SITE INDEX MODEL FOR NATURAL STANDS OF REBOLLO OAK (*Quercus pyrenaica* Willd.) IN GALICIA, NW IBERIAN PENINSULA

MODELO DE ÍNDICE DE SITIO EM POVOAMENTOS NATURAIS DE CARVALHO NEGRO (*Quercus pyrenaica* Willd.) DA GALICIA, NW DA PENÍNSULA IBÉRICA

Ignacio Javier Díaz-Maroto[1]   José Fernández-Parajes[2]   Pablo Vila-Lameiro[2]   Eva Barcala-Pérez[3]

ABSTRACT

Data from stem analysis of 56 dominant trees of *Quercus pyrenaica* Willd., in natural stands in Galicia (NW Iberian Peninsula), were used to evaluate four dynamic site equations derived with the Generalized Algebraic Difference Approach (GADA). All the equations are base-age invariant and directly estimate height and site index from any height and age. The fittings were made using a data structure involving all possible growth intervals. The GADA formulation derived on the basis of the Bertalanffy-Richards model by considering the asymptote and the initial pattern parameters as related to site productivity. It is therefore recommended for height growth prediction and site classification for natural stands of rebollo oak in Galicia. The autocorrelation was analyzed with a test of residuals using Durbin's t-test without reaching a manifest result of autocorrelation between managed data.

Keywords: *Quercus pyrenaica*; site index model; height growth; Generalized Algebraic Difference Approach.

RESUMO

Dados da análise de tronco de 56 árvores dominantes de *Quercus pyrenaica* Willd. de florestas naturais da Galicia (NW Península Ibérica) foram utilizados para avaliar quatro equações dinamicas de índice de sitio obtidas com a Abordagem Generalizada de Diferença Algébrica (GADA). Todas as equações são baseadas na idade invariante e diretamente estimam a altura e o índice de sitio. Os ajustes foram feitos utilizando uma estrutura de dados que incluiam todos os possíveis intervalos de crescimento. A formulação GADA teve base no modelo Bertalanffy-Richards, considerando a assíntota e do padrão inicial como parâmetros relacionados à produtividade do sitio. Assim, recomenda-se o modelo na previsão do crescimento em altura e classificação de sitios para povoamentos naturais de carvalho negro na Galicia. A autocorrelação foi analisada com um teste de resíduos utilizando Durbin's t-teste, sem chegar a um resultado significativo de autocorrelação entre os dados.

Palavras-chave: *Quercus pyrenaica*; modelo de índice de sitio; crescimento em altura; Abordagem das Diferenças Algébricas Generalizadas.

_____

1. Author for correspondence: Dr. Forestry Engineer, University of Santiago de Compostela, Department of Agroforestry Engineering, Campus Universitario s/n. E – 27002, Lugo, Spain. ignacio.diazmaroto@usc.es
2. Forestry Engineer, Dr., University of Santiago de Compostela, Department of Agroforestry Engineering, Campus Universitario s/n. E – 27002, Lugo, Spain.
3. Forestry Engineer, University of Santiago de Compostela, Department of Agroforestry Engineering, Campus Universitario s/n. E – 27002, Lugo, Spain.

INTRODUCTION

Rebollo oak (*Quercus pyrenaica* Willd.) is distributed throughout the Iberian Peninsula, it is probably found all over Spanish territory, although it is sometimes only present in small stands or represented by few individuals (MESON, 1983; CARVALHO *et al.*, 2005). In Galicia, the climax vegetation in large areas of the south of Lugo province and inland areas of Pontevedra and Ourense provinces is comprised of rebollo natural oaks (DÍAZ-MAROTO *et al.*, 2006) and, in other areas, it is rather more disperse and often mixed with *Quercus robur* L. (DÍAZ-MAROTO *et al.*, 2005). Rebollo oak is the dominant tree species in Galician native forest (NW Spain) with an oceanic climate and Mediterranean influence, where at present, covers 100,504 ha, in pure stands (DGCONA, 2001). Up to now, no studies have been carried out to determine the site quality of *Quercus pyrenaica* stands in the area being studied. Nevertheless, during recent decades, there has been increasing interest in the management of broadleaved tree species, due to the current emphasis on sustainable production of stands (DÍAZ-MAROTO *et al.*, 2007). In this sense, site quality is of primary concern in making forest management decisions (BARRIO and DIÉGUEZ-ARANDA, 2005).

An index which expresses site productivity is a required variable for the modelling of present and future growth and yield, and it can also be used for forest land stratification for purposes of forest inventory, and for forest exploitation on a sustainable yield basis (GARCÍA, 1983). Site quality should be measured directly in terms of volume yield. Nevertheless, the stand history affects the volume of timber in an area at any time, and we almost never know this history (BAILEY and CLUTTER, 1974; CLUTTER *et al.*, 1983). Stand density at establishment, past cutting, diseases, plagues, and weather can all affect the standing volume, as can the length of forest stand rotation (DAVIS *et al.*, 2001; KURINOBU *et al.*, 2007). Therefore, site quality is frequently made in terms of stand height, because, for many species, areas having good site quality are also where height growth is high (BAILEY and CLUTTER, 1974; CLUTTER *et al.*, 1983). Taking into account that the average height of the stand may be affected by certain types of thinning, but that within a wide range of stand densities the height growth does not seem to be affected – especially when the

comparison is restricted to dominant trees (BAILEY and CLUTTER, 1974; CLUTTER *et al.*, 1983) –, the most common approach in determining site productivity involves the use of a site index. This may be defined as the dominant height of the stand at a specific base age (BARRIO and DIÉGUEZ-ARANDA, 2005; ADAME *et al.*, 2006).

The aim of this work was to develop a site quality system (both for height growth and site index estimation) for single-species *Quercus pyrenaica* natural stands in Galicia (NW Iberian Peninsula).

MATERIAL AND METHODS

Description of sampling and data recording

In the winter of 2003-2004, a network of 40 permanent rectangular plots was established in even-aged of rebollo oak in Galicia (Figure 1). The plots were located throughout the area of distribution of this species in Galician territory, and were subjectively selected to represent the existing range of ages, stand densities and sites (DÍAZ-MAROTO *et al.*, 2006; KURINOBU *et al.*, 2007). The plot size ranged from 130 to 2,120 m^2, depending on stand density, in order to achieve a minimum of 50 trees per plot. Two dominant trees were destructively sampled at 28 locations; in the rest of plots it was not possible to fell no dominant tree, due to several reasons: 1) stand density was extreme, thus, the height growth of dominant trees was affected, 2) silvicultural treatments affected the upper canopy (e.g., pollarding and felling of trees to provide firewood and food for livestock) (CARVALHO *et al.*, 2005), and 3) the topography of some stands and the impossibility to locate to their owners. These trees were selected as the first two dominant trees found outside the plots but in the same stands within $\pm$ 5% of the mean diameter at breast height (dbh) and mean height of the dominant trees (considered as the 100 largest-diameter trees per hectare) (HUMMEL *et al.*, 1959).

The trees were felled leaving stumps of average height of 0.12 m; total stem length was measured to the nearest 0.01 m. The logs were cut at 1 m intervals except the first, which was cut at 1.3 m (dbh). Number of rings was counted at each cross-sectioned point and then converted to stump age. As cross-section lengths did not coincide with periodic height growth (NIGH, 1995), there was adjusted the height/age data from stem analysis to account for this bias using Carmean's method (CARMEAN, 1972),
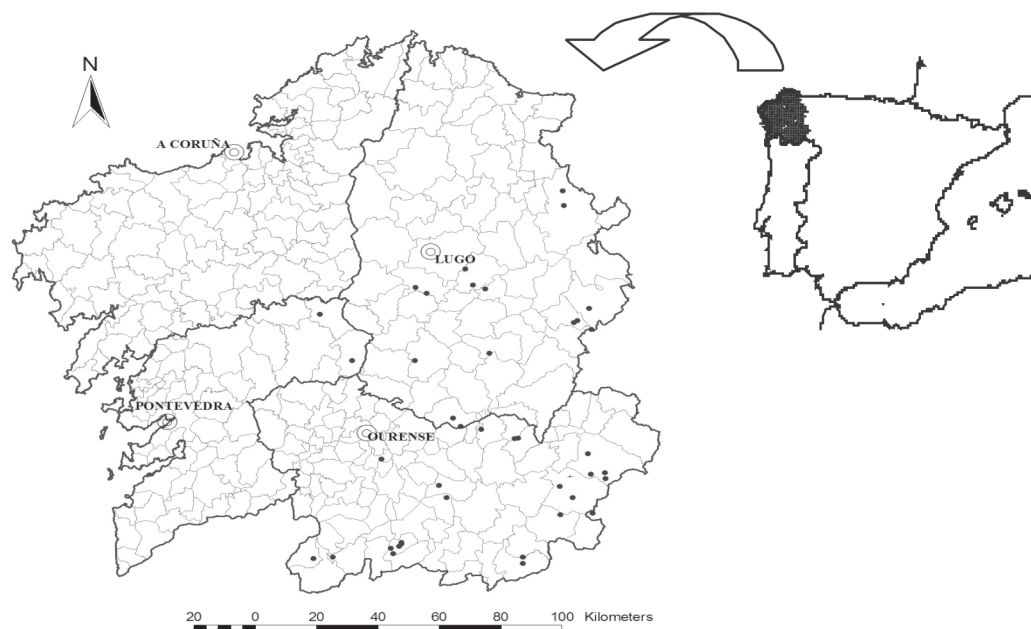
FIGURE 1: Location of *Quercus pyrenaica* plots in Galicia, NW Iberian Peninsula.
FIGURA 1: Localização das parcelas de *Quercus pyrenaica* na Galicia, noroeste da Península Ibérica.

and the modification by NEWBERRY (1991) for the topmost tree section, based on earlier studies of Dyer and Bailey (1987). After examination of data of trees with past height growth abnormalities, stem analyses of 56 trees were selected for modelling dominant height throughout the stand life. Descriptive statistics (number of sections, mean, minimum, maximum, and standard deviation) of the selected data were calculated for total tree height grouped by age classes (Table 1).

Methods for constructing a site index model

According to Clutter *et al.* (1983), most techniques for site index curves construction can be viewed as special cases of three general methods: 1)

the guide curve method, 2) the parameter prediction method, and 3) the Algebraic Difference Equation Method. Although the three methods are not mutually exclusive, the difference equation method has been the preferred form and widely used by numerous authors in order to develop site index curves (BAILEY and CLUTTER, 1974; CIESZEWSKI, 2003; CARVALHO and PARRESOL, 2005; DIÉGUEZ-ARANDA *et al.*, 2005a, 2005b; ADAME *et al.*, 2006; DIÉGUEZ-ARANDA *et al.*, 2006).

The difference equation method makes use of dynamic site equations for modelling the dominant height-age relationship (BAILEY and CIESZEWSKI, 2000a; NORD-LARSEN, 2006). Dynamic site equations have general form (omitting again the vector

TABLE 1: Total section statistics given in meters (age class 5 = 1-9 years,…).
TABELA 1: Estatísticas da seção total em metros (classe de idade 5 = 1-9 anos,....).

Age class	N. of sections	Average ± SD	Range
5	75	2.1 ± 1.0	0.9–4.8
15	118	4.9 ± 1.7	1.2–10.0
25	73	7.9 ± 2.1	2.2–15.1
35	29	9.7 ± 3.3	5.2–18.2
45	15	10.8 ± 4.4	5.0–21.3
55	9	11.4 ± 1.6	9.1–14.3
65	1	11.2 ± 0.0	11.2–11.2
75	2	12.8 ± 0.5	12.3–13.3

SD = Standard Deviation

of model parameters): $Y = f\left(t, t_0, Y_0\right)$, where Y is the function's value at age t, and Y_0 is the reference variable defined as the function's value at age t_0. Advantages of dynamic site equations have been pointed out in several studies (e.g., CIESZEWSKI, 2003): they are base-age-invariant, and they define height-growth and site index models as special cases of the same equation, two desirable properties in most applications. The invariant or unchanging property of dynamic equations refers to predicted heights: any number of points $\left(t_0, Y_0\right)$ on a specific site curve can be used to make predictions for a given age t and the predicted height Y will always be the same. This includes forward and backward predictions, and the path invariance property that ensures the result of projecting first from t_0 to t_1, and then from t_1 to t_2, being the same as that of the one-step projection from t_0 to t_2. Nord-Larsen (2006) noted that some models have been called "base-age or reference-age invariant", since do not have these comprehensive invariance properties.

Bailey and Clutter (1974) formalized the base-age invariance property and presented a technique for dynamic equation derivation which is known in forestry as the Algebraic Difference Approach (ADA), and that essentially consists of replacing a base-model's site-specific parameter with its initial-condition solution.

Cieszewski and Bailey (2000b) introduced a generalization of the ADA, the Generalized Algebraic Difference Approach (GADA), which can be used to derive the same models derived by the ADA. The main advantage of the GADA is that one can expand the base equations according to various theories about growth characteristics (e.g., asymptote, growth rate), allowing more than one parameter to be site-specific, and letting the derivation of more flexible dynamic equations (see CIESZEWSKI, 2002 and 2003; CIESZEWSKI *et al.*, 2007). It includes the ability to simulate concurrent polymorphism and multiple asymptotes, an important property of site equations (CIESZEWSKI, 2002; KURINOBU *et al.*, 2007).

Models considered

A large number of researchers have pointed to desirable attributes for site equations (BAILEY and CLUTTER, 1974; BAILEY and CIESZEWSKI, 2000a; CIESZEWSKI, 2002 and 2003; NORD-LARSEN, 2006). The most frequently listed criteria

are: polymorphism, sigmoid growth pattern with an inflexion point, horizontal asymptote at old ages, logical behaviour (height should be zero at age zero and equal to site index at reference age; the curve should never decrease), base-age invariance, and path invariance (KURINOBU *et al.*, 2007). The fulfilment of these attributes depends on both the model construction method and statistical procedures used, and it cannot always be achieved (BARRIO and DIÉGUEZ-ARANDA, 2005). Multiple asymptotes is also a desirable attribute (CIESZEWSKI, 2002), although some of the functions generally used have a single asymptote. This may not be an important problem whether the behaviour of curves is suitable for the range of ages that would be used in practice, and whether the single asymptote is achieved at very old ages (CIESZEWSKI, 2004).

With these criteria in mind, there were examined different base models and tested several variants for each one which were developed using the GADA. Both one and two parameters were considered to be site-specific. Finally, the efforts were focused efforts on four dynamic equations. As general notational convention, a_1, a_2… a_n, were used to denote parameters in base models, while b_1, b_2… b_m, were used for global parameters in subsequent GADA formulations. All the GADA based models have the general implicit form the equation:. $Y = f\left(t, t_0, Y_0, b_1, b_2, \ldots b_m\right)$

Two models were tested on the basis of the function proposed by Bartalanffy (1949 and 1957) and formulated by Richards (1959), whose form can be, for simplicity, represented as:

$$Y = a_1\left(1 - \exp\left(- a_2 t\right)\right)^{a_3} \qquad (1)$$

Where: a_1 is an asymptote or limiting value, a_2 is often called a "rate parameter", and a_3 is often referred to as "an initial pattern parameter". This model is very flexible and has been frequently used for site-index curve development, including with the GADA approach (e.g., GARCÍA, 1983; BIGING, 1985; CARVALHO and PARRESOL, 2005; DIÉGUEZ-ARANDA *et al.*, 2005b).

The first model derived by applying the GADA to equation (1) considers only the shape parameter a_2 to be site-specific (in this case GADA is equivalent to the ADA). Its GADA formulation results are:

$$Y = b_1\left(1 - \exp\left(- Xt\right)\right)^{b_3} \qquad (2)$$

The solution for X in equation (2) with initial

condition values t_0 and Y_0 $(X_0 = f(t_0, Y_0))$ for t and Y, respectively, is:

$$X_0 = -\ln\left(1 - \left(\frac{Y_0}{b_1}\right)^{1/b_3}\right)\Big/ t_0 \qquad (3)$$

Substituting X into equation (2) results in the following dynamic equation that provides polymorphic curves with a single asymptote:

$$Y = b_1\left(1 - \left(1 - \left(\frac{Y_0}{b_1}\right)^{1/b_3}\right)^{t/t_0}\right)^{b_3} \qquad (4)$$

To derive models with both polymorphism and variable asymptotes from equation (1), more than one parameter has to be a function of site productivity. When the parameter a_2 varies with site productivity none of the other two parameters a_1 or a_3 can be used (together with a_2) to model the site productivity responses because, when a_1 or a_3 is a function of X and the parameter a_2 is also a function of X, the model cannot be solved for X in a closed-form solution (CIESZEWSKI, 2004). Thus, in the second model derived from equation (1) using the GADA both the asymptote a_1 and the shape parameter a_3 are assumed to be dependent on X. To facilitate such derivation the base equation is re-parameterized into a form more suitable for manipulation of these two parameters (using $\exp(a_1{}')$ instead of a_1 and taking the log of the function) as follows:

$$Y = \exp(a_1{}')(1 - \exp(-a_2 t))^{a_3}$$
$$\ln Y = a_1{}' + a_3 \ln(1 - \exp(-a_2 t))$$

The site parameters can be conditioned to be consistently proportional to each other's inverse over the site productivity dimension by the following definitions:

$$a_1{}' = X \text{ and } a_3 = b_2 + b_3/X \text{ , while } a_2 = b_1 \qquad (5)$$

Which defines the following GADA formulation:

$$\ln Y = X + (b_2 + b_3/X)\ln(1 - \exp(-b_1 t)) \qquad (6)$$

For this base equation, the solution for X involves finding roots of a quadratic equation and selection of the most appropriate root to substitute into the dynamic equation. The selection of the most appropriate expression for X may depend on the equation parameters that in turn depend on the data and the domain of the applicable ages (NORD-

LARSEN, 2006). The solution for X in equation (4) with initial condition values t_0 and Y_0 is:

$$X_0 = \tfrac{1}{2}\left((\ln Y_0 - b_2 L_0) \pm \sqrt{(\ln Y_0 - b_2 L_0)^2 - 4b_3 L_0}\right)$$

Where:

$$L_0 = \ln(1 - \exp(-b_1 t_0))$$

Selecting the root more likely to be real and positive (i.e., the one involving addition rather than subtraction of the square root), and substituting it into equation (4) results in the following dynamic equation that provides polymorphic curves with variable asymptotes:

$$Y = Y_0\left(\frac{1 - \exp(-b_1 t)}{1 - \exp(-b_1 t_0)}\right)^{(b_2 + b_3/X_0)} \qquad (7)$$

The subsequent models may be viewed as special cases of the log-logistic class of models, which are equivalent to Hossfeld models (BAILEY and CIESZEWSKI, 2000a; NORD-LARSEN, 2006) and have a long history in describing a wide variety of population dynamics. Monserud (1984) applied a log-logistic model to describe the height growth of inland Douglas-fir in the northern Rocky Mountains based on data from stem analysis of dominant trees. Cieszewski (2002; 2003) examined several GADA formulations utilizing the log-logistic model, which can be expressed generally as:

$$Y = \frac{a_1}{1 + a_2 \exp(-u)} \qquad (8)$$

Where: a_1 is the asymptote parameter, a_2 is the half-saturation parameter that defines the value of $\exp(-u)$ at which, $Y(u) = a_1/2$ and u is a linear combination of log-transformations of the independent variable(s) (MONSERUD, 1984; CIESZEWSKI, 2002).

Within the different forms investigated, the GADA based formulation by Cieszewski (2002) was found to perform particularly well. It considers the above-mentioned u as a logarithmically transformed function of age, such as $u = a_3 \ln t$. Therefore, equation (6) can be simplified as the Hossfeld model:

$$Y = \frac{a_1}{1 + a_2 t^{-a_3}} \qquad (9)$$

Cieszewski (2002) replaced a_1 with a constant plus the unobserved site variable X, and a_2 by,

b_2/X becoming the GADA formulation:

$$Y = \frac{b_1 + X}{1 + b_2/X\, t^{-b_3}} \qquad (10)$$

The solution for X involves again finding roots of a quadratic equation. With initial condition values t_0 and Y_0 it is:

$$X_0 = \tfrac{1}{2}\left(Y_0 - b_1 \pm \sqrt{(Y_0 - b_1)^2 + 4b_2 Y_0 t_0^{-b_3}} \right)$$

Selecting the root more likely to be real and positive (i.e., the one involving addition rather than subtraction of the square root), and substituting it into equation (8) results in the following dynamic equation that provides polymorphic curves with variable asymptotes:

$$Y = \frac{b_1 + X_0}{1 + b_2/X_0\, t^{-b_3}} \qquad (11)$$

The fourth model is the dynamic equation proposed by KcDill and Amateis (1992). It is based on considering parameter a_2 in equation (7) as site-specific. After applying the GADA, solving for X and substituting for initial condition values t_0 and Y_0 results in the following dynamic equation that provides polymorphic curves with a single asymptote:

$$Y = \frac{b_1}{1 - \left(1 - \dfrac{b_1}{Y_0}\right)\left(\dfrac{t_0}{t}\right)^{b_3}} \qquad (12)$$

Data structure and model fitting

The data structure used for fitting the four models was arranged with all the possible combinations among height/age pairs for each tree, including descending growth intervals. All possible growth intervals typically produce fitted models with a better predictive performance as compared to, for example, forward moving first differences (GOELZ and BURK, 1992; HUANG, 1999). However, this data structure may lead to the rejection of the error assumptions because it automatically introduces a lack of independence among observations (GOELZ and BURK, 1992). Although the bibliography on the proper analysis of longitudinal data is available, the more complicated error structure of this type of data has often been ignored in forestry (GREGOIRE *et al.*, 1995), in part because least squares parameter estimates and resulting predictions are still unbiased

in the presence of this autocorrelation. However, least squares parameter estimates are no longer minimum variance estimators; that is, the model mean square error may underestimate the variance of the error terms, invalidating tests of significance (NETER *et al.*, 1990). Autocorrelation can therefore have an adverse impact where the goal is to identify statistically significant predictor variables. Because the aim of this study was to obtain the best site model for predicting height growth and site index estimation for even-aged, single-species *Quercus pyrenaica* natural stands in Galicia, initially we did not take into account the autocorrelation being short of independence among observations because of the use of longitudinal data with temporal dependence within each tree (NETER *et al.*, 1990; BARRIO and DIÉGUEZ-ARANDA, 2005).

Furthermore, non independent observations imply several disturbances as impartiality of parameters, skewed standard error estimation... In this case, autocorrelation is modelled by expanding the error term as an autoregressive process according to this data structure, fitting the model through generalized nonlinear least squares methods using the SAS/STATâ NLIN procedure adjusting data tree by tree (SAS INSTITUTE INC., 2004).

Model comparison

The comparison of the estimates for the different models was based on numerical and graphical analyses of the residuals. Two statistical criteria were examined: root mean square error (*RMSE*) and coefficient of determination for nonlinear models (R^2) (RYAN, 1997). Although several shortcomings have been stated against the use of the R^2 in nonlinear regression, the general usefulness of some global measure of model adequacy would seem to override some of those limitations (MYERS, 1990; RYAN, 1997). These expressions may be summarized as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n - p}} \qquad (13)$$

$$R^2 = r^2_{Y_i \hat{Y}_i} \qquad (12)$$

Where: $r_{Y_i \hat{Y}_i}$ is the correlation coefficient between the measured (Y_i) and estimated ($\hat{Y}_i$) values

of the dependent variable, *n* is the total number of observations used to fit the model, and *p* is the number of model parameters.

That is why the quality of fit does not necessarily reflect the future prediction quality (MYERS, 1990). Assessment of the validity of the model with an independent data set would be the most desirable (HUANG *et al.*, 2003; KOZAK and KOZAK, 2003). Due to the scarcity of such data, several methods have been proposed (e.g., splitting the data set or cross-validation and double cross-validation), although they seldom provide any additional information compared with the respective statistics obtained directly from models built from entire data sets (KOZAK and KOZAK, 2003). The last step in evaluating the different fitted models consisted of examining visually the residuals against the estimated values, and the fitted curves for different site indices overlaid on trajectories of the heights over time. Visual inspection is essential for selecting the most accurate model, since curve profiles may differ considerably, even though goodness-of-fit statistics are similar (NETER *et al.*, 1990).

Selection of reference age for site quality evaluation

Practical use of the model to estimate site quality from any given pair of height and age requires the selection of a base age to which site index will be referenced (CLUTTER *et al.*, 1983). Inversely, site index and its associated base age could be used to estimate dominant height at any desired age (BAILEY and CLUTTER, 1974). The base age should be selected so that it is a reliable predictor of height at other ages (GOELZ and BURK, 1996), taking into account the erratic height growth at young ages but considering that a young base age will help in earlier decision making of the silvicultural treatments to be stands applied (DAVIS *et al.*, 2001). In order to address this consideration, different base ages and their corresponding observed heights were used to estimate heights at other ages for each tree. The results were compared with the observed values from tree analyses, and the relative error in predictions (*RE%*) was then calculated as follows:

$$RE\% = \frac{\sqrt{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 / (n-p)}}{\overline{Y}} 100 \quad (15)$$

Where: Y_i, $\hat{Y}_i$ and $\overline{Y}$ are the observed,

predicted and average values of the tree height, respectively, *n* is the total number of observations used to fit the model, and *p* is the number of model parameters.

RESULTS AND DISCUSSION

The parameter estimates for each model and their corresponding goodness-of-fit statistics are shown in Tables 2 and 3, respectively. The two dynamic equations developed considering only one parameter to be site-specific, equations (3) and (10), in the base model described the data slightly worse than the two-site-specific parameter models (Table 2), and adjust the resultant model to managed data at 90.3% of accuracy. Within the group with two site-specific parameters, equation (5) provided the best results, as indicated by the graphs that represent the fitted curves for different site indices (Figure 2) and the goodness-of-fit statistics. This model explained 91.4% of the total variability, and provided a random pattern of residuals around zero with homogeneous variance and minor significant trends (by the immaturity stands at reference age of 30 years), both for predicted heights and site indices (Figure 3). The results suggest that the one site-specific parameter models could be adequate. Nevertheless, the two site-specific parameter equations fitted allow simulating concurrent polymorphism and multiple asymptotes, two desirable characteristics of site (CIESZEWSKI, 2002; CIESZEWSKI *et al.*, 2007).

Although this may be of secondary importance, equations that possess these characteristics are more flexible, and better describe a wide variety of height-age trends (NORD-LARSEN, 2006). Thus, among the four dynamic equations finally evaluated for height growth prediction and site classification of even-aged, single-species rebollo oak in Galicia, equation (5) – the GADA formulation derived from the Bertalanffy-Richards model by considering parameters a_1 and a_3 as related to site productivity – was selected:

$$Y = Y_0 \left(\frac{1 - \exp(-0.01794\,t)}{1 - \exp(-0.01794\,t_0)} \right)^{(-1.132 + 7.746/X_0)}$$

Where: Y_0 and t_0 represent the predictor height (meters) and age (years), *Y* is the predicted height age t, $X_0 = \frac{1}{2}\left((\ln Y_0 + 1.132 L_0) + \sqrt{(\ln Y_0 + 1.132 L_0)^2 - 4 \times 7.746 L_0} \right)$, and $L_0 = \ln(1 - \exp(-0.01794\,t_0))$.

TABLE 2: Parameter estimates and goodness-of-fit statistics for the different models evaluated.
TABELA 2: Estimativa dos parâmetros e qualidade estatística do ajuste para os diferentes modelos avaliados.

Model	RMSE	R_{adj}^2	Parameter	Estimate	Approx. SE	t-value
Equation (3)	1.203	0.9028	b_1	33.870	1.500	22.52
			b_3	1.102	0.011	103.60
Equation (5)	1.129	0.9143	b_1	0.018	0.001	18.09
			b_2	-1.132	0.201	-5.64
			b_3	7.746	0.674	11.49
Equation (9)	1.192	0.9046	b_1	58.200	2.740	21.23
			b_2	-946.800	87.200	-10.86
			b_3	1.100	0.009	122.60
Equation (10)	1.202	0.9029	b_1	51.020	2.880	17.74
			b_3	1.108	0.011	105.40

Statistical analysis precision: á = 0.05; Significance level in all the parameters: s = 0.001; SE = Standard Error

To estimate the dominant height (Y) of a stand for some desired age (t), given site index (S) and its associated base age (t_S), substitute S for Y_0 and t_S for t_0 in equation (14). Similarly, to estimate site index at some chosen base age, given stand height and age, substitute S for Y and t_S for t in the same equation (14).
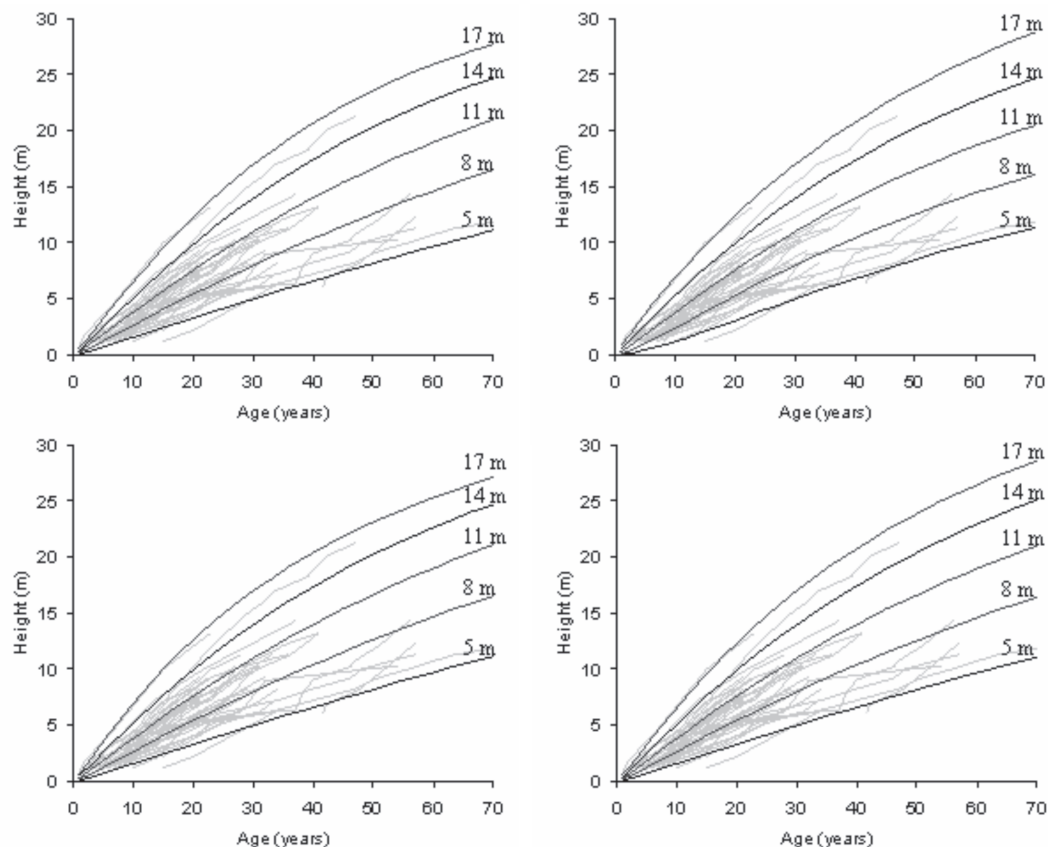

FIGURE 2: Curves for site indices of 5, 8, 11, 14 and 17 m at a reference age of 30 years overlaid on the trajectories of the observed heights over time for: top-left, equation (3); top-right, equation (5); bottom-left, equation (9); bottom-right, equation (10).
FIGURA 2: Curvas de índices de sítio 5, 8, 11, 14 e 17 m, para a idade de referência de 30 anos, sobrepostos sobre as alturas observadas: cima-esquerda, equação (3); cima-direita, equação (5), baixo-esquerda, equação (9), baixo-direita, equação (10).

TABLE 3: Goodness-of-fit autocorrelation statistics for the different models evaluated.
TABELA 3: Qualidade do ajuste de autocorrelação estatística para os diferentes modelos avaliados.

Model	c	t-value	P	R^2
Equation (3)	0.1	1.66	0.116	0.223
Equation (5)	0.1	1.85	0.231	0.375
Equation (9)	0.2	-5.63	0.761	0.153
Equation (10)	0.3	9.76	0.088	0.504

c = Autocorrelation statistical parameter; p = Autocorrelation significance level; R^2 = Coefficient of determination.
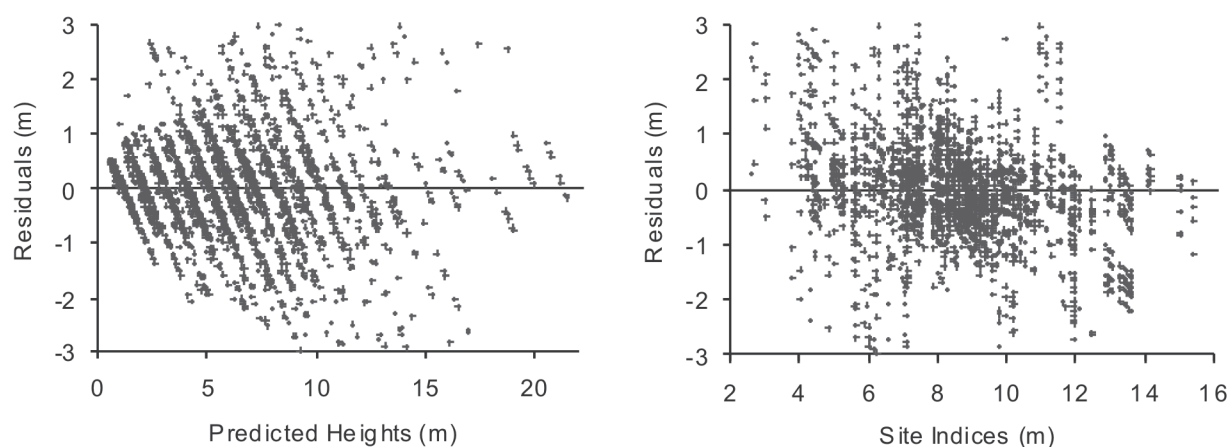

FIGURE 3: Residuals versus predicted heights as an illustration of the non-existent potential heteroscedasticity (left), and site indices at a reference age of 30 years (right) for equation (5).
FIGURA 3: Resíduos e alturas preditas ilustrando a inexistência heteroscedasticidade (esquerda), e índices de sitio á idade de referência de 30 anos (direita) para a equação (5).

Equation (14) fulfils all the desirable properties that a site equation should possess: it constitutes a parsimonious, tree parameter, dynamic site equation; it shows no trends in residuals; it predicts height zero at age zero and equal to site index at base age; it is polymorphic with sigmoid curve shapes and varying asymptotes; and, it has the base-age and path invariance properties providing consistent predictions (BAILEY and CLUTTER, 1974; CIESZEWSKI, 2002; 2003; CIESZEWSKI *et al.*, 2002).

In selecting the base age, it was found that an age around 30 years was superior for predicting height for other ages with a minimum degree of reliability (Figure 4). Even though at older ages the relative error in predicting height was also acceptable, the scarcity of data would lead to an incorrect decision as the data were not representative enough. Therefore, 30 years are proposed as the base age to which site index will be referenced in order to classify the stands according to their productivity.

Initially, all the models were fitted without the autocorrelation parameters, using weighted nonlinear least squares. Afterwards, a test of residuals for autocorrelation was applied using Durbin's t-test (DURBIN, 1970) with regression of the ordinary least squares residuals and testing the significance of the correlation parameter estimates.

As commented, the structure data made hope a high autocorrelation deficiency due to highly correlated residuals. However, the analyzed model does not allow affirm, clearly the existence of influence by data structure. For example, the selected model, equation (5), provided these results: c = 0.1, t = 1.85, P < 0.231, and R^2 = 0.375 (Table 3); this situation does not imply that if all models were refitted by modelling the error structure the selected was a better model.
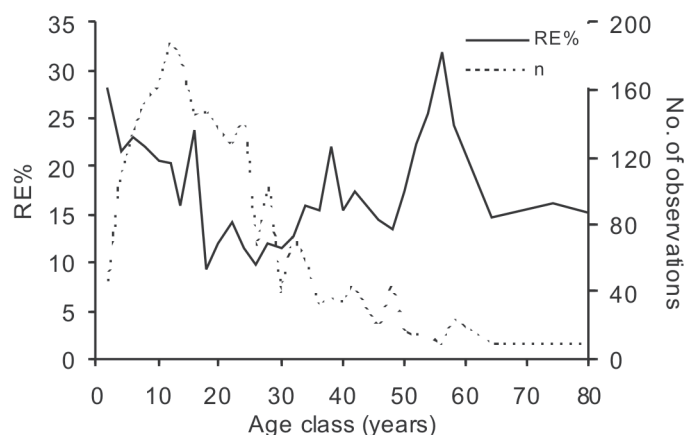
FIGURE 4: Relative error in height prediction related to choice of reference age for equation (5).
FIGURA 4: Erro relativo em altura relacionado com a escolha da idade de referência para a equação (5).

CONCLUSIONS

The typology of *Quercus pyrenaica* stands in Galicia (NW Iberian Peninsula) difficults the definition of the site equations. To solve this problem, data from stem analysis of 56 dominant trees were obtained, and evaluated with four dynamic site equations.

After the study of similar situations, the decision was to apply the Generalized Algebraic Difference Approach (GADA). This implies that all the equations are base-age invariant and directly estimate height and site index from any height and age.

The fittings were done using a data structure involving all possible growth intervals that may lead to the rejection of the error assumptions, because it automatically introduces a lack of independence among observations.

The GADA formulation derived on the basis of the Bertalanffy-Richards model by considering the asymptote and the initial pattern parameters as related to site productivity.

The resulting model is polymorphic with sigmoid curve shapes and varying asymptotes that fulfil all the desirable properties that a site equation should possess: a parsimonious tree parameter dynamic site equation, showing no trends in residuals providing consistent predictions.

The Durbin's residuals t-test was used to analyze the autocorrelation without reaching a manifest result of autocorrelation between managed data.

Thirty years are proposed as the base age to which site index will be referenced in order to classify the stands according to their productivity, because this age was superior for predicting height with a minimum degree of reliability although, at older ages, the relative error in predicting height was also acceptable.

ACKNOWLEDGEMENTS

_____

REFERENCES

ADAME, P. *et al.* Modelling dominant height growth and site index curves for rebollo oak (*Quercus pyrenaica* Willd.). **Annals of Forest Science**, Nancy, v. 63, p. 929-940, dec. 2006.

BAILEY, R. L.; CIESZEWSKI, C. J. Development of a well-behaved site-index equation: jack pine in north-central Ontario: comment. **Canadian Journal of Forest Research**, Ottawa, v. 30, p. 1667-1668, Oct. 2000a.

_____. Generalized Algebraic Difference Approach: Theory based derivation of dynamic site equations with polymorphism and variable asymptotes. **Forest Science**, Bethesda, v. 46, p. 116-126, Feb. 2000b.

BAILEY, R. L.; CLUTTER, J. L. Base-age invariant polymorphic site curves. **Forest Science**, Bethesda, v. 20, p. 155-159, June 1974.

BARRIO, M.; DIÉGUEZ-ARANDA, U. Site quality of pedunculate oak (*Quercus robur* L.) stands in Galicia (northwest of Spain). **European Journal of Forest Research**, New York, v. 124, p. 19-28, Apr. 2005.

BERTALANFFY, L. V. Problems of organic growth. **Nature**, London, v. 163, p. 156-158, Jan. 1949.

_____. Quantitative laws in metabolism and growth. **The Quarterly Review of Biology**, Chicago, v. 32, p. 217-231, Sept. 1957.

BIGING, G.S. Improved estimates of site index curves using a varying-parameter model. **Forest Science**, Bethesda, v. 31, p. 248-259, Mar. 1985.

CARMEAN, W. H. Site index curves for upland oaks in the Central States. **Forest Science**, Bethesda, v. 18, p. 109-120, June 1972.

CARVALHO, J. P. *et al.* **O Carvalho Negral**. Vila Real, Portugal: UTAD-CEGE, 2005. 206 p.

CARVALHO, J. P.; PARRESOL, B. R. A site model for Pyrenean oak (*Quercus pyrenaica*) stands using a dynamic algebraic difference equation. **Canadian Journal of Forest Research**, Ottawa, v. 35, p. 93-99, Jan. 2005.

CIESZEWSKI, C. J. Comparing fixed and variable-base-age site equations having single versus multiple asymptotes. **Forest Science**, Bethesda, v. 48, p. 7-23, Feb. 2002.

_____. Developing a well-behaved dynamic site equation using a modified Hossfeld IV function $Y^3 = (ax^m)/(c + x^{m-1})$, a simplified mixed-model and scant subalpine fir data. **Forest Science**, Bethesda, v. 49, p. 539-554, AUG. 2003.

_____. **GADA derivation of dynamic site equations with polymorphism and variable asymptotes from Richards, Weibull, and other exponential functions**. Georgia: University of Georgia PMRC-TR, 2004. 16 p.

CIESZEWSKI, C. J.; STRUB, M.; ZASADA, M. New dynamic site equation that fits best the Schwappach data for Scots pine (*Pinus sylvestris* L.) in Central Europe. **Forest Ecology and Management**, Amsterdam, v. 243, p. 83-93, May 2007.

CLUTTER, J. L. *et al.* **Timber management: a quantitative approach**. New York: Wiley & Sons, 1983. 333 p.

DAVIS, L.S. *et al.* **Forest management: to sustain ecological, economic, and social values**. New York, EEUU: McGraw-Hill, 2001. 886 p.

DGCONA. **Tercer Inventario Forestal Nacional, Galicia (A Coruña, Lugo, Ourense y Pontevedra)**. Madrid, España: Ministerio de Medio Ambiente, 2001. 122 p.

DÍAZ-MAROTO, I. J.; VILA-LAMEIRO, P.; SILVA-PANDO. F. J. Autoecology of oaks (*Quercus robur* L.) in Galicia (Spain). **Annals of Forest Science**, Nancy, v. 62, p. 737-749, Mar. 2005.

DÍAZ-MAROTO, I. J.; FERNÁNDEZ-PARAJES, J.; VILA-LAMEIRO, P. Autecology of rebollo oak (*Quercus pyrenaica* Willd.) in Galicia (Spain). **Annals of Forest Science**, Nancy, v. 63, p. 157-167, Mar. 2006.

DÍAZ-MAROTO, I. J. *et al.* A comparison of the autecology of *Quercus robur* L. and *Q. pyrenaica* Willd.: present habitat in Galicia, NW Spain. **Forestry**, Edinburgh, v. 80, p. 223-239, July 2007.

DIÉGUEZ-ARANDA, U. *et al.* Site quality equations for *Pinus sylvestris* L. plantations in Galicia (northwestern Spain). **Annals of Forest Science**, Nancy, v. 62, p. 143-152, Mar. 2005a.

DIÉGUEZ-ARANDA, U.; BURKHART, H. E.; RODRÍGUEZ-SOALLEIRO, R. Modelling dominant height growth of radiata pine (*Pinus radiata* D. Don) plantations in north-western Spain. **Forest Ecology Management**, Amsterdam, v. 215, p. 271-284, Aug. 2005b.

DIÉGUEZ-ARANDA, U. *et al.* Site quality curves for birch stands in north-western Spain. **Silva Fennica**, Vantaa, v. 40, p. 631-644, Dec. 2006.

DURBIN, J. Testing for serial correlation in least-square regression when some of the regressors are lagged dependent variables. **Econometrica**, New York, v. 50, p. 410-421, Mar. 1970.

DYER, M. E.; BAILEY, R. L. A test of six methods for estimating true heights from stem analysis data. **Forest Science**, Bethesda, v. 33, p. 3-13, Mar. 1987.

GARCÍA, O. A stochastic differential equation model for the height growth of forest stands. **Biometrics**, New York, v. 39, p. 1059-1072, Dec.1983.

GOELZ, J. C. G.; BURK, T. E. Development of a well-behaved site index equation: jack pine in north central Ontario. **Canadian Journal of Forest Research**, Ottawa, v. 22, p. 776-784, June1992.

_____. Measurement error causes bias in site index equations. **Canadian Journal of Forest Research**, Ottawa, v. 26, p. 1586-1593, Sept. 1996.

GREGOIRE, T.G.; SCHABENBERGER, O.; BARRETT, J. P. Linear modelling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. **Canadian Journal of Forest Research**, Ottawa, v. 25, p. 137-156, Jan. 1995.

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, Portugal: Edições Salamandra-Novas Tecnologias, 1999, p. 61-98.

HUANG, S.; YANG, Y.; WANG, Y. A critical look at procedures for validating growth and yield models. In: AMARO, A.; REED, D.; SOARES, P. (Eds.) **Modelling forest systems**. Wallingford, EEUU: CAB International, 2003, p. 271-293.

HUMMEL, F. C. *et al.* Code of sample plot procedure. London:, HMSO, 1959. 113 p. (Forestry Commission Bulletin 31).

KOZAK, A.; KOZAK, R. Does cross validation provide additional information in the evaluation of regression models? **Canadian Journal of Forest Research**, Ottawa, v. 33, p. 976-987, 2003.

KURINOBU, S. *et al.* A provisional growth model with a size-density relationship for a plantation of *Paraserianthes falcataria* derived from measurements taken over 2 years in Pare, Indonesia. **Journal of Forest Research**, Tokyo, v. 12, p. 230-236, June 2007.

McDILL, M. E.; AMATEIS, R. L. Measuring forest site quality using the parameters of a dimensionally compatible height growth function. **Forest Science**, Bethesda, v. 38, p. 409-429, Apr. 1992.

MESON, M. Aspectos autoecológicos de *Quercus pyrenaica* Willd.: distribución y climatología. **Boletín Estación Central Ecología**, Madrid, v. 12, p. 24-33, Enero 1983.

MONSERUD, R. A. Height growth and site index curves for inland Douglas-fir based on stem analysis data and forest habitat type. **Forest Science**, Bethesda, v. 30, p. 943-965, Dec. 1984.

MYERS, R. H. **Classical and modern regression with applications**. 2 nd ed. Belmont: Duxbury Press, 1990. 488 p.

NETER, J.; WASSERMAN, W.; KUTNER, M. H. Applied linear statistical models: regression, analysis of variance and experimental designs. 3th ed. Boston: Irwin, 1990. 842 p.

NEWBERRY, J. D. A note on Carmean's estimate of height from stem analysis data. **Forest Science**, Bethesda, v. 37, p. 368-369, Mar. 1991.

NIGH, G. D. **Compatibility improvements and bias reduction in height-age models**. British Columbia: Forestry Division Services Branch, Ministry of Forest, Victoria, 1995. 10 p.

NORD-LARSEN, T. Developing dynamic site index curves for European beech (*Fagus sylvatica* L.) in Denmark. **Forest Science**, Bethesda, v. 52, p. 173-181, Apr. 2006.

RICHARDS, F.J. A flexible growth function for empirical use. **Journal of Experimental Botany**, Oxford, v. 10, p. 290-300, June 1959.

RYAN, T. P. Modern regression methods. John Wiley & Sons, New York. 1996. 515 p.

SAS INSTITUTE INC. SAS/STAT® 9.1. **User's Guide**. Cary: SAS Institute Inc. 2004. 128 p.