Height-diameter models for maritime pine in Portugal: a comparison of basic, generalized and mixed-effects models

Esteban Gómez-García [1-2], Teresa F Fonseca [2], Felipe Crecente-Campo [1], Luís R Almeida [2], Ulises Diéguez-Aranda [1], Shongming Huang [3], Carlos P Marques [2]

Tree height is a key variable in forest monitoring studies and for forest management. However, tree height measurement is time consuming, and the recommended procedure is to use estimates from height-diameter models. Here, we compare height-diameter model forms and approaches for predicting tree height (*h*) as a function of tree diameter at breast height (*d*) and additional stand level covariates. Four model forms were evaluated: (i) basic models (which only used *d* as predictor variable); (ii) generalized models (which used *d* and stand-level predictor variables); (iii) a mixed-effects model based on the best basic model; and (iv) a mixed-effects model based on the best generalized model. Several alternatives aimed at minimizing height measurement effort were tested in terms of accuracy and applicability. From a practical point of view, the selected generalized model is recommended for estimating the height of maritime pine in Portugal. The results also show that a calibrated basic mixed model provides more accurate results than a basic model locally fitted when the number of *h-d* observations is limited.

Keywords: *Pinus pinaster* Ait., Sampling Design, Local Model, Stand Variables, Generalized *h-d* Relationship, Calibration

Introduction

Diameter at breast height (*d*) and total tree height (*h*) are key variables in forestry applications. These variables enable study of the horizontal and vertical forest structure, and they are used to estimate timber volume, site index and other important variables related to forest growth and yield, succession and carbon budget models (Peng 2001). Although *d* is routinely measured in current forest inventories, *h* is usually measured in a subsample of trees to save time and expense. Hence, the diameter to height (*h-d*) relationship is widely used to overcome the lack of information about tree height. In even-aged stands, differences in the *h-d* relationship are mainly related to age, density and site (Curtis 1967). In a specific stand, the *h-d* curve shifts upwards with increasing age until reaching a certain age, which is particular to the species and the site (*e.g.*, Loetsch et al. 1973, Pretzsch 2010). The curve is generally sigmoid in shape over the full range of diameters, and the slope depends on the site quality, with steeper slopes indicating better sites and flatter curves indicating poorer sites. Since the first freehand curves were developed in the early 1930s, many *h-d* models have been proposed for describing the relationship between both variables (for a review of the models, see Curtis 1967, Tomé 1988, Huang et al. 1992, Huang et al. 2000).

For height estimation using a basic model, which includes diameter at breast height as the sole independent variable [*h* = *f(d)*], the model must be fitted to data from each plot (representing a specific stand) and for each measurement occasion independently; this is usually referred to as a locally fitted basic model or simply a local model. Another option is the use of a generalized model that uses *d* and stand-specific variables as regressors [*h* = *f* (*d*, stand variables)], thus accounting for differences in the *h-d* relationship within stands and over time.

An alternative *h-d* modeling option is the development of a mixed-effects model (Lappi 1997, Calama & Montero 2004, Sharma & Parton 2007, Trincado et al. 2007). Mixed-effects models allow for both mean and subject-specific responses and they have been found to be useful for repeatedly measured data (Davidian & Giltinan 1995, Littell et al. 2006). The mean response considers only fixed parameters, common to the population, while the subject response considers both fixed parameters and random effects, specific to each subject. In forestry, the most common experimental subject is a permanent sample plot, in which some trees are measured repeatedly over time, or a stem analysis tree (Gregoire et al. 1995), in which the trajectories of height-age, diameter-age, and height-diameter can be reconstructed to determine how the height and diameter of trees change over time. As height and diameter data are generally obtained in plots, thus having a nested stochastic structure, there may be a lack of independence between observations, which yields biased estimates of confidence intervals for the parameter estimates (Searle et al. 1992). The use of mixed models is thought

☐ (1) Departamento de Ingeniería Agroforestal, Universidade de Santiago de Compostela, Escuela Politécnica Superior, R/ Benigno Ledo, Campus universitario, E-27002 Lugo (Spain); (2) Department of Forest Sciences and Landscape Architecture, University of Trás-os-Montes e Alto Douro, Apartado 1013, 5001-801 Vila Real (Portugal); (3) Forest Management Branch, Alberta Ministry of Sustainable Resource Development, 8th Floor, 9920-108 Street, Edmonton, Alberta T5K 2M4 (Canada)

@ Esteban Gómez-García (esteban.gomez@usc.es)

to help solve the lack of independence between observations (Huang et al. 2009). The inclusion of random effects, specific to each plot-inventory combination, allows the natural variability in the *h-d* relationship between different locations and time to be modeled, after defining a common fixed functional structure (Lindstrom & Bates 1990). In order to provide enhanced estimates, relative to those obtained when using a mean response in a new dataset, the random effects can be estimated when information about a sample of *h-d* pairs is available. This procedure is called calibration (or localization), and the resultant response is the calibrated response. Although there is no doubt about the advantages of the use of mixed models to avoid biased estimates for the covariance matrix, there is still some uncertainty about the advantage of its use for predictive purposes (De-Miguel et al. 2012).

This paper analyzes different model forms and approaches for developing a *h-d* relationship for maritime pine (*Pinus pinaster* Ait.) in Portugal, *i.e.*, the prevailing softwood species in this country. According to the National Forest Inventory (NFI - AFN 2010), maritime pine covers around 885 000 ha of land in Portugal (27% of the mainland forest area), with a wood volume of 85.8 × 10^6 m^3 and a total biomass of 49.7 × 10^6 Mg, which represents 24.8 × 10^6 Mg of sequestered carbon. Evaluation, comparison and selection of model forms and approaches were conducted with the purpose of identifying the optimal *h-d* model for incorporation in the MODISPINASTER growth and yield model (Fonseca 2004, Fonseca et al. 2012). Briefly, the model addresses forest growth and yield, risks (wind related) and management procedures (such as thinning and harvesting) for pure maritime pine stands. The level of detail of the output is the diameter class, with the diameter distribution recovered by the Johnson S$_B$ distribution (Parresol et al. 2010). The model can be downloaded from the CAPSIS (Computer-Aided Projection of Strategies In Silviculture) simulation platform (http://capsis.cirad.fr/capsis/model). Until recently, prediction of tree height using MODISPINASTER simulations was based on the following provisional equation parameterized by Almeida (1999 – eqn. 1):

$$h = hd \left(\frac{d}{dd} \right)^{0.09+0.04\,hd-0.049\,(N/1000+1.62\,t)}$$

where *N* represents the number of trees per hectare, *t* refers to the stand age in years, *hd* is the dominant height, defined as the mean height of the 100 largest-diameter trees per hectare, and *dd* is the dominant diameter, defined as the mean diameter of the 100 largest-diameter trees per hectare.

The overall objective of the present study was to select an *h-d* model form and a model-fitting approach for application to maritime pine in Portugal, to replace the model of eqn. 1 previously included in MODISPINASTER. The specific objectives of this research were as follows: (1) to compare basic, generalized and mixed-effects models; and (2) to test several alternatives for minimizing the height measurement effort in terms of accuracy and applicability.

Materials and methods

Data

This study used information from a large database on maritime pine collected in northern Portugal, mainly in the Tâmega Valley, which is the most representative continuous area in the state (latitude range: 41° 15′ - 41° 52′ N, longitude range: 7° 20′ - 8° 00′ W). In the Tâmega Valley, maritime pine occurs at an altitude between 100 and 900 m a.s.l. on hilly terrain. The soils are derived from granite and schist. The mean annual temperature varies between 13.1 °C and 16 °C at lower elevations (100-400 m), in an eastern direction. Above 400 m a.s.l, the mean annual temperature falls to 9.8 °C. The mean annual precipitation ranges between 660 and 1400 mm in the lower locations, and between 1000 mm and 2900 mm in higher locations (Marques 1991). Most forests and woodlands in the region are community areas and are managed by the Portuguese Forest Service (the National Forest Authority).

In each sampled stand, circular plots of 0.05 ha were established as permanent sample plots. The first measurements were carried out in 1997 and the most recent in 2010. A total of 387 inventories of 133 plots, with 1 to 4 measurements – depending on the year of plot installation – were available for study. The data set includes tree and stand variables. Available tree characteristics were diameter outside bark at breast height (*d*, cm) of all living trees taller than 1.30 m, and total height (*h*, m) for a subset of trees (including the proportion of the 100 largest-diameter trees per hectare). The diameters were measured with a diameter or girth tape, to the nearest 0.1 cm.

Criteria for selection of a subset of trees for height measurements varied over time. In the earlier measurements, a systematic layout was followed and heights were measured with a Blume-Leiss hypsometer to the nearest 0.5 m. Subsequently, 1-3 average trees within each diameter class of 5 cm width were selected and the heights were measured with a Vertex hypsometer to the nearest 0.1 m. For a subset of plot-inventory combinations (47 %), the diameter and height of all trees were measured. The average value of *h-d* observations per plot equals 27. Stand age (*t*, years) was evaluated from the dominant trees by use of an increment core taken at 30 cm from ground level. To estimate the total age, the number of years that a seedling takes to reach this height (usually 2-5 years in the region) was added to the number of rings counted. Site index (*S*, m) at a reference age of 35 years was estimated using the model proposed by Marques (1991). The diameter, height and stand variables are summarized in Tab. 1.

Basic models

The 27 models proposed by Huang et al. (2000) were evaluated in this study. The best locally fitted model for each plot-inventory combination was selected on the basis of the following criteria: (i) statistical significance of the parameters; (ii) visual inspection of the residuals; and (iii) goodness-of-fit statistics, which were averaged by model. The basic model with the best average goodness-of-fit statistics was again locally fitted by using different sampling designs (those evaluated for calibration of the basic mixed model, as explained below) for comparison with the other models and approaches tested.

Generalized models

Graphical analyses were performed to establish which stand variables were most closely related to the estimated local parameters of the best basic model, and to

Tab. 1 - Summary statistics for the fitting data (10580 *h-d* observations in 387 plot-inventory combinations). (d): diameter at breast height (1.3 m above ground level) outside bark; (*h*): total tree height; (*t*): stand age; (*N*): number of trees per hectare; (*dg*): quadratic mean diameter; (*G*): stand basal area; (*hm*): mean height; (*dd*): dominant diameter, defined as the mean diameter of the 100 largest-diameter trees per hectare; (*hd*): dominant height, defined as the mean height of the 100 largest-diameter trees per hectare; (*S*): Site index, at a reference age of 35 years.

Variable	Mean	Min.	Max.	Std. dev.
d (cm)	21.4	0.7	53.8	8.7
h (m)	14.1	1.7	29.2	4.4
t (years)	40.3	12	70	12.5
N (trees ha^{-1})	1055	120	7680	913
dg (cm)	23.1	5.2	43.6	7.7
G (m^2 ha^{-1})	32.1	6.0	57.3	9.9
hm (m)	14.8	4.8	26.7	4.3
dd (cm)	29.7	8.7	50.3	7.8
hd (m)	16.1	5.6	27.2	4.3
S (m)	15.9	9.7	30.4	2.6

identify the type of relationship (*e.g.*, linear, allometric, exponential). In addition, several generalized models selected from the literature (Krumland & Wensel 1988, Tomé 1988, Almeida 1999, López Sánchez et al. 2003, Sharma & Zhang 2004, Castedo Dorado et al. 2006, Sharma & Parton 2007, Crecente-Campo et al. 2010) were fitted to the dataset. These models include the most flexible equations for developing *h-d* relationships (*i.e.*, the Bertalanffy-Richards, Weibull and Schnute models), which have been demonstrated as suitable for several species and types of stands. Some modifications were also tested, *e.g.*, replacing the quadratic mean diameter and the mean height by dominant diameter and dominant height, respectively.

Mixed models

Mixed models can be applied to basic models (Trincado et al. 2007), thus generating a basic mixed model, and to generalized models (Calama & Montero 2004, Castedo Dorado et al. 2006), thus generating a generalized mixed model. The data used in this study comprised a sample of heights and diameters from different plots measured a maximum of four times. Thus, two hierarchical levels (trees in plots) and repeated measurements (several inventories of each tree) can be taken into account. Random effects for the tree measurement occasion were not included in this study, in accordance with Castedo Dorado et al. (2006). This is not likely to have a notable effect on fitted mixed models because the number of measurement-occasions for one tree was small in comparison with the number of trees analyzed (10580).

Determining which parameters should be considered fixed and which should be considered mixed (also including random effects) is a key question when developing mixed models. Some authors (Pinheiro & Bates 1998, Fang & Bailey 2001) suggest that all parameters in the model should be considered mixed if convergence is possible. Therefore, all parameters were first expanded with random effects in the selected basic and generalized models.

In this study, to evaluate the calibrated response the random effects were estimated using an approximate Bayesian estimator (Vonesh & Chinchilli 1997 – eqn. 2):

$$\hat{\mathbf{u}}_i = \hat{\mathbf{D}}\mathbf{Z}_i^T\left(\mathbf{Z}_i\hat{\mathbf{D}}\mathbf{Z}_i^T+\hat{\mathbf{R}}_i\right)^{-1}\hat{\mathbf{e}}_k =$$
$$= \hat{\mathbf{D}}\mathbf{Z}_i^T\left(\mathbf{Z}_i\hat{\mathbf{D}}\mathbf{Z}_i^T+\hat{\mathbf{R}}_i\right)^{-1}\left[\mathbf{h}_i-f\left(\mathbf{x}_i,\hat{\boldsymbol{\beta}},0\right)\right]$$

where $\hat{\mathbf{D}}$ is the estimated variance-covariance matrix for the random effects $\mathbf{u}_i$, $\hat{\mathbf{R}}_i$ is the estimated variance-covariance matrix for the error term, $\mathbf{Z}_i$ is the partial derivatives matrix with respect to random effects, and $\hat{\mathbf{e}}_k$ is the error matrix estimated using only the fixed parameters. When the random effects $\mathbf{u}_i$ are introduced linearly and when the nonlinear function is linearized by the zero expansion method, the partial derivatives of the $\mathbf{Z}_i$ matrix are equivalent to the partial derivatives with respect to the fixed parameters.

The calibrated response was then calculated as (eqn. 3):

$$\hat{\mathbf{h}}_i = f\left(\mathbf{x}_i,\hat{\boldsymbol{\beta}},0\right)+\mathbf{Z}_i\hat{\mathbf{u}}_i$$

The calibrated response was evaluated for different height sampling designs:
(i) Total height of randomly selected trees per plot-inventory. The mean and extreme values of the goodness-of-fit statistics after 100 simulations were obtained.
(ii) Total height of quantile trees of the diameter distribution per plot-inventory (*i.e.*, 1 tree = median-diameter tree; 2 trees = tercile-diameter trees; 3 trees = quartile-diameter trees; 4 trees = quintile-diameter trees, etc.).
(iii) Total height of the smallest diameter trees per plot-inventory (this approach was only considered for the selected generalized mixed model).

Model fitting and comparison

The basic and generalized models were fitted with the NLIN procedure, while the mixed models were fitted with the NLMIXED procedure, both of SAS/STAT® (SAS Institute Inc 2009). Statistical and graphical analyses were used to compare model performance. Three statistical criteria obtained from the residuals were examined: the coefficient of determination (R^2), which indicates the proportion of the total variance that is explained by the model, the root mean square error (RMSE), and the Schwarz's Bayesian Information Criteria (BIC – Schwarz 1978).

Results

Basic models

Fitting statistics showed that the model evaluated by Burkhart & Strub (1974) performed best (R^2 = 0.995, RMSE = 1.027 m – eqn. 4):

$$h = 1.3+b_0\exp\left(\frac{b_1}{d}\right)$$

where *h* represents total tree height (m), *d* is the diameter (cm) at breast height (1.3 m above ground level), and b_0 and b_1 are parameters to be estimated for each plot-inventory combination.

Model from eqn. 4 was also fitted locally using the different height sampling designs mentioned in the mixed models section. The use of quantile-diameter trees yielded a lower error than the average error obtained in the fitting with randomly selected trees up to five trees. When six or more trees were selected in both alternatives, the differences were not significant.

Generalized models

A modification of the model used by Tomé (1988), in which the age term was omitted, yielded the best fit to the dataset (R^2 = 0.933, RMSE = 1.149 m). The genera-
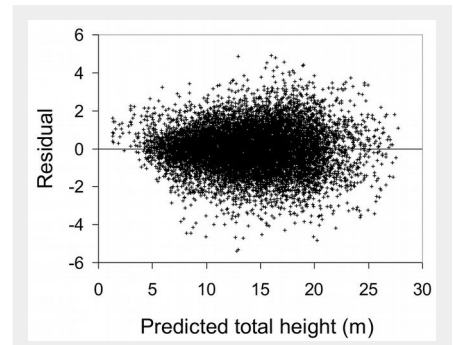
Fig. 1 - Plot of residuals against predicted values of height for eqn. 5 (modified model of Tomé 1988).

lization of the best basic model (eqn. 4) yielded poorer results than the modified model of Tomé, and, although the quality of fit in the former model (eqn. 1 - Almeida 1999) was greatly improved by enlarging the dataset (with R^2 from 0.805 to 0.929 and RMSE from 1.660 to 1.186m), it was poorer than the fit provided by the selected model. Therefore, the proposed generalized equation is as follows (eqn. 5):

$$h = 1.3+\left(hd-1.3\right)\cdot$$
$$\cdot \exp\left[\left(b_0-b_1\cdot hd+b_2\frac{N}{1000}\right)\cdot\left(\frac{1}{d}-\frac{1}{dd}\right)\right]$$

where b_0 = -3.861, b_1 = 0.2542 and b_2 = 0.5326. All parameters from eqn. 5 were significant at the 5% level. The plot of residuals against predicted heights for the fitting data set did not show clear evidence of variation in residuals over the full range of predicted values (Fig. 1). The plot of residuals against lagged residuals within each plot-inventory combination (with the data ordered by ascending diameter) did not show any correlated errors (Fig. 2). Moreover, the predicted *h-d* curves (Fig. 3) showed appropriate trends and logical asymptotes.
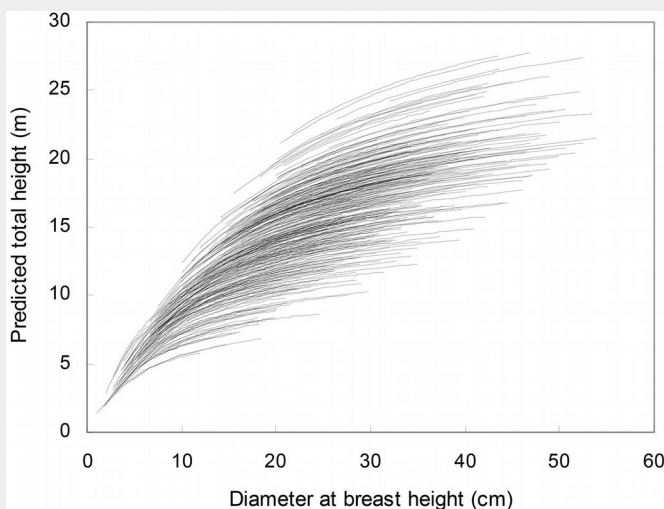
Mixed models

The model of Burkhart & Strub (eqn. 4) was also selected to develop a basic mixed
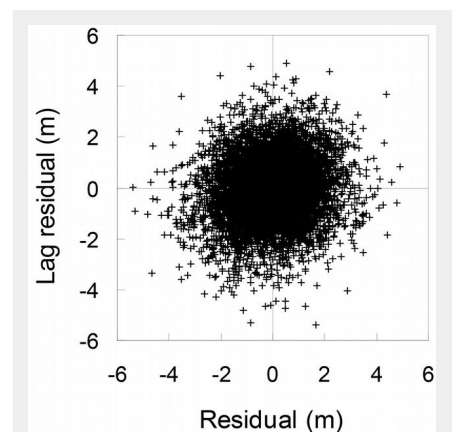
Fig. 2 - Plot of residuals against lagged residuals of height for eqn. 5.

Fig. 3 - Predicted height-diameter curves obtained using eqn. 5 and the 387 plot-inventory combinations from this study.

model. Both parameters of the model were linearly expanded with random effects as follows (eqn. 6):

$$h = 1.3 + (b_0 + u_1) \exp\left(\frac{b_1 + u_2}{d}\right)$$

Estimates of the fixed parameters and variance components for the random effects for eqn. 6 are shown in Tab. 2. All estimates were significant at the 5% level. In the calibration process, the use of quantile-diameter trees was preferable to the selection of random trees when only one tree is considered, but there were no differences in the other cases (note that the errors of randomly selected trees are average values).
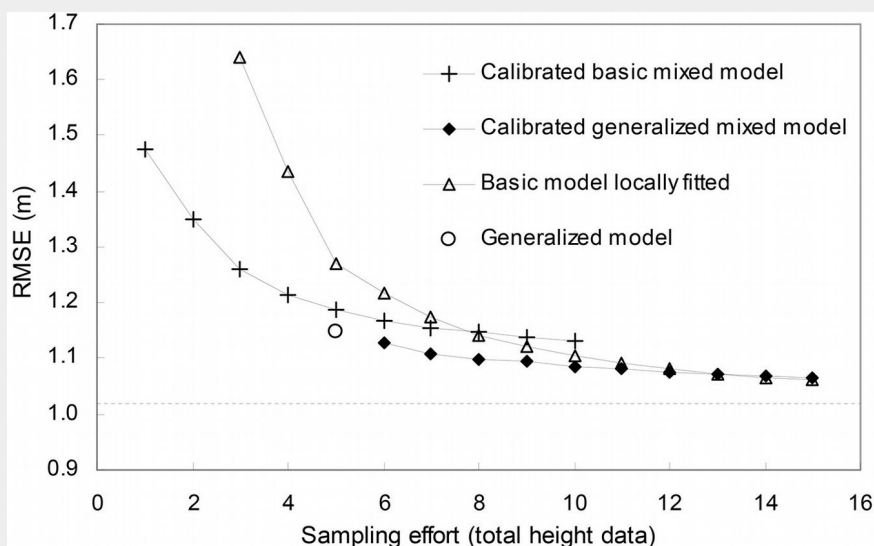
The selected generalized model (eqn. 5) was also used to develop a generalized mixed model. Convergence was not achieved when all parameters were considered as mixed and therefore all possible expansions of two fixed parameters with random effects were tested. The inclusion of random effects in the fixed parameters b_0 and b_2 generated the best results, producing the following model (eqn. 7):

$$h = 1.3 + (hd - 1.3) \cdot$$
$$\exp\left[\left(b_0 + u_1 + b_1 hd + (b_2 + u_2)\frac{N}{1000}\right)\left(\frac{1}{d} - \frac{1}{dd}\right)\right]$$

Estimates for the fixed parameters and variance components for the random effects for eqn. 7 are shown in Tab. 3. All estimates were significant at the 5% level.

Calibration of the generalized mixed model using the total height of the smallest diameter trees per plot-inventory yielded better results than calibration with other designs for the same sampling effort. For comparison of eqn. 7 with the other alternatives, five additional trees should be measured to compute dominant height in this study (the size of the plots was 500 m²).

The relation between RMSE and sampling effort (number of measurements of total height) for the selected model forms and

Tab. 2 - Fixed parameter estimates, approximate significance tests and variance components for eqn. 6.

Fixed Effect: Parameter	Estimate	Approx. std. error	t-value	Approx. p-value
b_0	19.04	0.2889	65.91	< 0.001
b_1	-7.352	0.1452	-50.62	< 0.001
Random Effect: Variance component	**Estimate**	**Approx. std. error**	**Z-value**	**Approx. Z-value**
$var(u_1)$	30.18	2.326	12.98	< 0.001
$var(u_2)$	5.111	0.5437	9.401	< 0.001
$cov(u_1, u_2)$	-9.378	0.9959	-9.416	< 0.001
σ^2(error variance)	1.140	0.01627	70.06	< 0.001

Tab. 3 - Fixed parameter estimates, approximate significance tests and variance components for eqn. 7.

Fixed Effect: Parameter	Estimate	Approx. std. error	t-value	Approx. p-value
b_0	-4.535	0.5459	-8.307	< 0.001
b_1	-0.2306	0.03012	-7.656	< 0.001
b_2	0.7753	0.08658	8.955	< 0.001
Random Effect: Variance component	**Estimate**	**Approx. std. error**	**Z-value**	**Approx. Z-value**
$var(u_1)$	8.870	0.4068	21.80	< 0.001
$var(u_2)$	0.6819	0.03157	21.60	< 0.001
$cov(u_1, u_2)$	-2.456	0.07077	-34.70	< 0.001
σ^2(error variance)	1.150	0.01609	71.43	< 0.001

Fig. 4 - Changes in RMSE with sampling effort for the different model forms and approaches tested in this study. The generalized mixed model was calibrated with data from the smallest diameter trees, and for calibration of the basic mixed model and local fit of the basic model were selected data from quantile-diameter trees. The horizontal dashed line represents the RMSE obtained in a local fit using all data and the selected basic model.

Tab. 4 - Model approach and sampling design yielding the lowest error in relation to sampling effort (number of measurements of total height). (*d*): diameter (cm) at breast height (1.3 m above ground level).

Sampling effort	Model approach	Sampling design	RMSE (m)
1	Calibrated basic mixed model	Median *d* tree	1.475
2	Calibrated basic mixed model	Quantile *d* or randomly selected trees	1.348
3	Calibrated basic mixed model	Quantile *d* or randomly selected trees	1.259
4	Calibrated basic mixed model	Quantile *d* or randomly selected trees	1.213
5	Generalized model	Dominant trees	1.149
6 to 13	Calibrated generalized mixed model	Dominant trees + smallest *d* trees	1.130 to 1.072
14 and so on	Basic model locally fitted	Quantile *d* or randomly selected trees	≤ 1.067

approaches is shown in Fig. 4. In each case, the option that generated the smallest RMSE is shown (*i.e.*, the locally fitted basic model and calibration of the basic mixed model with quantile-diameter trees, and calibration of the generalized mixed model with the smallest diameter trees). In summary, the combination of model approach and sampling design yielding the lowest RMSE in relation to sampling effort is shown in Tab. 4.

Discussion

Several model forms and approaches for developing an *h-d* model for *P. pinaster* in northern Portugal were compared. In the first step, 27 basic models (*h* as a function of just *d*) were locally fitted. Height predictions with the selected basic model (eqn. 4) yielded a RMSE value of 1.027 m, which constitutes a reference minimum error for comparison with the other model approaches analyzed.

The selected generalized model (eqn. 5) includes the variable dominant height in its formulation. The use of this covariate may be preferred as it requires less sampling effort than accurate measurement of the mean height (López Sánchez et al. 2003). Another advantage of this model is that it predicts a height equal to *hd* when the independent variable is a diameter equal to *dd,* and it predicts a height equal to 1.3 m when diameter is equal to 0. The model also includes a measure of stand density (*i.e.*, trees per hectare). Stand density is the most obvious factor affecting the *h-d* relationship in a stand (Zhang et al. 1997, Zeide & Vanderschaaf 2002): trees of the same height usually have smaller diameters in denser stands. Several variables have been proposed as additional predictor variables: stand age (Curtis 1967, Soares & Tomé 2002, López Sánchez et al. 2003), crown competition factor (Temesgen et al. 2007), geographical features (Hynynen et al. 2002, Russell et al. 2010, Schmidt et al. 2011) or wind speed (Meng et al. 2008). Incorporation of additional predictor variables can improve the predictions, but it requires a greater sampling effort and limits the application of the model.

The reduction in RMSE with calibration was more evident when the mixed model was based on a basic model rather than on a generalized model (Fig. 4), because most of the variability was accounted for by the stand level variables included in the generalized model. According to Trincado et al. (2007), the use of a basic mixed model in forest inventories by selecting a sub-sample of trees for height measurement enables the maintenance of a simple model structure without inclusion of additional predictor variables. The calibration of a basic mixed model with one tree (as done by Trincado et al. 2007) did not yield very good results in this study, but it could be used in the absence of more observations. On the other hand, local fitting using the model proposed by Burkhart & Strub (1974) is not possible with fewer than two observations because the model has two parameters. Although calibration using the total height of the smallest diameter trees per plot-inventory is not recommended for the basic mixed model, this calibration option produced the best results in relation to sampling effort when used with the generalized mixed model. This calibration method was also selected for other fast-growing species in the Iberian peninsula (Castedo Dorado et al. 2006, Crecente-Campo et al. 2010) with generalized mixed models that also included dominant height and dominant diameter as stand-predictor variables. This can be explained by the fact that these models are restricted to passing through the point *hd-dd*, implying that the models will behave as quite invariant in the upper part of the *h-d* relationship, and that small trees will provide much more information for calibration in this case (Crecente-Campo et al. 2010).

The use of the selected generalized model (eqn. 5) generated less error (RMSE) than the other options for the same sampling effort with five trees (Fig. 4 - note that in this study five trees were selected for estimating dominant height). Moreover, this equation has the advantage that it can be used for prediction purposes with data commonly measured in current forest inventories in Portugal, and it can easily be implemented in a statistical disaggregated dynamic growth model (Diéguez-Aranda et al. 2006, Castedo Dorado et al. 2007) based on the state-space approach (García 1994), in which the vector of state variables includes dominant height, as in MODISPINASTER.

Numerous studies have shown that in the development of height-diameter models, the best approach depends on: (1) the data available; (2) the intended accuracy; and (3) the interest (or otherwise) in obtaining additional information. From a practical point of view (related to sampling effort, accuracy and practical applications), the use of the selected generalized model is recommended in preference to the other approaches tested. The dominant height (*hd*) must be known when using this model, and therefore it should be considered during the data collection process. If dominant height is not known and the total height of randomly or quantile-diameter selected trees is available, calibration of the basic mixed model rather than a local fit may be recommended (Fig. 4), specifically for a short range (up to 7-8) of *h-d* observations. The inclusion of random effects accounts for the natural variability in height at the stand level or in different time series for the same stand. However, as the number of observations increases (above 8 according to the results from the real case study), a local model will provide more accurate results than a calibrated basic mixed model.

Acknowledgements

References

AFN (2010). 5° Inventário Florestal Nacional. Apresentação do Relatório Final [5th National Forest Inventory. Presentation of the Final Report]. National Forestry Authority, National Coordination of Forest Management, Lisbon, Portugal, pp. 14. [in Portuguese] [online] URL: http://www.icnf.pt/portal/florestas/ifn/resource /ficheiros/ifn/Apresenta-IFN5-AFN-DNGF-JP.pdf

Almeida LFR (1999). Comparação de metodologias para estimação de altura e volume em povoamentos de pinheiro bravo no Vale do Tâmega [Assessment of maritime pine forest stands height and volume estimation methodologies in Tâmega valley]. Graduate thesis,

UTAD, Vila Real, Portugal, pp. 109. [in Portuguese]

Burkhart HE, Strub MR (1974). A model for simulation of planted loblolly pine stands. In: "Growth models for tree and stand simulation" (Fries J ed). Royal College of Forestry, Stockholm, Sweden, pp. 128-135.

Calama R, Montero G (2004). Interregional nonlinear height-diameter model with random coefficients for stone pine in Spain. Canadian Journal of Forest Research 34: 150-163. - doi: 10.1139/x03-199

Castedo Dorado F, Diéguez-Aranda U, Barrio Anta M, Sanchez Rodríguez M, Von Gadow K (2006). A generalized height-diameter model including random components for radiata pine plantations in northwestern Spain. Forest Ecology and Management 229: 202-213. - doi: 10.1016/j.foreco.2006.04.028

Castedo Dorado F, Diéguez-Aranda U, Álvarez González JG (2007). A growth model for *Pinus radiata* D. Don stands in north-western Spain. Annals of Forest Science 64: 453-465. - doi: 10.1051/forest:2007023

Crecente-Campo F, Tomé M, Soares P, Diéguez-Aranda U (2010). A generalized nonlinear mixed effects height-diameter model for *Eucalyptus globulus* L. in northwestern Spain. Forest Ecology and Management 259: 943-952. - doi: 10.1016/j.foreco.2009.11.036

Curtis RO (1967). Height-diameter and height-diameter-age equations for second-growth Douglas-fir. Forest Science 13: 365-375. [online] URL: http://www.ingentaconnect.com/content/saf/fs/1967/00000013/00000004/art00007

Davidian M, Giltinan D (1995). Nonlinear models for repeated measurement data. Chapman and Hall, New York, USA, pp. 360. [online] URL: http://books.google.com/books?id=0eSIBPAL4qsC

De-Miguel S, Mehtälo L, Shater Z, Kraid B, Pukkala T (2012). Evaluating marginal and conditional predictions of taper models in the absence of calibration data. Canadian Journal of Forest Research 42: 1383-394. - doi: 10.1139/x2012-090

Diéguez-Aranda U, Castedo-Dorado F, Álvarez-González JG, Rodríguez-Soalleiro R (2006). Dynamic growth model for Scots pine (*Pinus sylvestris* L.) plantations in Galicia (north-western Spain). Ecological Modelling 191: 225-242. - doi: 10.1016/j.ecolmodel.2005.04.026

Fang Z, Bailey RL (2001). Nonlinear mixed effects modeling for slash pine dominant height growth following intensive silvicultural treatments. Forest Science 47: 287-300. [online] URL: http://www.ingentaconnect.com/content/saf/fs/2001/00000047/00000003/art00002

Fonseca TF (2004). Modelação do crescimento, mortalidade e distribuição diamétrica, do pinhal bravo no Vale do Tâmega [Modelling of growth, mortality and diameter distribution of maritime pine in the Valley of Tâmega]. PhD thesis, UTAD, Vila Real, Portugal, pp. 248. [in Portuguese]

Fonseca TF, Parresol B, Marques C, De Coligny F (2012). Models to implement a sustainable forest management - an overview of the Modis-Pinaster model. In: "Sustainable Forest Management - Current Research" (Garcia JM, Díez Casero JJ eds). InTech, Rijeka, Croatia, pp. 321-338. [online] URL: http://www.intechopen.com/books/sustainable-forest-management-current-research/

García O (1994). The state-space approach in growth modelling. Canadian Journal of Forest Research 24: 1894-1903. - doi: 10.1139/x94-244

Gregoire TG, Schabenberger O, Barret J (1995). Linear modelling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. Canadian Journal of Forest Research 25: 137-156. - doi: 10.1139/x95-017

Huang S, Titus SJ, Wiens DP (1992). Comparison of nonlinear height-diameter functions for major Alberta tree species. Canadian Journal of Forest Research 22: 1297-1304. - doi: 10.1139/x92-172

Huang S, Price D, Titus SJ (2000). Development of ecoregion-based height-diameter models for white spruce in boreal forests. Forest Ecology and Management 129: 125-141. - doi: 10.1016/S0378-1127(99)00151-6

Huang S, Meng SX, Yang Y (2009). Prediction implications of nonlinear mixed-effects forest biometric models estimated with a generalized error structure. In: Proceedings of the "Joint Statistical Meetings, Section on Statistics and the Environment". Washington (DC, USA), 1-6 Aug 2009. American Statistical Association, Alexandria, Virginia, USA, pp. 1174-1188.

Hynynen J, Ojansuu R, Hökkä H, Siipilehto J, Salminen H, Haapala P (2002). Models for predicting stand development in MELA system. Finnish Forest Research Institute, Research Paper 835, Helsinki, Finland, pp. 116.

Krumland BE, Wensel LC (1988). A generalized height-diameter equation for coastal California species. Western Journal of Applied Forestry 3: 113-115. [online] URL: http://www.ingentaconnect.com/content/saf/wjaf/1988/00000003/00000004/art00007

Lappi J (1997). A longitudinal analysis of height/diameter curves. Forest Science 43: 555-570. [online] URL: http://www.ingentaconnect.com/content/saf/fs/1997/00000043/00000004/art00013

Lindstrom MJ, Bates DM (1990). Nonlinear mixed-effects models for repeated measures data. Biometrics 46: 673-687. - doi: 10.2307/2532087

Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O (2006). SAS® for mixed models (2nd edn). SAS Institute Inc., Cary, NC, USA, pp. 840.

Loetsch F, Zöhrer F, Haller KE (1973). Forest Inventory (vol II). BLV Verlagsgesellschaft, München, Germany, pp. 469.

López Sánchez CA, Gorgoso JJ, Castedo Dorado F, Rojo A, Rodríguez R, Álvarez González JG, Sánchez F (2003). A height-diameter model for *Pinus radiata* D. Don in Galicia (northwest Spain). Annals of Forest Science 60: 237-245. - doi: 10.1051/forest:2003015

Marques CP (1991). Evaluating site quality of even-aged maritime pine stands in northern Portugal using direct and indirect methods. Forest Ecology and Management 41: 193-204. - doi: 10.1016/0378-1127(91)90103-3

Meng SX, Huang S, Lieffers VJ, Nunifu T, Yang Y (2008). Wind speed and crown class influence the height-diameter relationship of lodgepole pine: nonlinear mixed effects modeling. Forest Ecology and Management 256: 570-577. - doi: 10.1016/j.foreco.2008.05.002

Parresol BR, Fonseca TF, Marques CP (2010). Numerical details and SAS programs for parameter recovery of the SB distribution. Gen. Tech. Rep. SRS-122, Southern Research Station, USDA Forest Service, Asheville, NC, USA, pp. 27.

Peng C (2001). Developing and validating nonlinear height-diameter models for major tree species of Ontario's boreal forest. Northern Journal of Applied Forestry 18: 87-94. [online] URL: http://www.ingentaconnect.com/content/saf/njaf/2001/00000018/00000003/art00004

Pinheiro JC, Bates DM (1998). Model building for nonlinear mixed effects model. Department of Statistics, University of Wisconsin, Madison, WI, USA, pp. 11.

Pretzsch H (2010). Forest dynamics, growth and yield. Springer-Verlag, Berlin, Germany, pp. 664.

Russell MB, Amateis RL, Burkhart HE (2010). Implementing regional locale and thinning response in the loblolly pine height-diameter relationship. Southern Journal of Applied Forestry 34: 21-27. [online] URL: http://www.ingentaconnect.com/content/saf/sjaf/2010/00000034/00000001/art00004

SAS Institute Inc (2009). SAS/STAT® 9.2 User's Guide (2nd edn). SAS Institute Inc., Cary, NC, USA, pp. 7869.

Schmidt M, Kiviste A, Gadow Kv (2011). A spatially explicit height-diameter model for Scots pine in Estonia. European Journal of Forest Research 130: 303-315. - doi: 10.1007/s10342-010-0434-8

Schwarz G (1978). Estimating the dimension of a model. Annals of Statistics 6 (2): 461-464. - doi: 10.1214/aos/1176344136

Searle SR, Cassela G, McCulloch CE (1992). Variance components. Wiley, New York, USA, pp. 501.

Sharma M, Parton J (2007). Height-diameter equations for boreal tree species in Ontario using a mixed-effects modeling approach. Forest Ecology and Management 249: 187-198. - doi: 10.1016/j.foreco.2007.05.006

Sharma M, Zhang SY (2004). Height-diameter models using stand characteristics for *Pinus banksiana* and *Picea mariana*. Scandinavian Journal of Forest Research 19: 442-451. - doi: 10.1080/02827580410030163

Soares P, Tomé M (2002). Height-diameter equation for first rotation eucalypt plantations in Portugal. Forest Ecology and Management 166: 99-109. - doi: 10.1016/S0378-1127(01)00674-0

Temesgen H, Hann DW, Monleon VJ (2007). Regional height-diameter equations for major tree species of southwest Oregon. Western Journal of Applied Forestry 22: 213-219. [online] URL: http://www.ingentaconnect.com/content/saf/wjaf/2007/00000022/00000003/art00009

Tomé M (1988). Modelação do crescimento da árvore individual em povoamentos de Eucalyptus globulus Labill. (1ª rotação) na região centro de Portugal [Modelling of the individual tree growth in Eucalyptus globulus Labill. (1st rotation) in the central region of Portugal]. PhD thesis, Instituto Superior de Agronomía, Universidade Técnica de Lisboa, Lisbon, Portugal, pp. 256. [in Portuguese]

Trincado G, Curtis LV, Burkhart HE (2007). Regio-

nal mixed-effects height-diameter models for loblolly pine (*Pinus taeda* L.) plantations. European Journal of Forest Research 126: 253-262. - doi: 10.1007/s10342-006-0141-7

Vonesh EF, Chinchilli VM (1997). Linear and non-linear models for the analysis of repeated measurements. Marcel Dekker Inc, New York, USA, pp. 560.

Zeide B, Vanderschaaf C (2002). The effect of density on the height-diameter relationship. In: Proceedings of the "11th Biennial Southern Silvicultural Research Conference" (Outcalt KW ed). Knoxville (TN, USA), 20-22 March 2001. Gen. Tech. Rep. SRS-48, Southern Research Station, USDA Forest Service, Asheville, NC, USA, pp. 463-466. [online] URL: http://www.tree search.fs.fed.us/pubs/viewpub.jsp?index=4900

Zhang SA, Burkhart HE, Amateis RL (1997). The influence of thinning on tree height and diameter relationships in loblolly pine plantations. Southern Journal of Applied Forestry 21: 199-205. [online] URL: http://www.ingentaconne ct.com/content/saf/sjaf/1997/00000021/000000 04/art00010