ELSEVIER

Forest Ecology and Management 166 (2002) 99–109

Forest Ecology and Management

www.elsevier.com/locate/foreco

Height–diameter equation for first rotation eucalypt plantations in Portugal

Paula Soares*, Margarida Tomé

Department of Forestry, Instituto Superior de Agronomia, Universidade Tecnica de Lisboa, Tapada da Ajuda, 1349-017 Lisbon, Portugal

Received 22 June 2000; received in revised form 23 April 2001; accepted 31 May 2001

Abstract

A height–diameter equation for eucalypt plantations was developed based on a tree dataset from trials and permanent plots located in the north and central coastal regions of Portugal. The total dataset was split into two datasets through restricted random sampling at the plot level. The equations selected in one data subset were evaluated with the other subset and vice versa. Harrison equation, fitted with the iteratively reweighted least squares method, in both versions—restricted and not restricted to pass through the point diameter–height (0, 1.30)—was selected. The first version was recommended for young plantations; it is age dependent and requires a measure of stand productivity. The second version was appropriate to use in commercial forest inventory where trees smaller than 4 cm diameter are not measured; it is age independent, density dependent and, also requires a measure of stand productivity.

Keywords: Tree height prediction; Height–diameter equation; Plantations; *Eucalyptus globulus* Labill.

1. Height–diameter equations

The modelling of tree height growth can be one component of an individual tree model. However, the lack of successive height measurements of the same tree is commonly forcing the use of compatible height projection equations and height–diameter prediction equations (e.g. Lynch and Murphy, 1995). The usefulness of the last equation in forest inventories is also generally recognised; height–diameter equations are needed to obtain total and merchantable tree volumes. When combined with crown ratio data or models, height–diameter equations can be used to predict tree height and to estimate the change in crown ratio (e.g. Maguire and Hann, 1990). Height–diameter equations are also used in many stand growth and yield models to predict the mean height for a given diameter or diameter class (e.g. Lenhart and Clutter, 1971).

Height–diameter equations can be of local application or can have a more generalised use (Tomé, 1988). The first type is normally only dependent on tree diameter and can be applied to the stand where the data were gathered; regional height–diameter equations are a function of tree diameter, age, and other stand variables and can be applied at the regional level.

The height–diameter relationship has been expressed by a wide variety of equations (Table 1). Curtis (1967) presented an exhaustive list of the most common local equations. More recently, Lynch and Murphy (1995) presented a detailed discussion of previous work related to height–diameter–age curves.

Staebler (1954) and Curtis (1967) equations are essentially empirical, Staebler equation has a maximum at $d = -a_1/2a_2$, where $a_2 < 0$; Curtis equation

* Corresponding author. Tel.: +351-21-3653367; fax: +351-21-3645000.
E-mail address: paulasoares@isa.utl.pt (P. Soares).

0378-1127/02/$ – see front matter
PII: S0378-1127(01)00674-0

Table 1
Equations used to model the tree height–diameter relationship[a]

Equation	Author
Local height–diameter equations	
$h = a_0 + a_1 d + a_2 d^2$	Staebler (1954)
$h = A(1 - e^{-a_0 d})$	Meyer (1940)
$\ln h = a_0 + a_1 1/d$	Michailoff (1943)
$h = a_0 + a_1 \ln d$	Henriksen (1950)
$\ln h = a_0 + a_1 \ln d$	Stoffels and van Soest (1953)
$h = d/(a_0 + a_1 d)$	Prodan (1965)
$h = a_0 + a_1 1/d + a_2 1/d^2$	Curtis (1967)
$h - h_0 = h_d(1 + a_0 e^{a_1 h_d})(1 - e^{-a_2 d/h_d})$	Harrison et al. (1986)
Regional height–diameter equations	
$\ln h = a_0 + a_1 S_{h,t} + a_2 N/100 + a_3 1/t + a_4 1/d$	Bennett and Clutter (1968)
$\ln(h_d/h) = a_0 + (1/d - 1/d_{max})(a_1 + a_2 \ln N + a_3 1/t + a_4 \ln h_d)$	Lenhart (1968)
$\ln h = a_0 + a_1 \ln h_d + a_2 1/t + a_3 \ln N/d + a_4 1/(dt) + a_5 1/d$	Burkhart and Strub (1974)
$h = h_d\, e^{a_0(1/d - 1/d_d)}, \quad a_0 = f(\text{stand parameters})$	Michailoff modified by Tomé (1988)
$h = h_d(d/d_d)^{a_0}, \quad a_0 = f(\text{stand parameters})$	Stoffels and van Soest modified by Tomé (1988)
$h = h_d(1 + a_0 h_d(1/d - 1/d_d))^{-1}, \quad a_0 = f(\text{stand parameters})$	Prodan modified by Tomé (1988)

[a] h: total tree height; d: diameter at breast height; A: height asymptote; $S_{h,t}$: site index; N: stand density; t: age; d_{max}: maximum tree diameter; h_d: dominant height; d_d: dominant diameter; h_0: reference height $-$ 1.30 m to restrict the origin of the equation at the point $(d, h) = (0, 1.30)$; a_0, a_1, a_2: equation parameters.

presents an asymptote but, since both equations are not constrained to pass through the origin, may give negative estimates for small trees. Henriksen (1950) equation is an increasing function for values of $a_1 > 0$, but it is not an asymptotic function, however, it presents a biologically appropriate shape—height increment per diameter unit decreases with increasing diameter. Stoffels and van Soest (1953) equation corresponds, in the non-linear form, to a linear alometric relationship between diameter and height; like Henriksen equation it is not an asymptotic function; a_1 must be between 0 and 1 for the function to assume a biologically appropriate shape. Michailoff (1943) equation presents an asymptote (e^{a_0}), an inflexion point, and passes through the origin; after the inflection point it presents an adequate biological behaviour when $a_1 < 0$. Meyer (1940) equation is similar to Michailoff equation, but more flexible; the height increment is a_0 dependent. Prodan (1965) equation corresponds to the hyperbolic formulation; it has $1/a_1$ as asymptote and, since $a_0 > 0$, it presents an adequate biological behaviour.

Mendes (1989) fitted the local equations presented in Table 1 to data of eucalypt permanent plots located in the centre coastal area of Portugal and confirmed the adequacy of Michailoff (1943), Stoffels and van Soest (1953), and Prodan (1965) equations.

The first three regional height–diameter equations are empirical equations with a level of complexity that may induce the presence of colinearity. Harrison et al. (1986) equation was deduced based on the interpretation of the parameters of Meyer (1940) equation. The formulation guarantees that the asymptote is near to the dominant height (can be superior) and the height growth rate is smaller for the greatest dominant heights. Tomé (1988) deduced the three last regional equations presented in Table 1 by restriction of some local height–diameter equations to the point (dominant diameter, dominant height). The resulting equations depend on only one parameter that was expressed by the author as a linear combination of stand parameters.

Tomé (1988) and Ribeiro (1998) presented regional height–diameter equations for eucalypt in Portugal. However, the equation developed by Tomé was based on a restricted dataset characterised by small variation in planting density and site index; Ribeiro (1998) developed an equation for a specific Portuguese pulp company. The purpose of the present work was to develop a regional height–diameter equation, to be used as a component of a tree model—GlobTree, for first cutting cycle *Eucalyptus globulus* Labill. plantations located in the north and central coastal regions of Portugal (Soares, 1999).

2. Data

Eucalyptus globulus Labill. is one of the most important economic forest species in Portugal, occupying an area of 676,500 ha in a total forest area of 3,275,300 ha (DGF, 1999). It is a fast growing species mainly used by the pulp industry; the trees are planted at final density—thinning and pruning practices are unusual in first rotation stands. The stands are intensively managed in a short rotation coppice system in which the first cycle of planted seedlings (single stem) is followed by two or three coppiced stands, with an average cutting cycle of 10–12 years.

Data from permanent plots, five spacing trials, and a fertilised and irrigated experiment of eucalypt were used (Table 2). The dataset includes 10, 16, 36, 10 and 27 plots, respectively, of the Quinta Paço (QP), Vale Bezerra (VB), Vilar Luz (VL), Alto Vilão (AV) and Seixosa (SX) spacing trials (Soares, 1999), 2 control plots and 2 fertilised plots of the Furadouro fertilised and irrigated trial (Pereira et al., 1989), and 52 permanent plots. The principal criterion for the selection of these plots was the availability at the tree level of pairs of measurements (diameter, height). Most of the plots were measured annually; on young spacing trials, trees were measured every 3 months. On the fertilised and irrigated trial, trees were measured initially monthly, later twice per year and, at the end, annually, however, only measurements obtained on annual periods were considered for this study. Data from stands older than 15 years were obtained on permanent plots and the AV and SX spacing trials; data from stands younger than 4 years were from spacing trials (QP, VB and VL) and the fertilised and irrigated trial.

Total heights were gathered, on young stands, with a telescoping measuring rod and, on mature stands, with a hypsometer; in one of the permanent plots, heights were obtained on felled trees. Border trees, trees without simultaneous measurements of total height and diameter, and trees with height and/or diameter imperfections were eliminated from the available measurements at individual tree level.

Table 2
Characterisation of the plots used in the definition of the height–diameter equation[a]

	Plot characteristics				Tree variable		
	Plot area (m^2)	Spacing (m × m)	Site index	Number of measurements	Age (years)	*d* (cm)	*h* (m)
QP	648–2916	2 × 1–3 × 3	25.7–28.4	6	2.6	0.3	1.7
					4.4	10.1	13.7
					7.6	23.9	26.8
VB	765–2487	1 × 3.5–4 × 4	16.9–20.7	4	1.5	0.2	1.3
					3.0	6.0	6.8
					4.5	21.2	16.1
VL	470–2475	2 × 1–4 × 4	19.7–25.8	9	1.3	0.1	1.3
					2.5	5.2	6.6
					4.8	22.8	20.0
AV	1584–2464	3 × 2–5 × 4	20.4–23.7	12	5.9	1.9	4.0
					13.4	19.8	22.2
					17.9	38.6	32.0
SX	432–490	3 × 3–4 × 3	12.4–26.1	8	5.8	10.3	9.0
					11.2	22.3	22.1
					15.8	38.8	34.5
FR	1089	3 × 3	23.3–28.1	5	0.9	0.2	1.3
					2.9	8.8	9.4
					4.8	20.3	19.5
PP	243–2919	1.7 × 1.7–3.3 × 3.2	12.6–28.2	1–27	1.0	0.2	1.3
					8.8	12.8	14.8
					31.4	42.0	40.0

[a] Minimum, mean and maximum values; QP, VB, VL, AV and SX are Quinta Paço, Vale Bezerra, Vilar Luz, Alto Vilão and Seixosa spacing trials, respectively; FR: fertilised and irrigated trial; PP: permanent plots.

3. Methodology

3.1. Candidate models

From the works of Tomé (1988) and Ribeiro (1998) the best non-linear functions to express the height–diameter relationship were selected (Table 1).

1. Michailoff equation modified by Tomé (M); this function does not constrain the diameter–height relationship to pass through (0, 1.30).
2. Harrison equation (H); this function was the selected in both previous works but, it was fitted without considering the constraint of diameter–height to the point (0, 1.30); in this work both versions were tested: the first one corresponding to the original Harrison equation where the dependent variable was expressed as ($h - 1.30$), and a second version without constraint to the height 1.30.

The parameters a_0 in M equation and a_0, a_1 and a_2 in H equations were expressed as a combination of tree and stand variables. As an explanatory analysis, an all-possible-regressions algorithm, with tree total height as dependent variable, was used to select combinations of variables to express total height. The variables tested were representative of:

- age (t);
- tree dimension: diameter (d), height (h);
- maximum diameter of the stand (d_{max});
- mean tree dimension: quadratic mean diameter (d_g), quadratic mean diameter of dominant trees (d_d);
- stand density: number of planted trees per hectare (N_{pl}), number of living trees per hectare (N), basal area (G);
- site productivity: dominant height (h_d), site index ($S_{h,t}$); site index was expressed as the mean height of the dominant trees (100 largest dbh trees per hectare) at base age 10 years and it was obtained directly by interpolation or estimated according to Tomé (1990) site index curves.

As a consequence, and being Φ the vector of parameters and ϵ the error term, the general model can be written as

$$h = f(t, d, d_{max}, d_g \text{ or } d_d, N_{pl} \text{ or } N \text{ or } G, h_d \text{ or } S_{h,t}; \Phi) + \varepsilon$$

The selection was based on measures of multiple linear regression performance and prediction ability: adjusted-R^2, residual mean square (RMS), sum of PRESS residuals and sum of absolute PRESS residuals (Myers, 1986). The presence of colinearity was analysed on the basis of the values of the variance inflation factors (VIFs); values up to 10 were accepted (Myers, 1986). The number of variables was restricted by the presence of one of each above-mentioned group.

3.2. Model fitting and selection

In this work the total dataset was randomly split into two subsets and both were used for fit, select and validate the height–diameter equations. The randomisation was at plot level (and not at the measurement date or the tree level) to guarantee that all the trees of a certain plot were in the same data subset trying to increase the "independency" between the data subsets. To ensure that the data splitting was not affected by systematic influences, the equations selected in one data subset were evaluated with the other subset and vice versa.

West (1995) revised some of the techniques used to estimate the covariance matrix of the error term of the model being fitting. The author cited the works of Newberry and Burkhart (1986) and Gregoire (1987) that had made detailed comparisons of the results they obtained with generalised least squares regression and maximum likelihood estimation from ordinary least squares (OLS) regression. They concluded that there was little gain in using more complex techniques and the OLS parameter estimator served adequately. Therefore, in this work and, in spite of the fact that data contains successive measurements from individual plots, the OLS regression is used.

Different versions of Michailoff and Harrison equations, representing different formulations of the parameters as functions of stand variables, were fitted; the parameter estimation of these non-linear functions was based on the least squares method of the PROC NLIN procedure of the SAS/STAT (1989). The modified Gauss–Newton iterative method was applied in model fitting. The PROC MODEL procedure of the SAS/ETS (1993) was used to analyse the colinearity between the variables and to ensure that the solution was global rather than local. Multicolinearity was assessed in terms of the condition number of the correlation matrix; when this value exceeded 1000, the effect of multicolinearity was considered serious and the model discarded (Myers, 1986).

In each data subset, the functions were ranked according to the increasing values of the residual sum of squares (RSS). Ranks of each function in each subset were then summed up to obtain an overall rank. The functions with the smallest overall rank numbers were evaluated in terms of measures of fitting and prediction ability: adjusted-R^2, RMS, mean of PRESS residuals and, mean of absolute PRESS residuals. The PRESS residuals give indication about the predictive ability of the equations by cross-validation (Myers, 1986). This entails omitting each observation in turn from the data, fitting the model to the remaining observations, predicting the response for the omitted observation and comparing the prediction with the observed value: $y_i - \hat{y}_{i,-i} = e_{i,-i}\,(i = 1, 2, \ldots, n)$. Each candidate equation has n PRESS residuals associated with it.

The normality of the studentised residuals was analysed through normal QQ plots. The presence of non-normality was overcome using iteratively reweighted least squares regression; Huber function was selected as the influence function to reduce the influence of data points containing large errors on fit (Myers, 1986)

$$\begin{cases} \psi\left(\frac{e_i}{\sigma}\right) = \frac{e_i}{\sigma}, & \left|\frac{e_i}{\sigma}\right| \le r \\ \psi\left(\frac{e_i}{\sigma}\right) = r, & \frac{e_i}{\sigma} > r \\ \psi\left(\frac{e_i}{\sigma}\right) = -r, & \frac{e_i}{\sigma} < -r \end{cases}$$

where e_i/σ are the studentised residuals and r is the limit factor (a residual that exceeds $r\sigma$ will exert no more influence than a residual with a value of $r\sigma$).

The presence of heteroscedasticity associated with the error term of the models was checked by plotting the studentised residuals against the predicted values. The heteroscedasticity was only checked graphically because the frequent non-normality of the studentised residuals makes the use of statistical tests impractical (e.g. White, 1980). Both the significance and the stability of the parameters estimated in the two data subsets were ensured based on the asymptotic t-statistic.

3.3. Model evaluation

The bias and precision of the selected functions was analysed. Bias was assessed through histograms of the prediction residuals and computation of the mean of the prediction residuals. Precision was expressed by the interquantile range of the prediction residuals (Q99-Q1) and by the computation of the mean of the absolute value of the prediction residuals. Average model bias measures the error when several observations are combined by totalling or averaging, and mean absolute difference measures the average error associated with a single prediction (Soares et al., 1995). Plots of observed over predicted values were also analysed. The model efficiency was computed; this statistic provides a simple index of performance on a relative scale, where 1 indicates a perfect fit, 0 reveals that the model is not better than a simple average, and negative values indicate a poor model indeed (Vanclay and Skovsgaard, 1997).

4. Results and discussion

The random split of the total dataset resulted into two very similar subsets in spite of the fact that the randomisation was made at the plot level (Table 3). Data subset 2 included some old permanent plots, as reflected by the maximum value of the stand age (31.4 years). However, that was not reflected on the maximum values of tree height and diameter; one permanent plot included in data subset 1 presented 40 m of tree height at 14.6 years.

The all-possible-regression algorithm applied to the total dataset resulted in VIFs greater than 10 for all combinations analysed. Tree diameter was excluded from the colinearity analysis because it is a fundamental variable to define a height–diameter equation. However, in spite of these results, and as a consequence of the high colinearity showed by the linear models, several combinations of variables were tested on the fitting of Michailoff and Harrison equations. The number and type of variables were restricted by the presence of one variable of each group previously defined (tree dimension, stand density, site productivity, etc.) and a maximum of three variables; 32 and 54 versions of Michailoff and Harrison equations, respectively, were analysed. Each version of Harrison

Table 3
Characterisation of the two data subsets[a]

Variables	Minimum	Mean	Maximum	S.D.	Minimum	Mean	Maximum	S.D.
	Data subset 1 ($n = 25708$)				Data subset 2 ($n = 25347$)			
$S_{h,t}$	13.5	22.0	28.4	3.1	12.4	22.2	27.3	2.9
N_{pl} (ha^{-1})	500	1647	5078	1118	625	1571	5078	915
N (ha^{-1})	469	1551	5000	1064	586	1457	4922	842
t (years)	0.9	5.6	17.9	4.5	0.9	6.9	31.4	6.2
G (m^2 ha^{-1})	0.06	10.8	41.2	9.3	0.02	14.4	49.5	13.0
h_d (m)	2.2	13.6	36.6	8.0	2.0	15.1	36.4	9.0
d_d (cm)	1.7	13.4	35.4	8.0	1.2	14.7	38.9	8.7
h_m (m)	1.3	11.5	34.2	7.4	1.0	12.8	35.7	8.2
d_g (cm)	1.0	9.8	26.7	6.4	0.7	10.5	31.0	6.5
h (m)	1.3	11.6	40.0	7.8	1.3	12.9	39.0	8.6
d (cm)	0.1	10.0	38.6	7.6	0.1	10.9	42.0	7.9

[a] S.D.: standard deviation; $S_{h,t}$: site index (mean height at base age 10 years of the 100 largest dbh trees per hectare); N_{pl}: number of trees at plantation; N: number of alive trees; t: age; G: basal area; h_d: dominant height; d_d: dominant diameter; h_m: mean height; d_g: quadratic mean diameter; h: tree height; d: tree diameter.

equation was fitted with either $h_0 = 0$ or 1.30 ($y = h - h_0$).

During the model fitting stage, Harrison equation presented convergence problems when all the parameters (a_0, a_1 and a_2) were defined as a linear combination of tree and stand variables. So, only Harrison equation versions with the parameter a_0 expressed as a linear combination of variables were selected. Those equations showed better performance when compared with the two other alternatives: only a_1 or a_2 expressed as a linear combination of tree and stand variables, respectively.

Six equations were proposed for the evaluation stage (Table 4): two of them were age independent; the number of live trees instead of the basal area expressed the stand density in three of the equations; the dominant height/quadratic mean tree diameter was also present in three situations.

The hypothesis of the normality of the studentised residuals was rejected for all functions. In a previous

Table 4
Height–diameter equations submitted to the evaluation stage[a]

Equation	Subset	Adjusted-R^2	RMS	MPRESS	MAPRESS
Michailoff equation modified by Tomé (1988)					
M1: $h = h_d\, e^{(a_0+a_1t+a_2h_d+a_3d_g)(1/d-1/d_d)}$	1	0.98	0.85	−0.0395	0.671
	2	0.99	1.29	−0.0099	0.789
M2: $h = h_d\, e^{(a_0+a_1t+a_2N/1000+a_3h_d)(1/d-1/d_d)}$	1	0.99	0.87	−0.0354	0.684
	2	0.98	1.28	0.0012	0.791
Harrison et al. (1986) equation					
H1: $h = h_d(1 + (a_0 + a_1N/1000 + a_2d_{max})\, e^{bh_d})(1 - e^{cd/h_d})$	1	0.99	0.86	0.0083	0.667
	2	0.98	1.20	−0.0055	0.765
H2: $h = h_d(1 + (a_0 + a_1t + a_2N/1000 + a_3d_d)\, e^{bh_d})(1 - e^{cd/h_d})$	1	0.99	0.84	0.0125	0.665
	2	0.98	1.20	−0.0040	0.766
H3: $h - 1.30 = h_d(1 + (a_0 + a_1h_d + a_2d_g)\, e^{bh_d})(1 - e^{cd/h_d})$	1	0.98	1.28	0.0012	0.791
	2	0.98	1.27	−0.0435	0.799
H4: $h - 1.30 = h_d(1 + (a_0 + a_1t + a_2h_d + a_3d_g)\, e^{bh_d})(1 - e^{cd/h_d})$	1	0.98	1.29	−0.0099	0.789
	2	0.98	1.26	−0.0418	0.795

[a] RMS: residual mean square; MPRESS: mean PRESS residuals; MAPRESS: mean absolute PRESS residuals; N: number of alive trees (ha^{-1}); t: age (years); h_d: dominant height (m); d_d: dominant diameter (cm); d_{max}: maximum diameter (cm); d_g: quadratic mean diameter (cm); h: tree height (m); d: tree diameter (cm); a_0, a_1, a_2, a_3, b and c: equation parameters.

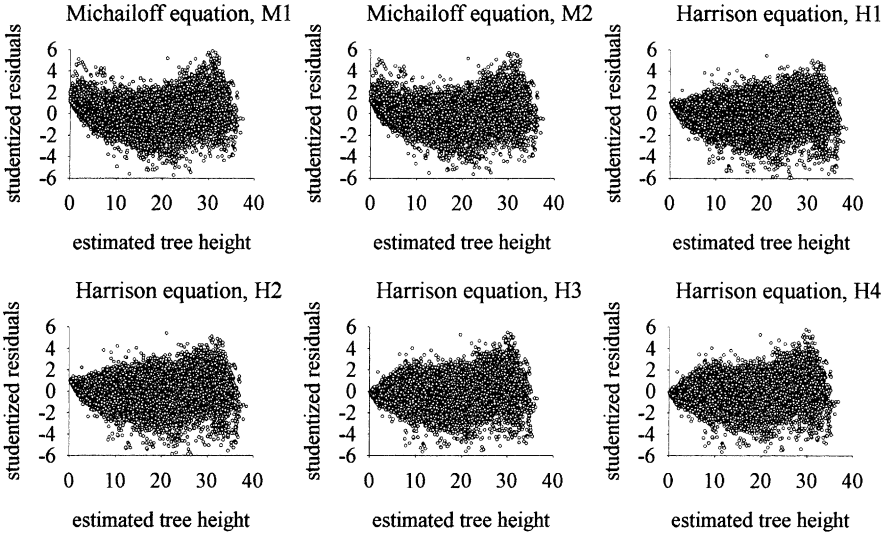

Fig. 1. Graphical relationship between the studentised residuals and the tree height values estimated with the different versions of Michailoff and Harrison equations fitted with the data subset 2 (see Table 4).

analysis, the QQ plots were used to detect outliers; when identified as measurement, handwriting or typing errors these were corrected or excluded. Fig. 1 exemplifies the graphical relationship between the studentised residuals and the height values estimated with the different versions of Michailoff and Harrison equations when fitted with data subset 2. A systematic pattern of the variation of residuals was not observed, although tree height estimates greater than 30 m were associated with a higher variation of the studentised residuals. The major difficulty of measuring old stands—some stands presented trees with more than 40 m of height and the presence of real anomalous situations—trees with decreasing height values in successive measurements as a consequence of decline of the stands—could justify some of the residual values associated with the greatest estimated tree heights.

The high level of accuracy associated with the measurement of tree heights in young stands justified the small residual values associated with the smallest estimates of tree height. This fact was particularly evident with Harrison equation restricted to the initial point diameter–height (0, 1.30)—versions H3 and H4 (Fig. 1, Table 4). With Michailoff and Harrison equations, versions M1, M2 and H1, H2, respectively, a positive bias associated with the smallest estimates of tree height was observed (Fig. 1, Table 4). A detailed analysis allowed the identification of two situations related with tree height estimates smaller than 1.30 m: very young stands, in which the most part of the diameters and heights, at the tree level, were between 0.1–2.1 cm and 1.3–2.1 m, respectively, and young stands in which anomalous relationships between tree diameter and height were observed.

The relationship defined between the studentised residuals and the tree height estimates, analysed for the six versions of Michailoff and Harrison equations and for the two data subsets, did not suggest the presence of heteroscedasticity associated with the error term.

Table 5 shows the mean values of the measures of precision and bias associated with the six analysed versions of Michailoff and Harrison equations as well as the correspondent model efficiencies (MEs). In general, the equations were negatively biased; when both data subsets were considered, the versions H1 and H2 ($y = h$) of Harrison equation were simultaneously more precise and less biased. The mean of the absolute value of the prediction residuals ranged between 67 and 80 cm. The bias expressed in Table 5 was confirmed by the histograms of the height prediction residuals. The graphs of the observed versus predicted tree height values, independently of the data subset and for the six versions of the analysed equations, revealed an approximately linear relation.

Table 5
Measures of precision and bias associated with the six versions of Michailoff and Harrison equations[a]

Equation	$\sum(y-\hat{y})/n$ (m)	$\sum\lvert(y-\hat{y})\rvert/n$ (m)	$RSS_p = \sum(y-\hat{y})^2$ (m^2)	MEs	Q99-Q1 (m)
Evaluation with the data subset 1 of the equations fitted with the data subset 2 ($n = 25707$)					
M1	−0.027	0.674	21951.2	0.986	2.331 − (−2.707) = 5.038
M2	−0.047	0.684	22439.6	0.986	2.289 − (−2.784) = 5.279
H1	0.022	0.669	21956.6	0.986	2.402 − (−2.699) = 5.101
H2	0.021	0.672	22103.4	0.986	2.388 − (−2.698) = 5.086
H3	−0.036	0.690	22752.8	0.985	2.289 − (−2.844) = 5.133
H4	−0.042	0.688	22711.0	0.985	2.248 − (−2.896) = 5.144
Evaluation with the data subset 2 of the equations fitted with the data subset 1 ($n = 25347$)					
M1	−0.012	0.791	33021.4	0.982	3.595 − (−3.112) = 6.707
M2	0.038	0.795	33118.1	0.982	3.673 − (−3.092) = 6.765
H1	−0.048	0.767	30812.7	0.984	2.811 − (−3.506) = 6.317
H2	−0.046	0.776	32023.9	0.983	3.038 − (−3.432) = 6.470
H3	−0.059	0.798	32382.4	0.983	3.114 − (−3.337) = 6.451
H4	−0.024	0.793	32686.8	0.983	3.422 − (−3.215) = 6.637

[a] y: observed height of tree i; $\hat{y}$: estimated height of tree i; n: total number of observations; RSS: residual sum of squares; MEs: model efficiencies; Q99: quantile 99; Q1: quantile 1; M1, M2, H1, H2, H3 and H4 are defined in Table 4.

The results of the analysis of the accuracy by age, site index and planting density classes associated to the height–diameter equations are exemplified in Table 6. The decrease of precision with increasing age was evident for all the equations and in both datasets. This tendency reflected the influence of the quality of the database on the fitting of the equations; the accuracy of the tree height measurements decreases with stand development. Both versions of Michailoff equation presented, for stands older than 4 years, the smallest mean of the absolute values of the prediction residuals. The bad behaviour presented in young stands reflected the fact that this equation is not restricted to the point diameter–height (0, 1.30). However, this fact was not a limitation for a better behaviour presented by versions H1 and H2 of Harrison equation. The versions of Michailoff equations did not present an identical behaviour in both data subsets. In a general way, both versions of Michailoff equation and versions H3 and H4 of Harrison equation were negatively biased, originating tree height estimates superior to the respective observed values; versions H1 and H2 were less biased.

A decrease of precision associated with an increase of site index was observed with all the versions of the two equations. The class that comprised site index values between 20 and 24 m was an exception to this gradient presenting, when compared with the other three classes, the greatest values for the mean of the absolute value of the prediction residuals; this class was associated with the highest mean height values. All the versions analysed were, generally, negatively biased.

Analysing model performance by planting density classes, all the versions of the two equations were less precise for the class with the widest spacings, associated to the greatest tree height values. This class comprises many observations of the old spacing trials (AV and SX) that were measured until the age of 17 years. Version H1 of Harrison equation was, in general more precise; versions of the two equations were in general, negatively biased for densities equal or less than 1667 trees per hectare and positively biased for densities greater than that limit. Not one of the equations assumed a superior behaviour in both data subsets.

From this analysis it was evident:

- that the strict relationship between the results of the evaluation stage and the quality of fitting/evaluation data; the equations are more biased and less precise when applied to high trees and, in a general way, to old stands; the measurement errors associated with the hypsometer and the operator, the anomalous relations between tree diameter and height that expressed particular stand development situations (dry periods, night-frosts, fires, strong winds, diseases and presence of insects) can result in incorrect

Table 6
Mean of the prediction residuals and mean of the absolute prediction residuals by age, site index and density at plantation classes for the equations fitted with data subset 2 and evaluated with data subset 1

	Mean of prediction residuals				Mean of absolute prediction residuals			
Age classes								
Equations	$t \leq 4$	$4 < t \leq 8$	$8 < t \leq 12$	$t > 12$	$t \leq 4$	$4 < t \leq 8$	$8 < t \leq 12$	$t > 12$
Nobs	13588	6059	2616	3432	13588	6059	2616	3432
M1	−0.0140	−0.0705	−0.0432	0.0109	0.479	0.786	0.872	1.097
M2	−0.0326	−0.0623	−0.0737	−0.0596	0.498	0.786	0.879	1.096
H1	0.0213	0.0146	0.0264	0.0350	0.446	0.812	0.882	1.133
H2	0.0208	0.0231	0.0422	−0.0001	0.447	0.822	0.890	1.129
H3	0.0303	−0.0111	−0.2469	−0.1832	0.483	0.810	0.895	1.139
H4	0.0321	−0.0246	−0.2231	−0.2279	0.483	0.806	0.896	1.135
Site index classes								
Equation	$S_{h,t} \leq 16$	$16 < S_{h,t} \leq 20$	$20 < S_{h,t} \leq 24$	$S_{h,t} > 24$	$S_{h,t} \leq 16$	$16 < S_{h,t} \leq 20$	$20 < S_{h,t} \leq 24$	$S_{h,t} > 24$
Nobs	910	5785	11582	7418	910	5785	11582	7418
M1	−0.0110	−0.1410	−0.0199	0.0489	0.600	0.647	0.699	0.663
M2	0.0620	−0.0755	−0.0808	0.0132	0.631	0.660	0.715	0.662
H1	0.0059	−0.0538	0.0244	0.0797	0.573	0.622	0.697	0.675
H2	0.0240	−0.0384	0.0154	0.0748	0.572	0.624	0.698	0.681
H3	−0.2265	−0.2184	−0.0365	0.1298	0.551	0.644	0.720	0.695
H4	−0.1312	−0.1685	−0.0500	0.0802	0.548	0.633	0.722	0.695
Density at plantation classes								
Equation	$N_{pl} < 1111$	$1111 \leq N_{pl} \leq 1667$	$N_{pl} > 1667$	$N_{pl} < 1111$	$1111 \leq N_{pl} \leq 1667$	$N_{pl} > 1667$		
Nobs	8216	9489	7990	8216	9489	7990		
M1	−0.0862	−0.1084	0.1306	0.769	0.683	0.565		
M2	−0.0876	−0.0729	0.0241	0.780	0.691	0.578		
H1	−0.0462	0.0221	0.0922	0.771	0.686	0.546		
H2	−0.0551	0.0378	0.0785	0.771	0.689	0.550		
H3	−0.2073	−0.0783	0.1898	0.772	0.707	0.585		
H4	−0.2152	−0.0940	0.1979	0.773	0.704	0.583		

evaluations of the measures of fit and the predictive ability of the equations;

- that the M1 and M2 versions of Michailoff equation and the H1 and H2 versions of Harrison equation are not the most appropriate for young stands because they do not consider the constraint of diameter–height to the point (0, 1.30);
- that the H4 version of Harrison equation evidence a better performance for young stands.

The final selection falls on versions H4 and H1 of Harrison equation for young and old stands, respectively. These equations were recalibrated with the total dataset ($n = 51055$) by the non-linear iteratively reweighted least squares method to reduce the limitations imposed by the non-normality of the errors detected during the fitting stage. The final equations are as follows. For eucalypt plantations with age less than 4 years

$$h = 1.30 + h_d(1 + (-0.43487 - 0.0108t + 0.09772h_d - 0.06021d_g)\,e^{-0.04864h_d}) \times (1 - e^{-1.58926d/h_d})$$

For eucalypt plantations with age greater than 4 years

$$h = h_d\left(1 + \left(0.10694 + 0.02916\frac{N}{1000} - 0.00176d_{max}\right) e^{0.03540h_d}\right)(1 - e^{-1.81117d/h_d})$$

where h is the tree height (m), d the tree diameter at breast height (cm), t the stand age (years), N the stand density (ha^{-1}), d_{max} the maximum stand diameter (cm), d_g the mean tree diameter (cm), and h_d is the dominant height (m).

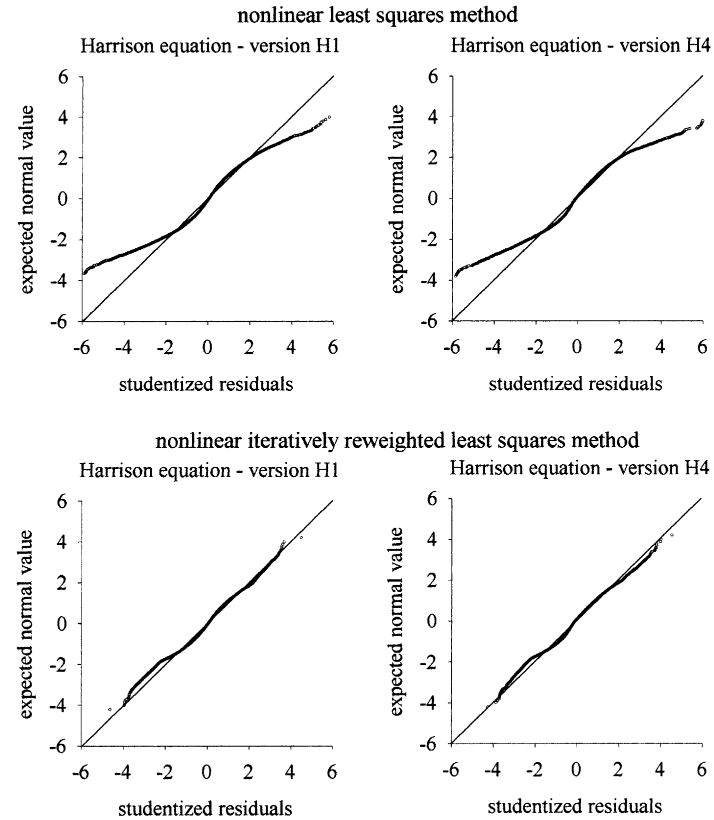

Fig. 2. Normal QQ plots for the studentised residuals, based on the tree height values estimated with the two versions of Harrison equation (H1 and H4) recalibrated with the total dataset by the non-linear least squares method and the non-linear iteratively reweighted least squares method.

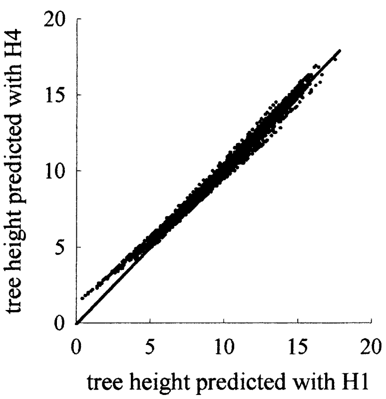

Fig. 3. Relationship between tree height estimates obtained with equations H1 and H4 in stands with ages between 3.5 and 4.5 years ($n = 5688$).

Huber influence function was defined for a limiting value of 2, according to the graphical development present on the normal QQ plots for the studentised residuals. The non-linear iteratively reweighted least squares method, reducing the influence of the points associated to the highest studentised residuals, normalised its distributions (Fig. 2).

To analyse the behaviour of both equations at age of 4 years, the relationship between tree height estimates obtained with equations H1 and H4 was analysed for a data subset restricted from the total data with tree age in the range 3.5–4.5 years (Fig. 3). The relationship observed between both estimates was approximately linear.

5. Conclusions

The final selection of a height–diameter equation for eucalypt plantations in Portugal with age less than

4 years strikes on Harrison equation version H4, while for stands with age greater than 4 years Harrison equation version H1 is recommended. Harrison equation version H4 is mean tree dimension and age dependent; Harrison equation version H1 is density dependent and age independent.

Acknowledgements

Financial support for this research was provided by the project PAMAF 4025 ''Influência do compasso no crescimento e produção de plantações de *Eucalyptus globulus* Labill. e *Pinus pinaster* Ait. em diferentes ambientes ecológicos''; the work of the first author was partially financed by the Programme Ciência/Praxis XXI, BD/2810/93-IE. We acknowledge unpublished data provided by the pulp company CELBI and the Association of Portuguese Pulp and Paper Industries—CELPA.

References

Bennett, F., Clutter, J., 1968. Multiple-product yield estimates for unthinned slash pine plantations-pulpwood, sawtimber, gum. USDA For. Serv. Res. Pap. SE-35.

Burkhart, H., Strub, M., 1974. A model for simulation of planted loblolly pine stands. In: Fries, J. (Ed.), Growth Models for Tree and Stand Simulation. Royal College Forestry, Res. Not. 30, Stockholm, pp. 128–135.

Curtis, R., 1967. Height–diameter and height–diameter–age equations for second-growth Douglas-fir. For. Sci. 13 (4), 365–375.

DGF, 1999. Inventário Florestal Nacional 1998 Distribuição Da Floresta Em Portugal Continental-Resumo. Direcção Geral Florestas, Lisboa, Portugal.

Gregoire, T., 1987. Generalised error structure for forestry yield models. For. Sci. 33, 423–444.

Harrison, W., Burk, T., Beck, D., 1986. Individual tree basal area increment and total height equations for Appalachian mixed hardwoods after thinning. South J. Appl. For. 10, 99–104.

Henriksen, H.A., 1950. Height–diameter curve for logarithmic diameter. Dansk. Skovforen. Tidsskr. 35, 193–202.

Lenhart, J., 1968. Yield of old-field loblolly pine plantations in the Georgia piedmont. Ph.D. Thesis, University of Georgia, Georgia.

Lenhart, J., Clutter, J., 1971. Cubic foot yield tables for old field loblolly pine plantations in the Georgia piedmont. Ga. For. Res. Council, Rep. 22, Series 3, 12 pp.

Lynch, T., Murphy, P., 1995. A compatible height prediction and projection system for individual trees in natural, even-aged shortleaf pine stands. For. Sci. 41 (1), 194–209.

Maguire, D.A., Hann, D.W., 1990. Constructing models for direct prediction of 5-year crown recession in southwestern Oregon Douglas-fir. Can. J. For. Res. 20, 1044–1052.

Mendes, M., 1989. Estudos Biométricos Em Eucaliptal. Forestry Graduate Report, ISA, Lisbon, 40 pp.

Meyer, W., 1940. A mathematical expression for height curves. J. For. 38, 415–420.

Myers, R., 1986. Classical and Modern Regression with Applications. Duxbury Press, Boston, MA, 359 pp.

Michailoff, I., 1943. Zahlenmässiges verfahren für die ausführung der bestandeshöhenkurven forstw. Clb. U. Thar. Forstl. Jahrb. 6, 273–279.

Newberry, J., Burkhart, H., 1986. Variable-form stem profile models for loblolly pine. Can. J. For. Res. 16, 109–114.

Pereira, J.S., Linder, S., Araújo, M.C., Pereira, H., Ericsson, T., Borralho, N., Leal, L.C., 1989. Optimisation of biomass production in *Eucalyptus globulus* plantations: a case study. In: Pereira, J.S., Landsberg, J.J. (Eds.), Biomass Production by Fast Growing Trees. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 101–121.

Prodan, M., 1965. Holzmesslehre. Sauerländers Verlag, Frankfurt.

Ribeiro, R., 1998. Elaboração De Uma Curva Hipsométrica Geral Para *Eucalyptus globulus* Labill. A Partir Dos Dados De Inventário Da Silvicaima. Forestry Graduate Report, ISA, Lisbon, 102 pp.

SAS/STAT, 1989. User's Guide, Version 6, 4th Edition, Vol. 2. SAS Institute Inc., Cary, NC, 846 pp.

SAS/ETS, 1993. Users Guide, Version 6, 2nd Edition. SAS Institute Inc., Cary, NC, 1022 pp.

Soares, P., 1999. Modelação Do Crescimento Da Árvore Em Eucaliptais Em 1ª Rotação Localizados Nas Regiões Norte E Centro Litoral. Ph.D. Thesis, ISA, Lisbon, 369 pp.

Soares, P., Tomé, M., Skovsgaard, J.P., Vanclay, J.K., 1995. Evaluating a growth model for forest management using continuous forest inventory data. For. Ecol. Manage. 71, 251–265.

Staebler, G., 1954. Standard computations for permanent sample plots. US For. Serv. Pacif. Nthwest. For. Range Expt. Sta., 15 pp.

Stoffels, A., van Soest, J., 1953. The main problems in sample plots. Ned. Boschb. Tijdschr. 25, 190–199.

Tomé, M., 1988. Modelação Do Crescimento Da Árvore Individual Em Povoamentos De *Eucalyptus globulus* Labill. (1ª Rotação). Região Centro De Portugal. Ph.D. Thesis, ISA, Lisbon, 256 pp.

Tomé, M., 1990. Modelação do crescimento em altura dominante em eucaliptais com recurso a análise multivariada e análise de regressão. In: Turkman M., Carvalho, M. (Eds.), Actas Da 1ª Conferência Em Estatística E Optimização. INIC & DEIO, Lisboa, pp. 443–462.

Vanclay, J.K., Skovsgaard, J.P., 1997. Evaluating forest growth models. Ecol. Model. 98, 1–12.

West, P., 1995. Application of regression analysis to inventory data with measurements on successive occasions. For. Ecol. Manage. 71, 227–234.

White, H., 1980. A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. Econometrica 48, 817–838.