RESEARCH PAPER

Comparison of two parameter recovery methods for the transformation of *Pinus sylvestris* yield tables into a diameter distribution model

Francisco Mauro[1] · Antonio García-Abril[2] · Esperanza Ayuga-Téllez[2] · Alberto Rojo-Alboreca[3] · Ruben Valbuena[4] · José Antonio Manzanera[2]

Abstract

• *Key message* We successfully transformed Pinus sylvestris yield tables into diameter distribution models. The best results were obtained with the parameter recovery method based on both mean and quadratic mean diameter, which explained 70% of the variability of frequencies by diameter classes and provided better results in the analysis of errors. On the other hand, the method based on stand density, dominant diameter and quadratic mean diameter explained less variability of frequencies by diameter classes (64.4%).

• *Context* Old datasets used to develop yield table models can be recovered to transform those yield tables into diameter distribution models that provide a more detailed description of size variability and forest structure.

• *Aims* We used archived measurements collected to develop yield table models for *Pinus sylvestris* L in central Spain, to transform those yield tables into a diameter distribution model by using parameter recovery methods.

• *Methods* We compared two different parameter recovery methods, one based on both mean and quadratic mean diameter and another one based on dominant diameter, stand density and quadratic mean diameter and used a set of 104 even aged plots to analyze the performance of the said methods for the transformation of *Pinus sylvestris* L yield tables in central Spain into a diameter distribution model.

• *Results* The parameter recovery method based on both mean and quadratic mean diameter explained 70% of the variability of frequencies by diameter classes and provided better results than the method based on stand density, dominant diameter and quadratic mean diameter that explained 64.4% of the variability of frequencies by diameter classes. However, more important than the method itself were the errors that propagated from the models predicting the different variables used in the parameter recovery.

• *Conclusion* Based on the results from the analysis of errors by diameter classes, the method using both mean and quadratic mean diameter outperformed the method using dominant diameter, stand density and quadratic mean diameter and is the best option to transform *P. sylvestris* yield tables into diameter distribution models.

Handling Editor: Céline Meredieu

Extended author information available on the last page of the article

INRAE   Springer

1 Introduction

Pinus sylvestris L is the pine species with the widest distribution range, occupying forested areas in Eurasia ranging from eastern Siberia, Mongolia and China to south-western Europe. As in other countries, in Spain, *P. sylvestris* is the most abundant species in terms of stocked volume, accounting for a 15% of the total volume in the National Forest Inventory (IFN 2015); and this proportion is larger in the mountains where *P. sylvestris* is native, typically forming forested areas just below the timberline. However, the natural range of *P. sylvestris* has been extended as this species has been traditionally used in the area for lumber production and now occupies areas at lower elevations.

Different forest management systems and silvicultural treatments are used with *P. sylvestris*, with management objectives ranging from pure conservation of biodiversity and natural assets to purely productive forest managed as even-aged forest with shorter rotations. The scope and detail of forest growth models can vary substantially. Following the nomenclature in Burkhart and Tomé (2012, p. 234), yield tables are whole-stand growth models that only provide the temporal evolution of aggregated stand attributes and represent the coarser level of detail for growth models. Individual-tree models are on the opposite side of the spectrum where the growth and mortality of individual trees are subject to modelling. An intermediate level of detail in the modelling scale is size distribution models. These models do not provide information at the tree level but augment the information recorded in whole-stand models with information disaggregated by size, with diameter at breast height (*dbh*) being the size variable typically considered. Smalley and Bailey (1974) growth model for *Pinus taeda* L. is an early example of this type of moderate level of modelling detail where a Weibull probability distribution describing *dbh* frequencies is provided for stands with a given age and site index. By providing *dbh* distributions, models of this type allow for a better characterization of forest structure and also for more accurate appraisals of the forest resources, especially of timber assets whose price is highly dependent on the actual log size (Breidenbach et al. 2008).

For many species and sites, whole-stand models are the only growth model available, and transforming these yield tables into a diameter distribution model may significantly improve the information available for some management purposes. Whole-stand models typically provide the temporal evolution of attributes that can be related to a size distribution. If a model considers the temporal evolution of a sufficient number of attributes related to the size distribution, they can be used to transform a whole-stand model into a size distribution model (Qin and Cao 2006). This transformation is oftentimes performed applying parameter recovery methods (PRM) (Strub and Burkhart 1975; Hyink and Moser 1983) with the estimated stand attributes for each considered age.

PRM methods bear a strong resemblance with the statistical method of moments, and for some PRM variants, they coincide. These methods are general and not limited to stand growth modelling. For example, they can be employed to characterize the current state of the forest using field and remotely sensed data (Gobakken and Næsset 2005; Maltamo et al. 2006; Bollandsås and Næsset 2007). PRM methods are based on establishing a system of equations between aggregated stand attributes and their expression based on a probability distribution model describing the tree-size frequencies (Knoebel et al. 1986; Weiskittel et al. 2011; Mehtätalo and Lappi 2019). PRM systems of equations require as many equations as parameters define the size distribution model (e.g., Del Río 1998). Therefore, prior to applying PRM methods for the transformation of a whole-stand model into a size distribution model, it might be necessary to model additional stand attributes related to the size distribution.

Rojo and Montero (1996) *P. sylvestris* yield tables have been in use for more than 20 years and are still one of the main references for the silviculture of *P. sylvestris* in central Spain. The dataset used to construct these yield tables has been digitized and archived and it contains measurements that allow transforming the yield tables into a *dbh* distribution model. Such transformation is of clear interest for forest managers as it would significantly enhance the set of management tools available in the area. The objective of this manuscript is to compare two alternative methods to transform *P. sylvestris* yield tables into a *dbh* distribution model. To achieve this objective, we first recovered the field measurements used to develop these yield tables. Then, we used the recovered dataset to develop models for additional forest attributes in order to make possible the application of two alternative PRM methods for the transformation of the yield tables into a *dbh* distribution model. Finally, we assessed the errors of the *dbh* distributions obtained by each PRM method when (1) directly applied to field measured attributes and (2) applied to predictions derived from the yield tables.

2 Material and methods

2.1 Yield tables for *Pinus sylvestris*

Yield tables for *P. sylvestris* in central Spain were constructed by Rojo and Montero (1996) using a set of 104 plots containing

field measurements of *dbh*, tree heights, stand density and age. These measurements were used to construct a system of four models or fundamental relationships describing the evolution over time of four aggregated stand attributes for ages above 20 years. These models can be grouped into silviculture-independent models and silviculture-dependent models (Fig. 1).

The first model describes the evolution over time of Assman's dominant height, H_{od}. This model considers two extreme site indexes corresponding to dominant heights of 17 m and 29 m, respectively, at age 100 years. Two Chapman-Richards models $H_{od}(t, I = 17)$ and $H_{od}(t, I = 29)$ provide, respectively, the temporal evolution of H_{od} for even-aged stands with site indexes 17, Eq. (1), and 29, Eq. (2).

$$H_{od}(t, I = 17) = 19.962(1 - e^{-0.02642t})^{2.1739} \quad (1)$$

$$H_{od}(t, I = 29) = 31.827(1 - e^{-0.03431t})^{2.8280} \quad (2)$$

The temporal evolution of H_{od} for stands with a generic site index I was obtained by linear interpolation between the 17 and 29 m curves, Eq. (3). This technique, referred by Gadow and Hui (1999) as "sectioning", resulted in polymorphic dominant height models.

$$H_{od}(t, I) = H_{od}(t, I = 17) + [H_{od}(t, I = 29) - H_{od}(t, I = 17)](\frac{I - 17}{12})$$
$$(3)$$

The second silviculture-independent model, Eq. (4), provided the mean tree height for a stand with site index I and age t, $\overline{H}(t, I)$, as a function of $H_{od}(t, I)$. Thus, $\overline{H}$ is only a function of the stand age and site index.

$$\overline{H}(t, I) = -0.8920 + 0.9603 H_{od}(t, I) \quad (4)$$

The third and fourth models in Rojo and Montero (1996) yield tables depend on the silvicultural treatments applied to a given stand. These yield tables consider three silvicultural or thinning regimes, S. The first one, $S = 1$, is the thinning regime used in the plots that were measured to develop the yield tables and it is referred as observed silviculture. Two more thinning regimes (i.e. $S = 2$, or moderate thinning, and $S = 3$, or intense thinning) are considered in the yield tables. Models for the intense and moderate thinning regimes were fit using information from thinning experiments conducted in the region by the Spanish National Institute of Agricultural Research (INIA). For stands with ages larger than 40 years, the third model, Eq. (5), provided stand density, N, as a function of age, site index and the thinning regime as:

$$N(t, I, S) = \begin{cases} \left[\dfrac{100}{0.5815 + 0.0106 \times H_{od}(t,I) t^{0.5424}}\right]^2 & if\ S = 1 \\[3ex] \left[\dfrac{100}{0.2949 + 0.0653 \times H_{od}(t,I) t^{0.2301}}\right]^2 & if\ S = 2 \\[3ex] \left[\dfrac{100}{-1.1035 + 0.3476 \times H_{od}(t,I) t^{-0.0235}}\right]^2 & if\ S = 3 \end{cases} \quad (5)$$

Fig. 1 Parameter recovery method (PRM) of tree size variables from yield tables of *P. sylvestris* in Guadarrama sites (Spain). *I*: site index. *t*: stand age. *N*: stand density. D_g: quadratic mean diameter. D_o: dominant diameter (mean diameter of the 100 thickest trees ha^{-1}). *S*: thinning regime. *PRM*: parameter recovery method
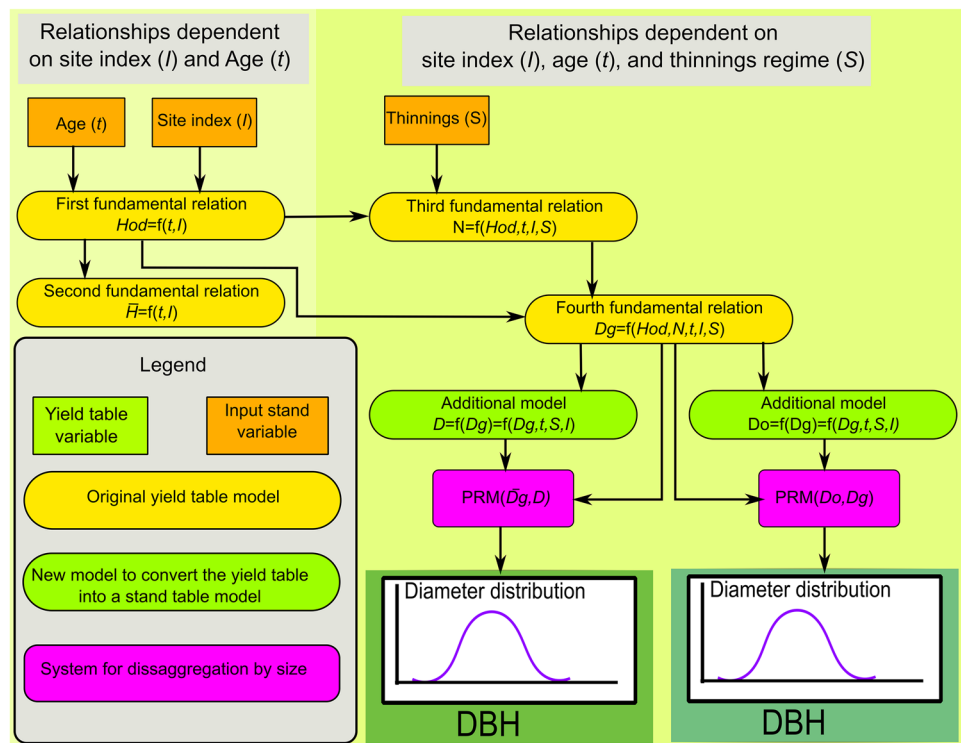
For ages below 40 years, $N(t, I, S = 2)$ and $N(t, I, S = 3)$ were computed using the curve for the observed silviculture, $N(t, I, S = 1)$, multiplied by a scaling ratio computed as the density for the considered thinning regime divided by the stand density for $S = 1$ both at age 40 years.

Finally, the fourth model provided the quadratic mean diameter D_g as a function of H_{0d} and N. Due to its dependence on N, D_g, Eq. (6), is ultimately a function of the thinning regime:

$$D_g(t, I, S) = -3.7795 + 6.9369 \frac{100}{\sqrt{N(t, I, S)}} + 0.5549 Ho(t, I, S). \quad (6)$$

2.2 PRM systems for transformation into *dbh* distribution models

We applied PRM methods assuming that the *dbh* distribution for an even-aged stand can be described using a Weibull probability distribution, Eq. (7). This distribution has been traditionally used to describe *dbh* (Bailey and Dell 1973; Cao 2004; Maltamo et al. 2004).

$$W(x | \alpha, \beta, \gamma) = \left(\frac{\alpha}{\beta} \right) \left(\frac{x - \gamma}{\beta} \right)^{\alpha - 1} e^{-\left(\frac{x - \gamma}{\beta} \right)^{\alpha}} \quad (7)$$

We used the three-parameter Weibull distribution functions where α and β are the shape and scale parameters, respectively, and the parameter γ is a location parameter that defines the minimum size. In this study, γ was fixed at 5 cm because that is the minimum *dbh* measured in the field plots and also a common threshold for tree measurements in the area. Models where γ was allowed to vary were initially considered but discarded as maximum likelihood (ML) estimates of Weibull *dbh* models with variable γ were significantly different, based on a likelihood ratio test, from the ML models with fixed γ in only 16.34% of the plots used to analyze *dbh* distributions (see Field data section). In addition, inspection of fitted models with fixed and variable γ only showed minor differences between frequencies in 10-cm *dbh* classes. Based on this preliminary result, we considered γ as fixed; thus, the number of stand attributes required by the PRM methods is at least two.

2.3 Recovery of *dbh* distributions using mean diameter

The studied *P. sylvestris* yield tables only provide the temporal evolution of one variable, D_g, directly related to the *dbh* distribution. Thus, to transform these yield tables into a *dbh* distribution model, we modeled the temporal evolution of extra variables that could be computed from the *dbh* distribution.

The first PRM system of equations, Eqs. (8) and (9), to transform the yield tables into a *dbh* distribution model was based on D_g and the mean diameter $\overline{D}$ as additional variable to obtain the parameters α and β of the *dbh* distribution.

$$D_g = \int_5^\infty x^2 \left(\frac{\alpha}{\beta} \right) \left(\frac{x - 5}{\beta} \right)^{\alpha_d - 1} e^{-\left(\frac{x - 5}{\beta} \right)^{\alpha_d}} dx \quad (8)$$

$$\overline{D} = \int_5^\infty x \left(\frac{\alpha}{\beta} \right) \left(\frac{x - 5}{\beta} \right)^{\alpha - 1} e^{-\left(\frac{x - 5}{\beta} \right)^{\alpha}} dx \quad (9)$$

This PRM system of equations, when applied to sampled values of D_g and $\overline{D}$, is equivalent to the commonly used method of moments as D_g and $\overline{D}$ are, respectively, second- and first-order moments with respect to zero of the *dbh* distribution. Then, we fitted a linear regression model where $\overline{D}$ was expressed as a function of D_g (Fig. 1) with the 104 plots used to construct the yield tables. Due to its dependence on D_g, $\overline{D}$ is a function of stand age, site index and applied silviculture, i.e. $\overline{D}(t, I, S)$. Finally, to solve for α and β, we applied the numerical methods to solve PRM systems of equations, described later.

2.4 Recovery of *dbh* distributions using dominant diameter

We tested a second PRM system for the transformation of the yield tables into a *dbh* distribution model. The additional *dbh* parameter used for the second PRM system was the dominant diameter D_o. Following Assmann's dominant height definition, D_o is the mean diameter of the largest 100 trees in a hectare. Using D_g and D_o, we established a system of PRM Eqs. (8) and (10) to obtain the parameters α and β.

$$D_o = \left(\frac{N}{100} \right) \int_{D_{lim}}^\infty x \left(\frac{\alpha}{\beta} \right) \left(\frac{x - 5}{\beta} \right)^{\alpha - 1} e^{-\left(\frac{x - 5}{\beta} \right)^{\alpha}} dx \quad (10)$$

where D_{lim} is the $\left[100 \left(1 - \frac{100}{N} \right) \right]^{th}$ percentile of the diameter distribution (Ayuga-Téllez et al. 2014).

As with $\overline{D}$, in order to incorporate D_o to the set of variables provided by the yield tables (Fig. 1), we fitted a linear model where D_o was expressed as a linear function of the stand variables D_g and N^{-1} included in the yield table models (Fig. 1). Due to its dependence on D_g and N, D_o is a function of stand age, site index and applied silviculture, i.e. $D_o(t, I, S)$. This model was fit using the *dbh* data from the 104 plots used to construct the yield tables. This system of PRM equations was solved using numerical methods described in next section.

2.5 Numerical methods to solve PRM systems of equations

Solutions for both PRM systems of equations were obtained using numerical methods. For each method, we defined a cost function of the form

$$cost(\alpha, \beta) = (L_1 - R_1(\alpha, \beta))^2 + (L_2 - R_2(\alpha, \beta))^2 \qquad (11)$$

where L_1 and L_2 were the left hand side values for the first and second equation of the corresponding PRM system of equations, and $R_1(\alpha, \beta)$ and $R_2(\alpha, \beta)$ their associated right hand sides. This solution has a minimum value of 0 when both PRM equations are verified (i.e. $L_1 = R_1(\alpha, \beta)$ and $L_2 = R_2(\alpha, \beta)$). For the PRM systems that use $\overline{D}$, β was directly deducted as a function of α. Then, substituting in the second equation reduced the problem to solving an equation of the form $L_2 - R_2(\alpha) = 0$.

As yield tables model the temporal evolution of even-aged stands, we restricted the search with $\alpha > 1$, to obtain unimodal distributions. We used the internal optimizer available in R (R Core Team 2018) through the optim and optimize functions and implemented a grid search method for the PRM method in section *Recovery of dbh distributions using dominant diameter*. R code implementing these methods can be obtained from https://doi.org/10.5281/zenodo.3934885.

2.6 Field data

We validated the PRM methods for *dbh* with a set of 104 field plots used in the construction of Rojo and Montero (1996) yield tables. These plots were measured in the valleys of Fuenfria and Valsain in Guadarrama Mountains. All trees with *dbh* 5 cm or larger were callipered and those trees identified as dominant were also measured for height to determine the dominant height of each plot (Table 1). In addition, the age of each plot was determined using tree cores and combined with H_o to determine age and site index of the plots. Tree measurements and field plot attributes (Rojo and Montero 2020) can be obtained from https://doi.org/10.5281/zenodo.3934885.

2.7 Validation of PRM methods

Two different assessments of the PRM methods previously described were considered. In the first assessment, we only evaluated the PRM systems and we did not consider additional errors derived from the chain of yield table models. Plot values of $D_{g,i}$ and N_i, where i indicates the ith plot, were directly computed from the field measurements and used to solve the PRM systems of equations (Eqs. 8, 9 and 8, 10). These PRM systems of equations will be, respectively,

Table 1 Summary of plot level values in the dataset used to analyze the PRM systems for *dbh*. Mean, minimum (Min), maximum (Max) and standard deviation (Sd) values of the plot level mean diameter D, quadratic mean diameter D_g, dominant diameter D_o and stand density N

Variable	Min	Mean	Max	Sd
$\overline{D}$(cm)	7.02	27.50	50.24	10.85
D_g(cm)	7.17	28.24	50.96	10.90
D_o(cm)	8.80	36.61	58.90	12.22
N (trees ha^{-1})	250.00	2058.98	11,154.00	2350.32
Trees per plot	24	71.61	193	33.15

indicated as $PRM(\overline{D}D_g)$ and $PRM(D_oD_g)$. Solutions of the PRM systems of equations provided two initial sets of *dbh* distribution parameters $\widehat{\alpha}_{i,\overline{D}}$ and $\widehat{\beta}_{i,\overline{D}}$ and $\widehat{\alpha}_{i,D_o}$ and $\widehat{\beta}_{i,D_o}$ based on field measurements of $D_{g,i}$, $\overline{D}_i$ and $D_{o,i}$, $D_{g,i}$ and N_i.

For the second assessment, the stand variables used by the PRM methods were deducted from different models using t, I and S as predictors, for a more realistic scenario. First, to account for the errors introduced by the models of the yield table, predictions of $D_{g,i}$ and N_i were computed for each field plot by entering in the yield table models in Eqs. (5) and (6) with the corresponding plot values t_i, I_i, S_i. To avoid over-optimistic predictions of $\overline{D}_i$ and $D_{o,i}$, these attributes were obtained using leave-one-out cross-validation with their corresponding models. For each plot, the models relating $\overline{D}$ with D_g and D_o with D_g and N^{-1} were refitted without plot i, then $\widehat{D}_{g,i}$ and $\widehat{N}_i$ were used as input to obtain predicted values $\overline{D}_i$ and $\widehat{D}_{o,i}$. The cross-validation was not applied to the yield table models for D_g and N because these models were validated when the yield tables were developed and because no significant lack of fit has been observed during more than 20 years of use of these models. Once predicted values of $\overline{D}_i$, $\widehat{D}_{o,i}$, $\widehat{D}_{g,i}$ and $\widehat{N}_i$ were obtained, they were used to solve the corresponding PRM systems of equations. These new solutions provided two additional sets of parameters denoted as $\widehat{\widehat{\alpha}}_{i,\overline{D}}$ and $\widehat{\widehat{\beta}}_{i,\overline{D}}$ and $\widehat{\widehat{\alpha}}_{i,D_o}$ and $\widehat{\widehat{\beta}}_{i,D_o}$, where the double hat is used to indicate that these parameters were obtained using predicted values $\overline{D}_i$, $\widehat{D}_{o,i}$, $\widehat{D}_{g,i}$ and $\widehat{N}_i$. These PRM systems of equations are, respectively, indicated as $PRM(\overline{D}\widehat{D}_g)$ and $PRM(\widehat{D}_o\widehat{D}_g)$.

Also, in order to have a baseline for comparisons with the PRM methods, we obtained, for each plot i, ML estimates, $\widehat{\alpha}_{i,ML}$ and $\widehat{\beta}_{i,ML}$, of the parameters of the Weibull distribution model describing the *dbh* distribution of each plot. For one plot, $\widehat{\alpha}_{i,ML}$ was smaller than one and the estimated distributions had a reverse J shape. As all plots were measured in even-aged stands, to obtain unimodal Weibull distributions, we re-fitted the model for this plot but only search for solution with $\widehat{\alpha}_{i,ML} > 1$ when maximizing the likelihood function.

INRAE ✷ Springer

Finally, for all plots and methods for estimating the parameters of the Weibull distribution, we computed the standard deviation of the corresponding *dbh* distribution as

$$\sigma = \sqrt{\beta^2 (\Gamma \left(1 + \frac{2}{\alpha}\right) - \Gamma \left(1 + \frac{1}{\alpha}\right)^2)}, \tag{12}$$

using the values of α and β obtained with each method. To differentiate between standard deviations obtained with different methods, we employ the notation that we used for α and β.

2.8 Similarity between PRM and ML *dbh* probability distributions

We evaluated the similarity between the parameters derived by each PRM method m and those obtained through ML by using the correlation coefficient between ML and PRM parameters as a measure of agreement:

$$\rho = \frac{\sum_{i=1}^{n} (y_{ML,i} - \bar{y}_{ML})(y_{m,i} - \bar{y}_m)}{\sqrt{\sum_{i=1}^{n} (y_{ML,i} - \bar{y}_{ML})^2} \sqrt{\sum_{i=1}^{n} (y_{m,i} - \bar{y}_m)^2}} \tag{13}$$

In Eq. (13), $y_{ML,i}$ represents either $\hat{\alpha}_{i,ML}$, $\hat{\beta}_{i,ML}$ or $\hat{\sigma}_{i,ML}$ and $y_{m,i}$ represents either, $\hat{\alpha}_{i,\overline{D}}$, $\hat{\alpha}_{i,D_o}$, $\hat{\hat{\alpha}}_{i,\overline{D}}$ or $\hat{\hat{\alpha}}_{i,D_o}$; $\hat{\beta}_{i,\overline{D}}$, $\hat{\beta}_{i,Do}$, $\hat{\hat{\beta}}_{i,\overline{D}}$ or $\hat{\hat{\beta}}_{i,D_o}$ or $\hat{\sigma}_{i,\overline{D}}$, $\hat{\sigma}_{i,D_o}$, $\hat{\hat{\sigma}}_{i,\overline{D}}$ or $\hat{\hat{\sigma}}_{i,Do}$. Finally, $\bar{y}_{ML}$ represents the average value of either $\hat{\alpha}_{i,ML}$, $\hat{\beta}_{i,ML}$ or $\hat{\sigma}_{i,ML}$ and $\bar{y}_m$ is the average of the estimated values of α, β, or σ obtained by each PRM method. We tested whether correlation coefficients were significantly different from zero and whether the ML estimates of α, β and σ were on average different from those obtained applying the four PRM methods under analysis.

2.9 Class-wise and global errors

To evaluate the performance of the PRM methods, rather than their similarity with ML, we computed two different groups of error metrics. These error metrics were also computed for ML as baseline values of reference. For each method, including ML, we computed class-wise error metrics for the proportion of trees by *dbh* classes. The number of *dbh* classes, n_c, was 8. The first class was defined by the interval (5–20 cm] and the last class contained trees with *dbh* larger than 80 cm. The remaining classes were 10 cm wide intervals open to the left and closed to the right. For each *dbh* class, j, and method, m, we computed

$$Bias_{j,m} = \frac{\sum_{i=1}^{n} (\hat{p}_{i,j,m} - p_{i,j})}{n} \tag{14}$$

$$RMSE_{j,m} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{p}_{i,j,m} - p_{i,j})^2}{n}} \tag{15}$$

$$MAE_{j,m} = \frac{\sum_{i=1}^{n} \left|\hat{p}_{i,j,m} - p_{i,j}\right|}{n} \tag{16}$$

$$R_{j,m}^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{p}_{i,j,m} - p_{i,j})^2}{\sum_{i=1}^{n} (\bar{p}_j - p_{i,j})^2} \tag{17}$$

In Eqs. (14), (15), (16) and (17), $p_{i,j}$ represents the proportion of trees in the jth*dbh* class in the ith plot, $\hat{p}_{i,j,m}$ represents the prediction of $p_{i,j}$ by method m, $\bar{p}_j$ is the average across plots of $p_{i,j}$, and n equals 104. In addition to $Bias_{j,m}$, $RMSE_{j,m}$ and $MAE_{j,m}$, we computed their relative counterparts denoted as $rBias_{j,m} = Bias_{j,m}/\bar{p}_j$, $rRMSE_{j,m} = RMSE_{j,m}/\bar{p}_j$ and $rMAE_{j,m} = MAE_{j,m}/\bar{p}_j$. Finally, we focused on the errors of $PRM(\hat{D}\hat{D}_g)$ and $PRM(\hat{D}_o\hat{D}_g)$. For each *dbh* class, we performed a paired sampled t test on the absolute values of the errors $\left|\hat{p}_{i,j,m} - p_{i,j}\right|$ obtained by each of these two methods.

The last set of error metrics consisted of the global versions of Eqs. (14), (15), (16) and (17) where all *dbh* classes were considered together. These error metrics are

$$gBias_m = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n_c} (\hat{p}_{i,j,m} - p_{i,j})}{n_c n} \tag{18}$$

$$gRMSE_m = \sqrt{\frac{\sum_{i=1}^{n} \sum_{j=1}^{n_c} (\hat{p}_{i,j,m} - p_{i,j})^2}{n_c n}} \tag{19}$$

$$gMAE_m = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n_c} \left|\hat{p}_{i,j,m} - p_{i,j}\right|}{n_c n} \tag{20}$$

$$gR_m^2 = 1 - \frac{\sum_{i=1}^{n} \sum_{j=1}^{n_c} (\hat{p}_{i,j,m} - p_{i,j})^2}{\sum_{i=1}^{n} \sum_{j=1}^{n_c} (\bar{p}_j - p_{i,j})^2} \tag{21}$$

where the prefix g stands for global to indicate that all classes were pooled together. Note that $gMAE_m$ and $gRMSE_m$ are proportional to the average of the Reynolds error index used in Gobakken and Næsset (2005) and Bollandsås and Næsset (2007) and to the I index proposed by Strunk et al. (2017), respectively. Finally, gR_m^2 equals Strunk et al. (2017) H index. We did not calculate relative versions of $gBias_m$, $gRMSE_m$ and $gMAE_m$ as they will be proportional to their absolute counterparts by a constant that is the same for all methods.

Finally, we considered all *dbh* classes together and performed a paired sampled *t* test on the absolute values of the errors $\left|\hat{p}_{i,j,m} - p_{i,j}\right|$ obtained for $PRM(\overline{D}\widehat{D}_{g})$ and $PRM(\widehat{D}_{o}\widehat{D}_{g})$.

3 Results

3.1 Auxiliary regression models to complete the yield tables

The analyzed systems to transform the yield tables into *dbh* distribution models required auxiliary regression models to obtain $\overline{D}$ and D_{o}. The developed model with $\overline{D}$ as response and D_{g} as predictor explained 99.95% of the variability of $\overline{D}$. This model provides values of $\overline{D}$ that are smaller than D_{g} for all possible values of D_{g}. The linear model for D_{o} with D_{g} and N^{-1} as predictors explained 97.99% of the variability of D_{o}. Coefficients of determination obtained by cross validation were 99.95% for $\overline{D}$ and 97.87% for D_{o} (Table 2). Once these models were chained with the yield table model to predict D_{g} and N from t, I and S, the percentage of explained variance for $\overline{D}$ and D_{o} dropped to 83.52% and 85.13%, respectively.

3.2 PRM and ML methods

Correlations between $\hat{\alpha}_{ML}$ and $\hat{\alpha}_{\overline{D}}$, $\hat{\beta}_{ML}$ and $\hat{\beta}_{\overline{D}}$, $\hat{\sigma}_{ML}$ and $\hat{\sigma}_{\overline{D}}$ were 0.99 or larger, while correlations of $\hat{\alpha}_{ML}$, $\hat{\beta}_{ML}$ and $\hat{\sigma}_{ML}$ with $\hat{\alpha}_{D_{o}}$, $\hat{\beta}_{D_{o}}$ and $\hat{\sigma}_{D_{o}}$ were 0.81, 0.99 and 0.89, respectively (Fig. 2). All these correlations were significantly different from zero at a 0.05 significance level. Correlations between $\hat{\alpha}_{ML}$, $\hat{\beta}_{ML}$ and $\hat{\sigma}_{ML}$ and $\hat{\hat{\alpha}}_{\overline{D}}$, $\hat{\hat{\beta}}_{\overline{D}}$ and $\hat{\hat{\sigma}}_{\overline{D}}$ were 0.82, 0.92 and 0.80, respectively, and correlations between $\hat{\alpha}_{ML}$, $\hat{\beta}_{ML}$ and $\hat{\sigma}_{ML}$ and $\hat{\hat{\alpha}}_{Do}$, $\hat{\hat{\beta}}_{D_{o}}$ and $\hat{\hat{\sigma}}_{Do}$ were 0.81, 0.92 and 0.80, respectively (Fig. 2). All correlations between ML parameters and those derived using PRM methods based

on predictions of $\overline{D}$ and D_{g} and D_{o}, D_{g} and N were still significantly different from zero at a 0.05 significance level but their magnitude was considerably smaller. These results showed that the chain of models to predict the attributes used in the PRM systems introduced significant errors in the estimation of the *dbh* distribution. No significant differences were observed between ML and any of the PRM solutions for the parameter β. However, differences between ML and $PRM(D_{o}D_{g})$ and $PRM(\widehat{D}_{o}\widehat{D}_{g})$ were significantly different from zero at a 0.05 significance level for σ and the parameter α (Table 3). While ML values are also estimates, ML method is a robust reference which indicates that the PRM method based on D_{o}, D_{g} and N is less efficient than the PRM method based on $\overline{D}$ and D_{g} and introduces some bias in the estimation of α that results in an underestimation of the variability of *dbh* measured by σ.

Maximum likelihood estimates $\hat{\alpha}_{ML}$, $\hat{\beta}_{ML}$ and $\hat{\sigma}_{ML}$ for *dbh* distribution showed a strong correlation with age and similar patterns were observed for $\hat{\alpha}_{\overline{D}}$, $\hat{\beta}_{\overline{D}}$ and $\hat{\sigma}_{\overline{D}}$, for $\hat{\alpha}_{D_{o}}$, $\hat{\beta}_{D_{o}}$ and $\hat{\sigma}_{D_{o}}$, for $\hat{\hat{\alpha}}_{\overline{D}}$, $\hat{\hat{\beta}}_{\overline{D}}$ and $\hat{\hat{\sigma}}_{\overline{D}}$ and for $\hat{\hat{\alpha}}_{D_{o}}$, $\hat{\hat{\beta}}_{D_{o}}$ and $\hat{\hat{\sigma}}_{D_{o}}$ (Fig. 3). Applying the PRM methods to the predicted values of $\overline{D}$ and D_{g} and D_{o}, D_{g} and N resulted on estimated values of the shape and scale parameter with a significantly lower variability (Fig. 3). A similar effect was observed for the standard deviation of the estimated Weibull distributions. The ranges of the standard deviations computed with $\hat{\hat{\alpha}}_{\overline{D}}$ and $\hat{\hat{\beta}}_{\overline{D}}$ and $\hat{\hat{\alpha}}_{D_{0}}$ and $\hat{\hat{\beta}}_{D_{0}}$ were 33.06% and 22.74% smaller than the range of standard deviations computed with $\hat{\alpha}_{ML}$ and $\hat{\beta}_{ML}$ (Fig. 3). This may be explained by the fact that the models for $\overline{D}$, D_{g}, D_{o} and N provided the average response for these variables for a given site index, age and thinning regime, and these variables did not fully explain the variability in the *dbh* distributions. An interesting topic for future research is to investigate how to further enhance the diameter distribution models derived from this study, so they might incorporate the unexplained variability when

Table 2 Regression models to obtain the mean diameter, $\overline{D}$ (cm), and the dominant diameter D_{0} (cm). The term ε represents a generic error term in each regression model. Coef std err: coefficient standard error. Res std err: residual standard error. Res std err CV and R^{2} CV are the residual standard error and coefficient of determination obtained using leave-one-out cross-validation, respectively. D_{g} (cm) is the quadratic mean diameter and N the stand density (trees ha^{-1})

Model	$\overline{D} = a + bD_{g} + \varepsilon$							
Coefficient	Estimate	Coef std err	*t* value	*p* value	R^{2}	R^{2} CV	Res std err	Res std err CV
a	− 6.23E-1	6.35E-2	− 9.95	<1E-5	99.95%		99.95%	0.23 cm	0.24 cm
b	9.96E-1	2.10E-3	474.8	<1E-5				
Model	$D_{o} = a + bD_{g} + c\frac{1}{N} + \varepsilon$							
Coefficient	Estimate	Coef std err	*t* value	*p* value	R^{2}	R^{2} CV	Res std err	Res std err CV
a	3.18	6.21E-1	5.12	<1.E-5	97.99%	97.87%	1.75 cm	1.80 cm
b	1.30	3.81E-2	34.11	<1E-5				
c	− 3017	544	5.54	<1E-5				

Correlation between parameters estimated by ML and PRM methods
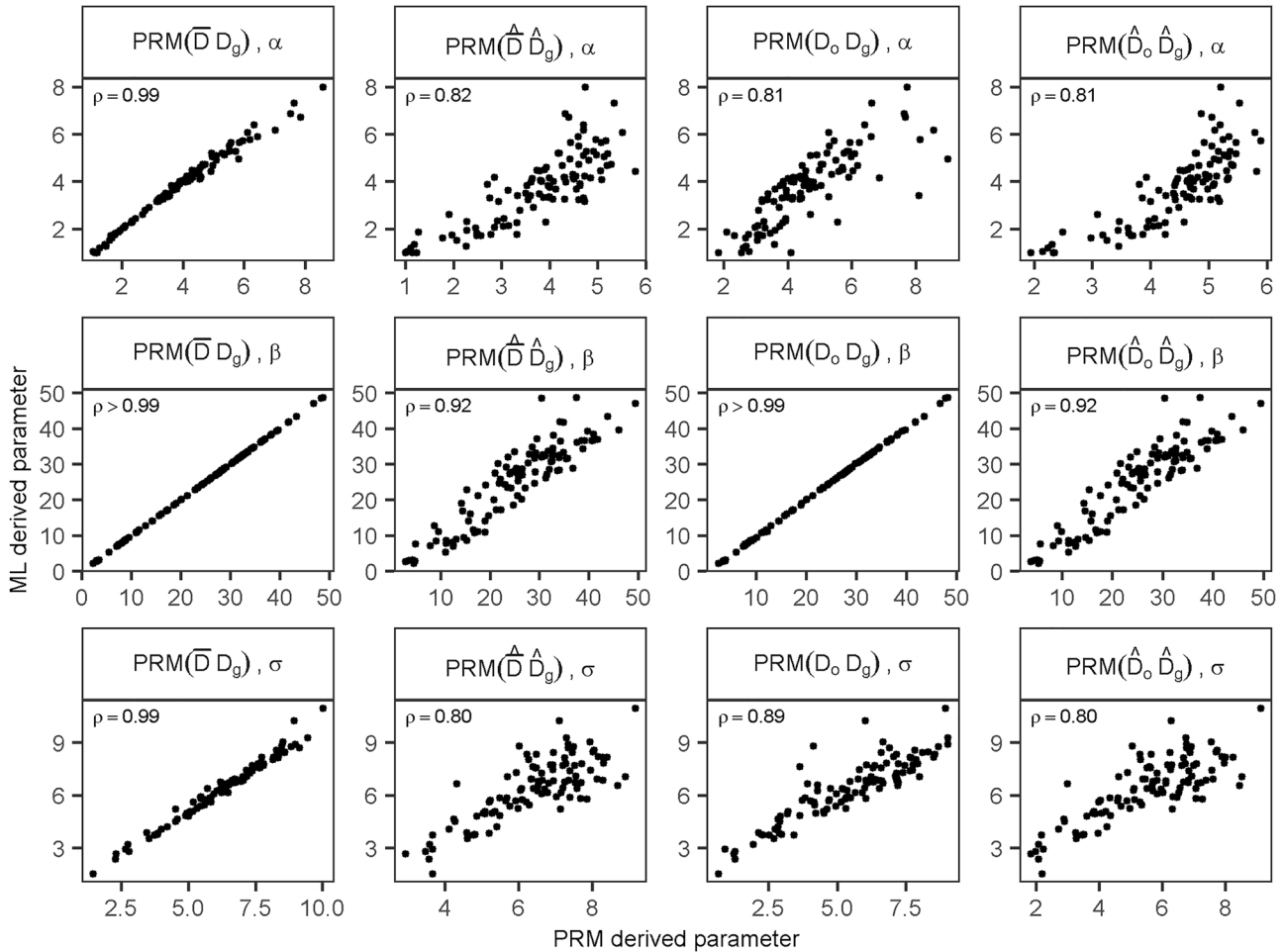
Fig. 2 Correlations between shape (α), scale (β) parameters and standard deviation (σ) of the *dbh* distributions obtained using maximum likelihood (ML) and PRM methods. $PRM(\overline{D}D_g)$: PRM system of Eqs. (8, 9) applied to the field values of $\overline{D}$ and D_g. $PRM(\widehat{\overline{D}}\widehat{D}_g)$: PRM system of Eqs. (8, 9) used with values of $\widehat{\overline{D}}$ and $\widehat{D}_g$ obtained from age and site

index. $PRM(D_oD_g)$ and $PRM(\widehat{D}_o\widehat{D}_g)$: PRM system of Eqs. (8, 10) using the field values of N, D_o and D_g and their predicted values $\widehat{N}$, $\widehat{D}_o$ and $\widehat{D}_g$, resp. N: stand density. D_g: quadratic mean diameter. D_o: dominant diameter (mean diameter of the 100 thickest trees ha^{-1})

generating predictions from site index, age and thinning regime.

3.3 Class-wise and global errors for *dbh* distributions

Table 3 Average difference between maximum likelihood (ML) estimates of α, β and σ and the estimates of these parameters using PRM methods and *t* test *p* values for those differences. $PRM(\overline{D}D_g)$: PRM system of Eqs. (8, 9) applied to the field values of $\overline{D}$ and D_g. $PRM(\widehat{\overline{D}}\widehat{D}_g)$: PRM system of Eqs. (8, 9) applied to values of $\widehat{\overline{D}}$ and $\widehat{D}_g$ obtained from age and site index using the corresponding models.

$PRM(D_oD_g)$ and $PRM(\widehat{D}_o\widehat{D}_g)$: PRM system of Eqs. (8, 10) using the field values of D_o, D_g and N and their predicted values $\widehat{D}_o$, $\widehat{D}_g$ and N, resp. $\overline{D}$, D_o, D_g and N, are the mean diameter, dominant diameter, quadratic mean diameter and stand density. $\widehat{\overline{D}}$, $\widehat{D}_o$ and $\widehat{D}_g$ represent predictions of $\overline{D}$, D_o and D_g obtained from the yield table models

Comparison	α		β		σ	
	Mean difference	*p* value	Mean difference	*p* value	Mean difference	*p* value
ML vs $PRM(\overline{D}D_g)$	− 0.38	0.70	0.01	0.99	0.39	0.70
ML vs $PRM(D_oD_g)$	− 4.18	0.00	− 0.04	0.97	3.52	0.00
ML vs $PRM(\widehat{\overline{D}}\widehat{D}_g)$	− 0.03	0.97	− 0.15	0.88	− 0.08	0.94
ML vs $PRM(\widehat{D}_o\widehat{D}_g)$	− 4.27	0.00	− 0.19	0.85	3.51	0.00

Age vs Weibull parameters and standard deviation of the dbh distributions obtained by different methods
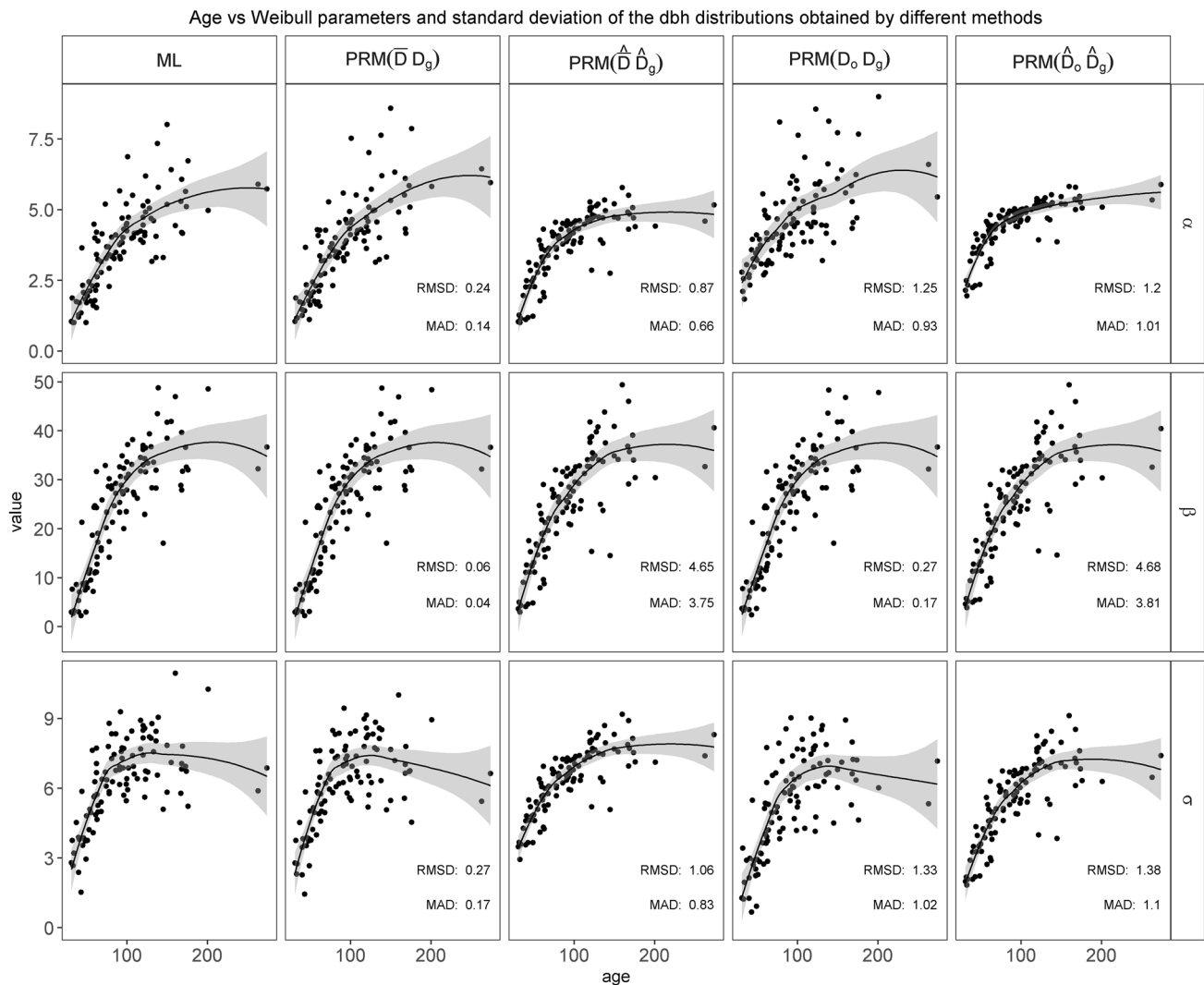

Fig. 3 Age vs. Weibull parameters and standard deviation of the *dbh* distributions obtained by: ML, maximum likelihood. *PRM*($\overline{D}D_g$): PRM system of Eqs. (8, 9) applied to the field values of $\overline{D}$ and D_g. *PRM*($\widehat{\overline{D}}\widehat{D}_g$): PRM system of Eqs. (8, 9) used with values of $\widehat{\overline{D}}$ and $\widehat{D}_g$ obtained from age and site index. *PRM*(D_oD_g) and *PRM*($\widehat{D}_o\widehat{D}_g$): PRM system of Eqs. (8, 10) using the field values of N, D_o and D_g and their predicted values $\widehat{N}$, $\widehat{D}_o$ and $\widehat{D}_g$, resp. N: stand density. D_g: quadratic mean diameter. D_o: dominant diameter (mean diameter of the 100 thickest trees ha^{-1})

The analysis of class-wise *Bias*, *RMSE* and *MAE* showed that the effect of chaining models to predict the input variables of the PMR methods had an important impact on the accuracy and precision of the recovered *dbh* distributions, which is more important than the PMR method itself (Fig. 4). For both PRM methods, class-wise *RMSE* and *MAE* became up to three times larger for *dbh*<30 cm when used with inputs predicted from site index and age than when applied with the field measurements while differences between PRM methods were of much smaller magnitude (Fig. 4). Even though there were some exceptions, (i.e. *dbh* between 50 and 60 cm and 60 and 70 cm), the class-wise error metrics showed that the PRM method based on $\widehat{\overline{D}}$ and $\widehat{D}_g$ outperformed the PRM method base $\widehat{D}_o$, $\widehat{D}_g$ and $\widehat{N}$ and provided

estimated frequencies that were closer to the observed ones for almost all *dbh* classes.

The two PRM methods applied to the field values of $\overline{D}$ and D_g or D_o, D_g and N provided class-wise R^2 and global gR^2 similar to those obtained when using ML (Fig. 5, Table 3). In general, the PRM method using D_g, D_o and N had a slightly worse performance than ML and the PRM method based on $\overline{D}$ and D_g. General trends for both PRM methods and ML showed that class-wise R^2 were high for the small and medium size classes (*dbh*<50 cm) but then rapidly decreased for classes with *dbh*>60 cm. However, trees with *dbh*>60 cm are relatively infrequent in the study area and their relative importance in gR^2 is small, being gR^2 larger than 0.94 for both ML and the PRM

Fig. 4 Class-wise error metrics (bias, root mean square error RMSE, mean absolute error MAE) for maximum likelihood (ML) and the studied PRM methods. $PRM(\overline{D}D_g)$: PRM system of Eqs. (8, 9) applied to the field values of $\overline{D}$ and D_g. $PRM(\widehat{\overline{D}}\widehat{D}_g)$: PRM system of Eqs. (8, 9) used with values of $\widehat{\overline{D}}$ and $\widehat{D}_g$ obtained from age and site index. $PRM(D_oD_g)$ and $PRM(\widehat{D}_o\widehat{D}_g)$: PRM system of Eqs. (8, 10) using the field values of N, D_o and D_g and their predicted values $\widehat{N}$, $\widehat{D}_o$ and $\widehat{D}_g$, resp. N: stand density. D_g: quadratic mean diameter. D_o: dominant diameter (mean diameter of the 100 thickest trees ha^{-1})

Class-wise error metrics for dbh by method
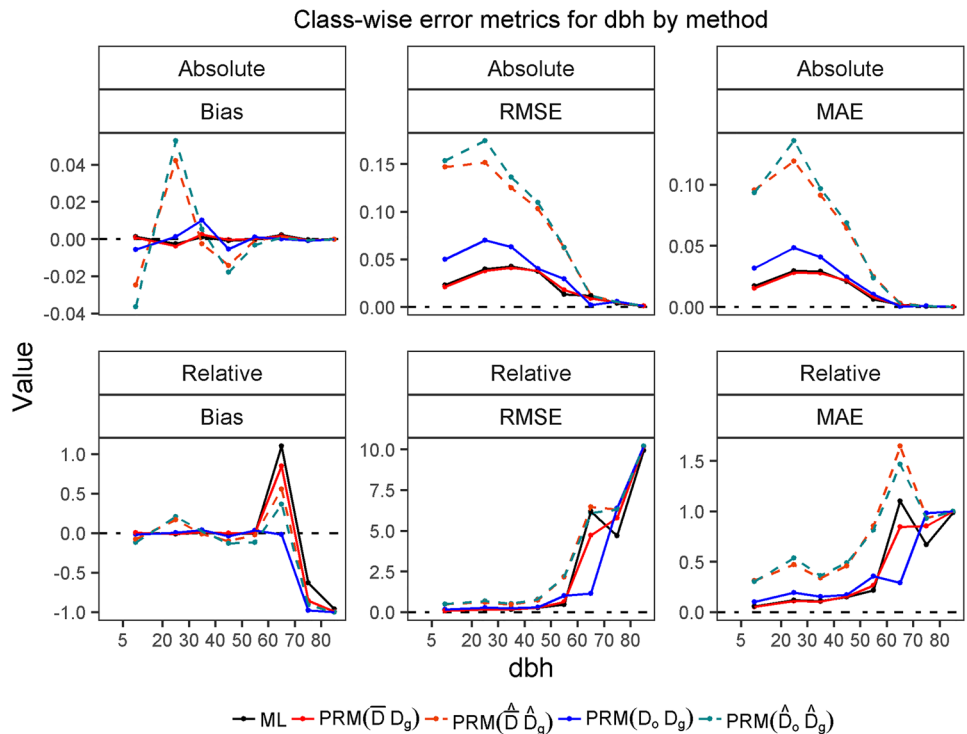
methods applied to the field values of $\overline{D}$ and D_g and D_o, D_g and N (Fig. 5). When the PRM methods were applied with the predicted values of $\overline{D}$ and D_g or D_o, D_g and N obtained from site index and age, gR^2 decreased from 97.80 to 69.92% for the PRM method using $\widehat{\overline{D}}$ and $\widehat{D}_g$ and from 94.41 to 64.39% for the PRM method using $\widehat{D}_o$, $\widehat{D}_g$ and $\widehat{N}$ (Fig. 5, Table 4). These values of gR^2 (69.92% and 64.39%) show that both PRM methods were able to explain a significant portion of the variance of the *dbh*-class frequencies even when applied to predicted values of $\overline{D}$ and D_g or D_o, D_g and N. Therefore, transforming *P. sylvestris* yield table using either of these PRM methods would provide useful information for forest management.

For all methods, *gBias* was negligible. In terms of *gRMSE* and *gMAE*, the PRM method using $\widehat{\overline{D}}$ and $\widehat{D}_g$ also outperformed the PRM method based on $\widehat{D}_o$, $\widehat{D}_g$ and $\widehat{N}$ and appeared as a better alternative to transform *P. sylvestris* yield tables into a diameter distribution model. However, PRM methods applied to field measurements of $\overline{D}$ and D_g and D_o, D_g and N had values of *gRMSE* and *gMAE* that were 288.07% and 28.47% and 169.35% and 31.84% smaller than their counterparts using predicted values of D_g and $\overline{D}$ and D_o, D_g and N (Table 4). This again indicates that the most important source of error when recovering the *dbh* distributions is the chain of models used to obtain attributes needed by each PRM method and not the PRM method itself.

Results from the *t* test did not show significant differences at the 0.05 level between the absolute errors for $PRM(\widehat{\overline{D}}\widehat{D}_g)$ and $PRM(\widehat{D}_o\widehat{D}_g)$ for the first and the last two *dbh* classes (Table 4). For the second, third and fourth *dbh* classes, absolute errors for $PRM(\widehat{\overline{D}}\widehat{D}_g)$ were significantly smaller than errors for $PRM(\widehat{D}_o\widehat{D}_g)$, for the fifth and sixth *dbh* class errors for $PRM(\widehat{D}_o\widehat{D}_g)$ were significantly smaller than errors for $PRM(\widehat{\overline{D}}\widehat{D}_g)$. Considering all *dbh* classes together, absolute errors for $PRM(\widehat{\overline{D}}\widehat{D}_g)$ were on average of smaller magnitude than errors obtained with $PRM(\widehat{D}_o\widehat{D}_g)$ (Table 5).

4 Discussion

The first step to transform the yield tables into *dbh* distribution models was the choice of the probability function to model *dbh* frequencies. Several probability functions, including the Gram-Charlier series (Meyer 1930), log-normal probability functions (Bliss and Reinker 1964), Jonshon's SB probability functions (Rennolls and Wang 2005; Gorgoso et al. 2012; Özçelik et al. 2016; Cosenza et al. 2019), beta probability functions (Gorgoso et al. 2012) and Weibull probability functions (e.g. Cao et al. 1982; Magnussen 1986; Poudel and Cao 2013; Mora et al. 2012; Cao and Coble 2014; Nanos and Sjöstedt de Luna 2017) have been used to describe *dbh* distributions for even-aged stands.
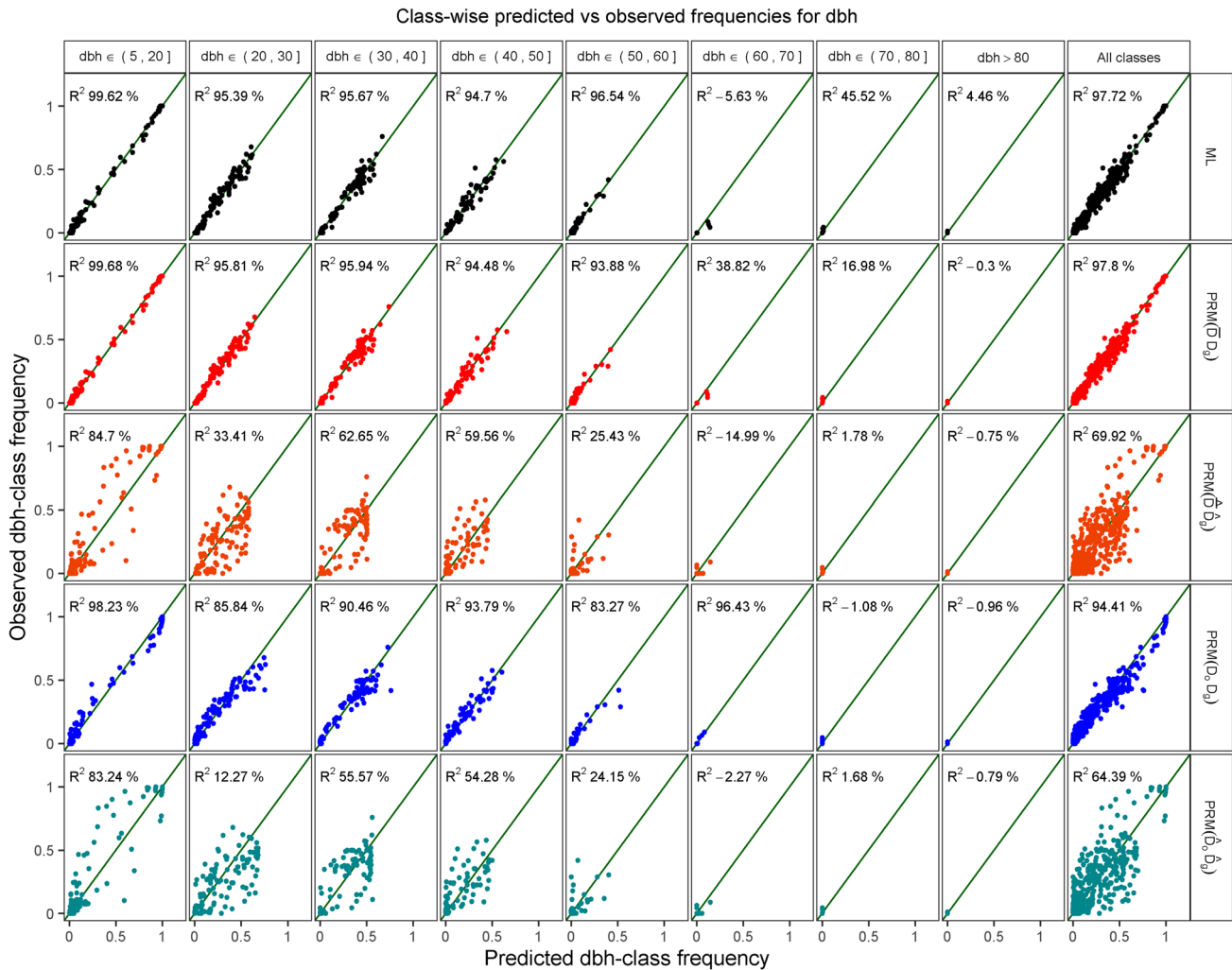
Fig. 5 Class-wise observed vs. predicted *dbh* frequencies for maximum likelihood (ML) and the studied PRM methods. $PRM(\overline{D}\widetilde{D}_g)$: PRM system of Eqs. (8, 9) applied to the field values of $\overline{D}$ and $\widetilde{D}_g$. $PRM(\widehat{\overline{D}}\widehat{D}_g)$: PRM system of Eqs. (8, 9) used with values of $\widehat{\overline{D}}$ and $\widehat{D}_g$ obtained from age and site index. $PRM(D_o D_g)$ and $PRM(\widehat{D}_o \widehat{D}_g)$:

PRM system of Eqs. (8, 10) using the field values of N, D_o and D_g and their predicted values $\widehat{N}$, $\widehat{D}_o$ and $\widehat{D}_g$, resp. N: stand density. D_g: quadratic mean diameter. D_o: dominant diameter (mean diameter of the 100 thickest trees ha^{-1})

Among the mentioned models, the Weibull probability function used in this study is presumably the most widely used to model *dbh* distributions. Several variants of the Weibull models can be found in the literature depending on (1) whether a 2-parameter (e.g. Bailey and Dell 1973) or a 3-parameter (e.g. Cao 2004) Weibull function was used to model the size distribution and (2) whether the distribution function included truncation points to accommodate for specific characteristics of the corresponding sampling design (Mehtätalo et al. 2011; Nanos and Sjöstedt de Luna 2017) or not (e.g. Bailey and Dell 1973). We used 3-parameter Weibull models without truncation points and with a fixed location parameter γ. Three-parameter Weibull distribution models with free γ have been developed for other pine species in the Iberian peninsula (e.g. Palahí et al. 2006; Piqué

et al. 2011) and have shown large flexibility because the size distribution model has an extra degree of freedom. However, a PRM system for Weibull distributions with free γ requires developing models for additional stand variables (e.g. the minimum *dbh*). Results from our study show that the most important source of error in the modeled sized distributions were derived or propagated from the regression models predicting the stand variables needed for the PRM methods. Thus, more parsimonious models would be less affected by the errors propagating from the regression models used to obtain stand attributes for the PRM systems, which, to some extent, can compensate their smaller flexibility.

Once a probability function has been chosen, it is necessary to apply a method to derive the *dbh* distribution parameters. In this study, we used PRM methods but other alternatives

Table 4 Global error metrics for maximum likelihood (ML) and PRM methods. $PRM(\overline{D}D_g)$: PRM system of Eqs. (8, 9) applied to the field values of $\overline{D}$ and D_g. $PRM(\widehat{\overline{D}}\widehat{D}_g)$: PRM system of Eqs. (8, 9) applied to values of $\widehat{\overline{D}}$ and $\widehat{D}_g$ obtained from age and site index using the corresponding models. $PRM(D_oD_g)$ and $PRM(\widehat{D}_o\widehat{D}_g)$: PRM system of Eqs. (8, 10) using the field values of D_o, D_g and N and their predicted values $\widehat{D}_o$, $\widehat{D}_g$ and N, resp. $\Delta PRM(\widehat{\overline{D}}\widehat{D}_g)$ and $\Delta PRM(\widehat{D}_o\widehat{D}_g)$: increase in each error metric and PRM method when attributes for the PRM sys-

tems of equations were predicted from age, site index and silvicultural regime using the yield table models. Increases are expressed in terms relative to the error metrics obtained using the field measurements of $\overline{D}$ and D_g and D_o, D_g and N. $\overline{D}$: mean diameter, D_o: dominant diameter, D_g: quadratic mean diameter and N : stand density. $\widehat{\overline{D}}$, $\widehat{D}_o$ and $\widehat{D}_g$ represent predictions of $\overline{D}$, D_o and D_g obtained from the yield table models

Metric	ML	$PRM(\overline{D}D_g)$	$PRM(\widehat{\overline{D}}\widehat{D}_g)$	$\Delta PRM(\widehat{\overline{D}}\widehat{D}_g)$	$PRM(D_oD_g)$	$PRM(\widehat{D}_o\widehat{D}_g)$	$\Delta PRM(\widehat{D}_o\widehat{D}_g)$
gBias	− 9.95E-19	− 1.13E-18	− 7.00E-19	− 37.87%	− 2.32E-19	− 2.77E-19	19.36%
gRMSE	0.0266	0.0261	0.0966	269.72%	0.0417	0.105	152.58%
gMAE	0.0132	0.0128	0.0498	288.07%	0.0196	0.0527	169.35%
gR^2	97.7%	97.8%	70.0%	-28.47%	94.4%	64.4%	-31.84%

such as parameter prediction methods, PPM (e.g. Gorgoso et al. 2007; Siipilehto and Mehtätalo 2013), could be used with similar purposes. An advantage of PRM approaches over PPM approaches is that PRM ensures compatibility with the forest attributes they use to derive the parameters of the *dbh* distributions. This advantage is very important in the context of our study, where the main objective is the transformation of existing yield table models into *dbh* distribution models. Using PPM could result in diameter distributions providing values of D_g different from those already present in the yield tables and that is clearly an undesired feature for the resulting models as we aim at preserving compatibility of the *dbh* distribution models with their associated yield tables.

Differences between the two PRM methods under analysis showed that using $\widehat{\overline{D}}$ and $\widehat{D}_g$ provided better results than using $\widehat{D}_o$, $\widehat{D}_g$ and $\widehat{N}$, globally and also for the *dbh* classes 20 cm to 30 cm, 30 cm to 40 cm and 40 cm to 50 cm. The PRM method based on $\widehat{D}_o$, $\widehat{D}_g$ and $\widehat{N}$ provided better results for the classes 50 cm to 60 cm and 60 cm to 70 cm but the absolute values of *RMSE* and *MAE* of both PRM methods for these classes were orders of magnitude smaller than those observed for the 20 cm to 30 cm, 30 cm to 40 cm and 40 cm to 50 cm *dbh* classes. While, in general, the PRM method based on $\widehat{\overline{D}}$

and $\widehat{D}_g$ provided better results, the PRM method, based on $\widehat{D}_o$, $\widehat{D}_g$ and $\widehat{N}$, was able to explain 64.39% of the variability of *dbh*-class frequencies. This result indicates that using a dominant variable as proposed by Ayuga-Téllez et al. (2014) in a PRM method can effectively recover a significant amount of information about a size distribution. Future research should explore other applications where PRM methods are applied to dominant attributes. For example, numerous yield tables such as Montero et al. (2008) yield tables for *Pinus halepensis* Mill.; Sánchez et al. (2008) yield tables for *Pinus radiata* or the tables analyzed here, contain models for stand density, mean tree height and dominant height and those attributes can be used to transform yield tables into tree-height distribution models. Similarly, numerous studies have shown that it is possible to accurately estimate dominant height and mean tree height using airborne laser scanning data (e.g. Næsset 2002; González-Ferreiro et al. 2012; Mauro et al. 2017). Those results suggest that there are potential applications of PRM methods based on dominant attributes to projects aiming at mapping forest structure and tree-height distributions.

One important feature of *dbh* distribution modelling studies is that, by definition, the response variable is a curve,

Table 5 Paired sample t test comparison of absolute value of errors obtained using $PRM(\widehat{\overline{D}}\widehat{D}_g)$ and $PRM(\widehat{D}_o\widehat{D}_g)$ for *dbh* classes. Differences between absolute value of errors are obtained subtracting the absolute value of the errors for $PRM(\widehat{D}_o\widehat{D}_g)$ to the absolute value of the errors for $PRM(\widehat{\overline{D}}\widehat{D}_g)$. Negative values indicate that errors in $PRM(\widehat{D}_o\widehat{D}_g)$ are

larger than those in $PRM(\widehat{\overline{D}}\widehat{D}_g)$. $\overline{D}$: mean diameter, D_o: dominant diameter, D_g: quadratic mean diameter and N : stand density. $\widehat{\overline{D}}$, $\widehat{D}_o$ and $\widehat{D}_g$ represent predictions of $\overline{D}$, D_o and D_g obtained from the yield table models

dbh class (cm)	dbh ε (5,20)	dbh ε (20,30)	dbh ε (30,40)	dbh ε (40,50)	dbh ε (50,60)	dbh ε (60,70)	dbh ε (70,80)	dbh > 80	All dbh classes
Mean difference between errors in absolute value	1.80E-3	− 1.64E-2	− 5.69E-3	− 4.43E-3	1.11E-03	3.61E-4	− 5.84E-6	− 7.71E-8	− 2.91E-3
p value	5.72E-1	1.23E-4	3.56E-2	7.46E-4	4.93E-2	2.67E-3	6.72E-1	3.39E-1	1.83E-4

or at least a set of frequencies or stocking levels by *dbh* classes. That characteristic of *dbh* distribution models has important implications for accuracy assessment purposes that typically require a step where the dimensionality of the problem is substantially reduced. Such dimensionality reduction can be achieved comparing the parameters of a distribution model obtained from field measurement to their respective counterparts derived from a PRM or PPM method (Gorgoso et al. 2007). This type of accuracy assessment reduces the problem to analyzing a small set of parameters, but this relatively straightforward set-up has interpretability problems because (1) for some *dbh* distribution models, i.e. Weibull models, it is difficult to translate errors or R^2 values for the estimation of a given distribution parameter to magnitudes with a physical interpretation and (2) in real application, it is not possible to know the true values that define a *dbh* distribution so comparisons are always made between two estimates. Goodness of fit statistical tests, such as the Kolmogorov–Smirnov test, are also used to evaluate *dbh* distribution models. However, if there is a high variability in the number of trees per plot, results from statistical tests can be also problematic because the power of these tests might substantially vary from plot to plot as a result of their different sample size. In addition, only rejections of the null hypothesis are meaningful when there is no control over the power of the test (Mauro et al. 2010). In our study, the number of trees per plot ranged from 24 to 193 (Table 1); for that reason, we limited the use of statistical tests to the initial inspection where Weibull distribution models with fixed γ were chosen and, for that step, we combined the use of statistical test with comparisons of the frequencies provided by models using fixed and variable γ for 10 cm *dbh* classes. Finally, the use of global indexes in combination with class-wise error metrics (e.g. Gorgoso et al. (2007) and Siipilehto and Mehtätalo (2013)) provides the best alternative to evaluate *dbh* distribution models. Global indexes provide summarizations that have a physical interpretation and can be easily used for model selection purposes and class-wise error metrics provide a detailed description of where the modelling errors occur that complement the information that global indexes cannot provide. In particular, we find that errors for *dbh* classes are the metrics preferred and better understood by forest managers as they allow them to evaluate if a model is accurate enough for specific *dbh* classes that are of interest for them.

Modelling of *dbh* distributions appear frequently in two different contexts in the forest research literature: on one hand, in growth and yield modelling applications where the *dbh* distributions are ultimately predicted from stand age (e.g. Smalley and Bailey (1974), Diéguez-Aranda et al. (2006) and Piqué et al. (2011)) and, on the other hand, in forest structure mapping applications where *dbh* distributions

are predicted from remotely sensed data (e.g. Maltamo et al. (2007); Breidenbach et al. (2008); Strunk et al. (2017); Arias-Rodil et al. (2018) and Mauro et al. (2019)). In the first type of studies, models are developed to describe reference conditions for forest stands subject to specific silvicultural treatments while the second type of study aims at describing the current state of the forests. Numerous studies have explored both contexts separately, but to the best of our knowledge, there is no study where both, reference condition from models like the one developed here and current state conditions derived from remote sensing methods, are combined to guide forest management activities. We believe that an analysis of that kind is clearly needed to bridge the gap and make use of the synergies between these two different modelling contexts.

5 Conclusions

The main conclusions from this study are first, both PRM methods for *dbh* allow explaining significant portions of the variability of the *dbh*-class frequencies and thus can provide useful information for forest managers in *P. sylvestris* stands with similar conditions to the study area. Second, the PRM method based on D_g and $\overline{D}$ was more accurate than the PRM method based on D_o, N and D_g. However, errors propagating from the chain of models predicting the stand attributes required by each PRM method were more important than the PRM method used to recover *dbh* distributions.

Data availability The datasets generated during and/or analyzed during the current study are available at the Zenodo repository: https://doi.org/10.5281/zenodo.3934885.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Arias-Rodil M, Diéguez-Aranda U, Álvarez-González JG et al (2018) Modeling diameter distributions in radiata pine plantations in Spain with existing countrywide LiDAR data. Ann For Sci 75:36. https://doi.org/10.1007/s13595-018-0712-z

INRAE Springer

Ayuga-Téllez E, Mauro-Gutiérrez F, García-Abril A et al (2014) Comparison of estimation methods to obtain ideal distribution of forest tree height. Comput Electron Agric 108:191–199. https://doi.org/10.1016/j.compag.2014.07.011

Bailey RL, Dell T (1973) Quantifying diameter distributions with the Weibull function. For Sci 19:97–104

Bliss CI, Reinker KA (1964) A lognormal approach to diameter distributions in even-aged stands. For Sci 10:350–360. https://doi.org/10.1093/forestscience/10.3.350

Bollandsås OM, Næsset E (2007) Estimating percentile-based diameter distributions in uneven-sized Norway spruce stands using airborne laser scanner data. Scand J For Res 22:33–47. https://doi.org/10.1080/02827580601138264

Breidenbach J, Gläser C, Schmidt M (2008) Estimation of diameter distributions by means of airborne laser scanner data. Can J For Res 38:1611–1620

Burkhart HE, Tomé M (2012) Modeling forest trees and stands. Springer Science & Business Media

Cao QV (2004) Predicting parameters of a Weibull function for modeling diameter distribution. For Sci 50:682–685

Cao QV, Burkhart HE, Lemin RC Jr (1982) Diameter distributions and yields of thinned loblolly pine plantations. Virginia Polytechnic Institute and State University, School of Forestry and Wildlife Resources, Blacksburg, VA

Cao QV, Coble DW (2014) Deriving a diameter distribution from stand table data. For Sci 60:628–635. https://doi.org/10.5849/forsci.12-150

Cosenza DN, Soares P, Guerra-Hernández J et al (2019) Comparing Johnson's SB and Weibull functions to model the diameter distribution of forest plantations through ALS data. Remote Sens 11:2792. https://doi.org/10.3390/rs11232792

Del Río M (1998) Régimen de claras y modelo de crecimiento para *Pinus Sylvestris* L. en los sistemas Central e Ibérico. Tesis-Universidad Politécnica de Madrid. E.T.S.I. Montes

Diéguez-Aranda U, Castedo Dorado F, Álvarez González JG, Rojo Alboreca A (2006) Dynamic growth model for Scots pine (*Pinus sylvestris* L.) plantations in Galicia (north-western Spain). Ecol Model 191:225–242. https://doi.org/10.1016/j.ecolmodel.2005.04.026

Gadow KV, Hui G (1999) Modelling stand development. Modelling Forest Development. Springer, Netherlands, pp 26–60

Gobakken T, Næsset E (2005) Weibull and percentile models for lidar-based estimation of basal area distribution. Scand J For Res 20:490–502. https://doi.org/10.1080/02827580500373186

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

Gorgoso JJ, Alvarez Gonzalez JG, Rojo A, Grandas-Arias JA (2007) Modelling diameter distributions of Betula alba L. stands in northwest Spain with the two-parameter Weibull function. For Syst 16:113–123. https://doi.org/10.5424/srf/2007162-01002

Gorgoso JJ, Rojo A, Cámara-Obregón A, Diéguez-Aranda U (2012) A comparison of estimation methods for fitting Weibull, Johnson's SB and beta functions to *Pinus pinaster*, *Pinus radiata* and *Pinus sylvestris* stands in northwest Spain. For Syst 21:446–459

Hyink DM, Moser JW (1983) A generalized framework for projecting forest yield and stand structure using diameter distributions. For Sci 29:85–95

IFN (2015) IFN. Principales especies arbóreas. Miscelánea

Knoebel BR, Burkhart HE, Beck DE (1986) A growth and yield model for thinned stands of yellow-poplar. For Sci 32:a0001-z0002

Magnussen S (1986) Diameter distributions in *Picea abies* described by the Weibull model. Scand J For Res 1:493–502

Maltamo M, Eerikäinen K, Pitkänen J et al (2004) Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. Remote Sens Environ 90:319–330

Maltamo M, Malinen J, Packalén P et al (2006) Nonparametric estimation of stem volume using airborne laser scanning, aerial photography, and stand-register data. Can J For Res 36:426–436. https://doi.org/10.1139/x05-246

Maltamo M, Suvanto A, Packalén P (2007) Comparison of basal area and stem frequency diameter distribution modelling using airborne laser scanner data and calibration estimation. For Ecol Manage 247:26–34

Mauro F, Frank B, Monleon VJ et al (2019) Prediction of diameter distributions and tree-lists in southwestern Oregon using LiDAR and stand-level auxiliary information. Can J For Res 49:775–787. https://doi.org/10.1139/cjfr-2018-0332

Mauro F, Monleon VJ, Temesgen H, Ruíz Fernández LÁ (2017) Analysis of spatial correlation in predictive models of forest variables that use LiDAR auxiliary information. Can J For Res 47:788–799. https://doi.org/10.1139/cjfr-2016-0296

Mauro F, Valbuena R, Manzanera JA, García-Abril A (2010) Influence of Global Navigation Satellite System errors in positioning inventory plots for tree-height distribution studies This article is one of a selection of papers from Extending Forest Inventory and Monitoring over Space and Time. Can J For Res 41:11–23. https://doi.org/10.1139/X10-164

Mehtätalo L, Comas C, Pukkala T, Palahí M (2011) Combining a predicted diameter distribution with an estimate based on a small sample of diameters. Can J For Res 41:750–762

Mehtätalo L, Lappi J (2019) Forest biometrics with examples in R. Chapman and Hall/CRC, Boca Raton, Florida

Meyer WH (1930) Diameter distribution series in even aged forest stands. Yale School of Forestry, New Haven, CT, US

Montero G, Cañellas I, Ruiz-Peinado R (2008) Growth and yield models for Pinus halepensis Mill. For Syst 10:179–201

Mora JV, del Rio M, Bravo-Oviedo A (2012) Dynamic growth and yield model for Black pine stands in Spain. For Syst 21:439–445

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

Nanos N, Sjöstedt de Luna S (2017) Fitting diameter distribution models to data from forest inventories with concentric plot design. Forest Systems 26(2):e015. https://doi.org/10.5424/fs/2017262-10486

Özçelik R, Fidalgo Fonseca TJ, Parresol BR, Eler Ü (2016) Modeling the diameter distributions of Brutian pine stands using Johnson's SB distribution. For Sci 62:587–593. https://doi.org/10.5849/forsci.15-089

Palahí M, Pukkala T, Trasobares A (2006) Modelling the diameter distribution of *Pinus sylvestris*, *Pinus nigra* and *Pinus halepensis* forest stands in Catalonia using the truncated Weibull function. Forestry 79:553–562

Piqué M, del Río M, Calama R, Montero G (2011) Modelling silviculture alternatives for managing" *Pinus Pinea* L." forest in North-East Spain. For syst 20:2–20

Poudel KP, Cao QV (2013) Evaluation of methods to predict Weibull parameters for characterizing diameter distributions. For Sci 59:243–252

Qin J, Cao QV (2006) Using disaggregation to link individual-tree and whole-stand growth models. Can J For Res 36:953–960

R Core Team (2018) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria

Rennolls K, Wang M (2005) A new parameterization of Johnson's SB distribution with application to fitting forest tree diameter data. Can J For Res 35:575–579

Rojo A, Montero G (2020) Tree measurements and summaries of the field plots used to develop Rojo and Montero (1996) yield tables for

Pinus sylvestris L. in central Spain [dataset] Zenodo. V1. https://doi.org/10.5281/zenodo.3934885

Rojo A, Montero G (1996) El pino silvestre en la Sierra de Guadarrama: historia y selvicultura de los Pinares de Cercedilla, Navacerrada y Valsain. Ministerio de Agricultura, Pesca y Alimentación, Secretaria General Tecnica, Centro de Publicaciones, Madrid

Sánchez F, Rodríguez R, Rojo A et al (2008) Crecimiento y tablas de producción de *Pinus radiata* D Don en Galicia. For Syst 12:65–83

Siipilehto J, Mehtätalo L (2013) Parameter recovery vs. parameter prediction for the Weibull distribution validated for Scots pine stands in Finland 47:1–22

Smalley GW, Bailey RL (1974) Yield tables and stand structure for shortleaf pine plantations in Tennessee, Alabama, and Georgia highlands. Res Pap SO-97 New Orleans, LA: US Department of Agriculture, Forest Service, Southern Forest Experiment Station 57 p 97:

Strub MR, Burkhart HE (1975) A class-interval-free method for obtaining expected yields from diameter distributions. For Sci 21:67–69

Strunk LJ, Gould JP, Packalen P et al (2017) An examination of diameter density prediction with k-NN and airborne lidar. Forests 8:444. https://doi.org/10.3390/f8110444

Weiskittel AR, Hann DW, Kershaw Jr JA, Vanclay JK (2011) Forest growth and yield modeling. John Wiley & Sons

Authors and Affiliations

Francisco Mauro[1] · **Antonio García-Abril[2]** · **Esperanza Ayuga-Téllez[2]** · **Alberto Rojo-Alboreca[3]** · **Ruben Valbuena[4]** · **José Antonio Manzanera[2]**

✉ José Antonio Manzanera
joseantonio.manzanera@upm.es

Francisco Mauro
francisco.mauro@oregonstate.edu

Antonio García-Abril
antonio.garcia.abril@upm.es

Esperanza Ayuga-Téllez
esperanza.ayuga@upm.es

Alberto Rojo-Alboreca
alberto.rojo@usc.es

Ruben Valbuena
r.valbuena@bangor.ac.uk

[1] Forest Engineering Resources and Management, College of Forestry, Oregon State University, 329 Peavy Forest Science Center 3100 SW Jefferson Way, Corvallis, OR, USA

[2] College of Forestry and Natural Environment, Universidad Politécnica de Madrid. Ciudad Universitaria, 28040 Madrid, Spain

[3] Dep. Enxeñaría Agroforestal, Universidade de Santiago de Compostela, 27002 Lugo, Spain

[4] School of Natural Sciences. Thoday Building, Bangor University, Bangor LL57 2UW, UK