Tree height estimation using a stochastic height-diameter relationship

M. Barrio¹, U. Diéguez-Aranda¹, F. Castedo², J.G. Álvarez-González¹ and A. Rojo¹

Unidade de Xestión Forestal Sostible
Web: http://www.lugo.usc.es/uxfs/

¹ Departamento de Ingeniería Agroforestal, Universidad de Santiago de Compostela. Escuela Politécnica Superior, Campus universitario, 27002 Lugo, Spain.
² Departamento de Ingeniería Agraria, Universidad de León. Escuela Superior y Técnica de Ingeniería Agraria, Campus de Ponferrada, 24400 Ponferrada, Spain.

Abstract
Measuring the height of a tree takes longer than measuring its diameter at breast height and often only the heights of a subset of trees of known diameter are measured in forest inventories. Since trees with the same diameter are not usually of the same height, even within the same stand, a stochastic component was added to the deterministic height function. This approach mimics the natural variability of heights and therefore provides more realistic height predictions, as demonstrated by the results of the Kolmogorov Smirnov test and by visual analysis of box plot graphs.

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). The time taken to measure tree heights takes longer than measuring the diameter at breast height. For this reason, often only the heights of a subset of trees of known diameter are measured, and accurate height-diameter equations must be used to predict the heights of the remaining trees to reduce the costs involved in data acquisition. If stand conditions vary greatly within a forest, a height regression may be derived separately for each stand, or a generalized function, which includes stand variables to account for the variability, may be developed (Curtis, 1967; Zhang et al., 1997; Sharma & Zhang, 2004).

Two trees within the same stand and that have the same diameter are not necessarily of the same height; therefore a deterministic model does not seem appropriate for mimicking the real natural variability in height (Parresol & Lloyd, 2004).

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.

Material and methods
We used data from two thinning trials installed in Pinus pinaster and Pinus sylvestris even-aged stands located in northwestern Spain. The stand conditions within each thinning trial were similar and thus we considered the data obtained as belonging only to two different stands.

Table 1. Characteristics of the tree samples used for model fitting.

<table>
<thead>
<tr>
<th>Thinning trial</th>
<th>Number of trees</th>
<th>d (cm)</th>
<th>h (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Min.</td>
</tr>
<tr>
<td>P. pinaster</td>
<td>1062</td>
<td>11.2</td>
<td>2.9</td>
</tr>
<tr>
<td>P. sylvestris</td>
<td>488</td>
<td>22.4</td>
<td>7.0</td>
</tr>
</tbody>
</table>
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 of the 12 plots belonging to each thinning trial. Summary statistics, including the mean, minimum, maximum and standard deviation (SD) of the above-mentioned variables are shown in Table 1.

A large number of local height-diameter equations have been reported in the forestry literature (e.g. Curtis, 1967; García, 1974; Wykoff et al., 1982; Huang et al., 1992, 2000; Fang & Bailey, 1998; Peng, 1999; Soares & Tomé, 2002), most of which exhibit S-shaped or concave-shaped patterns. From among all the models we selected the Schnute growth model (Schnute, 1981) for the purposes of our study. This function 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). The expression of Schnute’s function is as follows:

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

where: \(h\) is total height of the tree; \(d\) is tree diameter at breast height; \(d_2\) is diameter at breast height of a large tree (upper range of data); \(h_2\) is parameter representing mean tree height at \(d_2\); \(a\) and \(b\) are the model parameters to be estimated and \(\varepsilon\) the residual error.

Since the scatter plots of total tree height against diameter did not show an unequal error variance for any of the two data sets, parameters of equation (1) were estimated using non-linear least squares, using the NLIN procedure of SAS/STAT® (SAS Institute Inc., 2004).

The goodness-of-fit of the model was evaluated by means of the coefficient of determination for non-linear regression \((R^2)\) (see Ryan, 1997) and the root mean squared error \((RMSE)\).

The stochastic approach uses this standard error in a similar way as the prediction interval is obtained for a new individual in a regression model. Instead of using the \(t\) value corresponding to a fixed limit for all the trees, 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:

\[
\hat{y}_{i(\text{stoch})} = \hat{y}_i + F_U^{-1} s_{\hat{y}_{i(\text{new})}}
\]

where \(\hat{y}_{i(\text{stoch})}\) is the stochastic height estimation, \(\hat{y}_i\) is the deterministic height obtained from equation (1) (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 of a new observation and is expressed as:

\[
s_{\hat{y}_{i(\text{new})}} = \sqrt{\text{var}(\hat{y}_{i(\text{new})})} = \sqrt{\text{RMSE}^2 + z(b)^{\top} S^2(b) z(b)},
\]

where the scalar \(\text{RMSE}^2\) is the regression mean squared error, \(z(b)\) is the partial derivatives of the least squared estimator \(b\) of the parameter vector \(\beta\).

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**

The Schnute function provided a reasonable fit, taking into account the high variability in the height-diameter relationship in the databases (see Table 1). All the parameters were significant at the $P = 95\%$ level and showed reasonable values. The low variability explained by the local model indicates that stand variables must be taken into account to improve this relationship.

Deterministic values of the height $\hat{y}_i$ were obtained using equation (1) with the parameter estimates obtained for each data set (Table 2). The prediction of the stochastic values $\tilde{y}_{i(stoch)}$ involved the following steps: a) calculation of the standard error of the prediction $s_{\hat{y}_{i(new)}}$ by means of equation (3); b) calculation of the inverse of the standard normal distribution function for one pseudo-random number in the interval (0,1) and c) estimation of the $\tilde{y}_{i(stoch)}$ from equation (2) 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 (Figure 1, Table 3).

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

<table>
<thead>
<tr>
<th>Thinning trial</th>
<th>Parameter estimates</th>
<th>$R^2$</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a$</td>
<td>$b$</td>
<td>$h_2$</td>
</tr>
<tr>
<td>$P$. pinaster</td>
<td>0.0896 (0.0246)</td>
<td>1.9051 (0.3197)</td>
<td>10.9793 (0.1536)</td>
</tr>
<tr>
<td>$P$. sylvestris</td>
<td>0.1069 (0.0176)</td>
<td>0.9516 (0.2575)</td>
<td>21.9859 (0.2805)</td>
</tr>
</tbody>
</table>

The observed height distribution generally followed a normal distribution within diameter classes with a high degree of variability in the two data sets evaluated. 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 from the results of the Kolmogorov-Smirnov test for each diameter class (Table 3) and the interquartile and maximum ranges of the distributions (Figure 1).

**Figure 1.** Box plots of height distributions (Y-axis) against diameter classes (X-axis) for *Pinus sylvestris* (left) and *Pinus pinaster* (right) 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 in the two data sets.

<table>
<thead>
<tr>
<th>Diameter class (cm)</th>
<th>Pinus sylvestris</th>
<th>Pinus pinaster</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>5.5</td>
<td>15</td>
<td>0.334</td>
</tr>
<tr>
<td>7.5</td>
<td>64</td>
<td>0.344</td>
</tr>
<tr>
<td>9.5</td>
<td>111</td>
<td>0.324</td>
</tr>
<tr>
<td>12.5</td>
<td>121</td>
<td>0.256</td>
</tr>
<tr>
<td>14.5</td>
<td>96</td>
<td>0.229</td>
</tr>
<tr>
<td>17.5</td>
<td>81</td>
<td>0.432</td>
</tr>
</tbody>
</table>

** Indicates significance at $P = 0.05$. * Indicates significance at $P = 0.01$.

Conclusions
The relationship between tree height and diameter is one of the 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 northwestern Spain are disaggregated by diameter classes to estimate merchantable volume and biomass using taper functions and individual tree biomass equations.

The stochastic approach proposed may also be used in most cases where different predictions are required for the same values of the independent variables. In each situation, the process should involve the study of the specific distribution function to be used for generating variability in the simulated data (Parresol & Lloyd, 2004).

Acknowledgments
The fieldwork involved in the collection of the data was financed by research project PGIDIT02RF029103PR, entitled “Estudio de las cortas de mejora en las masas de coníferas de Galicia”. We thank Dr. Christine Francis for correcting the English grammar of the text.

Reference


