Corral-Rivas *et al. Forest Ecosystems* 2014, **1**:6
http://www.forestecosyst.com/1/1/6

Forest Ecosystems
a SpringerOpen Journal

RESEARCH **Open Access**

Local and generalized height-diameter models with random parameters for mixed, uneven-aged forests in Northwestern Durango, Mexico

Sacramento Corral-Rivas[1*], Juan Gabriel Álvarez-González[2], Felipe Crecente-Campo[2] and José Javier Corral-Rivas[3]

Abstract

Background: We used mixed models with random components to develop height-diameter (h-d) functions for mixed, uneven-aged stands in northwestern Durango (Mexico), considering the breast height diameter (d) and stand variables as predictors.

Methods: The data were obtained from 44 permanent plots used to monitor stand growth under forest management in the study area.

Results: The generalized Bertalanffy-Richards model performed better than the other generalized models in predicting the total height of the species under study. For the genera Pinus and Quercus, the models were successfully calibrated by measuring the height of a subsample of three randomly selected trees close to the mean d, whereas for species of the genera Cupressus, Arbutus and Alnus, three trees were also selected, but they are specifically the maximum, minimum and mean d trees.

Conclusions: The presented equations represent a new tool for the evaluation and management of natural forest in the region.

Keywords: Conifer and broadleaves forests; *h-d* relationship; Mixed models; Calibration

Background

Most forests in Durango State (Mexico) are comprised of a mixture of species of the genera *Pinus* and *Quercus* with an irregular distribution of trees of all size classes. However, species of the genera *Arbutus* and *Juniperus* are also found in most of these forests (Wehenkel *et al*. 2011). These forests, which cover an area of 5.4 million ha, are considered as the primary forest reserve at a national level, and they provide almost a quarter of the national forest production in Mexico (SRNyMA 2006). The forests also play an important role in providing environmental services, such as protection against soil erosion, biodiversity conservation, carbon capture and protection of water reserves; they also provide recreational areas and represent an important source of income for their owners and local inhabitants.

Forest management requires prediction tools that provide detailed information about the development of mixed, uneven-aged stands. Growth and production models are the most commonly used tools for this purpose. When the breast height diameter (d) and total height (h) are known, application of these models is relatively easy (Sharma and Parton 2007). Measuring diameter is simple, accurate and inexpensive, whereas measuring height is relatively more complex, time-consuming and expensive. Therefore, height-diameter (h-d) functions are often utilized, so that the height of an individual tree can be predicted only from the diameter. These relationships are also very useful for estimating individual volume, site index and for describing growth and production in forest stands over time when the height is not measured (Curtis 1967).

Most h-d functions have been developed for forest plantations (e.g. Soares and Tomé 2002; López Sanchez et al. 2003). However, the relationship between the diameter and height of a tree varies between stands (Calama and

* Correspondence: juangabriel.alvarez@usc.es
[1]Instituto Tecnológico de El Salto, Mesa del Tecnológico s/n. Apdo. Postal No. 2, El Salto P.N. 34950, Durango, México
Full list of author information is available at the end of the article

Montero 2004) because it depends on stand characteristics such as density and site index (Sharma and Zhang 2004). Moreover, the *h-d* relationship also varies over time within the same stand (Curtis 1967). Such considerations indicate that stand variables should be used to construct generalized functions that represent all possible conditions in forest stands (Temesgen and Gadow 2004). This is particularly important in mixed, uneven-aged stands in which different species, ages, structures and levels of competition coexist (Vargas-Larreta *et al.* 2009).

The hierarchical structure of *h-d* data (i.e. trees grouped in plots and plots grouped in stands) results in a lack of independence between measurements because the observations in each sampling unit will be correlated (Gregoire 1987). Mixed models have been successfully used to address this type of problem (e.g. Lappi 1997; Calama and Montero 2004; Castedo Dorado et al. 2006). This approach simultaneously estimates *fixed* parameters (parameters that are common to the entire population) and *random* parameters (parameters that are specific to each plot) within the same model and enables the variability between plots of the same population to be modelled.

The objectives of this study were as follows: i) to compare different local *h-d* equations for the mixed, uneven-aged forests in north-western Durango; ii) to develop new generalized *h-d* equations for different groups of species based on the best local model previously fitted iii) to use the local and the generalized equations to study the capacity of mixed models to explain the variability in the *h-d* relationship; and iv) to determine the most suitable size and type of sample for calibrating the functions fitted with mixed models.

Methods

Study area

The study was carried out in the *Ejido San Diego de Tezains*, Municipality of Santiago Papasquiaro, Durango State, Mexico (between 105° 53′ 36″ and 106° 12′ 40″ W and 24° 48′ 16″ and 25° 13′ 32″ N). The predominant vegetation in the area is mixed, uneven-aged forests of *Pinus* and *Quercus*. The altitude above sea level of the study area varies between 1,400 and 3,000 m. The prevailing climate is temperate: the annual precipitation ranges between 800 and 1,100 mm and the mean annual temperature varies between 8°C in the highest elevations and 24°C in the lowest elevations (García 1981).

Data

The data were obtained from 44 permanent plots used to monitor the growth and production of the forests in the *Ejido San Diego de Tezains.* These plots, which were established in 2008, were selected with the aim of representing all types of vegetation, site qualities and diameter distributions in managed stands. The plots, of size 50x50 m, are distributed under a systematic grid sampling approach that varies between 3 and 5 km, and will be remeasured at 5 year intervals. We recorded the following main variables: number of trees, species code, breast height diameter at 1.3 m (d, cm), total tree height (h, m), azimuth (°) and radius (m) from the centre of the plot (point where the diagonals cross) towards all trees of breast height diameter ≥5 cm.

The database included 25 species, which were classified on the basis of their growth patterns into the following 13 groups for posterior analysis: 1 (*Pinus arizonica*), 2 (*P. ayacahuite*), 3 (*P. durangensis*), 4 (*P. herrerae*), 5 (*P. lumholtzii*), 6 (*P. teocote*), 7 (*P. douglasina*), 8 (*Quercus sideroxyla*), 9 (**other species of *Quercus***: *Q. arizonica*, *Q. mcvaughii*, *Q. durifolia*, *Q. crassifolia*, *Q. jonesii*, *Q. rugosa* and *Q. laeta*), 10 ***Pinus* species** (all species of the genus *Pinus* [codes 1 to 7]), 11 ***Quercus* species** (all species of the genus *Quercus* [codes 8 and 9]), 12 **other conifer species** (*Juniperus deppeanna, J. durangensis and Cupresus lusitanica*) and 13 **other broadleaf species** (*Arbutus arizonica, A. bicolor, A. madrensis, A. tesselata, A. xalapensis* and *Alnus firmifolia*).

We examined the distribution of the pairs of *h-d* data for each species or group graphically to identify any possible anomalies. As extreme data points were observed, a systematic approach, similar to the one proposed by Bi (2000) for detecting abnormal data points, was applied to increase the efficiency of the process. A local quadratic equation with a smoothing parameter of 0.25 (selected after iterative fitting and visual examination of the smoothed curves for different smoothing parameters overlaid on the data), was fitted for each of the species or group. In this approach, the number of extreme values accounted for about 1% for all species together, which were excluded from the database used for fitting the equations. The main descriptive statistics for the breast height diameter and the total height of the main groups that included the species under study are shown in Table 1.

The following stand variables were calculated from the trees registered in each plot: number of trees per hectare (N, trees ha^{-1}), stand basal area (G, m^2 ha^{-1}), mean square diameter (d_g, cm), dominant height (estimated as the mean height of the 100 largest diameter trees per hectare, independently of the species [H_0, m]), dominant diameter (estimated as the mean diameter of the 100 largest diameter trees per hectare, independently of the species [D_0, cm]) and Hart's index (%) estimated as follows: $HI = {}^{10000}\big/_{\sqrt{N}} * H_0$.

Comparison of equations

We selected a total of 27 *local* equations (Huang *et al.* 2000) for data fitting. We also studied the relationship between the stand variables and the parameters of the

Table 1 Summary statistics of the database used in fitting the h-d equations

Group	Plots	Number of observations	*d* (cm)				*h* (m)			
			Mean	Max.	Min.	SD	Mean	Max.	Min.	SD
Pinus species	44	4033	18.0	99.5	6.9	11.0	12.1	38.1	2.3	5.7
Quercus species	44	1801	18.4	90.0	6.1	10.8	8.7	29.8	1.8	4.2
Other conifer spp.	31	188	17.6	75.0	7.5	12.2	7.9	16.5	2.5	2.9
Other broadleaf spp.	38	302	16.7	56.5	7.3	9.5	6.6	14.0	2.2	2.3

SD: standard deviation.

local equations that best described the *h-d* relationship, with the aim of improving the accuracy of the equation and developing new generalized functions.

For preliminary selection, we used ordinary non-linear least squares (ONLS) to fit each of the local equations to the data from the 13 established groups, with the MODEL procedure in SAS/ETS® statistical software package (SAS Institute Inc 2008). We evaluated the goodness of fit of the models by graphical analysis and by considering the following statistics, calculated from the residuals: root mean square error (*RMSE*), the coefficient of determination (R^2), bias, and Bayesian information criterion (*BIC*; Schwarz 1978). We used the following formulae to calculate these statistics:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{i=n}\left(h_i - \hat{h}_i\right)^2}{n-k}} \quad (1)$$

$$R^2 = \left[1 - \frac{\sum_{i=1}^{i=n}\left(h_i - \hat{h}_i\right)^2}{\sum_{i=1}^{i=n}\left(h_i - \bar{h}\right)^2}\right] \quad (2)$$

$$Bias = \frac{\sum_{i=1}^{i=n}\left(h_i - \hat{h}_i\right)}{n} \quad (3)$$

$$\mathrm{BIC} = n\ln\left(\sum_{i=1}^{i=n}\left(h_i - \hat{h}_i\right)^2/(n-k)\right) + k\cdot\ln(n) \quad (4)$$

where h_i, $\hat{h}_i$ and $\bar{h}$ are the observed and estimated heights and the mean of the observed heights, respectively; n is the number of observations used in the fitting; k is the number of parameters in the equation, and ln is the natural logarithm.

As each local equation has different strengths and weaknesses, which may lead to different goodness-of-fit results for each group of species, we used a Qualification Index (QI_t) to evaluate the goodness of fit by considering the values of R^2 (with high values representing good fits), Bias (with low absolute values representing good fits) and *RMSE* and *BIC* (with low values representing good fits). For this index, a value of 1 is assigned to the equation that was best for each group of species and a value of 0 to the others. The qualifications obtained for each equation and statistics were then summed as follows: $QI_{total} = \sum_i \sum_j QI_{ij}$; where QI_{ij} is the qualification for the j-th goodness of fit criterion in the i-th group of species.

For the local equation for which the QI_{total} was highest for the defined groups, we used graphical analysis and the CORR procedure in SAS (SAS Institute Inc 2008) to analyse the relation between each of the parameters and the main stand variables, with the aim of testing different forms of generalized equations.

Effect of mixed models

The *h-d* observations made in plots and stands may be highly correlated, thus violating the principle of independence of error terms (Calama and Montero 2004). One procedure used to deal with correlated observations is to fit mixed models, in which the variability between the sampling units can be explained by including random parameters, which are estimated at the same time as the fixed parameters (Lappi 1997; Calama and Montero 2004).

Basically, the parameter vector of a non-linear mixed model can be defined as follows (Pinheiro and Bates 1998):

$$\mathbf{\Phi}_j = \mathbf{A}_j\boldsymbol{\lambda} + \mathbf{B}_j\mathbf{b}_j \quad (5)$$

where Φ_j is the parameter vector $r \times 1$ (where r is the total number of parameters in the model) specified for the j-th plot, $\boldsymbol{\lambda}$ is the vector $p \times 1$ of the common fixed parameters for the whole population (p is the number of fixed parameters in the model), $\mathbf{b}_j$ is the vector $q \times 1$ of the random parameters associated with the j-th plot (q is the number of random parameters in the model), $\mathbf{A}_j$ and $\mathbf{B}_j$ are matrices of size $r \times p$ and $r \times q$ for specific and random effects for the j-th plot, respectively.

The basic theory of non-linear mixed models says that the residual vector ($\hat{\mathbf{e}}_{ij}$) and the random effects vector ($\mathbf{b}_j$) are often assumed to be uncorrelated and normally distributed with mean zero and variance-covariance matrices $\mathbf{R}_j$ and $\mathbf{D}$, respectively. The residual vector represents within subject (e.g., plot) variability and the random effects vector represents between subject variability (Littell *et al.* 1996).

We constructed the non-linear mixed effects model by selecting the local and generalized equations that yielded the best fits for the species groups defined using the NLMIXED procedure in SAS/ETS (SAS Institute Inc 2008). We tested different combinations of fixed and random parameters and compared the fitting statistics (*RMSE*, R^2, Bias and *BIC*), to determine which parameter (s) should be considered mixed.

Calibration

The inclusion of random parameters in *h-d* equations leads to two possible situations as regards prediction of the height of trees within a stand (Vonesh and Chinchilli 1997): i) a population mean response (PMR) when only diameters are measured (and the stand variables are included in the model in the case of generalized models) and the vector of random parameters is assumed to have an expected value of $E(b_j) = 0$; and ii) a calibrated response, when the height of a subsample of m_j trees is measured along with diameter measurement in each new plot *j* (and the stand variables in the case of generalized models) and is subsequently used to calculate the specific random parameters of the new sampling units (*Calibrated Response*; *CR*), i.e. vector $\mathbf{b}_j$, expressed as follows (Vonesh and Chinchilli 1997):

$$\hat{\mathbf{b}}_j \approx \hat{\mathbf{D}}\hat{\mathbf{Z}}_j^{\mathrm{T}}\left(\hat{\mathbf{R}}_j + \hat{\mathbf{Z}}_j\hat{\mathbf{D}}\hat{\mathbf{Z}}_j^{\mathrm{T}}\right)^{-1}\hat{\mathbf{e}}_{ij} \quad (6)$$

where $\hat{\mathbf{D}}$ is the matrix $q \times q$ of variances-covariances associated with the random parameters (q = number of random parameters included in the model), which is common to all plots and is estimated in the general model fitting procedure; $\hat{\mathbf{R}}_j$ is the $m_j \times m_j$ estimated matrix of variances-covariances of the error term; $\hat{\mathbf{e}}_{ij}$ is the residuals vector $m \times 1$, the components of which are obtained as the difference between the observed height of each tree and the value predicted using the model with fixed parameters only; and $\hat{\mathbf{Z}}_j$ is the matrix $m \times q$ of the partial derivatives of the random parameters evaluated in $\hat{\mathbf{b}}_j = 0$.

Two sampling options were considered for selecting the subsample of trees to measure within each for calibration of the local and generalized equations:

(i) CR1: Measuring the total height of between 1 and 5 randomly selected trees within each plot that are close (± 10%) to the mean breast height diameter.
(ii) CR2: Measuring the total height of the tree of mean breast height diameter, or measuring the height of two trees – the mean and minimum breast height diameters, or measuring the height of three trees – the mean, minimum and maximum breast height diameters within each plot.

We evaluated these two alternatives in terms of the previously defined goodness-of-fit statistics (*RMSE*, R^2 and Bias), which we compared with the statistics obtained for the equations fitted by the ONLS and NLMIXED procedures.

Results and discussion

Local equations

In the comparison of the goodness-of-fit statistics for the local *h-d* equations fitted to the data for the 13 previously defined groups of species, the Bertalanffy-Richards equation (Bertalanffy 1949; Richards 1959) consistently yielded the highest R^2 and lowest *RMSE* values; however, the equations proposed by Stage (1975) and Meyer (1940) yielded the lowest values of BIC, which gives preference to models containing few parameters over those containing several parameters (Table 2).

The Bertalanffy-Richards equation yielded the highest R^2 values for 6 of the 13 species groupings and the lowest *RMSE* values for 4 of the groups. Finally, comparison of the BIC values indicated that this was the preferred equation only for the *Quercus spp*. grouping.

In selecting the best local equation, we also examined graphs of the residuals, the significance of the parameters and the mean bias for each equation. In this respect, the Bertalanffy-Richards equation was the preferred model. The final structure of the local Bertalanffy-Richards equation used was as follows:

$$h = 1.3 + b_0(1-\exp(-b_1 \cdot d))^{b_2} \quad (7)$$

where b_0-b_2 are equation parameters and the rest of variables as defined in the data section.

The value of the goodness-of-fit statistics for Eq. (7) and the species groups are shown in Figure 1.

Considering that the fitting statistics for the different broad groupings (*Pinus* species, *Quercus* species, other conifer species and other broadleaf species) are similar to those obtained for each individual species and that some parameters were not significant in the individual fits for some of the species, we decided to

Table 2 Qualification index for the 8 best local equations for the 13 groups of species

Equation	Qualification index (QI_{ij})				QI_{total}	%
	R^2	*RMSE*	*Bias*	*BIC*		
Bates and Watts (1980)	0	1	1	1	3	5.8
Meyer (1940)	0	1	0	3	4	7.7
Stage (1975)	1	3	0	4	8	15.4
Logarithmic	0	0	0	1	1	1.9
Wykoff *et al.* (1982)	1	1	1	2	5	9.6
Bertalanffy-Richards	6	4	4	1	15	28.8
Hossfeld (1822)	3	1	7	0	11	21.2
Weibull (1951)	2	2	0	1	5	9.6

QI_{total}= $\Sigma\Sigma Q_{ij}$; where QI_{ij} is the qualification for the *j*-th goodness of fit criterion in the *i*-th group of species.

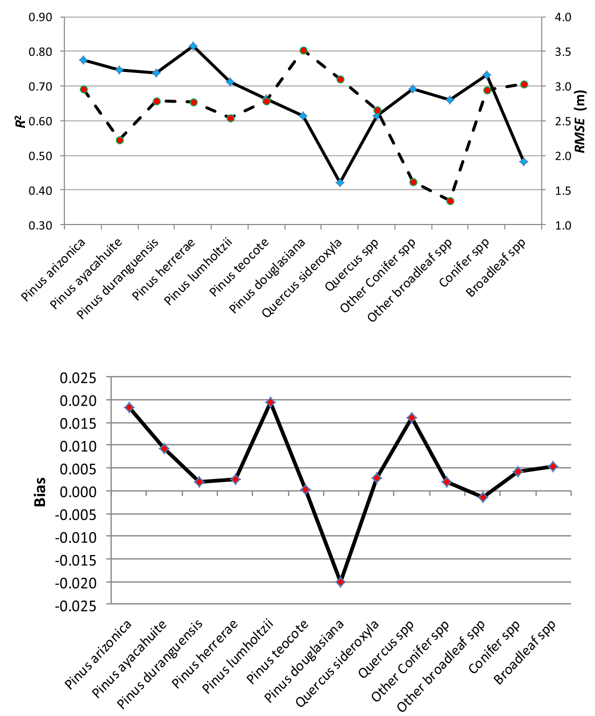

Figure 1 Values of R^2 (solid line), RMSE (dashed line), and mean bias (bottom) obtained by fitting Eq. (7) to the *h-d* relationship for the 13 groups of species considered.

use 4 different local equations, one for each broad grouping studied.

The parameters in Eq. (7) and the goodness-of fit statistics, obtained by the ONLS method for the 4 broad groupings, are shown in Table 3.

Generalized equations

On relating the parameters in Eq. (7) to the stand variables, we found that in the 4 groups that included all species, parameter b_0, representative of the asymptote of Eq. (7), was positively correlated (almost 52%) with H_0 *and* D_0, whereas parameter b_1, representative of "scale" was only positively correlated by more than 50% with Hart's index (*HI*) in the other conifer species and other broadleaf species and, finally, parameter b_2, representative of "shape" was also positively correlated (~ 51%) with d_g and N for the four groups that include all species.

To develop a generalized equation from Eq. (7), we tested various combinations of the stand variables to improve their efficacy in the fit, taking into account the previously mentioned correlations; this resulted in Eqs (8) and (9). Equation 8 yielded the best fitting statistics for the *Pinus* and *Quercus* groups, whereas Eq. (9) yielded the best fitting statistics for the species included in other conifer species and other broadleaf species.

$$h = 1.3 + b_0 \cdot {H_0}^{b_1}\left(1-\exp(-b_2\cdot d)\right)^{b_3\left({}^{N}/_{d_g}\right)^{b_4}} \tag{8}$$

$$h = 1.3 + {H_0}^{b_0}\left(1-\exp(-b_1\cdot HI\cdot d)\right)^{\left({}^{N}/_{d_g}\right)^{b_2}} \tag{9}$$

where b_0-b_2 are equation parameters and the rest of variables as defined in the data section.

On comparing the goodness-of-fit statistics for Eqs (8) and (9) with those obtained when fitting the 30 generalized equations used in other studies (e.g., López Sanchez et al. 2003; Sharma and Zhang 2004; Sharma and Parton 2007), we found that the value of *RMSE* for Eq. (8) was slightly lower than those obtained by fitting the above-mentioned generalized equations, whereas the value of the *RMSE* for Eq. (9) was only lower than some of these. However, the latter equations included parameters that were not significant at the 0.05 level. Another advantage of Eq. (9) is that the value of the *BIC* was lower than that obtained for the 30 generalized equations compared. Analysis of the residuals also revealed that there were no anomalies associated with Eqs (8) and (9) that would indicate non-compliance with the underlying hypotheses of normality, homogeneity of variance or independence of errors. We therefore

Table 3 Estimated parameters and fitting statistics obtained for the local model with and without mixed effects for the groups of species considered

Species	Equation	Fitting method	Parameters						Statistics			
			b_0	b_1	b_2	σ_u^2	σ_v^2	σ_{uv}	R^2	*RMSE*	*Bias*	*BIC*
Pinus species	(7)	ONLS	34.9600	0.0223	1.0148				0.73	2.95	0.004	8749.7
	(10)	NLMIXED	28.6890	0.0357	1.1148	47.9289	0.0002	−0.0849	0.85	2.21	0.005	6461.0
Quercus species	(7)	ONLS	21.1250	0.0240	0.9666				0.48	3.03	0.005	4021.3
	(11)	NLMIXED	16.2456	0.0317	0.9971	20.8757			0.71	2.25	0.014	2955.0
Other conifer species	(7)	ONLS	24.4720	0.0094*	0.6691				0.69	1.62	0.002	197.1
	(11)	NLMIXED	134.2700	0.0004	0.6062	115.6000			0.75	1.46	−0.013	170.8
Other broadleaf species	(7)	ONLS	24.8375*	0.0085*	0.7489				0.66	1.35	−0.002	197.8
	(11)	NLMIXED	94.3893	0.0005	0.6086	130.9100			0.79	1.07	−0.015	70.0

ONLS = fitted by ordinary least squares, NLMIXED = fitted by non-linear mixed effects; *estimated parameters not significant at the 0.05 level.

decided to use Eqs (8) and (9) as generalized models for the four groups that included all species considered. The values of the parameters of Eqs (8) and (9) obtained for fitting each group of species are shown in Table 4. The signs and values of all parameters are consistent with their biological interpretation and visual examination of the graphs of the *h-d* relationship indicates that its performance was consistent with the theory of growth.

The generalized *h–d* equations selected in this study included dominant stand height. This represents an advantage over equations that include the mean height because less effort is required in conventional inventories to estimate the dominant height than the mean height of the stand (López Sanchez et al. 2003). These functions also include the density of the stand in terms of number of trees per unit of area and mean square diameter. Stand density is the most obvious factor affecting the *h-d* relationship in a stand (Zeide and VanderSchaaf 2002); in other words, trees of the same diameter are generally taller in denser stands.

Various stand variables have been proposed as predictors of the *h-d* relationship: stand age (Curtis 1967; Soares and Tomé 2002; López Sanchez et al. 2003); crown competition index (Temesgen *et al.* 2007); geographic variables (Schmidt *et al.* 2011); and wind speed (Meng *et al.* 2008). Although the inclusion of other variables may improve the predictive capacity of the selected functions, this requires great sampling effort and limits the practical application of the functions and therefore we did not take such variables into account.

Effect of mixed models

Parameters b_0, b_1 and b_2 in Eqs (7) and (9) determine the asymptote, the scale and the shape of the *h-d* curves, respectively, whereas in Eq. (8), the parameters b_0 and b_1 define the asymptote, b_2 is the scale parameter and b_3 and b_4 define the shape of the curve. In fitting Eq. (7) to the data from each plot, we found that the parameter that affected the asymptote was the most variable, followed by the scale parameter. Therefore, in a first step, we fitted Eqs (7), (8) and (9) to the *h-d* data by considering the parameters that define the asymptote and scale as mixed, in other words, with a random parameter added. Pinheiro and Bates (1998) obtained similar results and found that the best results were obtained when the asymptote and scale of Eq. (7) were considered as random parameters. In most cases, the mixed model did not converge, so that we tested the inclusion of only mixed parameters associated with those parameters that define the asymptote of the *h-d* curve until reaching convergence. Similar results have been reported by Sharma and Parton (2007) and Vargas-Larreta *et al.* (2009). Some of the parameters were scaled so that all were of the same order of magnitude and to prevent instability in the fitting function (Calama and Montero 2004). The expressions of the mixed models finally obtained are Eqs (10) to (15):

$$\hat{h}_{ij} = 1.3 + (b_0 + u_j)\left(1 - \exp\left(-(b_1 + v_j)d_{ij}\right)\right)^{b_2} + e_{ij} \quad (10)$$

$$\hat{h}_{ij} = 1.3 + (b_0 + u_j)\left(1 - \exp\left(-b_1 \cdot d_{ij}\right)\right)^{b_2} + e_{ij} \quad (11)$$

$$\hat{h}_{ij} = 1.3 + (b_0 + u_j) \cdot Ho^{b_1}\left(1 - \exp\left(-b_2 \cdot d_{ij}\right)\right)^{b_3\left(N/_{dg}\right)^{b_4}} + e_{ij} \quad (12)$$

$$\hat{h}_{ij} = 1.3 + b_0 Ho^{(b_1 + u_j)}\left(1 - \exp\left(-b_2 \cdot d_{ij}\right)\right)^{b_3\left(N/_{dg}\right)^{b_4}} + e_{ij} \quad (13)$$

$$\hat{h}_{ij} = 1.3 + Ho^{(b_0 + u_j)}\left(1 - \exp\left(\left(b_1/_{100}\right) \cdot HI \cdot d_{ij}\right)\right)^{\left(N/_{dg}\right)^{b_2}} + e_{ij} \quad (14)$$

$$\hat{h}_{ij} = 1.3 + Ho^{(b_0 + u_j)}\left(1 - \exp\left(\left(b_1/_{10}\right) \cdot HI \cdot d_{ij}\right)\right)^{\left(N/_{dg}\right)^{b_2}} + e_{ij} \quad (15)$$

Table 4 Estimated parameters and fitting statistics obtained for the generalized model with and without mixed effects for the groups of species considered

Species	Equation	Fitting method	Parameters						Statistics			
			b_0	b_1	b_2	b_3	b_4	σ_u^2	R^2	RMSE	Bias	BIC
Pinus species	(8)	ONLS	4.4250	0.6125	0.0392	1.5073	−0.0847		0.82	2.41	−0.013	7141.6
	(12)	NLMIXED	4.1798	0.6330	0.0398	1.8370	−0.1428	0.1150	0.84	2.29	−0.011	6730.5
Quercus species	(8)	ONLS	1.1200	0.9694	0.0363	2.8325	−0.2697		0.66	2.45	0.031	3267.4
	(13)	NLMIXED	1.0046	0.9804	0.0375	2.5189	−0.2482	0.0017	0.72	2.23	−0.019	2941.0
Other conifer species	(9)	ONLS	1.2830	0.0002	−0.1036				0.64	1.74	0.118	224.5
	(14)	NLMIXED	1.2923	0.0231	−0.1038			0.0018	0.74	1.49	0.029	177.1
Other broadleaf species	(9)	ONLS	1.3120	0.0001	−0.1634				0.63	1.41	−0.011	226.0
	(15)	NLMIXED	1.1986	0.0017	−0.1352			0.0027	0.79	1.06	−0.094	63.8

ONLS = fitted by ordinary least squares, NLMIXED = fitted by non-linear mixed effects.

where $b_0 - b_4$ are the fixed parameters of the model (common to all plots); $(u_j, v_j) \sim N(0, \tau)$ are the random parameters (specific to each plot); and $\hat{h}_{ij}$ and e_{ij} are respectively the height and error estimated by the model for the i-th observation (tree) in the j-th plot.

The values of the parameters and goodness-of-fit statistics for the local mixed models (Eqs 10 and 11) and for the generalized mixed models (Eqs 12 to 15) are shown in Tables 3 and 4, respectively. We compared the *RMSE* values obtained with the mixed effects equations with those obtained with fixed effects equations (fitted by ONLS); the values obtained with the local mixed model (Eq. 10) and the generalized mixed model (Eq. 12) for the *Pinus* grouping were 25.0% and 5.2% lower than those obtained with the local model (Eq. 7) and the generalized model (Eq. 8) without random parameters, respectively. For the group of *Quercus* species, the *RMSE* values obtained with the local mixed model (Eq. 11) and the generalized mixed model (Eq. 13) were 26.0% and 9.0% lower than those obtained with the local (Eq. 7) and the generalized models without random parameters (Eq. 8), respectively. For the other conifers, the *RMSE* values obtained with the local mixed model (Eq. 11) and the generalized mixed model (Eq. 14), were 9.8% and 14.3% lower than those obtained with Eqs (7) and (9), respectively. For the group comprising other broadleaf species, the *RMSE* values were 20.9% and 25.0% lower with the local mixed model (Eq. 11) and the generalized mixed model (Eq. 15) than with Eqs (7) and (9). The results obtained for *BIC* and R^2 were similar to those obtained for *RMSE*.

On inspecting the graphs of the residuals for the heights estimated by the models for each species grouping, we did not find any anomalies that would suggest non compliance of underlying hypothesis of independence of errors or homogeneity of variance. The magnitude of the bias in the residual values estimated by the two fitting methods (ONLS and NLMIXED) was consistent for all ranges and classes of heights observed by the defined species groupings.

Calibrated response

Calibration option CR1 for Eq. (12) in *Pinus* species and Eq. (13) in *Quercus* species was the most accurate when the total height of a subsample of 3 trees close (± 10%) to the mean breast height diameter for the plot was measured (Figure 2), as indicated by the slight decrease in the *RMSE* by 0.4% and 3.0% respectively for the two groups relative to the values estimated by the generalized model (Eq. 8) fitted without random parameters. For the other broadleaf species, calibration option CR2 for Eq. (14) was the most accurate (in terms of *RMSE*) when a subsample of 3 trees of mean, minimum and maximum breast height diameter were measured in each plot, as the *RMSE* was 21% lower than that estimated with the generalized model fitted without random parameters (Eq. 9). Finally, in the group of other conifer species, the *RMSE* value obtained with Eq. (15) and calibration option CR2 was 13% lower than that obtained with the generalized model fitted without random parameters (Eq. 9). Vargas-Larreta *et al.* (2009) found that the decrease in *RMSE* varied between 3.7 and 13.3% on calibrating the model of Sharma and Parton (2007) with data from 1 tree selected at random in the plot. However, Calama and Montero (2004) observed that use of a subsample of the 5 trees with the largest diameters significantly decreased the *RMSE* value; Castedo Dorado et al. (2006) observed that the bias estimated with Schnute's generalized equation was lower when a subsample of the 3 trees with the smallest diameters was used for calibration than when the estimate was made without random parameters.

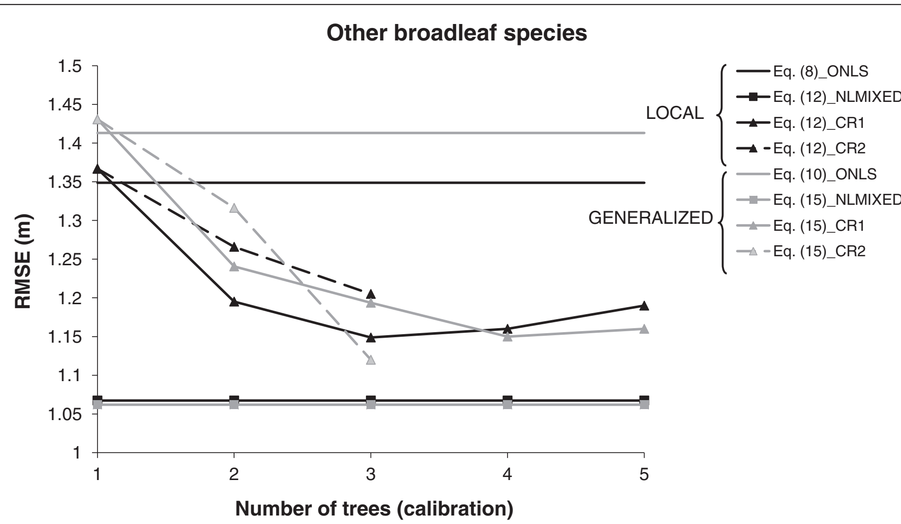

Figure 2 Value of RMSE for the calibrated equations (CR1 and CR2) fitted by NLMIXED and ONLS for the four groups of species shown.

The variation in the value of *RMSE* with respect to the number of trees used with the two calibration options for the four main groups of species studied is shown in Figure 2. This statistic was also compared with those values obtained when fitting the equations by the NLMIXED (minimum value of *RMSE* reached only using all trees as a calibration subsample) and ONLS (maximum value of *RMSE* using only fixed parameters) methods.

In the calibration process, the reduction in the *RMSE* value was particularly evident with the generalized mixed models for the other broadleaf species and other conifer species (21.0% and 13.0% respectively) compared with the generalized model fitted without random parameters; however, for the *Pinus* and *Quercus* groupings, the decrease in the value of this statistic was lower. Both calibration options resulted in an important reduction of *RMSE* for the local mixed model compared to the same model fitted without random parameters for all the groups analized. In accordance with Trincado *et al.* (2007), the use of a local mixed model in forest inventories with a subsample of trees to calibrate and then predict the total height of all trees not used in calibration allows retention of a simple model structure (i.e. without the need to include stand predictor variables) and may be an useful alternative to generalized mixed models when there is a lack of data to calculate stand variables.

Conclusions

Two generalized equations (Eqs 8 and 9) were derived from a local equation (Eq. 7) and used to estimate total tree height from breast height diameter and stand variables for the 25 species identified in the sample by using mixed models. The variability between plots is explained in terms of the random effect of each plot and from the stand variables included in the generalized models.

For species in the *Pinus* and *Quercus* groups, inclusion of the height measurements of 3 trees close (± 10%) of the mean breast height diameter from each plot improved the predictive capacity of the calibrated model. For the species included in other broadleaf species and other conifer species, the predictive capacity of the model was improved by including the total height measured in a subsample of 3 trees of minimum, mean and maximum breast height diameter. The possibility of using complementary data from the stands to calibrate the mixed models provides a clear advantage over models developed by other procedures, which require large amounts of data or are less accurate.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

SCR provied the experimental data, wrote the article and analyzed the data with JGAG and FCC. JJCR supervised the work and coordinated the research project. All authors read and approved the final manuscript.

Acknowledgements

The present investigation was financially supported by the "*Programa de Mejoramiento del Profesorado*" (project: Seguimiento y Evaluación de Sitios Permanentes de Investigación Forestal y el Impacto Socioeconómico del Manejo Forestal en Norte de México). The study was conducted during the doctoral studies of the first author at the *Universidad de Santiago de Compostela* USC (Spain), supported by "Programa Banco Santander – USC" (becas para estancias predoctorales destinadas a docentes e investigadores de América Latina).

Author details

[1]Instituto Tecnológico de El Salto, Mesa del Tecnológico s/n. Apdo. Postal No. 2, El Salto P.N. 34950, Durango, México. [2]Departamento de Ingeniería Agroforestal, Universidad de Santiago de Compostela. Escuela Politécnica Superior, Campus Universitario, 27002 Lugo, España. [3]Instituto de Silvicultura e Industria de la Madera, Universidad Juárez del Estado de Durango, Blvd. del Guadiana # 501., Cuidad Universitaria, 34160 Durango, México.

Received: 15 May 2013 Accepted: 19 September 2013
Published: 26 February 2014

References

Bates DM, Watts DG (1980) Relative curvature measures of nonlinearity. J R Stat Soc 42:1–16

Bertalanffy LV (1949) Problems of organic growth. Nature 163:156–158

Bi H (2000) Trigonometric variable-form taper equations for Australian eucalyptus. For Sci 46:397–409

Calama R, Montero G (2004) Interregional nonlinear height–diameter model with random coefficients for stone pine in Spain. Can J For Res 34:150–163

Castedo Dorado F, Diéguez-Aranda U, Barrio Anta M, Sánchez Rodríguez M, Gadow KV (2006) A generalized height–diameter model including random components for radiata pine plantations in northwestern Spain. For Ecol Manage 229:202–213

Curtis RO (1967) Height–diameter and height–diameter–age equations for second-growth Douglas-fir. For Sci 13:365–375

García ME (1981) Modificaciones al Sistema de Clasificación Climática de Köppen, 4ath edition. Instituto de Geografía, Universidad Nacional Autónoma de México, México D.F

Gregoire TG (1987) Generalized error structure for forestry yield models. For Sci 33:423–444

Hossfeld JW (1822) Mathematik für Forstmänner, Ökonomen und Cameralisten, 4th edition Gotha, Hennings p 472.

Huang S, Price D, Titus SJ (2000) Development of ecoregion-based height–diameter models for white spruce in boreal forests. For Ecol Manage 129:125–141

Lappi J (1997) A longitudinal analysis of height/diameter curves. For Sci 43:555–570

Littell RC, Milliken GA, Stroup WW, Wolfinger RD (1996) SAS System for Mixed Models. SAS Institute Inc., Cary

López Sanchez CA, Varela JG, Dorado FC, Alboreca AR, Soalleiro RR, Alvarez Gonzalez JG, Rodriguez FS (2003) A height–diameter model for *Pinus radiata* D. Don in Galicia (Northwest Spain). Ann For Sci 60:237–245

Meng SX, Huang S, Lieffers VJ, Nunifu T, Yang Y (2008) Wind speed and crown class influence the height-diameter relationship of lodgepole pine: Nonlinear mixed effects modeling. For Ecol Manage 256:570–577

Meyer HA (1940) A mathematical expression for height curves. J For 38:415–420

Pinheiro JC, Bates DM (1998) Model Building for Nonlinear Mixed Effects Model. Department of Biostatistics, University of Wisconsin, Madison, Wis.. 11

Richards FJ (1959) A flexible growth function for empirical use. J Exp Biol 10:290–300

SAS Institute Inc (2008) SAS/ETS® 9.2 User's Guide. SAS Institute Inc, Cary, NC, p 2861

Schmidt M, Kiviste A, Gadow KV (2011) A spatially explicit height-diameter model for Scots pine in Estonia. Eur J For Res 130:303–315

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 5(2):461–464

Sharma M, Parton J (2007) Height–diameter equations for boreal tree species in Ontario using a mixed-effects modeling approach. For Ecol Manage 249:187–198

Sharma M, Zhang SY (2004) Height–diameter models using stand characteristics for *Pinus banksiana* and *Picea mariana*. Scand J For Res 19:442–451

Soares P, Tomé M (2002) Height–diameter equation for first rotation eucalypt plantations in Portugal. For Ecol Manage 166:99–109

SRNyMA (2006) Programa Estratégico Forestal 2030. Secretaría de Recursos Naturales y Medio Ambiente del Estado de Durango, Victoria de Durango, Dgo, p 242

Stage AR (1975) Prediction of height increment for models of forest growth. USDA For Serv Res Pap INT-164:20

Temesgen H, Gadow KV (2004) Generalized height-diameter models–an application for major tree species in complex stands of interior British Columbia. Eur J For Res 123:45–51

Temesgen H, Hann DW, Monleon VJ (2007) Regional height-diameter equations for major tree species of southwest Oregon. West J Appl For 22:213–219

Trincado G, VanderSchaaf CL, Burkhart HE (2007) Regional mixed-effects height–diameter models for loblolly pine (*Pinus taeda* L.) plantations. Eur J For Res 126:253–262

Vargas-Larreta B, Castedo-Dorado F, Álvarez-González JG, Barrio-Anta M, Cruz-Cobos F (2009) A generalized height-diameter model with random coefficients for uneven-aged stands in El Salto, Durango (Mexico). Forestry 84(2):445–462

Vonesh EF, Chinchilli VM (1997) Linear and Nonlinear Models for the Analysis of Repeated Measurements. Marcel Dekker Inc., New York. 560 p

Wehenkel C, Corral-Rivas JJ, Hernández-Díaz JC, Gadow KV (2011) Estimating balanced structure areas in multi-species forests on the Sierra Madre Occidental, Mexico. Ann For Sci 68:385–394

Weibull W (1951) A statistical distribution function of wide applicability. J Appl Mech 18:293–297

Wykoff WR, Crookston NL, Stage AR (1982) User's Guide to the stand prognosis model. USDA For Serv Gen Tech Rep INT-133:122

Zeide B, Vanderschaaf C (2002) The Effect of Density on the Height-Diameter Relationship. In: Outcalt KW (ed) Proceedings of the 11th Biennial Southern Silvicultural Research Conference. 2001 March 20–22. USDA Forest Service, Gen. Tech. Rep. SRS–48, Asheville, NC, Knoxville, TN, pp 463–466

doi:10.1186/2197-5620-1-6
Cite this article as: Corral-Rivas *et al.*: Local and generalized height-diameter models with random parameters for mixed, uneven-aged forests in Northwestern Durango, Mexico. *Forest Ecosystems* 2014 1:6.