Forest Ecology and Management 529 (2023) 120711

Contents lists available at ScienceDirect

Forest Ecology and Management

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

Development of a site form equation for predicting and mapping site quality. A case study of unmanaged beech forests in the Cantabrian range (NW Spain)

Javier Castaño-Santamaría [a,b,c], Carlos A. López-Sánchez [a], José Ramón Obeso [b], Marcos Barrio-Anta [a,*]

[a] SMartForest Research Group, Department of Biology of Organisms and Systems, Oviedo University, Mieres, Spain
[b] IMIB Biodiversity Research Institute (UO-CSIC-PA), Oviedo University, Mieres, Spain
[c] Directorate General for Cadastre, Ministry of Finance, Regional Office of Asturias, Oviedo, Spain

ARTICLE INFO

Keywords:
Dominant height
Dominant diameter
LiDAR
ALS
Random forest
Site productivity

ABSTRACT

Site form (SF), expressed as the dominant height of a stand at a reference dominant diameter, is used less often than site index (SI) to estimate forest site quality. However, it has the advantage that it is age-independent and can therefore be applied in a wider set of situations in forestry practice. Like SI, elaboration of SF has traditionally required stem analysis or repeated measurements in permanent plots; however, Airborne Laser Scanning (ALS) can nowadays be used to generate site quality practical maps, thereby simplifying the method. The aims of this study were to fit a dynamic equation to stem analysis data in order estimate SF in natural beech forests in NW Spain, as well as to compare the performance of SF for site quality estimation and to analyze the possibility of using ALS data for site quality prediction in these forests. The Algebraic Difference Approach formulation of the Bertalanffy-Richards model provided the best results and defined four curves for dominant heights of 5, 10, 15 and 20 m at a reference dominant diameter of 20 cm. A significant relationship between SF and SI was observed, and we therefore believe that SF is a good approach for site quality estimation. These results can be used directly at inventory plot level for establishing site quality classes without the need to know the stand age, either in terms of SF or translated into SI. On the other hand, ALS data allowed estimation of both dominant height and dominant diameter, although better results were obtained with the former. Applying the SF dynamic model to both data sets enabled prediction of SF at 25x25 m/pixel (and SI using the SI-SF relationship). The overall accuracy of the relationship between the observed SI and that predicted from ALS metrics yielded a coefficient of determination of 0.456 without bias, heteroscedasticity or absence of normality. The results of this unbiased raster model were considered rather good, as predictions were obtained for a pixel size of 25x25 m (0.0625 ha). Prediction of mean SI value for one hectare would thus be necessary to average the values obtained in 16 pixels with an expected error compensation. Use of the raster model based on ALS metrics will enable site quality estimation for current beech stands at high spatial resolution without the need for fieldwork, providing very valuable information for forest managers and researchers.

1. Introduction

Efficient yield forecasting and sustainable forest management require reliable measures of site productivity. Forest site productivity refers to a quantitative estimate of the potential of a site (usually by designating and summarizing the local biophysical characteristics of a forest environment) to produce plant biomass (Bontemps and Bouriaud, 2014). The terms 'site productivity' and 'site quality' can be considered equivalent when only biophysical site variables drive tree growth (i.e. absence of vegetation control, irrigation, drainage…) (Skovsgaard and Vanclay, 2008) as, for instance, in unmanaged forests.

Site index (SI), defined as the average height of the dominant trees of the stand at a specific reference age (Carmean, 1975), is by far the most frequently used indicator of site productivity. It is derived from the fact that height growth is closely correlated with stand volume productivity and that dominant height is not greatly affected by stand density or

* Corresponding author at: Address: Polytechnic School of Mieres, E-33600 Mieres, Asturias, Spain.
E-mail address: barriomarcos@uniovi.es (M. Barrio-Anta).

https://doi.org/10.1016/j.foreco.2022.120711
Received 9 October 2022; Received in revised form 30 November 2022; Accepted 1 December 2022
Available online 14 December 2022
0378-1127/

thinning treatments (assuming thinning from below) (Burkhart and Tomé, 2012). However, the method is limited to evaluating the site productivity potential of individual species or mixed stands of known age (Sharma, 2013). Hence, to evaluate the site productivity of natural (i.e. unplanted) and mixed stands when information on tree age is not known, SI cannot produce the desired result (Vanclay, 1994) unless permanent plots are established or dominant trees are drilled for core extraction or felled and dated. This is a slow, costly process and not always possible. Likewise, the use of SI is questionable for uneven-aged stands, in which the trees have irregular or polymorphic growth patterns (e.g. Huang and Titus, 1993) that are not necessarily proportionally related to age.

As an alternative to SI, some authors have proposed the height-diameter relationship as an appropriate measure of site productivity for these stands (e.g. Trorey, 1932; Meyer, 1940). In particular, McLintock and Bickford (1957) found that the height-diameter relationship of dominant trees, expressed by the monomolecular function suggested by Meyer (1940), is a sensitive and reliable site productivity measure, as confirmed in other studies (e.g. Stout and Shumway, 1982; Vanclay and Henry, 1988; Beltran et al., 2016). Vanclay (1983) proposed the term site form (SF) for this concept and defined it as the dominant height of the stand at a reference dominant diameter. Unlike SI, SF is claimed to be unaffected by species composition and age-class structure, and it is assumed that *i)* a decreasing tree taper (diameter/height ratio) is associated with increased site productivity, and *ii*) stand density does not affect the height-diameter relationship of dominant and codominant trees in these stands (Huang and Titus, 1993).

Both SF and SI can be estimated by direct and indirect methods. Direct methods are based on the relationship between the height and diameter (in the case of SF) and between the height and age (in the case of SI) of dominant trees. These relationships can be obtained directly from single measurements (e.g. Aguirre et al., 2022), repeated measurements collected in traditional forest inventories on permanent plots (e.g. Molina-Valero et al., 2019) or by stem analysis (e.g. Lappi, 1997). By contrast, indirect methods allow estimation of SF or SI from site environmental variables (soil, climate and terrain parameters or the presence or abundance of ground vegetation) and can thus be used even when trees are absent (e.g. Álvarez-Álvarez et al., 2013; Castaño-Santamaría et al., 2019).

Remote sensing techniques such as Airborne Laser Scanning (ALS) have emerged in the past decades providing alternative approaches to forest inventories without the need for traditional field sampling (e.g. Hyyppä et al., 2008). This active remote sensing technique, also referred to as Light Detection and Ranging (LiDAR), allows distance ranges to be determined from the product of the speed of light and the time required for an emitted laser to travel to a target object (Lim et al., 2003), providing three-dimensional data. Processing LiDAR data provides information on forest measurements (e.g. Nilsson, 1996; Næsset, 2002), for analysing habitats (e.g. Hyde et al., 2005), predicting fire risk (e.g. González-Olabarria et al., 2012) and determining canopy structure (e.g. Zimble et al., 2003), amongst other applications. Thus, ALS has been successfully used to estimate dominant tree height and evaluate SI in stands of known age, with data obtained in one flight (e.g. Packalén et al., 2011) or two flights separated by several years (Socha et al., 2017). In fact, the latter method is the most similar to the traditional SI repeated observations method, which demonstrates the potential value of this technique in SI determination. ALS could possibly also be used for determining SF. In terms of stand metrics and tree-level statistics, several studies have been successful in obtaining values for dominant heights (e.g. Lovell et al., 2005) and dominant diameters (e.g. Heurich and Thoma, 2008) from ALS. Thus, it may be possible *a priori* to determine site quality from SF values calculated using these data (at least with successive flights). Nevertheless, most such studies have been conducted in boreal forests (dominated by coniferous species and with a relatively homogenous structure), and it must be taken into account that the metrics calculated from laser scanning data are heavily dependent on the tree species involved (Heurich and Thoma, 2008).

Fagus sylvatica L. (hereinafter "beech") is a climax species in the Cantabrian Range (NW Spain), considered the boundary between the Euro-Siberian and Mediterranean regions, where it is restricted to slopes of elevation higher than 600 m above sea level. The crown distribution and spatial arrangement of leaves in beech trees (e.g. Collet et al., 2001) hamper the use of ALS data in these forests. This has favoured replacement of ALS with terrestrial laser scanning (TLS) in beech forest inventories (e.g. Barbeito et al., 2017). However, Heurich and Thoma (2008) obtained relatively accurate stand and tree-level metrics in beech forests by using ALS. To the best of our knowledge, no previous studies have elaborated site quality maps using both SF values and LiDAR data. Thus, the overall aim of this study was to develop a method for predicting and mapping site quality in beech forests in the Cantabrian Range, from SF, without the need to know the stand age. The specific objectives were as follows: *i)* to develop a dynamic equation for estimating SF; *ii*) to estimate SI from SF; and *iii*) to develop a model for predicting and mapping site quality by using ALS public data.

2. Materials and methods

2.1. Data

Four different types of data were used in this study for different purposes: *i)* longitudinal tree height-diameter data, obtained by stem analysis, were used to develop a site form system; *ii*) dominant heights and dominant diameters from sample plots were used to calculate the respective site form values; *iii*) site index values, expressed as the dominant height at a reference age of 80 years (see Castaño-Santamaría et al., 2019), were used to develop a site index-site form relationship; and *iv*) ALS data were used to estimate both dominant height and dominant diameter and to map the site index on the basis of site form.

2.1.1. Research plot measurements

A total of 112 sample plots were established in 2010 and 2011 in natural beech-dominated stands ($\geq$90 % of standing basal area) throughout the northwestern Cantabrian Range (in the regions of Asturias and León, NW Spain), to cover the existing range of stand structures, densities and site qualities in the area. After this inventory, the dominant height of each plot was calculated as the arithmetic mean of the 100 thickest trees per hectare (Assmann, 1970), and the number of target trees was adapted to the plot size to maintain the ratio of the 100 thickest trees per hectare. In 50 of these plots, destructive sampling of dominant trees was planned. However, the measurements had to be made in a maximum sample of 30 plots due to forest/environmental policy restrictions. These beech forests form part of the habitats of endangered and emblematic species such as the Cantabrian capercaillie and the brown bear, leading to their inclusion in protected areas relatively unaffected by human influence. Finally, 30 plots of area between 400 and 900 m^2 were selected to represent all site qualities (Fig. 1), and two dominant trees were felled and destructively sampled in each plot in 2012. The felled trees were the first two dominant trees found outside the plots, but in the same stands, within $\pm$ 5 % of the mean diameter at 1.3 m above ground level and mean height of the dominant trees (considered as the 100 largest-diameter trees per hectare). All of these trees ($n = 60$) were cross-sectioned at stump height, at 0.50 m above ground, at breast height, and 1 m intervals thereafter along the stem. Each cross section was processed by electric brushing and sanding until the tree rings were clearly visible. The treated cross sections were scanned at 900 dpi (in an Epson Expression STD 1680 PLUS flatbed scanner) and the resulting images were analyzed using WinDENDRO® tree-ring increment measurement software (Regent Instruments Canada Inc.) to count and measure the tree rings (Fig. 2).

In this study, stem analysis was applied to dominant trees, to determine the diameter at breast height and total height at each age, as these variables are required for development of SF systems. As in SI

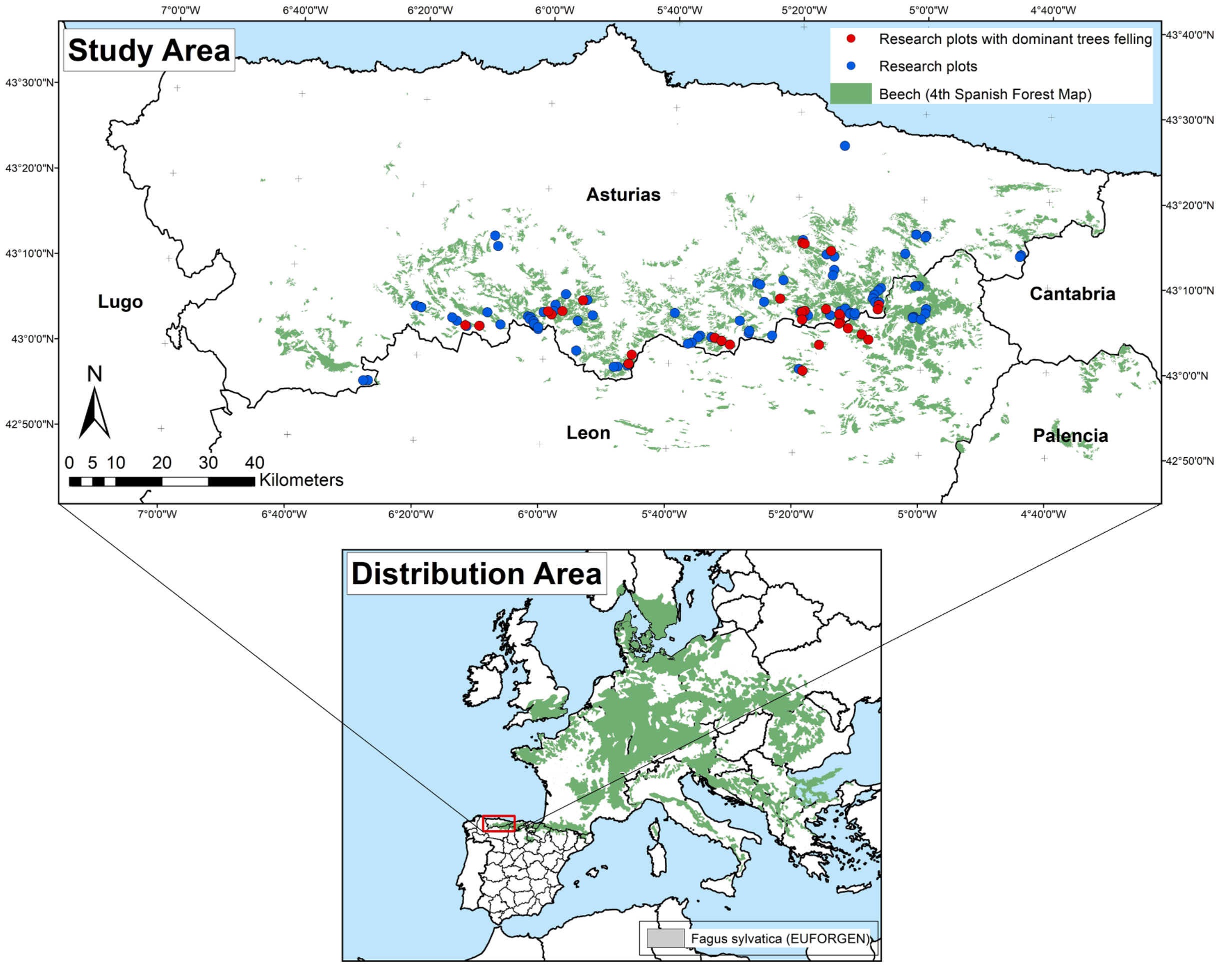

Fig. 1. Location of the study area.

studies, Carmean's algorithm, with the modification proposed by Newberry (1991), was used to reduce the bias when determining the height of each cross section at a given age. However, the practical difference from SI stem analysis was that instead of counting the number of rings (i.e. age), the diameter corresponding to that number of rings was measured. Therefore, dominant diameter was calculated as the mean of the sum of the annual diametric increments for that age in two fixed directions, on each stem disk obtained at 1.3 m above ground section, without bark. Height and diameter were plotted against age for each tree to detect any abnormal growth patterns. As a result, one tree was rejected.

We used the dynamic equation (Eq. (1)) previously developed by Castaño-Santamaría et al. (2019), with the same destructively sampled trees and the same fitting methodology, to calculate the SI in each plot:

$$SI = \frac{23.8753 + X_0}{1 + 20526.03/X_0 \bullet 80^{-1.51}}$$

$$X_0 = \frac{1}{2}\left(H_0 - 23.8753 + \sqrt{(H_0 - 23.8753)^2 + 4 \bullet 20526.03 \bullet H_0 \bullet t^{-1.51}}\right) \quad (1)$$

where SI = site index (m) (dominant height at a reference age of 80 years), H_0 = dominant height (m) and t = age of dominant trees (years).

In addition to the information obtained from dominant trees destructively sampled in 30 plots, the stand-related variables dominant height and dominant diameter were also recorded, as these are the most appropriate variables for relating to the ALS data. Summary statistics, including mean, maximum, minimum and standard deviation values for the main tree and stand variables, are shown in Table 1.

2.1.2. Airborne laser scanning data

The Airborne Laser Scanning (ALS) survey took place in the period May-October 2012 as part of the Spanish PNOA-LiDAR project. The ALS data set corresponds to the first round of countrywide ALS measurements, which are publicly available in Spain through the National Plan for Aerial Orthophotography (hereafter referred to as PNOA-LiDAR). Square ALS tiles of area 2 km side in LASer (LAS) binary files were obtained from the National Geographic Information Centre (CNIG, 2022) computer server (http://centrodedescargas.cnig.es/CentroDescargas/index.jsp). The scanning sensor used to collect the ALS data was a RIEGL LMS-Q680i. The point cloud was captured with up to four returns measured per pulse and a mean density of 0.5 points/m^2 and vertical RMSE $\leq$ 0.20 m. The ALS data sets were processed using several processing programmes implemented in the FUSION/LDV software (McGaughey, 2014). A detailed description of the software parametrization and all the processing workflow of ALS point cloud is given by Novo-Fernández et al. (2019). Briefly, all ALS echoes classified as ground were used to create a digital elevation model (DEM) considering a spatial resolution of 5 m, and a predefined threshold of between 2 and 50 m above ground level was used to compute canopy cover metrics using a buffer of 25 m of radius from the coordinates of the sample plots. In total, 36 ALS metrics widely used as effective variables for height and diameter estimations (e.g. Næsset, 2002, 2004) were computed as independent variables (Table 2, Fig. 2).

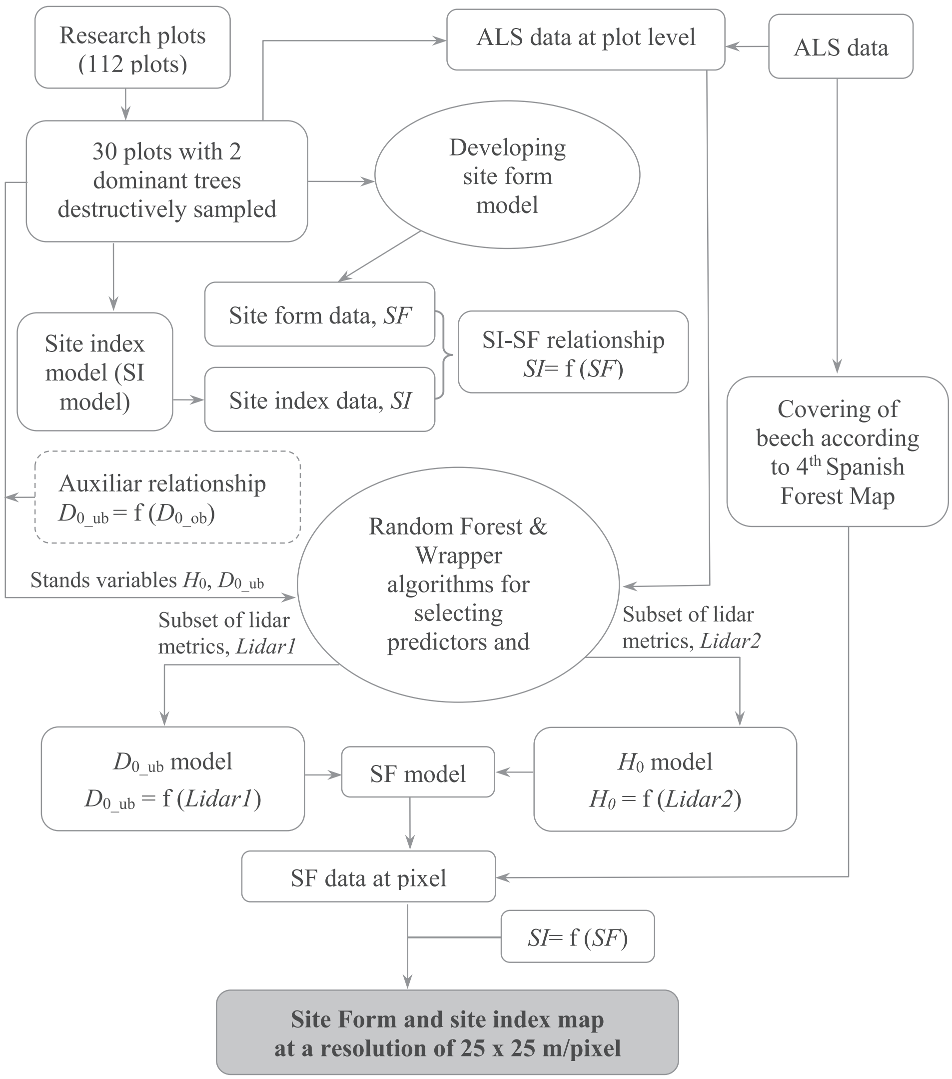

Fig. 2. Workflow adopted for modelling and mapping the site quality throughout ALS metrics in this study.

Table 1
Summary statistics for individual tree and stand variables.

Statistic	Tree variable (59 dominant trees)				Stand variable (30 plots)		
	Diameter at 1.3 m (cm)	Height (m)	Age (years)	Site index (m at 80 yr)	Number of trees (ha^{-1})	Basal area (m^2/ha)	Dominant height (m)
Minimum	12.90	9.77	43	7.62	300	22.51	10.37
Maximum	69.60	38.63	215	26.76	2445	94.95	35.90
Mean	39.10	20.21	109.17	15.81	1073.70	44.04	19.25
Standard deviation	14.20	6.50	41.88	4.81	575.40	13.46	6.02

2.2. Developing a dynamic site form equation

The hierarchical structure of the data sets used in dominant height–diameter models (i.e. repeated measurement data from dominant or codominant trees in different stands) usually leads to a lack of independence among the observations (West et al., 1984). In order to solve this problem, the algebraic difference approach (ADA; Bailey and Clutter, 1974) and its generalization (GADA; Cieszewski and Bailey, 2000), which includes an autoregressive error structure model, were used to develop the site form system. We tested three well-known base models traditionally used in the development of site quality models: Korf (cited in Lundqvist, 1957), Hossfeld IV (Hossfeld, 1882) modified by Molina-Valero et al. (2019), and Bertalanffy-Richards (Bertalanffy, 1949; 1957; Richards, 1959) (Table 3). In all models, an additional constant of 1.3 was added to the right-hand side of the equation to force the curves to pass through the point (D_0 = 0, H_0 = 1.3) (Moreno-Fernández et al., 2018; Molina-Valero et al., 2019).

The dummy variables method (Cieszewski et al., 2000), in which a second-order continuous-time autoregressive error structure (CAR2) is included to account for auto-correlation, was used to estimate the model parameters (see Diéguez-Aranda et al., 2005, for more details). The dummy variables method and the CAR2 error structure were programmed using the SAS/ETS® MODEL procedure (SAS Institute Inc., 2004), which allows dynamic updating of the residuals. We carried out a cross-validation, estimating the residuals in dominant height estimation for both trees from one plot by fitting the model without these two trees from that plot. The root mean square error (RMSE), which assesses the precision of the estimates, and the adjusted pseudo-coefficient of

Table 2
Summary of Airborne Laser Scanning (ALS) metrics for each plot.

	ALS Metrics		Description
Height metrics	Metrics expressing the central trend in ALS height distribution	h_{mean}	mean
		h_{mode}	mode
	Metrics expressing the dispersion of ALS height distribution	h_{SD}	Standard deviation
		h_{VAR}	variance
		h_{AAD}	absolute average deviation
		h_{IQ}	interquartile range
		h_{CV}	coefficient of variation
		h_{max}, h_{min}	maximum and minimum
	Metrics expressing the shape of ALS height distribution	h_{SKW}	Skewness
		h_{Kurt}	Kurtosis
		CRR	canopy relief ratio ((mean height – min height)/(max height – min height)
	Percentiles of the ALS height distribution	h_{01}, h_{05}, … h_{99}	1st, 5th, 10th, 20th, 25th, 30th, 40th, 50th, 60th, 70th, 75th, 80th, 90th, 95th, 99th percentiles
Canopy cover metrics	Fixed height break threshold (HBT)	CC	percentage of first returns above 2.00 m/total all returns
		PARA2	percentage of all returns above 2.00 m/total all returns
		ARA2/TFR	ratio between all returns above 2.00 m and total of first returns
	Variable HBT	PFRAM	percentage of first returns above mean/total all returns
		PARAM	percentage of all returns above mean/total all returns
		PARAMO	percentage of all returns above mode/total all returns
		PFRAMO	percentage of first returns above mode/total all returns
		ARAM/TFR	ratio between all returns above mean and total first returns
		ARAMO/TFR	ratio between all returns above mode and total first returns

determination (Ryan, 1997) for nonlinear regression (R^2), which indicates the proportion of the variance of the dependent variable explained by the model, were calculated from the residuals obtained from cross-validation. In addition, graphical analysis of the residuals and of the appearance of the fitted curves overlaid on the trajectories of the dominant height of the plots was also conducted.

Finally, we used the method proposed by Molina-Valero et al. (2019) to select the reference dominant diameter. This consists of using different base diameters and their corresponding observed heights to estimate heights at other diameters (both forward and backward) for each tree; it also involves comparing the results with the values obtained from stem analysis by using the relative error in predictions.

2.3. Comparative performance of site form as a site quality estimator and development of an SI-SF relationship

Vanclay and Henry (1988) indicated the four characteristics that an index (SF in this case) should have in order to be considered a good measure of site quality: *i*) it must be reproducible and consistent over time, *ii*) it must be indicative of the site and not influenced by the stand conditions or management history, *iii*) it must be correlated with the productive potential of the site, and *iv*) it must be at least as good as any other available measures of productivity.

To analyze the consistency of the index over time and its capacity to estimate productivity, we considered that SI is the most widely used and recognized indicator of site quality, and we then compared similarities in the prediction uncertainties of SI and SF. For this purpose, we first searched for a relationship between SI and SF, conducting a SI-SF correlation analysis at two different levels: *i*) at individual tree level for measurement values and *ii*) at plot level for values obtained with SI and SF equations. For the first analysis, we had available 46 dominant trees of age (*t*) and diameter under bark (D_{0_ub}) greater than 80 years and 20 cm (the reference age and diameter required to determine SI and SF respectively) for the stem analysis. On the individual trajectories H_0-D_{0_ub} and H_0-*t* of each tree, we interpolated the height values corresponding to 80 years and 20 cm to obtain the true values of SI and SF of each tree and then conduct correlation analysis. For the second analysis, we had available 30 plots in which dominant height, dominant diameter and age (mean of the age of two dominant trees) were known. These data were used to calculate SI (with Eq.1) and SF (with Eq.3).

The second characteristic states that stand density does not affect the height-diameter relationship of dominant and codominant trees in these stands (Huang and Titus, 1993). To assess this criterion, we computed the Pearson correlation coefficient (*r*) of the predicted SI and SF against Relative Spacing index (RS) (Hart, 1928; Becking, 1953), as proposed by Molina-Valero et al. (2019). The RS index is used to characterize the growing stock level and is calculated by dividing the average distance between trees by the dominant height and expressing this as a percentage. The RS index is a useful parameter in stand density management because it is generally independent of site quality and stand age, and because dominant height growth is one of the best criteria, from a biological point of view, for establishing thinning intervals (Barrio-Anta and Álvarez-González, 2005). As the beech forests under study are natural forests, we assumed triangular spacing between trees, so that RS can be expressed as follows:

$$RS = \frac{\sqrt{20000/N\sqrt{3}}}{H_0} \times 100 \tag{2}$$

where *RS* is the relative spacing index (%), *N* is the number of stems per hectare, and H_0 is the dominant height (m).

On the other hand, as indicated in the introduction, SF is usually used as a site quality index under the assumption that the tree taper (diameter/height ratio) decreases with increasing site productivity. Therefore,

Table 3
Dynamic equations used for fitting SF curves.

Base equation	Site-related parameters	Solution for X with initial values (D_0, H_0)	Dynamic equation
Korf: $H = a_1 exp(-a_2 D^{-a_3})$	$a_2 = X$	$X_0 = -ln\left(\frac{H_0 - 1.3}{a_1}\right) {D_0}^{a_3}$	$H = 1.3 + b_1 \left(\frac{H_0 - 1.3}{b_1}\right)^{\left(\frac{D_0}{D}\right)^{b_3}}$
Hossfeld IV: $H = \frac{a_1}{1 + a_2 D^{-a_3}}$	$a_1 = X a_2 = \frac{b_2}{X}$	$X_0 = \frac{1}{2}\left[(H_0 - 1.3) + \sqrt{(H_0 - 1.3)^2 - 4 b_2 {D_0}^{-b_3} (1.3 - H_0)}\right]$	$H = 1.3 + \frac{X_0}{1 + \frac{b_2}{X_0} {D_0}^{-b_3}}$
Bertalanffy-Richards: $H = a_1 (1 - exp(-a_2 D))^{a_3}$	$a_2 = X$	$X_0 = \frac{-ln\left(1 - \left(\frac{H_0 - 1.3}{b_1}\right)^{\frac{1}{b_3}}\right)}{D_0}$	$H = 1.3 + b_1 \left(1 - \left(1 - \left(\frac{H_0 - 1.3}{b_1}\right)^{\left(\frac{1}{b_3}\right)}\right)^{\left(\frac{D}{D_0}\right)}\right)^{b_3}$

; where *H* is the dominant height (m), *D* is the dominant diameter (cm) and a_x and b_x are model parameters to be estimated.

to evaluate the third characteristic we used the Pearson correlation coefficient (*r*) between tree taper and predicted SF for each tree.

Conversion of SF to SI values may become necessary on some occasions, such as in growth and yield studies because most growth and yield models require SI as an input variable (e.g. for yield tables). For this purpose, the pairs of SF-SI values were plotted and a model relating the two variables was fitted after visual inspection of the scatter plot.

2.4. Model development for predicting and mapping site quality from ALS data

As numerous PNOA-LiDAR flights were not available in this case, the ALS public data had to be combined with data from the inventory and dominant tree felling phases. To predict site quality by means of SF from ALS data, the following steps must be carried out: *i*) development of a dynamic SF equation and a relationship between SI and SF, *ii*) development of a relationship between D_0 over and under bark, and *iii*) prediction of H_0 and D_0 over bark in the ALS data.

At this point we have already developed a dynamic SF equation and the relationship between SI and SF is available. As we used stem analysis, the dominant diameters are under bark, and the dynamic equation developed refers to the dominant height at a reference dominant diameter under bark. The ALS point cloud hits the external surface of the trees, and thus estimates heights and diameters over bark (e.g. Næsset, 2002). Therefore, the relationship between D_0 over bark (D_{0_ob}) and D_0 under bark (D_{0_ub}) must be established.

For predicting H_0 and D_0 over bark through ALS metrics, ideally there should be temporal coincidence between the research plot measurements (H_0 and D_0) and the PNOA-LiDAR data acquisitions. In this case, the research plots were measured in 2010 and 2011 and the ALS data were acquired in 2012. Harmonization procedures would be necessary if the models were intended to be run with data from new available PNOA-LiDAR flights, e.g. as ALS-based-models for yield estimation (e.g. Novo-Fernández et al., 2019). This small temporary discrepancy does not represent a methodological problem as beech is a slow growing forest species and also because our model will not run with new ALS data, so dependent variables (H_0 and D_0) can be used with predictor variables that are not exactly temporally coincident. Therefore, we predicted H_0 and D_0 from the 36 ALS metrics indicated in Table 2, by using the Random Forest (RF) non-parametric ensemble learning method (Breiman, 2001).

RF is a widely used non-parametric classification and regression approach that consists of building an ensemble of decision trees (Gislason et al., 2006). The success of this technique is based on the use of numerous trees and different independent variables that are randomly selected from the complete original set of features (e.g. Deschamps et al., 2012). For this purpose, WEKA open-source software (Hall et al., 2009) was used to fit the RF algorithm by implementing a wrapper methodology to select the subsample of variables that usually produces the best results (Zhiwei and Xinghua, 2010). This method selects the subsample of variables by using a learning algorithm as part of the evaluation function. The final fitted models were applied to ALS data, to generate a spatially continuous map of H_0 and D_0 at a resolution of 25 m/pixel.

The 10-fold cross-validation approach was used to test the accuracy of the algorithm. This process consists of the following four steps: *i*) splitting the data set into 10 random subsets of roughly the same size; *ii*) fitting the model 10 times, sequentially omitting one subset each time; *iii*) repeating step *i* 10 times, and *iv*) using each of the fitted models to produce pseudo-independent predictions on the omitted subset, as an indicator of how well the classifier will perform on unseen data. The pseudo-coefficient of determination (R^2) (Ryan, 1997) and the root mean squared error (RMSE) were used to assess the model performance. For implementation of machine learning algorithms, WEKA has an embedded feature-ranking technique called the variable importance measure (VIM), which was used to guide selection of predictors for the final model. To ensure that values of variable importance were expressed on comparable scales, the VIM values were normalized so that they summed to a unit value (normalized importance) and were also expressed in relative values (relative importance).

Finally, considering the predictive algorithms for H_0 and D_0, the dynamic SF equation, the D_{0_ub}-D_{0_ob} relationship and the SI-SF relationship, it was possible to generate a raster map of resolution 25 m/pixel for both SF and SI for the area currently occupied by beech forests, obtained from the most recent Spanish Forest Map (scale 1:25000) (MITECO, 2012). We graphically summarize the main methodological steps of the approach used in the present study in Fig. 2.

3. Results

3.1. Dynamic site form equation

All of the models tested achieved convergence in the Marquardt iterative fitting process, but one parameter from the derived Korf-based model was not significant, and therefore this model was discarded. The GADA formulation of the Hossfeld IV equation modified by Molina-Valero et al. (2019) and the ADA formulation of the Bertalanffy-Richards equation yielded good fits and were considered for further analysis as all of their parameter estimates were significant at the 5 % level (Table 4). Visual comparison of the fitted curves overlaid on the trajectories of the observed dominant heights from stem analysis revealed the good performance of both models. However, the Bertalanffy-Richards' curves fitted better to the real trajectories, while the Hossfeld IV model was less realistic in terms of growth. Together with slightly better fitting statistics and higher RMSE value (Fig. S1), this led us to choose the Bertalanffy-Richards' model (Fig. 3). The plot of residuals against estimated values showed a random pattern of residuals around zero, with homogeneous variance and no detectable significant trends, after modelling the error structure following Diéguez-Aranda et al. (2005) (Fig. S2). The autocorrelation parameters were all significant, but are not included in Table 4 because the sole purpose of correction for autocorrelation was to obtain unbiased and efficient estimates of the parameters (Parresol and Vissage, 1998), and it would have no use in practical applications unless the same individual was being measured repeatedly.

In selecting the reference dominant diameter, a diameter of 20 cm was superior for predicting height at other diameters (Fig. S3), because it presents a good compromise between a low relative error (RE%) and the number of observations. Therefore, the dynamic equation for determining the site form (SF) is as follows:

$$SF = 1.3 + 46.00879\left(1-\left(1-\left(\frac{H_0-1.3}{46.00879}\right)^{\left(\frac{1}{1.00830}\right)}\right)^{\left(\frac{20}{D_{0\_ub}}\right)}\right)^{1.00830} \quad (3)$$

where $SF =$ site form (m) (dominant height at a reference dominant diameter under bark of 20 cm), $H_0 =$ dominant height (m) and $D_{0\_ub} =$ dominant diameter under bark (cm).

As practical use of dynamic SF equations requires knowledge of dominant diameter under bark, a model is required for predicting this parameter from the usual determined dominant diameter over bark. The

Table 4
Parameter estimates and goodness-of-fit statistics for the three models tested.

Model	Parameter	p-value	R^2adj	RMSE (m)
Korf	$b_1 = 8095.32$	0.3843	0.9851	0.8217
	$b_2 = 0.127237$	<0.0001		
Hossfeld IV GADA	$b_1 = 0$	–	0.9854	0.8148
	$b_2 = 10182.15$	<0.0001		
	$b_3 = 1.001685$	<0.0001		
Bertalanffy-Richards ADA	$b_1 = 46.00879$	<0.0001	0.9857	0.8070
	$b_3 = 1.00830$	<0.0001		

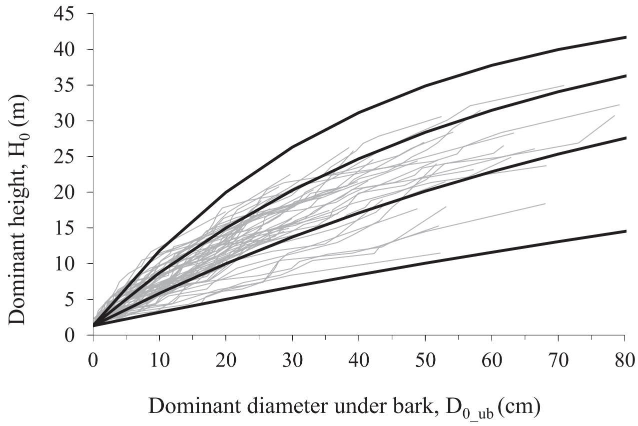

Fig. 3. Dominant height-dominant diameter growth curves for SF values of 5, 10, 15 and 20 m at a reference dominant diameter of 20 cm overlaid on the trajectories of the observed values.

model developed for this purpose was fitted using the data pairs of the 60 felled trees, yielding a strongly linear relationship ($R^2 = 0.9999$; $p < 0.0001$), expressed as follows:

$$D_{0\_ub} = 0.980972 \bullet D_{0\_ob} \quad (4)$$

where D_{0_ub} = dominant diameter under bark and D_{0_ob} = dominant diameter over bark.

3.2. Consistency of SF as a site quality indicator

Analysis of the standardized true values of SI and SF, obtained directly by stem analysis of the 46 dominant trees, revealed a positive correlation ($r = 0.6669$; $p < 0.0001$). On the other hand, analysis of the standardized values of SI and SF at plot level (obtained from the respective equations) again revealed a strong positive correlation between both variables ($r = 0.783$; $p < 0.0001$).

Regarding the influence of stand density, as expected, the RS index was not significantly correlated with either SI ($r = -0.331$; $p = 0.079$) or SF ($r = -0.361$; $p = 0.063$). These results indicate that SF can be considered a consistent indicator of site quality in beech forests across different management regimes.

On the other hand, the diameter/height ratio was strongly negatively correlated with SF ($r = -0.664$; $p < 0.0001$). This result can be expected a priori because like SF, tree taper is calculated by an allometric relationship between diameter and height.

Finally, and as a result of the SI-SF relationship at plot level, we propose the model shown in Fig. 4 for conversion of SF into SI values. This model performs well ($R^2 = 0.6524$, $p < 0.0001$) with absence of bias.

3.3. Predicting and mapping site form and site index from ALS data

As a result of the ALS variable selection process, dominant height was only able to be estimated from height-related metrics (specifically, 95th, 99th and 90th percentiles and h_{CV} of height distribution), and the model performed very well after 10-fold cross-validation ($R^2 = 0.92$; RMSE = 1.57 m). The variable that contributed most to the model was the 95th percentile, which contributed 36.5 % of the VIM_R (Table 5). The next most important variables, with an VIM_R of 30.3 %, were the 99th and 90th percentiles; the tree variables together accounted for 97.2 % of the VIM_R.

D_{0_ob} was related to an optimal subset size of 8 of the 36 variables, and the model performance was poorer ($R^2 = 0.48$; RMSE = 10.13 cm) (Table 5). The features that contribute most to estimating D_{0_ob} are height-related metrics, which accumulated 85.9 % of the VIM_R. The contribution of the canopy cover metrics was rather lower, with 14.1 % of accumulated VIM_R. The variable that contributed most to the model was the standard deviation of ALS heights (h_{SD}), which contributed 21.2 % of VIM_R. The ALS height 99th, 75th and 70th percentiles were the next most important variables, with an accumulated VIM_R of 49.5 %.

Table 5
Variables included in the H_0 and D_0 models developed from ALS data, including their variable importance. To ensure values of variable importance are expressed on a comparable scale for each of the response variables, the scores of all the predictors selected were normalized or are expressed as relative values. Normalized importance (VIM_N) = (VIM-VIMmin)/(VIMmax-VIMmin), Relative importance (VIM_R) = (VIM/ΣVIM)·100. R^2 and RMSE are the goodness-of-fit statistics obtained from 10-fold cross validation with 10 repetitions.

Model	Variable	VIM	VIM_N	VIM_R (%)	R^2	RMSE
H_0	h_{95}	654	0.92	36.5	0.9168	1.5719
	h_{90}	543	0.75	30.3		
	h_{99}	543	0.75	30.3		
	h_{CV}	51	0.00	2.8		
D_0	h_{SD}	4124	1.00	21.2	0.4849	10.1257
	h_{99}	4028	0.97	20.7		
	h_{75}	2988	0.63	15.4		
	h_{70}	2604	0.50	13.4		
	ARA2/TFR	1682	0.20	8.7		
	h_{20}	1539	0.16	7.9		
	h_{min}	1388	0.11	7.1		
	ARAMO/TFR	1064	0.00	5.5		

where h_{99}, h_{95}, h_{90}, h_{75}, h_{70} and h_{20} are the corresponding percentiles of ALS height distribution, h_{CV} is the coefficient of variation of ALS heights, h_{SD} is the standard deviation of ALS heights, ARA2/TFR is the ratio between all returns above 2.00 m and total of first returns, h_{min} is the minimum of height distribution and ARAMO/TFR is the ratio between all returns above mode and total first returns.

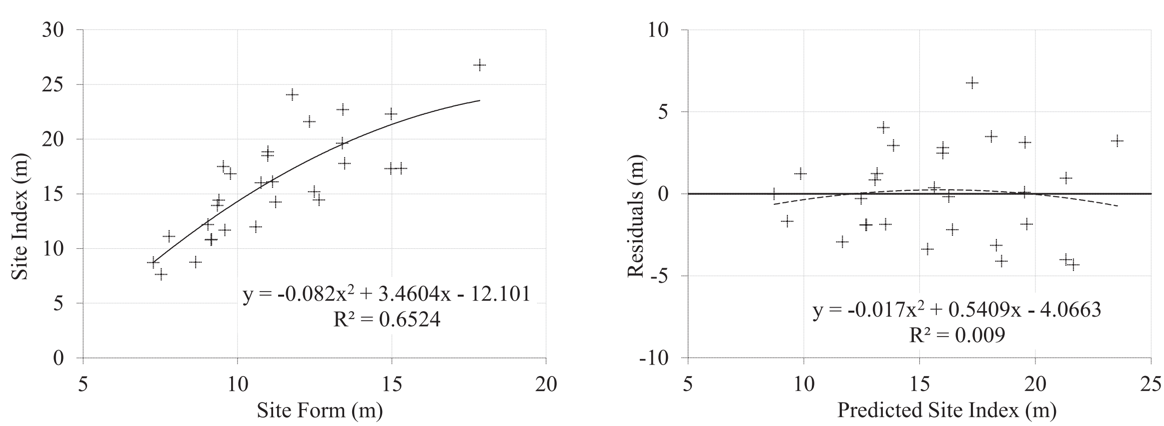

Fig. 4. Relationships between SI and SF (left) and model residuals (right). SI and SF were calculated at plot level using Eq.1 and Eq.3 respectively.

Together, these four most important variables accumulated a VIM_R of 70.8 % (Table 5). Finally, the D_{0_ob} obtained through ALS metrics must be converted into D_{0_ub} by means of Eq. 4 to be used to determine SF.

After prediction of SF (by applying Eq. 3 with predictor variables H_0 and D_{0_ub} derived from ALS metrics), SI was subsequently obtained for each pixel by using the previously obtained SI-SF relationship (Fig. 4). At this point and considering that several equations were used to estimate the SI estimation at the pixel level, it would be very useful to estimate the overall accuracy of SI. For this purpose, we carried out graphical analysis of estimated against predicted values of SI for each of the 30 experimental plots (Fig. 5). In these plots, estimated site index (SI) values were obtained using Eq. (1) (dominant height and stand age known) and SI was predicted from ALS site form, as explained above. The linear model fitted to the scatter plot did not reveal any important problems related to bias, heteroscedasticity or lack of normality, yielding a coefficient of determination of 0.45, which can be considered quite good, as predictions are made for a resolution of 25x25 m/pixel. Finally, Fig. 6 shows the raster map of SI from SF for the current beech forests.

4. Discussion

4.1. Dynamic site form equation

In this study, a reference dominant diameter was used instead of the reference age for constructing site quality models. An adaptation of the Bertalanffy-Richards' model was chosen to describe SF relationships for beech, fulfilling most of the desirable properties that a site quality equation should possess (Diéguez-Aranda et al., 2006). There is no consensus in the literature regarding which particular function form is most appropriate for developing SF models. However, the Bertalanffy-Richards' equation has traditionally been used with good results both in SI (e.g. Monserud, 1984; Nord-Larsen, 2006) and in SF studies (e.g. Beltran et al., 2016; Ahmadi et al., 2017), as its biological behaviour is as important as its goodness of fit metrics (Ivancich et al., 2011). Statistically, our model performed well ($R^2 = 0.9857$ and RMSE $= 0.8070$ m), yielding more accurate values than those obtained in other SF studies using the Bertalanffy-Richards' equation (none for beech) (e.g. Beltran et al., 2016; Ahmadi et al., 2017).

Although beech is a widespread forest tree in Europe, very few SF curves have been developed for the species. However, in Spain two studies have elaborated SF models at a national scale for beech (among many other species) based on National Forest Inventory (NFI) data (Moreno-Fernández et al., 2018; Aguirre et al., 2022). Both studies used a single observation of D_0 and H_0 for each plot and developed the model

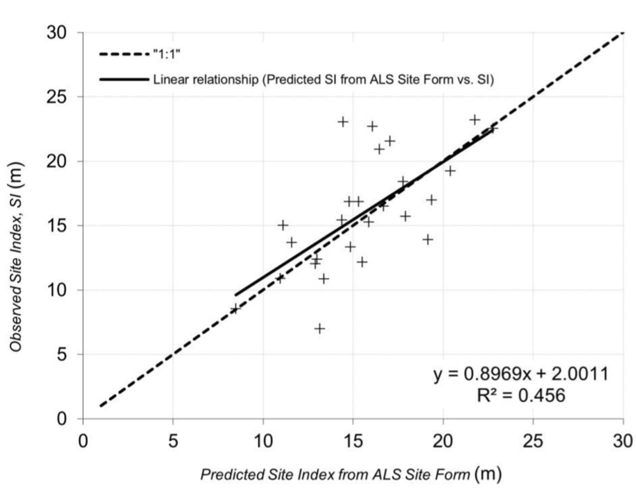

Fig. 5. SI observed in each plot (calculated by applying Eq.1) *vs* SI predicted or estimated from ALS metrics (right).

by using the guide-curve method. Our SF model fitted well to the observed values of the stem analysis and distinguished four site qualities defined by heights of 5, 10, 15 and 20 m at a reference dominant diameter of 20 cm. Moreno-Fernández et al. (2018) did not indicate the reference diameter, and therefore comparison with our results was not possible. However, Aguirre et al. (2022) used the same equation as Moreno-Fernández et al. (2018) (Hossfeld II) but with a different parameter estimate and a reference dominant diameter of 30 cm. Both mathematical and visual comparison of the curve with the model developed by Aguirre et al. (2022) shows that intermediate qualities have heights around 6–7 m higher at a dominant diameter of 80 cm. This difference was reduced to around 4 m in the worst quality and was reversed in the best quality (height around 2 m lower) (see Fig. S4). The differences (and magnitude of these) can mainly be attributed to two factors: *i*) the different methods of site index curve construction (guide curve *vs* stem analysis) and *ii*) differences in the study area. From our point of view, the first factor is the most important as the guide curve method only fits height-diameter pairs on temporary plots and does not adequately capture trend of the data. This generally leads to underestimation of dominant height, and therefore of SF, for diameters larger than the reference diameter (see Figure S4) as demonstrated in numerous studies (e.g. Monserud, 1985; Thrower and Goudie, 1992). Considering the second factor, our study is restricted to the Cantabrian Range in NW Spain, while the two aforementioned studies consider the beech forests of the whole country, including rather different environmental conditions and higher and lower site qualities (see Sánchez-Palomares et al., 2004).

4.2. Using site form as a site quality indicator

Our results fulfil the four criteria suggested by Vanclay and Henry (1988) for considering SF a good measure of site quality. First of all, the trees used in the study were aged between 43 and 215 years, and thus SF can be used over the entire rotation of the species in Spain, between 100 and 150 years (Madrigal et al., 2008). Moreover, analysis of the standard deviations in SI and SF estimates for each plot showed that the uncertainties associated with both methods were usually similar, which indicates similar uncertainty in both indices when predicting site quality for different stand development stages of the same plot. Secondly, the correlations indicate that stand density does not affect the height-diameter relationship of dominant and codominant trees in these beech stands, thus corroborating the results obtained in other studies (e.g. Duan et al., 2018; Fu et al., 2018; Molina-Valero et al., 2019). Thirdly, the assumption that the tree taper decreases as site quality increases was also fulfilled by SF ($r = -0.664$; $p < 0.0001$). This correlation was expected because, by definition, higher values of SF represent higher values of dominant height at equal dominant diameter. This is a common result (e.g. Larson, 1963; Molina-Valero et al., 2019), although not obvious, as positive correlations (e.g. Buda and Wang, 2006) and even no correlation (Wang, 1998) have also been reported. Finally, the significant relationship between SF and SI are shown by Fig. 4 and the correlation results. Duan et al. (2018) added a fifth criterion: height growth over time is asymptotic, whereas diameter is not. The selected model (see Fig. 3) also fulfils this criterion.

Some studies have analyzed the relationship between SI and SF for a given species (e.g. Huang and Titus, 1993; Wang, 1998; Beltran et al., 2016; Duan et al. 2018 or Molina-Valero et al., 2019 among others). Relative to SI, the number of site quality studies using SF is much lower, and among these, studies using stem analysis are scarce (e.g. Wang, 1998; Buda and Wang, 2006; Beltran et al., 2016). From an academic point of view, the two methodologies do not share the same target forest stand. SI refers essentially to even-aged stands (Skovsgaard and Vanclay, 2008), while SF was originally proposed for uneven-aged or mixed-species stands (Huang and Titus, 1993). However, several studies have shown the usefulness of SF as a reliable estimator of site quality in pure even-aged stands (e.g. Beltran et al., 2016; Moreno-Fernández et al.,

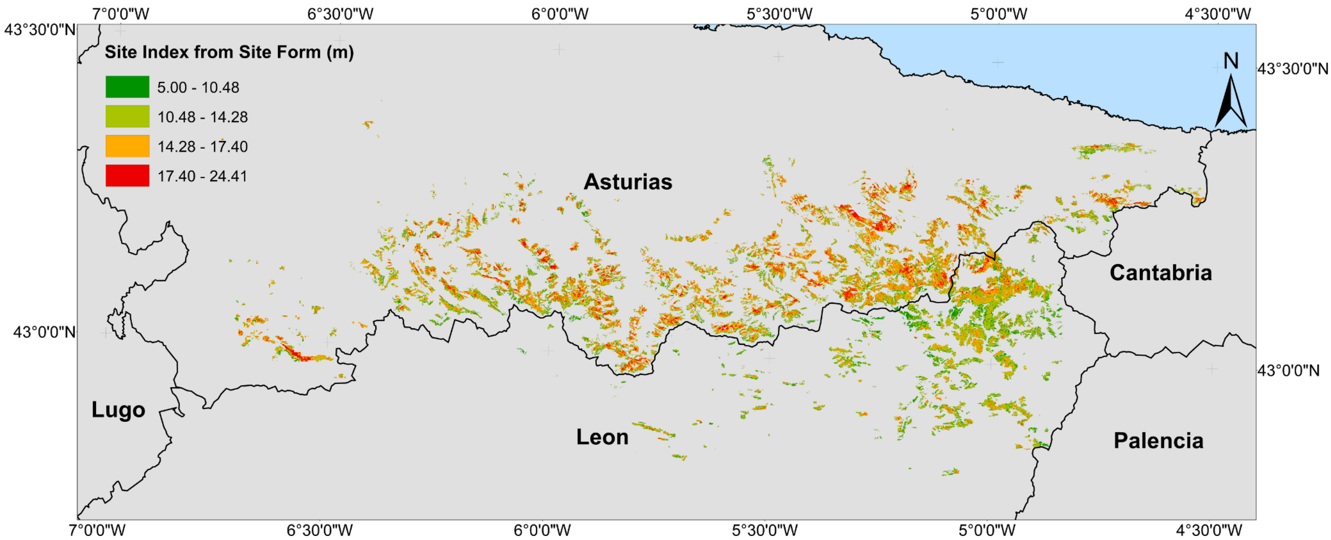

Fig. 6. SI predictions for current beech forests in the study area.

2018; Molina-Valero et al., 2019). In fact, our results are very similar to those obtained for the SI model developed by Castaño-Santamaría et al. (2019) ($R^2 = 0.9882$ and RMSE = 0.74 for SI model *vs* $R^2 = 0.9857$ and RMSE = 0.807 for SF model).

The main advantage of SF over SI is that it does not require information about age, measurement of which is very difficult and costly in natural forests. Only dominant height and dominant diameter values are needed, and these are easily obtained in traditional field inventories. This makes SF suitable for estimating site quality from existing NFI data in most countries, where stand age is not generally recorded (Molina-Valero et al., 2019). Therefore, SF model and SI-SF relationship may have two important practical applications: *i*) SF can be used directly to establish site class classification in the field (Eq. 3), converting previously known D_{0_ob} into D_{0_ub} (Eq. 4) and *ii*) users can also convert SF to SI by using this relationship (Fig. 4). The first application may be sufficient for researchers who only require establishing site quality classes for use as independent variables to carry out further ecological studies. The second application is necessary for users who are also interested in predicting growth and yield variables (i.e. volume, average annual volume increment, biomass, carbon content, etc.) as growth models including the SF index have not yet been developed.

4.3. Predicting and mapping site quality from ALS data

In this study, SI (pixel level) was predicted in four stages: *i*) H_0 and D_{0_ob} estimation from ALS data, *ii*) conversion of D_{0_ob} into D_{0_ub} (Eq. 4), *iii*) SF calculation (Eq. 3), and *iv*) SI prediction (Fig. 4). According to the results, dominant height estimation is the most accurate, showing a strong correlation with the 90, 95 and 99th height distribution percentiles. This has already been reported for different species and is based on the dominant height concept (e.g. Holopainen et al., 2010; Packalén et al., 2011; González-Ferreiro et al., 2012; Socha et al., 2017). On the other hand, dominant diameter estimation also yielded acceptable results, although not as accurate as the dominant height estimation. According to Vauhkonen et al. (2014), prediction of tree diameter using ALS metrics is a challenging task due to several uncertainties (e.g. accuracy in the measurement of ground height, and its effect on breast height determination), as well as the limited detection rates of trees below the dominant tree layer. This is evident in beech forests due to their closed canopies and spatial positioning of the leaves after leaf emergence (e.g. Collet et al., 2001), as in this case. Several studies use low density ALS data to estimate diameter at breast height (e.g. Næsset, 2002; Fu et al., 2020) or diameter distributions (e.g. Packalén and Maltamo, 2008; Räty et al., 2020). However, the number of studies that calculate the dominant diameter using ALS metrics is much smaller. For example, Heurich and Thoma (2008) estimated dominant diameter as a function of height measurements, as we did, although the variables are not identical. These researchers reported a strong correlation between dominant diameter and 90th, 40th and 20th height distribution percentiles for German beech forests, with $R^2 = 0.67$ for a laser point density of 5 points m^{-2}. We obtained lower values for model precision. As our laser point density was lower (0.5 points m^{-2}), the difference in model precision may be due to this difference in point density, although numerous studies have shown that the influence of point density on the estimates is negligible (e.g. Strunk et al., 2012; Jakubowski et al., 2013).

In addition to the low ALS point density, the discrepancy between the pixel size used to obtain ALS metrics (25x25 m) and the field plots (range between 20x20 m and 30x30 m) could also lead to some inconsistencies in the results. We choose this pixel size (25 m × 25 m) because we found it was the best compromise for dealing with the range of plot size.

The overall accuracy of our SI estimation approach ($R^2 = 0.456$) would be considered rather low if we were dealing with traditional growth and yield models (not spatially explicit). However, our SI model, which uses ALS metrics as predictors, is an unbiased raster model and predictions are obtained by applying the model over the territory (spatially explicit model). Considering that predictions are applied for a pixel size of 25x25 m (0.0625 ha), prediction of SI in one hectare would require averaging the values obtained in 16 pixels. This average result (in one hectare) can be considered a rather good estimation taking into account the compensation of errors that occur in unbiased models.

This methodology therefore provides the site quality estimation with an adequate resolution at the forest scale.

Castaño-Santamaría et al. (2019) developed a SI raster model as a function of environmental variables for the same study area with a resolution of 250x250 m/pixel (i.e. 6.25 ha), which is an adequate resolution at landscape level but not at forest scale. There are advantages and disadvantages to both the previous and the present approaches. Thus, the environmental-variables-based methodology allows site quality estimates to be obtained without the need for the species to be present, although with low spatial resolution, providing very useful information for expanding beech forest through reforestation. By contrast, the present methodology enables site quality estimation at high spatial resolution but only for current beech stands, providing very valuable information for carrying out forest management plans or for further research purposes.

Socha et al. (2017) showed that the use of multi-temporal ALS data allows site index to be modelled, although it yielded slightly poorer results than those obtained by the stem analysis method that they also developed for the same Norway spruce stands in Poland. If the stand age and the time elapsed between the two flights are known, pairs of dominant height-age values can be obtained, in the same way as if permanent sample plots were measured twice. However, stand age is essential data. Several flights in successive years would allow us to obtain pairs of H_0-D_0 values and thus to estimate site quality in terms of site form, but without the need to know the stand age. As to date we only have the coverage used in this study, testing this approach will be the subject of a future study.

5. Conclusions

The present study showed that SF model is a reliable site quality estimator for natural beech forests in north-western Spain. Our findings can be used directly to establish site quality classes at inventory plot level without the need to know the stand age (in terms of SF), or by translating it into the more widely used SI, by applying the SI-SF relationship. The advantage of the SF model is that it is age-independent and can be used in a wide range of situations where age is unknown or/and costly to obtain. The findings also showed that dominant height and dominant diameter can be correctly estimated using ALS metrics computed from the first coverage of the Spanish PNOA-LiDAR project. Despite the low point density of ALS, we were able to use these data to develop an unbiased site form raster model with a resolution of 25x25 m/pixel, allowing estimation of site quality in beech forest stands in the study area without the need for fieldwork. The approach developed thus allow users to obtain good site quality estimates for beech forests at two levels: *i*) at inventory plot level by using the SF equation (and eventually the SI-SF relationship), *ii*) at forest scale using the SF and/or SI raster map at a resolution of 25x25 m/pixel. Remeasurement of new experimental plots combined with new ALS data, such as the recent 2021 s coverage (data not yet available) with a higher laser density point (2 points m^{-2}), will enable exploration of improvements in this type of site quality model.

CRediT authorship contribution statement

Javier Castaño-Santamaría: Data curation, Formal analysis, Writing – original draft. **Carlos A. López-Sánchez:** Conceptualization, Formal analysis, Supervision, Writing – review & editing. **José Ramón Obeso:** Supervision. **Marcos Barrio-Anta:** Conceptualization, Formal analysis, Supervision, Writing – review & editing.

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

While undertaking the present study, the first author was in receipt of a Severo Ochoa Fellowship from the Asturias Government-FICYT (code 09/111). The authors are grateful to the Asturias Government Forest Service for valuable help and also for financing the felling of the trees used in the study. The authors also thank the Asturias and Castilla y León Government Forest and Environmental Services for access to the forests and for granting permission to carry out inventories. Our thanks to Mr José Manuel Álvarez González and Dr María Castaño Díaz for invaluable help during field sampling and laboratory tasks. Finally, the authors thank two anonymous referees for their highly constructive, valuable comments and suggestions.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foreco.2022.120711.

References

Aguirre, A., Moreno-Fernández, D., Alberdi, I., Hernández, L., Adame, P., Cañellas, I., Montes, F., 2022. Mapping forest site quality at national level. For. Ecol. Manage. 508, 120043 https://doi.org/10.1016/j.foreco.2022.120043.

Ahmadi, K., Alavi, S.J., Kouchaksaraei, M.T., 2017. Constructing site quality curves and productivity assessment for uneven-aged and mixed stands of oriental beech (*Fagus orientalis* Lipsky) in Hircanian forest. Iran. Forest Sci. Technol. 13, 41–46. https://doi.org/10.1080/21580103.2017.1292959.

Álvarez-Álvarez, P., Barrio-Anta, M., Cámara-Obregón, A., dos Santos Bento, J.M.R., 2013. Ground vegetation as an indicator of site quality: effect of non-site factors on the productivity of newly established chestnut plantations in northwestern Spain. J. For. Res. 18 (5), 407–417. https://doi.org/10.1007/s10310-012-0361-2.

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

Bailey, R.L., Clutter, J.L., 1974. Base-age invariant polymorphic site curves. For. Sci. 20 (2), 155–159. https://doi.org/10.1093/forestscience/20.2.155.

Barbeito, I., Dassot, M., Bayer, D., Collet, C., Drössler, L., Löf, M., del Rio, M., Ruiz-Peinado, R., Forrester, D.I., Bravo-Oviedo, A., Pretzsch, H., 2017. Terrestrial laser scanning reveals differences in crown structure of *Fagus sylvatica* in mixed vs. pure European forests. For. Ecol. Manage. 405, 381–390. https://doi.org/10.1016/j.foreco.2017.09.043.

Barrio-Anta, M., Álvarez-González, J.G., 2005. Development of a stand density management diagram for even-aged pedunculate oak stands and its use in designing thinning schedules. Forestry 78, 209–216. https://doi.org/10.1093/forestry/cpi033.

Becking, J.H., 1953. Einige Gesichtspunkte für die Durchführung von vergleichenden Durchforstungsversuchen in gleichaltrigen Beständen. En: Proc. 11th Congress IUFRO. Rome, pp. 580-582.

Beltran, H.A., Chauchard, L., Velásquez, A., Sbrancia, R., Martínez Pastur, G., 2016. Diametric Site Index: an alternative method to estimate Site Quality in *Nothofagus obliqua* and *N. alpina* forests. CERNE 22 (3), 345–354. https://doi.org/10.1590/01047760201622032207.

Bertalanffy, L.v., 1949. Problems of organic growth. Nature 163, 156–158. https://doi.org/10.1038/163156a0.

Bertalanffy, L.v., 1957. Quantitative laws in metabolism and growth. Q. Rev. Biol. 32, 217–231. https://doi.org/10.1086/401873.

Bontemps, J.D., Bouriaud, O., 2014. Predictive approaches to forest site productivity: recent trends, challenges and future perspectives. Forestry 87 (1), 109–128. https://doi.org/10.1093/forestry/cpt034.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Buda, N.J., Wang, J.R., 2006. Suitability of two methods of evaluating site quality for sugar maple in central Ontario. For. Chron. 82, 733–744. https://doi.org/10.5558/tfc82733-5.

Burkhart, H.E., Tomé, M., 2012. Modeling Forest Trees and Stands. Springer, London.

Carmean, W.H., 1975. Forest site quality evaluation in the United States. Adv. Agron. 27, 209–269. https://doi.org/10.1016/S0065-2113(08)70011-7.

Castaño-Santamaría, J., López-Sánchez, C.A., Obeso, J.R., Barrio-Anta, M., 2019. Modelling and mapping beech forest distribution and site productivity under different climate change scenarios in the Cantabrian Range (North-western Spain). For. Ecol. Manage. 450, 117488 https://doi.org/10.1016/j.foreco.2019.117488.

Cieszewski, C.J., Bailey, R.L., 2000. Generalized algebraic difference approach: theory based derivation of dynamic equations with polymorphism and variable asymptotes. For. Sci. 46, 116–126. https://doi.org/10.1093/forestscience/46.1.116.

Cieszewski, C.J., Harrison, M., Martin, S.W., 2000. Practical methods for estimating non-biased parameters in selfreferencing growth and yield models. University of Georgia, Athens, Ga. PMRC-TR 2000–7.

CNIG, 2022. Spanish National Geographic Information Centre. ALS data available at http://centrodedescargas.cnig.es/CentroDescargas/buscadorCatalogo.do?codFamilia=LIDAR#). last accessed on 18 July 2022.

Collet, C., Lanter, O., Pardos, M., 2001. Effects of canopy opening on height and diameter growth in naturally regenerated beech seedlings. Ann. For. Sci. 58 (2), 127–134. https://doi.org/10.1051/forest:2001112.

Deschamps, B., McNairn, H., Shang, J., Jiao, X., 2012. Towards operational radar-only crop type classification: comparison of a traditional decision tree with a random forest classifier. Can. J. Remote Sens. 38, 60–68. https://doi.org/10.5589/m12-012.

Diéguez-Aranda, U., Burkhart, H.E., Rodríguez-Soalleiro, R., 2005. Modeling dominant height growth of radiata pine (*Pinus radiata* D. Don) plantations in north-western Spain. For. Ecol. Manage. 215, 271–284. https://doi.org/10.1016/j.foreco.2005.05.015.

Diéguez-Aranda, U., Burkhart, H.E., Amateis, R.L., 2006. Dynamic site model for Loblolly pine (*Pinus taeda* L.) plantations in the United States. For. Sci. 52 (3), 262–272. https://doi.org/10.1093/forestscience/52.3.262.

Duan, G., Gao, Z., Wang, Q., Fu, L., 2018. Comparison of Different Height-Diameter Modelling Techniques for Prediction of Site Productivity in Natural Uneven-Aged Pure Stands. Forests 9, 63. https://doi.org/10.3390/f9020063.

Fu, L., Lei, X., Sharma, R.P., Li, H., Zhu, G., Hong, L., You, L., Duan, G., Guo, H., Lei, Y., Li, Y., Tang, S., 2018. Comparing height–age and height–diameter modelling approaches for estimating site productivity of natural uneven-aged forests. Forestry 91, 419–433. https://doi.org/10.1093/forestry/cpx049.

Fu, L., Duan, G., Ye, Q., Meng, X., Luo, P., Sharma, R.P., Sun, H., Wang, G., Liu, Q., 2020. Prediction of individual tree diameter using nonlinear mixed-effects modelling approach and airbourne LiDAR data. Remote Sens. 12, 1066. https://doi.org/10.3390/rs12071066.

Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R., 2006. Random Forests for land cover classification. Pattern Recogn. Lett. 27 (4), 294–300. https://doi.org/10.1016/j.patrec.2005.08.011.

González-Ferreiro, E., Diéguez-Aranda, U., Miranda, D., 2012. Estimation of stand variables in *Pinus radiata* D. Don plantations using different LiDAR pulse densities. Forestry 85, 281–292. https://doi.org/10.1093/forestry/cps002.

González-Olabarria, J.R., Rodríguez, F., Fernández-Landa, A., Mola-Yudego, B., 2012. Mapping fire risk in the Model Forest of Urbión (Spain) based on airborne LiDAR measurements. For. Ecol. Manage. 282, 149–156. https://doi.org/10.1016/j.foreco.2012.06.056.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: An update. SIGKDD Explorations 11 (1), 10–18. https://doi.org/10.1145/1656274.1656278.

Hart, H.M.J., 1928. Stamtal en Dunning. Proefstation Boschwesen, Batavia, Mededelingen, p. 21.

Heurich, M., Thoma, F., 2008. Estimation of forestry stand parameters using laser scanning data in temperate, structurally rich natural European beech (*Fagus sylvatica*) and Norway spruce (*Picea abies*) forests. Forestry 81 (5), 645–661. https://doi.org/10.1093/forestry/cpn038.

Holopainen, M., Vastaranta, M., Haapanen, R., Yu, X., Hyyppä, J., Kaartinen, H., Viitala, R., Hyyppä, H., 2010. Site-type estimation using airborne laser scanning and stand register data. Photogramm. J. Finl. 22, 16–32.

Hossfeld, J.W., 1882. Mathematik für Forstmänner, Ökonomen und Cameralisten (Gotha, 4 Bd., S. 310).

Huang, S., Titus, S.J., 1993. An index of site productivity for uneven-aged or mixed-species stands. Can. J. For. Res. 23, 558–562. https://doi.org/10.1139/x93-074.

Hyde, P., Dubayah, R., Peterson, B., Blair, J., Hofton, M., Hunsaker, C., Knox, R., Walker, W., 2005. Mapping forest structure for wildlife habitat analysis using waveform LiDAR: Validation of montane ecosystems. Remote Sens. Environ. 96, 427–437. https://doi.org/10.1016/j.rse.2005.03.005.

Hyyppä, J., Hyyppä, H., Leckie, D., Gougeon, F., Yu, X., Maltamo, M., 2008. Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. Int. J. Remote Sens. 29, 1339–1366. https://doi.org/10.1080/01431160701736489.

Ivancich, H., Martínez-Pastur, G., Peri, P.L., 2011. Modelos forzados y no forzados para el cálculo del índice de sitio en bosques de *Nothofagus antarctica* en Patagonia Sur. Bosque 32 (2), 135–145. https://doi.org/10.4067/S0717-92002011000200004.

Jakubowski, M.K., Guo, Q., Kelly, M., 2013. Tradeoffs between LiDAR pulse density and forest measurement accuracy. Remote Sens. Environ. 130, 245–253. https://doi.org/10.1016/j.rse.2012.11.024.

Lappi, J., 1997. A longitudinal analysis of height/diameter curves. For. Sci. 43 (4), 555–570 https://doi.org/forestscience/43.4.555.

Larson, P.R., 1963. Stem form development of forest trees. For. Sci. 9, 1–42. https://doi.org/10.1093/forestscience/9.s2.a0001.

Lim, K., Treitz, P., Wulder, M., St-Onge, B., Flood, M., 2003. LiDAR remote sensing of forest structure. Prog. Phys. Geogr. 27 (1), 88–106. https://doi.org/10.1191/0309133303 pp360ra.

Lovell, J.L., Jupp, D.L.B., Newnham, G.J., Coops, N.C., Culvenor, D.S., 2005. Simulation study for finding optimal LiDAR acquisition parameters for forest height retrieval. For. Ecol. Manage. 214, 398–412. https://doi.org/10.1016/j.foreco.2004.07.077.

Lundqvist, B., 1957. On the height growth in cultivated stands of pine and spruce in Northern Sweden. Medd. Fran Statens Skogforsk. 47 (2), 1–64.

Madrigal, A., Calama, R., Madrigal, G., Aunós, A., Reque, J.A., 2008. Selvicultura de *Fagus sylvatica* L. In: Serrada, R., Montero, G., Reque, J.A. (Eds.), CompEndio De Selvicultura Aplicada En España. Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria-Ministerio de Educación y Ciencia-Fundación Conde del Valle Salazar, Madrid, pp. 155–185.

McGaughey, R.J., 2014. FUSION/LDV: Software for LiDAR Data Analysis and Visualization. March 2014, v. 3.42. US Department of Agriculture, Forest Service. Pacific Northwest Research Station: Seattle, USA.

McLintock, T.F., Bickford, C.A., 1957. A proposed site index for red spruce in the Northeast. U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station. Upper Darby, PA.

Meyer, H.A., 1940. A mathematical expression for height curves. J. For. 38, 415–420. https://doi.org/10.1093/jof/38.5.415.

MITECO, 2012. Mapa Forestal de España de maxima actualidad. Ministerio para la Transición Ecológica y el Reto Demográfico. https://www.miteco.gob.es/es/cartografia-y-sig/ide/descargas/biodiversidad/mfe.aspx.

Molina-Valero, J.A., Diéguez-Aranda, U., Álvarez-González, J.G., Castedo-Dorado, F., Pérez-Cruzado, C., 2019. Assessing site form as an indicator of site quality in even-aged *Pinus radiata* D. Don stands in north-western Spain. Ann. For. Sci. 76, 113. https://doi.org/10.1007/s13595-019-0904-1.

Monserud, R.A., 1984. Height growth and site-index curves for inland Douglas-fir based on stem analysis data and forest habitat type. For. Sci. 30, 943–965. https://doi.org/10.1093/forestscience/30.4.943.

Monserud, R.A., 1985. Comparison of Douglas-fir site index and height growth curves in the Pacific Northwest. Can. J. For. Res. 15, 673–679. https://doi.org/10.1139/x85-110.

Moreno-Fernández, D., Álvarez-González, J.G., Rodríguez-Soalleiro, R., Pasalodos-Tato, M., Cañellas, I., Montes, F., Díaz-Varela, E., Sánchez-González, M., Crecente-Campo, F., Álvarez-Álvarez, P., Barrio-Anta, M., Pérez-Cruzado, C., 2018. National-scale assessment of forest site productivity in Spain. For. Ecol. Manage. 417, 197–207. https://doi.org/10.1016/j.foreco.2018.03.016.

Næsset, E., 2002. Predicting forest stand characteristics with airborne 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., 2004. Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. Scand. J. For. Res. 19, 164–179. https://doi.org/10.1080/02827580310019257.

Newberry, J.D., 1991. A note on Carmean's estimate of height from stem analysis data. For. Sci. 37, 368–369. https://doi.org/10.1093/forestscience/37.1.368.

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

Nord-Larsen, T., 2006. Developing dynamic site index curves por European beech (*Fagus sylvatica* L.) in Denmark. For. Sci. 52, 173–181. https://doi.org/10.1093/forestscience/52.2.173.

Novo-Fernández, A., Barrio-Anta, M., Recondo, C., Cámara-Obregón, A., López-Sánchez, C.A., 2019. Integration of National Forest Inventory and nationwide airborne laser scanning data to improve forest yield predictions in North-Western Spain. Remote Sens. 11 (14), 1693. https://doi.org/10.3390/rs11141693.

Packalén, P., Maltamo, M., 2008. Estimation of species-specific diameter distributions using airborne laser scanning and aerial photographs. Can. J. For. Res. 38, 1750–1760. https://doi.org/10.1139/X08-037.

Packalén, P., Mehtätalo, L., Maltamo, M., 2011. ALS-based estimation of plot volume and site index in a eucalyptus plantation with nonlinear mixed-effect model that accounts for the clone effect. Ann. For. Sci. 68, 1085–1092. https://doi.org/10.1007/s13595-011-0124-9.

Parresol, B.R., Vissage, J.S., 1998. White pine site index for the southern forest survey. Research Paper SRS-10. U.S. Department of Agriculture, Forest Service. Southern Research Station, Asheville, NC, p. 8.

Räty, J., Packalén, P., Kotivuori, E., Maltamo, M., 2020. Fusing diameter distributions predicted by and area-based and individual-tree detection in coniferous dominated forests. Can. J. For. Res. 50, 113–125. https://doi.org/10.1139/cjfr-2019-0102.

Richards, F.J., 1959. A flexible growth function for empirical use. J. Exp. Bot. 10, 290–300. https://doi.org/10.1093/jxb/10.2.290.

Ryan, T.P., 1997. Modern Regression Methods. John Wiley & Sons, New York.

Sánchez-Palomares, O., Rubio Sánchez, A., Blanco, A., 2004. Definición y cartografía de las áreas potenciales fisiográfico-climáticas de hayedo en España. Invest. Agrar.: Sist. y Recur. For. Fuera de Serie, 13-62.

SAS Institute Inc., 2004. SAS/ETS 9.1 User's Guide. SAS Institute Inc., Cary, NC, USA.

Sharma, R.P., 2013. Modelling height, height growth and site index from national forest inventory data in Norway. Norwegian University of Life Sciences. PhD Thesis.

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.

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.

Stout, B.B., Shumway, D.L., 1982. Site Quality Estimation Using Height and Diameter. For. Sci. 28 (3), 639–645. https://doi.org/10.1093/forestscience/28.3.639.

Strunk, J., Temesgen, H., Andersen, H.E., Flewelling, J.P., Madsen, L., 2012. Effects of LiDAR pulse density and sample size on a model-assisted approach to estimate forest inventory variables. Can. J. Remote Sens. 38, 644–654. https://doi.org/10.5589/m12-052.

Thrower, J.S., Goudie, J.W., 1992. Estimating dominant height and site index for even-aged interior Douglas-fir in British Columbia. West. J. Appl. For. 7, 20–25. https://doi.org/10.1093/wjaf/7.1.20.

Trorey, L.G., 1932. A mathematical method for the construction of diameter height curves based on site. For. Chron. 8, 121–132. https://doi.org/10.5558/tfc8121-2.

Vanclay, J.K., 1983. Techniques for modelling timber yield from indigenous forests with special reference to Queensland. University of Oxford. MSc. Thesis.

Vanclay, J.K., 1994. Modelling forest growth and yield: applications to mixed tropical forests. CAB International, Wallingford.

Vanclay, J.K., Henry, N.B., 1988. Assessing site productivity of indigenous cypress pine forest in southern Queensland. Commonw. For. Rev. 67, 53–64.

Vauhkonen, J., Packalén, P., Malinen, J., Pitkänen, J., Maltamo, M., 2014. Airborne laser scanning-based decision support for wood procurement planning. Scan. J. For. Res. 29, 132–143. https://doi.org/10.1080/02827581.2013.813063.

Wang, G.G., 1998. Is height of dominant trees at a reference diameter an adequate measure of site quality? For. Ecol. Manage. 112, 49–54. https://doi.org/10.1016/S0378-1127(98)00315-6.

West, P.W., Ratkowsky, D.A., Davis, A.W., 1984. Problems of hypothesis testing of regressions with multiple measurements from individual sampling units. For. Ecol. Manage. 7, 207–224. https://doi.org/10.1016/0378-1127(84)90068-9.

Zhiwei, X., Xinghua, W., 2010. Research for information extraction based on wrapper model algorithm. 2010 Second International Conference on Computer Research and Development. Kuala Lumpur, Malaysia, pp. 652–655.

Zimble, D.A., Evans, D.L., Carlson, G.C., Parker, R.C., Grado, S.C., Gerard, P.D., 2003. Characterizing vertical forest structure using small-footprint airborne LiDAR. Remote Sens. Environ. 87, 171–182. https://doi.org/10.1016/S0034-4257(03)00139-1.