

Validating generalised diameter-height models for several species and heterogeneous crop structures

M.J. Hawkins^{a*}, K. Black^{a,b} and J. Connolly^c

Abstract

An examination of the suitability of generalised height-diameter models for growth modelling and augmenting inventory measurements was undertaken. A large database of repeated measurements taken from crop structure experiments since 1963 to the present in Ireland was used. We used a distance independent individual tree height-diameter model to investigate whether inclusion of competition variables can be used to predict variations in height across a wide range of species and silvicultural management regimes. To this end, we stratified the heterogeneous dataset post-hoc into a variety of constituent species, management and silvicultural strata. In addition, we attempted to control for site-specific effects and serial correlation by using a mixed-effects framework in an effort to identify site specific height-diameter variables not explained by the model. The generalised model typically performed well for each species and silvicultural treatment. The most noticeable impact of treatment was observed in plots of Sitka spruce (*Picea sitchensis* (Bong.) Carr.) with differing spacing. The magnitude of inter-plot variability as modelled by a random effect related to the height asymptote varied between species, possibly as a result of inter-species differences in tolerance to variability in environmental growing conditions. Following validation against external data, we show that these generalised models could be used when, in the case of growth modelling for example, it is sometimes necessary to derive individual tree heights from individual tree diameters, perhaps in standard inventory plots where tree height is not measured on every instance that DBH is measured.

Keywords: *Tree height-diameter modelling, individual-tree model, tree competition.*

Introduction

The goal of this study was to find individual-tree, age- and location-independent, species-specific prediction equations that can be used for plots at any stage in their lifecycle under a wide variety of management regimes. Forest inventory datasets usually contain many more measurements of diameter at breast height (DBH, cm) than tree height (H, m). This practice often comes about because it is the DBH distribution which is more variable than the H distribution, and because it may be assumed that the DBH-H relationship can be modelled for the unmeasured heights to be predicted with this model. A common forest inventory approach to DBH-H modelling is to use Chapman-Richard models based on species and plot-specific predictions (Wykoff et al. 1982). However, Chapman-Richards and similar functions are problematic when used as generalised models because the solved function approaches the asymptote too rapidly, particularly when there is a weak relationship between DBH and H in larger trees and across different sites (Temesgen and von Gadow 2004).

^a UCD Forestry, School of Agriculture and Food Science, Dublin 4.

^b FERS Ltd., 117 East Courtyard, Tullyvale, Cabinteely, Dublin 18.

^c UCD Statistics and Actuarial Science, School of Mathematical Science, Dublin 4.

* Corresponding author: michael.hawkins@ucd.ie

The height-diameter relationship can vary between individual trees in a dataset due to competition, structural allocation variations across different silviculture management types (Cameron and Watson 1999), or variations in site conditions, such as degrees of exposure (Brüchert and Gardiner 2006). Distance independent DBH-H models fitted on the scale of the tree – incorporating information on tree size, inter-tree competition and site differences – have been successfully used to describe variations in height across sites varying in respect of environmental and competitive conditions (Monserud and Sterba 1996, Temesgen and von Gadow 2004, Uzoh and Oliver 2006).

In Ireland, there is an increasing need to develop individual tree growth models and height-diameter functions to facilitate the projection of volume or carbon stock changes using the National Forestry Inventory (NFI 2007). Projections and annual interpolation of tree diameter and height between repeated inventory cycles are particularly relevant for reporting annual carbon stock changes to the United Nations Framework Convention on Climate Change. In this context, height-diameter functions are required to derive height estimates from individual tree diameter increment models, such as those described by Monserud and Sterba (1996).

Temesgen and von Gadow (2004) derived nonlinear regression models that estimate height of individual trees in a stand or plot as a function of DBH, using covariates of competition proxies that are calculated without using the spatial coordinates of trees in the stand or plot (i.e. models that are both age and distance independent). They found that using these competition covariates produced DBH-H models with improved accuracy of prediction, compatibility among the various estimates in a growth and yield model, and maintained projections within reasonable biological limits.

In this study, we used a heterogeneous database pertaining to experimental plots to investigate if the inclusion of these described competition variables can be incorporated to accurately predict, with minimal bias, variations in height across a wide range of species, sites and silvicultural management regimes. For our DBH-H model we followed the approach developed by Temesgen and von Gadow (2004) who described competition using plot density (DENS, trees ha⁻¹), plot basal area (BA, m² ha⁻¹), and basal area in larger trees (BAL, m² ha⁻¹). (We calculated BAL as the basal area of all trees in the plot whose DBH were greater than the target tree. BAL calculation was made for each measurement occasion, as it was used for DENS and BA.) Unlike Temesgen and von Gadow, we incorporated random site effects as well. The performance of these models was assessed using a randomly sampled, independent and external datasets. The performance of these models in mixed species stands, such as the Sitka spruce¹ and southern provenances of lodgepole pine (*Pinus contorta* Dougl.) mixture on planted blanket peats in Ireland were assessed.

Methods

Datasets for model development and validation

We describe here the relevant aspects of the data in our study. Interested readers will find additional detail pertaining to the datasets in Broad and Lynch (2006a). The data in

¹ The full botanical names and authorities for all species are listed in Table 1.

our study were obtained from Coillte Teoranta's permanent sample plot record system. The dataset contains records from many spacing, respacing and thinning trials (as well as unreplicated sample plots) established during the period 1963 to 2001. The trials were initially established in the 1960s, 1970s and 1980s as replicated experimental designs with consecutive measurements typically made up to 2001. It was common practice to record diameter measurements on an annual basis in the early years of the experiment, with full plot repeated measurements occurring at 3-5 year thinning cycle intervals; thereafter DBH was typically measured for all trees on all measurement occasions. In each permanent sample plot, from 10 to 20 tree heights (depending on plot size) were measured. The experiments were laid out in plots (varying from 0.01 to 0.21 ha) and plot stocking was known at the time of measurement. The species are listed in Table 1.

The permanent sample plot (PSP) trials were set out in ca. 2,900 permanent sample plots with various species and silvicultural treatments, including thinning, spacing and pruning. The dataset used to develop the models described here (the calibration dataset) contained 1170 permanent sample plots. Data from plots were excluded where simultaneous measurements of DBH and H were missing. In addition, all pruning experimental data were excluded from the calibration dataset.

Plots used for thinning and spacing experiments were included in the dataset to which the models were fitted (Tables 2 and 3). The thinning treatments contained in the dataset included (Table 3): no thinning (NTH), no thinning with removal of dead trees (NLT), line thinning -1st cycle only, subsequently selective thinning (LS); selective thinning (SEL); systematic thinning (SYS) and thinning of dominant trees (DOM). Thinning intensity was generally constant across all treatments, using a moderate intensity prior to the advent of the marginal thinning intensity concept introduced

Table 1: Details of species sampled in the PSP database and the grouping of species or provenances in modelling exercise. Note: both larch species were grouped together.

Common name	Binomial name	Provenance (if known)
Common alder	<i>Alnus glutinosa</i> L.	
Common ash	<i>Fraxinus excelsior</i> L.	
Douglas fir	<i>Pseudotsuga menziesii</i> (Mirb.) Franco	Washington and coastal provenances
Japanese larch	<i>Larix kaempferi</i> Fortune ex Gord. <i>Larix decidua</i> Mill.	
Lodgepole pine	<i>Pinus contorta</i> Loud.	South and north coastal provenances
Monterey pine	<i>Pinus radiata</i> D. Don	
Norway spruce	<i>Picea abies</i> Karst.	
Pedunculate oak	<i>Quercus robur</i> L.	
Scots pine	<i>Pinus sylvestris</i> L.	Scottish
Sitka spruce	<i>Picea sitchensis</i> (Bong.) Carr.	QCI

by Bradley et al. (1966) and further developed by Hamilton and Christie (1971). Thinning intensity was also investigated in a small set of experiments (Gallagher 1966, 1969, 1972, Gallagher et al. 1987) wherein heavy, light and moderate thinning were applied, based on the Forestry Commission A-D thinning grades. These thinning grades determined by height or age class, or a relative proportion of basal area at the time of thinning (Gallagher 1969, 1972). The data were not categorised by thinning intensity because these were not documented in the PSP database.

External independent data were used to validate the fitted models. These data came from a cross-sectional sample in 2003 of plots within Coillte production stands (Broad and Lynch 2006a). Such cross-sectional comparisons are particularly useful for assessing inter-plot bias due to a more randomized sampling approach in the validation dataset. These sample plots comprised thinned and un-thinned stands, which had been initially planted at a spacing of 2,500 stems ha⁻¹ for coniferous species. External validation data were not available for all species.

Site-to-site variations in the relationship between height and diameter are often well described by inclusion of dependent variables such as aspect, slope or exposure (Uzoh and Oliver 2006). These data are not always captured in sample plot databases, so they cannot be included as dependent variables in the model. These variables were not included in our dataset, so we accounted for plot-to-plot variability with a site variability parameter (Equation 1) in a mixed-effects model (McCulloch et al. 2008).

Table 1 presents the species represented in the PSP database. Table 2 presents summary statistics for the height and diameter at breast height and illustrate the two main features of the repeated sampling structure of the dataset. These tables show the number of repeated measurements on individual trees, which varied from 1 to 18, classified by silvicultural treatment and species. They also showed the degree to which repeated H measurements on sample trees occurred less frequently than their repeated DBH measurement counterparts. Table 3 shows the heterogeneity of experimental treatments present in the dataset, in terms of different thinning intensities and planting spacing distances, for Sitka spruce and lodgepole pine. All the other species were insufficiently represented in the dataset to allow a comparison of different silvicultural regimes.

Table 2: Summary of the parameterisation dataset, showing range of DBH (cm) and H (m) measurements and summary statistics on counts of series length for consecutive H and DBH measurements on individual trees. Also shown is the number of plots and trees stratified by species.

Species	Common ash	Norway spruce	Douglas fir	Monterey pine	Sitka spruce	Pedunculate oak	Japanese larch	Lodgepole pine	Scots pine	
DBH value (cm)	min	7	2.6	5.3	7.6	4.3	9	7.1	4.4	7.4
	max	15.1	46.7	50.8	38.7	91.2	33.5	38.3	37.7	38.9
	mean	10.3	18.1	15.0	17.3	19.2	17.5	21.9	14.1	17.8
	sd	2.1	6.5	5.9	6.7	8	6.1	5.1	4.3	5.3
DBH (nos. of measurements per tree)	min	2	1	1	5	1	4	1	1	1
	max	3	16	14	6	18	6	6	11	14
	mean	2.79	7.8	3.85	5.9	7.3	5.4	5.5	4.3	9.1
Height value (m)	min	8.5	5	3.8	7.5	3.3	8.8	6.1	1.9	7.7
	max	17.1	26.8	33.6	17.2	47	16.5	26.4	22.5	22.2
	mean	13.2	14.4	12.3	11.3	14.9	13.7	17.2	10.1	15.4
	sd	1.8	3.9	3.9	2.7	5.7	1.8	3.1	2.8	2.3
Height (nos. of measurements per tree)	min	1	1	1	1	1	1	1	1	1
	max	1	3	3	2	6	2	2	3	5
Nos. of plots	min	1	1.1	1.1	1.1	1.2	1.1	1.01	1.1	1.14
	max	12	71	98	1	424	1	98	415	18
Nos. of trees	71	1288	1239	10	5631	10	969	4959	417	

Table 3: The number of trees of Sitka spruce and lodgepole pine classified by plot treatment. DOM = dominant trees removed in early thinning, LS = line selection first thinning, Nxx = no thinning at planting spacing of xx metres, NLT = only dead trees are removed, NTH = no thinning, Sxx = selective thinning at planting spacing of xx metres, SEL = selective thinning, SYS = systematic thinning. Explanatory Note: total number of observations exceeds the number of trees (cf. Table 2) because of repeated measurements.

Treatment	Sitka spruce	Lodgepole pine
DOM	143	0
LS	520	718
N1.22	350	278
N1.83	112	347
N2.44	159	573
N3.05	607	466
N3.67	275	291
NLT	283	41
NTH	2036	1080
S1.22	327	416
S1.83	278	353
S2.44	278	379
S3.05	150	282
SEL	4219	1661
SYS	5976	1580

Nonlinear mixed-effects model

As noted earlier, series of consecutive height measurements for individual trees were much shorter on average (e.g. Series mean ≈ 1) than series of consecutive DBH measurements (Table 2). Such short series made it unfeasible to parameterise the DBH-H models at the level of the individual tree, e.g. to estimate tree-specific coefficients for each tree in the dataset. The parameters of the model equation that were initially fitted to each species in turn is given in Equation 1. This equation corresponds in most respects to Model 7 from Temesgen and von Gadow (2004), but differs slightly on account of the inclusion of a plot-level random effect related to the asymptote.

$$E(H_{ijk} | u_i) = (u_i + a + a_{ba} \cdot BA_{ijk} + a_{dens} \cdot DENS_{ijk} + a_{bal} \cdot BAL_{ijk}) (1 - \exp(b \cdot DBH_{ijk}^{(c + c_{bal} \cdot BAL_{ijk})})) \quad [1]$$

In Equation 1, plots, trees and measurement occasions are indexed by i , j , and k , respectively. $E(\cdot)$ is the expectation operator. Equation 1 shows the modelled mean

of a conditional Gaussian distribution with mean zero and variance σ_e^2 (Table 4). Symbolically, the terms in the model equation are interpreted as “*coefficient.variable*”, i.e. *aba* is the coefficient of the variable *BA*, and so on. The plot level random effect $u_i \sim \text{Gaussian}(0, \sigma_u^2)$ accounts for correlations between consecutive measurements on the same plot (McCulloch et al. 2008) as well as quantifying inter-plot variability. The model in Equation 1 was selected through a process of model fitting which involved evaluating different model equations for suitability. For example, we tried introducing further parameters to describe correlation and inter-subject variability, i.e. inter-tree variance parameters, but these models were not an improvement on Equation 1.

The *b* Parameter is *a priori* negatively valued (Table 4) so the model equation realistically represents the shape of the empirical DBH-H relationship. The estimated asymptote for trees in the i^{th} plot as DBH tends to infinity, all covariates having been set to zero, is given by $a + u_i$. Models were fitted to each species dataset separately using the algorithms in the SAS NLMIXED procedure (SAS Institute Inc. 2009). The *lattice* library in R v2.10.1 was used to produce the graphical summaries (Sarkar 2008).

Results

The parameter values of the best-fitting models are given in Table 4. The empty cells in Table 4 refer to those parameters that were either not statistically significant (at significance level $\alpha = 0.05$), or that the model did not converge with that parameter included, or that the fitted parameters were inconsistent with the results obtained by Temesgen and von Gadow (2004). (Their results were used as an external benchmark check, which was particularly useful in those cases where convergence was difficult to achieve or the parameter estimates were dubious.) We set the significance level for model selection at $\alpha = 0.05$, but parameters were typically significant below this level. For common ash, the inter-plot variance parameter was not statistically significant, but was retained in the final model because, while its inclusion did not affect the estimated values of the coefficients of the other covariates, we considered it desirable to include inter-plot variability in the estimates of the standard errors of the coefficients.

External validation

External validation data were available for a subset of the species grown in pure plots: Douglas fir, lodgepole pine, Norway spruce, Scots pine, and Sitka spruce. Comparisons of the external validation data and modelled data are shown in Table 5. We also show the empirical distribution of the external validation residuals in Figure 1. Plot size effects were looked for in the external validation residuals but none were found. Residuals plotted against DBH and BAL for all species tested similarly and showed no correlation (data not shown). However, it was evident (Figure 2) that height was overestimated in Sitka spruce stands with a low stocking density (ca. <200 stems ha^{-1} residuals were greater than 5 m). Residuals derived from validation plots with stands of a stocking density below 200 stem ha^{-1} were more likely to have been artefacts of differences in the respective ranges of the DENS variable in the fitting and validation data. In any event, such densities might be considered very atypical in practice.

Mixed plots

RMSE and bias for intimate mixtures of lodgepole pine and Sitka spruce are presented in Table 5. There were few mixture plots present in the dataset and no consecutive height measurements were present in the data for the mixture plots. There were a total of 185 height measurements from nine mixed lodgepole pine and Sitka spruce plots. As such, the results pertaining to mixtures in Table 5 are given in the interest of completeness, rather than as a conclusive or extensive analysis of DBH-H relationships in mixed-species plots.

Management and thinning effects

The dataset consisted of experimental data from many different types of silvicultural and thinning trials (Table 3). Smoothed density estimates of the empirical residual distributions for these trials are shown in Figures 3a and 3b. In general, all experimental types were modelled in an unbiased way, and the residual distributions are symmetrical around zero.

Figures 3a and 3b and Table 5 shows that the lodgepole pine model was more accurate than the Sitka spruce model. The differences in species models were most pronounced in plots where both spacing and selective thinning treatments had been applied and where the spacing levels varied across plots (cf. the S_{xx} panels in Figure 3b). There is some evidence that the models for each species performed best at intermediate spacing levels, both in thinned and unthinned plots (Figure 3b). In both lodgepole pine and Sitka spruce, model accuracy was greater in plots where spacing was the only treatment applied, than where spacing and selective treatments were applied (compare rows in Figure 3b).

Table 4: Estimated model parameters significant at least at $\alpha = 0.05$. Parameter symbols are explained in Equation 1 and related text.

Species	Parameter								
	a	a_{ba}	a_{dens}	a_{bal}	b	c	c_{bal}^a	σ_u^2	σ_e^2
Common alder	10.9	0.18	-0.001		-0.13	0.8		2.300	1.01
Common ash	13.3	0.14			-0.18			0.016	0.99
Douglas fir	11.3	0.34	-0.001		-0.07			2.900	1.94
Japanese larch	19.6	0.17	-0.003		-0.07				3.94
Lodgepole pine	11.1	0.11	-0.001		-0.11			3.900	1.04
Monterey pine	19.7			-0.098	-0.06				1.90
Norway spruce	34.5	0.33	-0.003		-0.07	0.6		31.040	1.18
Pedunculate oak	6.3	0.29		0.050	-0.17				0.96
Scots pine	26.2		-0.003		-0.16	0.7		6.700	1.20
Sitka spruce	12.7	0.26	-0.002	-0.003	-0.07			12.100	2.20

^a This parameter was not significant; however, it was included in the parameter list here because it featured in Equation 1.

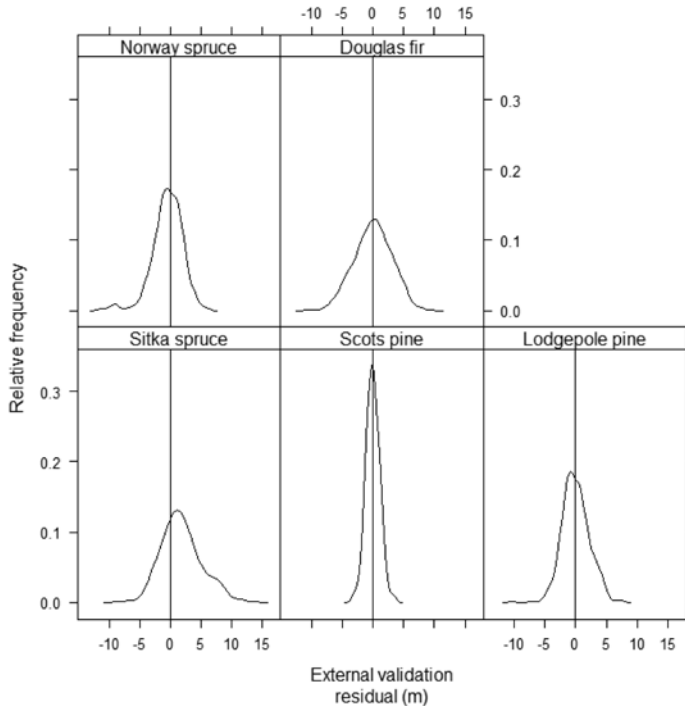


Figure 1: Smoothed empirical frequency distributions of external validation residuals (Actual height - Predicted height).

Discussion

Many previously published DBH-H models have not been validated against an independent (external) dataset to investigate the presence of sampling bias in the model parameterisation dataset, as done here. This issue was also highlighted by Broad and Lynch (2006b). Prediction models tend to perform better on data from which they were constructed than on new data. Results are often accepted without sufficient regard to the importance of external validation. The limitations of internal validation are acknowledged and this work incorporates an external validation to demonstrate the potential generalisability of a diagnostic prediction model to future settings or independently sampled data.

Temesgen and von Gadow (2004) defined generalised DBH-H models as equations that predict tree heights using information on both individual tree DBH and plot or stand level information, such as stand basal area or plot density. Individual-tree distance-independent DBH-H models that do not incorporate information on the plot make the implicit assumption that competition (as measured on the scale of the plot/stand by DENS, and BA, and on the scale of the tree by BAL) does not affect the DBH-H relationship over the lifetime of a tree that is subject to management influences, e.g. spacing and thinning. Plot-specific DBH-H models, that do not condition on plot/stand covariates, are often fitted by a multi-step approach wherein

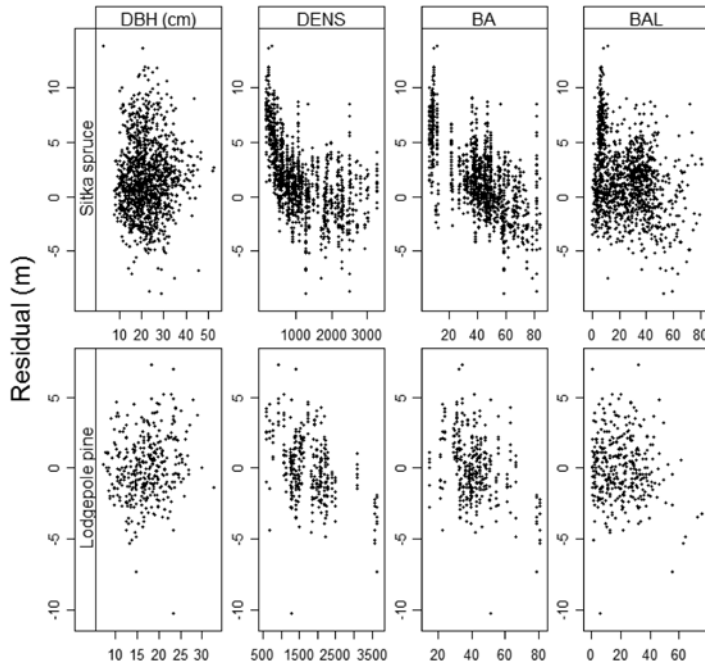


Figure 2: External validation residuals from the lodgepole pine and Sitka spruce models plotted against a subset of covariates. Note, not all covariates shown here were included in the final models (cf. also Table 4). Scale varies among panels. *DENS* is plot density (trees ha^{-1}), *BA* is plot basal area (*BA*, $m^2 ha^{-1}$), *BAL* is basal area in larger trees (*BAL*, $m^2 ha^{-1}$).

separate DBH-H relationships for each plot are fitted. Competition effects are thus at best implicitly described by the variation in each plot-specific fitted parameter.

Generalised models, as defined above, attempt to deal with the broadest response range, so perform better on plots that are near the centre of the sample space dataset rather than plots subject to relatively atypical management conditions. (In fact, this is similar in principle to a standard result in regression modelling, whereby accuracy is maximised at the mean).

Generalised models tend to borrow strength across plots/stands, meaning that issues related to data sparseness on individual plots are mitigated. A plot-specific approach can encounter problems if data for a given plot are so sparse as to not support model fitting. When this occurs in practice, parameters are sometimes pinned at their generalised values, i.e. they are fixed at their value estimated using data from all plots (this approach was used in Ireland's NFI, for example), and the remainder of the parameters are estimated with whatever plot-specific data exist. If there are many such plots in the dataset, this process of estimating parameters plot-by-plot can be tedious; hence the appeal of our approach, which conflates the better aspects of the generalised and plot-specific approaches.

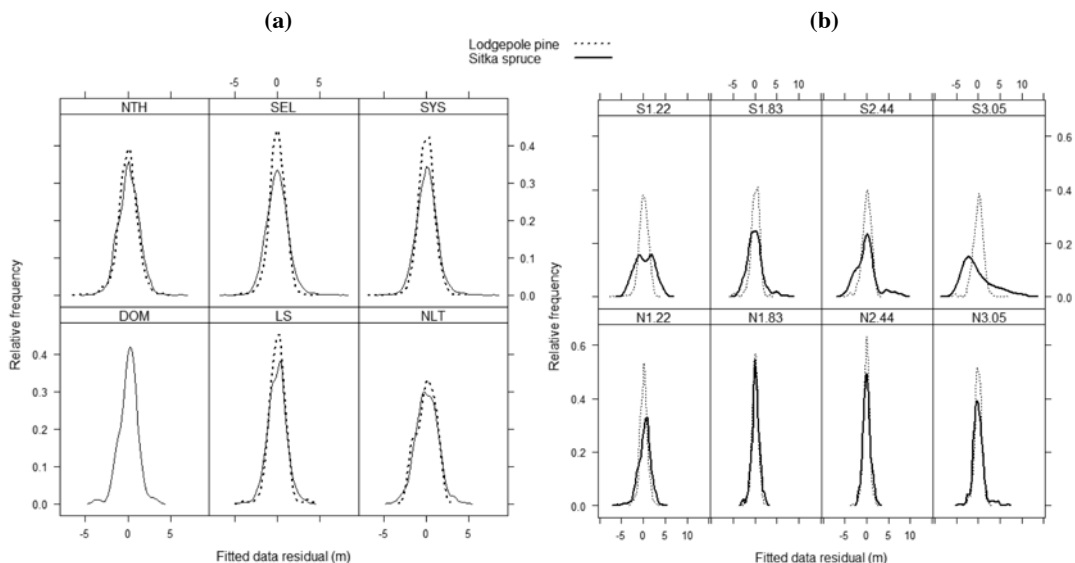


Figure 3: Classifying fitted-data residuals by experiment type and treatments applied to plots. Only lodgepole pine (broken line) and Sitka spruce (continuous line) are shown because they were the most abundant species in the dataset.

Data depicted in (a) relates to thinning experiments treatments, where “NTH” denotes unthinned plots, “SEL” denotes selective thinning treatments, “SYS” denotes systematic thinning, “DOM” denotes dominant tree removal, “LS” denotes line and selective thinning plots, and “NLT” denotes removal of dead trees only. See also Table 3.

Data shown in (b) relates to spacing and thinning experiments. The top row shows selectively thinned plots at an initial spacing of (from left to right) 1.22, 1.83, 2.44 and 3.05 m. The bottom row shows non-thinning treatments at an initial spacing of (from left to right) 1.22, 1.83, 2.44 and 3.05 m.

We modelled inter-plot variability through a plot-specific random effect related to the asymptote using mixed-model technology (McCulloch et al. 2008). This model feature forms a logical link between generalized and plot-specific models because plot-specific parameters are estimated for plots where sufficient data exist (the u_i terms in Equation 1.). The resulting model equation, if there is an estimated plot-specific effect, predicts for a specific plot. To predict for plots not included in the dataset, or for those plots with insufficient data for plot-specific effect estimation, u_i was set at 0. The mixed model approach used also imposes a common correlation between each measurement on a given plot, and observations on different plots are independent. More complex correlation models (the correlation model, nested within the overall model, cf Equation 1, dealing with how measurements on the same plot or tree are related) did not improve the overall model and the selected final models, therefore fulfil the goal of finding individual-tree, age- and location-independent, prediction equations that can be used for each species for plots at any stage in their lifecycle under a wide variety of management regimes was not fulfilled.

Table 5: Measures of model performance based on fit to the data and external validation data.

$$\text{Bias} = \sum_1^n \frac{H_i - \hat{H}_i}{n} \text{ and } \text{RMSE} = \sqrt{\sum_1^n \frac{(H_i - \hat{H}_i)^2}{n - p}}$$

where \hat{H}_i denotes height (m) predicted by the model, p represents the model dimension and n = the total number of times in a given species dataset, that DBH and H were measured on the same tree on the same measurement occasion. Validation data were not available for all species.

Species	Fitted data			Validation data (where applicable)		
	Bias/RMSE ^a			Bias/RMSE ^a		
	Bias ^a	RMSE	(%)	Bias ^a	RMSE	(%)
Common alder	-0.00	0.99	-0.06			
Common ash	0.00	1.10	0.18			
Japanese larch	0.01	1.99	0.25			
Lodgepole pine	0.01	0.99	0.90	-0.00	2.30	0.20
Lodgepole pine (in mixture with Sitka spruce)	0.00	1.87	0.13			
Scots pine	0.00	1.11	0.00	-0.04	1.20	0.30
Norway spruce	0.00	1.10	0.30	-0.48	2.50	1.92
Douglas fir	0.10	1.40	3.60	0.01	3.10	0.16
Monterey pine	0.03	1.60	1.90			
Sitka spruce (pure stand)	0.04	1.50	2.70	1.80	3.80	47.00
Sitka spruce (in mixture with lodgepole pine)	-0.01	1.74	0.34			
Pedunculate oak	-0.00	1.20	0.10			

^a Rounded to two decimal places.

The mixed-effects model framework enables the user to estimate parameters for potentially complex datasets with hierarchical samples and groups. For example, the modelling task might have been accomplished with species-specific random effects models that also incorporated plot-specific effects. Such a nonlinear model with multiple groups and levels of hierarchy is very complex, as is model-fitting and parameter selection, particularly for very large datasets, such as those analysed in this study. Within our framework, the potential for other parameters to vary between plots, i.e. the b and c parameters in Equation 1, were tested but they did not improve the final model. This suggests that there is overlap between the competition proxy variables, BAL etc., and plot-specific effects that essentially “mop up” residual variability. That random b and c parameters were not statistically significant, given the prior presence of competition variables, is an indication that the competition covariates account for residual inter-plot variability. By extension, the significance of the random asymptote term, indicates a potential shortcoming of those same covariates.

Table 5 shows that RMSE is typically higher for the validation data than the calibration data. The Bias/RMSE ratio increases dramatically for Sitka spruce. This is due to a combination of reasons, primary among them being the different distribution of plot density values in the calibration and validation datasets. In the validation dataset, the median plot density of Sitka spruce plots is 900 trees ha⁻¹, compared with 1,600 trees ha⁻¹ in the calibration dataset. The Bias/RMSE ratio for the upper 50% values in the validation data was only 6%. Therefore, the increase in the ratio was largely due to an inflation in the bias value caused by a mismatch between the calibration and validation data, given that the RMSE for the Sitka spruce validation dataset is of the same order of magnitude as it is for the other species. The susceptibility of the model to this kind of mismatch is illustrated in Figure 2, where external validation residual associations with the covariates are shown for Sitka spruce and lodgepole pine. This phenomenon also partly derives from the inclusion of an asymptote in the model, because mature stands usually contain tall trees at low densities, and it is in that region of the sample space that the fitted curve begins to level off towards the asymptote. This feature is observable in the fitted data also, but at a smaller magnitude than when the model is used “out of sample”. We believe that a constant asymptote is necessary in the model so that out-of-sample predictions are robust, in the sense that out-of-sample predictions can potentially become negative without a constant asymptote term.

In the case of larch, the random asymptote model (Equation 1) did not converge. To achieve convergence we could have either omitted the asymptote constant (parameter a) from the model, or omitted the inter-plot variance component (parameter σ_u^2). If the model equation does not have a constant parameter (i.e. parameter a), implausible model estimates of values less than zero can arise. The inter-plot variance component in the final larch model was omitted. For completeness, we note that the estimated inter-plot variance parameter for the model fitted without the constant term (parameter a) was 41.87 (s.e. 6.2). This is large compared to the majority of the estimates of inter-plot variability presented in Table 4. However, the estimated values of this parameter are not directly comparable across different species because they are conditional on different subsets of covariates being included in the models.

The observed level of the inter-plot variation (as measured by the parameter σ_u^2 in Table 4) in Norway spruce may be related to species specific responses to exposure (Horgan et al. 2004, Ray et al. 2009) and other factors. Norway spruce is generally considered as very intolerant to exposure (Horgan et al. 2004), typically showing a marked reduction in the slenderness ratio, i.e. the ratio of DBH to H (Wang et al. 1998, Bröchert and Gardiner 2006). By contrast, lodgepole pine is considered to be relatively more tolerant to exposure (Horgan et al. 2004, Ray et al. 2009). In apparent concordance with that relationship, our model estimated relatively lower inter-plot variability for lodgepole pine than for Norway spruce (Table 4).

The mechanism for species-specific variations in the interplot-variability in the relevant DBH-H models is not obvious in our models. Our models do not incorporate quantifiable variables, such as aspect or level of exposure that may reinforce our posited links between species and environment, such as those mentioned in the previous paragraph. However, if additional information describing varying plot and

site conditions became available, variables such as site slope, elevation, exposure or aspect could easily be incorporated in the model (e.g. Uzoh and Oliver 2006, Monserud and Sterba 1996) or could be compared with the individual plot-effect estimates in the manner of a residual analysis since, after all, the estimated plot effects are simply residuals related to individual plots.

Conclusions and practical implications

We fitted a generalised DBH – H model incorporating covariates pertaining to tree size and competition to a dataset that encompassed a wide range of silvicultural management conditions and tree species. We post-stratified the dataset into its constituent species and experiment-type groupings, examined the model fit using an external validation dataset, and found that the generalised model performed well in the vast majority of cases. The incorporation of variables that describe site-specific conditions and how such models might relate to the relatively more empirical mixed model approach implemented here may be investigated in the future.

The practical implications of the study were:

- The generalised DBH-H models presented here can be used in forest inventories to derive height, if not available, based on DBH measurements. The derived height measurements can assist in more accurate determination of single tree volumes, top height or taper equations.
- The advantage of using single tree models, which are calibrated across a range of spacing and thinning treatments, is that one model can be used. In contrast, traditional stand-based models, which are parameterised for different silvicultural treatments (e.g. GROWFOR), use separate models for thinning and non-thinning scenarios.
- The Irish national GHG reporting system, CARBWARE, uses the described model to derive height increment based on DBH growth models. The same modeling approach could be used for timber forecasting at the single tree level. Timber projections at the single tree level provide a more accurate estimation of timber assortment distribution, when compared to stand-based models.

Acknowledgements

The work carried out for this paper was funded by the Irish Council for Forest Research and Development (COFORD). We would like to thank Mr. Ted Lynch from Coillte for the provision of the data. We acknowledge helpful comments by reviewers and editors.

References

- Bradley, R.T., Christie, J.M. and Johnson, D.R. 1966. *Forest Management Tables*. For. Comm. Book. 16. HMSO. London. 219 p.
- Broad, L. and Lynch, T. 2006a. Growth models for Sitka spruce in Ireland. *Irish Forestry* 63: 53-79.
- Broad, L. and Lynch, T. 2006b. Panel data validation using cross-sectional methods. *Irish Forestry* 63: 80-95.

- Brüchert, F. and Gardiner, B. 2006. The effect of wind exposure on the tree aerial architecture and biomechanics of Sitka spruce (*Picea sitchensis*, Pinaceae). *Journal of Botany* 93: 1512-1521.
- Cameron, A.D. and Watson, B.A. 1999. Effect of nursing mixtures on stem form, crown size, branching habit and wood properties of Sitka spruce (*Picea sitchensis* (Bong.) Carr.). *Forest Ecology and Management* 122: 113-124.
- Edwards, P.N. and Christie, J.M. 1981. *Yield models for forest management*. Forestry Commission Booklet no. 48. HMSO, London.
- Gallagher, G.J. 1966. Thinning. In *Forest Research Review 1957-64*. Dept. Lands, Forestry Division, Dublin. p 69-75.
- Gallagher, G.J. 1969. The thinning of Sitka spruce - two experiments. *Irish Forestry* 26: 4-21.
- Gallagher, G.J. 1972. Pattern in crop structure and productivity for unthinned Sitka spruce. *Irish Forestry* 29: 33-53.
- Gallagher, G.J., Lynch, T.J. and Fitzsimons, B. 1987. Lodgepole pine in the Republic of Ireland II. Yield and management of coastal lodgepole pine. *Forest Ecology and Management* 22: 185-203.
- Hamilton, G.J. and J.M. Christie. 1971. *Forest Management Tables* (Metric). Forestry Commission Booklet. 34. HMSO. London. 201 p.
- Horgan, T., Keane, M., McCarthy, R., Lally, M., Thompson, D. 2004. *A Guide to Forest Tree Species Selection and Silviculture in Ireland*. Ed. O'Carroll, J.L., COFORD, Dublin.
- McCulloch, C. E., Searle, S.R. and Neuhaus, J.M. 2008. *Generalized, Linear and Mixed Models*. Hoboken, New Jersey, John Wiley and Sons, Inc., 2nd Ed.
- Monserud, R.A. and Sterba, H. 1996. A basal area increment model for individual trees growing in even- and uneven-aged forest stands in Austria. *Forest Ecology and Management* 80: 57-80.
- NFI. 2007. *Proceedings to Ireland's National Forest Inventory Conference*. Forest Service, Dublin.
- Ray, D., Xenakis, G., Tene, A. and Black, K. 2009. Developing a site classification system to assess the impact of climate change on species selection in Ireland. *Irish Forestry* 66: 101-122.
- Sarkar, D. 2008. *Multivariate Data Visualization with R*. Springer, New York.
- SAS Institute Inc. 2009. *SAS/STAT® 9.2 User's Guide*. Cary, NC: SAS Institute Inc.
- Temesgen, H. and von Gadow, K. 2004. Generalized height-diameter models: An application for major tree species in complex stands of interior British Columbia. *European Journal of Forest Research* 123: 45-51.
- Uzoh, F.C.C. and Oliver, W.O. 2006. Individual tree height increment model for managed even-aged stands of ponderosa pine throughout the western United States using linear mixed effects models. *Forest Ecology and Management* 221: 147-154.
- Wang, Y., Titus, S.J. and LeMay, V.M. 1998. Relationships between tree slenderness coefficients and tree or stand characteristics for major species in boreal mixedwood forests. *Canadian Journal of Forest Research* 28: 1171-1183.
- Wykoff, W.R., Crookston, N.L. and Stage, A.R. 1982. *User's Guide to the Stand Prognosis Model Ogden*, GTR INT-133, USDA Forest Service, 112 pp.