



## Evaluation and validation of forest models: Insight from Mediterranean and scots pine models in Spain

A. Vázquez-Veloso<sup>a,b,\*</sup>, V. Pando<sup>a,c</sup>, C. Ordóñez<sup>a,b</sup>, F. Bravo<sup>a,b</sup>

<sup>a</sup> Instituto de Investigación en Gestión Forestal Sostenible (iuFOR), ETS de Ingenierías Agrarias, Universidad de Valladolid, Avda. de Madrid 44, 34004 Palencia, Spain

<sup>b</sup> Vegetal Production and Forest Resources Department. Higher Technical School of Agricultural Engineering (Palencia), University of Valladolid, Spain

<sup>c</sup> Statistics and Operations Research Department. Higher Technical School of Agricultural Engineering (Palencia), University of Valladolid, Spain

### ARTICLE INFO

Original content: [IBERO evaluation and validation \(Original data\)](#)

**Keywords:**  
Survival  
Ingrowth  
Growth  
Modelling  
Accuracy

### ABSTRACT

Forest models predict tree and stand evolution under different scenarios, thereby supporting decision-making in forest management. Models are complex structures composed of sub-models that estimate forest variables at tree and stand levels. Prediction accuracy has generally been evaluated independently of the model. Integrated sub-models make forest models easier to use and provide predictions for growth, survival, ingrowth and many other tree and stand variables with reduced effort. However, while individual submodel validation is widely practiced and normally done by each author individually, joint model validation remains less explored. This study deploys a useful methodology for evaluating and validating models. After comparing observed and predicted data, several case studies were then proposed to improve the accuracy of the joint model. We used the IBERO model, data from the Spanish National Forest Inventory and the SIMANFOR simulator platform. The accuracy of growth submodels was improved by calibrating their equations, though accuracy was not improved in survival and ingrowth submodels.

### 1. Introduction

Quantifying and classifying are tasks that simplify forest dynamics to understand and manage them more easily and make predictions under different scenarios. Forest modelling has been evolving steadily in recent decades with the development of submodels to predict growth and yield for the main forest species (Bravo et al., 2011; Pretzsch, 2009c). Different simulation platforms have also been developed to make the models easier to use (Pretzsch, 2009a). Examples include the SILVA simulator, a single tree-based stand simulator developed for species in North Europe (Pretzsch et al., 2002); SIMANFOR, a web-based application for simulating silvicultural alternatives (Bravo et al., 2012, 2023); or Capsis open software for modelling forest growth (Dufour-Kowalski et al., 2012). The forestry sector is actively encouraging the establishment and advancement of extensive forestry databases, such as national forest inventories (Tomppo et al., 2010) and long experimental networks across climate condition gradients (Pardos et al., 2021). These facilitate the development, testing and use of growth and yield forest models at larger scales, making them more robust.

Accurate predictions are key to robust outcomes. While the

prediction accuracy of each submodel (defined here as each individual equation or group of equations to estimate a process, such as growth or survival, or a variable, such as height or biomass) is normally evaluated as part of the model development process, it is not a common practice at the whole-model level. Combining the different submodels into a single model allows predictions to be made based on all the processes that occur at tree and/or stand level but increases the difficulty of evaluating their performance. Model validation is defined as the procedure used to calculate the accuracy of a model in representing intended use in the real world (Ling and Mahadevan, 2013). Weiskittel et al. (2011) offer some guidance about how to approach this task, and other authors (Aldea et al., 2023; Pretzsch, 2009b; Zhu et al., 2019) have developed examples of how to validate different submodels. However, model validation must combine the evaluation of each fundamental area: growth, survival, and ingrowth. The goal of this study is to establish the basis for developing a methodology for validating forest models using direct quantitative comparison of model predictions and observed values for diameter, height growth, survival, and ingrowth submodels. To do so, we designed a method for validating the IBERO model (Bravo, 2005) parameterizations for *Pinus pinaster* IBERO<sup>PT</sup> (Lizarralde et al., 2010a) and *Pinus*

\* Corresponding author at: Instituto de Investigación en Gestión Forestal Sostenible (iuFOR), ETS de Ingenierías Agrarias, Universidad de Valladolid, Avda. de Madrid 44, 34004 Palencia, Spain.

E-mail address: [aitor.vazquez.veloso@uva.es](mailto:aitor.vazquez.veloso@uva.es) (A. Vázquez-Veloso).

<https://doi.org/10.1016/j.ecoinf.2023.102246>

Received 19 May 2023; Received in revised form 31 July 2023; Accepted 31 July 2023

Available online 1 August 2023

1574-9541/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

*syvestris* IBERO<sup>PS</sup> (Lizarralde et al., 2010b) on the SIMANFOR simulator (Bravo et al., 2010, 2012, 2023; SIMANFOR, 2023).

## 2. Material and methods

### 2.1. Data

Data from the 2nd (NFI2) and 3rd (NFI3) edition of the Spanish National Forest Inventory were used. This dataset is independent from the one used to fit the original models but has the same circular concentric plot structure (Alberdi et al., 2010). Data from NFI2 were used to develop projections and generate NFI3 predicted data (NFI3<sub>predicted</sub>), while actual data from NFI3 (NFI3<sub>observed</sub>) was used to check the accuracy of those predictions. Data were filtered to fulfill the following criteria:

- Plots must be within the defined geographical boundaries of the two parameterizations.
- Dominant species for each model must be *Pinus pinaster* or *Pinus sylvestris*, accordingly (dominant plots species means that the basal area of the target species represents >90% of the total basal area in the plot).
- Plots must be measured in both inventories (NFI2 and NFI3) and have had no structure-modifying events (i.e., harvests or wildfires).
- Mean plot age must be known for both inventories due to the model requirements.

After filtering the database, a total of 49 plots with 575 *P. pinaster* trees and 136 plots with 2810 *P. sylvestris* trees were selected. Since the time between measurements was not the same for each plot, data from NFI3 plots were modified to make the time lapse between inventories equivalent to 10 years and facilitate analysis of growth submodels. This was done by recalculating NFI3 tree growth for diameter and height in a linear way, dividing total growth by the years between NFI3 and NFI2 measurements and multiplying the result by 10 (years). The same approach was used on SMC-ORGANON (Hann et al., 2006); it makes the plot age of the predicted data (2 projections of 5 years each) compatible with that of the observed data (NFI3<sub>observed</sub>) to enable comparison. For survival and ingrowth submodels, original data was used without time-lapse modifications because there is no data for the moment when a tree dies or reaches a certain diameter category.

Annual rainfall (mm·year<sup>-1</sup>) and mean annual temperature (°C·year<sup>-1</sup>) were obtained from the State Meteorological Agency of the Spanish Government (AEMET, 2021) to evaluate the possibility of improving the survival equation predictions. Relevant information was obtained from the closest meteorological station to each plot for the measurement year. Forestry and climate data were scrubbed and curated using R (R Core Team, 2021) and tidyverse software packages (Wickham et al., 2019).

### 2.2. Model and parameterizations

The IBERO model (Bravo, 2005) and its parameterizations for *Pinus pinaster* IBERO<sup>PT</sup> (Lizarralde et al., 2010a) and *Pinus sylvestris* IBERO<sup>PS</sup> (Lizarralde et al., 2010b) were used. IBERO is an individual-tree growth model and independent from the distance originally parameterized for *Pinus pinaster mesogeensis* and *Pinus sylvestris*. The core model integrates equations to estimate forest productivity (Bravo and Montero, 2001; Bravo-Oviedo et al., 2004), diameter and height growth (Lizarralde, 2008), survival (Bravo-Oviedo et al., 2006) and ingrowth (Bravo et al., 2008). It also includes imputation equations such as crown and height-diameter equations (Lizarralde, 2008), taper equations to estimate volume (Lizarralde, 2008), biomass submodels (Ruiz-Peinado and del Río, 2011) and others. The IBERO model is integrated with the SIMANFOR simulator (Bravo et al., 2012, 2023) and adapted to the needs of this study for each analysis, as described in the analysis section. IBERO was

parameterized for *P. pinaster* stands located in the Iberian Meridional Mountains (Soria, Guadalajara, Cuenca and Teruel provinces) of Eastern Spain (IBERO<sup>PT</sup>, full parameterization content on SIMANFOR (2022a)) and for *P. sylvestris* in the Iberian and Central Mountains (Ávila, Burgos, Segovia and Soria provinces) of Central Spain (IBERO<sup>PS</sup>, full parameterization content on SIMANFOR (2022b)).

Both original models were used to assess the differences between predicted and observed data, then modified after analyzing each case study. Submodels with the correct shape but predictions that needed improvement were calibrated by modifying their scalar value (Bravo and Montero, 2003). The original submodels and their parameterizations are summarized on the Supplementary Material section.

Growth submodels were evaluated first by direct quantitative comparison between predicted and observed data. Since survival and ingrowth submodels are complex, we looked at case studies that involved modifying both original models. We suggested adding a threshold value based on stand density capacity to the survival submodel, to determine when it needs to be applied, following the methodology of Rodríguez de Prado et al. (2020). It operates as ‘border control’ in deciding when an event must occur: when the threshold condition is surpassed, the survival submodel is activated. This methodology is based on the Stand Density Index (SDI) (Reineke, 1933) and the Maximum Stand Density Index (SDI<sub>max</sub>) (Rodríguez de Prado et al., 2020). The specific coefficients used in our submodels were developed by Del Río et al. (2006) for *P. pinaster* and *P. sylvestris*. In that case study, if  $SDI > SDI_{max}$ , then the survival submodel was applied. For the ingrowth submodel, we decided to study the influence of “technical ingrowth” on the predictions. Technical growth is the ingrowth that affects the outer circles of the concentric plots and the higher diameter classes by extension. To do this, we used the ingrowth of trees in the first measurable diameter class, which ranges from 7.5 to 12.5 cm diameter at breast height (dbh). In three other case studies, the original submodel threshold values were modified to 0.25, 0.5 and 0.75. The original ingrowth submodel threshold values for were established at 0.38 for IBERO<sup>PT</sup> and 0.43 for IBERO<sup>PS</sup> by Bravo et al. (2008).

### 2.3. Simulations

To perform growth simulations, the SIMANFOR platform (Bravo et al., 2010, 2012, 2023; SIMANFOR, 2023) was used. SIMANFOR allows users to run previously parameterized models with datasets and silviculture defined by the users. It generates predictions of stand dynamics at tree and stand levels after each step (projection or harvest) in the silvicultural scenario. For faster results, we ran the simulations using the SIMANFOR version installed on the Caléndula High-Performance Computer (SCAYLE, 2019) at the University of León.

The simulations were developed by following a sequence of case studies to validate the models. The first step evaluated diameter growth and calibrates the data as needed. The second step evaluated height growth; then the third and fourth steps evaluated survival and ingrowth submodels in a parallel manner based on the case studies mentioned in the previous section. After evaluating each case study, a modified submodel with improvements (i.e., submodel calibration) or test implementations (i.e., modification of the ingrowth threshold value) was created to develop a new case study and evaluate it. The simulation flowchart for the full model is shown in Fig. 1.

In each simulation, plots from NFI2 are projected over 10 years (divided into 2 projections of 5 years each) to obtain the NFI3<sub>predicted</sub> data; plots from NFI3 were modified to make the time elapsed between NFI2 and NFI3 inventories equivalent to 10 years (as explained in the data section) and obtain the NFI3<sub>observed</sub> data. NFI3<sub>observed</sub> data was run on the SIMANFOR initialization process, which provides a larger dataset by inputting missing variables (i.e., dominant height) while preserving the original inventory data (i.e., diameter and height). While NFI2 data must be simulated again under the conditions of the different case studies to obtain NFI3<sub>predicted</sub>, NFI3<sub>observed</sub> data remains the same for all

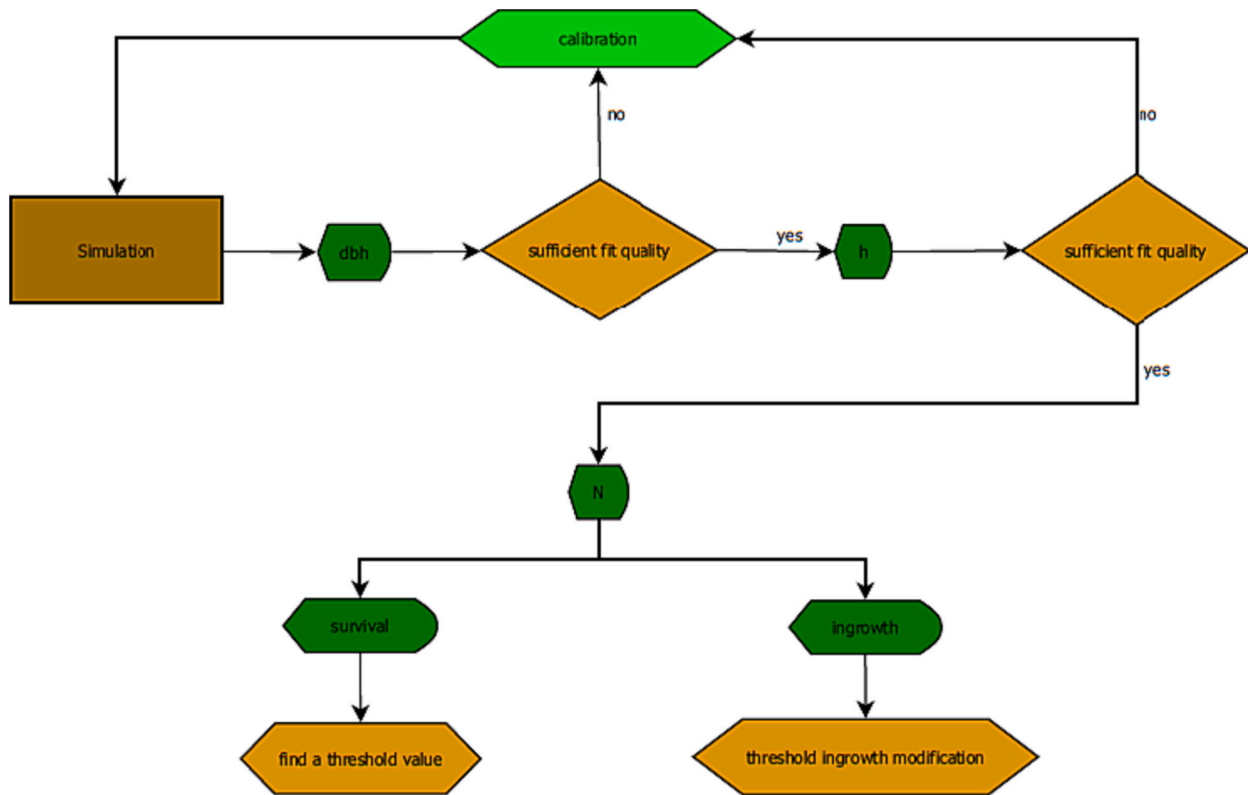


Fig. 1. Flowchart used to validate IBERO<sup>PT</sup> and IBERO<sup>PS</sup> models.

of them.

With these results, NFI3<sub>predicted</sub> and NFI3<sub>observed</sub> were ready for direct comparison. Although SIMANFOR also includes harvest simulation possibilities, this was not needed in our case studies.

#### 2.4. Statistical analysis

The statistical analysis is based on direct quantitative comparison of the observed data (NFI3<sub>observed</sub>) with the predicted data (NFI3<sub>predicted</sub>). It follows a simple but efficient linear regression that generates a simple performance index indicating the prediction accuracy (Vanclay and Skovsgaard, 1997). We chose the concordance correlation coefficient index for continuous variables, also known as the Lin coefficient (Lawrence and Lin, 1989; Lin, 2000), to develop our analysis. With a random bivariate sample (x<sub>1</sub>, y<sub>1</sub>), (x<sub>2</sub>, y<sub>2</sub>), ..., (x<sub>n</sub>, y<sub>n</sub>) from a random vector (X, Y) in R<sup>2</sup>, the concordance correlation coefficient index r<sub>c</sub> is defined as:

$$r_c = \frac{2 \cdot s_{xy}}{s_x^2 + s_y^2 + (\bar{x} - \bar{y})^2}$$

where:

- s<sub>xy</sub> is the sample covariance.
- s<sub>x</sub><sup>2</sup> and s<sub>y</sub><sup>2</sup> are the sample variances.
- $\bar{x}$  and  $\bar{y}$  are the means of the two vector random samples.

If r is shown to be Pearson's sample correlation coefficient, then:

$$r_c = r \cdot C_b$$

Where:

$$C_b = \left[ \left( v + \frac{1}{v} + u^2 \right) / 2 \right]^{-1}$$

$$v = s_x / s_y$$

$$u = (\bar{x} - \bar{y}) / \sqrt{s_x \cdot s_y}$$

In addition, it has been verified that 0 < C<sub>b</sub> ≤ 1, when C<sub>b</sub> is the accuracy factor, and thus |r<sub>c</sub>| ≤ |r|. In other words, the concordance correlation coefficient index is always lower than or equal (module) to Pearson's correlation coefficient. Thus, C<sub>b</sub> is a value that measures how far is the best-fit line is from the 45° line (where observations = predictions), measuring accuracy; r is a value that represents how far is each observation from the best-fit line, measuring precision; and r<sub>c</sub> contains both information, accuracy and precision, in one single metric (Lawrence and Lin, 1989).

The Lin coefficient value is related to a concordance degree between predictions and observations. Following the indications of Camacho-Sandoval (2008), a Lin coefficient value >0.99 in a qualitative classification indicates almost perfect concordance; from 0.95 to 0.99 indicates a substantial concordance level; from 0.90 to 0.95 indicates a moderate concordance level; and < 0.90 indicates a low concordance level between predictions and observations. Using this approach, we can qualitatively cluster the results of our predictions for easier visualization of the model improvement.

To improve the concordance level of the predictions, we decided to calibrate the submodels to the correct shape following the method proposed by Bravo and Montero (2003). It consists of adding a calibration factor to the original submodel to adjust the scale.

For this analysis, the diameter at breast height (dbh) and the total height (h) were studied to assess growth submodels using the tree datasets. Density (N), basal area (G) and dominant height (H<sub>0</sub>) were chosen as the most representative variables for assessing survival and ingrowth submodels using the plot datasets. SAS software was used to develop the statistical analysis (SAS-Institute, 2011).

### 3. Results

#### 3.1. Diameter and height growth submodels

The simulations developed with the original models showed some diameter and height growth projection biases for both the Maritime pine and Scots pine models. In both instances, the correlation between predicted (NFI3<sub>predicted</sub>) and observed (NFI3<sub>observed</sub>) data was high, indicating that the models had a correct shape, but the accuracy of predictions had a low improvement margin (Table 1) that could be increased. To do that, both models were calibrated jointly in two steps. The diameter growth was calibrated first because it is needed to calculate tree height, and then height growth was calibrated using the diameter growth predictions obtained from the first step. This calibration strategy follows the same procedure as previous works in this field (Bravo and Montero, 2003).

The calibrated equations are shown next, with the resulting calibrations highlighted in bold for Maritime pine (eq. 6) and Scots pine (eq. 7) diameter growth submodels, and for Maritime pine (eq. 8) and Scots pine (eq. 9) height growth submodels. The calibrations improved the estimations (Table 1) based on the Pearson correlation coefficient (*r*), Lin concordance correlation coefficient index (*r<sub>c</sub>*), and accuracy factor (*C<sub>b</sub>*), offering more accurate results.

$$dbh_{i5} = 1.18 + e^{0.2030 \cdot \ln(dbh-10) + 0.4414 \cdot \ln((cr+0.2)/1.2) + 0.8379 \cdot \ln(SI) - 0.1295 \cdot \sqrt{G - (0.0007 \cdot bal^2) / (\ln(dbh-10))}} / 10$$

$$dbh_{i5} = 1.70 + e^{-0.3711 + 0.2525 \cdot \ln(dbh-10) + 0.7090 \cdot \ln((cr+0.2)/1.2) + 0.9087 \cdot \ln(SI) - 0.1545 \cdot \sqrt{G - (0.0004 \cdot bal^2) / (\ln(dbh-10))}} / 10$$

$$h_{i5} = 0.42 + e^{0.2160 + 0.4033 \cdot \ln(dbh_{i5}/2) - 1.1272 \cdot \ln(dbh-10) + 1.1810 \cdot \ln(h-100) + 3.0162 \cdot cr} / 100$$

$$h_{i5} = 1.14 + e^{3.1222 - 0.4939 \cdot \ln(dbh_{i5}/2) + 1.3763 \cdot \ln(SI) - 0.0061 \cdot bal + 0.1876 \cdot \ln(cr)} / 100$$

where:

- *dbh* is the diameter at breast height (1.30 m) (cm)
- *dbh<sub>i5</sub>* is the 5-year increment in diameter (cm)
- *h* is the total height (m)
- *h<sub>i5</sub>* is the 5-year increment in height (m)
- *bal* is the accumulated basal area of trees thicker than the subject tree (m<sup>2</sup> ha<sup>-1</sup>)
- *cr* is the crown ratio (%)
- *G* is the stand basal area (m<sup>2</sup> ha<sup>-1</sup>)
- *SI* is the Site Index (m)

**Table 1**

Results obtained for the Pearson correlation coefficient (*r*), Lin concordance correlation coefficient index (*r<sub>c</sub>*), and accuracy factor (*C<sub>b</sub>*) of the different growth equation case studies. Values modified after calibrations are in bold.

	IBERO <sup>PT</sup>		IBERO <sup>PS</sup>		
	dbh	h	dbh	h	
original model	<i>r</i>	0.92962	0.86296	0.96995	0.85744
	<i>r<sub>c</sub></i>	<b>0.86221</b>	<b>0.82141</b>	<b>0.83102</b>	<b>0.65214</b>
	<i>C<sub>b</sub></i>	<b>0.92749</b>	<b>0.95185</b>	<b>0.85677</b>	<b>0.76057</b>
dbh calibrated	<i>r</i>	0.92973	0.86325	0.96990	0.85761
	<i>r<sub>c</sub></i>	<b>0.92727</b>	0.80137	<b>0.96317</b>	0.65380
	<i>C<sub>b</sub></i>	<b>0.99735</b>	0.92832	<b>0.99306</b>	0.76235
dbh and h calibrated	<i>r</i>	0.92973	0.86336	0.97051	0.86006
	<i>r<sub>c</sub></i>	0.92726	<b>0.85454</b>	0.96389	<b>0.85719</b>
	<i>C<sub>b</sub></i>	0.99734	<b>0.98978</b>	0.99318	<b>0.99666</b>

Calibrating the diameter and height growth submodels increases the accuracy of the whole model, since *dbh* and *h* are used as independent variables in other submodels (i.e., crown, biomass or volume). A graphic description of that improvement is shown in Fig. 2, where the original and predicted data show the same distribution but different accuracies due to the calibration process.

#### 3.2. Survival and ingrowth submodels

Once the growth submodels were calibrated, the survival and ingrowth submodels were evaluated for accuracy. Prediction accuracy for both submodels was assessed using *N*, *G* and *H<sub>0</sub>* stand variables. The initial simulation after growth submodel improvement (calibrated model) indicated that high correlation and accuracy were high for the *H<sub>0</sub>* prediction (*r* = 0.79681 and 0.89247, *C<sub>b</sub>* = 0.98760 and 0.99264 for IBERO<sup>PT</sup> and IBERO<sup>OS</sup>, respectively), but lower for *N* and *G* (Table 1). Since *H<sub>0</sub>* only represents a portion of the trees in the stand, that variable was not considered in the subsequent analysis.

Evaluating the survival submodel involved establishing a threshold value based on stand density capacity, to decide when the submodel should be applied. The results obtained showed no improvements in the correlation coefficient or prediction accuracy between observed and predicted values (Table 2). The predictions also had a different density

distribution than observed data, as can be seen in Fig. 3. The differences were related to the combined effects of the survival and ingrowth submodels, which have independent effects on stand density but for the same projection period.

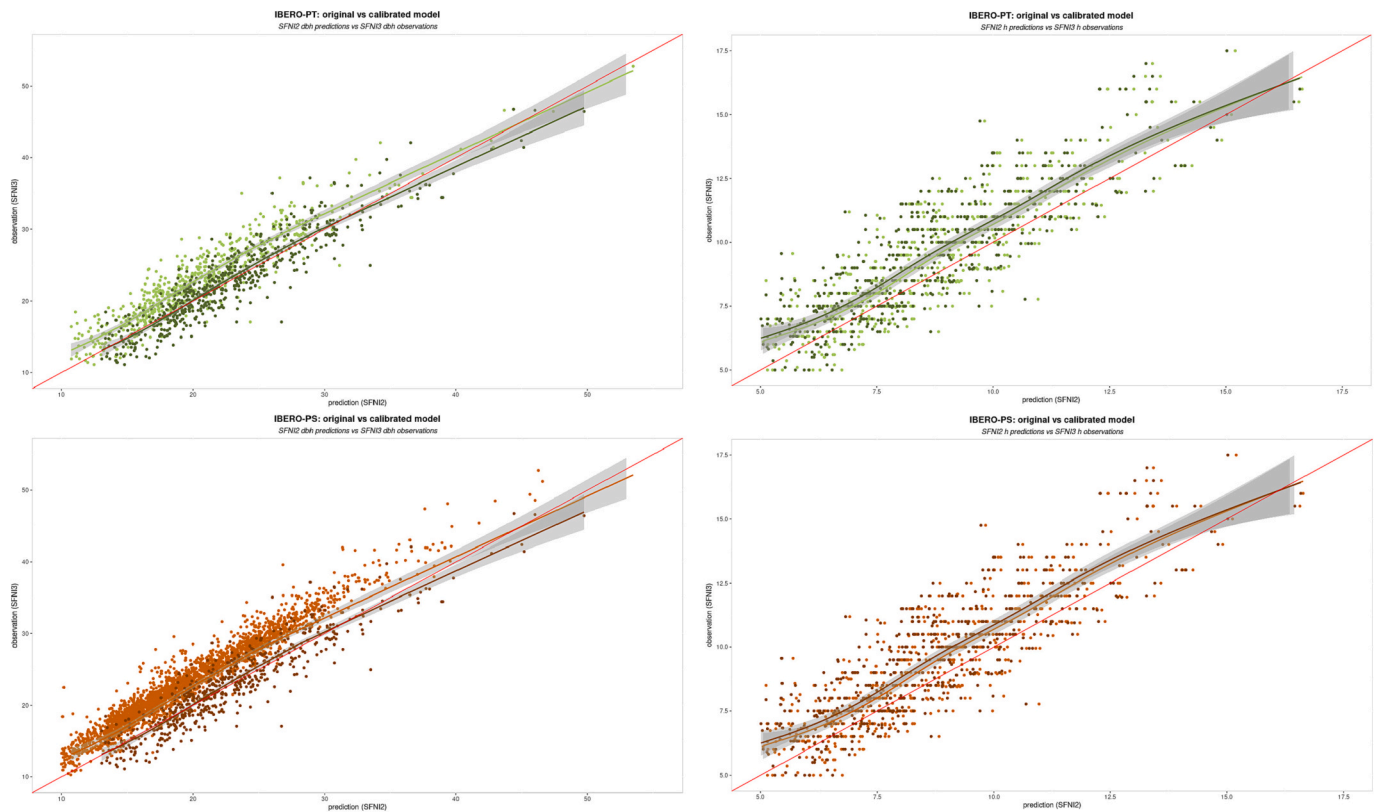
Ingrowth submodels were evaluated based on two types of case studies. In one, removing the technical ingrowth slightly increased the accuracy of the *N* and *G* variable predictions of both models (Table 2) and the density distribution was closer to the observed data than prior case studies (Fig. 3). In the other, the case studies focused on changing the threshold value to 0.25, 0.5 and 0.75 did not alter prediction correlation and accuracy. This implies that modification of the threshold value has a weaker effect on predictions than expected. As in the survival case study, the predictions also showed a different density distribution compared to observed data (Fig. 3).

### 4. Discussion

The methodology is based on direct quantitative comparisons of model predictions and observed values in the analysis of submodels for diameter, height growth, survival, and ingrowth. By comparing quantitative predictions and observed data, we were able to adequately calibrate diameter and height growth submodels. Nonetheless, our approach was not effective for improving the accuracy of survival and ingrowth submodels.

The approach used here is similar to that of other model evaluation studies (Moore, 2010; Roxburgh et al., 2019) and effectively improved growth submodel predictions. Bias errors were detected in the diameter and height growth submodels, which tended to underestimate tree growth. Schmid et al. (2006) and Pretzsch et al. (2002) reported similar findings using the SILVA simulator. When detected, two-step calibration was carried out to correct the biases of the diameter growth submodel first, and subsequently the biases of the height growth submodel. The





**Fig. 2.** Predicted (x-axis) and observed (y-axis) results developed with R for dbh and h of IBERO<sup>PT</sup> (green) and IBERO<sup>PS</sup> (orange), where both original and calibrated data are shown. The point cloud represents the value of each tree before (light green/light orange) and after (dark green/dark orange) calibration. The green/orange lines indicate the mean value and the red line is a reference of the zero error. Graph design is based on the recommendations of Piñeiro et al. (2008). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 2**

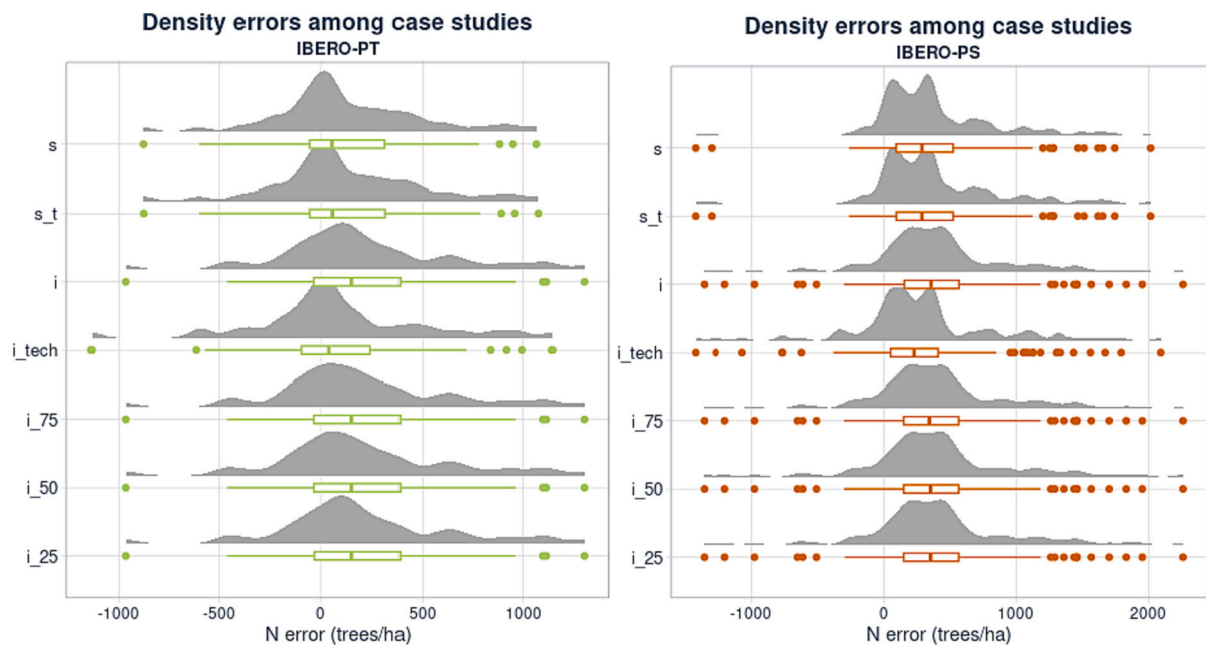
Results obtained for the Pearson correlation coefficient ( $r$ ), Lin concordance correlation coefficient index ( $r_c$ ), and accuracy factor ( $C_b$ ) of each case study of survival and ingrowth submodels: cal (model with growth submodels calibrated, used as reference); s\_t (survival submodel with threshold); i\_tech (ingrowth submodel without technical ingrowth); i\_25, i\_50, i\_75 (ingrowth submodel with threshold as 0.25, 0.50 and 0.75, respectively).

		IBERO <sup>PT</sup>		IBERO <sup>PS</sup>				IBERO <sup>PT</sup>		IBERO <sup>PS</sup>	
		N	G	N	G			N	G	N	G
cal	$r$	0.63443	0.71958	0.73423	0.79962	i_25	$r$	0.59762	0.74044	0.63154	0.78366
	$r_c$	0.59384	0.59026	0.56554	0.60565		$r_c$	0.51420	0.49817	0.48155	0.52338
	$C_b$	0.93602	0.82028	0.77025	0.75742		$C_b$	0.86041	0.67280	0.76250	0.66787
s_t	$r$	0.63581	0.72010	0.73348	0.80058	i_50	$r$	0.60060	0.74348	0.63148	0.78360
	$r_c$	0.59561	0.59149	0.57580	0.61323		$r_c$	0.51817	0.50412	0.48151	0.52345
	$C_b$	0.88959	0.82140	0.78502	0.76598		$C_b$	0.86275	0.67805	0.76251	0.66801
i_tech	$r$	0.57210	0.71917	0.63340	0.78816	i_75	$r$	0.60013	0.74483	0.63180	0.78077
	$r_c$	0.54384	0.58506	0.53583	0.61580		$r_c$	0.51825	0.50601	0.48260	0.52379
	$C_b$	0.95060	0.81352	0.84596	0.78131		$C_b$	0.86356	0.67936	0.76385	0.67086

resulting calibration of both submodels led to a substantial improvement in the Lin's correlation coefficient value and prediction accuracy. Nevertheless, individual errors were still present in predictions, even when the mean growth of the whole population improved. Qualitatively, the concordance between predictions and observations increased from low to moderate and substantial levels for dbh, while height improvement was insufficient to reach the next level. In this case, the qualitative classification categories were originally developed for laboratory devices (Camacho-Sandoval, 2008), which require much higher levels of accuracy than can be obtained in forestry. So, this improvement is still interesting even though the level of agreement did not change. Some of the existing bias can also be linked to natural events. Temporal biases such as differences in weather conditions interacting annually on the study period can lead to overestimations or underestimations (Pretzsch and Dursky, 2001; Sterba and Monserud, 1997; Vospernik et al., 2010).

Spatial biases are also frequent in relation to site-specific variability not assessed by the model (Sterba and Monserud, 1997).

Estimation deviations related to tree size were present in both models. The original submodels showed a higher bias in the predictions for extreme diameter classes, which have a smaller database (<10 and > 40 cm), and greater accuracy in the remaining diameter classes (>10 and < 40 cm) (Lizarralde, 2008). In our dataset, most individual trees were distributed across the intermediate diameter classes, so we were unable to assess predictions for extreme distributions. Extreme diameter predictions might be improved by applying the same methodology to data from extra-mature and young stands, thus dividing the sample into strata. In this way, the main submodel could be parameterized for each of the diameter classes studied, as proposed by Vanclay and Skovsgaard (1997). Calibrating these growth submodels is crucial, because diameter and height variables are used as independent variables to calculate most



**Fig. 3.** Errors between predicted and observed N stand values for IBERO<sup>PT</sup> (green) and IBERO<sup>PS</sup> (orange) among case studies: s (original survival submodel); s<sub>t</sub> (survival submodel with threshold); I (original ingrowth submodel); i<sub>tech</sub> (ingrowth submodel without technical ingrowth); i<sub>25</sub>, i<sub>50</sub>, i<sub>75</sub> (ingrowth submodel with threshold as 0.25, 0.50 and 0.75, respectively). Each case study is shown as a density and boxplot graph. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of the basic variables, such as basal area, volume, or tree biomass. They are also the basis for estimating stand variables.

A first analysis of the predictions for the stand variables studied (N, G and H<sub>0</sub>) showed a poor fit for the N and G variables and greater accuracy for the H<sub>0</sub> variable, though a moderate fit was not achieved for any of the case studies. Since H<sub>0</sub> provides information about productivity and is more dependent on height predictions (calibrated in the previous section), we did not include it in subsequent analysis.

Stochastic mortality algorithms like the one studied here are more popular for assessing tree mortality than deterministic approaches, even though their calibration is more difficult to assess (Hawkes, 2000). In the original models, Bravo-Oviedo et al. (2006) reported good fits for the survival submodels studied here, though with some overestimation for the IBERO<sup>PT</sup> model and underestimation for the IBERO<sup>PS</sup> model. In our study, we assessed the implementation of a threshold factor based on stand density to determine when the submodel should be applied. The threshold value is based on the maximum stand carrying capacity under specific climate conditions (Rodríguez de Prado et al., 2020), thus making it possible to introduce climate limitations into the survival submodel. Our results showed negligible change in the accuracy of the predictions (Table 2), which can be linked to the model structure. Both IBERO<sup>PT</sup> and IBERO<sup>PS</sup> are distance-independent models, and inserting a threshold factor into the survival submodel follows a structure corresponding to distance-dependent models (Bravo-Oviedo et al., 2006). Fig. 3 shows a dissimilar prediction error distribution between the original and modified survival submodels, suggesting that inclusion of the threshold value had no effect on most of the plots. While Bravo-Oviedo et al. (2006) reported deviations in the original submodel predictions, growth-dependence and disturbance-induced tree mortality can explain a portion of the biases (Bugmann et al., 2019; Zhu et al., 2019).

Similarly, the ingrowth submodel case studies showed no improvements with respect to the original model. The original ingrowth submodels have a two-step structure of demonstrated utility (Bravo et al., 2008) involving a probabilistic and a deterministic equation, with a threshold value that decides when the second one will be applied. Ingrowth in nature is the result of accumulated events and conditions

from previous estimation years (Yang and Huang, 2015), which increase modelling difficulty. Bravo-Oviedo et al. (2006) reported low error in the original IBERO<sup>PT</sup> ingrowth submodel, and higher error in IBERO<sup>PS</sup>. We used case studies over the threshold value that followed the idea of Lexerod and Eid (2005) and Muhairwe (2003, in Yang and Huang (2015)), who used random numbers generated between zero and one to establish the threshold value instead of a pre-defined criterion. Our findings indicated low differences among the threshold values tested, as no important statistical differences were noticed among case studies. The higher values predicted by the probabilistic equation surpassed the various threshold values studied and provided similar results, as no modifications were introduced to the deterministic equation. Future work in this area should explore the possibility of refitting both submodels using other approaches, such as Zero Inflated Poisson models (Bravo et al., 2017; Calama et al., 2011).

In the case study where technical ingrowth was removed, the analysis results were quite different. The degree of correlation between predicted and observed values was lower than the reference model and similar to Lin's coefficient value, thus increasing prediction accuracy. Though increased accuracy seems like a better result than the original model, it was not, because correlation decreased. Qualitatively, all case studies assessed for the survival and ingrowth submodels showed a low level of correlation between observations and predictions.

One of the important differences in the study of survival and ingrowth submodels compared to growth models is the type of variables to which they are applied. Growth submodels are applied to tree-level continuous variables, allowing direct quantitative comparison of predicted and observed values for each tree. Meanwhile, survival and ingrowth submodels are applied to tree-level variables but assessed by binary and continuous stand variables in a two-step process. Thus, both the errors and the lack of accuracy in these processes are the result of the accumulated individual errors for each tree in the stand. In addition, survival and ingrowth submodels both modify the same variable (N) at each projection, and events that can affect reality – climate disturbances, plant regeneration, grazing – are not included in them. This fact helps explain the high difference in the values of stand variable statistics (N, G and H<sub>0</sub>) compared to those of tree variables (dbh and h).

However, the strength of this work is the development of a methodology to evaluate and validate models, thus including all the key submodels (growth, survival, and ingrowth), using statistics to assess both accuracy and precision of predictions in one single metric. For future work, this study has uncovered a series of points to improve:

- Data variability must be reduced for better control of each situation. For example, a ‘dummy’ variable could be included to determine whether the stand is natural or a plantation (Palahí et al., 2003).
- Stand age can be useful as a group data variable that facilitates adjustment of the submodels to different stand ages (Vanclay and Skovsgaard, 1997).
- Sample size must be greater (Vanclay and Skovsgaard, 1997). In future work, data must be included from the 4th National Forest Inventory (NFI4), which was not available at the time of this analysis. This will increase the amount of data and expand the simulation period from 10 to 20 years.
- Submodels can also be rebuilt to improve their accuracy. For example, a combined survival-ingrowth submodel could be constructed to manage stand density and reduce validation effort, as was done for a *Quercus robur* model (Anta, 2003).

## 5. Conclusion

In this work, a methodology for validating forest models (thus including growth, survival, and ingrowth) was designed and applied. Using consecutive measurements of the same dataset, a simulation was developed to compare predicted and observed values through a metric that assesses accuracy and precision in one single value. The IBERO model parameterizations for *Pinus pinaster* (IBERO<sup>PT</sup>) and *Pinus sylvestris* (IBERO<sup>PS</sup>) were evaluated. Diameter and height growth submodels were calibrated for various case studies and performance improved, but survival and ingrowth submodel performance did not improve due to their complexity. Nonetheless, the improvement in both growth submodels enhances general model accuracy due to the importance of predicting tree and stand variables. Further development in this area is needed to establish a robust methodology capable of evaluating and calibrating forest growth models.

## Author contributions

Vázquez-Veloso, A.: data curation, formal analysis, writing.

Pando, V.: conceptualization, formal analysis, methodology, supervision.

Ordóñez, C.: data curation, resources, supervision.

Bravo, F.: conceptualization, funding acquisition, methodology, supervision.

## Data availability

Data, code and additional information here: [https://github.com/aitorvv/IBERO\\_evaluation\\_and\\_validation/tree/1.1](https://github.com/aitorvv/IBERO_evaluation_and_validation/tree/1.1) - <https://zenodo.org/record/8199180>

IBERO evaluation and validation (Original data) (Zenodo-GitHub)

## Acknowledgements

This research work was developed through resources provided by the Supercomputing Center of Castilla and León (SCAYLE, [www.scayle.es](http://www.scayle.es)), financed by the European Regional Development Fund (FEDER). Funds were received from the European Union and Junta de Castilla y León Education Council (ORDEN EDU/842/2022). This study was also subsidized by the Junta de Castilla y León through the projects “CLU-2019-01 and CL-EI-2021-05 – iuFOR Institute Unit of Excellence” of the University of Valladolid, co-financed by the European Regional Development Fund (ERDF “Europe drives our growth”), and through the

projects “SOE4/PA/E1012 - COMFOR-SUDO: Integrated and intelligent management of complex forests and mixed-species plantations in Southwest Europe”, “VA183P20 – SMART: Bosques mixtos: selvicultura, mitigación, adaptación, resiliencia y trade-offs” and “PID2021-1262750B-C22 – Integrated Forest Management along complexity gradients (IMFLEX)”.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ecoinf.2023.102246>.

## References

- AEMET, 2021. AEMET: Agencia Estatal de Meteorología. [WWW Document]. URL. [www.aemet.es/](http://www.aemet.es/).
- Alberdi, I., Condes, S., Martínez, J., Martínez, S.S., Sánchez, G., Pérez, F., Villanueva, M. J.A., Vallejo, R., 2010. Spain, in: National Forest Inventories. Pathways for Common Reporting. European Science Foundation. Springer, p. 533. <https://doi.org/10.1007/978-90-481-3233-1>.
- Aldea, J., Bianchi, S., Nilsson, U., Hynynen, J., Lee, D., Holmström, E., Huuskonen, S., 2023. Evaluation of growth models for mixed forests used in Swedish and Finnish decision support systems. *For. Ecol. Manag.* 529, 120721 <https://doi.org/10.1016/j.foreco.2022.120721>.
- Anta, M.B., 2003. Crecimiento y producción de masas naturales de *Quercus robur* L. en Galicia. Doctoral dissertation., Universidade de Santiago de Compostela.
- Bravo, F., 2005. Dinámica de rodales de pino negral (*Pinus pinaster* Ait.) en el Sistema Ibérico Meridional: Estructura genética, regeneración y dinámica forestal. Informe final del proyecto AGL-2001-1780.
- Bravo, F., Montero, G., 2001. Site index estimation in scots pine (*Pinus sylvestris* L.) stands in the high Ebro Basin (northern Spain) using soil attributes. *Forestry* 74, 395–406. <https://doi.org/10.1093/forestry/74.4.395>.
- Bravo, F., Montero, G., 2003. High-grading effects on scots pine volume and basal area in pure stands in northern Spain. *Ann. For. Sci.* 60, 11–18. <https://doi.org/10.1051/forest:2002069>.
- Bravo, F., Pando, V., Ordóñez, C., Lizarralde, I., 2008. Modelling ingrowth in Mediterranean pine forests: a case study from scots pine (*Pinus sylvestris* L.) and mediterranean maritime pine (*Pinus pinaster* Ait.) stands in Spain. *Investig. Agrar. Sist. y Recur. For.* 17, 250–260. <https://doi.org/10.5424/srf/2008173-01039>.
- Bravo, F., Rodríguez, F., Ordóñez, C., Broto, M., Lizarralde, I., Ruano, I., Del Río, M., Calama, R., 2010. Simanfor: aplicación web para la simulación de alternativas selvícolas. *Divulg.* 100, 22–26.
- Bravo, F., Álvarez-González, J.G., Del Río, M., Barrio, M., Bonet, J.A., Bravo-Oviedo, A., Calama, R., Castedo-Dorado, F., Crecente, F., Condes, S., 2011. Growth and yield models in Spain: historical overview, contemporary examples and perspectives. *For. Syst.* 20, 315–328. <https://doi.org/10.5424/fs/2011202-11512>.
- Bravo, F., Rodríguez, F., Ordóñez, C., 2012. A web-based application to simulate alternatives for sustainable forest management: SIMANFOR. *For. Syst.* 21, 4–8. <https://doi.org/10.5424/fs/2112211-01953>.
- Bravo, F., Maguire, D.A., González-Martínez, S.C., 2017. Factors affecting cone production in *Pinus pinaster* Ait.: lack of growth-reproduction trade-offs but significant effects of climate and tree and stand characteristics. *For. Syst.* 26, e075. <https://doi.org/10.5424/fs/2017262-11200>.
- Bravo, F., Ordóñez, C., Vázquez-Veloso, A., Michalakopoulos, S., 2023. Forest Stand Simulations in the Cloud: Use of the IBERO Growth Model in the SIMANFOR Decision Support System. Manuscr. Prep.
- Bravo-Oviedo, A., Del Río, M., Montero, G., 2004. Site index curves and growth model for Mediterranean maritime pine (*Pinus pinaster* Ait.) in Spain. *For. Ecol. Manag.* 201, 187–197. <https://doi.org/10.1016/j.foreco.2004.06.031>.
- Bravo-Oviedo, A., Sterba, H., Del Río, M., Bravo, F., 2006. Competition-induced mortality for Mediterranean *Pinus pinaster* Ait. and *P. sylvestris* L. *For. Ecol. Manag.* 222, 88–98. <https://doi.org/10.1016/j.foreco.2005.10.016>.
- Bugmann, H., Seidl, R., Hartig, F., Bohn, F., Brūna, J., Cailleret, M., François, L., Heinke, J., Henrot, A.-J., Hickler, T., Hülsmann, L., Huth, A., Jacquemin, I., Kollas, C., Lasch-Born, P., Lexer, M.J., Merganić, J., Merganićová, K., Mette, T., Miranda, B.R., Nadal-Sala, D., Rammer, W., Rammig, A., Reineking, B., Roedig, E., Sabatés, S., Steinkamp, J., Suckow, F., Vacchiano, G., Wild, J., Xu, C., Reyer, C.P.O., 2019. Tree mortality submodels drive simulated long-term forest dynamics: assessing 15 models from the stand to global scale. *Ecosphere* 10, e02616. <https://doi.org/10.1002/ecs2.2616>.
- Calama, R., Mutke, S., Tomé, J., Gordo, J., Montero, G., Tomé, M., 2011. Modelling spatial and temporal variability in a zero-inflated variable: the case of stone pine (*Pinus pinea* L.) cone production. *Ecol. Model.* 222, 606–618. <https://doi.org/10.1016/j.ecolmodel.2010.09.020>.
- Camacho-Sandoval, J., 2008. Coeficiente de concordancia para variables continuas. *Acta Med. Costarric.* 50, 211–212.
- Del Río, M., López, E., Montero, G., 2006. Manual de gestión para masas procedentes de repoblación de *Pinus pinaster* Ait., *Pinus sylvestris* L. y *Pinus nigra* Arn. en Castilla y León. Junta de Castilla y León, Castilla y León (España).
- Dufour-Kowalski, S., Courbaud, B., Dreyfus, P., Meredieu, C., De Coligny, F., 2012. Capsis: an open software framework and community for forest growth modelling. *Ann. For. Sci.* 69, 221–233.

- Hann, D., Marshall, D.D., Hanus, M.L., 2006. Reanalysis of the SMC-ORGANON Equations for Diameter-Growth Rate, Height-Growth Rate, and Mortality Rate of Douglas-Fir.
- Hawkes, C., 2000. Woody plant mortality algorithms: description, problems and progress. *Ecol. Model.* 126, 225–248. [https://doi.org/10.1016/S0304-3800\(00\)00267-2](https://doi.org/10.1016/S0304-3800(00)00267-2).
- Lawrence, I., Lin, K., 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 255–268. <https://doi.org/10.2307/2532051>.
- Lexerød, N., Eid, T., 2005. Recruitment models for Norway spruce, scots pine, birch and other broadleaves in young growth forests in Norway. *Silva Fenn.* 39, 391–406.
- Lin, K., 2000. A note on the concordance correlation coefficient. *Biometrics* 56, 324–325. <https://doi.org/10.1177/1536867X0200200206>.
- Ling, Y., Mahadevan, S., 2013. Quantitative model validation techniques: New insights. *Reliab. Eng. Syst. Saf.* 111, 217–231. <https://doi.org/10.1016/J.RESS.2012.11.011>.
- Lizarralde, I., 2008. Dinámica de rodales y competencia en las masas de pino silvestre (*Pinus sylvestris* L.) y Pino Negral (*Pinus pinaster* Ait.). In: de los Sistemas Central e Ibérico Meridional. (Tesis doctoral) Departamento de Producción Vegetal y Recursos Forestales. Universidad de Valladolid.
- Lizarralde, I., Ordóñez, C., Bravo-Oviedo, A., Bravo, F., 2010a. IBERO-PT: Modelo de dinámica de rodales de *Pinus pinaster* Ait. en el sistema ibérico meridional.
- Lizarralde, I., Ordóñez, C., Bravo-Oviedo, A., Bravo, F., 2010b. IBERO-PS: Modelo de dinámica de rodales de *Pinus sylvestris* L. en el sistema ibérico meridional.
- Moore, J.R., 2010. Allometric equations to predict the total above-ground biomass of radiata pine trees. *Ann. For. Sci.* 67, 806. <https://doi.org/10.1051/forest/2010042>.
- Muhairwe, C.K., 2003. Biometric models. Nandewar bio-region. NSW West. Reg. Assessments. Proj. no. NAND 08 (vol 5). Resour. Conserv. Assess. Council.
- Palahí, M., Pukkala, T., Miina, J., Montero, G., 2003. Individual-tree growth and mortality models for scots pine (*Pinus sylvestris* L.) in north-East Spain. *Ann. For. Sci.* 60, 1–10.
- Pardos, M., del Río, M., Pretzsch, H., Jactel, H., Bielak, K., Bravo, F., Brazaitis, G., Defossez, E., Engel, M., Godvod, K., Jacobs, K., Jansone, L., Jansons, A., Morin, X., Nothdurft, A., Oreti, L., Ponette, Q., Pach, M., Riofrío, J., Ruíz-Peinado, R., Tomao, A., Uhl, E., Calama, R., 2021. The greater resilience of mixed forests to drought mainly depends on their composition: analysis along a climate gradient across Europe. *For. Ecol. Manag.* 481, 118687 <https://doi.org/10.1016/J.FORECO.2020.118687>.
- Piñeiro, G., Perelman, S., Guerschman, J.P., Paruelo, J.M., 2008. How to evaluate models: observed vs. predicted or predicted vs. observed? *Ecol. Model.* 216, 316–322. <https://doi.org/10.1016/j.ecolmodel.2008.05.006>.
- Pretzsch, H., 2009a. Forest Dynamics, Growth, and Yield., in: Forest Dynamics, Growth and Yield. Springer, Berlin, Heidelberg, pp. 1–39. [https://doi.org/10.1007/978-3-540-88307-4\\_1](https://doi.org/10.1007/978-3-540-88307-4_1).
- Pretzsch, H., 2009b. Application of Forest simulation models for decision support in practice. In: Dynamics, Forest (Ed.), Growth and Yield. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp. 515–552. [https://doi.org/10.1007/978-3-540-88307-4\\_1](https://doi.org/10.1007/978-3-540-88307-4_1).
- Pretzsch, H., Dursky, J., 2001. Evaluierung von Waldwachstumssimulatoren auf Baum- und Bestandesebene. *Allg. Forst. Jagdzeitung* 172, 146–150.
- Pretzsch, H., 2009c. Forest Growth Models. In: Forest Dynamics, Growth and Yield. Springer, Berlin, Heidelberg, pp. 423–492. [https://doi.org/10.1007/978-3-540-88307-4\\_1](https://doi.org/10.1007/978-3-540-88307-4_1).
- Pretzsch, H., Biber, P., Dursky, J., 2002. The single tree-based stand simulator SILVA: construction, application and evaluation. *For. Ecol. Manag.* 162, 3–21. [https://doi.org/10.1016/S0378-1127\(02\)00047-6](https://doi.org/10.1016/S0378-1127(02)00047-6).
- R Core Team, 2021. R: A Language and Environment for Statistical Computing.
- Reineke, L.H., 1933. Perfection a stand-density index for even-aged forest. *J. Agric. Res.* 46, 627–638.
- Rodríguez de Prado, D., San Martín, R., Bravo, F., Herrero, C., 2020. Potential climatic influence on maximum stand carrying capacity for 15 Mediterranean coniferous and broadleaf species. *For. Ecol. Manag.* 460, 117824 <https://doi.org/10.1016/j.foreco.2019.117824>.
- Roxburgh, S.H., Karunarathne, S.B., Paul, K.I., Lucas, R.M., Armston, J.D., Sun, J., 2019. A revised above-ground maximum biomass layer for the Australian continent. *For. Ecol. Manag.* 432, 264–275. <https://doi.org/10.1016/j.foreco.2018.09.011>.
- Ruiz-Peinado, R., del Río, M., 2011. New models for estimating the carbon sink capacity of Spanish softwood species. *For. Syst.* 20, 176–188.
- SAS-Institute, 2011. SAS/STAT 9.3 User's Guide: Survey Data Analysis (Book Excerpt). SCAYLE, 2019. Supercomputación Castilla y León [WWW Document]. Accedido: 09-05-2021.
- Schmid, S., Zingg, A., Biber, P., Bugmann, H., 2006. Evaluation of the forest growth model SILVA along an elevational gradient in Switzerland. *Eur. J. For. Res.* 125, 43–55. <https://doi.org/10.1007/s10342-005-0076-4>.
- SIMANFOR, 2022a. IBERO-PT, an individual tree growth model independent from distance for maritime pine (*Pinus pinaster*). In: Sistema Ibérico Meridional (Spain)..
- SIMANFOR, 2022b. IBERO-PS, an individual tree growth model independent from distance for scots pine (*Pinus sylvestris*). In: Sistema Ibérico and Sistema Central (Spain).
- SIMANFOR, 2023. Sistema de Apoyo para la Simulación de Alternativas de Manejo Forestal Sostenible [WWW Document]. URL. <https://www.simanfor.es/>.
- Sterba, H., Monserud, R.A., 1997. Applicability of the forest stand growth simulator prognas for the Austrian part of the bohemian massif. *Ecol. Model.* 98, 23–34. [https://doi.org/10.1016/S0304-3800\(96\)01934-5](https://doi.org/10.1016/S0304-3800(96)01934-5).
- Tomppo, E., Gschwantner, T., Lawrence, M., McRoberts, R.E., 2010. National forest inventories. In: Pathways for Common Reporting, 1st ed. Springer, Dordrecht. <https://doi.org/10.1007/978-90-481-3233-1>.
- Vanclay, J.K., Skovsgaard, J.P., 1997. Evaluating forest growth models. *Ecol. Model.* 98, 1–12. [https://doi.org/10.1016/S0304-3800\(96\)01932-1](https://doi.org/10.1016/S0304-3800(96)01932-1).
- Vospernik, S., Monserud, R.A., Sterba, H., 2010. Do individual-tree growth models correctly represent height:diameter ratios of Norway spruce and scots pine? *For. Ecol. Manag.* 260, 1735–1753. <https://doi.org/10.1016/J.FORECO.2010.07.055>.
- Weiskittel, A.R., Hann, D.W., Kershaw Jr., J.A., Vanclay, J.K., 2011. Forest growth and yield modeling. John Wiley & Sons. <https://doi.org/10.1002/9781119998518>.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the Tidyverse. *J. Open Source Softw.* 4, 1686. <https://doi.org/10.21105/joss.01686>.
- Yang, Y., Huang, S., 2015. Two-stage ingrowth models for four major tree species in Alberta. *Eur. J. For. Res.* 134, 991–1004. <https://doi.org/10.1007/S10342-015-0904-0/FIGURES/4>.
- Zhu, Y., Liu, Z., Jin, G., 2019. Evaluating individual-based tree mortality modeling with temporal observation data collected from a large forest plot. *For. Ecol. Manag.* 450, 117496 <https://doi.org/10.1016/J.FORECO.2019.117496>.