite, B., and Braastad, H. 1981. Bonitering av gran, furu og bjork. Technical Report 4, Norsk Institutt for Skogforskning, Äs, Norway.

Wang, B., Jia, K., Liang, S., Xie, X., Wei, X., Zhao, X., et al. 2018. Assessment of Sentinel-2 MSI spectral band reflectances for estimating fractional vegetation cover. Remote Sens. **10**(12): 1927. doi:10.3390/rs10121927.

Watt, M.S., Palmer, D.J., Kimberley, M.O., Höck, B.K., Payn, T.W., and Lowe, D.J. 2010. Development of models to predict *Pinus radiata* productivity throughout New Zealand. Can. J. For. Res. **40**(3): 488–499. doi:10.1139/X09-207.

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

Weiskittel, A.R., Crookston, N.L., and Radtke, P.J. 2011. Linking climate, gross primary productivity, and site index across forests of the western United States. Can. J. For. Res. **41**: 1710–1721. doi:10.1139/x11-086.