

Site index model for *Pinus sylvestris* in north-east SpainM. Palahí^{a,*}, M. Tomé^b, T. Pukkala^c, A. Trasobares^a, G. Montero^d^aCentre Tecnològic Forestal de Catalunya, Fundació Catalana per la Recerca, Pg. Lluís Companys, 23, 08010 Barcelona, Spain^bDepartamento de Engenharia Florestal, Instituto Superior de Agronomia, Tapada de Ajuda, 1349-017 Lisboa, Portugal^cFaculty of Forestry, University of Joensuu, P.O. Box 111, 80101 Joensuu, Finland^dDepartamento de Silvicultura, CIFOR-INIA, Carretera de la Coruña, Km 7. 28080 Madrid, Spain

Received 20 June 2001; received in revised form 5 February 2003; accepted 12 June 2003

Abstract

The McDill–Amateis difference model and difference forms of the Chapman–Richards, Hossfeld, Lundqvist–Korf and Schumacher growth models were derived to model the dominant height growth of Scots pine (*Pinus sylvestris* L.) in north-east Spain. Data from permanent sample plots and stem analysis were combined and used for modelling. Both qualitative and quantitative criteria were used to compare alternative models. The PRESS (prediction sum of squares) statistic was used to analyse the predictive ability of the models. The polymorphic difference equation derived from the Hossfeld function resulted in the best compromise between biological and statistical aspects, producing the most adequate site index curves. This dynamic equation can compute predictions directly from any age–dominant height pair without compromising consistency of the predictions. The predictions of dynamic equations are unaffected by arbitrary changes in base-age. The presented Hossfeld equation is recommended for site classification and dominant height prediction in Scots pine stands in north-east Spain.

© 2003 Published by Elsevier B.V.

Keywords: Scots pine; *Pinus sylvestris*; Hossfeld equation; Non-linear difference equations; Site classification

1. Introduction

Scots pine (*Pinus sylvestris* L.) is a species of great economic, ecological and social importance in Spain, as it forms large forests in most mountain ranges (Pirineos, Sistema Central, Sistema Ibérico, etc.). One of the major needs in forest management planning is to predict forest stand development under different treatment alternatives. Construction of site index curves is a fundamental task for site-quality differentiation and for predicting stand development.

According to Clutter et al. (1983), most techniques for site index curve construction can be viewed as special cases of three general methods:

1. the guide curve method;
2. the parameter prediction method; and
3. the difference equation method.

Previous site index curves in Spain have been constructed using the guide curve method (García Abejon, 1981; García Abejon and Gomez Loranca, 1984; García Abejón and Tella Ferreiro, 1986), which assumes proportionality among curves of different site index. An average curve is fitted, and a set of anamorphic site index curves can then be generated. Rojo and Montero (1996) developed the latest site index curves for Scots pine in Spain. They used a

* Corresponding author. Tel.: +34-93-268-7700;

fax: +34-93-315-01-40.

E-mail address: marc.palahi@ctfc.es (M. Palahí).

technique known as sectioning, which first determines the parameters of a guide curve for the poorest and the best sites, respectively. Then height values for intermediate sites are obtained by interpolation between the two fitted curves, resulting in a set of polymorphic curves.

The difference equation method makes direct use of the fact that observations on a given plot should belong to the same site index curve (Elfving and Kiviste, 1997). Height H_2 at age t_2 is expressed as a function of t_2 , height H_1 at age t_1 , and t_1 . The expression is obtained through substitution of one parameter in the growth model. Substitution of the asymptotic parameter produces anamorphic curves. Substitution of any other parameter produces polymorphic curves with a common asymptote. Difference equations are dynamic equations representing a continuous four-variable prediction system directly interpreting three-dimensional surfaces without explicit knowledge of the third dimension, which depends on an unobservable variable (Cieszewski and Bailey, 2000).

The objective of this study is to develop a site index model for Scots pine that is applicable in north-east Spain. The difference equation method was used.

2. Material and methods

2.1. Data

The data were measured in 24 permanent sample plots (Table 1) established in 1964 by the Instituto Nacional de Investigaciones Agrarias (INIA) to represent most Scots pine sites in north-east Spain. The plots were located in the provinces of Huesca, Lérida and Tarragona. The plots were naturally regenerated and thinned during the second measurement. All plots were thinned from below using a moderate intensity.

Table 1
Summary of the characteristics of the 24 permanent plots used in the study

Variable	Average	Minimum	Maximum	Standard deviation
Stand age (years)	67.7	33	148	24.56
Dominant height (m)	15.8	8	25	3.47
Number of measurements	5.3	2	7	1.48

Table 2
Characteristics of the stem analysis data^a

Variable	Average	Minimum	Maximum	Standard deviation
Tree age ^b (years)	28.3	3	100	20.14
Height (m)	7.4	0.3	21.5	5.19
Number of cross-sectional cuts	6.4	4	10	1.50

^a The total number of trees was 36 (two stems per plot).

^b The age of the tree at the height of the cross-sectional cut.

The mean plot area was 0.1 ha. The plots were last remeasured during the year 2000; there were an average of five measurements per plot. Dominant height was defined as the average total height of the 100 thickest trees per hectare.

Stem analysis data (Table 2) from 18 of the permanent sample plots were also used in the study. In each plot two trees were selected as sample trees. Only co-dominant and dominant trees free of damage and without obvious history of suppression were selected. Each tree was felled and height–age pairs were determined by making cross-sectional cuts every meter in the first 10 m of the stem and every 2 m after the first 10 m. Precise measurements of annual rings at each cross-sectional cut were made. The tree age, at the height where the disc was taken, was concluded by subtracting the number of rings from the total age of the tree. It is true that this method can give some biased height–age relationships. However, due to the fact that plots were naturally regenerated and *P. sylvestris* is a rather slow growing species the correction seemed to be unnecessary. Several correction methods presented by Dyer and Bailey (1986) were tested but the height–age relationships did not vary significantly. The mean height of the sample trees was as considered as the dominant height of the plot.

2.2. Candidate functions

Five equations were selected as candidate functions to model dominant height growth. Four of them were based on various transformations of the exponential function: the Chapman–Richards function (Pienaar and Turnbull, 1973), the Lundqvist–Korf function (cited in Tomé, 1988), the McDill–Amateis function (McDill and Amateis, 1992), and the Schumacher function (Schumacher, 1939). The last one,

the Hossfeld function, was based on a fractional function. The Hossfeld function has been utilised in Spain (Pita, 1964) for developing the first site index curves for Scots pine, and more recently Elfving and Kiviste (1997) used a special formulation of the Hossfeld model proposed by Cieszewski and Bella (1989) for constructing site index equations for Scots pine in Sweden. The Hossfeld model has been usually expressed in integral form:

$$H = \frac{t^2}{a + bt + ct^2} \quad (1)$$

where t is the age of the stand (years), H the dominant height and a , b and c the parameters to be estimated.

The Chapman–Richards function has been widely utilised in modelling biological growth phenomena (Rennolls, 1995; Rojo and Montero, 1996; Amaro et al., 1998) and is usually expressed in integral form:

$$H = A(1 - e^{-kt})^{1/(1-m)} \quad (2)$$

where A is the asymptote of dominant height (H , m), k a growth rate related parameter, t the stand age (years), and m a shape parameter.

The Schumacher function has been commonly used with the guide curve method (Clutter et al., 1983) to

produce anamorphic site index curves (Bennet et al., 1959), but it can also be transformed to obtain a polymorphic system (Bailey and Clutter, 1974). Its integral form is

$$\ln H = a + b \left(\frac{1}{t} \right) \quad (3)$$

where a is the logarithm of the maximum dominant height and b the slope parameter. Cieszewski and Bailey (2000) defined t as a simple exponential function of age (Age^c where c was an estimable parameter).

The Lundqvist–Korf function is an S-shaped curve and it has been used often for modelling dominant height (Tomé, 1988; Carvalho et al., 1994). Its integral form is

$$H = A e^{-k/t^d} \quad (4)$$

where A is the asymptote of dominant height, k a growth rate related parameter, t the stand age (years), and d a shape parameter. The Lundqvist–Korf function can be seen as a generalisation of the Schumacher (1939) function (e.g., the Schumacher function presented by Cieszewski and Bailey (2000) with a linearisation parameter c). In this study, both functions were grouped as the same functional form (Table 3),

Table 3
Candidate equations for dominant height modelling

Designation	Function	Free/site specific parameter	Difference equation	No.
Lundqvist–Korf	$H = A e^{-k/t^d}$	k	$H_2 = A \left(\frac{H_1}{A} \right)^{(t_1/t_2)^d}$	F01
		A	$H_2 = H_1 e^{k((1/t_1^d) - (1/t_2^d))}$	F02
		d	$H_2 = \frac{A}{\ln t_2 \sqrt{(1/k)^{\ln(t_2/t_1)} \ln[A/H_1]}}$	F03
Schumacher	$\ln H = a + b \left(\frac{1}{t^c} \right)$	a	$H_2 = e^{\ln(H_1) + b[(1/t_1)^c - (1/t_2)^c]}$	F09
		b	$H_2 = e^{[a + (1/t_2)^c] \ln(H_1) - a}$	F10
		a, b	$H_2 = e^{a - (b/t_2) + (\ln(H_1) - a + (b/t_1)) \times [d - (1/t_2)] \times [d - (1/t_1)]}$	F11
Chapman–Richards	$H = A(1 - e^{-kt})^{1/(1-m)}$	k	$H_2 = A \left\{ 1 - \left[1 - \left(\frac{H_1}{A} \right)^{1-m} \right]^{t_2/t_1} \right\}^{1/(1-m)}$	F04
		A	$H_2 = H_1 \left(\frac{1 - e^{-kt_2}}{1 - e^{-kt_1}} \right)^{1/(1-m)}$	F05
		m	$H_2 = A^{1 - (\ln(1 - e^{-kt_2}) / \ln(1 - e^{-kt_1}))} \times H_1^{\ln(1 - e^{-kt_2}) / \ln(1 - e^{-kt_1})}$	F06
McDill–Amateis			$H_2 = \frac{M}{1 - (1 - (M/H_1)) \times (t_1/t_2)^a}$	F07
Hossfeld	$H = \frac{t^2}{a + bt + ct^2}$	b	$H_2 = \frac{t_2^2}{a + t_2[(t_1/H_1) - c \times t_1 - (a/t_1) + c \times t_2]}$	F08

but since we analysed some of the difference equations derived by Cieszewski and Bailey (2000) from the Schumacher formula with a linearisation parameter c , we refer to them as derivations of the Schumacher function.

All the functions presented above, are usually presented in integral or differential forms. However, when using data from remeasured plots and to project growth over discrete time intervals such as years, difference forms are more appropriate (Amaro et al., 1998). In deriving difference forms of the Chapman–Richards and Lundqvist–Korf functions, it is used an approach developed by Amaro et al. (1998) consisting in equating subdefined ratios of base equations. Using this approach three difference equations were derived for both Chapman–Richards and Lundqvist–Korf models. In deriving difference forms of Schumacher and Hossfeld functions, the algebraic difference approach (ADA) (Bailey and Clutter, 1974) and the generalised algebraic difference approach (GADA) (Cieszewski and Bailey, 2000) were used. The ADA has been applied successfully in various modelling contexts, e.g., by Clutter et al. (1983), Borders et al. (1984, 1988), Lappi and Bailey (1988), McDill and Amateis (1992), Cao (1993), Cao et al. (1997), and Begin and Schutz (1994). Using the ADA approach polymorphic equations of both the Schumacher and Hossfeld models, and an anamorphic equation of the Schumacher model were derived. The GADA is a new approach to equation derivation developed by Cieszewski and Bailey (2000), who demonstrated its advantages arising from greater flexibility in the model analysis. The GADA approach was used to derive a generic form of the Schumacher model. Table 3 summarises all candidate functions derived by these methods.

The McDill–Amateis function, which has proved to fit dominant height data very well (McDill and Amateis, 1992; Amateis, 1994), has no integral form. This equation is given by

$$H_2 = \frac{M}{1 - (1 - (M/H_1)) \times (t_1/t_2)^a} \quad (5)$$

where subscripts 1 and 2 refer to time points 1 and 2, M the asymptote, and a a growth rate/shape parameter (Table 3).

All these candidate functions have the property of time invariant predictions ensuring that projections

over the same period of time are equivalent, regardless of the length or number of projection intervals.

2.3. Parameter estimation

Let f represent any of the eleven non-linear candidate functions (Table 3), Y the dependent variable Y_2 , X the vector of explanatory variables (i.e. X_1 , t_1 and t_2), and ϕ the vector of parameters:

$$Y = f(X, \phi) + \varepsilon \quad (6)$$

Most statistical procedures to estimate ϕ assume that the error term, ε , is normally distributed with zero mean, constant variance, and independence between cases.

The observations in the estimation data (Tables 1 and 2) allow the generation of a first-order difference equation data structure that contains observations of initial age and dominant height along with final age and dominant height for each measurement period according to the candidate functions considered in Table 3.

We performed a non-linear regression analysis using a non-linear least squares (NLS) technique in SPSS (SPSS, 1999). SPSS software uses the Levenberg–Marquard algorithm to obtain the final parameter estimates. The loss function was defined as the sum of squared residuals (observed minus predicted values). Due to the characteristics of the modelling data (multiple measurements are taken from individual sampling units, and measurements are combined across sampling units), a non-random pattern may exist in the residuals of each sampling unit if differences among sampling units cannot be fully incorporated in regression models (Fox et al., 2001). The presence of a non-random pattern amongst the residuals of any one sampling unit would mean that those residuals tend to be positively correlated. Therefore, the basic assumption in ordinary least squares of independence of the error term is likely to be violated (West, 1995). The violation of this assumption does not lead to any problem with the parameter estimates of the equation, which are unbiased, but the estimator of the covariance matrix of the parameter estimates of the equation will be negatively biased (West, 1995). Such bias would lead to invalidation of the normal statistical hypothesis test about the fitted equation.

Difference equations are recursive equations for which the regression assumptions are violated due to the fact that the explanatory variables are not independent variates (Seber and Wild, 1989). In fact, in this type of equation, the response variable, Y_2 , becomes the explanatory variable for a different observation (Y_3) that could result in a large likelihood of serial correlation (Amaro et al., 1998).

At first, the fitting process for each of the 11 candidate difference equations was performed using only the permanent plots as the modelling data. However, all models fitted using permanent plots data produced unrealistic results at young ages. Therefore, another data set consisting of stem analyses, which included height observations at young ages was added. The fitting process was then repeated using the combined data of permanent sample plots and stem analysis (Tables 1 and 2). A residual analysis was performed for each of the fitted equations.

2.4. Model selection criteria

The procedures used to evaluate the models included qualitative as well as quantitative examinations, and both theoretical and empirical issues were addressed. A three-step procedure was used to select the most appropriate model. Step 1 was based on the model fitting statistics and other desirable characteristics, like biological realism and graphical appearance. This analysis was based on six model performance evaluation criteria described by Amaro et al. (1998)

(Table 4), resulting in a selection of the apparently best equations. The asymptotic parameter was examined in order to discard biologically non-realistic models. Step 1, resulted in four remaining equations: two derived from the Schumacher function, one from the Hossfeld model, and one from the McDill–Amateis model. Step 2 analysed the characterisation of model error, based on a validation technique. Since a validation data set was not available, the PRESS (prediction sum of squares) statistic was used:

$$\text{PRESS} = \sum_{i=1}^n (Y_i - \hat{Y}_{ip}^*) \quad (7)$$

where n is the number of observations, Y_i the observed value of observation i , and $\hat{Y}_{ip}^*$ the predicted value for observation i as calculated from a regression equation derived through fitting the p parameter model to a data set obtained by deleting observation i from the original data set (Montgomery and Peck, 1982).

Calculation of the PRESS statistic requires fitting of the p parameter model to each of the n different data sets (there are as many data sets as observations) obtained by deleting observation i from the original data set. Each of the regression equations is used to calculate a single predicted value (Clutter et al., 1983). However, it is possible to calculate PRESS from the results of a single least squares fit to all observations, since PRESS is just a weighted sum of squares of the residuals, where the weights are related to the leverage of the observations (Montgomery and Peck, 1982). The latter technique was used.

Table 4
Model performance evaluation criteria (estimation and validation procedures)^a

Performance criterion	Symbol	Formula	Ideal
Mean residual	Mres	$\sum_{i=1}^n \frac{\text{est}_i - \text{obs}_i}{n}$	0
Variance ratio	VR	$\frac{\sum_{i=1}^n (\text{est}_i - \overline{\text{est}})^2}{\sum_{i=1}^n (\text{obs}_i - \overline{\text{obs}})^2}$	1
Residual mean of squares	RMS	$\frac{\sum_{i=1}^n (\text{est}_i - \text{obs}_i)^2}{n - p}$	0
Absolute mean residual	Amres	$\sum_{i=1}^n \frac{ \text{est}_i - \text{obs}_i }{n}$	0
Coefficient of determination/model efficiency	R^2/Mef	$1 - \frac{\sum_{i=1}^n (\text{est}_i - \text{obs}_i)^2}{\sum_{i=1}^n (\text{obs}_i - \overline{\text{obs}})^2}$	1
Linear regression	α, β and R_{adj}^2	$\text{obs}_i = \alpha + \beta \text{est}_i + e_i$	$\alpha = 0, \beta = 1$ and $R_{\text{adj}}^2 = 1$

^a est_i : i th estimated value; obs_i : i th observed value; n : number of observations; p : number of parameters.

There are numerous graphical and statistical methods to support model validation (e.g., Soares et al., 1995; Tomé and Soares, 1999; Huang, 1999). Quantitative evaluation using the predicted residuals included the six model performance evaluation criteria summarised in Table 4. Plots of model bias (mean residual) and model precision (mean absolute residual) across age classes and site classes were also assessed graphically.

Step 3 assessed the correctness of the theoretical and biological aspects of the five models. This was done interactively with Step 2. Taking into account the information available on Scots pine growth patterns and the intended use of the model the following biological aspects were examined:

- (1) Signs and values of the coefficients in the model components, especially the asymptotes.
- (2) Quality of extrapolations outside the range of the site indexes of the modelling data as well as outside the age range.
- (3) Height curve development at young ages.

After all these considerations, the most adequate equation for describing dominant height growth of Scots pine for north-east Spain was selected.

3. Results

3.1. Need for including stem analyses in the modelling data

All candidate equations were first fitted based only on the permanent plots data (Table 1). This data set did not include plots younger than 33 years old. The models based on these data gave unrealistic predictions at young ages. Fig. 1 shows, for the Hossfeld model (F08), that the model based on combined permanent plot and stem analyses data produced more realistic height curves specially at very young ages. The model based on permanent plots gave biased predictions at very young ages (overestimating dominant height) and had a high absolute error (AMres) (Fig. 2). However, when stem analyses data were included in the modelling

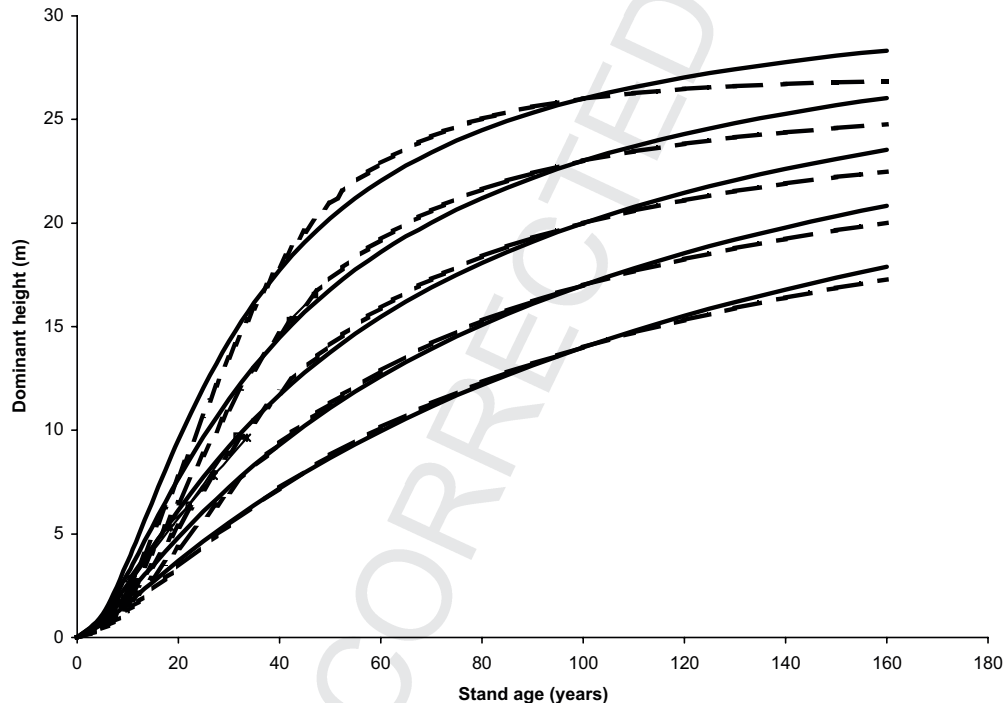


Fig. 1. Comparison of site index curves produced by the Hossfeld model (F08) when fitted with two different modelling data sets, based on permanent plots (---) and stem analyses combined with permanent plots (—).

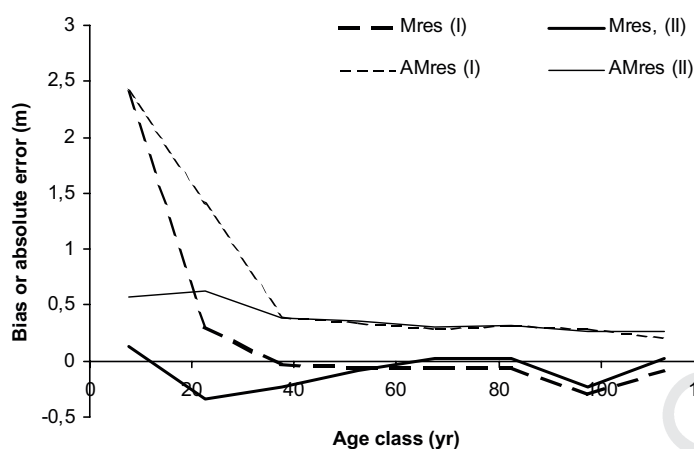


Fig. 2. Comparison of the bias (Mres) and absolute error (AMres) of the Hossfeld model (F08) when fitted with two different data sets: (I) using only permanent plots and (II) permanent plots combined with stem analysis as the modelling data.

data set, the accuracy of the model improved substantially for young ages (Fig. 2).

In the other candidate equations, the problems (low accuracy at young ages) of using only permanent plots as the modelling data were even more serious. Therefore, it was decided to use the combined permanent plots and stem analyses data set for fitting all models and conducting the model performance analyses.

3.2. Performance of candidate equations

The models with the best performance results (Table 5), were the McDill–Amateis function (F07), the generic dynamic equation derived from the Schumacher function (F11), the Chapman–Richards equation with k as a free parameter (F04), the polymorphic

equation derived from the Schumacher function (F10), and the polymorphic equation derived from the Hossfeld function (F08). However, the Chapman–Richards equation (F04) was not further analysed due to its low asymptote ($A = 24$ m). The non-linear regression analysis using the Lundqvist–Korf equation with d as a free parameter (F03) did not converge in model fitting. The other equations derived from the Lundqvist–Korf function (F01, F02) had the poorest fitting statistics. Plots of residuals of equations F01 and F02 showed a poor fit, especially at young stand ages.

3.3. PRESS validation and biological considerations

For the validation procedure, some of the performance evaluation criteria (Table 4) were computed

Table 5
Performance results for the estimation data set

	F01	F02	F04	F05	F06	F07	F08	F09	F10	F11
Fitting statistics										
Mres	−0.711	−0.165	−0.088	−0.178	−0.140	−0.123	−0.171	−0.172	−0.134	−0.082
VR	0.759	1.110	1.050	1.071	1.047	1.041	1.058	1.080	1.054	1.068
RMS	2.384	0.438	0.339	0.398	0.388	0.324	0.351	0.408	0.347	0.325
Amres	1.336	0.551	0.487	0.527	0.544	0.473	0.491	0.528	0.493	0.481
R^2	0.921	0.985	0.988	0.986	0.987	0.989	0.988	0.986	0.988	0.989
Linear regression $\text{obs}_i = \alpha + \beta \text{est}_i + \varepsilon_i$										
R^2	0.94	0.99	0.99	0.98	0.98	0.99	0.99	0.98	0.99	0.99
α	0.84	0.84	0.44	0.64	0.48	0.42	0.56	0.69	0.51	0.54
β	0.94	0.94	0.97	0.96	0.97	0.97	0.96	0.95	0.96	0.96

Table 6
Performance criteria by using the PRESS validation technique

	F07	F11	F08	F10
PRESS	65.328	64.188	70.277	71.861
MEf	0.988	0.988	0.987	0.987
Mres	0.108	0.074	0.153	0.117
AMres	0.438	0.445	0.456	0.459

with the PRESS technique for each of the four models selected in the first evaluation step. Plots of the mean residual and absolute mean residual across different age and site classes showed a similar trend in the four selected models. The McDill–Amateis function (F07) and the generic equation derived from the Schumacher function (F11) gave the best performance results (Table 6), but the results for all four functions were quite similar. The McDill–Amateis model had a rather low asymptote ($A = 27$ m), since previous studies by Rojo and Montero (1996) have shown that Scots pine in Spain can exceed 30 m in height. Furthermore, dominant height curves using the McDill–Amateis

model did not perform well when extrapolating outside the age and site index range of the data (Fig. 3).

The generic equation derived from the Schumacher function using the GADA approach (F11) was polymorphic and had different asymptotes, two desirable characteristics when constructing site index curves (Cieszewski and Bailey, 2000). However, graphical analysis showed that dominant height growth produced with this equation at young ages (see Fig. 3) did not reflect the growth pattern observed in our empirical data (see Fig. 5). Furthermore, at old ages for the best sites the generic form of the Schumacher equation as well as the polymorphic form (F10) produced a non-asymptotic trend (Fig. 3) which might not be realistic. However, due to the lack of very old permanent plots in the modelling data this cannot be confirmed.

The polymorphic form of the Schumacher function (F10) produced excessive growth rates at young ages when extrapolating outside the range of site classes of our data (Fig. 3). The polymorphic equation derived from the Hossfeld function (F08) produced more S-

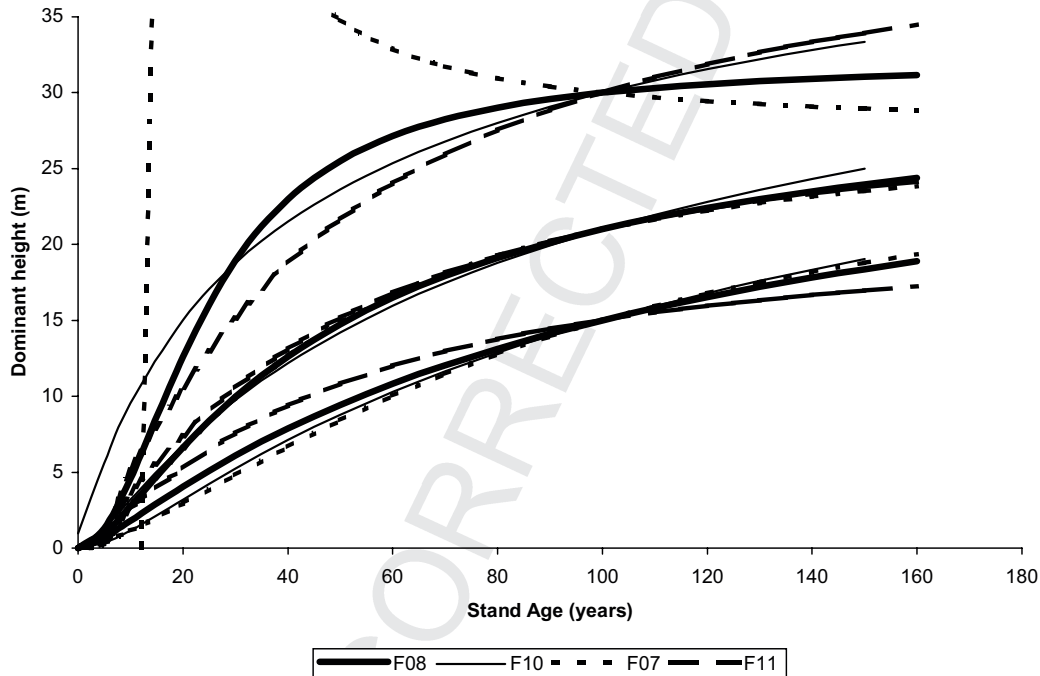


Fig. 3. Comparison of height curves produced by the four selected models: F07 (McDill–Amateis), F08 (polymorphic form of Hossfeld model), F10 (polymorphic form of Schumacher model) and F11 (generic dynamic form of Schumacher model). The curves represent dominant height development for site indices 30, 21 and 15 m at 100 years.

shaped curves giving a realistic development at young ages. The Hossfeld equation produced sensible curves when extrapolating outside the age and site index range of our data. This equation generated a set of polymorphic curves with a common asymptote, which was independent of site. Whether or not the latter property has biological significance is not important as the effective limit will depend on the life span of the species under consideration, and therefore will not be constant across site values. The polymorphic equation derived from the Hossfeld model was finally:

$$H_2 = \frac{t_2^2}{a + t_2[(t_1/H_1) - c \times t_1 - (a/t_1) + c \times t_2]} \quad (8)$$

The parameters of the model are as follows: $a = 18.6269$; $c = 0.03119$.

4. Discussion

We were aware of the restrictions of our data, having only three plots with measurements over the age of 100 years. The lack of permanent sample plots measured at young ages was also an important restriction. Tree analysis data combined with permanent plots data had already been used by Rojo and Montero (1996) for creating site index curves. In our study, the inclusion of tree analysis data was crucial to develop a model that could predict accurately dominant height development at young ages (Fig. 2). The fact that models based on permanent plots and stem analyses data give higher dominant heights at young ages (e.g., Fig. 1 for the Hossfeld equation) than models based only on permanent plots supports the conclusion that the analysed sample trees were dominant rather than co-dominant trees. It also supports the use of stem analyses data to produce more realistic site index curves, especially, at young ages.

Due to the characteristics of the modelling data, the assumption of the independence of the error term is likely to be violated. The violation of this assumption may produce an estimator of the covariance matrix of the parameter estimates that is negatively biased (West, 1995), leading to the invalidation of the normal statistical hypothesis test about the fitted equation. However, hypothesis test may be quite unnecessary when the form of the model being fitted is limited

(West, 1995). For example, Alemdag (1988) developed a model to predict height growth of individual *Picea glauca* trees in terms of their present height and age. He used the two-stage approach to account for the nested structure of his data and avoid the violation of the assumption of independent residuals. However, the model form seemed so simple and obvious that no detailed or formal hypothesis testing seemed necessary. In our study, the model to predict dominant height in terms of current height and age seemed to be rather obvious for any hypothesis testing. Furthermore, in our study new functional forms to model dominant height were not developed, but well-known equations were used as candidate equations. The selected Hossfeld equation (F08) was also fitted as a mixed model (Fox et al., 2001) to be able to characterise and explicitly incorporate the nested stochastic error structure (plot and measurement factors) in the model estimation. Mixed models use random parameters to characterise nested variability. However, in our study, the random plot and measurement factors included for fitting the Hossfeld model using mixed modelling methodology were not significant. These could be explained by the fact that differences among plots and measurements were already fully incorporated in the regression model through the explanatory variables. Another factor that could affect these results could be that only 24 permanent plots composed the modelling data, which in turn were treated using a rather homogenous silviculture. Palahí et al. (2003) used the same permanent sample plots as in this study to develop individual-tree diameter growth models using a mixed model approach. In their study, the diameter growth model included significant plot- and tree-random factors due to the existence of nested stochastic structure in model residuals, indicating that the variables in the model did not explain all variation in diameter growth among plots and among trees within each plot.

The assumption concerning the non-independence of explanatory variables and the possibility of serial correlation, might be also violated. However, the plot of residuals versus lag residuals did not show a strong temporal autocorrelation (Fig. 4), and the Durbin–Watson test ($d = 1.995$) was not significant at $*P < 0.05$.

Due to the fact that only 24 permanent plots were available for fitting the models, all data were used for

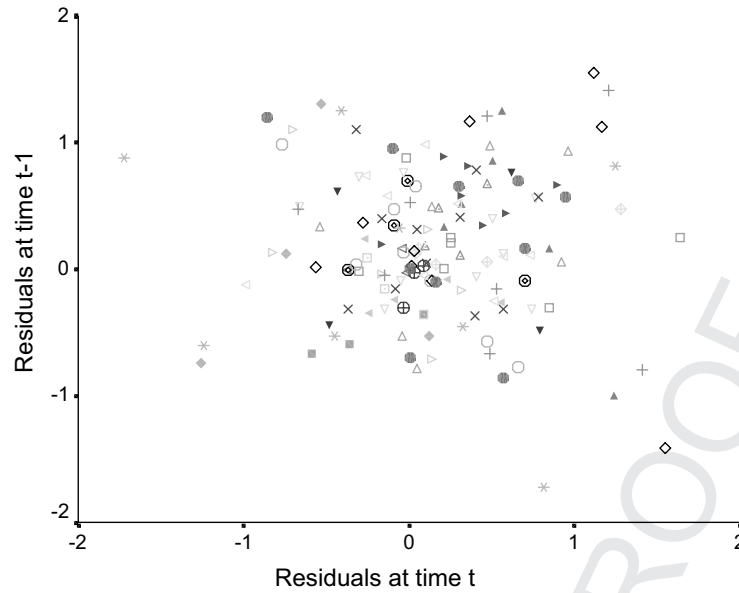


Fig. 4. Lag residual graph for the temporal correlation of the residuals of selected model (Hossfeld). Points with the same symbol represent the same plot.

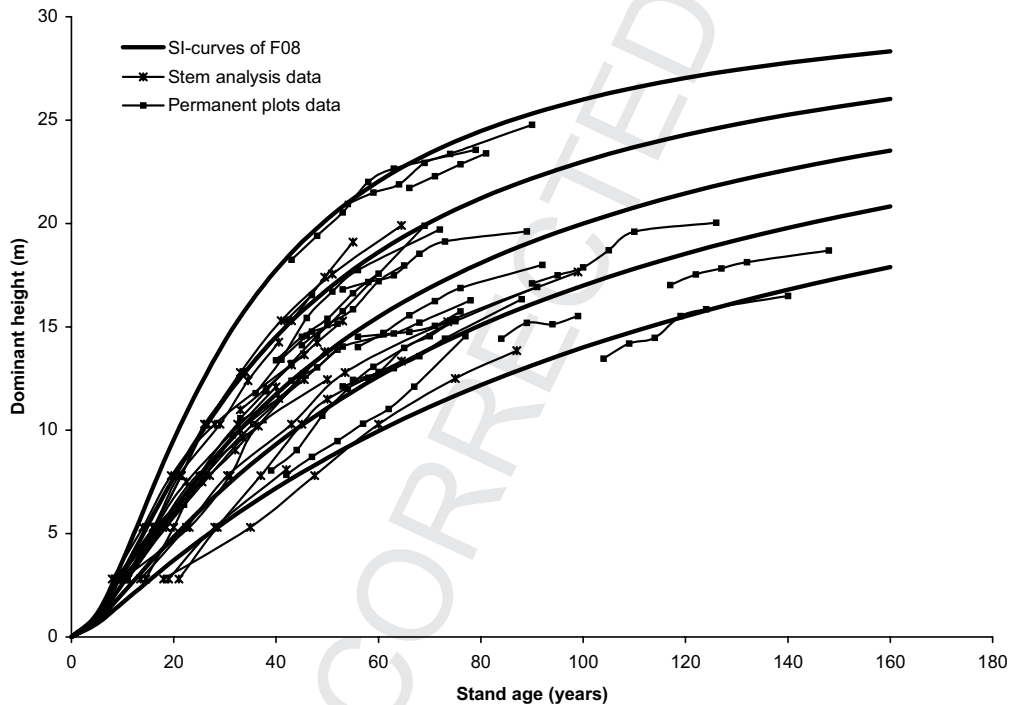


Fig. 5. Height curves produced by the selected Hossfeld equation (F08) together with the empirical modelling data from permanent plots and stem analyses.

modelling and no independent data set was used for validation. However, the PRESS technique was used to analyse the predictive ability of the models. Although the PRESS results presented in the study considered each observation as the deleting unit, the PRESS was also calculated considering each plot as the deleting unit. However, there were no significant differences between the two approaches used.

Model selection was viewed as a compromise between biological and statistical considerations, rather than as a pure exercise in statistical inference. The model selection procedure indicated that the polymorphic equation derived by the ADA approach from the Hossfeld function resulted in the best compromise between biological and statistical aspects, producing the most adequate site index curves (Fig. 5).

In Fig. 6, the site index curves of F08 (Hossfeld model) are compared with previous site index curves for Scots pine developed in Spain. The site index curves developed for the same region by Garcia Abejón and Tella Ferreiro (1986) show that for the

best sites there was a non-asymptotic trend that produces non-realistic growth at old ages (Fig. 6), and underestimate dominant height at younger ages. The site index curves recently developed by Rojo and Montero (1996) for the Sierra de Guadarrama (in central part of Spain) (Fig. 6) were also recommended by Madrigal et al. (1999) to be use in other regions (Pirineos and Sistema Ibérico). It can be seen from Fig. 6 that these curves seriously underestimate dominant height at young and medium ages, and do not indicate at those stages the dominant height development pattern that was present in our empirical data (see Fig. 5). Therefore, we recommend the use of equation F08 for site classification and height growth prediction in Scots pine stands in north-east Spain instead of the earlier models.

This dynamic equation can compute predictions directly from any age-dominant height pair without compromising consistency of the predictions. The predictions are unaffected by changes in the base-age. The site equation is mathematically sound, and always computes consistent predictions.

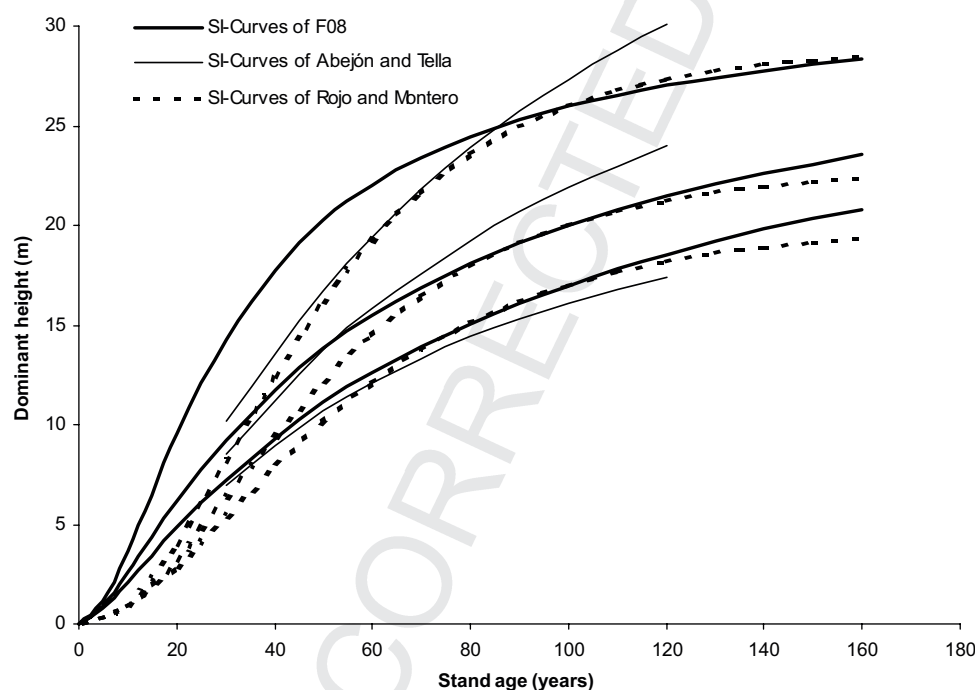


Fig. 6. Comparison of site index curves developed in this study (F08) to previous site index curves developed for the same region by Garcia Abejón and Tella Ferreiro (1986) and by Rojo and Montero (1996) for the Sierra de Guadarrama.

Acknowledgements

Financial support for this project was given by the Forest Technology Centre of Catalonia (Solsona, Spain). We gratefully acknowledge the data provided by the Instituto Nacional de Investigaciones Agrarias (Spain). We thank Mr. Kalle Eerikäinen (Finland) for his helpful suggestions and also thank Mr. Tim Green for the linguistic revision of the manuscript.

References

- Alemdag, I.S., 1988. Site index equations for White Spruce in the Northwest Territories, Canada. *For. Ecol. Manage.* 23, 61–71.
- Amaro, A., Reed, D., Tomé, M., Themido, I., 1998. Modeling dominant height growth: eucalyptus plantations in Portugal. *For. Sci.* 44 (1), 37–46.
- Amateis, R.L., 1994. An approach developing process-oriented growth and yield models. *For. Ecol. Manage.* 69, 7–20.
- Bailey, R.L., Clutter, J.L., 1974. Base-age invariant polymorphic site curves. *For. Sci.* 20, 155–159.
- Begin, J., Schutz, J.P., 1994. Estimation of total yield of Douglas-fir by means of incomplete growth series. *Ann. Sci. For.* 51, 345–355.
- Bennet, F.A., McGee, C.E., Clutter, J.L., 1959. Yield of old-field slash pine plantations. USDA Forest Service, South-east Forest Experiment Station Paper 107.
- Borders, B.E., Bailey, R.L., Ware, K.D., 1984. Slash pine site-index from a polymorphic model by joining (splining) nonpolynomial segments with an algebraic difference method. *For. Sci.* 30, 411–423.
- Borders, B.E., Bailey, R.L., Clutter, M.L., 1988. Forest growth models: parameter estimation using real growth series. In: Ek, A.R., Shifley, S.R., Burk, T.E. (Eds.), *Forest Growth Modeling and Prediction*. USDA Forest Series General Technical Report NC-120, pp. 660–667.
- Cao, Q.V., 1993. Estimating coefficients of base-age invariant site index equations. *Can. J. For. Res.* 23, 2343–2347.
- Cao, Q.V., Baldwin, V.C., Lohrey, R.E., 1997. Site index curves for direct-seeded loblolly and longleaf pines in Louisiana. *North J. Appl. For.* 21, 134–138.
- Carvalho, A., Falcão, A., Carvalho, J.L., Coucelo, F., 1994. TWIGGY um modelo de produção para a Portucel. Twin globulus growth yield. In: Páscoa, F. (Ed.), *Proceedings of the III Congresso Florestal Nacional. Os recursos no desenvolvimento rural*, pp. 222–230.
- Cieszewski, C.J., Bailey, R.L., 2000. Generalized algebraic difference approach: theory based derivation of dynamic equations with polymorphism and variable asymptotes. *For. Sci.* 46 (1), 116–126.
- Cieszewski, C.J., Bella, I.E., 1989. Polymorphic height and site index curves for lodgepole pine in Alberta. *Can. J. For. Res.* 19, 1151–1160.

- Clutter, J.L., Forston, J.C., Piennar, L.V., Brister, G.H., Bailey, R.L., 1983. *Timber Management—A Quantitative Approach*. Wiley, New York, p. 333.
- Dyer, M., Bailey, R.L., 1986. A test of six methods for estimating tree heights from stem analyses data. *For. Sci.* 33, 3–13.
- Elfving, B., Kiviste, A., 1997. Construction of site index equations for *Pinus sylvestris* L. using permanent sample plots data in Sweden. *For. Ecol. Manage.* 98, 125–134.
- Fox, J., Ades, P., Bi, H., 2001. Stochastic structure and individual-tree growth models. *For. Ecol. Manage.* 154, 261–276.
- García Abejón, J.L., 1981. Tablas de producción de densidad variable para *Pinus sylvestris* en el Sistema Ibérico. *Comunicaciones I.N.I.A Serie: Recursos Naturales*, No. 10, p. 47.
- García Abejón, J.L., Gómez Loranca, J.A., 1984. Tablas de producción de densidad variable para *Pinus sylvestris* en el sistema Central. *Comunicaciones I.N.I.A Serie: Recursos Naturales*, No. 29, p. 36.
- García Abejón, J.L., Tella Ferreiro, G., 1986. Tablas de producción de densidad variable para *Pinus sylvestris* L. en el Sistema Pirenaico. *Comunicaciones I.N.I.A. serie: Recursos Naturales*, N-43, p. 28.
- Huang, S., 1999. Development of compatible height and site index models for young and mature stands within an ecosystem-based management framework. In: Amaro, A., Tomé, T. (Eds.), *Empirical and Process Based Models for Forest Tree and Stand Growth Simulation*. Edições Salamandra, pp. 61–98.
- Lappi, J., Bailey, R.L., 1988. A height prediction model with random stand and tree parameters: an alternative to traditional site-index methods. *For. Sci.* 34, 907–927.
- Madrigal, A., Álvarez González, J.G., Rodríguez Soalleiro, R., Rojo, A., 1999. Tablas de producción para los montes españoles. Fundación Conde del Valle de Salazar, Madrid, 253 pp.
- McDill, M.E., Amateis, R.L., 1992. Measuring forest site quality using the parameters of a dimensionally compatible height growth function. *For. Sci.* 38, 409–429.
- Montgomery and Peck, 1982. *Introduction to Linear Regression Analysis*. Wiley, New York, pp. 430–434.
- Palahí, M., Pukkala, Miina, J., Montero, G., 2003. Individual-tree growth and mortality models for Scots pine in north-east Spain. *Ann. For. Sci.*, in press.
- Piennar, L.V., Turnbull, K.J., 1973. The Chapman–Richards generalization of Von Bertalanffy's growth model for basal area growth and yield in even-aged stands. *For. Sci.* 19, 2–22.
- Pita, P.A., 1964. La calidad de la estación en las masas de *Pinus sylvestris* de la Península Ibérica. *Anales del Instituto Forestal de Investigaciones y experiencias*, No. 9, pp. 5–28.
- Rennolls, K., 1995. Forest height growth modeling. *For. Ecol. Manage.* 71, 217–225.
- Rojo, A., Montero, G., 1996. El pino silvestre en la Sierra de Guadarrama. Centro de publicaciones del Ministerio de Agricultura, Pesca y Alimentación, p. 293.
- Schumacher, F.X., 1939. A new growth curve and its application to timber yield studies. *J. For.* 37, 819–820.
- Seber, G.A.F., Wild, C.J., 1989. *Nonlinear Regression*. Wiley, New York, p. 768.

- 677 Soares, P., Tomé, M., Skovsgaard, J.P., Vanclay, J.K., 1995. Evaluating
678 a growth model for forest management using continuous forest
679 inventory data. *For. Ecol. Manage.* 71, 251–265.
- 680 SPSS, 1999. SPSS Base System Syntax Reference Guide. Release
681 9.0. SPSS Inc.
- 682 Tomé, M., 1988. Modelação do crescimento da árvore individual
683 em povoamentos de eucalyptus globulus Labill. (1ª rotação).
684 Região Centro de Portugal. Ph.D. Dissertation. Technical
685 University of Lisbon.
- 686 Tomé, M., Soares, P., 1999. A comparative evaluation of three
687 growth models for eucalypt plantation management in coastal
688 Portugal. In: Amaro, A., Tomé, T. (Eds.), *Empirical and Process*
689 *Based Models for Forest Tree and Stand Growth Simulation.*
690 *Edição Salamandra*, pp. 517–533.
- 691 West, P.W., 1995. Application of regression analysis to inventory
692 data with measurements on successive occasions. *For. Ecol.*
693 *Manage.* 71, 227–234.

UNCORRECTED PROOF