

Height-Diameter Models with Random Coefficients and Site Variables for Tree Species of Central Maine

Mike R. SAUNDERS^{1*}, Robert G. WAGNER²

¹ Department of Forestry and Natural Resources, Purdue University, 715 West State Street, West Lafayette, IN, USA

² School of Forest Resources, University of Maine, 5755 Nutting Hall, Orono, ME, USA

(Received 12 December 2006; accepted 23 October 2007)

Abstract – Height-diameter models with random coefficients and site variables for tree species of Central Maine. Height-diameter models were developed for nine tree species common to the northeastern United States: *Abies balsamea*, *Acer rubrum*, *Betula papyrifera*, *B. populifolia*, *Picea rubens*, *P. mariana*, *Pinus strobus*, *Populus tremuloides*, and *Tsuga canadensis*. Stem heights and diameters were collected from 6 146 trees (between 136 and 2 615 trees per species) on 50 plots within 10 structurally diverse stands that are part of a long-term silvicultural experiment in central Maine. The models were developed using both generalized nonlinear least squares (GNLS) and multi-level, mixed-effects approaches. Mixed-effects approaches were superior to GNLS, with inclusion of site covariates (tree density and basal area) accounting for some of the variability explained by the random coefficients in the full mixed-effect models. Analysis of plot-level parameter estimates suggested that differences in stand structure (even-aged vs. uneven-aged silvicultural practices) had a significant influence on the height-diameter relationships.

height-diameter model / mixed-effects modeling / site effects / stand structure / R

Résumé – Modèles hauteur–diamètre avec coefficients aléatoires et variables stationnelles pour des essences forestières du centre du Maine. Des modèles hauteur–diamètre ont été développés pour neuf essences courantes dans le Nord Est des USA : *Abies balsamea*, *Acer rubrum*, *Betula papyrifera*, *Betula populifera*, *Picea rubens*, *Pinus strobus*, *Populus tremuloides* et *Tsuga canadensis*. Le diamètre et la hauteur des troncs ont été recueillis sur 6 146 arbres (entre 136 et 2 615 arbres par espèce) dans 50 placeaux de 10 des peuplements de diverses structures qui font partie d'une expérimentation à long terme dans le centre du Maine. Les modèles ont été développés en utilisant à la fois les moindres carrés non linéaires généralisés (GNLS) et des approches multi-niveaux à effets mixtes. Les démarches à multi-niveaux ont été supérieures au GNLS, avec des covariables de station (densité et surface terrière) représentant une partie de la variabilité expliquée par les coefficients aléatoires dans l'ensemble des modèles à effets mixtes. L'analyse des estimations des paramètres au niveau des placeaux suggère que les différences de structure de peuplement (pratiques sylvicoles en peuplement équennes ou inéquennes) ont un effet significatif sur les relations hauteur –diamètre.

modèle hauteur–diamètre / modélisation des effets mixtes / effets de station / structure des peuplements

1. INTRODUCTION

The allometric relationship between tree diameter and total tree height is commonly used to estimate tree volume, and thus is a fundamental component of many growth and yield, process, and forest planning models [23, 28]. This relationship is highly site-dependent and not constant over time, even in the same stand [8]. Height-diameter models are often specific to localities, site fertility classes, and/or structural stand types (e.g., even-aged plantations), or include numerous site and regional adjustments to the base parameters [11, 16]. These allometric relationships also change with silvicultural practices, such as thinning, that alter competitive relationships among neighboring trees [9, 13]. Application of regional models that do not adequately consider the effect of silvicultural practices on stem height and diameter can lead to inaccurate and imprecise estimates of tree volume growth and value, particularly

when those models do not include variables that account for the dynamics of these stand relationships [6, 42].

Earlier modeling efforts focused on describing the mean parameter values of the height-diameter relationship and developing functional model forms that performed well and were biologically meaningful [8, 12, 17, 32, 43]. Recently, focus has shifted towards understanding the variability of parameter estimates over the spectrum of stand conditions. Although other techniques have been used, many recent modeling efforts have used mixed-effects models that simultaneously include both fixed coefficients to account for population-wide average response and random coefficients that account for variability in the response of a particular sampling unit (e.g., study, stand, or plot) [4, 6, 34].

Mixed-effects models offer several advantages over ordinary linear (OLS), generalized nonlinear least squares (GNLS), and other approaches. First and foremost, mixed-effects models can incorporate the hierarchical structure of data into the analysis and thus reduce interdependence among measurements from the same sampling unit

* Corresponding author: msaunder@purdue.edu

by defining a covariance matrix among random parameters within and among sampling levels [4, 6, 10]. Second, mixed-effects models are a compromise between fitting, with either OLS or GNLS, global models with few parameters and that do not include variability among sampling units (i.e., stand or plot), and local models specific to each sample unit that have numerous and often intercorrelated parameters [34]. Third, because the variation in the parameter estimates is known at each level of the hierarchical sampling structure, mixed-effects models provide an unbiased estimation of model parameters for sample units with small sample sizes. Lastly, mixed-effects models can be calibrated for new, previously unsampled plots or stands quickly and effectively [20, 25, 27], particularly if values of the random parameters can be predicted from covariates. Some height-diameter mixed-effects models can be calibrated with as few as from 4–10 tree heights per sample unit with minimal introduction of bias [4, 6, 25]. An OLS or GNLS model would require a much larger sample of tree heights and all related covariate parameters in order to create a new localized equation for these same plots or stands [25, 42].

We developed height-diameter models for nine tree species found throughout northern New England using both GNLS and mixed-effects modeling approaches. The trees were selected from a long-term experiment where different silvicultural methods had been used to create stands with highly varied structural conditions. Inventories in this experiment did not consistently measure tree height; therefore analyses of changes in forest structure had been largely diameter-based [41]. Equations were developed to estimate height in the most common species: balsam fir (*Abies balsamea* (L.) Mill.), red (*Picea rubens* Sarg.) and black spruce (*P. mariana* (Mill.) B.S.P.), eastern hemlock (*Tsuga canadensis* (L.) Carr.), eastern white pine (*Pinus strobus* L.), red maple (*Acer rubrum* L.), gray (*Betula populifolia* Marsh.) and paper birch (*B. papyrifera* Marsh.), and quaking aspen (*Populus tremuloides* Michx.). We hypothesized that: (1) models fit with random coefficients would dramatically outperform those fit only with fixed coefficients; (2) height-diameter relationships would differ by structural condition, and (3) additional plot-level covariants, specifically tree density or plot basal area, would further improve model fits and simplify model structure by accounting for variability captured by random coefficients in the original model.

2. METHODS

2.1. Study Area and Field Measurements

The data used in this study came from the Penobscot Experimental Forest near the town of Bradley, Maine ($44^{\circ} 52' N, 68^{\circ} 38' W$). This 1 550 ha area lies on soil types derived from glacial till and ranging from well-drained loams and sandy loams on glacial till ridges to poorly and very poorly drained loams and silt loams in flat areas between the ridges [41]. Cover types are dominated by Acadian Region softwoods including red, white (*P. glauca* (Moench) Voss) and black spruce, balsam fir, eastern white pine, eastern hemlock, and northern white cedar (*Thuja occidentalis* L.). Common hardwoods in

these types include red maple, paper and gray birch, and quaking and bigtooth aspen (*P. grandidentata* Michx.). Natural stand structures in this region are typically uneven-aged and diverse with windstorms and insect epidemics as the major disturbance events. Stand replacing fires are thought to occur less than once per 1 000 years in these forest types [24].

Tree height and stem diameter data were obtained from measurements on a subset of long-term sample plots established in 1952 by the USDA Forest Service (USFS) to monitor the effects of five silvicultural and three exploitative harvest systems on stand growth, yield, and structure within the Acadian Forest region. Beginning in June 2001 and continuing through August 2002, 7938 trees were measured and stem-mapped on 50 sample plots – 5 plots in each of the two replicates within the unharvested natural area control (NA), the unregulated commercial clearcut (CC), the fixed diameter limit (DL), the 5-year selection (5S), and the 3-stage shelterwood (SW), one replicate with and one without precommercial spacing, treatments. Generally, the commercial clearcut, with both replicates last harvested in the 1980s, and the fixed diameter limit, with one replicate harvested in 1993 and the second in early 2001, were exploitative in nature and created an irregularly-aged structure with many cull individuals. Replicates for both the unharvested natural area control and the 5-year selection are uneven-aged with multiple canopy strata. Lastly, the 3-stage shelterwood is quite even-aged and had a final overstory removal in 1975; this stand was split into the two current replicates in 1982 for a test of precommercial thinning on early stand development. For more detailed marking prescriptions, harvesting techniques and timings for each treatment see Sendak et al. [41] and Saunders and Wagner [40].

Plots were nested with all trees > 11.45 cm diameter at breast height (DBH) measured within a 0.081 ha plot and trees > 1.27 cm DBH measured on a smaller, interior 0.020 ha plot. Heights were measured to the nearest 0.1 m either directly, using 10 and 15 m telescoping height poles, or as an average of 2–4 readings from a Haglöf hypsometer [14].

With the full dataset, basal area and tree density were calculated for each plot (Fig. 1). The dataset was then trimmed of all cull, dying and leaning trees. Of the 26 species recorded, only 9 species—balsam fir, red maple, paper and gray birch, red and black spruce, white pine, trembling aspen, and eastern hemlock—occurred in high numbers ($n > 100$) and across a majority of the stands and sample plots. Red and black spruce hybridize extensively within the PEF and were grouped together for this analysis. Summary statistics for each species are given in Table I.

2.2. Statistical Analysis

Several nonlinear functional forms have been used to model height-diameter relationships [12, 17, 32]. Although other mixed-modeling efforts have used functional forms that could be linearized and/or exponential [4, 27, 29], sigmoid models are more biologically appropriate for height-diameter relationships [23]. In a preliminary study using weighted OLS, fits of the data with a 3-parameter form of the Chapman-Richards function [7, 33, 39] were slightly better and converged more often than 3-parameter Weibull [44], modified logistic [38], or exponential [37] functions. In this study, the Chapman-Richards function was parameterized as (hereafter referred to as Model I):

$$HT = 1.35 + a_0 \left[1 - e^{(b_0 \cdot DBH)} \right]^{c_0} + \varepsilon \quad (1)$$

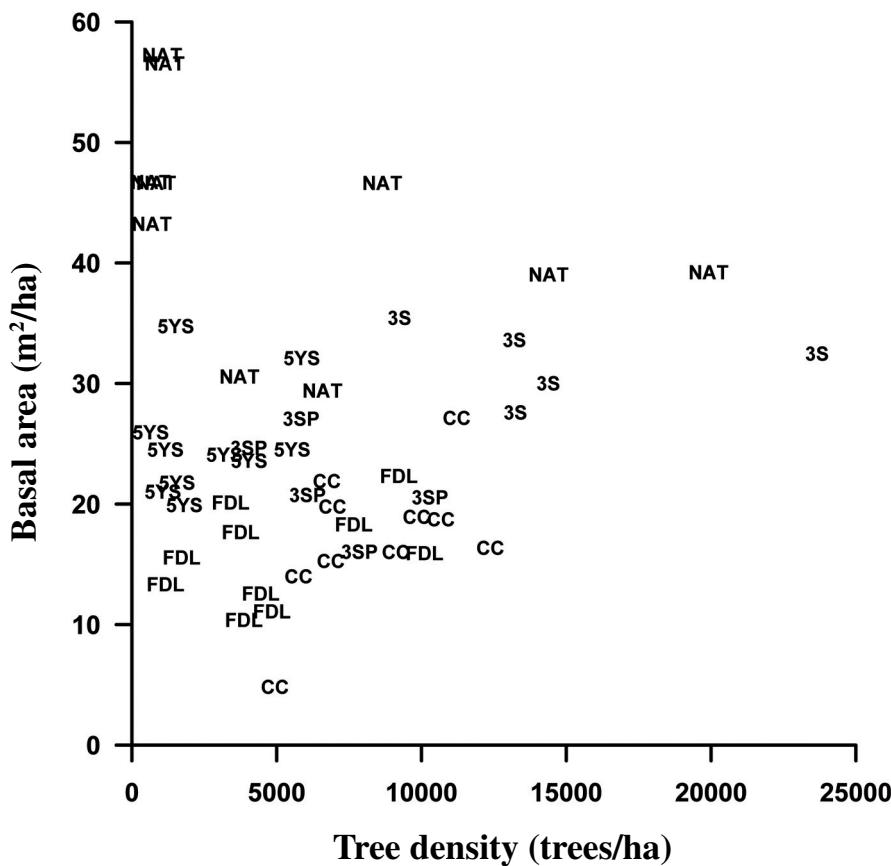


Figure 1. Distribution of tree density and basal area within sample plots as separated by silvicultural treatment (NA = unharvested natural area control, CC = unregulated commercial clearcut, DL = fixed diameter limit, 5S = 5-year selection system, and SW = 3-stage shelterwood).

Table I. Summary statistics of the diameter at breast height (DBH) and height of trees (HT) by species. Total sample size (N) and the number of stands and plots for each species is also given.

Species	N	Stands	Plots	DBH (cm)			HT (m)		
				Mean	S.D.	Range	Mean	S.D.	Range
Balsam fir	2615	10	47	5.7	4.9	1.3–29.2	5.35	3.34	1.65–20.25
Red & black spruce	1415	10	43	7.1	6.9	1.3–47.2	6.05	4.00	1.70–27.00
Eastern hemlock	831	10	41	12.0	12.5	1.3–67.8	8.55	6.46	1.98–27.90
Eastern white pine	181	10	26	21.8	19.2	1.4–81.5	13.32	9.44	2.26–33.73
Red maple	547	10	45	9.1	8.2	1.3–43.5	9.50	5.42	2.01–24.65
Gray birch	238	6	21	4.4	2.3	1.3–13.0	6.82	2.31	2.06–12.22
Paper birch	183	9	29	5.8	6.0	1.3–30.1	6.91	4.25	2.37–23.70
Quaking aspen	136	6	15	9.3	8.8	1.3–45.4	10.16	4.38	2.80–24.50

where HT is tree height (m), DBH is tree diameter at breast height (cm), a_0 , b_0 , and c_0 are estimated, fixed (population-wide) parameters, and ε are distributed with variance σ^2 as $\varepsilon \sim N(0, \sigma^2)$. In a multi-level, mixed-effects model, random coefficients can be assigned to each parameter at each level (stand and plot in this study), and have a complex variance-covariance structure that varies among parameters and levels. However, this complexity can lead to nonidentifiability and ill-conditioning, with convergence becoming computationally intensive and difficult [15]. Trial analyses with random coefficients at both stand and plot levels for all three fixed parameters only converged with large sample sizes ($n > 1000$). Therefore, we investigated the variation within and correlation among parameters by

fitting Model I to each plot within the study. Regardless of species, the variation about a_0 and c_0 was markedly higher, often by a magnitude or more, than for b_0 at the plot level. Furthermore, parameter estimates for b_0 and c_0 were nearly always highly correlated ($\rho = 0.763 - 0.989$); the single exception occurred with spruce ($\rho = 0.357$) where the b_0 parameter was commonly nonsignificant at the plot level. Therefore based on this evidence, we simplified the full random model to Model II:

$$HT = 1.35 + (a_0 + \nu_S + \nu_P) [1 - e^{(b_0 \cdot DBH)}]^{(c_0 + \omega_S + \omega_P)} + \varepsilon \quad (2)$$

where ν_S and ν_P , and ω_S and ω_P are random coefficients at the stand (S) and plot (P) levels for a_0 and c_0 , respectively, and b_0 was not

allowed to vary randomly. The variance-covariance structures were positive-definite at both the stand (Ψ_S) and plot (Ψ_P) levels, and specified as:

$$\Psi_S = \begin{pmatrix} \sigma_{ys}^2 & \sigma_{y\varpi_S}^2 \\ \sigma_{y\varpi_S}^2 & \sigma_{\varpi_S}^2 \end{pmatrix} \quad \text{and} \quad \Psi_P = \begin{pmatrix} \sigma_{yp}^2 & \sigma_{y\varpi_P}^2 \\ \sigma_{y\varpi_P}^2 & \sigma_{\varpi_P}^2 \end{pmatrix},$$

and distributed bivariate normally with normal random errors:

$$\begin{pmatrix} v_S \\ \varpi_S \end{pmatrix} \sim N(0, \Psi_S), \quad \begin{pmatrix} v_P \\ \varpi_P \end{pmatrix} \sim N(0, \Psi_P), \quad \varepsilon \sim N(0, \sigma^2). \quad (3)$$

The next phase in developing a mixed-effects model is determining whether covariates can help explain or potentially eliminate random coefficients from a model [29, 34]. We tested the inclusion of the natural logarithm of plot-level tree density ($\ln[\text{trees/ha}]$; Model III) and total basal area (m^2/ha ; Model IV) as fixed parameters to help explain variability. These two covariates were chosen for two reasons. First, graphs of the random coefficients in Model II against these covariates suggested that these two covariates could explain some of the random variability at either the stand or plot level for several species. Second, these covariates could be reliably calculated from the historical records for the USFS study and have been found to affect the height-diameter relationship [11, 42, 47]. Other authors have suggested dominant height and/or some index of the dispersion of the diameter or height distribution as potential covariates [4, 12], but it is not clear if these covariates have been applied when the data source is obtained from a mix of normally distributed, even-aged and non-normally distributed, two- or uneven-aged stands. The original fixed parameters a_0 , b_0 and c_0 were assumed to depend linearly on the covariates (COV). The model form was:

$$HT = 1.35 + (a_0 + a_1 \cdot COV + v_S + v_P) \\ [1 - e^{-(b_0 + b_1 \cdot COV) \cdot DBH}]^{(c_0 + c_1 \cdot COV + \omega_S + \omega_P)} + \varepsilon \quad (4)$$

where a_1 , b_1 and c_1 are estimated parameters for the covariate, and with a variance-covariance structure identical to Model II (Eq. (3)). We tested the inclusion of an interaction term between tree density and basal area, but the interaction was not significant ($p > 0.05$) for many of the species and often caused the model to not converge, generally due to the high correlation between tree density and basal area (Fig. 1).

Heteroscedasticity was observed in the plot-level residuals for all models during preliminary runs. Therefore, Model I was fit with generalized nonlinear least-squares (the *gnls* function) and Models II–IV were fit with pseudo-likelihood approach for nonlinear mixed-effect models (the *nlme* function) within the *nlme* package for the R programming language [35, 36]. We used the *varPower* function to weight the variance of the residuals by a power of the diameter or:

$$Var(\varepsilon_{ijk}) = \sigma^2 |DBH_{ijk}|^\Delta \quad (5)$$

where DBH_{ijk} are the i tree diameters in plot j of stand k , and Δ is the power of the variance covariate [34].

Models were simplified by removing terms stepwise and assessing significance of the parameter (fixed or random) by log-likelihood ratio tests (*LRT*) between the original and reduced models or:

$$LRT = 2(\log L_F - \log L_R) \quad (6)$$

where L_F is the maximum likelihood of the full model, L_R is the maximum likelihood of the reduced model, and *LRT* is tested on $\chi^2_{df_F - df_R}$,

where df_F and df_R are the degrees of freedom associated with the full and reduced model, respectively. Significant but ill-conditioned models with $\sigma_{y\varpi_S}^2$ or $\sigma_{y\varpi_P}^2 \approx 1$ were also discarded [34]; this commonly occurred with Model III and IV because the covariates were estimated at the plot- and not tree-level. Model fit was assessed with the coefficient of determination (R^2), the mean difference (MD) and the mean absolute difference (MAD) which were calculated, respectively, as:

$$R^2 = 1 - \frac{\sum_{i=1}^n (HT_i - HT_i^*)^2}{\sum_{i=1}^n (HT_i - \bar{HT})^2} \\ MD = \frac{\sum_{i=1}^n (HT_i - HT_i^*)}{n} \\ MAD = \frac{\sum_{i=1}^n |HT_i - HT_i^*|}{n} \quad (7)$$

where HT_i^* is the predicted height of the i th tree, HT_i is the observed height of the i th tree, and \bar{HT} is the mean height of all trees of that species [32]. MD values far from 0 would indicate a consistent bias with the model, whereas large MAD values would indicate bias in the model through a restricted range of the dependent variable (i.e., diameter).

Beyond these assessments, validation of these models was not performed. Although we could have split the data for model validation, we deemed that this was unnecessary given that datasets were already quite small for a mixed modeling approach for several species. Kozak and Kozak [19] have suggested that model validation methods, such as cross validation by data splitting and double cross validation, provide little additional information beyond the general fit statistics; in fact, Huang et al. [18] argue that these approaches validate the sampling techniques and not the model itself. Yang et al. [46] show that the numerous diagnostic tests used for model validation have limited use. All agree that only independently collected data will help. Unfortunately, we did not have a compatible, independent dataset available to us, and must wait until new inventory data is available that includes height information from other unsampled stands in this region [5].

3. RESULTS

Without exception, the mixed-effects modeling approach (Model II) outperformed a generalized least squares approach (Model I; Tab. II). Inclusion of random coefficients in the height-diameter model always lead to increases in both *LL* and R^2 values, decreases MAD , and comparable values for MD . Further, *LRT* tests showed without exception ($p < 0.0001$) that Model II was more appropriate than Model I for these data. An example of the difference in fits between the two approaches is shown for red and black spruce in Figure 2. Visually, Model II did a better job at fitting the data specific to each stand, even when there were few data points (e.g., the commercial clearcut stands in Fig. 2). Model II also captured a wider range of behavior in the height-diameter relationship than Model I. Model II, however, sometimes lead to unrealistically parameter values particularly if the stand structure did not include a wide variation in tree heights and diameters. For example, the lack of large diameter spruce trees in the shelterwood stands lead to an estimated asymptote that

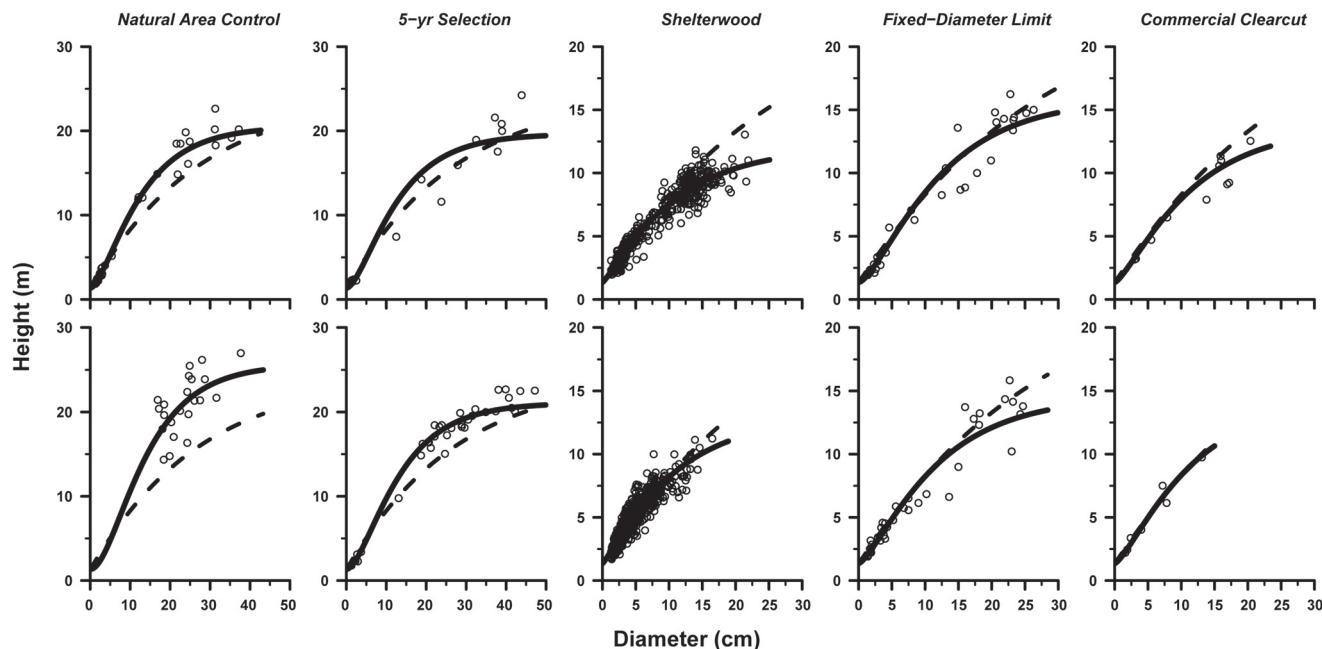


Figure 2. Mixed-effects models (solid line) and a population-wide general least squares model (dashed line) for the relationship between height and diameter at breast height for red and black spruce as measured across the two replicates (top and bottom rows) for each of the five treatments. These fits include only the stand-level random coefficients within Model II.

would unrealistically low for predicting future conditions in this structure (Tab. II, Fig. 2).

Both stand- and plot-level random coefficients were significant ($p < 0.05$) for many of the tree species, although the random effects associated within the “shape” of the height-diameter relationship (i.e., the c_0 parameter and associated σ_{ω_s} and σ_{ω_p} variance components) were less commonly found to be significant, particularly for hardwood species. Correlations between variance components (i.e., $\sigma_{v\omega_s}^2$ or $\sigma_{v\omega_p}^2$) were commonly approaching 1, reflecting the inherent correlations between parameters in a Chapman-Richman model, and, with one exception, only significant at the stand level and in species with large sample sizes. Not surprisingly, the variance components associated with the stand-level random effects (σ_{v_s} and σ_{ω_s}) were several times larger than those associated with the plot-level random effects (σ_{v_p} and σ_{ω_p}). This difference suggested that broad differences in structure and competitive conditions among the stands may have been driving the height-diameter relationships for most species, rather than intra-stand variation in environmental conditions. For example, a graph of the estimated plot-level coefficients of Model II suggested that plots with uneven-aged structures (i.e., NA and 5S) occupied different areas of parameter space than plots with even-aged (i.e., SW) or highly irregular (i.e., CC and DL) structures (Fig. 3). For conifers and red maple, the height models both had higher asymptotes (the a parameter) and more sigmoidal shapes (the c parameter) in uneven-aged structures than other structures. Eastern white pine, in particular, showed a strong difference in the height-diameter relationships between uneven-aged structures and the shelterwood treatment.

However, there was no discernable pattern in the coefficients for early-successional hardwoods (i.e., aspen and birch).

The addition of tree density (Model III) and basal area (Model IV) plot-level covariates to the mixed-model generally improved model fits for all species (Tab. II). However, improvements were not dramatic and the additional covariate generally accounted for only part of the variation captured by random coefficients in Model II. Model IV was superior to Model III for only three species—red maple, paper birch and quaking aspen. With some obvious exceptions, increasing plot-level tree density within Model III negatively adjusted the inflated base parameters (a_0 , b_0 , and c_0) for most covariate parameters (a_1 , b_1 , and c_1) that were significant (Tab. II). Basal area effects were not as prevalent or consistent in Model IV, although for five of the species a_1 was a positive adjustment to a reduced estimated asymptote (a_0) from Model II. Lastly, both Models III and IV were sensitive to the range of the covariate, particularly for species with smaller sample sizes. For example, Model IV for quaking aspen gave realistic heights only within a diameters of 2–30 cm and a basal areas of 20–40 m²/ha.

4. DISCUSSION

Silviculture has strong influences on the height-diameter relationship, primarily through controlling the allocation of light and resources to individual growing spaces. For example, we saw strong differences between even-aged and uneven-aged treatments in growth of most species, but most notably with red/black spruce and white pine. For white pine, we suggest

Table II. Weighting power (Δ), model parameters, variance components for the random effects, root mean squared error (RMSE), and fit statistics for the four models described in the text. All parameters are significant at $\alpha = 0.05$ unless noted by “n.s.” Fit statistics include the log-likelihood (LL), the coefficient of determination (R^2), the mean difference (MD), and the mean absolute difference (MAD).

Species	Model	Δ	Fixed effects					
			a_0	a_1	b_0	b_1	c_0	c_1
Balsam fir	I	1.13	14.05	—	0.105	—	1.438	—
	II	1.02	13.53	—	0.115	—	1.570	—
	III	1.02	25.76	-1.45	0.115	n.s.	3.187	-0.190
	IV	1.04	9.96	0.13	0.114	n.s.	1.522	n.s.
Red and black spruce	I	1.23	23.13	—	0.038	—	1.038	—
	II	0.88	15.72	—	0.100	—	1.551	—
	III	0.89	35.00	-2.25	-0.093	0.022	1.480	n.s.
	IV	0.88	16.35	n.s.	0.042	0.002	1.077	0.015
Eastern hemlock	I	1.26	30.11	—	0.025	—	0.941	—
	II	1.04	30.18	—	0.018	—	0.844	—
	III	1.02	50.18	-3.81	-0.032	0.009	0.935	n.s.
	IV	1.04	23.75	0.31	0.017	n.s.	0.852	n.s.
Eastern white pine*	I	-**	39.97	—	0.026	—	1.312	—
	II	0.98	21.56	—	0.030	—	0.741	—
	III	0.93	21.50	n.s.	0.104	-0.008	0.819	n.s.
Red maple	I	0.92	24.16	—	0.049	—	0.935	—
	II	0.61	16.60	—	0.103	—	1.093	—
	III	0.60	27.37	-1.30	0.106	n.s.	1.097	n.s.
	IV	0.61	10.02	0.26	0.136	-0.001	1.057	n.s.
Gray birch	I	0.66	10.24	—	0.268	—	1.487	—
	II	0.68	9.41	—	0.257	—	1.377	—
	III	0.68	32.70	-2.54	-1.042	0.143	1.376	n.s.
	IV	0.66	12.43	-0.16	0.124	0.008	1.419	n.s.
Paper birch	I	1.03	23.21	—	0.049	—	0.917	—
	II	0.63	15.18	—	0.121	—	1.040	—
	III	0.68	15.18	n.s.	0.125	n.s.	2.760	-0.190
	IV	0.63	9.76	0.19	0.125	n.s.	1.178	-0.007
Quaking aspen	I	0.65	19.87	—	0.079	—	0.994	—
	II	0.66	18.03	—	0.076	—	0.957	—
	III	0.73	14.77	n.s.	1.142	-0.109	8.191	-0.752
	IV	0.65	6.46	0.31	0.394	-0.009	2.441	-0.044

* Model IV was not significant.

** Model would not converge with variance weighting by *varPower*.

that this difference may have been driven by slower height development from repeated attacks by white-pine weevil in the more open shelterwood (*Pissodes strobi* Peck.) [45]. For the other species, one may preclude that these are entirely site differences. However, all stands in this study were observed to have roughly the same overstory structure and age class distribution at the study’s onset in the 1950s [40, 41]. Further, attempts to include a site quality covariate derived from soil drainage classes using a soil type map of the PEF [2, 3] within models failed, as the covariate was not significant ($p > 0.05$) for any species. Further quantification of silvicultural effects on height-diameter relationships was difficult because of the relatively few stands used and confounding between tree sizes and stand structures.

Although mixed models are more complex than GNLS models, their increased precision would significantly increase the performance of growth and yield models dependent on these height-diameter relationships [4, 25]. GNLS models

were not biased, but had very high *MADs* suggesting that their accuracy often decreased when predicting interior sampling levels (e.g., stand or plot). For example, Model I consistently and dramatically underestimated spruce heights (Fig. 2) and white pine (not shown) in both natural area controls. GNLS models would only avoid this if specifically parameterized to include covariates that accounted for variation at each level. Mixed-effects model can avoid this behavior by incorporating stand- and plot-level variability into the model [15, 21] and then optionally using covariates to account for variability captured by the random effects [34].

We suggest that a mixed modeling approach is attractive when stand- and plot-level covariates are not always obvious [29]. Both tree density and basal area are plot-level covariates that can influence the height-diameter relationship [9, 31, 47]. For example, our study generally agreed with that of Ek et al. [11] who found that total tree height increased with basal area (Tab. II). However, other potential covariates were

Table II. Extended.

σ_{v_S}	σ_{ϖ_S}	Random effects			RMSE	Fit statistics				
		$\sigma_{v_{\omega_S}}$	σ_{v_P}	σ_{ϖ_P}		LL	R^2	MD	MAD	
—	—	—	—	—	0.9462	-2634	0.920	-0.0019	0.6363	
2.307	0.184	0.924	1.140	0.132	0.610	0.7280	-2262	0.953	0.0040	0.4979
1.250	n.s.	n.s.	0.809	0.106	n.s.	0.7289	-2252	0.952	0.0042	0.5011
0.898	0.954	n.s.	0.106	0.108	n.s.	0.7388	-2266	0.951	0.0046	0.5016
—	—	—	—	—	—	1.3726	-1731	0.882	-0.0045	0.8300
4.267	0.296	0.962	0.744	0.053	n.s.	0.7889	-1415	0.961	0.0025	0.5415
3.921	0.211	n.s.	0.585	0.061	n.s.	0.7859	-1406	0.961	0.0030	0.5384
3.263	0.220	n.s.	0.412	0.049	n.s.	0.8025	-1414	0.960	0.0009	0.5468
—	—	—	—	—	—	1.9000	-1239	0.913	0.0011	1.2430
6.487	0.080	0.913	n.s.	0.037	n.s.	1.3509	-1127	0.956	0.0005	0.8839
3.501	0.105	0.848	1.389	n.s.	n.s.	1.3045	-1115	0.959	0.0300	0.8585
2.142	0.036	n.s.	n.s.	0.028	n.s.	1.3597	-1121	0.956	0.0019	0.8977
—	—	—	—	—	—	2.8127	-444	0.911	0.2319	2.2169
9.194	0.285	0.844	0.534	n.s.	n.s.	1.3686	-295	0.979	0.0332	0.9150
8.483	0.320	0.922	0.337	n.s.	n.s.	1.3203	-289	0.980	0.0253	0.9094
—	—	—	—	—	—	1.6298	-910	0.910	-0.0062	1.1500
3.108	0.153	0.953	1.223	n.s.	n.s.	1.0657	-819	0.961	0.0088	0.8226
2.514	0.154	0.971	1.166	n.s.	n.s.	1.0621	-815	0.962	0.0083	0.8193
1.543	0.072	n.s.	1.176	n.s.	n.s.	1.0615	-812	0.962	0.0084	0.8202
—	—	—	—	—	—	0.9060	-303	0.846	-0.0009	0.6917
1.242	0.229	n.s.	0.477	0.067	n.s.	0.7227	-279	0.902	0.0043	0.5431
0.806	0.092	n.s.	0.436	n.s.	n.s.	0.7286	-271	0.900	0.0053	0.5452
0.500	n.s.	n.s.	0.537	n.s.	n.s.	0.7502	-275	0.894	0.0055	0.5628
—	—	—	—	—	—	1.1947	-227	0.921	-0.0081	0.7947
2.996	0.075	n.s.	1.202	0.041	n.s.	0.5435	-189	0.984	0.0074	0.3987
3.085	0.083	n.s.	0.992	n.s.	n.s.	0.5621	-186	0.982	0.0097	0.4152
1.899	n.s.	n.s.	1.020	n.s.	n.s.	0.5636	-180	0.982	0.0075	0.4144
—	—	—	—	—	—	1.4895	-237	0.884	-0.0044	1.1381
2.569	0.116	n.s.	n.s.	n.s.	n.s.	1.0795	-210	0.939	0.0052	0.8419
2.971	n.s.	n.s.	n.s.	n.s.	n.s.	1.1084	-209	0.936	0.0111	0.8628
2.750	n.s.	n.s.	n.s.	n.s.	n.s.	1.0502	-204	0.942	0.0068	0.8120

not apparent, particularly since the dataset was collected from mixed-species stands that included a wide array of structural conditions and site qualities. For example, dominant height, although obviously correlated with site quality by Eichorn's rule [1], could not be used as a covariate in this study because canopy position is a function of both growth potential and a cumulated history of suppression by older cohorts in multi-aged stands [30]. For these same reasons, direct estimates of site index could not be determined from these multi-aged stands.

This study also highlighted the more intensive data requirements for mixed-effects approaches to model height-diameter relationships. First, multi-level, mixed-effects models with more than two random parameters and/or with complex variance-covariance structures may have difficulty converging unless sample sizes are large and well dispersed among all sampling units and sampling levels. Most published mixed-effects models have used datasets with >1 000 trees and at least 10–20 trees within the intermost sampling units [4]. Longitudinal models are commonly developed using fewer trees per sampling unit, but the multiple measurements per tree in order to increase the sample size [21, 22]. In this study, only balsam fir occurred in sufficient numbers and was

distributed across enough sampling units (Tab. I) for the more complex version of Model II, with all three parameters as random and an unrestricted variance-covariance matrix, to converge. In all other cases, correlation at the plot level could not be estimated reliably.

Second, a mixed-effects modeling approach can be quite sensitive to the range of tree sizes used to develop the model. This same concern is apparent when modeling with OLS or GNLS approaches, but for the mixed-effects models, the concern can be considerable since the range of tree sizes in each sampling unit can influence the variance components associated with each random effect. In this study, young even-aged stands had substantially different parameter estimates than uneven-aged stands (Figs. 2 and 3), suggesting that some type of structural covariate should be included as a fixed effect in more extensive modeling efforts. The choice of a different functional model form (i.e., exponential as opposed to sigmoidal) also may help reduce this sensitivity [26].

Although the height-diameter models developed in this study were intended primarily for descriptive purposes, they can be easily calibrated for other stands using the methods of Lappi [20]. As suggested earlier, calibration data for

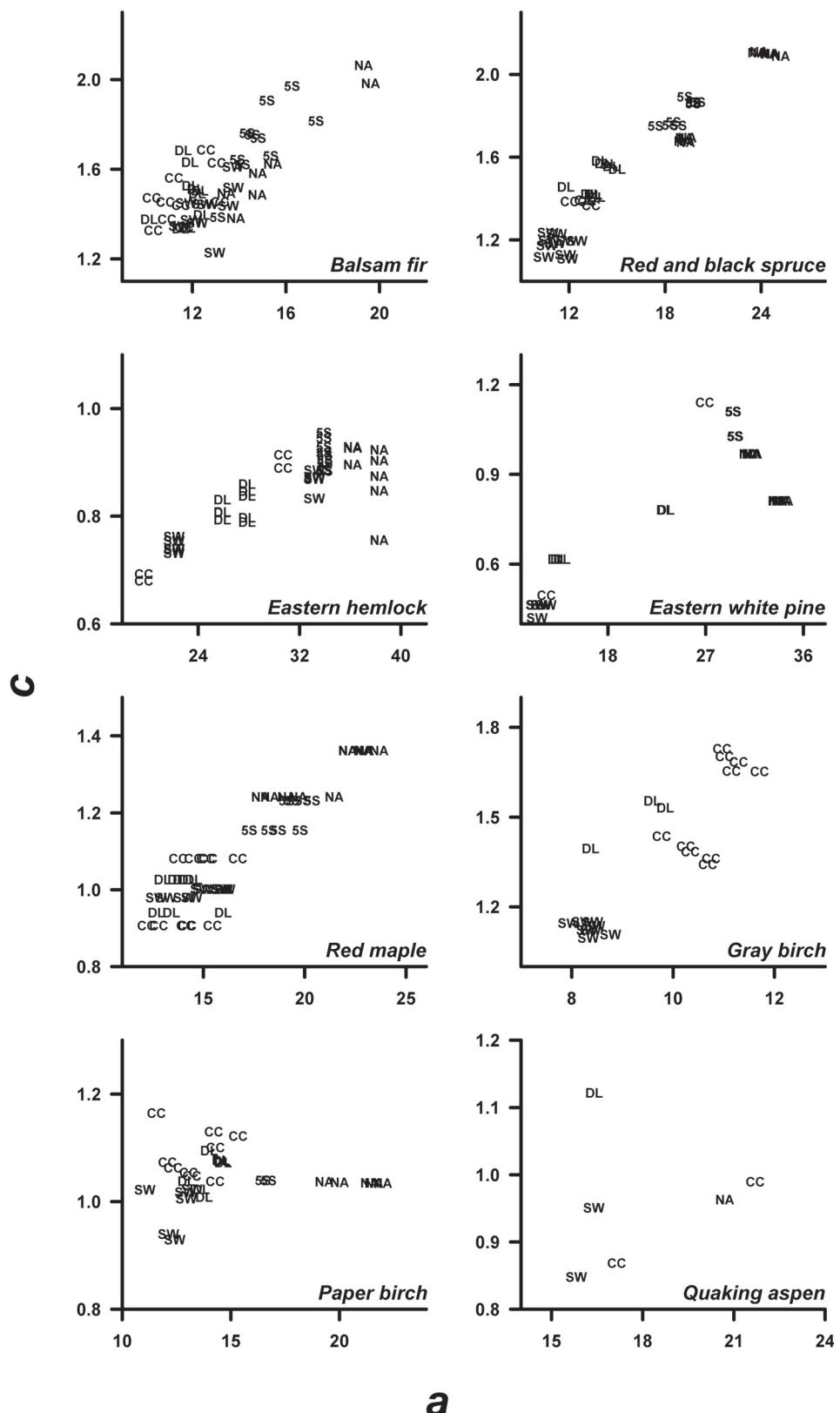


Figure 3. Comparison of plot-level estimates of a and c parameters from Model II as they vary by silvicultural treatment (CC = unregulated commercial clearcut, DL = fixed diameter limit, 5S = 5-year selection system, SW = 3-stage shelterwood, and NA = unharvested natural area control) for nine tree species.

these models should include a wider range of diameters and larger numbers of trees than that recommended by Calama and Montero [4] or Castedo Dorado et al. [6]. The broader range of stand ages and structural classes used in this study should make these models more widely applicable than many other height-diameter models developed for northeastern tree species. However, since these models have not been validated with an independent data set, we suggest that models should be applied only after calibration.

Acknowledgements: We would like to thank Robert Seymour, John Brissette, Malcolm Hunter Jr., Jeremy Wilson and two anonymous individuals for reviewing early versions of this manuscript. We would also like to thank the US Forest Service at the Penobscot Experimental Forest for use of field sites and access to their long-term datasets. This work was supported by the Maine Agricultural and Forest Research Station at the University of Maine (MAFES #2926) and USDA National Research Initiative (Project # ME0-2000-0700).

REFERENCES

- [1] Assmann E., The principles of forest yield study, Pergamon Press, Oxford, 1970.
- [2] Briggs R.S., Site classification field guide, Maine Agri. Forest. Exp. Station Misc. Publ. 724, 1994.
- [3] Briggs R.S., Lemkin R.C. Jr., Soil drainage class effects on early response of balsam fir to precommercial thinning, *Soil Sci. Soc. Am. J.* 58 (1994) 1231–1239.
- [4] Calama R., Montero G., Interregional nonlinear height-diameter model with random coefficients for stone pine in Spain, *Can. J. For. Res.* 34 (2004) 150–163.
- [5] Castedo Dorado F., Barrio Anta M., Parresol B.R., Álvarez González J.G., A stochastic height-diameter model for maritime pine ecoregions in Galicia (northwestern Spain), *Ann. For. Sci.* 62 (2005) 455–465.
- [6] Castedo Dorado F., Diéguez-Aranda U., Barrio Anta M., Sánchez Rodríguez M., von Gadow K., A generalized height-diameter model including random components for radiata pine plantations in northwestern Spain, *For. Ecol. Manage.* 229 (2006) 202–213.
- [7] Chapman D.G., Statistical problems in population dynamics, Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability, Univ. Calif. Press, Berkeley and Los Angeles, 1961, pp. 153–168.
- [8] Curtis R.O., Height-diameter and height-diameter-age equations for second-growth Douglas-fir, *For. Sci.* 13 (1967) 365–375.
- [9] Daggett H., Long-term effects of herbicide and precommercial thinning treatments on species composition stand structure, and net present value in spruce-fir stands in Maine: the Austin Pond study, M.S. Thesis, University of Maine, 2003.
- [10] Demidenko E., Mixed models: Theory and applications, John Wiley & Sons, Inc., Hoboken, New Jersey, 2004.
- [11] Ek A.R., Birdsall E.T., Spears R.J., A simple model for estimating total and merchantable tree heights, USDA Forest Service, North Central Forest Experiment Station, St. Paul, Minnesota, Res. Note NC-309, 1984.
- [12] Fang Z., Bailey R.L., Height-diameter models for tropical forests on Hainan Island in southern China, *For. Ecol. Manage.* 110 (1998) 315–327.
- [13] Fang Z., Bailey R.L., Shiver B.D., A multivariate simultaneous prediction system for stand growth and yield with fixed and random effects, *For. Sci.* 47 (2001) 550–562.
- [14] Haglöf AB, Users guide Vertex III and Transponder T3, Haglöf Sweden, AB, Långsele, Sweden, 2002.
- [15] Hall D.B., Bailey R.L., Modeling and prediction of forest growth variables based on multilevel nonlinear mixed models, *For. Sci.* 47 (2001) 311–321.
- [16] Huang S., Price D., Titus S.J., Development of ecoregion-based height ± diameter models for white spruce in boreal forests, *For. Ecol. Manage.* 129 (2000) 125–141.
- [17] Huang S., Titus S.J., Wiens D.P., Comparison of nonlinear height-diameter functions for major Alberta tree species, *Can. J. For. Res.* 22 (1992) 1297–1304.
- [18] Huang S., Yang Y., Wang Y., A critical look at procedures for validating growth and yield models, in: Amaro A., Reed D., Soares P., (Eds.), Modelling forest systems, CABI International, Cambridge, Massachusetts, pp. 271–293.
- [19] Kozak A., Kozak R., Does cross validation provide additional information in the evaluation of regression models? *Can. J. For. Res.* 33 (2003) 976–987.
- [20] Lappi J., Calibration of height and volume equations with random parameters, *For. Sci.* 37 (1991) 781–801.
- [21] Lappi J., Bailey R.I., Height prediction model with random stand and tree parameters: An alternative to traditional site index methods, *For. Sci.* 34 (1988) 907–927.
- [22] Lappi J., Malinen J., Random-parameter height/age models when stand parameters and stand age are correlated, *For. Sci.* 40 (1994) 715–731.
- [23] Lei Y., Parresol B.R., Remarks on height-diameter modeling, USDA Forest Service, Southern Research Station, Asheville, North Carolina, Res. Note SRS-10, 2001.
- [24] Lorimer C.G., The presettlement forest and natural disturbance cycle of northeastern Maine, *Ecology* 58 (1977) 139–148.
- [25] Lynch T.B., Holley A.G., Stevenson D.J., A random-parameter height-dbh model for cherrybark oak, *South. J. Appl. For.* 29 (2005) 22–26.
- [26] MacFarlane D.W., Ecologically stratified height-diameter models for hardwood species in northwestern lower Michigan, in: Yaussi D.A., Hix D.M., Long R.P., Goebel P.C. (Eds.), Proc. 14th Central Hardwood Forest Conf., USDA Forest Service, Newtown Square, Pennsylvania, Gen. Tech. Rep. NE-316, 2004, pp. 87–93.
- [27] Mehtätalo L., A longitudinal height-diameter model for Norway spruce in Finland, *Can. J. For. Res.* 34 (2004) 131–140.
- [28] Meyer H.A., A mathematical expression for height curves, *J. For.* 38 (1940) 415–420.
- [29] Nanos N., Calama R., Montero G., Gil L., Geostatistical prediction of height/diameter models, *For. Ecol. Manage.* 195 (2004) 221–235.
- [30] Oliver C.D., Larson B.C., Forest stand dynamics, (update edition), John Wiley & Sons, Inc., New York, 1996.
- [31] Parresol B.R., Baldcypress height-diameter equations and their prediction confidence intervals, *Can. J. For. Res.* 22 (1992) 1429–1434.
- [32] Peng C., Zhang L., Liu J., Developing and validating nonlinear height-diameter models for major tree species of Ontario's boreal forests, *North. J. Appl. For.* 18 (2001) 87–94.
- [33] Pienaar L.V., Turnbull K.J., The Chapman-Richards generalization of von Bertalanffy's growth model for basal area growth and yield in even-aged stands, *For. Sci.* 19 (1973) 2–22.

- [34] Pinheiro J.C., Bates D.M., Mixed-effects models in S and S-PLUS, Springer, New York, 2000.
- [35] Pinheiro J.C., Bates D.M., DebRoy S., Sarkar D., nlme: Linear and nonlinear mixed effects models, R package version 3.1-65, 2005, available online at <http://www.R-project.org>.
- [36] R Development Core Team, R: A language and environment for statistical computing (Version 2.2), R Foundation for Statistical Computing, Vienna, Austria (ISBN 3-900051-07-0), 2005, available online at <http://www.R-project.org>.
- [37] Ratkowsky D.A., Nonlinear regression modeling: a unified practical approach, Marcel Dekker, Inc., New York, 1983.
- [38] Ratkowsky D.A., Reedy T.J., Choosing near-linear parameters in the four-parameter logistic model for radioligand and related assays, *Biometrics* 42 (1986) 575–582.
- [39] Richards F.J., A flexible growth function for empirical use, *J. Exp. Bot.* 10 (1959) 290–300.
- [40] Saunders M.R., Wagner, R.G., Long-term spatial and structural dynamics in Acadian mixedwood stands managed under various silvicultural systems, *Can. J. For. Res.*, in press.
- [41] Sendak P.E., Brissette J.C., Frank R.M., Silviculture affects composition, growth, and yield in mixed northern conifers: 40-year results from the Penobscot Experimental Forest, *Can. J. For. Res.* 33 (2003) 2116–2128.
- [42] Sharma M., Zhang S.Y., Height-diameter models using stand characteristics for *Pinus banksiana* and *Picea mariana*, *Scand. J. For. Res.* 19 (2004) 442–451.
- [43] Trorey L.G., A mathematical method for the construction of diameter height curves based on site, *For. Chron.* 8 (1932) 121–132.
- [44] Weibull W., A statistical distribution function of wide applicability, *J. Appl. Mech.* 18 (1951) 293–296.
- [45] Wilkinson R.C., Leader and growth characteristics of eastern white pine associated with white pine weevil attack susceptibility, *Can. J. For. Res.* 13 (1983) 78–84.
- [46] Yang, Y., Monserud, R.A., Huang, S., An evaluation of diagnostic tests and their roles in validating forest biometric models, *Can. J. For. Res.* 34 (2004) 619–629.
- [47] Zhang S., Burkhart H.E., Amateis R.L., The influence of thinning on tree height and diameter relationships in loblolly pine plantations, *South. J. Appl. For.* 21 (1997) 199–205.