

Site Quality Curves for Birch Stands in North-Western Spain

Ulises Diéguez-Aranda, José Antonio Grandas-Arias, Juan Gabriel Álvarez-González and Klaus von Gadow

Diéguez-Aranda, U., Grandas-Arias, J.A., Álvarez-González, J.G. & Gadow, K.v. 2006. Site quality curves for birch stands in north-western Spain. *Silva Fennica* 40(4): 631–644.

A model for predicting the height growth of even-aged, birch (*Betula pubescens* Ehrh.) dominated stands in Galicia (north-western Spain) was developed. Data from stem analysis of 214 trees were used for model construction. Two dynamic site equations derived with the generalized algebraic difference approach (GADA) were tested, which combine compatible site index and height models in one common equation. Both equations are base-age invariant and directly estimate height and site index from any height and age. The fittings were done in one stage using the base-age-invariant dummy variables method. A second-order continuous-time autoregressive error structure was used to correct the inherent autocorrelation of the longitudinal data used in this study. Cieszewski's model best described the data. This model is therefore recommended for height growth prediction and site classification of birch stands in Galicia.

Keywords site quality, generalized algebraic difference approach, base-age invariant dynamic equation, Hossfeld model, *Betula pubescens*, Galicia

Addresses Departamento de Ingeniería Agroforestal, Universidad de Santiago de Compostela. Escuela Politécnica Superior, Campus universitario, 27002 Lugo, Spain; *Gadow*: Institut für Waldinventur und Waldwachstum, George-Augusts-Universität Göttingen. Büsingenweg 5, D-37077 Göttingen, Germany

E-mail udieguez@lugo.usc.es

Received 17 January 2006 **Revised** 24 May 2006 **Accepted** 6 July 2006

Available at <http://www.metla.fi/silvafennica/full/sf40/sf404631.pdf>

1 Introduction

The *Betula* genus is distributed throughout most of Europe, where it is mainly represented by two stand-forming tree species, *Betula pubescens* Ehrh. and *Betula pendula* Roth. In the Iberian

Peninsula, these two taxa are freely hybridized, and their taxonomy is somewhat confused, with each possessing a number of different nomenclatural identities (Castroviejo et al. 1990). Of the two taxa, *Betula pubescens* is the more oceanic, westerly distributed taxon while *Betula pendula*

is rarer and occupies more easterly and southerly, high altitude, mesic locations (Stevenson 2000). In Galicia, in north-western Spain, *Betula pubescens* Ehrh. (also referred to as *Betula alba* L. and *Betula pubescens* subsp. *celtiberica* Rothm. & Vasc. – Castroviejo et al. 1990) grows from 0–1700 m above sea level, although it is more abundant in the north-eastern mountains of this region at altitudes above 400–500 m. *Betula pubescens* requires high moisture all year round. Therefore, in the south of Galicia, where the summers are dry, it appears naturally only in areas near the streams or in wet locations. This species may be considered a fast-growing pioneer tree, which readily colonises open ground originated by human activity (burning, cutting or grazing) or natural disturbances. In Galicia there are currently 32 000 hectares of stands that include *Betula pubescens* as the main tree species (Xunta de Galicia 2001). The management of broadleaf species, including birch, is important for preserving biodiversity. Although birch is a native broadleaf species which grows well on soils of low fertility, no studies have been carried out to date to assess the basic productivity of the birch stands in this region, which is of primary concern in making forest management decisions.

To quantify site productivity, foresters usually evaluate the relationship between tree height and age. A specific model of this relationship is known as site index (S , usually defined as the dominant height of a population of trees at a specific base or reference age). Although true site productivity may not be fully represented by site index, S is the most widely accepted method for estimating site productivity of even-aged forests. The development of dominant height over time may be modeled using self-referencing functions (Northway 1985), which use the current states of the considered phenomena (in this case height and age), to determine their past or predict their future states (Cieszewski 2003). For this purpose, so called “panel data” (i.e., pooled cross-sectional and time series data) are required. A pooled database blends characteristics of both cross-sectional and time series data. Like cross-sectional data, it describes each of a number of individuals. Like time series data, it describes each single individual over time. Pooled data are important because they contain the information necessary

to deal with both the inter-temporal dynamics and the individuality of the entities being investigated (Dielman 1989, p. 2). These sequences may be obtained by repeated measurements on permanent sample plots, or by stem analysis in which the past growth of trees is reconstructed from the annual growth ring patterns of the tree. The latter enables a faster way of constructing site curves. Its primary disadvantage is that height growth of individual trees may not represent height growth of stands, depending on the criterion used to select the trees.

The dominant height-age relationship can be modeled using either static or dynamic site equations. Static equations have the following general form (omitting the vector of model parameters): $Y=f(t,S)$, where Y is the value of the function (e.g., predicted height) at age t , and S is a fixed base-age (of say, 20 years) site index. Dynamic site equations have general form (again omitting the vector of model parameters): $Y=f(t,t_0,Y_0)$, where Y is the value of the function at age t , and Y_0 is the reference variable defined as the value of the function at age t_0 . Advantages of dynamic site equations have been pointed out in several studies (e.g., Cieszewski and Bailey 2000, Cieszewski 2001, 2003): they are base-age-invariant, and define both 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 (t_0,Y_0) 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 is valid for both forward and backward predictions, which include the path invariance property, ensuring that the result of projecting first from t_0 to t_1 , and then from t_1 to t_2 , is the same as that of the one-step projection from t_0 to t_2 . Bailey and Cieszewski (2000) noted that some models that have been referred to as ‘base-age invariant’ or ‘reference-age invariant’ 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 that is known in forestry as the Algebraic Difference Approach (ADA), and that essentially involves replacing a base-model’s site-specific parameter

with its initial-condition solution. However, the concept of base-age-invariance of Bailey and Clutter (1974) also involves the estimation of the model parameters using a base-age-invariant (BAI) method (e.g., Bailey and Clutter 1974, García 1983, DuPlat and Tran-Ha 1986, 1997, Tait et al. 1988, Bégín and Schütz 1994, Gregoire et al. 1995, Gregoire and Schabenberger 1996, Cieszewski et al. 2000, Hall and Bailey 2001), which does not depend on the selection of any base age and which allows the use of all the data available. The main limitation of the ADA is that most models derived with it are either anamorphic or have single asymptotes (Bailey and Clutter 1974, Cieszewski and Bailey 2000).

Cieszewski and Bailey (2000) 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 the base equations can be expanded according to various theories about growth characteristics (e.g., asymptote, growth rate), allowing more than one parameter to be site-specific, and also the derivation of more flexible dynamic equations (see Cieszewski and Bailey 2000, Cieszewski 2001, 2002, 2003). This approach also includes simulation of concurrent polymorphism and multiple asymptotes, important properties of site equations (Cieszewski 2002).

The objective of the present study was to develop a base-age invariant equation, *sensu* Bailey and Clutter (1974), which realistically describes the dominant height growth of single-species, even-aged birch natural stands in Galicia.

2 Data

In the winter of 1998–1999 a network of 124 permanent plots was established in even-aged, birch-dominated stands (85% or more of the standing basal area consisting of birch – *Betula pubescens*). The plots were located throughout the area of distribution of this species in Galicia, and were subjectively selected to represent the existing range of ages, stand densities and sites. The plot size ranged from 200 m² to 1000 m², depending on stand density, in order to achieve

Table 1. Total tree height statistics given in meters (age class 5=3–7 years, etc.).

Age class	No of observations	Mean	Standard deviation	Minimum value	Maximum value
5	800	2.50	1.40	0.3	6.9
10	611	5.97	1.81	1.1	11.8
15	401	8.77	2.30	2.0	15.6
20	273	11.33	2.53	3.6	17.8
25	116	13.33	2.53	7.5	20.9
30	44	13.74	3.77	8.3	24.4
35	52	16.62	2.33	8.9	20.9
40	27	17.35	1.46	14.2	19.8
45	15	18.71	2.53	14.8	22.9
50	3	18.17	2.64	16.4	21.2
55	5	18.14	1.19	16.8	19.8
65	1	15.85		15.9	15.9

a minimum of 30 trees per plot. Although the smallest plot size (200 m²) may appear to be too small, only 2 plots were of this size. More than 50 trees were measured in these plots, which were selected to include extremes of site and stand conditions. The rest of the plots were larger than 500 m². In the cases where was possible, two undamaged dominant trees were destructively sampled, constituting a final sample of 214 dominant trees. 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 1.3 m above ground level and mean height of the dominant trees (considered as the 100 largest-diameter trees per hectare). An exchange of trees over time within this group of dominant trees could be expected in forest stands (Eriksson et al. 1997). This so-called “rank effect” may be a problem when using height growth for single trees, reconstructed by means of stem analysis, for site curve modelling. Because of the strong light-demanding behaviour of birches the rank effect was assumed to be of minor importance and was thus neglected in this study. Therefore, according to Tennent and Burkhart (1981), the two trees selected in each stand following the above procedure were considered adequate for representing mean dominant height development of the stand with reasonable fidelity. The trees were felled leaving stumps of average height 0.20 m; total bole length was measured to the nearest 0.01 m. The logs were cut at approximately 1 m intervals until the diameter was 7 cm and at 2

m intervals thereafter. The 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, we adjusted the height/age data from stem analysis to account for this bias using Carmean's method (Carmean 1972), and the modification proposed by Newberry (1991) for the topmost section of the tree, based on earlier studies (Dyer and Bailey 1987, Fabbio et al. 1994). The data were further examined for evidence of trees with past height growth abnormalities. Summary statistics including number of observations, mean, standard deviation, minimum, and maximum values of the selected data for modelling mean dominant height throughout the life of the stand were calculated for total tree height grouped by age classes (Table 1).

3 Methods

Cieszewski (2003) proposed a simplified approach of mixed-effects model for site curve development based on an adaptation of the nonlinear mixed-effects model defined by Lindstrom and Bates (1990). The method of Cieszewski (2003), which we employed in this study, is based on a one-point principle, i.e., it involves finding a model form that can describe the data well using only one parameter that varies with sites. If more than one parameter has to be site-specific, the varying parameters must be correlated with each other and defined as functions of fixed parameters and just one common varying parameter. After such re-parameterization, all except the varying parameter are uncorrelated parameters of the base model, and can be defined as pure fixed effects. The varying parameter is the only parameter related to the random site effects (Lindstrom and Bates 1990), and can be defined as a mixed-effect parameter (including both fixed and random effects) comprising the sum of the global mean for this parameter (ψ) (plus a site-specific effect (ζ_i)). The next steps involve: i) applying the GADA to $\psi + \zeta_i$ defined as X (a theoretical vector variable defining site quality with mean ψ and site-specific adjustments ζ_i with unknown distributional properties and an unknown magnitude

of values); ii) solving for X ; iii) substituting the solution for X in the explicit three-dimensional site equation ($Y=f(t,X)$) for initial conditions t_0 and Y_0 , ($Y=f(t,t_0,Y_0)$), where Y_0 comprises for each individual the sum of the observed height at age t_0 plus a site-specific effect e_{0i} interpreted as the random measurement and environmental errors); and iv) fitting in one stage using a stochastic regression approach, or a varying parameter regression with desired fitting criteria. In the final fitting procedure, the estimated heights (i.e., Y_0 at t_0) obtained by fitting a curve to the observed trend in the data for each individual growth series are used as predictors rather than the observed heights. This direct use of data, as constants, does not violate regression assumptions because the measurement and environmental errors associated with these data are estimated simultaneously with all the other parameters of the model (Cieszewski 2003).

3.1 Models

A number of researchers have pointed out desirable attributes of site quality equations (Bailey and Clutter 1974, Gadow and Hui 1999, Cieszewski and Bailey 2000, Cieszewski 2002, 2003). The most frequently listed criteria are: polymorphism, sigmoid growth pattern with an inflexion point, horizontal asymptote at old ages, logical behavior (height should be zero at age zero and equal to site index at the reference age; the curve should never decrease), theoretical basis or interpretation of model parameters derived by analytically tractable algebraic operations, base-age invariance, and path invariance. The fulfillment of these attributes depends on both the model construction method and the statistical procedures used for model fitting, and cannot always be achieved.

With these criteria in mind, we examined different three-parameter base models and tested several variants for each one using the simplified approach of the mixed-effects model of Cieszewski (2003). Both one and two parameters were considered to be site-specific. Some of the models tested were discarded in the first steps because of their poor performance when applied to the data, especially those which considered only one site-specific parameter in the base model. Therefore,

at least two parameters had to vary with site quality to fit all data well, while one parameter was common for all sites without compromising the overall fit before it was used for derivation of the dynamic equations using the GADA. Equations that include two site-specific parameters allow simulation of concurrent polymorphism and multiple asymptotes, which are both desirable characteristics of site equations (Cieszewski 2002). 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.

We finally focused our efforts on two dynamic equations. Using general notational convention, $a_1, a_2 \dots a_n$ denoted parameters in base models, while $b_1, b_2 \dots b_m$ were used for global parameters in subsequent GADA formulations. All the GADA based models have the general implicit form $Y=f(t,t_0,Y_0,b_1,b_2 \dots b_m)$.

The first model tested is based on the function proposed by Bertalanffy (1949, 1957) and studied by Richards (1959), whose integral form can, for simplicity, be represented as

$$Y = a_1 [1 - \exp(-a_2 t)]^{a_3} \tag{1}$$

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 (e.g., Burkhardt and Tennent 1977, Powers and Oliver 1978, García 1983, Biging 1985, Diéguez-Aranda et al. 2005).

To derive a model with both polymorphism and variable asymptotes from Eq. 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 response, 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 model derived from Eq. 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 $\exp(a_1')$ rather than a_1 and taking the natural logarithm – \ln – 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:

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

which defines the following GADA formulation:

$$\ln(Y) = X + (b_2 + b_3 / X) \ln[1 - \exp(-b_1 t)] \tag{2}$$

For this base equation, the solution for X involves finding roots of a quadratic equation and selecting the most appropriate root for substituting 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 (Cieszewski and Bailey 2000). The solution for X in Eq. 2 with initial condition values t_0 and Y_0 is

$$X_0 = \frac{1}{2} \left\{ \ln(Y_0) - b_2 L_0 \pm \sqrt{[b_2 L_0 - \ln(Y_0)]^2 - 4 b_3 L_0} \right\} \tag{3}$$

$$\text{where } L_0 = \ln[1 - \exp(-b_1 t_0)]$$

Selecting the root most likely to be real and positive (i.e., that involving addition rather than subtraction of the square root), and substituting it into Eq. 2 results in the following dynamic equation that provides polymorphic curves with variable asymptotes:

$$Y = Y_0 \left[\frac{1 - \exp(-b_1 t)}{1 - \exp(-b_1 t_0)} \right]^{b_2 + b_3 / X_0} \tag{4}$$

The second model may be viewed as a special case of the log-logistic class of models, which are equivalent to Hossfeld models (Cieszewski 2000) 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 (United States of

America) based on data from stem analysis of dominant trees. Cieszewski (2000, 2001, 2002, 2003) examined several GADA formulations using the log-logistic model as a base equation, generally expressed as

$$Y = \frac{a_1}{1 + a_2 \exp(-u)} \quad (5)$$

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 usually a linear combination of log-transformations of the independent variable(s) (Monserud 1984, Cieszewski 2000, 2001, 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, so that $u = a_3 \ln t$. Therefore, Eq. 5 can be simplified as the Hossfeld (1822) model

$$Y = \frac{a_1}{1 + a_2 t^{-a_3}} \quad (6)$$

Cieszewski (2002) replaced a_1 with a constant plus the unobserved site variable X , and a_2 by b_2/X , while $a_3 = b_3$, becoming the GADA formulation

$$Y = \frac{b_1 + X}{1 + b_2 / X t^{-b_3}} \quad (7)$$

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

$$X_0 = \frac{1}{2} \left[Y_0 - b_1 \pm \sqrt{(b_1 - Y_0)^2 + 4b_2 Y_0 t_0^{-b_3}} \right] \quad (8)$$

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

$$Y = \frac{b_1 + X_0}{1 + b_2 / X_0 t^{-b_3}} \quad (9)$$

3.2 Analysis

Parameter estimation for site models involves many different statistical considerations such as fitting criteria, error structures and independence of errors, as well as homogeneity of variance. An effective system for parameter estimation must be based on identifying individual trends represented in the data (Bailey and Clutter 1974, García 1983, Cieszewski et al. 2000, Cieszewski 2003). These trends can be modeled by considering that individuals' responses all follow a similar functional form with parameters that vary among individuals (local parameters) and parameters that are common for all individuals (global parameters). In practice both base-age specific (BAS) and base-age invariant (BAI) methods can be used. The assumption behind the BAS methods, which use selected data (i.e., site indices or heights at the given base age) as site-specific constants, is that the data measurements simultaneously do and do not contain measurement and environmental errors (on the left and right hand sides of the model respectively; e.g., Curtis et al. 1974, Monserud 1984), which is clearly untenable (Bailey and Clutter 1974). The assumption behind the BAI methods, which estimate site-specific effects, is that the data measurements always contain measurement and environmental errors (both on the left and right hand sides of the model) that must be modeled (Cieszewski 2003). In the BAI methods, the observation units – study plots or trees – of the panel data are assumed to constitute a random sample from a population of interest, while the observations within each unit are not random. Among the different BAI parameter estimation techniques reported in the introductory section of the present study, we estimated the random site-specific effects simultaneously with the fixed effects using the dummy variables method described by Cieszewski et al. (2000). In this method the initial conditions are specified as identical for all the measurements belonging to the same tree. The initial age can be, within limits, arbitrarily selected, even for each tree; however, age zero is not allowed. The height corresponding to the initial age is then simultaneously estimated for each tree along with all of the global model parameters during the fitting process. Unlike the traditional technique (BAS method) that requires

an arbitrary choice of base age prior to fitting the model, and then forcing the model through the chosen height/age point (site index), the BAI methods recognize that each measurement is made with error and, therefore, it seems unreasonable to force the model through any given measurement. Instead, the curve is fitted to the observed individual trends in the data. With this method all the data can be used, and there is no need to make any arbitrary choice regarding measurement intervals (i.e., it is a BAI method).

There are particular problems associated with the estimation of regression equations describing the behavior of individuals over time. In a single time series equation we often encounter the problem of autocorrelated errors. In a single cross-sectional regression equation we may have problems with a non-constant error variance. When combining cross-sectional and time series data we may find both problems occurring simultaneously. In addition, we may find that cross-sectional disturbances for different individuals at the same point in time may be correlated. Such contemporaneous correlation is an added problem specific to the analysis of pooled data (Dielman 1989, p. 2).

In the general formulation of the nonlinear mixed-effects model used in this study the error terms e_{ij} are assumed to be independent and identically distributed with zero mean. Nevertheless, because of the longitudinal nature of the data set used for model fitting, serial correlation (i.e., correlation between the residuals within the same individual) was expected. To account for this possible autocorrelation we modeled the error terms using a continuous-time autoregressive error structure (CAR(x)). This error structure permits the model to be applied to irregularly spaced, unbalanced data (Gregoire et al. 1995, Zimmerman and Núñez-Antón 2001), both of which are characteristics of many forestry data sets (West et al. 1984). The CAR(1) expands the error terms in the following way:

$$e_{ij} = d_1 \rho_1^{t_{ij} - t_{ij-1}} e_{ij-1} + \varepsilon_{ij}$$

where e_{ij} is the j th ordinary residual on the i th individual (i.e., the difference between the observed and the estimated heights of the tree i at age measurement j), $d_1 = 1$ for $j > 1$ and it is zero for $j = 1$, ρ_1 is the first-order autoregressive parameter

to be estimated, and $t_{ij} - t_{ij-1}$ is the time distance separating the j th from the j th-1 observations within each tree, $t_{ij} > t_{ij-1}$. In CAR(2) the error terms are expanded as:

$$e_{ij} = d_1 \rho_1^{t_{ij} - t_{ij-1}} e_{ij-1} + d_2 \rho_2^{t_{ij} - t_{ij-2}} e_{ij-2} + \varepsilon_{ij}$$

consequently, $d_2 = 1$ for $j > 2$ and it is zero for $j \leq 2$, ρ_2 is the second-order autoregressive parameter to be estimated, and $t_{ij} - t_{ij-2}$ is the time distance separating the j th from the j th-2 observations within each tree, $t_{ij} > t_{ij-2}$. The error terms for continuous-time autoregressive error structures of higher-order would be expanded similarly. In these cases ε_{ij} are now independent and identically distributed.

Visual examination of relevant graphs provided valuable insight into the statistical properties of the data. To evaluate the presence of autocorrelation and the order of the CAR(x) to be used, graphs representing residuals versus lag-residuals from previous observations within each tree were examined visually. The dummy variables method including the CAR(x) error structure was programmed using the SAS/ETS® MODEL procedure (SAS Institute Inc. 2004), which allows for dynamic updating of the residuals. To examine if cross-sectional correlation existed, we first ordered the residuals in ascending order by age and height, and then examined visually the graphs of residuals versus lag-residuals from previous observations within the same age. The heteroscedasticity was also examined visually, by plotting residuals as a function of predicted values, in order to investigate possible weighting factors.

3.3 Model Comparison

Comparison of the estimates for the two models tested was based on numerical and graphical analyses of the residuals. Two statistical criteria were examined: coefficient of determination for nonlinear regression (R^2) and root mean square error (RMSE). Although there are several shortcomings associated with 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 (Ryan 1997,

p. 424). The expressions of these statistics are summarized as follows:

$$R^2 = r_{Y_i \hat{Y}_i}^2$$

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

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.

Because the quality of fit does not necessarily reflect the quality of future prediction (Myers 1990, p. 168), assessment of the validity of the model with an independent data set is recommended (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, 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). Thus, because decisions have to be made with available information, it is better to wait to obtain new data before such validation is carried out.

The final step in evaluating the different fitted models involved visual examination of the residuals against the estimated values, and the fitted curves for different site indices overlaid on the profile plots of the trees. Visual inspection is an essential point in selecting the most accurate model because curve profiles may differ considerably, even though goodness-of-fit statistics are similar.

3.4 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. Inversely, site index and its associated base age may be used to estimate dominant height at any desired age. Therefore, the selection of a base age becomes an important issue when only one observation of a new individual is available.

The base age should be selected so that it is a reliable predictor of height at other ages, 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 applied to the stand.

In order to address the first consideration, different base ages and their corresponding observed heights were used to estimate heights at other ages (both forward and backward) for each tree. The results were compared with the values obtained from stem analysis, and the relative error in predictions (RE%) was then calculated as follows:

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

where Y_i , \hat{Y}_i and \bar{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.

It is important to note that the statistic described in this section is only meaningful if the site-specific parameters are discarded because of the lack of repeated data within the same individual. Otherwise, the site-specific parameters could be estimated and all the predictions would be identical whatever base age was selected (this is the essence of the BAI method used for model fitting).

4 Results and Discussion

We first fitted Eqs. 4 and 9 using nonlinear least squares without expanding the error terms to account for autocorrelation. A trend in residuals as a function of age-lag-residuals within the same tree was evident in both equations, as expected because of the longitudinal nature of the data used for model fitting. Fig. 1 (first row) provides an example with Eq. 9. After correction for autocorrelation using a second-order continuous-time autoregressive error structure, the trends in residuals disappeared (Fig. 1 third row). The fitted curve shapes were not significantly altered as compared with the models fitted without considering such correction. It should be noted that, from a practi-

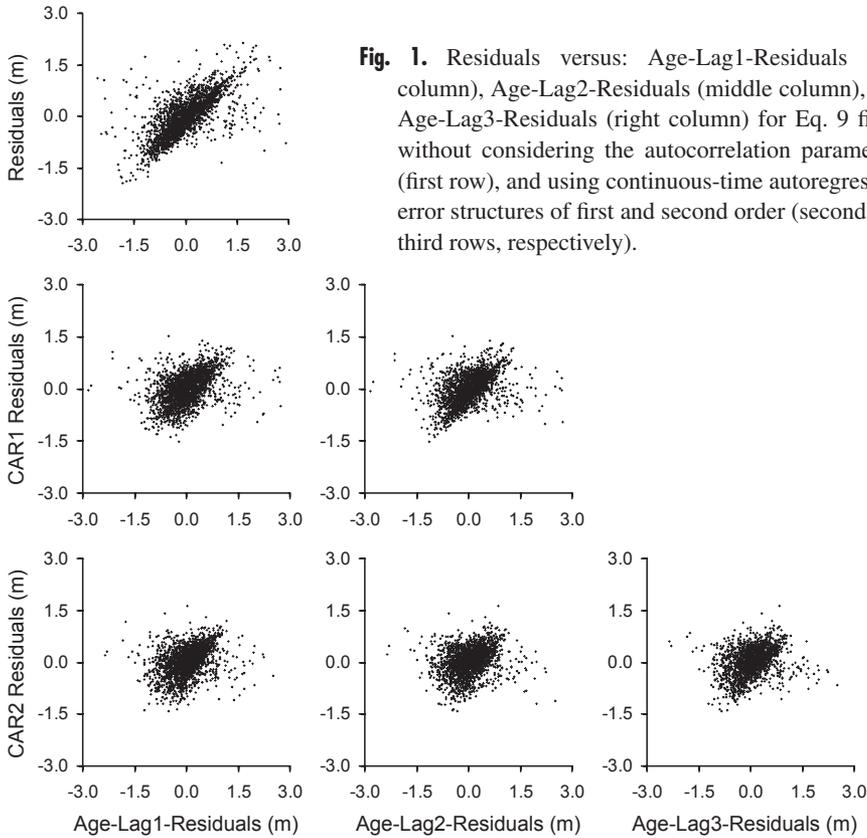


Fig. 1. Residuals versus: Age-Lag1-Residuals (left column), Age-Lag2-Residuals (middle column), and Age-Lag3-Residuals (right column) for Eq. 9 fitted without considering the autocorrelation parameters (first row), and using continuous-time autoregressive error structures of first and second order (second and third rows, respectively).

Table 2. Parameter estimates and goodness-of-fit statistics for the two models tested.

Model	Parameter	Estimate	Approx. Std. Err.	t-value	Approx. p-value	RMSE	R ²
Eq. 4	b_1	0.0480	0.0013	38.3	<0.0001	0.5043	0.9892
	b_2	-1.37	0.28	-4.9	<0.0001		
	b_3	8.65	0.89	9.7	<0.0001		
	ρ_1	0.940	0.012	79.4	<0.0001		
	ρ_2	0.852	0.013	65.1	<0.0001		
Eq. 9	b_1	19.8	1.7	11.7	<0.0001	0.5050	0.9892
	b_2	758	148	5.1	<0.0001		
	b_3	1.40	0.02	80.9	<0.0001		
	ρ_1	0.942	0.012	81.1	<0.0001		
	ρ_2	0.847	0.013	63.7	<0.0001		

cal point of view, the autocorrelation parameters are generally ignored when using the site model for predicting height and site index. The quantities e_{ij-1} and e_{ij-2} will probably not be known unless one is working with stem analysis data or with repeated measurements on the same plot, in which case they could be determined and used to

predict a future value. The main purpose in using the autocorrelation error structure was to obtain consistent estimates of the parameters and their standard errors.

The parameter estimates for each model and their corresponding goodness-of-fit statistics are shown in Table 2. All the parameters were signifi-

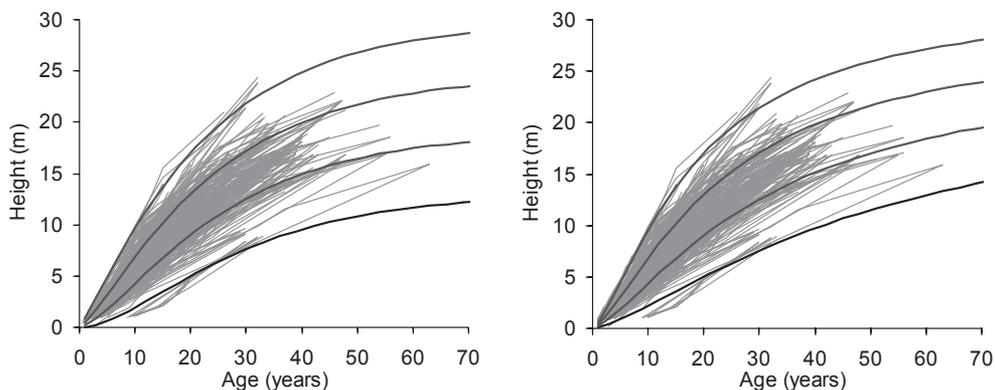


Fig. 2. Curves for site indices of 5, 9, 13 and 17 m at a reference age of 20 years overlaid on the trajectories of the observed heights over time for Eq. 4 (left) and Eq. 9 (right). The fittings were carried out using the dummy variables method and correcting autocorrelation with a second-order continuous-time autoregressive error structure.

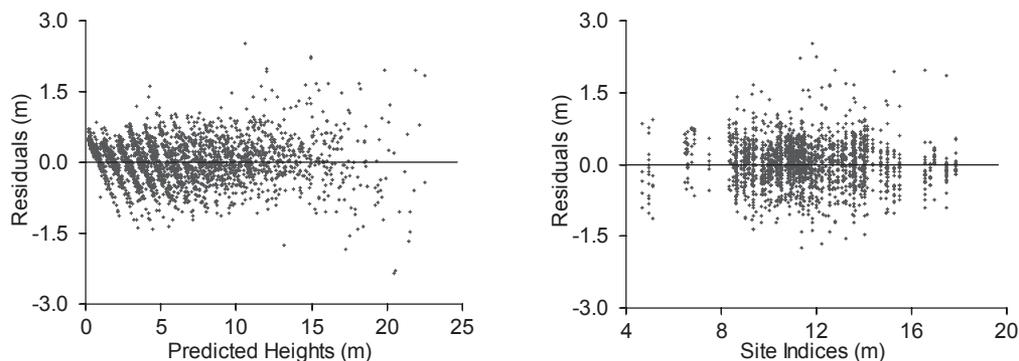


Fig. 3. Residuals versus: predicted heights as an illustration of potential heteroscedasticity (left), and site indices at a reference age of 20 years (right) for Eq. 9 fitted with a second-order continuous-time autoregressive error structure.

cant at the 1% level (including the site-specific parameters for each tree). The two models provided similar results, as indicated by the graphs that represent the fitted curves for different site indices overlaid on the trajectories of the observed heights over time (Fig. 2) and the goodness-of-fit statistics (Table 2). These models explained 98.9% of the total variance, and provided a random pattern of residuals around zero with homogeneous variance and no detectable significant trends, both for predicted heights and site indices at a reference age of 20 years (Fig. 3 exemplified with Eq. 9). The use of two-site-specific parameter equations allowed simulation of concurrent polymorphism

and multiple asymptotes, a desirable characteristic of site equations (Cieszewski 2002).

Graphs showing bias and root mean square errors in the height estimates for different age classes were analyzed (Fig. 4). Both models behaved similarly for juvenile and intermediate age classes, and also appeared to describe dominant height growth realistically for older age classes. Selection of the model was thus based on visual inspection of the graph of the fitted curves for different site indices overlaid on the trajectories of the observed heights over time (Fig. 2). In this figure a slight underestimate of dominant height at old ages can be inferred for Eq. 4.

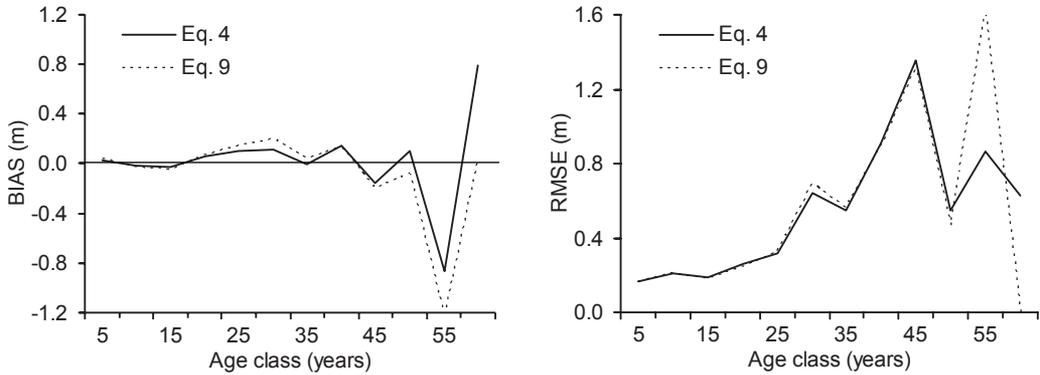


Fig. 4. Bias and root mean square error (RMSE) in height estimation for Eqs. 4 and 9 by 5-year age classes.

Thus, from the two dynamic equations finally evaluated for height growth prediction and site classification of birch stands in Galicia, Eq. 9, i.e. the GADA formulation derived by Cieszewski (2002) from the Hossfeld (1822) model by considering parameters a_1 and a_2 as related to site productivity, was selected:

$$Y = \frac{19.8 + X_0}{1 + 758 / X_0 t^{-1.40}} \tag{10}$$

where Y is the predicted height (meters) at age t (years),

$$X_0 = \frac{1}{2} \left[Y_0 - 19.8 + \sqrt{(19.8 - Y_0)^2 + 4 \times 758 Y_0 t_0^{-1.40}} \right]$$

and Y_0 and t_0 represent the predictor height and age. Note that the site-specific parameters are discarded in a way similar to that used to discard the autocorrelation coefficients (Cieszewski 2001), because the general use of the model involves making predictions using observed height and its associated measurement age in new individuals.

As already mentioned, dynamic equations derived with the GADA allow direct prediction of heights at any age from any other reference height and age. Therefore, 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 Eq. 10. 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 Eq. 10.

Eq. 10 fulfills most of the desirable properties that a site equation should possess, that is: it constitutes a parsimonious, three parameter, dynamic

site equation; it shows no trends in residuals; it predicts height equal to site index at base age and, although it is not defined for age zero *, the height limit when age tends to zero is zero; it is polymorphic with sigmoidal curve shapes and varying asymptotes; and it has the base-age and path invariance properties providing consistent predictions. Furthermore, it has been fitted using a base-age invariant method that accounts for site-specific and global effects, while addressing the error structure of the data.

The present investigation was based primarily on trees of height at 20 years of between 4 and 18 m. The ages ranged from 2 to more than 45 years. Therefore, the curves may be used over the entire rotation of the species in the region of study. The asymptotic values of the curves for site indices of 5, 9, 13 and 17 m at a reference age of 20 years resulted 23.0, 25.9, 29.1 and 32.4 m, respectively, which seems realistic. The curves should be used for ages greater than 5 years, since for younger ages erratic height growth may lead to erroneous classifications.

In selecting the base age, it was found that an age of 20 years was superior for predicting height at other ages (Fig. 5). Therefore, it is proposed as the base age to which site index will be referenced in order to classify the stands according to its productivity.

* This is not a fundamental property, albeit an identical model which is defined for $t=0$ is the generalization of this equation in Cieszewski (2001) (Eq. 21 in the original publication).

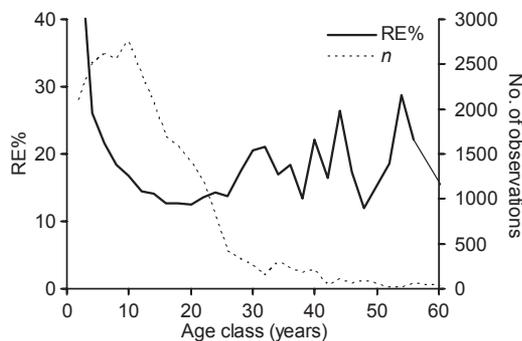


Fig. 5. Relative error in height prediction (solid line) related to choice of reference age for Eq. 9 by 2-year age classes and number of observations n (dotted line).

Finally, the curves generated using Eq. 10 were compared with the curves developed by Eriksson et al. (1997) for pure and mixed stands of *Betula pendula* and *Betula pubescens* in Sweden, to evaluate possible similarities in dominant height growth of birch stands at contrasting latitudes in Europe. A wider spectrum of site qualities was found in Sweden than in Galicia (site indices of between 3.5 and 21 m at a reference age of 20 years compared with site indices of between 4 and 18 m at the same reference age, respectively). The better site qualities found in Sweden may be explained by the presence of some birch stands on abandoned farmland (Karlsson et al. 1998), i.e. on relatively fertile soils, whereas in Galicia the stands are mainly growing on relatively infertile forest soils. Furthermore, the poorest site qualities found in Sweden may be explained by the shorter annual growth period in boreal areas. The two site quality systems for intermediate qualities were also compared (Fig. 6). The curves are quite similar until ages of 25–30 years, but culmination of growth is reached at younger ages in the Galician stands. Nevertheless, these trends should be corroborated using the trajectories of the observed heights over time, used to develop the site quality curves for the two regions analyzed.

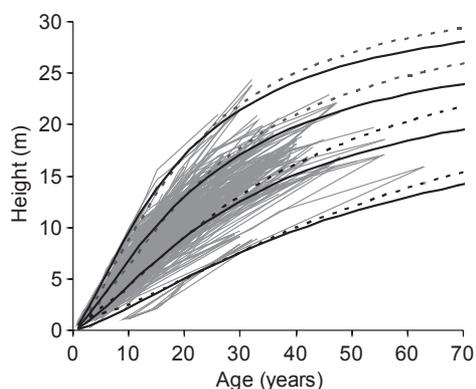


Fig. 6. Comparison of the Galicia curves (solid lines) for site indices of 5, 9, 13 and 17 m at a reference age of 20 years for Eq. 9 against the curves developed by Eriksson et al. (1997; dotted lines)* for birch stands in Sweden, overlaid on the profile plots of the data set used in this study.

* Because the curves of Eriksson et al. (1997) were developed for breast height age rather than stump age, we adjusted them for comparison by assuming that 2–3 years (depending on site quality) is the time that a tree needs to grow from stump height to breast height.

References

- Bailey, R.L. & Cieszewski, C.J. 2000. Development of a well-behaved site-index equation: jack pine in north-central Ontario: comment. *Canadian Journal of Forest Research* 30: 1667–1668.
- & Clutter, J.L. 1974. Base-age invariant polymorphic site curves. *Forest Science* 20: 155–159.
- Bégin, J. & Schütz, J.P. 1994. Estimation of total yield of Douglas-fir by means of incomplete growth series. *Annals of Forest Science* 51: 345–355.
- Bertalanffy, L.v. 1949. Problems of organic growth. *Nature* 163: 156–158.
- 1957. Quantitative laws in metabolism and growth. *Quarterly Review of Biology* 32: 217–231.
- Biging, G.S. 1985. Improved estimates of site index curves using a varying-parameter model. *Forest Science* 31: 248–259.
- Burkhart, H.E. & Tennent, R.B. 1977. Site index equations for radiata pine in New Zealand. *New Zealand Journal of Forestry Science* 7: 408–416.
- Carmean, W.H. 1972. Site index curves for upland oaks in the Central States. *Forest Science* 18: 109–120.

- Castroviejo, S., Laínz, M., López González, G., Montserrat, P., Muñoz Gardemena, F., Paiva, J. & Villar, L. 1990. Flora Ibérica: Plantas vasculares de la Península Ibérica e Islas Baleares. Volumen II: Platanaceae–Plumbaginaceae (partim.). Madrid: RJB (CSIC).
- Cieszewski, C.J. 2000. Analytical solution to the generalized log-logistic equation. *Forest Science* 46: 291–296.
- 2001. Three methods of deriving advanced dynamic site equations demonstrated on inland Douglas-fir site curves. *Canadian Journal of Forest Research* 31: 165–173.
- 2002. Comparing fixed- and variable-base-age site equations having single versus multiple asymptotes. *Forest Science* 48: 7–23.
- 2003. 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* 49: 539–554.
- 2004. GADA derivation of dynamic site equations with polymorphism and variable asymptotes from Richards, Weibull, and other exponential functions. University of Georgia PMRC-TR 2004-5.
- & Bailey, R.L. 2000. Generalized Algebraic Difference Approach: Theory based derivation of dynamic site equations with polymorphism and variable asymptotes. *Forest Science* 46: 116–126.
- , Harrison, M. & Martin, S.W. 2000. Practical methods for estimating non-biased parameters in self-referencing growth and yield models. University of Georgia PMRC-TR 2000-7.
- Curtis, R.O., DeMars, D.J. & Herman, F.R. 1974. Which dependent variables in site index-height-age regressions? *Forest Science* 20: 74–87.
- Diéguez-Aranda, U., Burkhart, H.E. & Rodríguez-Soalleiro, R. 2005. Modeling dominant height growth of radiata pine (*Pinus radiata* D. Don) plantations in north-western Spain. *Forest Ecology and Management* 215(1–3): 271–284.
- Dielman, T.E. 1989. Pooled cross-sectional and time series data analysis. Marcel Dekker Inc., New York.
- Duplat, P. & Tran-Ha, M. 1986. Modèles de croissance en hauteur dominante pour le hê tre, le sapin pectiné et le pin sylvestre (dans le Massif de l'Aigoual). Office National des Forêts, Section Technique, Paris, Document n° 86.1.
- & Tran-Ha, M. 1997. Modelling the dominant height growth of sessile oak (*Quercus petraea* Liebl.) in France. Inter-regional variability and effect of the recent period (1959–1993). *Annals of Forest Science* 54: 611–634.
- Dyer, M.E. & Bailey, R.L. 1987. A test of six methods for estimating true heights from stem analysis data. *Forest Science* 33: 3–13.
- Eriksson, H., Johansson, U. & Kiviste, A. 1997. A site-index model for pure and mixed stands of *Betula pendula* and *Betula pubescens* in Sweden. *Scandinavian Journal of Forest Research* 12: 149–156.
- Fabbio, G., Frattegiani, M. & Manetti, M.C. 1994. Height estimation in stem analysis using second differences. *Forest Science* 40: 329–340.
- Gadow, K.v. & Hui, G.Y. 1999. Modelling forest development. Kluwer Academic Publisher, Dordrecht. 213 p.
- García, O. 1983. A stochastic differential equation model for the height growth of forest stands. *Biometrics* 39: 1059–1072.
- Gregoire, T.G. & Schabenberger, O. 1996. A non-linear mixed-effects model to predict cumulative bole volume of standing trees. *Journal of Applied Statistical Science* 23: 257–271.
- , Schabenberger, O. & Barrett, J.P. 1995. Linear modelling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. *Canadian Journal of Forest Research* 25: 137–156.
- Hall, D.B. & Bailey, R.L. 2001. Modeling and prediction of forest growth variables based on multi-level nonlinear mixed models. *Forest Science* 47: 311–321.
- Hossfeld, J.W. 1822. *Mathematik für Forstmänner, Ökonomen und Cameralisten* (Gotha, 4. Bd., S. 310).
- Karlsson, A., Albrektson, A., Forsgren, A. & Svensson, L. 1998. An analysis of successful natural regeneration of downy and silver birch on abandoned farmland in Sweden. *Silva Fennica* 32(3): 229–240.
- Kozak, A. & Kozak, R. 2003. Does cross validation provide additional information in the evaluation of regression models? *Canadian Journal of Forest Research* 33: 976–987.
- Lindstrom, M.J. & Bates, D.M. 1990. Nonlinear mixed effects models for repeated measures data. *Biometrics* 46: 673–687.
- Monserud, R.A. 1984. Height growth and site index curves for inland Douglas-fir based on stem analysis data and forest habitat type. *Forest Science* 30: 943–965.

- Myers, R.H. 1990. Classical and modern regression with applications. 2nd edition. Duxbury Press, Belmont, California.
- Newberry, J.D. 1991. A note on Carmean's estimate of height from stem analysis data. *Forest Science* 37: 368–369.
- Northway, S.M. 1985. Fitting site index equations and other self-referencing functions. *Forest Science* 31: 233–235.
- Powers, R.F. & Oliver, W.W. 1978. Site classification of ponderosa pine stands under stocking control in California. USDA Forest Service Res. Pap. PSW-128.
- Richards, F.J. 1959. A flexible growth function for empirical use. *Journal of Experimental Botany* 10: 290–300.
- Ryan, T.P. 1997. Modern regression methods. John Wiley & Sons, New York.
- SAS Institute Inc. 2004. SAS/ETS® 9.1 User's Guide. Cary, NC: SAS Institute Inc.
- Stevenson, A.C. 2000. The Holocene forest history of the Montes Universales, Teruel, Spain. *The Holocene* 10: 603–610.
- Tait, D.E., Cieszewski, C.J. & Bella, I.E. 1988. The stand dynamics of lodgepole pine. *Canadian Journal of Forest Research* 18: 1255–1260.
- Tennent, R.B. & Burkhart, H.E. 1981. Selection of stem analysis trees for site index curve construction. New Zealand Forest Service, Forest Research Institute, Forest Mensuration Report 56. 12 p. (unpublished).
- West, P.W., Ratkowsky, D.A. & Davis, A.W. 1984. Problems of hypothesis testing of regressions with multiple measurements from individual sampling units. *Forest Ecology and Management* 7: 207–224.
- Xunta de Galicia 2001. O monte galego en cifras. Dirección Xeral de Montes e Medio Ambiente Natural. Conseller'a de Medio Ambiente. Santiago de Compostela.
- Zimmerman, D.L. & Núñez-Antón, V. 2001. Parametric modelling of growth curve data: An overview (with discussion). *Test* 10: 1–73.

Total of 47 references