Modelling tree height–diameter relationships in multi-species and multi-layered forests: A large observational study from Northeast China

H. Temesgen [a,*], C.H. Zhang [b], X.H. Zhao [b]

[a] Department of Forest Resources, Oregon State University, 237 Peavy Hall, Corvallis, OR 97331-5703, USA
[b] Key Laboratory for Forest Resources & Ecosystem Processes of Beijing, Beijing Forestry University, Beijing 100083, China

ARTICLE INFO

ABSTRACT

Projection of stand development over time relies on accurate height–diameter functions. In complex forests where trees of various ages, species, sizes, vigor and crown classes, and shade tolerance levels co-exist, tree height predictions are more difficult than in uniform conditions. In such complex situations, individual tree height–diameter relationships and predictions are often further complicated by structure (e.g., relative position of trees) and stand density which sometime are exacerbated by selective harvesting and other disturbances, and changing stand dynamics. Height–diameter (H–D) relationships are significantly modulated by microsite (elevation, aspect and slope), climate, and competition. Traditionally, tree height–diameter relationships are modeled or examined using linear or nonlinear regression models, in which the microsite influences or species–habitat relationships are largely ignored. Studies that have examined H–D relationships for multi-species and multi-layered forests using species–habitat relationships or microsite influences are limited or non-existent.

Using an extensive dataset, we evaluated two sets of height–diameter equations for estimating tree heights in multi-species and multi-layered forests in north-eastern China. The first set included three equations for estimating height as a function of individual tree diameter, the second set included nine equations for estimating height as a function of individual tree diameter, relative tree position, and stand density variables. The inclusion of relative position of a tree and stand density into the base height–diameter equations increased the accuracy of prediction for all species. On the average, root mean square values were reduced by 12.2% by including these variables. After comparing and selecting a base H–D equation, we evaluated its statistical attributes among habitat types constructed for forest a large observational study site in north-eastern China.

Nonlinear mixed effect (NLMEM) was used to account for the hierarchical data structure and to provide predictive equations. Dummy variables were used to indicate species effect and allow simultaneous estimation of all species. The NLMEM framework along with species effect made tree height predictions compatible for 23 tree species found in multi-layered forests in NE China. Model coefficients for selected equations are documented for future use.

© 2013 Published by Elsevier B.V.

1. Introduction

In multi-species and multi-layered forests, tree heights and diameter (H–D) relationships are more variable than in even-aged, single-layered, and/or single-species stands. Yet, the estimation of growth, yield, and stand development over time in multi-layered stands heavily relies on accurate height–diameter models, which are critical building block in quantifying tree volume, form, and weight. In multi-species and multi-layered forests, individual tree height–diameter relationships and predictions are often further complicated by spatial distribution patterns and relative position of trees in the stand which sometime are exacerbated by selective harvesting and other disturbances and changing stand dynamics, making tree height predictions are more difficult than in uniform or single-species stands (Schmidt et al., 2011).

The mixed deciduous forests in northeastern China are complex, with many species and a wide range of tree sizes. Species include shade-tolerant conifers (e.g., *Abies holophylla* and *Abies nephrolepis*), semi-tolerant conifers (e.g., *Sorbus alnifolia*), and shade-intolerant conifers (e.g., *Acer barbinerve*, and *Tilia*

---

* Corresponding author. Tel.: +1 541 737 8549; fax: +1 541 737 3049.
E-mail address: hailemariam.temesgen@oregonstate.edu (H. Temesgen).

mandshurica). These stands may also contain shade intolerant hardwoods (e.g., *Betula dahurica* and *Carpinus cordata*). Recognizing the difficulty of identifying tree species in tropical forest in Hainan Province of China, Fang and Bailey (1998) fitted height diameter models for all species, without accounting for the species-specific differences in H–D relationships. This approach might not lead to precise height prediction, and makes it difficult to predict stand and stock tables in preparing forest management plans.

The variation of H–D relationship is of scientific interest from a variety of perspectives. From an ecophysiological perspective, the variability in H–D may provide clues to how plants acclimate to their environment or how trees compete for resources (Canham et al., 2006). It has been proposed, for example, that variability in H–D within and among habitat types is adaptive or simply a passive response to an environmental gradient, such as light, elevation, slope, aspect or proximity to the coast (Schmidt et al., 2011). Strong empirical relationships have been noted between H–D and stand density and relative position (Temesgen and von Gadow, 2004; Temesgen et al., 2007) and also across a broad spectrum of different ecosystems (Huang et al., 2000).

Tree height–diameter relationships are commonly modeled or examined using species-specific linear or nonlinear regression models. Height–diameter (H–D) relationship is significantly modulated by site (e.g., elevation, aspect and slope), climate, and competition. Several studies have fitted H–D relationships vary by eco-region (e.g., Huang et al., 2000), used local site and competition measures to mimic the height–diameter allometry. However, none of the studies have examined H–D relationships for multi-species and multi-layered forests using species–habitat relationships.

The relationship between *h* and *d* is reported to vary by habitat (Huang et al., 2000; Temesgen and von Gadow, 2004), indicating the need for varying parameters among habitat types. Mixed model approaches formally incorporate the between-habitat type variability of *h–d* relationship into the model. Estimation of *h* using nonlinear mixed-effects models has been previously reported by Lappi (1991, 1997), Mehtätalo (2004), and Castedo et al. (2006). The mixed model NMEM estimates are effectively empirical Bayes estimates, and can be motivated from a Bayesian perspective. The prediction of *h* for trees from a new stand is based on the prior information, from the training dataset, actualized with new data collected as a subsample of trees from the stand or habitat type.

Despite the growing research interest on mixed model to predict tree heights, detailed analyses that quantify the gains obtained by using these methods over fixed effect models are lacking. Also lacking are analyses that examine the ability, efficiency and suitability of any of these approaches for predicting missing height measurements for multi-species and multi-layered forests using species–habitat relationships. Eerikäinen (2009) was the first to develop "localized species-specific" height diameter curves for more than three species groups.

The primary objectives of this study were to examine (1) the predictive abilities of commonly used height–diameter models in predicting tree heights in multi-species and multi-layered forests in NE China; (2) the contributions of relative position of trees and stand density variables in predicting tree height; and (3) the differences in height–diameter relationships and equations among habitat types constructed for multi-species and multi-layered forests of NE China. Based on an evaluation of the relative performance of a variety of potential diameter functions, a height–diameter equation was selected and its responses to habitat types were evaluated.

2. Methods

2.1. Study site

The study was carried out in a temperate forest at *Jiaohe* in Jilin Province, in Northeastern China. The forest type is a mixed broadleaf–conifer forest, in which the average annual temperature is 3.8 °C, and the hottest month is July with an average day temperature of 21.7 °C. The coldest month is January with an average day temperature of −18.6 °C. The average annual precipitation is 695.9 mm. The soil is a brown forest soil with a rootable depth ranged from 20 to 100 cm.

The observations for this study were obtained in a fully censused, unmanaged old-growth forest plot (OGF), covering 30 ha (500 × 600 m). The OGF plot was established in the summer of 2010, and is located at 43°57.928′–43°58.214′N, 127°45.287′–127°45.790′E. The OGF study area is situated in a protected area, far away from villages, where human disturbance has been virtually unknown. The elevation of ranges from 577.8 to 780.7 m above sea level and the topography is characterized by a valley between two slopes.

2.2. Data

Data for this study were collected as part of long-term forest observational study. The vegetation type represents a mixed broadleaf–conifer forest with 23 species. Altogether 20,254 individual trees with a breast height diameter (dbh, measured at 1.3 m above ground level) exceeding 5 cm were tagged and mapped, and their species was identified. The species were identified according to the records in the Chinese Virtual Herbarium (see http://www.cvh.org.cn/cms/), and are listed in Appendix A.

The dominant tree species by basal area are *Ulmus davidiana* var. *japonica* (Rehder) Nakai, *Pinus koraiensis* Siebold & Zucc., *Juglans mandshurica* Maxim., *Tilia mandshurica* Rupr.er Maxim., *C. cordata* Bl., *Acer mono* Maxim., *Fraxinus mandshurica* Rupr., *Tilia amurense* Rupr. and *Ulmus laciniata* (Trautv.) Mayr. The top five species in stem density are *Acer mandshuricum* Maxim., *Syringa reticulata* var. *amurensis* (Rupr.), *U. davidiana* var. *japonica*, *C. cordata* and *Acer mono*, respectively. The total basal area of dominant tree species and stem density of the top five species are shown in Appendix B.

To examine tree-height diameter relationships, the 30 ha OGF plot was divided in 20 m × 20 m cell. The elevation of a particular cell was calculated as the mean of the elevations of its four corner nodes. The slope for each cell was estimated as the mean angular deviation from the horizontal plane of each of the four triangular planes which were formed by connecting three of its adjacent corners (Harms et al., 2001). The convexity of a cell was calculated as the elevation of the focal cell minus the mean elevation of the eight surrounding cells (cf. Yamakura et al., 1995). For the edge cells, convexity was taken as the elevation of the center point minus the mean of the four corners. Positive and negative convexity values respectively indicate convex (ridge) and concave (valley) land surfaces. The aspect of a cell can be obtained from the average angle of the four triangular planes that deviate from the north direction. Four maps show the spatial pattern of the four topographic variables using 20 m × 20 m cells (Fig. 1). Each cell shows the altitude (ranging from 577.8 m to 780.7 m above sea level with 202.9 m difference in altitude between the highest and lowest cells), the convexity (ranging from −6.8 m to 8.6 m), the slope (ranging from 2.4° to 45.7°) and the aspect (ranging from 72.8° to 299.9°) (Table 1).
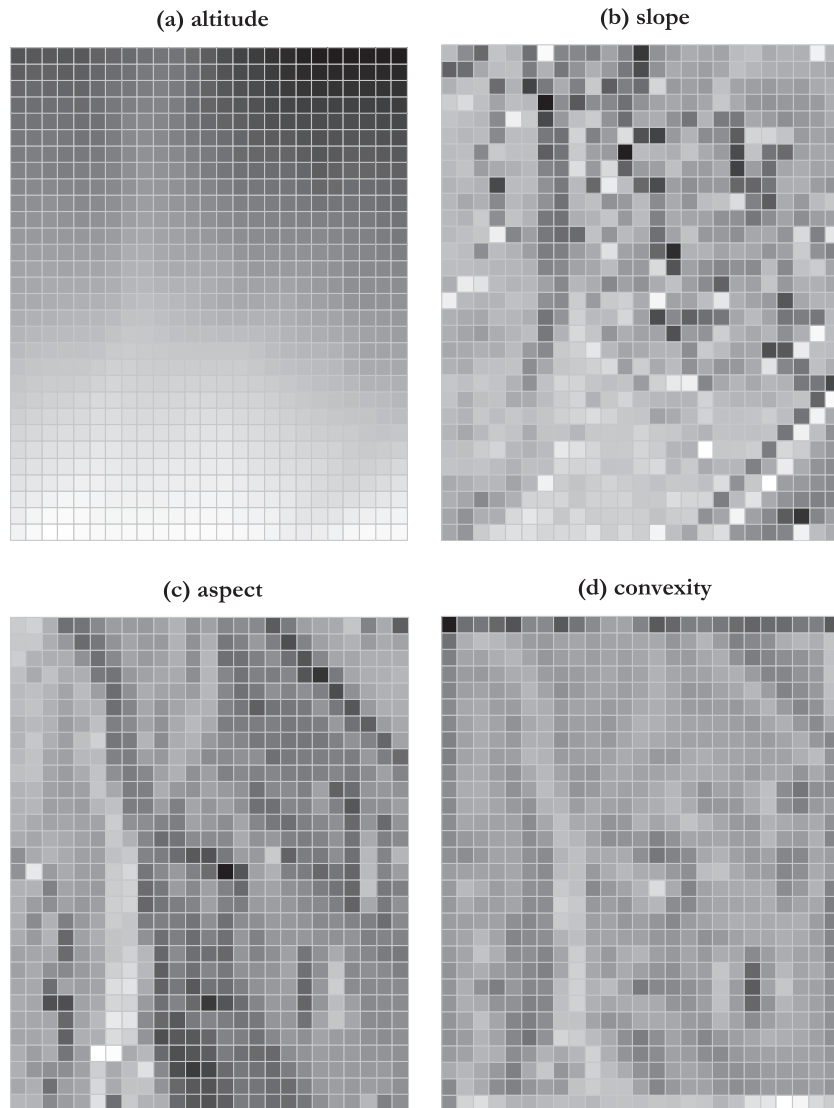
(a) altitude

(b) slope

(c) aspect

(d) convexity

Fig. 1. Maps depicting four topographic variables at the scale of 20×20 m cells. (a) Altitude from 577.8 m (white) to 780.7 m (black) above sea level. (b) Slope from 2.4° to 45.7°. (c) Aspect from 72.8° to 299.9°. (d) Convexity from −6.8 m (white) to 8.6 m (black).

Table 1
Minimum, average (mean), maximum, and standard deviation (std) of selected attributes for the study site.

Attribute	Minimum	Mean	Maximum	Std
Total basal area (m²/ha)	4.2	30.1	78.3	10.6
Stems (ha)	150.0	775.8	1475.0	268.7
Elevation (m)	577.8	669.9	780.7	52.2
Slope	2.4	19.8	45.7	7.5
Aspect	72.8	186.3	299.9	34.8
Convexity	−6.8	0.2	8.6	1.5

2.3. Height–diameter models by tree species

A literature review yielded three widely used height–diameter models that had been found useful in other studies because of their predictive abilities. Because height–diameter relationships are expected to vary by species, the selected models were fitted by species. The models were then compared for accuracy and examined for known biological trends. In order to obtain stable parameter estimates, only species with more than 30 trees were included in this study.

2.3.1. Model 1: Chapman–Richards (Richards, 1959)

Garman et al. (1995) used the Chapman–Richards (1959) function for 24 western Oregon tree species. Eq. (1) in Table 2 is the most commonly used height–diameter equation for estimating tree heights for the forests of Washington and Oregon.

2.3.2. Model 2: Yang et al. (1978)

Eq. (2) in Table 2 is a Weibull and a Chapman–Richards type model which produces a graph that is concave downward without an inflection point. This model was used by Yang et al. (1978) to estimate tree heights for some of the major tree species in British Columbia. It has also been extensively used in describing height–age relationships. Huang et al. (1992) cautioned that this model approaches the asymptote too quickly when there is a weak relationship between the dependent and predictor variables.

2.3.3. Model 3: Ratkowsky (1990)

In Eq. (3) in Table 2, $e^{\beta 0}$ represents the asymptotic height. Parameters β_1 and β_2 represent steepness and curvature respectively. A modified variant of Model 3 in which β_2 is fixed to 1.0 has been used for the 11 tree species found in the Northern Idaho

Table 2
List of equations examined in this study.

Equation	Form
[1] Chapman–Richards (Richards, 1959; Garman et al., 1995)	$\hat{h} = 1.3 + \beta_0[1 - e^{\beta_1 d}]^{\beta_2}$
[2] Yang et al. (1978)	$\hat{h} = 1.3 + \beta_0[1 - e^{\beta_1 d^{\beta_2}}]$
[3] Ratkowsky (1990)	$\hat{h} = 1.3 + e^{(\beta_0 + \frac{\beta_1}{d + \beta_2})}$
[4] Eq. (1) with BAL and BA	$\hat{h} = 1.3 + \beta_0[1 - e^{\beta_1 d}]^{\beta_2}$, where $\beta_0 = a_0 + a_1 \times BAL + a_2 \times BA$
[5] Eq. (2) with BAL and BA	$\hat{h} = 1.3 + \beta_0[1 - e^{\beta_1 d^{\beta_2}}]$, where $\beta_0 = a_1 + a_2 \times BAL + a_3 \times BA$
[6] Eq. (3) with BAL and BA	$\hat{h} = 1.3 + e^{(\beta_0 + \frac{\beta_1}{d + \beta_2})}$, where $\beta_0 = a_1 + a_2 \times BAL + a_3 \times BA$
[7] Eq. (1) with BAL and SPH	$\hat{h} = 1.3 + \beta_0[1 - e^{\beta_1 d}]^{\beta_2}$, where $\beta_0 = a_0 + a_1 \times BAL + a_2 \times SPH$
[8] Eq. (2) with BAL and SPH	$\hat{h} = 1.3 + \beta_0[1 - e^{\beta_1 d^{\beta_2}}]$, where $\beta_0 = a_0 + a_1 \times BAL + a_2 \times SPH$
[9] Eq. (3) with BAL and SPH	$\hat{h} = 1.3 + e^{(\beta_0 + \frac{\beta_1}{d + \beta_2})}$, where $\beta_0 = a_0 + a_1 \times BAL + a_2 \times SPH$
[10] Eq. (1) with Ht and BA	$\hat{h} = 1.3 + \beta_0[1 - e^{\beta_1 d}]^{\beta_2}$, where $\beta_0 = a_0 + a_1 \times Ht + a_2 \times BA$
[11] Eq. (2) with Ht and BA	$\hat{h} = 1.3 + \beta_0[1 - e^{\beta_1 d^{\beta_2}}]$, where $\beta_0 = a_0 + a_1 \times Ht + a_2 \times BA$
[12] Eq. (3) with Ht and BA	$\hat{h} = 1.3 + e^{[\beta_0 + \frac{\beta_1}{d + \beta_2}]}$, where $\beta_0 = a_0 + a_1 \times Ht + a_2 \times BA$

Where: a, b, and c are species dependent coefficients, $\hat{h}$ is estimated tree height in m, d is observed diameter outside bark at breast height in cm, e is the Naperian constant (i.e., 2.718), BAL is basal area in larger trees (m^2/ha), and BA is basal area per hectare (m^2/ha), SPH is number trees/ha, and Ht is height of the tallest tree for each species.

variant of Prognosis (Wykoff et al., 1982). In Model 3 parameter β_2 varies by tree species. Flewelling and de Jong (1994) also used this model to estimate missing tree heights for British Columbia's forests.

2.4. Extended height–diameter equations

Given that the height–diameter models are intended for multi-species and multi-layered stands, it is expected that a decrease in relative position or an increase in stand density would result in an increase in predicted height for a given dbh. The base height–diameter models were extended by including stand density attributes (basal area in m^2/ha, BA and number of trees/ha, SPH), and a relative position (basal area in larger trees in m^2/ha, BAL) in the base models. We also examined the impacts of including the height of the most dominant trees in predicting tree heights.

The contributions of stand density and relative position were evaluated for potential improvement in the predictive abilities of each of the four base equations. This was accomplished by expressing the asymptotic parameter in each equation (the "β_0" parameter) as a function of these variables (Table 2). For example, inserting a relative position (BAL) and a stand density measure (BA) into Eq. (1) can be expressed as in Eq. (4) in Table 2.

In Eq. (4) in Table 2, the asymptote parameter of the Chapman–Richards (1959) equation is assumed to vary linearly with BAL of a subject tree and BA of stand in which the tree resides. In multi-species and multi-layered stands, tree ages are highly variable and are measured only on a small subset of trees. As a result, tree age and site index were intentionally excluded from the model. Extended height–diameter models with BAL, BA, and SPH are listed in Table 2.

All models were fitted with weighted nonlinear regression techniques (weight = 1.0/dbh) (Temesgen et al., 2007). Fit statistics and model parameters were estimated using the Gauss–Newton optimization technique (Gallant, 1987) in a nonlinear least squares procedure (PROC NLIN) on SAS software (SAS Institute Inc., 2005). In order to find a global minimum, the starting value of each parameter was varied and several runs were obtained. Initial approximations for each parameter were obtained from linear transformations of the models, where possible.

2.5. Equation comparison and selection

The predictive performances of the height–diameter equations were compared using bias and RMSE. The fits of the equations were examined by randomly splitting the data for model development (80%) and for validation (20%) 200 times. In the evaluation, 20% of the plots (20×20-m cell) were excluded from the dataset and the selected models were fitted to the rest of the plots. Then, the models were used to predict the height of all the trees in the excluded plots. The rationale behind this method is that trees from the same plot tend to be correlated. Hence, excluding 20% plots, from the dataset, to examine the performance of the models provides stronger model evaluation for new plots that were not included in the original dataset.

To evaluate the performance of the selected height–diameter equations, the following cross-validation statistics were computed for each realization of randomly selected data, and averaged over the 200 realizations.

$$\text{Bias} = \frac{1}{R} \sum_{r=1}^{R} \frac{\sum_i^n (h_i - \hat{h}_i)}{n} \tag{1}$$

where R = number of resamplings (200); n is the number of trees of a given species per resampling r; H_i is measured height; and $\hat{H}_i$ is estimated height from the cross-validation study.

$$\text{RMSE} = \sqrt{\frac{1}{R} \sum_{r=1}^{R} \frac{\sum_i^n (h_i - \hat{h}_i)^2}{n}} \tag{2}$$

where RMSE is the estimate of root mean square error (a measure of accuracy of the prediction); R = number of resamplings (200); n is the number of trees of a given species per resampling r. It should be noted that, because RMSE does not remove the effect of bias in its calculation, any bias in the equation will increase its RMSE. Both statistics were rounded to the nearest 0.1 m, which corresponds to the precision one could reasonably expect when measuring H.

The 'best' equation was selected in the following fashion:

1. The equation with the lowest RMSE was noted for each species.
2. If the equation with the lowest RMSE was severely biased (i.e., more than 10 cm), then the unbiased equation with the lowest RMSE was also recorded for that species.
3. Finally, the equations were then evaluated for the consistency of their performance across all the species.

After evaluating the results, one equation was selected to examine height–diameter relationships and variation by habitat types and species–habitat type interactions.

2.6. Height–diameter relationships and variation by habitat types

Multivariate regression tree (MRT) analysis was performed, following De'ath (2002, 2010), to classify habitat types using topographic conditions and species composition. Distance-based MRT was used to describe relationships between multispecies data and environmental characteristics. The dissimilarities used in distance-based MRT are calculated by Euclidean distances. Thus, one obtains clusters of sites by repeated splitting of the data, which are chosen to minimize the dissimilarity of sites within clusters (Zhang et al., 2012).

In multi-species and multi-cohort stands, missing heights can be predicted from species dependent height–diameter equations. Given that habitat types vary in their microsite attributes (e.g., elevation, aspect, and slope) and dynamics regarding interspecies competition, height–diameter relationships and model coefficients are expected to vary by habitat types. Because trees from the same habitat type tend to be more similar to each other than to trees from different habitat types, the classical regression assumption that observations are independent does not hold (Gregoire et al., 1995; Lappi, 1997, 2006). Hence, accounting for the differences in height–diameter relationship for a selected model is warranted.

While the 12 equations listed in Table 2 are flexible and useful to compare different model forms and predictor variables, the equations do not account for the clustered and nested structure of the data. Recognizing this lack of independence, several studies have explored the use of linear mixed effects models (LMEM) (Lappi, 1991, 1997; Monleon, 2003; Mehtätalo, 2004; Eerikäinen, 2009), and non nonlinear mixed models (NMEM) (Calama and Montero, 2004; Hall and Clutter, 2004; Castedo et al., 2006; Temesgen et al., 2008) to estimate height as a function diameter and to account for a stand or habitat effect.

2.7. Including habitat and species effects

To examine the height–diameter relationships and variation by habitat types and to account for habitat and species effects, we used a nonlinear mixed effect model (NMEM). The nonlinear mixed model can be motivated as a hierarchical model (Pinheiro and Bates, 2000). The height of the jth tree from the ith habitat type is modeled as (after Temesgen et al., 2008):

$$h_{ij} = f(\boldsymbol{\beta_{ij}}, d_{ij}) + \varepsilon_{ij}, \quad i = 1, \ldots, M, \quad j = 1, \ldots, n_i, \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \tag{3}$$

where $\boldsymbol{\beta_{ij}}$ is a vector of p habitat type, and possibly tree, specific parameters, f is a real-valued function that relates the height of a tree to its diameter, and ε_{ij} is a within-habitat error term that is assumed to be independent and normally distributed. In this study, f can be one of the 12 equations listed in Table 2.

The kth element of the parameter vector $\boldsymbol{\beta_{ij}}$ is then modeled as a linear function of fixed and random effects:

$$\beta_{ijk} = \mathbf{x}'_{ijk}\boldsymbol{\beta_k} + \mathbf{z}'_{ijk}\mathbf{b}_{ik}, \quad k = 0, \ldots, p-1, \quad \mathbf{b}_{ik} \sim N(\mathbf{0}, \boldsymbol{\Psi_k}) \tag{4}$$

$\boldsymbol{\beta_k}$ is a vector of fixed effects and $\mathbf{b}_{ik}$ is a vector of random effects associated with the ith habitat type. It is assumed that $\mathbf{b}_{ki}$ is normally distributed, with mean $\mathbf{0}$, covariance matrix ψ_k and independent of ε_{ij}. $\mathbf{x}_{ijk}$ and $\mathbf{z}_{ijk}$ are vectors of covariates associated with the ith habitat type and possibly with the jth tree. Allowing for these vectors to depend on both the habitat type and trees within habitat type permits the use of positional variables such as BAL and Ht.

In matrix form, Eqs. (3) and (4) can be written as:

$$\mathbf{h_i} = \mathbf{f}(\boldsymbol{\beta_i}, \mathbf{d_i}) + \boldsymbol{\varepsilon_i}$$

$$\boldsymbol{\beta_i} = \mathbf{X_i}\boldsymbol{\beta} + \mathbf{Z_i}\mathbf{b_i}, \quad \mathbf{b_i} \sim N(\mathbf{0}, \boldsymbol{\Psi})$$

These two equations can be combined, so that:

$$\mathbf{h_i} = \mathbf{f}(\boldsymbol{\beta}, \mathbf{b_i}, \mathbf{X_i}, \mathbf{Z_i}) + \boldsymbol{\varepsilon_i} \tag{5}$$

where h_i is a vector of tree heights from habitat type i, $\boldsymbol{\beta_i}$ is a vector of habitat-level parameters, d_i is a vector of tree diameters from habitat type i, ε_i is a vector of within-habitat type error terms, $\boldsymbol{\beta}$ is a vector of fixed parameters that do not depend on the habitat type, $\mathbf{b_i}$ a vector of habitat type random effects, and $\mathbf{X_i}$ and $\mathbf{Z_i}$ are matrices of explanatory variables. The random effects for habitat type enter into Eq. (5) non-linearly, making the model a multilevel nonlinear mixed model.

After examining the between habitat variability of the coefficients of the base equations, we added a single random effect to the parameter that controls the model asymptote. For example, when a random effect is included in the asymptote of the Ratowski function, Eq. (4) becomes:

$$\beta_{0ij} = \beta_0 + b_i, \quad b_i \sim N(0, \sigma_b^2) \tag{6}$$

if only the tree diameter is included in the model, or:

$$\beta_{0ij} = \beta_{00} + \beta_{01}\mathrm{Ht}_{ij} + \beta_{02}BA_i + b_i, \quad b_i \sim N(0, \sigma_b^2) \tag{7}$$

if relative position and stand density variables are included. Inserting Eq. (6) as the asymptote of the Ratowski equations in Table 2 yields:

$$h_{ij} = 1.3 + e^{\left((\beta_0 + b_i) + \frac{\beta_1}{d_{ij} + \beta_2}\right)} + \varepsilon_{ij} \tag{8}$$

$$h_{ij} = 1.3 + e^{\left((\beta_{00} + \beta_{01}\mathrm{Ht}_{ij} + \beta_{02}BA_i + b_i) + \frac{\beta_1}{d_{ij} + \beta_2}\right)} + \varepsilon_{ij} \tag{9}$$

We deliberately excluded cell (plot)-level random effect because variables such as BA and SPH are normally estimated at forest or stand-level; they are not plot-level covariates. In addition, plot-level random effect did not provide stable convergence, and would not be ideal for future use of the models examined in this study.

In addition to the random asymptote, one may consider a random steepness (β_1) or curvature parameter (β_2), or both. The correlation between the steepness parameter and the asymptote is very high, 0.82. As a consequence, modeling two or all three parameters as random effects resulted in convergence problems. Hence, we did not consider making the steepness (β_1) or curvature parameter (β_2) random.

The steepness (β_1) coefficient varied widely by tree species. To account for species effect and develop species or species-group specific equations, tree species were grouped by genius, shade tolerance, and total number of trees by species (species with more than 2000 trees were assigned by themselves). We used dummy variables to include species-effects by including indicator variables the steepness (β_1) coefficient. To develop localized, species-specific height–diameter curves, Eerikäinen (2009) also used dummy variables to account for species effect.

When adding the species indicator variables, Eq. (8) becomes:

$$h_{ij} = 1.3 + e^{\left((\beta_0 + b_{0i}) + \frac{\beta_1 \gamma_2 + \cdots + \gamma_{15}}{d_{ij} + \beta_2}\right)} + \varepsilon_{ij} \tag{10}$$

When adding the relative position and stand density variables, Eq. (9) becomes:

$$\beta_{0ij} = \beta_{00} + \beta_{01}\mathrm{Ht}_i + \beta_{02}BA_i + b_i$$

Combining those two equations yields:

$$h_{ij} = 1.37 + e^{\left((\beta_{00} + \beta_{01}\mathrm{Ht}_{ij} + \beta_{02}BA_i + b_{1i}) + \frac{\beta_1 \gamma_2 + \cdots + \gamma_{15}}{d_{ij} + \beta_2}\right)} + \varepsilon_{ij} \tag{11}$$

where $\gamma_1, \ldots, \gamma_{15}$ are dummy variables for species or species groups listed in Appendix A.

2.7.1. Parameter estimation

For these models, parameter estimates and their fit statistics were obtained by using *R* and the asymptotic *t*-statistics of the parameters and asymptotic 95% confidence intervals were then evaluated. Each equation's parameters were evaluated to determine if they were significantly different from zero using the asymptotic *t*-test. Parameters not significantly different from 0 at $p = 0.05$ were set to zero and the remaining parameter re-estimated. Final parameter estimates and their fit statistics are reported for predicting height of trees growing in multi-species and multi-layered forests of NE China.

2.7.1.1. Predicting heights of trees for a new habitat.
Prediction of heights of trees for a new domain is detailed in Temesgen et al (2008). A brief excerpt follows. Let that the heights of a subsample of *m* trees from a new habitat type are known. Let $\mathbf{h_m}$, $\mathbf{X_m}$ and $\mathbf{Z_m}$ be the height vector and covariate matrices from those trees. Then, the height of another tree from the same habitat type can be estimated by first predicting the random effects of the stand, $\mathbf{b_m}$, based on the subsample of *m* trees of known heights, and then calculating the new height as:

$$\hat{h}_{new} = f(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}_{\mathbf{m}}, \mathbf{x}_{\mathbf{new}}, \mathbf{z}_{\mathbf{new}})$$

where $\hat{\boldsymbol{\beta}}$ are the estimated fixed-effects parameters, $\hat{\mathbf{b}}_{\mathbf{m}}$ the predicted random effects for habitat type *m*, and $\mathbf{x}_{\mathbf{new}}$ and $\mathbf{z}_{\mathbf{new}}$ the vector of covariates for the new tree that does not have a measured *h*.

Prediction of the random effects follows directly from Lindstrom and Bates' algorithm to estimate the model parameters (Lindstrom and Bates, 1990; Pinheiro and Bates, 2000). The NMEM is linearized using a Taylor expansion about the current estimate of θ and **b**. At each iteration, the model is approximated by the following linear mixed effects model:

$$\hat{\mathbf{h}}_{\mathbf{i}} = \hat{\mathbf{X}}_{\mathbf{i}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{Z}}_{\mathbf{i}}\hat{\mathbf{b}}_{\mathbf{i}}$$

where the transformed variables $\hat{\mathbf{h}}_{\mathbf{i}}$, $\hat{\mathbf{X}}_{\mathbf{i}}$ and $\hat{\mathbf{Z}}_{\mathbf{i}}$ are defined as:

$$\hat{\mathbf{h}}_{\mathbf{i}} = \mathbf{h}_{\mathbf{i}} - \mathbf{f}(\mathbf{X}_{\mathbf{i}}, \mathbf{Z}_{\mathbf{i}}, \hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}_{\mathbf{i}}) + \hat{\mathbf{X}}_{\mathbf{i}}\hat{\boldsymbol{\beta}} + \hat{\mathbf{Z}}_{\mathbf{i}}\hat{\mathbf{b}}_{\mathbf{i}} \quad (12)$$

$$\widehat{\mathbf{X}}_{\mathbf{i}} = \left.\frac{\partial \mathbf{f}(\mathbf{X}_{\mathbf{i}}, \mathbf{Z}_{\mathbf{i}}, \boldsymbol{\beta}, \mathbf{b}_{\mathbf{i}})}{\partial \boldsymbol{\beta}'}\right|_{\hat{\beta}}, \hat{\mathbf{b}}_{\mathbf{i}} \quad (13)$$

$$\widehat{\mathbf{Z}}_{\mathbf{i}} = \left.\frac{\partial \mathbf{f}(\mathbf{X}_{\mathbf{i}}, \mathbf{Z}_{\mathbf{i}}, \boldsymbol{\beta}, \mathbf{b}_{\mathbf{i}})}{\partial \mathbf{b}'_{\mathbf{i}}}\right|_{\hat{\beta}, \hat{\mathbf{b}}_{\mathbf{i}}}$$

The random effects for a new habitat type are approximated using an empirical BLUP on this LMEM approximation at convergence:

$$\hat{\mathbf{b}}_{\mathbf{m}} \approx \boldsymbol{\Psi}_{\hat{\sigma}^2}\widehat{\mathbf{Z}}'_{\mathbf{m}}\mathbf{V}_{\hat{\sigma}^2}^{-1}\left[\hat{\mathbf{h}}_{\mathbf{m}} - \widehat{\mathbf{X}}_{\mathbf{m}}\hat{\beta}\right] \quad (14)$$

where $\hat{\beta}$ is the estimates of the fixed parameters and $\boldsymbol{\Psi}_{\hat{\sigma}^2}$ is the matrix $\boldsymbol{\Psi}$ evaluated at the estimated variance components $\hat{\sigma}^2$, and

$$\mathbf{V}_{\hat{\sigma}^2}^{-1} = \left[Var(\hat{\mathbf{h}}_{\mathbf{m}})|_{\hat{\sigma}^2}\right]^{-1} = \left[\hat{\sigma}_{\varepsilon}^2\mathbf{I}_{\mathbf{m}} + \widehat{\mathbf{Z}}_{\mathbf{m}}\boldsymbol{\Psi}_{\hat{\sigma}^2}\widehat{\mathbf{Z}}'_{\mathbf{m}}\right]^{-1}$$

where $\mathbf{I}_{\mathbf{m}}$ is the $m \times m$ identity matrix. Both $\hat{\beta}$ and $\hat{\sigma}^2$ are estimated from the training dataset.

Substituting $\mathbf{h}_{\mathbf{m}}$ from Eq. (12) into Eq. (14), yields:

$$\hat{\mathbf{b}}_{\mathbf{m}} \approx \boldsymbol{\Psi}_{\hat{\sigma}^2}\widehat{\mathbf{Z}}'_{\mathbf{m}}\mathbf{V}_{\hat{\sigma}^2}^{-1}[\mathbf{h}_{\mathbf{m}} - \mathbf{f}(\mathbf{X}_{\mathbf{m}}, \hat{\beta}, \hat{\mathbf{b}}_{\mathbf{m}}) + \widehat{\mathbf{Z}}_{\mathbf{m}}\hat{\mathbf{b}}_{\mathbf{m}}] \quad (15)$$

Note that the terms involving $\widehat{\mathbf{X}}_{\mathbf{new}}$ cancel out, so that Eq. (13) does not have to be calculated. In general, a solution to Eq. (15) has to be obtained iteratively.

$$h_{m,new} = 1.37 + e^{\left((\hat{\beta}_0 + \hat{b}_m) + \frac{\hat{\beta}_1 + \hat{\gamma}_1 + \ldots + \hat{\gamma}_{15}}{d_{m,new} + \hat{\beta}_2}\right)} \quad (16)$$

When relative position and stand density variables are included, Eq. (16) has the same form but with $\hat{\beta}_0$ substituted by $\hat{\beta}_{00} + \hat{\beta}_{01}\mathrm{Ht}_{mj} + \hat{\beta}_{02}BA_m$.

3. Results

3.1. Data summary

Altogether 23 tree species were present on the 30 ha study site. The number of species found on a single 20 m by 20 m plot ranged from 1 to 12 and averaged to almost five species. While CARCOR

Table 3
Number of trees (NTREE); minimum (min), average (mean), maximum (max) and standard deviation (std) of the diameter at breast height, height, basal area in larger tree by species.

| Species	NTREE	Diameter (cm)				Height (m)				Basal area in larger trees (m²/ha)			
		Min	Mean	Max	Std	Min	Mean	Max	Std	Min	Mean	Max	Std
ABIHOL	248	5.4	29.8	105.4	23.1	2.0	14.1	27.8	6.6	0.0	18.0	50.2	12.4
ABINEP	127	5.1	14.4	42.3	8.7	2.4	9.2	26.1	4.9	0.0	29.5	56.6	9.8
ACEBAR	776	5.1	7.1	46.5	3.7	1.7	6.4	20.0	2.6	2.4	25.1	78.0	10.6
ACEMAN	2319	5.1	14.0	78.3	9.9	1.8	10.2	24.6	4.5	0.0	24.4	73.5	10.0
ACEMON	2488	5.1	19.2	88.9	14.9	1.4	12.5	26.5	4.9	0.0	23.6	68.7	11.7
ACETEG	1394	5.1	12.0	48.1	5.3	2.1	10.5	22.0	3.9	2.2	25.6	62.8	10.0
BETCOS	775	5.3	27.9	96.5	19.5	3.8	16.0	28.0	4.3	0.0	18.9	58.8	11.9
BETDAH	54	5.9	37.5	91.2	27.2	3.9	15.0	25.3	7.1	0.0	14.3	48.8	12.8
BETPLA	38	5.5	15.9	57.4	13.6	2.3	11.3	24.3	4.4	0.0	22.6	44.7	9.8
CARCOR	4024	5.1	11.1	51.9	5.4	2.1	9.3	23.5	3.2	0.0	29.9	78.3	12.0
FRAMAN	372	5.1	24.4	78.0	16.4	3.5	14.8	27.0	5.0	0.0	20.6	56.9	11.8
JUGMAN	804	5.9	26.7	77.8	10.9	3.5	16.1	25.2	3.9	0.0	15.7	44.9	8.4
PADRAC	207	5.1	7.9	26.0	2.8	1.7	6.7	18.0	2.6	4.0	22.3	55.5	9.6
PHEAMU	790	5.1	22.1	62.5	10.0	2.3	14.5	27.0	4.3	0.0	21.3	57.3	9.9
PINKOR	731	5.1	31.9	78.9	22.8	2.1	13.6	28.6	6.9	0.0	18.7	76.0	13.7
POPKOR	39	9.7	23.2	62.1	11.7	9.0	16.3	21.2	3.7	0.0	17.1	34.9	7.9
QUEMON	137	5.4	34.5	114.5	25.8	2.6	14.9	26.0	4.9	0.0	18.7	62.3	13.4
SORALN	758	5.1	16.1	80.8	11.7	2.4	11.6	28.7	4.4	0.0	27.1	76.6	11.3
TILAMU	1771	5.1	24.0	99.0	15.2	1.9	14.3	27.3	4.7	0.0	24.5	76.8	12.9
TILMAN	102	5.4	22.7	61.0	13.4	4.0	12.7	24.0	4.9	0.0	22.1	43.8	9.4
ULMDAV	59	5.1	13.8	71.0	11.5	3.2	10.5	18.9	4.0	4.8	23.4	49.0	10.6
ULMLAC	2088	5.1	23.4	88.0	15.5	2.0	13.5	26.3	5.0	0.0	19.7	67.3	10.6
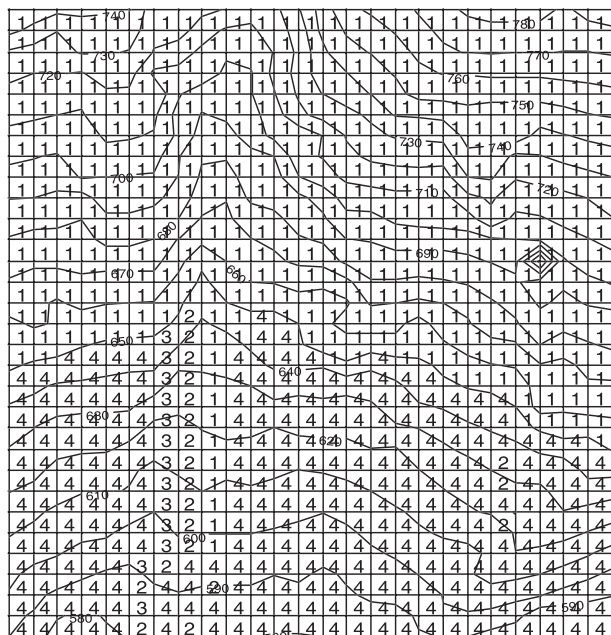
Fig. 2. Distribution of the four habitat types at the 20 m × 20 m scale. The lines show the elevation contours at 5 m intervals.

and ACEMON were the two most common species that are found on most of the 750 plots, PINKOR and TILAMU are the most valuable commercial species. Accordingly, the first six dominant and commercially valuable species will be the focus of this study. Since

one of the main objectives of this study was to evaluate the predictive performance of height–diameter equations by species, only stands with at least 30 trees sampled were included. The data set covered a wide array of stand densities, with basal areas ranging from 4.2 to 78.3 m²/ha; number of trees/ha ranging from 150 to 1475 trees/ha; tree diameters from 5.1 to 114.5 cm; and tree heights from 1.4 to 28.7 m (Table 1).

Selecting species with more than 30 trees resulted in a total sample of 20, 254 individual trees. The census trees were used to estimate the model parameters for 23 tree species. The data set covered a wide range of tree sizes with diameters ranging from 5.1 to 114.5 cm and standard deviations from 3.7 to 27.2 cm. Tree heights ranged from 1.4 to 28.7 m and their standard deviations ranged from 2.6 to 7.01 m (Table 3).

Plots of tree height versus diameter and other variables indicate nonlinear trends (Figs. 3 and 4). Simple correlations were obtained to describe the relationships between tree height and other variables. Based on simple correlations, tree height was negatively related to the BAL, as expected. In addition, the low correlation between tree height and convexity, which expresses the relative altitudinal difference between the focal cell and its surrounding neighbors, indicate convexity did not correlate well or influence height–diameter relationships (Table 4).

3.2. Evaluation of base equations by tree species

Noticeable differences were found among the predictive abilities of the height–diameter equations by species. The RMSE values ranged from 2.15–3.43, 2.11–3.42, 2.1–3.7, and 2.1–3.71 m for Models 1–3, respectively. For extended models that included
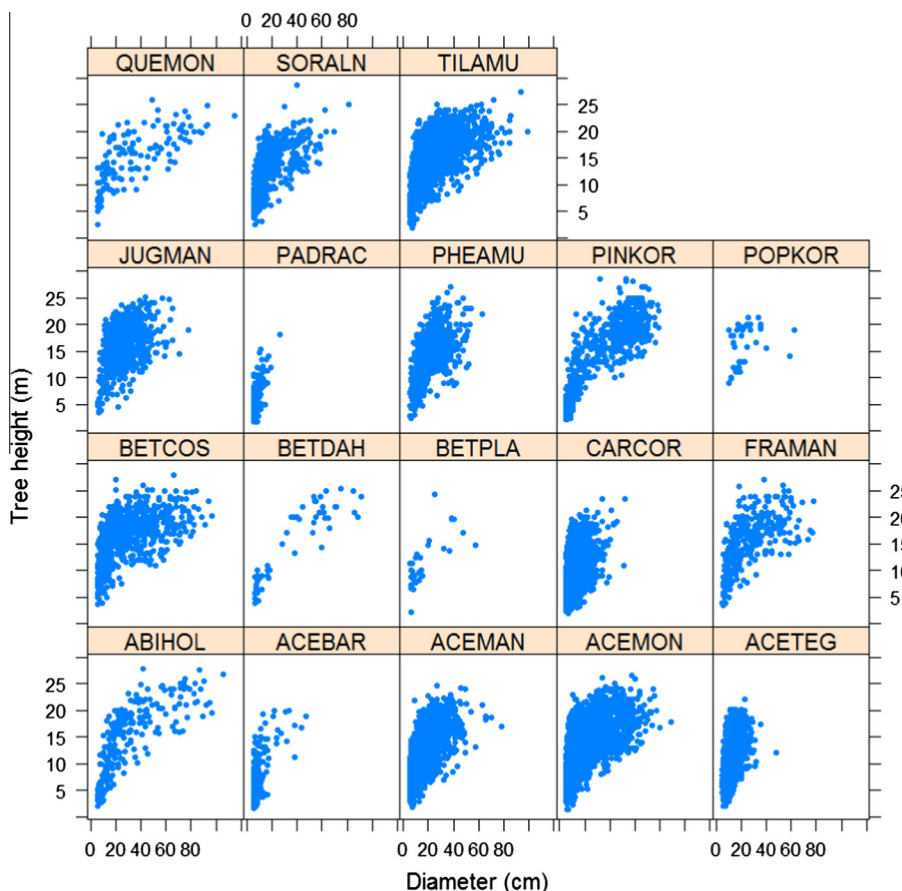

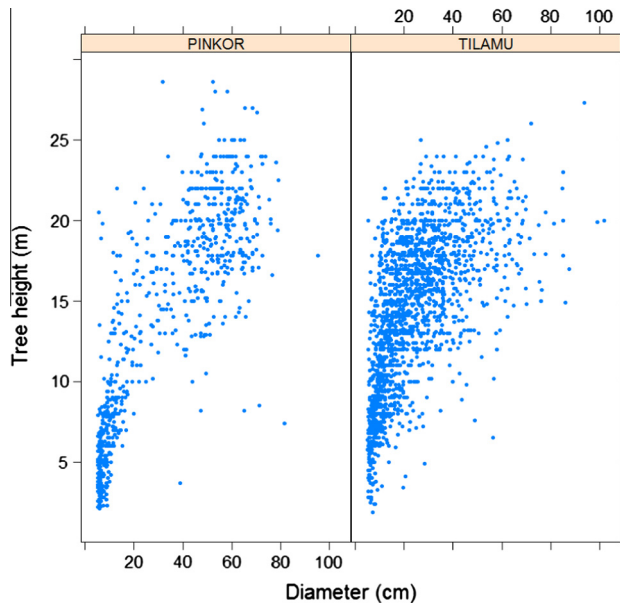**Fig. 3.** Height–diameter relations by tree species.

Fig. 4. Frequency distributions of diameter and height of *Pinus koraiensis* and *Tilia amurensis*.

BAL and BA, BAL and SPH, and Ht and BA, the RMSE values ranged from 2.11–3.66, 2.09–4.32, 1.78–3.4 m respectively (Table 7). Some of the equations examined were biased for sine species (Table 6).

All three base equations provided adequate performance across species (Tables 5 and 6). Eqs. (3) and (2) provided the lowest RMSE for most tree species, while Eqs. (1) and (4) provided the lowest RMSE for some tree species. However, Eq. (1) resulted in biased estimates for five tree species. Considering both bias and RMSE, Eq. (3) provided the lowest RMSE for most tree species, followed by Eq. (2) (Table 5). Based on these results, we chose Eq. (3) as being the best amongst the base equations.

The Chapman–Richards function (Eq. (1)) has been extensively used in describing height–age relationship. Huang et al. (1992) gave a cautionary note, however, stating that this function approaches the asymptote too quickly when there is a weak relationship between the dependent and independent variables. Hanus et al. (1999) reported similar problems when applying the Chapman–Richards function to data from young Douglas-fir stands. Similar problems were observed in this study.

3.3. Extended height–diameter equations

Biases for the expanded equations ranged from −0.1 m to +0.1 m. None of the equation was unbiased for all tree species groups (Tables 5 and 6). Examining equations that provided the lowest RMSE for a tree species (as indicated by the bolded values in Table 7) shows that Eq. (1) followed by Eqs. (11) and (12) resulted in the lowest RMSE values for most tree species. Eqs. (5), (8), and (11), extended from Eq. (2), failed to converge for 3, 4, and 5 species, respectively. Eqs. (4), (7), and (10) did not converge for one tree species. The differences in the extended models that used Ht and BA were small, While Eq. (10) provided the lowest RMSE values for most tree species, its associated base model resulted in biased estimates for five tree species. Based on these results and future application, we chose Eq. (12) as being the best amongst the extended equations to match with the selected base model (Eq. (3)).

3.4. Model comparison and selection

The inclusion of the relative position of trees and stand density measures improved the predictive abilities of most base equations examined in this study. The expanded equations reduced the lowest RMSE for a species by a minimum of 1.0% for ULMDAV and 1.3% for ACEBAR to a maximum of 31.9% for POPKOR, and 31.6% for BET-PLA. An average reduction of 12.2% was found across all tree species. For all of the species, the effect of these measures on predicted height is as expected: they increase predicted height for a decrease in relative position in the stand (as indicated by an increase in BAL) or an increase in stand density (as indicated by an increase in BA).

The best expanded equation (Eq. (13)) uses height of the tallest tree by species as measure of tree species position in multi-species and multi-layered forests, and it uses BA as a measure of stand density. Table 7 indicates that the combination of Ht and BA produced superior results more often than the combinations of BAL and BA, and BAL and SPH. By including the tallest tree height by species and BA, the RMSE decreased more than 5% for 15 of the 23 tree species (Table 5 vs. Table 7). This indicates that the Ht and BA combination captured H–D variations well in in multi-species and multi-layered forests. The extended H–D equations failed to converge for PADRAC (Table 9), and we suggest using the base equation for this species.

Parameters for base and extended height–diameter equations are provided in Tables 8 and 9. The reported values and their fit statistics provide starting values and baseline values for future modeling endeavors.

3.5. Height–diameter relationships and variation by habitat types

The experimental area was subdivided into four habitat types using a MRT method. The tree size was selected using a cross-validation procedure, with the four-leaf tree clearly identified as having the smallest cross-validated relative error (CV error = 0.906; see Appendix C). The geographical profile of the four habitat types is shown (Fig. 2). Each 20 m × 20 m cell is assigned to one specific habitat, as indicated by the numbers 1, 2, 3 and 4.

Table 4
Pearson correlation coefficients for selected variables.

Attribute	Height	Diameter	Elevation (m)	Slope	Aspect	Convexity	Basal area (m²/ha)	Stems (ha)
Diameter (cm)	0.7254							
Elevation (m)	0.0381	−0.0878						
Slope	0.0128	−0.0542	0.3086					
Aspect	0.0372	0.0223	0.0209	0.0710				
Convexity	0.0323	0.0094	0.2083	0.0289	−0.0521			
Basal area (m²/ha)	0.0908	0.1169	−0.1684	−0.0854	−0.1330	0.2093		
Stems (ha)	−0.0394	−0.1489	0.2811	0.1202	−0.2546	0.1695	0.3319	
BAL (m²/ha)	−0.3625	−0.5222	−0.1271	−0.0536	−0.0988	0.1501	0.7462	0.2629

Table 5

Bias and root mean square error (RMSE) for base height–diameter models by tree species. **Bold** indicates the lowest RMSE value across all equations for a given species. Species codes are listed in Appendix A.

Species	NTREE	Bias for base equations			RMSE (m) for base equations		
		1	2	3	1	2	3
ABIHOL	248	0.1			3.2	3.2	**3.1**
ABINEP	127				2.4	**2.4**	**2.4**
ACEBAR	789				2.2	**2.1**	**2.1**
ACEMAN	2325				2.6	**2.6**	2.6
ACEMON	2495				3.1	3.1	**3.0**
ACETEG	1401				**3.0**	**3.0**	**3.0**
BETCOS	776	0.1			3.3	3.3	**3.2**
BETDAH	55				2.3	2.3	**2.2**
BETPLA	38				**2.7**	**2.7**	**2.7**
CARCOR	4030				2.7	2.7	**2.6**
FRAMAN	372				**3.3**	**3.3**	**3.3**
JUGMAN	805	0.1			3.4	3.4	**3.3**
PADRAC	209				2.3	2.2	**2.1**
PHEAMU	795				3.2	**3.2**	3.3
PINKOR	733				2.9	2.9	**2.8**
POPKOR	39				**3.4**	**3.4**	**3.4**
QUEMON	137	0.1			3.3	**3.2**	**3.2**
SORALN	761				**2.8**	**2.8**	**2.8**
TILAMU	1776				3.2	3.2	**3.1**
TILMAN	104				**3.4**	**3.4**	**3.4**
ULMDAV	59	0.1			2.9	**2.8**	**2.8**
ULMLAC	2094				**3.0**	**3.0**	**3.0**
ULMMAC	154				2.9	2.9	**2.8**

Table 6

Bias for extended height–diameter models by tree species. Species codes are listed in Appendix A.

Species	NTREE	Base equations + BAL & BA			Base equations + BAL & SPH			Base equations + Ht & BA		
		4	5	6	7	8	9	10	11	12
ABIHOL	248									
ABINEP	127									
ACEBAR	776									
ACEMAN	2319									
ACEMON	2488									
ACETEG	1394									
BETCOS	775	n.c.			n.c.			n.c.		
BETDAH	54									
BETPLA	38							n.c.		
CARCOR	4024				n.c.					
FRAMAN	372									
JUGMAN	804						n.c.			
PADRAC	207				n.c.			n.c.	n.c.	
PHEAMU	790									
PINKOR	731									
POPKOR	39							n.c.		
QUEMON	137	n.c.			n.c.			n.c.		
SORALN	758									
TILAMU	1771									
TILMAN	102	n.c.	n.c.							
ULMDAV	59				n.c.					
ULMLAC	2088									
ULMMAC	153						n.c.			

n.c.: Indicates lack of convergence. BAL, basal area per hectare in larger trees; BA, basal area per hectare; SPH, number of trees per hectare; and Ht is tallest tree for associated species.

The topographic attributes of the four habitat types are presented in Appendix C. Habitat type 1, which occurs in 418 cells, occupies the high altitudes. It is separated from habitat type 2, 3 and 4 by the lower altitudinal boundary of 650 m. Habitat type 2 ($n = 20$) occupies the cells with convexity less than 0.7378, while

Habitat type 3 ($n = 13$) occupies the cells with convexity greater than 0.7378. Habitat type 4 ($n = 299$) are found in the westward-facing cells with aspects exceeding 135.4°, in the upper right and upper left of the plot. Habitat type 3 occupies aspects below 135.4°, while habitat type 4 occupies aspects above 135.4°.

3.6. Including habitat and species effects

Based on Table 9, the 15 tree species or species group depict different relationships among the four habitat types, with different sets of equations. There were substantial and significant differences in the indicator variable coefficients ($I2, \ldots, I15$) of the height–diameter equations among the 15 species (Table 10). For example, the indicator coefficients were negative for genus *Pinaceae*, *Fagaceae*, and species *C. cordata*. The significant variations of the indicator coefficients indicate that different H–D relationships vary among species and habitat types.

4. Discussion

4.1. Comparison of predictive performance and selection of strategies

The relative position of each tree in the population and the stand density measures used in this study are easily obtained and are available in most forest inventory databases and growth and yield models. When these variables were added to the base equation, the root mean square values were reduced by 12.2% on average. Where possible, the use of the height–diameter function with these attributes is suggested.

4.1.1. Extended models and habitat types

For over 65% of the tree species, more than 5% gain in RMSE were noted between using extended H–D equations that include stand density and relative position variables, versus using the base equations the gain in RMSE was less than 2% for five tree species (Table 5 vs. Table 7). When the difference between a base and extended equation for a given species is less than 5%, we do not suggest using the extended equations for these species.

The experimental plot was subdivided into four habitat types sharing topographic characteristics. Habitat type 1 has the largest area and occupies the low altitudes. Almost half of all the significant species–habitat associations were identified in habitat type 1, when considering both habitat and species autocorrelation. Habitat type 1 occupies 56% of the total experimental area; habitat type 4 occupies 40%. Habitat types 2 and 3 represent the transitional zones between the low and high plateau, and occupy only 5% of the area. This skewed distribution is simply an effect of the species–area relationship. However, this effect does not contradict the occurrence of species specialization.

Including both random effects and stand-level variables had a relatively small effect on the RMSE, compared with including random effects and diameter alone. This indicates that predicting a random habitat effect can be seen as an empirical surrogate for measuring stand-level variables that affect the d–h relationship (Table 10).

In this article we have not examined the presence of spatial autocorrelation of height to diameter relationships which may confuse the contribution of microsite variation to tree distribution patterns. Yet, the documented parameter estimates are unbiased and can be used for predicting tree heights as inputs to tree volume and biomass estimation for multi-species and multi-layered forests in Northeast China.

Table 7

Root mean square error (RMSE) for extended height–diameter models by tree species. **Bold** indicates the lowest RMSE value across all equations for a given species. Species codes are listed in Appendix A.

Species	NTREE	Base equations + BAL & BA			Base equations + BAL & SPH			Base equations + Ht & BA		
		4	5	6	7	8	9	10	11	12
ABIHOL	248	3.2	3.2	3.2	3.2	3.2	3.1	**2.8**	**2.8**	**2.8**
ABINEP	127	2.4	2.4	2.4	2.4	2.4	2.4	**2.3**	**2.3**	2.3
ACEBAR	776	2.1	2.1	2.1	2.1	2.1	2.1	**2.1**	**2.1**	**2.1**
ACEMAN	2319	2.6	2.6	2.6	2.6	2.6	2.6	**2.5**	**2.5**	2.6
ACEMON	2488	3.1	3.1	3.1	3.1	3.1	3.1	**2.9**	**2.9**	**2.9**
ACETEG	1394	3.0	3.0	3.0	2.9	2.9	2.9	**2.8**	2.9	2.9
BETCOS	775	3.3	n.c.	3.2	3.3	n.c.	3.3	**2.6**	nc	2.7
BETDAH	54	2.3	2.3	2.3	2.2	2.2	2.3	**1.8**	**1.8**	**1.8**
BETPLA	38	2.6	2.6	2.5	2.6	2.6	2.6	**1.9**	nc	**1.9**
CARCOR	4024	2.7	2.7	2.7	2.7	n.c.	2.7	**2.5**	2.6	2.6
FRAMAN	372	3.3	3.3	3.3	3.3	3.3	3.3	**2.7**	**2.7**	**2.7**
JUGMAN	804	3.3	3.4	3.3	3.3	3.4	n.c.	**2.7**	**2.8**	**2.7**
PADRAC	207	2.1	2.1	2.1	n.c.	2.1	2.6	n.c.	n.c.	2.1
PHEAMU	790	3.3	3.3	3.3	3.2	3.2	3.3	**2.8**	**2.8**	2.9
PINKOR	731	2.9	2.9	2.8	2.9	2.9	2.8	**2.3**	2.4	2.4
POPKOR	39	3.4	3.4	3.4	3.4	3.4	3.4	**2.3**	n.c.	2.5
QUEMON	137	3.2	n.c.	3.2	3.2	n.c.	3.2	**2.5**	n.c.	2.7
SORALN	758	2.8	2.8	2.8	2.8	2.8	4.1	**2.7**	**2.7**	**2.7**
TILAMU	1771	3.2	3.2	3.2	3.2	3.2	3.2	**2.7**	**2.7**	2.8
TILMAN	102	n.c.	n.c.	3.3	3.4	3.4	4.3	**2.8**	nc	2.9
ULMDAV	59	2.8	2.8	2.8	2.9	n.c.	2.9	**2.8**	2.9	**2.8**
ULMLAC	2088	3.0	3.0	3.0	3.0	3.0	3.0	**2.5**	2.6	2.7
ULMMAC	153	2.9	2.9	2.9	2.9	2.9	n.c.	**2.6**	**2.6**	2.7

n.c.: Indicates lack of convergence. BAL, basal area per hectare in larger trees; BA, basal area per hectare; SPH, number of trees per hectare; and Ht is tallest tree for associated species.

Table 8

Parameter estimates for base Eq. (3) (i.e., using diameter only) by species. All estimated parameters were significantly different from zero ($p < 0.005$). Species codes are listed in Appendix A.

Species	NTREE	RMSE (m)	β_0	β_1	β_2
ABIHOL	248	3.196	3.2025	−13.0064	−0.3265
ABINEP	127	2.414	3.3729	−21.4450	3.3136
ACEBAR	776	2.106	3.4795	−33.6826	11.2591
ACEMAN	2319	2.622	3.0748	−12.9432	3.1127
ACEMON	2488	3.082	2.9785	−8.0431	0.9204
ACETEG	1394	2.975	2.8659	−6.7197	−0.1822
BETCOS	775	3.263	2.9347	−3.8050	−1.8714
BETDAH	54	2.278	3.3690	−26.0821	7.6025
BETPLA	38	2.735	3.0814	−13.7241	5.3540
CARCOR	4024	2.677	2.7368	−8.0719	2.2748
FRAMAN	372	3.270	3.0357	−7.4574	−0.0158
JUGMAN	804	3.354	2.8956	−3.5637	−3.6845
PADRAC	207	2.100	5.1601	−129.5220	29.2232
PHEAMU	790	3.262	2.9929	−7.3352	−0.6310
PINKOR	731	2.859	3.1709	−14.6703	1.4694
POPKOR	39	3.397	2.9729	−4.7921	−1.2157
QUEMON	137	3.202	3.0649	−13.3138	5.9449
SORALN	758	2.821	2.9962	−9.4487	1.6561
TILAMU	1771	3.155	3.0085	−7.8296	0.0203
TILMAN	102	3.434	3.1951	−21.4130	8.7972
ULMDAV	59	2.827	2.8739	−7.3690	0.5614
ULMLAC	2088	3.030	3.0312	−9.7346	1.5838
ULMMAC	153	2.914	3.0724	−9.7447	1.1128

4.2. Final parameter estimates

After evaluating the equations, Eqs. (3) and (12) are selected, and the entire data were used to obtain parameter estimates, and documented in Tables 8 and 9. All estimated parameters are significantly different from zero ($p < 0.005$). The final parameter estimates do vary considerably between species. Graphs of residuals from the refitted equations showed approximately homogenous variances over the full range of predicted values.

The use of Eq. (16) and the coefficients listed in Table 10 will increase the precision and the resolution of the height–diameter equations. Using dummy variables to reflect species effect allow simultaneous estimation of all species, making the process compact and the "predictions more compatible" (Eerikäinen, 2009) for commonly found tree species in multi-species and multi-layered forests in NE China.

Except β_{00} and β_{01} in Table 10, all estimated parameters are significantly different from zero ($p < 0.005$). The final parameter estimates do vary considerably between species. Graphs of residuals from the refitted equations showed approximately homogenous variances over the full range of predicted values. Residual plots also indicated that tree height was well predicted across diameters.

5. Summary

Measuring tree heights is costly however, and foresters usually welcome an opportunity to estimate this variable with an acceptable accuracy. Based on a comprehensive data set which includes very small diameters, such height–diameter equations were fitted for 23 major tree species in multi-species and multi-layered forests in NE China. The inclusion of the height of tallest tree by species, relative position of a tree, and stand density variables into the base height–diameter function increased the accuracy of prediction for most species. For both conifers and hardwood tree species, the fit statistics indicated that Eqs. (3) and (12) are suitable for predicting tree heights in multi-species and multi-layered forests. The parameters estimates using Eqs. (3) and (12) will provide reasonable precision when there is no sample tree height value to estimate random coefficients.

The results of this study are sufficiently convincing to advocate the use of Eq. (16) which accounts for data structure and for the significant differences in microsite via habitat effect. Therefore, the use of Eq. (16) and the coefficients listed in Table 12 will increase the precision and the resolution of the height–diameter

Table 9
Parameter estimates for extended Eq. (12) (i.e., using dbh, height of the tallest tree by species, and basal area/ha) by tree species. All estimated parameters were significantly different from 0 ($p < 0.005$). Species codes are listed in Appendix A.

Species	NTREE	RMSE (m)	a_0	a_1	a_2	β_1	β_2
ABIHOL	248	2.7912	2.3559	0.0415	−0.0030	−11.412	−0.9835
ABINEP	127	2.3064	2.8533	0.0246	−0.0019	−19.505	2.8037
ACEBAR	776	2.0822	2.7831	0.0228	0.0021	−25.608	8.4779
ACEMAN	2319	2.5875	2.6855	0.0201	0.0001	−13.854	3.5567
ACEMON	2488	2.9160	2.4042	0.0296	−0.0009	−8.298	1.1049
ACETEG	1394	2.8936	2.3884	0.0281	−0.0024	−7.146	−0.0170
BETCOS	775	2.7266	2.1207	0.0412	−0.0008	−4.421	−1.0449
BETDAH	54	1.8255	2.7150	0.0360	−0.0052	−25.122	9.2637
BETPLA	38	1.9492	2.7342	0.0527	−0.0210	−20.670	10.1290
CARCOR	4024	2.6275	2.3044	0.0248	−0.0029	−7.396	1.7951
FRAMAN	372	2.7197	2.2323	0.0409	−0.0013	−7.637	−0.0682
JUGMAN	804	2.7250	2.0934	0.0397	−0.0004	−3.330	−4.0043
PADRAC	207	2.1183	5.1825	0.0055	−0.0001	−141.794	31.2307
PHEAMU	790	2.9292	2.3513	0.0353	−0.0017	−8.297	−0.0591
PINKOR	731	2.4255	2.3427	0.0393	−0.0007	−13.947	1.2339
POPKOR	39	2.5306	2.0365	0.0474	−0.0015	−2.353	−5.8612
QUEMON	137	2.7359	2.1243	0.0456	−0.0016	−9.435	2.7313
SORALN	758	2.7081	2.5371	0.0257	−0.0018	−9.741	1.8301
TILAMU	1771	2.8327	2.2645	0.0384	−0.0014	−7.559	−0.3600
TILMAN	102	2.9425	2.1114	0.0498	0.0027	−20.996	9.9514
ULMDAV	59	2.7789	2.4445	0.0207	−0.0007	−6.576	−0.1724
ULMLAC	2088	2.7288	2.3769	0.0352	−0.0019	−10.145	1.7886
ULMMAC	153	2.7360	2.4369	0.0364	−0.0050	−9.001	0.4887

Table 10
Estimated parameters and associated standard errors for mixed models with diameter only and with stand-level variables. Species group I (ACEBAR and ACETEG) were default species. $I2, \ldots, I15$ indicates the coefficient for each respective species or species groups listed in column 6 of Appendix A. NS indicates coefficient was not significant at $p < 0.05$ level.

Model with diameter only			Model with diameter, tree position, and stand-level variables		
Estimate	Value	Std. error	Estimate	Value	Std. error
β_0	3.060	0.007	β_0	2.98	0.014
β_1	−12.590	0.331	β_{00}	0.001 NS	0.010
β_2	2.910	0.212	β_{01}	0.001 NS	0.040
I2	3.990	0.212	β_1	−11.51	0.435
I3	−1.230	0.340	β_2	2.38	0.256
I4	−0.740	0.342	I2	3.83	0.210
I5	2.070	0.185	I3	−1.28	0.328
I6	1.280	0.371	I4	−0.74	0.327
I7	−1.060	0.283	I5	1.88	0.178
I8	0.970	0.219	I6	1.21	0.356
I9	−0.320	0.142	I7	−1.22	0.274
I10	1.850	0.227	I8	0.84	0.210
I11	3.000	0.237	I9	−0.40	0.136
I12	0.030	0.160	I10	1.83	0.219
I13	1.560	0.158	I11	3.11	0.232
I14	1.200	0.174	I12	0.05	0.152
I15	2.660	0.290	I13	1.45	0.152
σ_b	0.37		I14	1.21	0.167
σ_ε	2.39		I15	2.57	0.280
			σ_b	0.36	
			σ_ε	2.38	

equations for commonly found tree species in multi-species and multi-layered forests in NE China. While Eq. (16) is suggested for future use, additional data is needed.
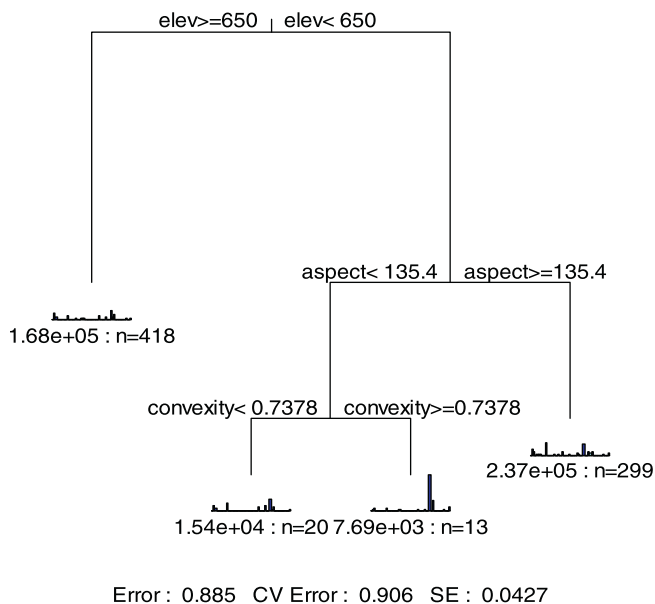
Acknowledgements

Appendix A. List of the 23 tree species and species group used in the analysis

Species name	Latin code	Family	Light (L)/ shade (S)	NTREE	Species group
Acer barbinerve	ACEBAR	Aceraceae	L	776	1
Acer tegmentosum	ACETEG	Aceraceae	L	1394	1
Acer mandshuricum	ACEMAN	Aceraceae	L	2319	12
Acer mono	ACEMON	Aceraceae	L	2488	13
Betula costata	BETCOS	Betulaceae	L	775	2
Betula dahurica	BETDAH	Betulaceae	L	54	2
Betula platyphylla	BETPLA	Betulaceae	L	38	2
Carpinus cordata	CARCOR	Betulaceae	L	4024	9
Abies holophylla	ABIHOL	Pinaceae	S	248	3
Abies nephrolepis	ABINEP	Pinaceae	S	127	3
Pinus koraiensis	PINKOR	Pinaceae	L	731	7
Quercus mongolica	QUEMON	Fagaceae	L	137	4
Padus racemosa	PADRAC	Rosaceae	L	207	4
Juglans mandshurica	JUGMAN	Juglandaceae	L	804	11
Fraxinus mandshurica	FRAMAN	Oleaceae	L	372	15
Sorbus alnifolia	SORALN	Rosaceae	M	758	8
Phellodendron amurense	PHEAMU	Rutaceae	L	790	10
Populus koreana	POPKOR	Salicaceae	L	39	5
Tilia amurensis	TILAMU	Tiliaceae	L	1771	5
Tilia mandshurica	TILMAN	Tiliaceae	L	102	5
Ulmus davidiana	ULMDAV	Ulmaceae	L	59	6
Ulmus macrocarpa	ULMMAC	Ulmaceae	L	153	6
Ulmus laciniata	ULMLAC	Ulmaceae	L	2088	14

Appendix B. Basal area and number of trees/ha by dominant tree species

Tree species	Total basal area (m²/ha)	Stems (ha)	%Basal area	%SPH
Acer mandshuricum	1.9	77.6	6.6	11.4
Acer mono	3.9	83.3	13.8	12.3
Carpinus cordata	2.3	134.6	8.0	19.9
Pinus koraiensis	2.5	24.5	8.9	3.6
Tilia amurensis	3.8	59.3	13.2	8.7
Ulmus laciniata	4.0	69.9	14.0	10.3
Total	18.4	449.1	64.5	66.3

Appendix C. Multivariate regression tree for the species composition data. Euclidean distance was used for splitting. Bar plots show the multivariate species mean at each node, and the numbers of sites are shown in parentheses

elev>=650 | elev< 650

1.68e+05 : n=418

aspect< 135.4 | aspect>=135.4

convexity< 0.7378 | convexity>=0.7378

2.37e+05 : n=299

1.54e+04 : n=20 7.69e+03 : n=13

Error : 0.885   CV Error : 0.906   SE : 0.0427

References

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.

Canham, C.D., Papaik, M.J., Uriarte, M., McWilliams, W.H., Jenkins, J.C., Twery, M.J., 2006. Neighborhood analysis of tree competition along environmental gradients in New England forests. Ecol. Appl. 16, 540–554.

Castedo, F., Dieguez-Aranda, U., Barrio, M., Sanchez, M.R., von Gadow, K., 2006. A generalized height–diameter model including random components for radiata pine plantations in northwestern Spain. For. Ecol. Manage. 229, 202–213.

De'ath, G., 2002. Multivariate regression trees: a new technique for modeling species–environment relationships. Ecology 83, 1105–1117.

De'ath, G., 2010. mvpart: multivariate partitioning. R Package Version 1.3-1. hhttp://cran.r-project.org/.

Eerikäinen, K., 2009. A multivariate linear mixed-effects model for the generalization of sample tree heights and crown ratios in the Finnish National Forest Inventory. For. Sci. 55 (6), 480–493.

Fang, Z., Bailey, R.L., 1998. Height–diameter models for tropical forests on Hainan Island in Southern China. For. Ecol. Manage. 110, 315–327.

Flewelling, J.W., de Jong, R.D., 1994. Considerations in simultaneous curve fitting for repeated height–diameter measurements. Can. J. For. Res. 24, 1408–1414.

Gallant, A.R., 1987. Nonlinear Statistical Models. John Wiley and Sons, New York, 610pp.

Garman, S.L., Acker, S.A., Ohmann, J.L., Spies, T.A., 1995. Asymptotic height–diameter equations for twenty-four tree species in Western Oregon. Res. Cont. 10, For. Res. Lab., Oregon State Univ., Corvallis, OR, 22p.

Gregoire, T.G., Schabenberger, O., Barrett, J.P., 1995. Linear modelling of irregularly spaced, unbalanced, longitudinal data from permanent plot measurements. Can. J. For. Res. 25, 137–156.

Hall, D.B., Clutter, M., 2004. Multivariate multilevel nonlinear mixed effects models for timber yield predictions. Biometrics 60 (1), 16–24.

Hanus, M.L., Marshall, D.D., Hann, D.W., 1999. Height–diameter equations for six species in the coastal regions of the Pacific Northwest. Res. Cont. 25, For. Res. Lab., Oregon State Univ., Corvallis, OR, 11p.

Harms, K.E., Condit, R., Hubbell, S.P., Foster, R.B., 2001. Habitat associations of trees and shrubs in a 50-ha neotropical forest plot. J. Ecol. 89, 947–959.

Huang, S., Titus, S.J., Wiens, D.P., 1992. Comparison of nonlinear height–diameter functions for major Alberta tree species. Can. J. For. Res. 22, 1297–1304.

Huang, S., Price, D., Titus, S.J., 2000. Development of ecological based height–diameter models for white spruce in boreal forests. For. Ecol. Manage. 129, 125–141.

Lappi, J., 1991. Calibration of height and volume equations with random parameters. For. Sci. 37 (3), 781–801.

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

Lappi, J., 2006. A multivariate, nonparametric stem-curve prediction method. Can. J. For. Res. 36, 1017–1027.

Lindstrom, M.J., Bates, D.M., 1990. Nonlinear mixed effects models for repeated measurements data. Biometrics 46, 673–687.

Mehtätalo, L., 2004. A longitudinal height-diameter model for Norway spruce in Finland. Can. J. For. Res. 34, 131–140.

Monleon, V.J., 2003. A hierarchical linear model for tree height prediction. *In*: Proceedings of the 2003 meeting of the American Statistical Association, Section on Statistics and the Environment. Alexandria, VA p. 2865–2869.

Pinheiro, J.C., Bates D.M., 2000. Mixed effects models in S and S-plus. Springer, New York, 528 pp.

Ratkowsky, D.A., 1990. In: Handbook of Nonlinear Regression. Marcel and Dekker, Inc., New York.

Richards, F.J., 1959. A flexible growth function for empirical use. J. Exp. Biol. 10, 290–300.

SAS Institute Inc., 2005, . fourth ed. SAS/STAT User's Guide, Version 9 fourth ed., vol. 2 SAS Institute, Inc., Cary, NC.

Schmidt, M., Kiviste, A., von Gadow, K., 2011. A spatially explicit height–diameter model for Scots pine in Estonia. Eur. J. For. Res. 130 (2), 303–315.

Temesgen, H., von Gadow, K., 2004. Generalized height–diameter models for major tree species in complex stands of interior British Columbia, Canada. Euro. J. For. Res. 123 (1), 45–51.

Temesgen, H., Hann, D.W., Monleon, V.J., 2007. Regional height–diameter equations for major tree species of southwest Oregon. West. J. Appl. For. 22 (3), 213–219.

Temesgen, H., Monleon, V.J., Hann, D.W., 2008. Analysis of nonlinear tree height prediction strategies for Douglas-fir forests. Can. J. For. Res. 38, 553–565.

Wykoff, W.R., Crookston, N.L., Stage, A.R., 1982. User's guide to the stand prognosis model. Gen. Tech. Rprt. INT-133. USDA For. Serv., Ogden, UT, 112pp.

Yamakura, T., Kanzaki, M., Ohkubo, T., Ogino, K., Chai, E.O.K., Lee, H.S., Ashton, P.S., 1995. Topography of a large-scale research plot established within a tropical rain forest at Lambir, Sarawak. Tropics 5, 41–56.

Yang, R.C., Kozak, A., Smith, J.H.G., 1978. The potential of Weibull-type functions as flexible growth curves. Can. J. For. Res. 8, 424–431.

Zhang, C., Zhao, Y., Zhao, X., von Gadow, K., 2012. Species–habitat associations in a northern temperate forest in China. Silva Fenn. 46 (4), 501–519.