

Ecoregional site index models for *Pinus pinaster* in Galicia (northwestern Spain)

Juan Gabriel ÁLVAREZ GONZÁLEZ*, Ana Daría RUÍZ GONZÁLEZ, Roque RODRÍGUEZ SOALLEIRO, Marcos BARRIO ANTA

Departamento de Ingeniería Agroforestal, Universidad de Santiago de Compostela, Campus Universitario s/n, 27002 Lugo, Spain

(Received 30 October 2003; accepted 6 April 2004)

Abstract – Ten algebraic difference equations were used to develop site index models for even-aged stands of *Pinus pinaster* in two ecoregions of Galicia (northwestern Spain). Data from 204 stem analyses were obtained and a data structure involving all possible growth intervals was used to fit the equations. Generalized nonlinear least square methods were applied to take into account the error structure. Autocorrelation was corrected expanding the error term to allow a first-order autoregressive model adequate for the data structure. Different weighting factors were employed to satisfy the equal error variance assumption. Bias, root mean square error and Akaike's information criterion were calculated and cross-validation residuals were used to evaluate the performance of the equations. Ecoregional differences in the site index models were analysed using the non-linear extra sum of squares method and Lakkis-Jones test. The parameters of the models were significantly different between ecoregions. Relative error in site index predictions was used to select 20 years as the best reference age. Based on the analysis, an algebraic difference equation derived from the base model of Chapman-Richards with a different set of parameters for each ecoregion can be recommended. This model is polymorphic and with multiple asymptotes. It provides compatible site index and height growth estimates.

site index model / ecoregion-based / *Pinus pinaster* / generalized nonlinear regression

Résumé – Modèles écorégionaux de site index pour *Pinus pinaster* en Galice (nord-ouest de l'Espagne). Dix équations en différences algébriques ont été utilisées pour développer des courbes de croissance pour futaies régulières de pin maritime en deux éco-régions de la Galice (nord-ouest de l'Espagne). Les données utilisées pour ajuster les équations proviennent d'analyse de tiges de 204 arbres dominants avec une structure de tous les intervalles de croissance possibles. Les méthodes des minima quadratiques généralisés ont été considérées pour tenir en compte la structure des erreurs. On a corrigé l'auto-corrélation avec un terme additionnel de l'erreur qui donne un modèle autorégressif de premier ordre qui s'adapte à la structure des données. Différents facteurs de pondération ont été employés pour satisfaire l'hypothèse de variance semblable. Biais, erreur moyenne quadratique et le critère d'information d'Akaike ont été calculés et les résidus de la validation croisée ont été utilisés pour évaluer le comportement des équations. On a analysé les différences des modèles de croissance entre éco-régions avec la méthode de la somme additionnelle des carrés des résidus et le test de Lakkis-Jones. Les paramètres des modèles sont significativement différents entre éco-régions. L'erreur relative pour la prédiction de l'indice de station a été employée pour sélectionner 20 années comme l'âge de référence optimale. Une équation en différences algébriques dérivée du modèle de Chapman-Richards avec un ensemble différent de paramètres pour chaque éco-région est proposée d'après les résultats. Le modèle est polymorphe et avec de multiples asymptotes. Il restitue des estimations compatibles des indices de station et des hauteurs dominantes.

modèle de site index / écorégion / *Pinus pinaster* / régression généralisée non linéaire

1. INTRODUCTION

Maritime pine is the most important coniferous species of northern Spain, where more than 650 000 ha of pure or mixed stands are present, derived both from plantations or natural regeneration. Its wide distribution and variety of growing sites have made *Pinus pinaster* a species of high relevance in Galician forestry with more than 2.4 million cubic meters of roundwood produced each year [56].

The silviculture of this species reached importance three centuries ago, when agricultural landowners started to sow pine nuts of Portuguese provenance in their intermittently worked rye lands. The grain was harvested and the pine seedlings were left to grow for a very short rotation [51]. This type of culture and the special ability of maritime pine to regenerate naturally, especially after burning, lead to a rapid expansion and naturalization in the coastal areas. Another important factor promoting the expansion was the intensive afforestation program developed

* Corresponding author: algonjg@lugo.usc.es

Table I. Summary of some stand-level variables for the sample data used for fitting site index equations for *Pinus pinaster* in Galicia (north-western Spain).

Ecoregion	Statistic	Age (years)	Density (stems/ha)	Quadratic mean diam. (cm)	Basal area (m ² /ha)	Dominant height (m)	Site index (m)*
Coast	Mean	19.34	1261.17	20.26	34.11	13.94	15.77
	Maximum	39.00	3237.00	35.01	56.53	24.03	21.89
	Minimum	8.00	423.00	6.33	5.86	4.67	7.96
	CV%	39.19	52.85	32.77	33.36	29.77	21.18
Interior	Mean	19.26	1763.38	14.86	26.28	10.64	11.94
	Maximum	50.00	3142.00	33.88	72.48	24.58	17.03
	Minimum	9.00	363.00	5.16	5.12	4.55	6.72
	CV%	48.03	38.66	45.02	50.11	40.98	22.53
All combined	Mean	19.30	1522.96	17.45	30.03	12.22	13.77
	Maximum	50.00	3237.00	35.01	72.48	24.58	21.89
	Minimum	8.00	363.00	5.16	5.12	4.55	6.72
	CV%	43.88	47.15	41.15	43.02	37.32	25.93

* Site index was calculated as the height of each tree at the reference age of 20 years for all trees exceeding this age.

by the Forest Administration on Communal Lands from 1940 to 1970, that took the pine to the interior areas by using not well adapted provenances.

Nowadays, maritime pine populations from Galicia show a high level of genetic diversity due to the use of seed from different origins. This lack of genetic homogeneity coupled with a genotype-by-environmental interaction which favours the adaptation to local ecological conditions [1, 2] is causing important differences in the growth pattern among ecoregions.

To solve this problem, it is necessary to adopt the principles of ecologically based forest management. Therefore, the development of growth and yield models should be based on the ecoregion classification system developed by Vega et al. [55] for *Pinus pinaster* in Galicia. This system differentiates the interior and the coastal ecoregions based on both environmental conditions and seed origin.

The growth and yield of an even-aged stand is mainly determined by the productive capacity of the growing site, which includes many variables that collectively determine the site quality [25]. Considerable effort has been devoted to the development of methods for quantifying site quality. For most species, dominant height growth is independent of stocking over a quite wide range of stand density, thus is often used as a measure of site quality [41]. Site index, defined as the height of trees that have always been dominant or codominant and healthy at a reference age, is the most widely used method of site quality evaluation for even-aged forest stands [14]. Therefore, reliable height prediction based on unbiased and accurate site index models is essential on growth and yield models.

The objective of this paper is to develop ecoregion-based site index models for *Pinus pinaster* in Galicia and to compare the differences of dominant height growth between the two ecoregions.

2. MATERIALS AND METHODS

2.1. Data set

A total of 102 permanent sample plots of even-aged *Pinus pinaster* stands were used in this study. These plots were subjectively selected throughout the inventory areas of Galicia to provide representative information of site quality, age and stand density. From these, 52 sample plots (50.98%) were located in the coast ecoregion and the rest in the interior ecoregion. Two dominant trees were destructively sampled at each location. The trees were sectioned at the stump, at breast height and 2.0 m, and 1-meter intervals. The age at each section height was determined in the laboratory. As cross section lengths do not coincide with periodic height growth, height values at 2 year-increments were estimated using the method of Carmean [13] with the modification proposed by Newberry [44] for the topmost section of the tree. A comparative study between six methods of height data correction in stem analysis showed that the Carmean algorithm had the best performance [22].

Summary statistics, including the mean, minimum, maximum, and coefficient of variation of the main stand variables for total plots and by ecoregion are shown in Table I. Site index was calculated as the height of each tree at the reference age of 20 years for all trees exceeding this age.

2.2. Equations considered

The most important desirable attributes of site index equations are: (1) a logical behavior (height should be zero at age zero and equal to site index at reference age), (2) a sound theoretical basis, (3) polymorphism, (4) asymptote is a function of site index (increases with increasing site index), (5) existence of an inflection point and (6) base-age invariance [4, 24, 26, 46]. Whether or not these requirements can be met depends on both, the construction method and the mathematical function used to develop the curves. According to Clutter et al. [20] most of the approaches used to fit site index curves can be viewed as

Table II. Algebraic difference models used in this study.

No.	Algebraic difference models from differential equations	Differential equation
M1	$H_2 = \exp(\ln(H_1) \cdot (t_1 / t_2)^{b_1} \cdot \exp[b_0 \cdot (1 / t_2 - 1 / t_1)])$	$d \ln(H) / d(1 / t) = b_0 \cdot \ln(H) + b_1 \cdot \ln(H) \cdot t$
M2	$H_2 = \exp(b_0 + b_1 / t_2 + [\ln(H_1) - b_0 - b_1 / t_1] \cdot z)$ with $z = \exp[b_2 \cdot (1 / t_2 - 1 / t_1)]$	$d \ln(H) / d(1 / t) = \alpha + \beta \cdot \ln(H) + \delta / t$ $b_0 = -(\alpha + \delta / \beta) / \beta; b_1 = -\delta / \beta; b_2 = \beta$
M3	$H_2 = b_0 / [1 - (1 - b_0 / H_1) \cdot (t_1 / t_2)^{b_1}]$	$dH / dt = (1 - H / b_0) \cdot b_1 \cdot (H / t)$
M4	$H_2 = b_0 \cdot (H_1 / b_0)^{\exp(z)}$ with $z = [b_1 / (b_2 - 1) \cdot t_2^{(b_2 - 1)} - b_1 / (b_2 - 1) \cdot t_1^{(b_2 - 1)}]$	$dH / dt = \ln(b_0 / H) \cdot b_1 \cdot (H / t^{b_2})$
No.	Algebraic difference models from height-age equations	Height-age equation
M5	$H_2 = b_0 \cdot (1 - [1 - (H_1 / b_0)^{1 / b_2}]^{t_2 / t_1})^{b_2}$	Chapman-Richards $H = b_0 \cdot [1 - \exp(-b_1 \cdot t)]^{b_2}$ solved by b_1
M6	$H_2 = b_0 \cdot (H_1 / b_0)^{\ln[1 - \exp(-b_1 \cdot t_2)] / \ln[1 - \exp(-b_1 \cdot t_1)]}$	Chapman-Richards $H = b_0 \cdot [1 - \exp(-b_1 \cdot t)]^{b_2}$ solved by b_2
M7	$H_2 = b_0 \cdot (H_1 / b_0)^{(t_1 / t_2)^{b_2}}$	Korf $H = b_0 \cdot \exp(-b_1 / t^{b_2})$ solved by b_1
M8	$H_2 = b_0 \cdot \exp(-b_1 / t_2^z)$ with $z = \ln[-b_1 / \ln(H_1 / b_0)] / \ln(t_1)$	Korf $H = b_0 \cdot \exp(-b_1 / t^{b_2})$ solved by b_2
No.	Algebraic difference models from height-age equations by relating parameters with H_1 , t_1 or S	Height-age equation
M9	$H_2 = H_1 \cdot ([1 - \exp(-z \cdot t_2)] / [1 - \exp(-z \cdot t_1)])^{b_2}$ with $z = b_3 \cdot (H_1 / t_1)^{b_4} \cdot t_1^{b_5}$	Chapman-Richards $H = b_0 \cdot [1 - \exp(-b_1 \cdot t)]^{b_2}$ solved by b_0 and assuming $b_1 = b_3 \cdot (H_1 / t_1)^{b_4} \cdot t_1^{b_5}$
M10	$H_2 = (H_1 + d + r) / [2 + (4 \cdot b_3 / t_2^{b_2}) / (H_1 - d + r)]$ with $d = b_3 / Asi^{b_2}$ and $r = \sqrt{(H_1 - d)^2 + 4 \cdot b_3 \cdot H_1 / t_1^{b_2}}$	Hossfeld IV $H = b_0 / (1 + b_1 / t^{b_2})$ solved by b_0 and assuming $b_1 = b_3 / S$

H_1 and H_2 are dominant height (m) at age t_1 and t_2 (years), respectively; Asi is an age ranged from 5 to 50 years to reduce the mean square error; \ln is natural logarithm and b_0 , b_1 , b_2 , b_3 , b_4 and b_5 are parameters to be estimated.

special cases of three general development techniques: (1) the guide-curve method, (2) the parameter-prediction method, and (3) the difference-equation method.

The guide curve method assumes proportionality (anamorphism) among curves for different site quality and is used to generate a set of anamorphic site index curves. This method has the disadvantage that correlation between site index and stand age may disturb the statistical analyses [38], and this correlation is very common when data are derived from stem analysis [35].

The parameter prediction method is based on fitting a growth function tree-by-tree or plot-by-plot and relating the parameters of the fitted curves to site index (e.g. [23, 45, 47]). The height-over-age series are generally obtained from stem analysis or from long-term growth trials.

The difference equation method is based on the fact that observations of the same plot or dominant tree should belong to the same site index curve. A difference algebraic form of a height-age or differential equation is developed where height at remeasurement (H_2) is expressed

as a function of the remeasurement age (t_2), the initial age (t_1) and the height at the initial measurement (H_1). The algebraic difference form is obtained through substitution of one parameter in the height-age or differential equation [24].

The advantages of the difference equation method in comparison with the parameter prediction method are: (1) short observation periods of temporary plots or stem analysis data from trees whose total age was under the reference age can be used, (2) the curves pass through site index at the reference age and, (3) the equations can be base-age invariant so the height at any age can be predicted given the height at any other age [6, 12, 17, 20]. The difference equation method has been widely used to develop site index curves (e.g. [4, 8, 11, 18, 37]) and it will be used in this study.

A total of 10 algebraic difference models were selected for evaluation from those most commonly used in forest research (Tab. II). The models were classified in three groups depending on the approach used to derive them: (1) Models from differential equations, (2) Models from height-age equations and (3) Models from height-age equations by relating parameters with S , H_1 and/or t_1 .

Models M1 to M4 belong to the first group and they were formulated based on the differential equations proposed by Amateis and Burkhart [3], Clutter and Lenhart [19], McDill and Amateis [40], and Sloboda [53], respectively. Models M5 to M8 belong to the second group and they were formulated based on the well-known height-age equations of the Chapman-Richards generalization of Bertalanffy [15, 50] and Korf (cited by Lundqvist [39]). Model M10 was proposed by Cieszewski and Bella [17] from the height-age equation of Hossfeld IV (cited by Peschel [48]) by relating a model parameter to site index. Model M9, proposed by Goelz and Burk [26], was formulated from the height-age equation of Chapman-Richards by relating a model parameter to H_1 and t_1 .

These algebraic difference equations are base-age invariant (except equation M9), polymorphic and the models M2, M9 and M10 have multiple asymptotes. All the models have been widely used to develop height-age curves (e.g. [12, 16, 24, 27, 32, 42, 46, 54]).

2.3. Data structure

The data structure used for fitting the difference algebraic models was arranged with all the possible combinations among height-age pairs for each tree, including descending growth intervals. All possible intervals may lead to the rejection of the error assumptions but, on the other hand, will produce fitted models with a better predictive performance [26, 29, 32].

The potential problem of heteroscedasticity and lack of independence among observations can be solved using generalised nonlinear least squares (GNLS) methods [26, 31, 41]. In this case, autocorrelation was modelled as a first-order autoregressive process where the error term was expanded to represent the autocorrelation structure inherent in fitting site index models to an all possible growth intervals data structure [26, 27, 46]:

$$H_{ij} = f(H_j, t_i, t_j, \beta) + e_{ij}$$

$$\text{with } e_{ij} = \rho \cdot e_{i-1,j} + \gamma \cdot e_{i,j-1} + \varepsilon_{ij} \quad (1)$$

where H_{ij} represents prediction of height i by using height j , age t_i and age t_j as predictor variables; ρ is a parameter that accounts for the autocorrelation between the current residual and the residual from estimating H_{i-1} using H_j as a predictor; γ is a parameter which accounts for the autocorrelation between the current residual and the residual from estimating H_i using H_{j-1} as a predictor; and ε_{ij} are independently and identically distributed errors.

To avoid the problem of heteroscedasticity the error variance was assumed to be a power function of the predicted dominant height [32, 33]. The weighting factors used were $\text{weight}_i = \text{pred}.h_i^k$, where k is a constant (e.g. $k = -2, -3/2, -1, -1/2, 1/2, 1, 3/2, 2$). Since the predicted dominant heights are initially unknown, weighting is an iterative process.

All the models were fitted to the total data and to each ecoregion separately. The fittings were done by modelling the mean and the error structure simultaneously using the MODEL procedure in the SAS/ETS system [52]. For the M10 model, the parameter As_i was ranged from 5 to 50 years to reduce the mean square error [24, 54].

When using the all possible growth intervals data structure, the number of observations is increased considerably, although no additional information is obtained. Thus, the resulting standard errors for the parameters estimates would be too small. The standard errors should be expanded by $\sqrt{n(\text{apd})/n(\text{fd})}$ where $n(\text{apd})$ is the number of observations using all possible differences and $n(\text{fd})$ is the number of observations if using only first differences [27].

2.4. Model comparison and cross-validation

The accuracy and precision of dominant height estimates of each model were compared using graphic and numeric analysis of the residuals (e_i). The plots of studentized residuals against the predicted dominant

height were examined for detection of possible systematic discrepancies and to select the weighting factor [43]. Also, three statistical criteria obtained from the residuals were examined: bias (\bar{E}); root mean square error ($RMSE$) and the adjusted coefficient of determination (R^2_{adj}). These expressions may be summarized as follows:

$$\text{Bias } \bar{E} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)/n}{n} \quad (2)$$

$$\text{Root mean square error } RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n-p)}{n-p}} \quad (3)$$

Adjusted coefficient of determination

$$R^2_{adj} = 1 - (n-1) \cdot \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n-p)}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

where $e_i = y_i - \hat{y}_i$, and y_i , \hat{y}_i and \bar{y} are the measured, predicted and average values of the dependent variable, respectively; n is the total number of observations used to fit the model and p is the number of model parameters.

Akaike's information criterion differences ($AICd$), which is an index to select the best model based on minimising the Kullback-Liebler distance, was used in order to compare models with a different number of parameters [10]:

$$AICd = n \cdot \ln \hat{\sigma}^2 + 2 \cdot (p+1) - \min(n \cdot \ln \hat{\sigma}^2 + 2 \cdot (p+1)) \quad (5)$$

where p , is the number of parameters of the model and $\hat{\sigma}^2$ is the estimator of the error variance of the model: $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}$.

Finally, a cross-validation approach was used to evaluate the prediction performance of the models. The bias, root mean square error (RMSE) and model efficiency of the estimates (ME), calculated by equation (4) were estimated using the residuals for fitting the model to a new data set obtained by deleting the observations of the tree i from the original data set. Also, plots of the studentized residuals against the predicted dominant height and plots showing the observed against the predicted dominant heights in cross-validation were analysed to detect systematic trends.

2.5. Comparison of site index models between ecoregions

To compare the differences of site index models between ecoregions, two tests for detecting simultaneous homogeneity among parameters were used: the non-linear extra sum of squares method [5] and the χ^2 test proposed by Lakkis and Jones [36]. These tests are frequently applied to analyse differences among different geographic regions [12, 30, 33, 49].

Both methods require the fitting of reduced and full models. The reduced model corresponds to the same set of parameters for the two ecoregions. The full model corresponds to different sets of parameter for each ecoregion and it is obtained by expanding each parameter, including an associated parameter and a dummy variable to differentiate the two ecoregions:

$$b_i + c_i \cdot I \quad i = 0, \dots, 5 \quad (6)$$

where b_i is a parameter of the models M1 to M10; c_i is the associated parameter of the full model and I is a dummy variable the value of which is equal to 0 for the interior ecoregion and 1 for the coastal ecoregion. The appropriate test statistics are given by:

Non-linear extra sum of squares

$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} \quad (7)$$

Table III. Parameter estimates and related statistics obtained for each ecoregion using the ten algebraic difference equations.

Model	Ecoregion	Parameter estimate						Fit				Cross-validation				
		b_0	b_1	b_2	b_3	b_4	b_5	Asi	R^2_{adj}	Bias	$RMSE$	$AICd$	ME	Bias	$RMSE$	$AICd$
M1	Interior	-2.9678	-0.1757						0.8972	0.5416	1.8939	23684.32	0.7550	0.9439	2.9231	24096.99
	Coast	-3.2962	-0.1767						0.8796	0.5126	1.8068	24223.67	0.7137	1.1034	2.7856	24179.96
M2	Interior	-0.3077	-22.4352	2.3344					0.9747	-0.0037	0.9398	5961.01	0.9454	-0.0433	1.3799	5111.88
	Coast	0.7381	-19.9880	2.5592					0.9834	-0.0020	0.6715	4614.46	0.9669	-0.0060	0.9473	2813.35
M3	Interior	31.8130	1.3683						0.9832	0.0017	0.7655	770.75	0.9615	-0.0947	1.1587	689.80
	Coast	32.2340	1.4442						0.9886	0.0108	0.5570	909.38	0.9751	0.0754	0.8220	0.00
M4	Interior	55.7048	0.3027	0.8082					0.9831	-0.0090	0.7687	877.07	0.9595	-0.1423	1.1885	1333.84
	Coast	44.3157	0.3001	0.6957					0.9896	-0.0012	0.5320	0.00	0.9722	0.1004	0.8687	1096.89
M5	Interior	25.8547		1.3784					0.9830	0.0013	0.7699	913.90	0.9612	-0.0967	1.1637	798.05
	Coast	25.7546		1.4979					0.9885	0.0112	0.5588	972.47	0.9751	0.0803	0.8222	4.1705
M6	Interior	36.1407	0.0224						0.9796	-0.0003	0.8444	3252.04	0.9485	-0.1339	1.3400	4365.92
	Coast	32.1946	0.0396						0.9878	0.0025	0.5759	1572.27	0.9611	0.1208	1.0271	4413.04
M7	Interior	147.2046		0.3203					0.9817	0.0124	0.7995	1867.94	0.9573	-0.0911	1.2202	1998.09
	Coast	271.9729		0.2889					0.9868	0.0206	0.5980	2317.03	0.9693	0.1191	0.9117	2052.09
M8	Interior	51.3595	5.7809						0.9578	0.1442	1.2134	12423.21	0.9055	0.2754	1.8150	12042.56
	Coast	74.8696	5.8181						0.9560	0.1590	1.0928	14261.21	0.9030	0.1875	1.6211	13454.80
M9	Interior			1.4214	0.1351	0.9819	-0.2104		0.9842	0.0913	0.7425	0.00	0.9636	0.0993	1.1274	0.00
	Coast			1.5019	0.1643	1.0734	-0.3071		0.9889	0.0806	0.5483	598.22	0.9746	0.0973	0.8299	192.67
M10	Interior			1.3686	579.35			10	0.9826	0.0173	0.7786	1197.78	0.9598	-0.0545	1.1835	1225.31
	Coast			1.4191	683.10			10	0.9876	0.0220	0.5792	1684.74	0.9733	0.0900	0.8502	668.96

Lakkis-Jones test

$$L = (SSE(F) / SSE(R))^{n/2} \quad (8)$$

where $SSE(R)$ is the error sum of square of the reduced model; $SSE(F)$ is the error sum of squares of the full model; df_R and df_F are the degrees of freedom of the reduced and full model, respectively; $-2 \cdot \ln(L)$ follows a χ^2 -distribution with $v = df_R - df_F$ degrees of freedom and F^* follows an F -distribution.

If the homogeneity of parameters test reveals significant differences between ecoregions, three different approaches can be used to model the site index curves: (1) to use the reduced model, (2) to use the full model and (3) to use different models for each ecoregion. To determine which was better, the accuracy and precision by age classes of height and site index predictions of cross-validation were calculated. Also, the relative errors ($RE\%$) and the critical errors (E_{crit}) in predictions were obtained according to Equations (9) and (10), respectively [34].

$$RE\% = 100 \cdot \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n-p) / \bar{y}}{}} \quad (9)$$

$$E_{crit} = \sqrt{\tau^2 \cdot \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / \chi_{crit}^2 / \bar{y}}{}} \quad (10)$$

where y_i , \hat{y}_i and \bar{y} are the observed, predicted in cross-validation and average values of the dependent variable, respectively; n is the total number of observations; p is the number of model parameters and τ and χ_{crit}^2 are a standard normal deviate and a χ^2 -distribution with n degrees of freedom at the specific probability level, respectively.

3. RESULTS AND DISCUSSION

3.1. Model comparison

At first, all the models were fitted to each ecoregion without the autocorrelation parameters using weighted least squares. The residuals were related using the hypothesized autoregressive error structure to test autocorrelation using the Durbin's t -test [21]. The test showed that the residuals were highly correlated for all the models and ecoregions. All the models were refitted, this time modelling the error structure using generalized non-linear least square and the results of fitting and cross-validation for each ecoregion and model are shown in Table III. All the parameters were found significant at a 5% level when the expansion factor proposed by Goelz and Burk [27] was applied.

Table IV. Parameter estimates and related statistics for the reduced model and the full model obtained using the ten algebraic difference equations.

Model	Parameter estimate							Associated parameter estimate					Cross-validation			
	b_0	b_1	b_2	b_3	b_4	b_5	Asi	c_0	c_1	c_2	c_3	c_4	c_5	Bias	RMSE	AICd
M1	Reduced	-3.1539	-0.1694											2.9139	50731.73	
	Full	-2.9801	-0.1751					-0.2964	-0.0025 ^a					0.9392	2.9012	51756.1
M2	Reduced	-0.3077	-22.4352	2.3344										0.0104	1.1381	8074.93
	Full	0.0556	-22.6363	2.5287				0.5917	2.4817	-0.0017 ^a				-0.0376	1.1131	8545.52
M3	Reduced	31.3251	1.4134											0.0125	0.9561	469.98
	Full	31.8405	1.3672					0.3458 ^a	0.0784					-0.0173	0.9407	953.47
M4	Reduced	56.3899	0.3274	0.8236										-0.0244	0.9874	1917.46
	Full	55.7185	0.3020	0.8075				-11.5487	-0.0015 ^a	-0.1122				-0.0333	0.9306	468.97
M5	Reduced	25.6066		1.4386										0.0134	0.9560	457.81
	Full	25.8647		1.3771				-0.1293 ^a		0.1225				-0.0175	0.9421	1021.66
M6	Reduced	32.7780	0.0306											0.0005	1.1178	7511.90
	Full	36.1379	0.0224					-3.9713	0.0173					-0.0157	1.0080	4068.70
M7	Reduced	153.4297		0.3265										0.0039	1.0159	3199.65
	Full	147.8383		0.3197				123.0849		-0.0303				0.0038	0.9837	2967.27
M8	Reduced	57.4088	5.7494											0.2287	1.6181	24197.38
	Full	51.4071	5.7782					23.2856	0.0398 ^a					0.2268	1.6101	25196.21
M9	Reduced			1.5039	0.1445	0.8373	-0.2436							0.0624	0.9462	0.00
	Full			1.4202	0.1352	0.9831	-0.2108				0.0801	0.0276	0.0940	-0.0929	0.0984	0.9210
M10	Reduced			1.4052	620.48		10							0.0107	0.9799	1570.78
	Full			1.3674	578.48		10				0.0529	105.27		0.0112	0.9686	2273.21

^a Indicates not significance at a 5% level when the expansion term proposed by Goelz and Burk [27] is used.

In general, weightings factors of $w_i = 1/\text{pred} \cdot ht_i^{3/2}$ and $w_i = 1/\text{pred} \cdot ht_i^{1/2}$ showed the best results when plots of studentized residuals against the predicted dominant height were examined for detection of possible systematic discrepancies.

The values of the statistics used to compare the models indicate that all the models, except model M1 [3] produced a reasonable performance with small bias and root mean square error on both ecoregions for fitting and cross-validation. These results are consistent with those obtained by Cao [11] where the root mean square error of the Amateis and Burkhart model increased much more quickly than Clutter and Lenhart [19] and height-age equation-based models when the time projection length increased, indicating that the estimation capabilities of this model are strongly dependent on the data structure used.

The best results were obtained with equations M9 [26] and M3 [40] for interior and coast ecoregions, respectively. Although model M5 represented the data almost equally well as models M9 and M3 for both ecoregions.

The models derived from the Chapman-Richards and Korf height-age equations based on solving by parameter b_2 (M6 and M8) performed relatively poorly when compared with another models based on the same base equations (M5, M7 and M9). These results suggest that, for this species, the b_2 parameter of the Chapman-Richards and Korf height-age equations does not depend on site quality. Similar results were obtained by Beck [7], Graney and Burkhart [28], Burkhart and Tennent [9] and Goelz and Burk [26] using the Chapman-Richards equation.

3.2. Comparisons between ecoregions

All the models were fitted to both ecoregions combined using the same set of parameters (reduced model) and a different set of parameters for each ecoregion (full model) by expanding each one including a dummy variable to differentiate the two ecoregions using equation (6). The weighting factors used were the same that gave the best results when the models were

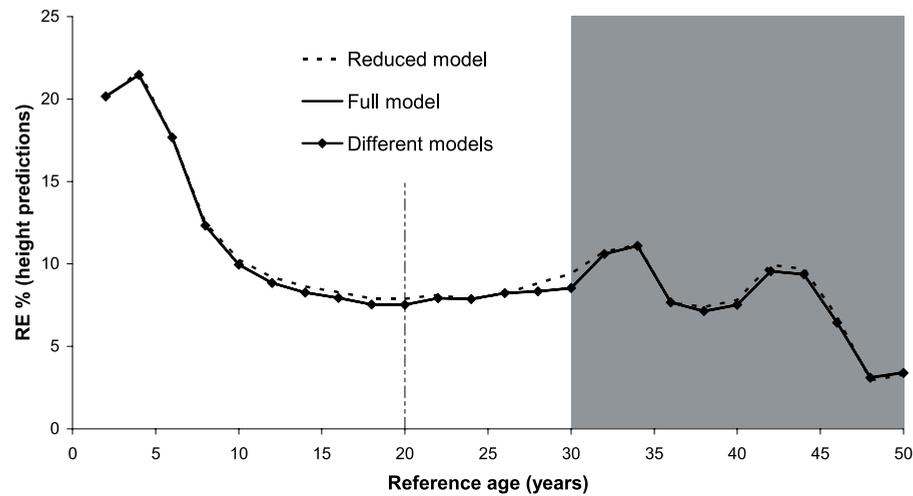


Figure 1. Relative error in dominant height predictions related to choice of reference age for reduced, full and different models for each ecoregion. The shadow zone is not representative due to the lack of trees at these ages (lower than 30 trees).

Table V. Results of the Lakkis and Jones (L -value) and non-linear extra sum of squares test (F -value) of the ecoregional differences for the ten algebraic difference models.

Model	Reduced model			Full model			n	L -value	F -value
	SSE	df	MSE	SSE	df	MSE			
M1	78865.60	22550	3.497	78171.92	22548	3.467	22554	199.26**	100.04**
M2	16195.79	22549	0.718	15665.35	22546	0.695	22554	751.05**	254.48**
M3	10858.40	22550	0.482	10511.11	22548	0.466	22554	733.14**	372.50**
M4	11606.79	22549	0.515	10309.53	22546	0.457	22554	2673.14**	945.66**
M5	10930.76	22550	0.485	10615.62	22548	0.471	22554	659.79**	334.68**
M6	15161.29	22550	0.672	12327.44	22548	0.547	22554	4666.82**	2591.69**
M7	12437.85	22550	0.552	11660.27	22548	0.517	22554	1456.02**	751.83**
M8	30842.47	22550	1.368	30536.56	22548	1.354	22554	224.82**	112.94**
M9	10533.60	22548	0.467	9978.23	22544	0.443	22554	1221.64**	313.69**
M10	11275.75	22550	0.500	11018.16	22548	0.489	22554	521.21**	263.57**

fitted to each ecoregion separately. The estimates of the parameters and the values of the statistics obtained in the cross-validations are shown in Table IV.

Again, models M9, M3 and M5 presented the highest accuracy and precision with the minimum value of the Akaike's information criterion for model M9 on both the reduced and the full approaches. A t -test indicated that the estimates of some associated parameters of the full models M1, M2, M3, M4, M5 and M8 were not significant at a level of 5% when the correction term proposed by Goelz and Burk [27] was used.

The values of the Lakkis-Jones test (see [36]) and non-linear extra sum of squares method [5] are presented in Table V. The results reveal that there are differences for all the site-index models between the two different ecoregions.

Three different approaches to develop the site index equations for the two ecoregions were compared. The first was to use the reduced model; the second was to use the full model;

the third was to use the best model for each ecoregion based on the results of fitting each one separately (models M9 and M3 for interior and coast ecoregion, respectively). To determine which approach was better, the accuracy and precision of height and site index predictions of cross-validation were compared by age classes.

The first step was to determine the best reference age to define the site index. In accordance with Goelz and Burk [26] the reference age should be selected taking into account three considerations: (1) the reference age should be less than or equal to the younger rotation age for common silvicultural treatments; (2) the base age should be close to the rotation age and (3) the base age should be selected such that it is a reliable predictor of height at other ages.

For each tree, the height at different reference ages was calculated using the other pairs dominant height-age of the same tree. The estimated heights were compared with the observed

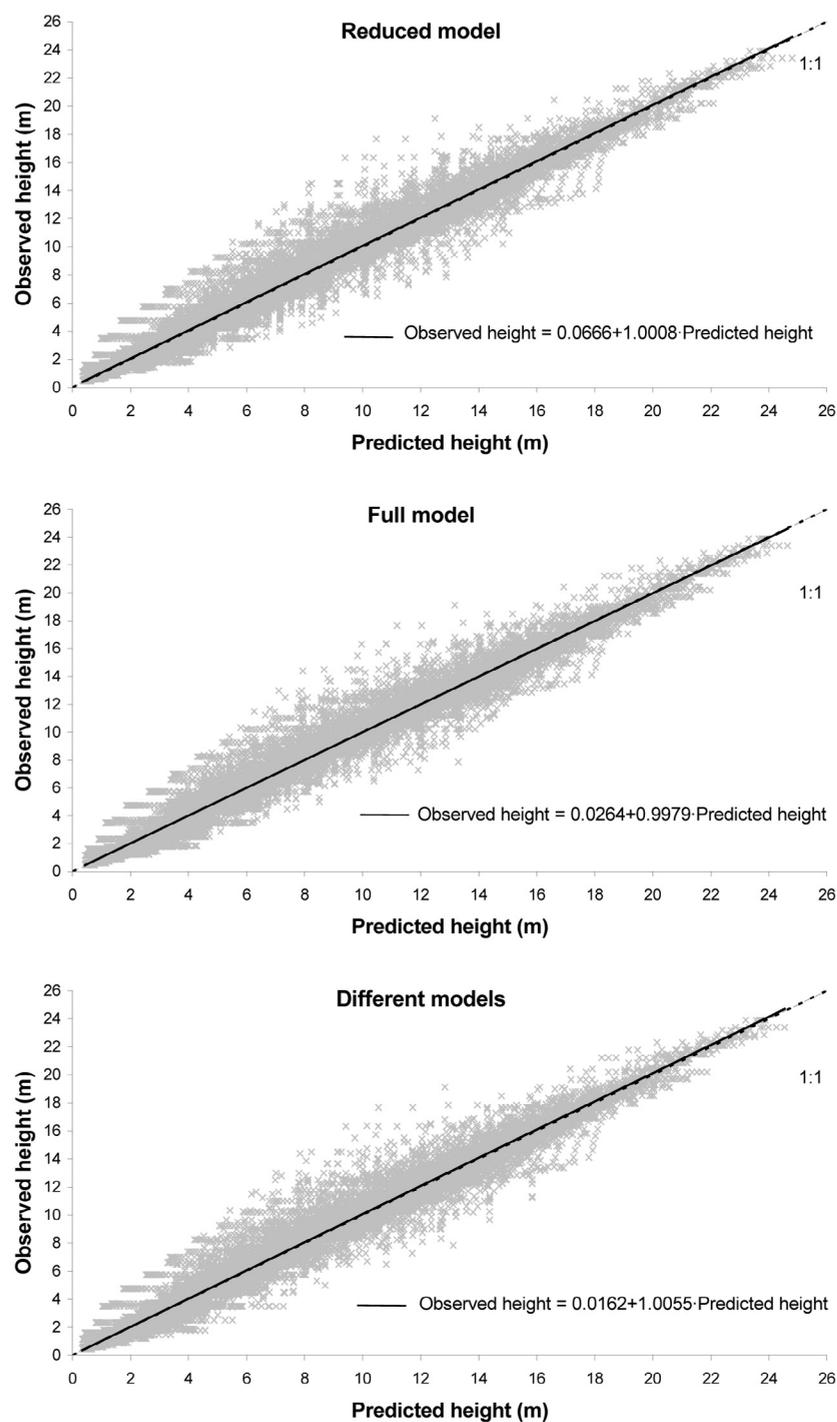


Figure 2. Observed against predicted dominant height obtained in cross-validation for the reduced model, the full model and a different model for each ecoregion. The solid line represents the linear equation fitted to the scatter plot of data and the dotted line is the diagonal.

heights from the stem analysis. The relative error in predictions ($RE\%$) calculated using equation (9) was used to select the best reference age. In Figure 1 the results for the three different approaches to develop the site index equations explained above, are displayed. The lowest relative error for the three approaches was found at a reference age of 20 years. Over 30 years the sample is not representative (less than 30 trees). Although, according

to Goelz and Burk [26] this selection procedure should be devised such that the error of predicting stand volume is minimised, the lack of necessary information forced us to conclude that a reference age of 20 years is appropriate for *Pinus pinaster* in Galicia.

In Figure 2 the observed dominant heights are plotted against the predicted dominant heights obtained in a cross-validation

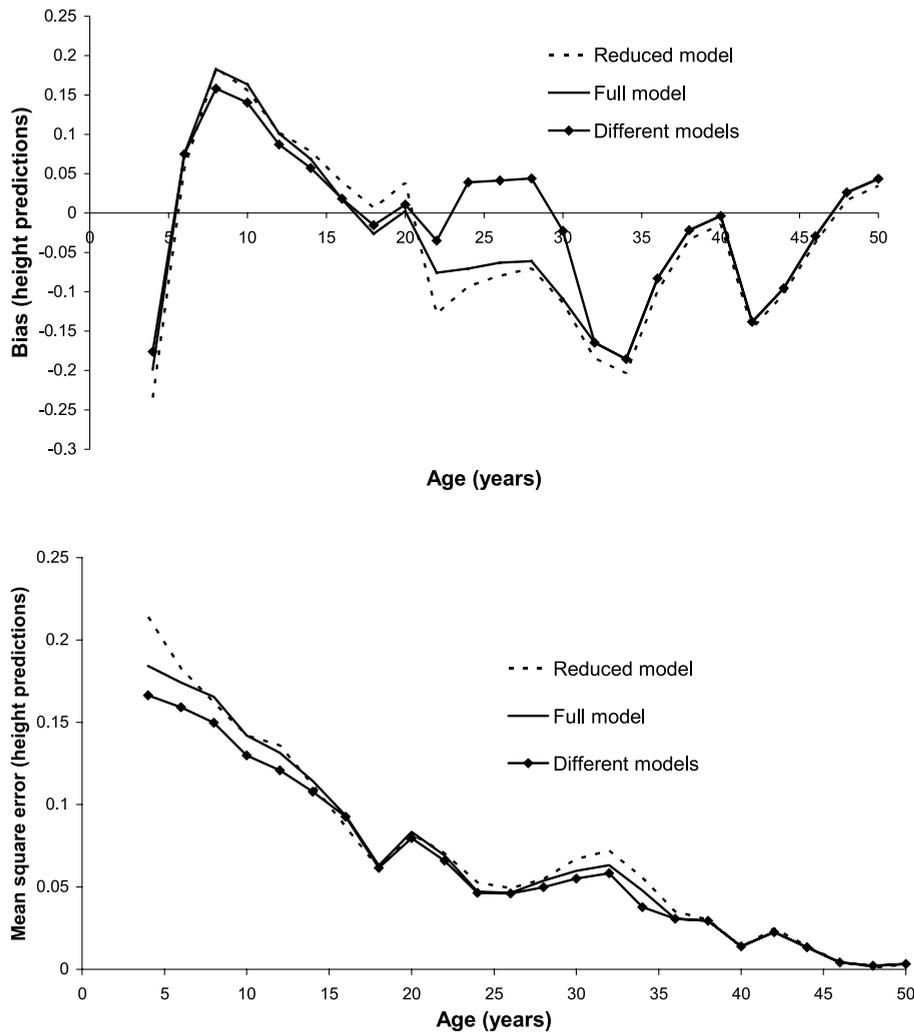


Figure 3. Bias and mean square error of height predictions obtained in cross-validation by age for the reduced model, the full model and for a different model for each ecoregion.

for the reduced model, the full model and a different model fitted to each ecoregion separately. A linear equation was fitted in each scatter plot to allow a comparison with the diagonal pattern [34]. The results show that all the approaches had a good overall performance.

In accordance with Huang [34], to evaluate the accuracy and precision of height and site index predictions, at a reference age of 20 years, plots of bias and mean square error obtained in the cross-validation across age of the three approaches were also compared (Figs. 3 and 4, respectively). It can be inferred that the reduced model performed worse than the other two approaches. The values of bias and root mean square error of height and site index predictions of the full model are very close to those obtained using different models for each ecoregion.

The critical errors (E_{crit}) were calculated using equation (10) for height and site index predictions of each approach. The best results were obtained using different models for each ecoregion with the lowest critical error for both the overall height prediction (11.77%) and the overall site index prediction (14.79%). However, the result of the full model for both height and site

index predictions were very close to those (11.78%, 14.84%, respectively) and this model presents the advantage of using a unique equation with an asymptote value changing with the site quality (model M3 has a unique asymptote). All these results suggest that a full model with a different set of parameters for each ecoregion based on the algebraic difference equation proposed by Goelz and Burk [26] is likely to be successful as a predictor.

Since site index is a fixed stand attribute which should be stable over time, a plot of the site index predictions against total age using the full model and the stem analysis data was developed (Fig. 5). The graph reveals the consistency of site index predictions over time except at young ages where the site index is underestimated for the higher site qualities and overestimated for the remaining site qualities.

Site indices of 6, 10, 14 and 18 m at a reference age of 20 years and 9, 13, 17 and 21 m at the same reference age for the interior and coast ecoregion, respectively, were used to develop the site index curves shown in Figure 6. These curves were plotted over the stem analysis data. For both ecoregions the curves are a realistic representation of the overall growth pattern of the stem analysis data.

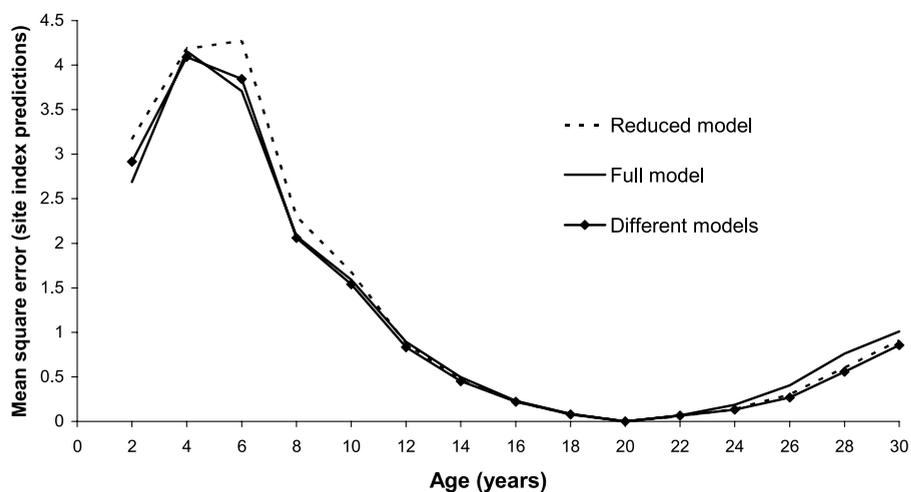
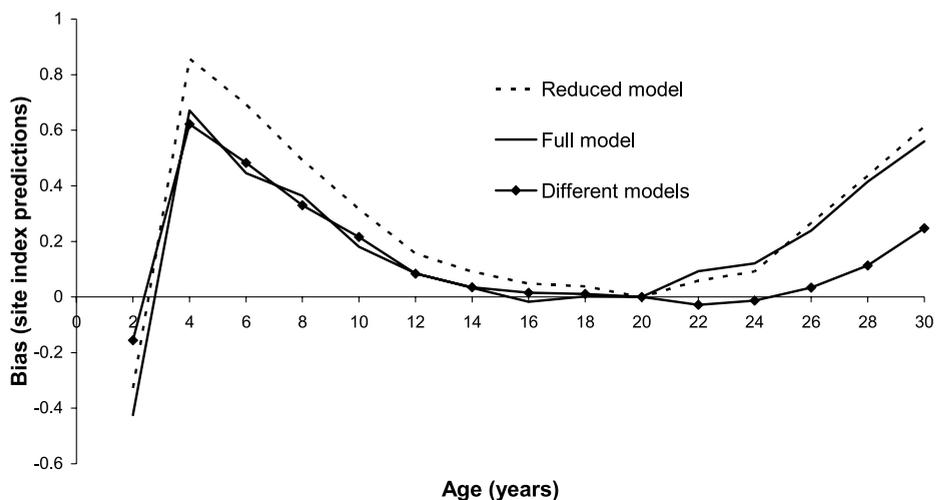


Figure 4. Bias and mean square error of site index predictions obtained in cross-validation by age for the reduced model, the full model and for a different model for each ecoregion.

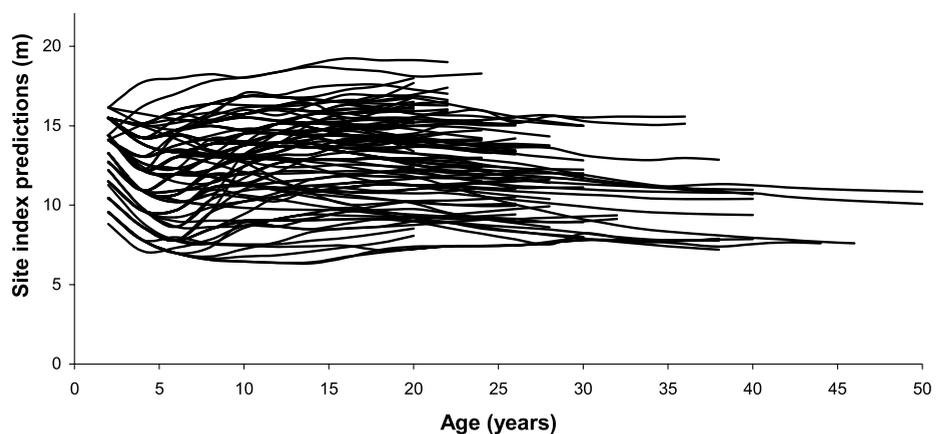


Figure 5. Site index predictions against total age using the full model and the stem analysis data.

The mathematical expression of the site index model for *Pinus pinaster* in Galicia is the following:

$$H_2 = H_1 \cdot ([1 - \exp(-z \cdot t_2)] / [1 - \exp(-z \cdot t_1)])^{(1.4202 + 0.0801 \cdot I)}$$

$$\text{with } z = (0.1352 + 0.0276 \cdot I) \cdot (H_1 / t_1)^{(0.9831 + 0.0940 \cdot I)} \cdot t_1^{(-0.2108 - 0.0929 \cdot I)} \tag{11}$$

where I is a dummy variable which assumes a value of 0 for the interior ecoregion and 1 for the coastal ecoregion.

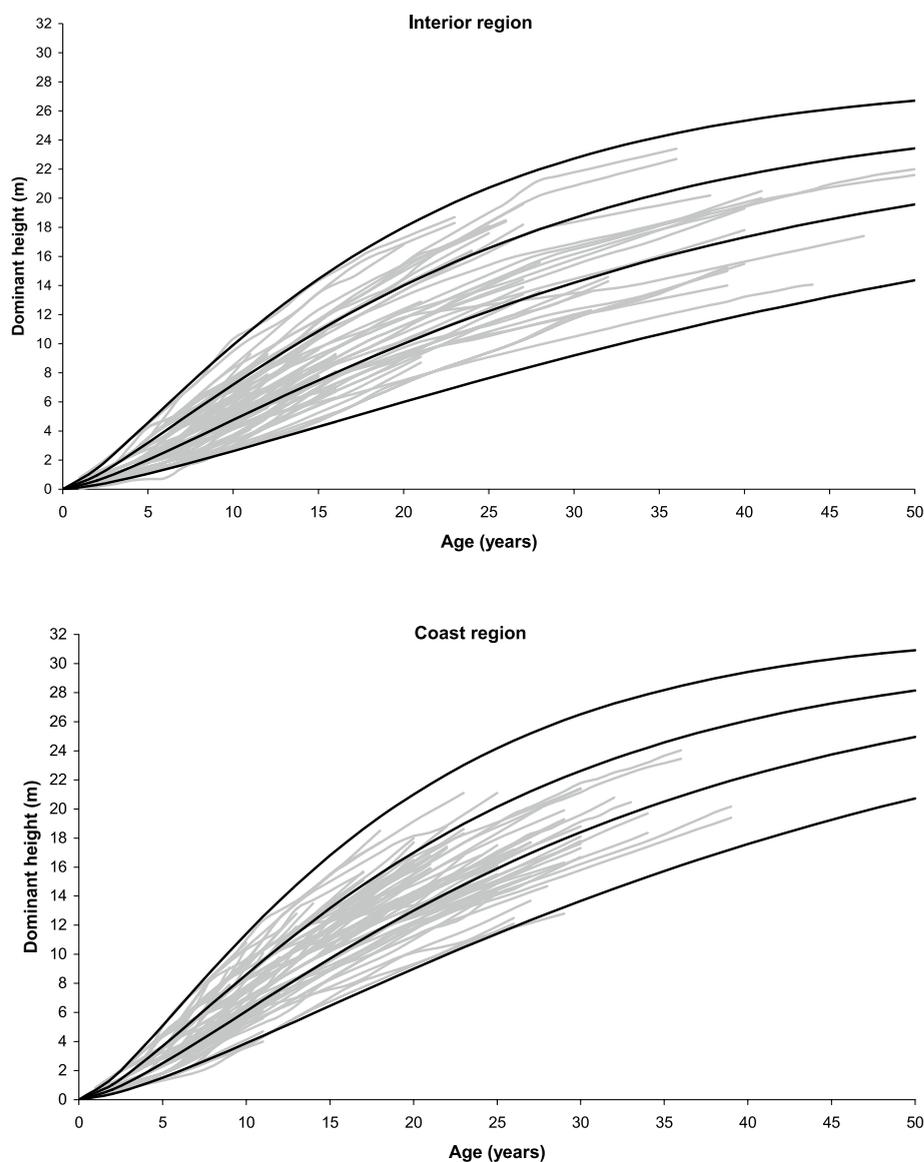


Figure 6. Site index curves generated by the algebraic difference equation proposed by Goelz and Burk [26] for four different site index (6, 10, 14 and 18 m at a reference age of 20 years and 9, 13, 17 and 21 m at the same reference age for interior and coast ecoregion, respectively).

The values of the parameters of the model are consistent with the empirical evidence that for the same site index, the growth rate ($dH/H \cdot dt$) and the asymptote value for the interior ecoregion are lower than those obtained for the coast ecoregion.

4. CONCLUSIONS

Many of the ten algebraic difference models tested gave an accurate site index system on both ecoregions for fitting and cross-validation. The model proposed by Amateis and Burkhart [3] showed the lowest accuracy and precision probably due to the strong dependence of this model on the data structure used [11]. Models based on the Chapman-Richards and Korf height-age equations solved by b_1 or b_0 (M5, M7 and M9) performed better than those solved by b_2 , meaning that this parameter of the Chapman-Richards and Korf height-age equations is not directly related with site quality differences in this species.

Based on the statistics which were used to compare and to validate the models, equation M9 [26] and M3 [40] showed, in general, the best results for interior and coast ecoregions, respectively.

This result was achieved using a generalized nonlinear least squares method to fit the models and a first-order autoregressive process to describe the error structure and different weighting factors. The results of the Durbin test and the analysis of the studentized residuals show that errors were independent and identically distributed with a common variance.

Ideally, the appropriate reference age should be selected such that the variance of the volume estimates for a particular stand was minimized, but in this study for practical use a reference age of 20 years was selected, based on the relative error in height predictions obtained when comparing the estimated value with the real value from the stem analysis for different reference ages.

The Lakkis and Jones and the nonlinear extra sum of squares tests indicated that the site index models are different between ecoregions. Therefore, ecoregion-based site index models were developed. Although the best results were obtained using different models for each ecoregion (M9 and M3 for interior and coast ecoregion, respectively), the critical error for both the overall height prediction and the overall site index prediction of a unique model with a different set of parameters for each ecoregion were very similar. In addition, the use of a unique model based on equation M9 presents the advantage that it has multiple asymptotes. Based on these results, a site index model with a different set of parameters for each ecoregion based on the algebraic difference equation proposed by Goelz and Burk [26] is recommended for site index and height-growth estimates in even-aged stands of *Pinus pinaster* in Galicia.

Acknowledgements: This study was financed by the Comisión Interministerial de Ciencia y Tecnología (CICYT), project No. AGL2001-3871-C02-01 and was developed in the University of Göttingen with a grant of the Ministerio de Educación, Cultura y Deporte of Spain.

REFERENCES

- [1] Alía R., Gil L., Pardos J.A., Performance of 43 *Pinus pinaster* provenances on 5 locations in Central Spain, *Silvae Genet.* 44 (1995) 75–81.
- [2] Alía R., Moro J., Denis J.B., Performance of *Pinus pinaster* Ait. provenances in Spain: interpretation of the genotype-environment interaction, *Can. J. For. Res.* 27 (1997) 1548–1559.
- [3] Amateis R.L., Burkhart H.E., Site index curves for loblolly pine plantations on cutover-site prepared lands, *South. J. Appl. For.* 9 (1985) 166–169.
- [4] Bailey R.L., Clutter J.L., Base-age polymorphic site curves, *For. Sci.* 20 (1974) 155–159.
- [5] Bates D.M., Watts D.G., *Nonlinear regression analysis and its applications*, John Wiley & Sons, New York, 1988.
- [6] Beaumont J.F., Ung C.H., Bernier-Cardou M., Relating site index to ecological factors in black spruce stands: test of hypotheses, *For. Sci.* 45 (1999) 484–491.
- [7] Beck D.E., Height-growth patterns and site index of white pine in the southern Appalachians, *For. Sci.* 17 (1971) 252–260.
- [8] Borders B.E., Bailey R.L., Ware K.D., Slash pine site index from a polymorphic model by joining (splining) non-polynomial segments with an algebraic difference method, *For. Sci.* 30 (1984) 411–423.
- [9] Burkhart H.E., Tennent R.B., Site index equations for radiata pine in New Zealand, *N.Z. J. For. Sci.* 7 (1977) 408–416.
- [10] Burnham K.P., Anderson D.R., *Model selection and inference. A practical information-Theoretic approach*, Springer-Verlag, New York, 1998.
- [11] Cao Q.V., Estimating coefficients of base-age-invariant site index equations, *Can. J. For. Res.* 23 (1993) 2343–2347.
- [12] Calama R., Cañadas N., Montero G., Inter-regional variability in site index models for even-aged stands of stone pine (*Pinus pinea* L.) in Spain, *Ann. For. Sci.* 60 (2003) 259–269.
- [13] Carmean W.H., Site index curves for upland oaks in the Central States, *For. Sci.* 18 (1972) 109–120.
- [14] Carmean W.H., Forest site quality evaluation in the United States, *Adv. Agron.* 27 (1975) 209–267.
- [15] Chapman D.G., Statistical problems in population dynamics, In: *Proceedings of 4th Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 1961, pp. 153–168.
- [16] Chen B., *Applications of optimizations methods to forest planning problems*, Cuvillier Verlag, Göttingen, 2003.
- [17] Cieszewski C.J., Bella I.E., Polymorphic height and site index curves for lodgepole pine in Alberta, *Can. J. For. Res.* 19 (1989) 1151–1160.
- [18] Cieszewski C.J., Bailey R.L., Generalized algebraic difference approach: theory based derivation of dynamic site equations with polymorphism and variable asymptotes, *For. Sci.* 46 (2000) 116–126.
- [19] Clutter J.L., Lenhart J.D., Site index curves for old-field loblolly pine plantations in Georgia Piedmont, Ga. *For. Res. Counc. Rep.* 22, Ser. 1, 1968.
- [20] Clutter J.L., Fortson J.C., Piennar J.C., Brister L.V., Bailey R.L., *Timber management: A quantitative approach*, John Wiley & Sons, New York, 1983.
- [21] Durbin J., Testing for serial correlation in least-square regression when some of the regressors are lagged dependent variables, *Econometrica* 50 (1970) 410–421.
- [22] Dyer M.E., Bailey R.L., A test of six methods for estimating true heights from stem analysis data, *For. Sci.* 33 (1987) 3–13.
- [23] Ek A.R., A formula for white spruce site index curves, *Univ. Wisc. For. Res. Note*, No. 161, 1971.
- [24] Elfving B., Kiviste A.K., Construction of site index equations for *Pinus sylvestris* L. using permanent plot data in Sweden, *For. Ecol. Manage.* 98 (1997) 125–134.
- [25] Gadown K., Bredenkamp B., *Forest Management*, Academica, Pretoria, 1992.
- [26] Goelz J.C.G., Burk T.E., Development of a well-behaved site index equation: jack pine in north central Ontario, *Can. J. For. Res.* 22 (1992) 776–784.
- [27] Goelz J.C.G., Burk T.E., Measurement error causes bias in site index equations, *Can. J. For. Res.* 26 (1996) 1585–1593.
- [28] Graney D.L., Burkhart H.E., Polymorphic site index curves for shortleaf pine in the Ouachita Mountains, *US For. Serv. South. For. Exp. Stn. Res. Pap.* SO-85, 1973.
- [29] Huang S., Titus S.J., An index of site productivity for uneven-aged and mixed-species stands, *Can. J. For. Res.* 23 (1993) 558–562.
- [30] Huang S., Ecologically-based individual tree volume estimation for major Alberta tree species, Land and Forest Service, Tech. Rep. Pub. No. T/288, Edmonton, Alberta, 1994.
- [31] Huang S., Development of a subregion-based compatible height-site index-age model for black spruce in Alberta, Alberta Land and Forest Service, *For. Manage. Res. Note* No. 5, Pub. No. T/352, Edmonton, Alberta, 1997.
- [32] Huang S., Development of compatible height and site index models for young and mature stands within an ecosystem-based management framework, in: Amaro A., Tomé M. (Eds.), *Empirical and process-based models for forest tree and stand growth simulation*, Lisboa, 1999, pp. 61–98.
- [33] Huang S., Price D., Titus S.J., Development of ecoregion-based height-diameter models for white spruce in boreal forests, *For. Ecol. Manage.* 129 (2000) 125–141.
- [34] Huang S., Validating and localizing growth and yield models: procedures, problems and prospects, in: Amaro A. (Ed.), *Reality, models and parameter estimation – the forestry scenario*, Sesimbra, 2002.

- [35] Johansson T., Site index curves for common alder and grey alder growing on different types of forest soil in Sweden, *Scan. J. For. Res.* 14 (1999) 441–453.
- [36] Khattree R., Naik D.N., *Applied multivariate statistics with SAS software*, 2nd ed., SAS Institute Inc., Cary, NC, 1999.
- [37] Lappi J., Bailey R.L., A height prediction model with random stand and tree parameters: an alternative to traditional site index methods, *For. Sci.* 34 (1988) 907–927.
- [38] Lappi J., Malinen J., Random parameter height/age models when stand parameters and stand age are correlated, *For. Sci.* 40 (1994) 715–731.
- [39] Lundqvist B., On the height growth in cultivated stands of pine and spruce in northern Sweden, *Medd. Fran Statens Skogforsk. Band 47* (1957) 1–64.
- [40] McDill M.E., Amateis R.L., Measuring forest site quality using the parameters of a dimensionally compatible height growth function, *For. Sci.* 38 (1992) 409–429.
- [41] Monserud R.A., Height growth and site index curves for inland Douglas-fir based on stem analysis data and forest habitat type, *For. Sci.* 30 (1984) 943–965.
- [42] Nagel J., Albert M., Schmidt M., Das waldbauliche Prognose- und Entscheidungsmodell BWINPro 6.1, *Forst u. Holz* 57 (2002) 486–493.
- [43] Neter J., Wasserman W., Kutner M., *Applied linear statistical models*, 3rd ed., Irwin, Homewood, IL, 1990.
- [44] Newberry J.D., A note on Carmean's estimate of height from stem analysis data, *For. Sci.* 37 (1991) 368–369.
- [45] Newnham R.M., A modification of the Ek-Payandeh nonlinear regression for site index curves, *Can. J. For. Res.* 18 (1988) 115–120.
- [46] Parresol B.R., Vissage J.S., White pine site index for the southern forest survey, Res. Pap. SRS-10, Asheville, NC: US Department of Agriculture, Forest Service, Southern Research Station, 1998.
- [47] Payandeh B., Wang Y., A site index model remodified, *Can. J. For. Res.* 24 (1994) 197–198.
- [48] Peschel W., Die mathematischen Methoden zur Herteitung der Wachstums-gesetze von Baum und Bestand und die Ergebnisse ihrer Anwendung, *Tharandter Forstl. Jarb.* 89 (1938) 169–274.
- [49] Pillsbury N.H., McDonald P.M., Simon V., Reliability of tanoak volume equations when applied to different areas, *West. J. Appl. For.* 10 (1995) 72–78.
- [50] Richards F.J., A flexible growth function for empirical use, *J. Exp. Bot.* 10 (1959) 290–300.
- [51] Ruíz Zorrilla P., *Notas para una historia del pino en Galicia*, Publicaciones del Centro Forestal de Lourizán, Pontevedra, 1981.
- [52] SAS Institute Inc., *SAS/ETS User's Guide*, Version 8, Vols. 1 and 2, SAS Institute Inc., Cary, NC, 2000.
- [53] Sloboda B., Zur Darstellung von Wachstumprozessen mit Hilfe von Differentialgleichungen erster Ordnung. Mitteilungen der Badenerbergischen Forstlichen Versuchs und Forschungsanstalt, 1971.
- [54] Trincado G., Kiviste A.K., Gadow K.v., Preliminary site index models for native Roble (*Nothofagus obliqua*) and Raulí (*Nothofagus alpina*) in Chile. *N.Z. J. For. Sci.* 33 (2003) 322–333.
- [55] Vega P., Vega G., González M., Rodríguez A., Mejora del *Pinus pinaster* Ait. en Galicia, in: Silva Pando J. (Ed.), *I Congreso Forestal Español*, Vol. 2, 1993, pp. 129–134.
- [56] Xunta de Galicia, *Anuario de estadística agraria 2000*, Santiago de Compostela, 2002.