MIMICKING NATURAL VARIABILITY IN TREE HEIGHT OF PINE SPECIES USING A STOCHASTIC HEIGHT-DIAMETER RELATIONSHIP

MARCOS BARRIO-ANTA*, ULISES DIÉGUEZ-ARANDA

Departamento de Ingeniería Agroforestal, Universidad de Santiago de Compostela, Escuela Politécnica Superior, Campus universitario, 27002 Lugo, Spain.

FERNANDO CASTEDO-DORADO

Departamento de Ingeniería Agraria, Universidad de León, Escuela Superior y Técnica de Ingeniería Agraria, Campus de Ponferrada, 24400 Ponferrada, Spain.

JUAN GABRIEL ÁLVAREZ-GONZÁLEZ, and ALBERTO ROJO-ALBORECA

Departamento de Ingeniería Agroforestal, Universidad de Santiago de Compostela, Escuela Politécnica Superior, Campus universitario, 27002 Lugo, Spain.

(Received for publication 16 November 2005; revision 20 March 2006)

ABSTRACT

Measuring the height of a tree takes longer than measuring its diameter at breast height and often the heights of only a subset of trees of known diameter are measured in forest inventories. Accurate height-diameter equations must therefore be used to predict the heights of the remaining trees. Two trees within the same stand and that have the same diameter are not necessarily of the same height; therefore we developed a deterministic equation, using the Schnute function, and then added a stochastic component to it, to mimic the real natural variability in height. The stochastic approach uses the standard error of a new observation in a similar way to the method of obtaining the prediction interval for an individual (new) in a regression model, but rather than using the *t* value corresponding to a fixed limit for all the trees, it uses a pseudo-random number having a normal distribution N(0,1) for each observation. The stochastic approach was evaluated with data from four thinning trials located in single-species, even-aged stands of the most commercially important pines in Galicia (north-western Spain). More realistic height predictions were obtained than with the deterministic model for individual diameter classes, as demonstrated by the results of the Kolmogorov Smirnov test and by visual analysis of box plot graphs.

Keywords: trees; height diameter-models; stochastic component.

* Corresponding author: barrio@lugo.usc.es

INTRODUCTION

Individual tree heights and diameters are essential measurements in forest inventories, and are used for estimating timber volume, site index, and other important variables related to forest growth and yield, succession, and carbon budget models (Peng 2001). Moreover, when combined with crown ratio data or models, height-diameter equations can be used to predict tree height and to estimate the change in crown ratio to be incorporated in individual tree growth models. Although the time taken to measure tree heights has been reduced by the use of devices that employ ultrasound or laser pulses to measure distances, it still takes longer than measuring the diameter at breast height. For this reason, often the heights of only a subset of trees of known diameter are measured, and accurate local height-diameter equations must be used to predict the heights of the remaining trees to reduce the costs involved in data acquisition. These equations use diameter at breast height as the only predictor variable for estimating total height. Therefore, accurate prediction of tree heights is critical in forest inventory, model simulation, and sustainable forest management decision-making (Curtis 1967; Botkin *et al.* 1972).

If stand conditions vary greatly within a forest, a height regression may be derived separately for each stand, or a generalised function, which includes stand variables to account for the variability, may be developed (Curtis 1967; Zhang *et al.* 1997; Schröeder & Álvarez 2001; López-Sánchez *et al.* 2003; Sharma & Zhang 2004).

Natural processes can be considered as the sum of deterministic and stochastic components. Knowledge of the deterministic component is obtained from the model's functional relationships. The stochastic component represents effects beyond our present predictive capability, or that are deliberately omitted from the model (Stage 2003). Thus, if assessment of the variability of the outcomes is one of the objectives of the modelling procedures, most of the predictive equations would require a random component.

In this sense, it is well known that trees with the same diameter within the same stand are not necessarily of the same height. Therefore, a deterministic model does not seem appropriate for mimicking the real natural variability in height (Castedo *et al.* 2005). To deal with this, an unstructured random component can be added to the deterministic model predictions. This approach assumes that the stochastic effects are entirely random and unstructured, and adds a normally distributed random component with variance equal to the residual mean square of a previously fitted model to the model estimations (Dennis *et al.* 1985; Fox *et al.* 2001). Many other studies have found that incorporation of this random component is important for preserving variability in predictions (Stage 1973, 2003; Daniels & Burkhart 1975; Stage & Wykoff 1993; Castedo *et al.* 2005).

In view of the previous considerations, the objective of the present study was to evaluate the performance of a stochastic height-diameter approach in mimicking

the observed natural variability in tree height. For that purpose, a large data set of height-diameter pairs corresponding to stands of the most commercially important pine species in Galicia (north-western Spain) was analysed.

MATERIAL

We used data from the first measurement of four thinning trials established in single-species, even-aged stands located throughout Galicia to evaluate the height-diameter stochastic approach. The first trial was established in *Fonsagrada* (province of Lugo), in a 40-year-old stand of Scots pine (*Pinus sylvestris* L.). The second trial was established in *O Carballiño* (province of Pontevedra) in a 12-year-old stand of maritime pine (*Pinus pinaster* Ait.). The third and fourth thinning trials were established in *Begonte* (province of Lugo) in two 12-year-old stands of radiata pine (*Pinus radiata* D. Don), which differed in seed origin (New Zealand and local). Each thinning trial consisted of 12 plots of 625–900 m^2, in which four thinning regimes were evaluated on three different occasions. The stand conditions within each thinning trial were similar and thus we considered the data obtained as belonging to only four different stands corresponding to the four different thinning trials.

Diameter at breast height (*d*) was measured to the nearest 0.1 cm in all the trees. Total tree height (*h*) was measured to the nearest 0.1 m in a sample of one-third of the trees in each plot. Summary statistics, including the mean, minimum, maximum, and standard deviation (SD) of these variables are shown in Table 1.

TABLE 1–Characteristics of the tree samples used for model fitting.

Thinning trial	Number of trees	Diameter at breast height (cm)				Total height (m)			
		Mean	Min.	Max.	SD	Mean	Min.	Max.	SD
Pinus sylvestris	488	22.4	7.0	34.8	7.2	19.6	9.6	25.4	2.8
Pinus pinaster	1062	11.2	2.9	19.8	3.3	9.2	4.8	12.4	1.2
Pinus radiata New Zealand	651	13.7	5.0	26.6	4.8	11.4	4.7	17.0	2.3
Pinus radiata Local	668	14.6	3.1	29.5	5.6	12.2	5.1	18.8	2.6

METHODS

Candidate Models

A large number of local height-diameter equations have been reported (e.g., Curtis 1967; García 1974; Wykoff *et al.* 1982; Huang *et al.* 1992, 2000; Fang & Bailey 1998; Peng 1999; Soares & Tomé 2002; Temesgen & Gadow 2004), most of which

exhibit S-shaped or concave-shaped patterns. The selection of a functional form for the height-diameter relationship should consider the following mathematical properties: (1) monotonic increment, (2) functional inflection point, and (3) asymptotic value (Lei & Parresol 2001). The number of parameters (flexibility), possible biological interpretation of the parameters (e.g., upper asymptote, maximum or minimum growth rate), and satisfactory predictions for height-diameter relationships are also important features (Peng 2001). Sigmoid or S-shaped functions are preferred because they posses all three of the above properties. However, if a data set includes only larger or older trees beyond the inflection point, then a model generating a concave curve will probably work best, although it may perform poorly in the lower range where there are no data (poor extrapolation properties) (Lei & Parresol 2001).

Initially, we considered four non-linear growth functions — Bertalanffy-Richards (Bertalanffy 1949, 1957; Richards 1959), Korf (cited by Lundqvist 1957), Weibull (1951), and Schnute (1981); these have frequently been used for describing the height-diameter relationship (*see* Yang *et al.* 1978; Zeide 1989; Zhang 1997; Fang & Bailey 1998; Peng *et al.* 2001; Sharma & Zhang 2004; Temesgen & Gadow 2004). Preliminary analysis of these models showed that they all provided similar results. Schnute's model has an advantage over the other models in that it is easy to fit and quick to achieve convergence for any database (Bredenkamp & Gregoire 1988; Lei & Parresol 2001), even with small data sets (Castedo *et al.* 2005), and thus it was the only one considered for further study:

$$h = \left[h_1^b + (h_2^b - h_1^b)\frac{1 - e^{-a(d-d_1)}}{1 - e^{-a(d_2-d_1)}} \right]^{1/b} + \varepsilon \qquad (1)$$

where: h = total height of the tree
d = tree diameter at breast height
d_1 = diameter at breast height of a small tree (lower range of data)
d_2 = diameter at breast height of a large tree (upper range of data)
h_1 = parameter representing mean tree height at d_1
h_2 = parameter representing mean tree height at d_2
b = incremental acceleration in growth rate
a = constant acceleration in growth rate
ε = residual error.

In the context of height-diameter modelling it is common practice to force the curve to pass through the point (0, 1.3) to prevent negative height estimates for small trees, although in reality, when d is zero, h can take any value between 0 and 1.3. This assumption does not affect the height-diameter relationship because data corresponding to heights lower than 1.3 are not used in the fitting process. Taking

into account these considerations, we let $d_1 = 0$ and $h_1 = 1.3$. This resulted in the modified Schnute model:

$$h = \left[1.3^b + (h_2^b - 1.3^b)\frac{1 - e^{-ad}}{1 - e^{-ad_2}} \right]^{1/b} + \varepsilon \quad (2)$$

Model Fitting and Evaluation

A fundamental assumption for the least squares method is that the errors (ε) in regression models are independent and identically distributed (constant variance) with a mean value of zero. However, forest modellers are often faced with heteroscedasticity in their data, which would lead to non-minimum variance parameter estimates and unreliable prediction intervals. The solution to this problem is to weight each observation during the fitting process by the inverse of its variance ($\sigma^2_{\varepsilon i}$). If the variance is unknown, the problem becomes one of estimating the proper weight for each observation (Parresol 1993).

Scatter plots of total tree height against diameter at breast height showed an increasing variance in height as the values of the independent variable increased, for the *P. radiata* data sets. This type of heteroscedasticity is common and is usually modelled as a power function (Draper & Smith 1981; Neter *et al.* 1996). The base of the exponent was selected as d, that is, $\sigma^2_{\varepsilon i} = d^k_i$. The most reasonable value of the exponential term k should provide the most homogeneous studentised residual plot (Huang *et al.* 2000), with a reduction in the standard errors of the estimates as compared with the parameter estimates of the unweighted model (Parresol 1993). The k number can be obtained by iteratively testing different values (e.g., from 0.1 to 2). An alternative theoretical optimisation of k can be achieved using the method suggested by Harvey (1976), which consists of using the estimated errors of the unweighted model ($\hat{\varepsilon}_i$) as the dependent variable in the error variance model, that is:

$$\hat{\varepsilon}^2_i = \gamma d^k_i \quad (3)$$

or

$$\ln\hat{\varepsilon}^2_i = \ln\gamma + k\ln d_i \quad (4)$$

The k parameter of Equation (4) was estimated using linear regression. Parameters of Equation (2) were estimated using generalised non-linear least squares (GNLS), also known as weighted non-linear regression, using the NLIN procedure of SAS/STAT® (SAS Institute Inc. 2004). The weight function was specified with the option _WEIGHT_=$1/d^k_i$.

The goodness-of-fit of the model was evaluated using two statistics: the coefficient of determination for non-linear regression (R^2) and the root of the weighted mean squared error ($RMSE_\psi$). Although several shortcomings have been stated against the use of the R^2 in non-linear regression (Neter *et al.* 1996), the general usefulness

of some global measure of model adequacy would seem to override some of those limitations (Ryan 1997, p. 424). The expressions for these statistics are as follows:

$$R^2 = r^2_{h_i\hat{h}_i} \tag{5}$$

$$RMSE_{\psi} = \sqrt{\frac{\sum_{i=1}^{n} \psi_i (h_i - \hat{h}_i)^2}{n-p}} \tag{6}$$

where $r_{h_i\hat{h}_i}$ is the correlation coefficient for a linear regression between the measured (h_i) and estimated ($\hat{h}_i$) values of the dependent variable (*see* Ryan 1997, pp. 419 and 424); ψ_i is the weight function; n is the total number of observations used to fit the model; and p is the number of model parameters.

Another important step in the evaluation of the fitted models was to perform a graphical analysis of the residuals, searching for dependencies or patterns that indicated systematic discrepancies.

Stochastic Height Prediction

Basically, the variance components from a regression model linked with random numbers with a certain distribution can be used to create a stochastic prediction.

In matrix notation, a general non-linear model can be written as:

$$\mathbf{y} = f(\mathbf{X},\boldsymbol{\beta}) + \boldsymbol{\varepsilon} \tag{7}$$

where $\mathbf{y}$ is a vector of the dependent variable (total height of the tree in our case), $\mathbf{X}$ is a non-stochastic matrix of tree variables, $\boldsymbol{\beta}$ is a vector of unknown parameters, and $\boldsymbol{\varepsilon}$ is a vector of random errors.

The generalised non-linear least squares (GNLS) estimate of the vector $\boldsymbol{\beta}$ is the value that minimises the sum of squared errors:

$$SSE(\boldsymbol{\beta}) = \boldsymbol{\varepsilon}'\boldsymbol{\psi}(\boldsymbol{\theta})^{-1}\boldsymbol{\varepsilon} = [\mathbf{y} - \mathbf{f}(\mathbf{X},\boldsymbol{\beta})]'\boldsymbol{\psi}(\boldsymbol{\theta})^{-1}[\mathbf{y} - \mathbf{f}(\mathbf{X},\boldsymbol{\beta})] \tag{8}$$

where $\boldsymbol{\psi}(\grave{\mathbf{e}})$ is a diagonal matrix of weights that is dependent on a fixed number q of parameters denoted by the $(q \times 1)$ vector $\boldsymbol{\theta}$. The dimension of $\boldsymbol{\theta}$ and the precise way in which $\boldsymbol{\psi}$ depends on $\boldsymbol{\theta}$, depend on the assumptions made about the error process. Under suitable conditions, the GNLS estimate $\mathbf{b}$ will be approximately normally distributed, with mean $\boldsymbol{\beta}$ and variance-covariance matrix that is consistently estimated by:

$$\mathbf{S}^2(\mathbf{b}) = RMSE_{\psi}^2 \left[\mathbf{Z}(\mathbf{b})'\boldsymbol{\psi}(\hat{\theta})^1\mathbf{Z}(\mathbf{b})\right]^{-1} \tag{9}$$

where the scalar $RMSE_{\psi}^2$ is the weighted regression mean squared error, i.e., Equation (8) divided by the degrees of freedom:

$$RMSE_{\psi}^2 = \frac{SSE(\mathbf{b})}{n-p} = \frac{[\mathbf{y} - \mathbf{f}(\mathbf{X},\mathbf{b})]'\boldsymbol{\psi}(\hat{\theta})^{-1}\,[\mathbf{y} - \mathbf{f}(\mathbf{X},\mathbf{b})]}{n-p} \tag{10}$$

and $\mathbf{Z}(\boldsymbol{\beta})$ is the partial derivatives matrix with respect to each parameter:

$$\mathbf{Z}(\boldsymbol{\beta}) = \begin{bmatrix} \frac{\partial f(\mathbf{x}_1,\boldsymbol{\beta})}{\partial \beta_1} & \cdots & \frac{\partial f(\mathbf{x}_1,\boldsymbol{\beta})}{\partial \beta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{x}_n,b)}{\partial \beta_1} & \cdots & \frac{\partial f(\mathbf{x}_n,\boldsymbol{\beta})}{\partial \beta_p} \end{bmatrix} \quad (11)$$

This information can be used to form tests of hypotheses and interval estimates for **b** in a similar way to that for linear least squares methods.

In the special case where we want to know the prediction interval for an individual (new) outcome drawn from the distribution of y_i, the variance is:

$$\text{var}(\hat{y}_{i(\text{new})}) = \text{RMSE}_{\psi}^2 \, \psi_i(\hat{\boldsymbol{\theta}}) + \mathbf{z}(\mathbf{b})'_i \, \mathbf{S}^2 \, (\mathbf{b}) \, \mathbf{z} \, (\mathbf{b})_i \quad (12)$$

where $\psi_i(\hat{\boldsymbol{\theta}})$ is the ith diagonal element of the estimated weight matrix $\psi(\hat{\boldsymbol{\theta}})$ i.e., the value of the weight function at observation i, and the partial derivatives vector $\mathbf{z}(\mathbf{b})'_i$ is the ith row of $\mathbf{Z}(\mathbf{b})$ (*see* Equation (11)).

Thus, the prediction interval for a new observation drawn from the distribution of y_i is:

$$\hat{y}_i \pm t_{\alpha/2,n-p} s_{\hat{y}_{i(\text{new})}} \quad (13)$$

where $s_{\hat{y}_i}$ is the standard error of prediction of a new observation, obtained from Equation (12) as:

$$s_{\hat{y}_{i(\text{new})}} = \sqrt{\text{var}(\hat{y}_{i(\text{new})})} = \sqrt{RMSE_{\psi}^2 \, \psi_i(\hat{\boldsymbol{\theta}}) + \mathbf{z}(\mathbf{b})'_i \, \mathbf{S}^2(\mathbf{b})\mathbf{z}(\mathbf{b})_i} \quad (14)$$

The stochastic approach uses this standard error in a similar way to how the prediction interval is obtained for an individual (new) in a regression model. Instead of using the t value corresponding to a fixed limit for all the trees, e.g., –1.96 and 1.96 respectively for a probability of 2.5% and 97.5% and infinite degrees of freedom, it is substituted by a value randomly generated from the inverse of the normal distribution function for each individual. Thus, the expression used to assign stochastically the height to each tree in a sample is:

$$\hat{y}_{i(\text{stoch})} = \hat{y}_i + F_U^{-1} \, s_{\hat{y}_{i(\text{new})}} \quad (15)$$

where $\hat{y}_{i(\text{stoch})}$ is the stochastic height estimation, $\hat{y}_i$ is the deterministic height obtained from Equation (2) (the fitted model without considering the stochastic component), F_U^{-1} is the inverse of the standard normal distribution function for U — a uniform random variable in the interval (0,1) — and $s_{\hat{y}_{i(\text{new})}}$ is the standard error of prediction for a new individual.

The F_U^{-1} values were obtained through the NORMAL(SEED) function in SAS/STAT® package (SAS Institute Inc. 2004). The NORMAL function is a scalar function that returns a pseudo-random number having a normal distribution, with a mean of 0 and a standard deviation of 1. The function requires an initial starting

point, called a seed, which either the user or the computer clock supplies, and which must be a non-negative integer: if a positive seed is used, it is possible to replicate the stream of random numbers, while if zero is used as the seed, the computer clock initialises the stream and it is not replicable. A zero value was used as seed in the present study.

The performance of the stochastic approach proposed was evaluated for diameter classes of 5 cm width and for each thinning trial. A comparison between the observed height frequencies and the variability of the deterministic and stochastic approaches was carried out. The Kolmogorov-Smirnov test was used to compare the observed and the estimated height frequencies of the two approaches. Box plot graphs were used to visually contrast the means, medians, and 25/75 percentiles of the height distributions.

RESULTS AND DISCUSSION

Model Fitting and Evaluation

As unequal error variance occurred in the *P. radiata* height-diameter data, GNLS techniques with the weights chosen as _WEIGHT_=$1/d^{0.336}$ and _WEIGHT_=$1/d^{0.526}$ were applied for stands of New Zealand and local seed origin, respectively. These weight functions stabilised the variance and provided a homogeneous residual plot over the full range of the predicted values (Fig. 1).

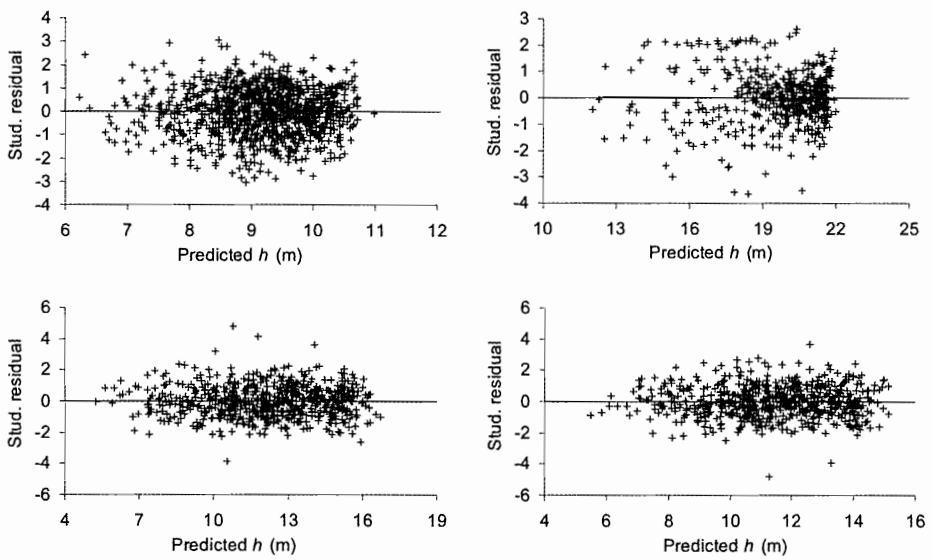

FIG. 1–Plots of the studentised residuals against the predicted heights for the four data sets.

Although in general the Schnute function provided a reasonable fit, taking into account the great variability in the height-diameter relationship in the databases, notable differences were found among the predictive abilities for each species. The models produced satisfactory fits for the *P. radiata* thinning trials; however, the percentage of variability explained by the model was low for the *P. pinaster* and *P. sylvestris* data sets (*see* Table 2). All the parameters were significant at the p = 95% level and showed reasonable values.

TABLE 2–Parameter estimates (standard error in brackets) and goodness-of-fit statistics for the Schnute (1981) model and the four data sets under analysis.

Thinning trial	Weight function	Parameter estimates			R^2	$RMSE_\psi$
		a	b	h_2		
Pinus sylvestris	1	0.1069 (0.0176)	0.9516 (0.2575)	21.9859 (0.2805)	0.5188	1.918
Pinus pinaster	1	0.0896 (0.0246)	1.9051 (0.3197)	10.9793 (0.1536)	0.5091	0.828
Pinus radiata New Zealand	$1/d^{0.336}$	0.0618 (0.0155)	1.2779 (0.1817)	15.1430 (0.2136)	0.7513	1.134
Pinus radiata Local	$1/d^{0.526}$	0.0551 (0.0120)	1.2979 (0.1432)	16.7094 (0.2545)	0.7880	1.214

Stochastic Height Prediction

Deterministic values of the height $\hat{y}_i$ were obtained using Equation (2) with the parameter estimates obtained for each data set (*see* Table 2). The prediction of the stochastic values $\hat{y}_{i(\text{stoch})}$ also involved the following steps:

(1) Calculation of the standard error of the prediction $s_{\hat{y}_{i(\text{new})}}$ by means of Equation (14), which requires obtaining the variance-covariance matrix of the parameter estimates $\mathbf{S}^2(\mathbf{b})$ (Equation (9)). The partial derivatives vector $\mathbf{z}(\mathbf{b})_i$ and its transpose $\mathbf{z}(\mathbf{b})'_i$ can be obtained by substituting the estimated parameters in the partial derivatives of Equation (2) shown in the Appendix.

(2) Calculation of the inverse of the standard normal distribution function for one pseudo-random number in the interval (0,1).

(3) Estimation of the stochastic height $\hat{y}_{i(\text{stoch})}$ from Equation (15) by substituting the values obtained in the previous steps.

Comparison of the observed, deterministic, and one random stochastic height distribution per diameter class showed substantial differences between the two

approaches, even though the mean height values per diameter class obtained by each approach were very similar (Fig. 2, Table 3). The observed height distribution generally followed a normal distribution within diameter classes with a high degree of variability in the four data sets evaluated. This was particularly true for the *P. radiata* data sets. This may be explained by the earlier differentiation of trees in crown classes in *P. radiata* than in *P. pinaster* plantations. In the latter, the delay of differentiation of crown classes due to lower height growth and lower sensitivity to competition resulted in a more similar growth pattern for all trees. The *P. sylvestris* stand showed intermediate behaviour; although the biological rate of height development of *P. sylvestris* is lower than that of *P. pinaster*, the former is more affected by competition, and the age of the stand is almost three times that of the *P. pinaster* stand. The deterministic estimate provided a height distribution with low variability located around the observed mean value, whereas the random stochastic approach provided greater variability, consistent with the observed distribution. This is evident in the values of the Kolmogorov-Smirnov test for each diameter class (Table 3) and in the interquartile and maximum ranges of the distributions (Fig. 2).

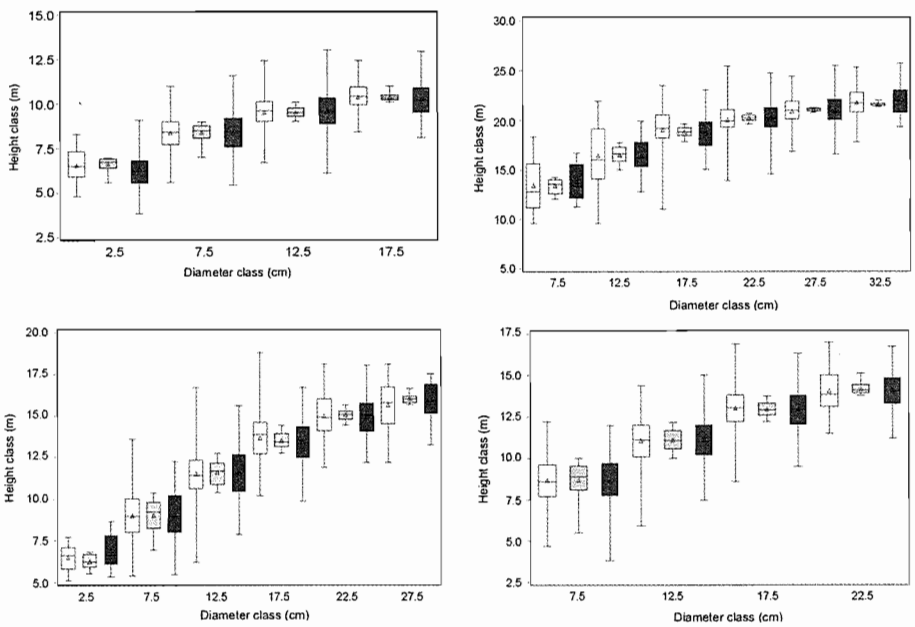

FIG. 2–Box plots of height distributions (Y-axis) against diameter classes (X-axis) for the four data sets. The triangles represent the means of height estimates. The boxes represent the interquartile ranges. The maximum and minimum height estimates are represented by the upper and lower small horizontal lines crossing the vertical bars, respectively. In white: observed distribution; in light grey: deterministic estimation; in dark grey: stochastic estimation.

TABLE 3–Results of the Kolmogorov-Smirnov test for diameter classes, comparing the predicted height distributions of the deterministic and stochastic approaches with the observed height distribution. Significant values obtained in this test mean that the predicted distribution does not differ significantly from the observed distribution.

Class centre (cm)	*Pinus sylvestris*			*Pinus pinaster*			*Pinus radiata* New Zealand			*Pinus radiata* local		
	n	Determ.	Stoch.	*n*	Determ.	Stoch.	*n*	Determ.	Stoch.	*n*	Determ.	Stoch.
2.5		—	—	15	0.334	0.134*		—	—	10	0.300**	0.200*
7.5	15	0.334	0.267**	390	0.297	0.049**	154	0.169	0.045*	138	0.079**	0.058*
12.5	64	0.344	0.219	509	0.261	0.080	253	0.205	0.039*	218	0.128	0.046*
17.5	111	0.324	0.189	148	0.263	0.169	158	0.234	0.051*	170	0.247	0.071**
22.5	121	0.256	0.074*		—	—	86	0.232	0.047*	110	0.254	0.073*
27.5	96	0.229	0.062*		—	—		—	—	22	0.273	0.091*
32.5	81	0.432	0.062*		—	—		—	—		—	—

n = number of trees
* Indicates significance at $p = 0.05$
** Indicates significance at $p = 0.01$

The above results suggest that the stochastic approach allows mimicking of the natural variability in tree height within diameter classes, providing more realistic height predictions at stand level. This is necessary when height-diameter models are used to fill in the missing height measurements, because deterministic models smooth the tree height estimations and eliminate the observed variation.

This is also important in stand-level growth models (e.g., Knoebel *et al.* 1986; Diéguez-Aranda *et al.* 2006), which use height-diameter functions to estimate the height of the average tree in each diameter class. By using the stochastic approach it is possible to obtain more realistic predictions of variables that depend on the estimated heights (e.g., volume, biomass, carbon pools).

CONCLUSIONS

The relationship provided by the Schnute function is biologically reasonable. Unrealistic height predictions should not occur beyond the range of the empirical observations.

The relationship between tree height and diameter is one of most important elements in natural and managed forest dynamics and structure. However, a deterministic model is not capable of completely describing the nature of the height-diameter relationship. Thus, if real variability is required, a stochastic component should be added to the deterministic estimation.

The suggested approach improves the realism and accuracy of height predictions at stand level. This feature is considered very important because stand growth level models used for these species in Galicia are disaggregated by diameter classes to estimate merchantable volume and biomass using taper functions and individual tree biomass equations (Merino *et al.* 2005).

Many models in forestry are deterministic and predict the most likely outcome. While this may not matter much in models corresponding to homogeneous industrial plantations, it may be a serious deficiency in sustainable forest management in which decisions are based on models that attempt to address natural variability or ecological succession (Vanclay 2003).

The stochastic approach proposed can also be used in most situations where different predictions are required for the same values of the independent variables. Otherwise, the mean and distribution of the final estimates would be biased. In each situation, the process should involve the study of the specific distribution function to be used for generating variability in the simulated data because models with a strong basis in scientific knowledge of the processes represented should behave more realistically than models with a structure determined by statistical analysis that has not been guided by the same scientific knowledge (Stage 2003).

ACKNOWLEDGMENTS

The fieldwork involved in the collection of the data was financed by research project PGIDIT02RFO29103PR, entitled "Estudio de las cortas de mejora (claras) en las masas de coníferas de Galicia". We thank Dr Christine Francis for correcting the English grammar of the text.

REFERENCES

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.

BOTKIN, D.B.; JAMAK, J.F.; WALLIS, J.R. 1972: Some ecological consequences of a computer model of forest growth. *Journal of Ecology 60*: 849–873.

BREDENKAMP, B.V.; GREGOIRE, T.G. 1988: A forestry application of Schnute's generalized growth function. *Forest Science 34*: 790–797.

CASTEDO, F.; BARRIO, M.; PARRESOL, B.R.; ÁLVAREZ-GONZÁLEZ, J.G. 2005: A stochastic height-diameter model for maritime pine ecoregions in Galicia (northwestern Spain). *Annals of Forest Science 62*: 455–465.

CURTIS, R.O. 1967: Height-diameter and height-diameter-age equations for second-growth Douglas-fir. *Forest Science 13*: 365–375.

DANIELS, R.F.; BURKHART, H.E. 1975: Simulation of individual tree growth and stand development in managed loblolly pine plantations. *Virginia Polytechnic Institute and State University, Blacksburg, VA, FWS 5-75*.

DENNIS, B.; BROWN, B.E.; STAGE, A.R.; BURKHART, H.E.; CLARK, S. 1985: Problems of modeling growth and yield of renewable resources. *The American Statistician 39*: 374–383.

DIÉGUEZ-ARANDA, U.; CASTEDO DORADO, F.; ÁLVAREZ GONZÁLEZ, J.G.; ROJO ALBORECA, A. 2006: Dynamic growth model for Scots pine (*Pinus sylvestris* L.) plantations in Galicia (north-western Spain). *Ecological Modelling* 191: 225–242.

DRAPER, N.R.; SMITH, H. 1981: "Applied Regression Analysis". John Wiley & Sons, New York.