Using Nonlinear Mixed Model Technique to Determine the Optimal Tree Height Prediction Model for Black Spruce

Shongming Huang, Shawn X. Meng & Yuqing Yang

Biometrics Unit, Forest Management Branch

Alberta Department of Sustainable Resource Development

8th Floor, 9920 -108 Street, Edmonton, Alberta T5K 2M4, Canada

Tel: 1-780-422-5281   E-mail: shongming.huang@gov.ab.ca

Abstract

Based on the nonlinear mixed model technique, four base height-diameter models were evaluated for black spruce. The Chapman-Richard model was chosen. Top height and basal area were incorporated into the base model. Comparison of the base and expanded models showed that, although the goodness-of-fit measures on the modelling data were improved with the inclusion of top height and basal area, the predictive accuracy of the expanded models at the subject-specific level where the predominant interest of nonlinear mixed models lies, was reduced when tested on the model validation data. This has important practical implications because more accurate individual tree height predictions can be better achieved using the base height-diameter model without requiring the addition of other variables. It also reaffirms that determining the adequacy of a model on model fitting statistics alone can be misleading. A fitted model is best judged on separate validation data.

Keywords: Individual tree height prediction, Height-diameter model, Nonlinear mixed model, Model validation, Local prediction, Error distribution

1. Introduction

Total tree height (H) and tree diameter at breast height (DBH) are two of the most fundamental variables in forestry. The estimation of tree volume and biomass, the description of stand conditions and their changes over time, as well as the development of growth and yield projection systems all rely heavily on the availability of a complete set of tree heights and diameters.

However, since measuring tree height is time consuming and costly, typically in various data collection programs in forest management, all trees are measured for diameter but only a portion of the trees are measured for height. Predicting the "missing" tree heights is therefore routinely required in forest research and operations (Curtis 1967, Wykoff et al. 1982, Arabatzis and Burkhart 1992, Huang et al. 1992).

Although different options exist, the most commonly used method of predicting tree heights is to develop height-diameter models. From the trees that have both height and diameter measured, a height-diameter model can be developed to express tree height as a function of tree diameter. This model can then be used to predict the "missing" heights from measured DBHs.

In numerous studies related to the development of height-diameter models, it has been almost unanimously shown that the predictive ability and accuracy of the base height-diameter models can be improved if additional (statistically significant) tree and stand level variables are incorporated into the models. These variables may include tree or stand age, basal area per hectare (ha), stems per ha, crown ratio, crown competition factor, site index, dominant/co-dominant height, top height, crown class, wind speed, and different competition indices and species composition measures (e.g., Curtis 1967, Huang and Titus 1994, Huang 1999, Eerikäinen 2003, Temesgen et al. 2008, Meng et al. 2008).

The conclusions reached in previous studies were mostly derived from the goodness-of-fit measures on the modelling data, which, as will be shown later, could be unreliable and even misleading sometimes if one is not careful. Furthermore, in the cases where the nonlinear mixed model (NLMM) technique was applied (Castedo Dorado et al. 2006, Sharma and Parton 2007, Temesgen et al. 2008, Meng et al. 2008), no frequency-based comparison of the predictive performance between the base and expanded models was made on the plot-specific level where the main interest of NLMMs lie.

The prime objectives of this study were to: (1) select the optimal base height-diameter model for black spruce (*Picea mariana* (Mill.) B.S.P.); (2) compare the predictive accuracy of the base model with that of the expanded models on both the model fitting and validation data sets; and (3) evaluate the impacts of including or not including statistically

significant stand level variables into the base model. The selection, comparison and evaluation were all conducted within the NLMM framework.

2. Material and methods

2.1 Data

The black spruce data used in this study were collected by the Alberta Forest Service as a part of the provincial Permanent Sample Plots (PSP) database. The data encompass three key forest ecoregions (also called subregions) in Alberta: the lower foothills, the upper foothills, and the boreal mixedwood. The PSPs range in sizes from 200 to 2000 m^2, with the most common one being 1000 m^2. Within each plot or subplot (usually a quarter of the full PSP), diameters (always at a breast height of 1.3 m above ground) were measured for all trees taller than 1.3 m, and heights were measured for approximately 10 to 20% of the representative trees in a random or systematic fashion. A detailed description of the data collection procedures is provided in Alberta Forest Service (2000).

To avoid the potential serial correlation from repeated measurements of the same PSPs, only the initial measurements were retained in this study, which resulted in a total of 164 plots (1703 trees). The data were divided into two independent data sets by ecoregion: 93 plots (917 trees) from the lower foothills ecoregion were used as modeling data, and the other 71 plots (786 trees) from the boreal mixedwood and upper foothills ecoregions were used as model validation data. For the reasons discussed in Huang et al. (2002), we purposely avoided "random splitting" of the data. Summary statistics for the tree level variables H and DBH, and the plot level variables top height (i.e., mean height of the 100 largest DBH trees per ha) and stand basal area, are listed in Table 1. The data are also shown in Figure 1.

2.2 Specification of base height-diameter models

Based on a detailed examination of the height-diameter data shown in Figure 1, a large number of base height-diameter models were examined in preliminary analyses. Four promising candidate models were selected for further evaluation:

(1) $H = 1.3 + \exp(\beta_1 + \beta_2/(DBH + 1))$

(2) $H = 1.3 + \beta_1 DBH^{\beta_2}$

(3) $H = 1.3 + \beta_1/(1 + \exp[\beta_2 + \beta_3 \ln(DBH + 1)])$

(4) $H = 1.3 + \beta_1[1 - \exp(-\beta_2 DBH)]^{\beta_3}$

where H is total tree height (m), DBH is tree diameter (cm) at breast height, and β_1, β_2 and β_3 are model parameters. These models and their slight modifications have been ranked favorably among alternative models in other relevant studies, particularly the Chapman-Richards model (4) (e.g., Wykoff et al. 1982, Huang et al. 1992, Sharma and Parton 2007, Temesgen et al. 2008).

Following the NLMM technique (Davidian and Giltinan 1995, Vonesh and Chinchilli 1997), one or more parameters in the above models were assumed to be mixed (i.e., they consist of a fixed component and a random component). Likelihood ratio tests were conducted to determine if this assumption was true (at $\alpha = 0.05$ throughout this study). It was found that the two parameters in (1) and (2) were both mixed, but parameter β_2 in (3) and (4) was fixed. Hence, the NLMM formulations of the above base models are:

(5) $H_{ij} = 1.3 + \exp((\beta_1 + b_{1i}) + (\beta_2 + b_{2i})/(DBH_{ij} + 1)) + \varepsilon_{ij}$

(6) $H_{ij} = 1.3 + (\beta_1 + b_{1i})DBH_{ij}^{(\beta_2 + b_{2i})} + \varepsilon_{ij}$

(7) $H_{ij} = 1.3 + (\beta_1 + b_{1i})/(1 + \exp[\beta_2 + (\beta_3 + b_{2i})\ln(DBH_{ij} + 1)]) + \varepsilon_{ij}$

(8) $H_{ij} = 1.3 + (\beta_1 + b_{1i})[1 - \exp(-\beta_2 DBH_{ij})]^{(\beta_3 + b_{2i})} + \varepsilon_{ij}$

where H_{ij} and DBH_{ij} are observed height and DBH for the *j*th tree in the *i*th plot (subject), $i = 1, 2, \ldots m$, $j = 1, 2, \ldots n_i$, m is the total number of plots in the population, n_i is the number of trees in plot *i*, β_1, β_2 and β_3 are fixed parameters common to all plots, b_{1i} and b_{2i} are random parameters specific to plot *i*, and ε_{ij} is a normally distributed within-plot error term. Models (5)-(8) can be expressed in matrix forms. For example, the Chapman-Richards model [8] can be written for plot *i* as

(9) $\mathbf{H}_i = 1.3 + (\beta_1 + b_{1i})[1 - \exp(-\beta_2 \mathbf{DBH}_i)]^{(\beta_3 + b_{2i})} + \boldsymbol{\varepsilon}_i$

where $\mathbf{H}_i$ and $\mathbf{DBH}_i$ are vectors of observed heights and DBHs, respectively, for the *i*th plot with n_i observations, and $\boldsymbol{\varepsilon}_i = [\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{in_i}]'$ is a vector of the within-plot errors.

2.3 Expanded models

Individual tree height-diameter relationship could be affected by many factors. Of which, site quality and stand density were found to be most common (Huang and Titus 1994, Sharma and Parton 2007, Meng et al. 2008). Typically, site quality is represented by site index or dominant/co-dominant height (H_{dc}), with the latter appeared more common, likely due to the fact that site index is often unobservable and must be predicted from site index models, which could bring in

additional prediction errors and artificially reduce the variation in site quality. The use of H_{dc}, however, is not without its problems. The definition of "dominant"/"co-dominant" is not always clear-cut in practice, and how many "dominants" and/or "co-dominants" are needed to get a stable H_{dc} for a stand is unclear. Nicholas at al. (1991) found that the reliability of tree crown position classification was very low, and "only 38% of dominant trees and 81% of co-dominant trees were similarly re-classified" during visits in the same growing season.

In this study top height was used as an indicator of site quality. Top height is defined as the average height of the 100 largest DBH trees per ha. Given a plot size, a fixed number of the largest DBH trees corresponding to the plot size were selected and the average height of these trees was obtained to represent the site quality of the plot. The use of top height removed the inconsistencies associated with the use of H_{dc}.

For stand density, basal area per ha (BA, m^2/ha) was used. Basal area per ha is a simple and objective measure of the degree of crowding within a stand. It is generally better than the stand density measure based on trees per ha because basal area combines tree number with tree size.

Top height and basal area per ha were incorporated into the base H-DBH models. Different functional forms and various combinations of the variables and their transformations were tried and evaluated based on the standard NLMM goodness-of-fit measures such as -2 times log-likelihood (-2LL), Akaike Information Criterion, Schwarz's Bayesian Information Criterion, significance of the parameters, and studentized residual plots (Littell et al. 2006). It was found that simple linear combinations were adequate. As examples, two expanded models, one with top height incorporated and the other with both top height and BA incorporated, were formed based on the Chapman-Richards base model (9):

(10) $\mathbf{H}_i = 1.3 + (\beta_1 + \beta_4 \text{TopH}_i + b_{1i})[1 - \exp(-\beta_2 \mathbf{DBH}_i)]^{(\beta_3 + b_{2i})} + \varepsilon_i$

(11) $\mathbf{H}_i = 1.3 + (\beta_1 + \beta_4 \text{TopH}_i + \beta_5 \text{BA}_i + b_{1i})[1 - \exp(-\beta_2 \mathbf{DBH}_i)]^{(\beta_3 + b_{2i})} + \varepsilon_i$

where TopH_i is top height (m), BA_i is basal area per ha (m^2/ha), and β_1-β_5 are fixed parameters common to all plots, and b_{1i} and b_{2i} are random parameters specific to the ith plot. Attempts to relate β_4 and β_5 to additional random parameters in linear or nonlinear fashions failed to achieve convergence.

2.4 Parameter estimation

All base and expanded models can be written as a generalized NLMM of the form:

(12) $\mathbf{y}_i = f(\mathbf{x}_i, \boldsymbol{\beta}, \mathbf{b}_i) + \boldsymbol{\varepsilon}_i$

where $\mathbf{y}_i = [y_{i1}, y_{i2}, ..., y_{in_i}]'$ is a vector of tree height measurements in plot i, $\mathbf{x}_i$ is a known design matrix of the covariates (DBH, and where applicable also TopH and/or BA), $\boldsymbol{\beta}$ is a vector of fixed parameters, $\mathbf{b}_i = [b_{1i}, b_{2i}]'$ is a vector of random parameters, and $\boldsymbol{\varepsilon}_i = [\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{in_i}]'$ is the error vector. The $\mathbf{b}_i$ and $\boldsymbol{\varepsilon}_i$ are assumed to be uncorrelated and normally distributed with mean zero and variance-covariance matrices $\mathbf{D}$ and $\mathbf{R}_i$, respectively, that is:

$$E\begin{bmatrix} \mathbf{b}_i \\ \boldsymbol{\varepsilon}_i \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \qquad Var\begin{bmatrix} \mathbf{b}_i \\ \boldsymbol{\varepsilon}_i \end{bmatrix} = \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_i \end{bmatrix}$$

Furthermore, since no obvious pattern of unequal error variance was detected from studentized residual plots, which is typical of height-diameter models *after* the inclusion of random parameters (e.g., Castedo Dorado et al. 2006, Meng et al. 2008), $\mathbf{R}_i$ is assumed to be $\sigma^2 \mathbf{I}_{n_i}$, where σ^2 is the error variance and $\mathbf{I}_{n_i}$ is a $n_i \times n_i$ identity matrix, and $\mathbf{D}$ is assumed to be an unstructured covariance matrix that is the same for all i:

(13) $\mathbf{D} = \begin{bmatrix} \sigma_{b_1}^2 & \sigma_{b_1 b_2} \\ \sigma_{b_1 b_2} & \sigma_{b_2}^2 \end{bmatrix}$

where $\sigma_{b_1}^2$ and $\sigma_{b_2}^2$ are the variances for random parameters b_1 and b_2, respectively, and $\sigma_{b_1 b_2}$ is the covariance between b_1 and b_2.

The parameters of the generalized NLMM (12) can be estimated using different methods (Davidian and Giltinan 1995, Vonesh and Chinchilli 1997). In this study we used the first-order (FO) method of Beal and Sheiner (1982) and the first-order conditional expectation (FOCE) method of Lindstrom and Bates (1990). We found that the FOCE method often failed to achieve convergence, and was less stable and more prone to floating errors than the FO method. In the cases where both FO and FOCE achieved convergence, the FO method was found to give more accurate predictions with smaller mean error and error variance. Therefore, we limited ourselves in this analysis to the FO method.

The FO method employs a first-order Taylor series expansion of (12) around a $\boldsymbol{\beta}^*$ close to $\boldsymbol{\beta}$ and a $\mathbf{b}_i^*$ set to zero, the expected value of the random-parameters (i.e., $\mathbf{b}_i^* = \text{E}(\mathbf{b}_i) = \mathbf{0}$), to approximate the nonlinear function (12), with the quadratics and cross-products dropped:

(14) $\mathbf{y}_i \approx f(\mathbf{x}_i, \boldsymbol{\beta}^*, \mathbf{b}_i^*) + \mathbf{X}_i(\boldsymbol{\beta} - \boldsymbol{\beta}^*) + \mathbf{Z}_i(\mathbf{b}_i - \mathbf{b}_i^*) + \boldsymbol{\varepsilon}_i$

where the derivative matrices $\mathbf{X}_i$ and $\mathbf{Z}_i$ are defined by

(15) $\mathbf{X}_i = \dfrac{\partial f(\mathbf{x}_i, \boldsymbol{\beta}, \mathbf{0})}{\partial \boldsymbol{\beta}'}\bigg|_{\boldsymbol{\beta}^*, \mathbf{0}}$ $\mathbf{Z}_i = \dfrac{\partial f(\mathbf{x}_i, \boldsymbol{\beta}, \mathbf{0})}{\partial \mathbf{b}_i'}\bigg|_{\boldsymbol{\beta}^*, \mathbf{0}}$

Since $\mathbf{b}_i^* = \mathbf{0}$, if we define a pseudo-response function $\mathbf{y}_i^*$ as: $\mathbf{y}_i^* = \mathbf{y}_i - f(\mathbf{x}_i, \boldsymbol{\beta}^*, \mathbf{0}) + \mathbf{X}_i \boldsymbol{\beta}^*$, eq. (14) can be written as a standard linear mixed model:

(16) $\mathbf{y}_i^* = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i$

Following the linear mixed model theory, the generalized least squares estimator $\hat{\boldsymbol{\beta}}$ of the fixed parameters $\boldsymbol{\beta}$ in (16) can be obtained as described in Fitzmaurice et al. (2004). The random parameters predictor $\hat{\mathbf{b}}_i$ of the $\mathbf{b}_i$ is (Davidian and Giltinan 1995, Vonesh and Chinchilli 1997):

(17) $\hat{\mathbf{b}}_i = \hat{\mathbf{D}}\mathbf{Z}_i'(\mathbf{Z}_i\hat{\mathbf{D}}\mathbf{Z}_i' + \hat{\mathbf{R}}_i)^{-1}[\mathbf{y}_i - f(\mathbf{x}_i, \hat{\boldsymbol{\beta}}, \mathbf{0})]$

where $\hat{\mathbf{D}}$ and $\hat{\mathbf{R}}_i$ are estimates of $\mathbf{D}$ and $\mathbf{R}_i$, respectively, and $\mathbf{Z}_i$ is defined in (15).

It is worthwhile to point out that eq. (17) is adequate only for the FO method. Had the FOCE method been used, we must use the following equation to predict the $\hat{\mathbf{b}}_i$ and solve it iteratively (Lindstrom and Bates 1990, Schabenberger 1994, Huang 2008, Temesgen et al. 2008, Meng et al. 2008):

$$\hat{\mathbf{b}}_i = \hat{\mathbf{D}}\mathbf{Z}_i'(\mathbf{Z}_i\hat{\mathbf{D}}\mathbf{Z}_i' + \hat{\mathbf{R}}_i)^{-1}[\mathbf{y}_i - f(\mathbf{x}_i, \hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}_i) + \mathbf{Z}_i\hat{\mathbf{b}}_i]$$

Unfortunately, many forest modellers used eq. (17) to predict the $\hat{\mathbf{b}}_i$ while implementing the FOCE method, without realizing that this could seriously bias the results.

All base and expanded NLMMs were estimated using the SAS macro %NLINMIX, with the keyword EXPAND = ZERO for the FO method (Littell et al. 2006).

2.5 Model prediction

Since model fitting statistics alone can sometimes be unreliable, a fitted model is more appropriately judged through cross-validation or using separate model validation data (Picard and Cook 1984, Huang et al. 2002, Yang et al. 2004). To evaluate the predictive accuracy of the base and expanded H-DBH models fitted in this study on the validation data at plot- or subject-specific level where the main interest of NLMMs lies, the random parameters $\hat{\mathbf{b}}_i$ must be predicted first based on eq. (17). Once the $\hat{\mathbf{b}}_i$ are available, predictions for $\mathbf{y}_i$ are obtained by eq. (18), which is derived by equating (16) to the pseudo function $\mathbf{y}_i^*$ (Davidian and Giltinan 1995, Vonesh and Chinchilli 1997):

(18) $\hat{\mathbf{y}}_i = f(\mathbf{x}_i, \hat{\boldsymbol{\beta}}, \mathbf{0}) + \mathbf{Z}_i\hat{\mathbf{b}}_i$

The errors (residuals) associated with the predictions are

$$\mathbf{e}_i = \mathbf{y}_i - \hat{\mathbf{y}}_i \qquad \text{or} \qquad e_{ij} = y_{ij} - \hat{y}_{ij}$$

where y_{ij} and $\hat{y}_{ij}$ are the jth observed and predicted values for subject i with n_i observations. For a population of N observations from m subjects, where N = $\sum_{i=1}^{m} n_i$, the overall mean bias ($\bar{e}$) is

(19) $\bar{e} = \dfrac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i} e_{ij} = \dfrac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \hat{y}_{ij})$

Since in practice, many researchers (e.g., Fang and Bailey 2001, Calama and Montero 2004) have used the fixed parameters estimated as a part of a NLMM to represent the "typical" population-averaged responses, we also calculated such "typical" responses:

(20) $\hat{\mathbf{y}}_{i\_fix} = f(\mathbf{x}_i, \hat{\boldsymbol{\beta}}, \mathbf{0})$

where $\hat{\mathbf{y}}_{i\_fix}$ are the "typical" population-averaged responses based on the fixed parameters $\hat{\boldsymbol{\beta}}$ only (with the $\hat{\mathbf{b}}_i$ set to zero). The associated "typical" population-averaged residuals are:

(21) $\mathbf{e}_{i\_fix} = \mathbf{y}_i - \hat{\mathbf{y}}_{i\_fix}$

In addition to the overall mean bias averaged over all observations from all subjects (as in eq. (19)), which could be misleading sometimes because the positive and negative errors from individual subjects could cancel one another out, the mean bias (or simply bias) for each individual subject was also calculated one plot at a time, together with the standard deviation (SD):

(22) $\bar{e}_i = \dfrac{1}{n_i}\sum_{j=1}^{n_i} e_{ij} = \dfrac{1}{n_i}\sum_{j=1}^{n_i}(y_{ij} - \hat{y}_{ij})$

(23) $SD_i = \sqrt{\dfrac{1}{n_i}\sum_{j=1}^{n_i}(e_{ij} - \bar{e}_i)^2}$

where $\bar{e}_i$ is the bias for subject i with n_i (or n_{ij}) observations, and all other variables are as defined before. Two other conventional goodness-of-fit statistics were also calculated by plot (subject). One is the percent bias (Bias% or $\bar{e}_i\%$) and the other is the root mean square error (RMSE):

(24) $\text{Bias\%}_i = \dfrac{\bar{e}_i}{\bar{y}_i} \times 100\%$

(25) $\text{RMSE}_i = \sqrt{\dfrac{1}{n_i} \sum\limits_{j=1}^{n_i} (y_{ij} - \hat{y}_{ij})^2}$

where $\bar{y}_i$ is the average of the observed values for subject i, and all others are as defined before.

Once the values of $\bar{e}_i$, $\text{Bias\%}_i$ and RMSE_i were calculated for each of the 71 plots of the validation data, their frequency distributions were examined and compared to assess the goodness-of-prediction of the base and expanded height-diameter models on the validation data.

For a comparison, the biases for each of the 93 plots of the modelling data were also calculated following the same procedures outlined above and their frequency distributions were also examined.

2.6 Maximum height-diameter relationship

Subject-specific local predictions obtained from the NLMM technique attempt to "best" mimic the local data. Sometimes, due to the scarcity and the limited range of the data, as well as the flexibility of the specified model, local predictions may be unrealistic when extrapolated beyond the original data range (especially when more covariates are present). This is illustrated in Figure 2. To prevent the potential extrapolation errors due to data limitation, a maximum height-diameter curve was developed for black spruce based on all PSP data, plus the stem analysis data from sectioned trees (Huang et al. 1992). The combined data of 14,345 trees (Figure 2) were grouped by a 2-cm DBH class and the tallest tree in each DBH class was used to fit the Chapman-Richards model (4) by the ordinary nonlinear least squares method. This produced the "mean" maximum height-diameter curve for black spruce:

(26) $H_{\max} = 1.3 + 34.3451[1 - \exp(-0.04892\text{DBH})]^{0.9042}$

which, if necessary, would be used to constrain the height predictions beyond the observed data range. In the cases where negative predictions are obtained, which is possible for the FO method since eq. (18) is used to predict the response variable, negative values are set to zero to be biologically meaningful.

3. Results and discussion

3.1 Choice of the base and expanded models

Table 2 lists the parameter estimates and relevant fit statistics for the base and expanded models, where all fixed parameters are highly significant at $\alpha = 0.05$. The Akaike Information Criterion (AIC) and the Schwarz's Bayesian Information Criterion (BIC) are computed by

(27) $\text{AIC} = -2LL + 2P \qquad \text{BIC} = -2LL + P\ln(m)$

where -2LL is -2 times the log-likelihood function of the model, P is the total number of effective parameters (includes fixed parameters, variance-covariance components of the random parameters, plus the residual variance component), and m is the number of effective subjects.

Among the four base models, the Chapman-Richards model (9) has the smallest AIC and BIC values (Table 2), suggesting (9) is the most reasonable base model for black spruce. Therefore, only (9) is expanded to include top height and/or basal area. The two expanded models (10) and (11) have smaller AIC and BIC values than their counterparts from (9), suggesting (10) and (11) are better than the Chapman-Richards base model (9).

Between the two expanded models, model (11), which included both top height and basal area, has smaller AIC and BIC values (Table 2). Hence, model (11) is considered better than model (10). This is expected because (10) included top height only. The estimated fixed parameters related to top height (β_4) and basal area (β_5) are both positive, indicating that in general, site quality and stand density have positive impacts on black spruce height growth.

Following the standard model ranking and selection practice in regression analysis, models (9), (10) and (11) would be ranked 3, 2 and 1 (best), respectively. Model (11) would be selected as the best model among the three models. The base model (9) would be considered as the worst model.

3.2 Predictive performance of the models

To assess the predictive performance of the models in real-world applications, height predictions from the base and expanded models were made based on the validation data (as well as on the modeling data for a comparison). To illustrate the computations involved, an example plot was used. This plot has three height-diameter observations. They are listed in Table 3 and shown in Figures 3(a)-3(b).

The height predictions based on the fixed parameters only are obtained directly from the models, with the random parameters set to zero. For instance, for the base model (9):

$$\hat{H}_{i\_fix} = 1.3 + \hat{\beta}_1[1 - \exp(-\hat{\beta}_2 \mathbf{DBH}_i)]^{\hat{\beta}_3}$$

where the estimated fixed parameters $\hat{\beta}_1$ - $\hat{\beta}_3$ are given in Table 2. For the FO method, the derivatives of (9) with respect to the two random parameters are calculated by

$$\text{der\_b}_1 = \partial f(\mathbf{x}_i, \hat{\boldsymbol{\beta}}, \mathbf{0})/\partial b_1 = [1 - \exp(-\hat{\beta}_2 \mathbf{DBH}_i)]^{\hat{\beta}_3}$$

$$\text{der\_b}_2 = \partial f(\mathbf{x}_i, \hat{\boldsymbol{\beta}}, \mathbf{0})/\partial b_2 = \hat{\beta}_1 \ln[1 - \exp(-\hat{\beta}_2 \mathbf{DBH}_i)][1 - \exp(-\hat{\beta}_2 \mathbf{DBH}_i)]^{\hat{\beta}_3}$$

The calculated results at the given DBHs are listed in Table 3. They constitute the $\mathbf{Z}_i$ matrix defined in (15). In addition, from Table 2, we also know the $\hat{\mathbf{D}}$ and $\hat{\mathbf{R}}_i$ matrices. Therefore:

$$\mathbf{Z}_i = \begin{bmatrix} 0.286 & -8.475 \\ 0.371 & -8.709 \\ 0.596 & -7.306 \end{bmatrix} \quad \hat{\mathbf{D}} = \begin{bmatrix} 10.7688 & 0.2834 \\ 0.2834 & 0.02546 \end{bmatrix} \quad \hat{\mathbf{R}}_i = \begin{bmatrix} 1.2931 & 0 & 0 \\ 0 & 1.2931 & 0 \\ 0 & 0 & 1.2931 \end{bmatrix}$$

Having the $\hat{\mathbf{H}}_{i\_fix}$, $\mathbf{Z}_i$, $\hat{\mathbf{D}}$ and $\hat{\mathbf{R}}_i$ values, the random parameters can be predicted from (17):

$$\hat{\mathbf{b}}_i = \hat{\mathbf{D}}\mathbf{Z}_i^{'}(\mathbf{Z}_i\hat{\mathbf{D}}\mathbf{Z}_i^{'} + \hat{\mathbf{R}}_i)^{-1}(\mathbf{H}_i - \hat{\mathbf{H}}_{i\_fix}) = [-2.0180, 0.06947]'$$

Once the $\hat{\mathbf{b}}_i$ are known, the predicted heights are obtained by (18): $\hat{\mathbf{H}}_i = \hat{\mathbf{H}}_{i\_fix} + \mathbf{Z}_i\hat{\mathbf{b}}_i$. Results are listed in Table 3, along with the associated errors. The standard deviation of the errors and an overall accuracy measure (δ) combining bias and precision were also calculated (Cochran 1977):

(28) $\delta = (\text{mean bias})^2 + (\text{standard deviation of the errors})^2$

Height predictions from the other models were obtained in a similar manner. Results are also listed in Table 3. Height prediction trajectories for all 71 plots of the validation data across the possible predictive range where future predictions are likely to be made, are shown in Figures 3(c)-3(d). These trajectories, sometimes referred to as "spaghetti plots" or "chow-mein plots" (Huang 2008), can be particularly useful in assessing the predictive behaviors of NLMMs within and beyond the observed data range. All trajectories for black spruce (Figure 3) were found to be within the bounds defined by the maximum H-DBH curve shown in Figure 2, rendering the constraint unnecessary for this data set.

The results in Table 3 suggest that all three models over-predicted the height for the example plot (i.e., the mean biases are all negative). Among the three models, the expanded model (11) produced the largest bias (-0.633) and δ (1.156), whereas the base model (9) produced the smallest bias (-0.347) and δ (0.748). Note that this is completely opposite to the ranking of the models based on the goodness-of- fit measures on the modelling data (Table 2), where the expanded model (11) was the best and the base model (9) was the worst.

Following the same procedure illustrated for the example plot, random parameters were calculated and height predictions were made for each of the 71 plots of the validation data. In addition, "typical" population-averaged (PA) height predictions were also made using the estimated fixed parameters only given in Table 2. For the purpose of comparison, similar predictions were made for each of the 93 plots of the modelling data. Summary statistics of the prediction biases from different models, by different methods, and on different data sets are provided in Table 4.

For the PA predictions based on the fixed parameters only, the expanded models are more accurate with smaller values of δ and Bias% than those from the base model (Table 4). This implies that if we use the fixed parameters only to make the PA predictions, the inclusion of additional variables will improve the accuracy of the predictions. This is consistent with the traditional least squares regression, in which only fixed parameters are present, and additional variables may need to be incorporated into the base model to explain additional variations that may arise from the between-subject differences in, for instance, site quality and stand density.

However, under the NLMM framework, or whenever the NLMM technique is applied, subject-specific local predictions become the predominant focus. The fixed parameters estimated as a part of a NLMM do not fully characterize the mean responses of the population, especially when the inter-subject variability is large (Hu et al. 1998, Davidian and Giltinan 2003, Fitzmaurice et al. 2004, Young et al. 2007). As a consequence, they generally provide a biased partial representation of the true population mean responses, which, unfortunately, is hardly ever recognized in forest modelling, where many forest modellers used the predictions obtained from the fixed parameters only to represent the "unbiased" PA responses. Corrections (or separate fits by nonlinear least squares) are necessary to obtain the unbiased PA predictions unless it can be shown that the biases are of little or no practical consequence (Chiswell and Monahan 2004, Kjellsson et al. 2004, Funatogawa and Funatogawa 2007, Huang 2008). The simplest and the most effective correction procedure for obtaining the unbiased PA predictions under the NLMM framework can be derived based on the ratio estimate method of Hartley and Ross (1954), which is frequently employed in sample surveys to obtain unbiased estimates of the population and subpopulation means. Since a correctly specified and fitted NLMM possesses the ability to closely follow the population-averaged and subject-specific trends exhibited by the data, the proportionality between the observed and predicted values can be assumed and the ratio estimate method can be used to adjust the predictions. Various ratio estimators can be formed of which the most common one, when calculated for the entire population based on the fixed parameters only, is

(29) $\quad r = \dfrac{\bar{y}}{\bar{\hat{y}}} = \left(\dfrac{1}{N} \sum\limits_{i=1}^{m} \sum\limits_{j=1}^{n_i} y_{ij} \right) \Big/ \left(\dfrac{1}{N} \sum\limits_{i=1}^{m} \sum\limits_{j=1}^{n_i} \hat{y}_{ij\_fix} \right) = \sum\limits_{i=1}^{m} \sum\limits_{j=1}^{n_i} y_{ij} \Big/ \sum\limits_{i=1}^{m} \sum\limits_{j=1}^{n_i} \hat{y}_{ij\_fix}$

where $\bar{y}$ and $\bar{\hat{y}}$ are the grand mean of the observed (y_{ij}) and predicted ($\hat{y}_{ij\_fix}$) values, respectively. The r defined in eq. (29) follows that: 1) if $\bar{\hat{y}} = \bar{y}$, r = 1, indicating unbiased prediction; 2) if $\bar{\hat{y}} > \bar{y}$, r < 1, indicating over-prediction; and 3) if $\bar{\hat{y}} < \bar{y}$, r > 1, indicating under-prediction. Cochran (1977) showed that the expectation:

$$E\left(\bar{y} - r \cdot \bar{\hat{y}} \right) = 0$$

Therefore, in the case of over-prediction (r < 1), the ratio estimate will adjust the predicted values downward by the ratio $\bar{y} / \bar{\hat{y}}$ to remove the bias. When under-prediction occurs (r > 1), the ratio estimate will adjust the predicted values upward by the same ratio $\bar{y} / \bar{\hat{y}}$. In both cases, the adjusted (unbiased) predictor $\hat{y}_{ij\_adj}$, can be written as

$$\hat{y}_{ij\_adj} = r \cdot \hat{y}_{ij\_fix}$$

Once the adjusted predictor is obtained, it can be used to generate an unbiased PA curve for the population. For the base model (9) fitted in this study, this unbiased PA curve is given by

$$\hat{\mathbf{H}}_{i\_adj} = r(1.3 + \hat{\beta}_1 [1 - \exp(-\hat{\beta}_2 \mathbf{DBH}_i)]^{\hat{\beta}_3})$$

where $\hat{\mathbf{H}}_{i\_adj}$ are the adjusted unbiased PA predictions, and r is ratio estimate defined in eq. (29).

For subject-specific predictions, the results in Table 4 suggest that the overall mean biases are much smaller than those for the PA predictions. This is expected because the PA predictions were based on the fixed parameters only, whereas each plot in the subject-specific predictions has its own localized curve obtained from both fixed and random parameters.

The results in Table 4 also suggest that for subject-specific predictions, the mean bias and the standard deviation of the errors from the base model (9) are smaller than those from the expanded models. The bias (-0.023) and the SD (0.870) from the base model (9) are the smallest, whereas the bias (-0.068) and the SD (0.903) from the expanded model (11) are the largest. The values of the overall accuracy (δ) are 0.757, 0.798 and 0.820 for models (9), (10) and (11), respectively, indicating that as the number of variables incorporated into the base model increases, the accuracy of the predictions decreases. A similar trend is also observed on the modeling data (Table 4). Clearly, the height predictions from the base model are the most accurate. This is again entirely opposite to the ranking of the models based on the model fitting statistics. It is unexpected because the expanded models, which included one additional variable (top height) in (10), and two additional variables (top height and basal area) in (11), were supposed to provide incrementally more accurate predictions with smaller biases and SDs (to offset the incremental time and costs associated with measuring the additional variables).

For subject-specific local predictions, the unbiasedness of the predictions is not guaranteed because, depending on many factors (e.g., number of subjects, inter-subject variability, observations per subject, approximation process, model specification, errors-in-variables, assumptions, etc.), a NLMM provides varying degrees of goodness-of-fits to individual subjects within the population (Huang 2008). In fact, the biases for many subjects of an appropriately fitted NLMM could still be substantial (e.g., |Bias%| exceeding ±2.5 or ±5.0% of the observed mean, see Tables 5-6). To ensure the unbiasedness of subject- specific local predictions, a ratio estimate can also be calculated for each subject within the population:

(30) $\quad r_i = \dfrac{\bar{y}_i}{\bar{\hat{y}}_i} = \sum\limits_{j=1}^{n_i} y_{ij} \Big/ \sum\limits_{j=1}^{n_i} \hat{y}_{ij}$

where $\bar{y}_i$ is the mean of y_{ij} for subject i and $\bar{\hat{y}}_i$ is the mean of $\hat{y}_{ij}$ predicted from eq. (18). The calculated r_i can be used to adjust the $\hat{y}_{ij}$ to obtain unbiased predictions for subject i:

$$\hat{y}_{ij\_adj} = r_i \cdot \hat{y}_{ij}$$

Essentially, both the population-based and subject-specific correction procedures adjust the predicted values from a NLMM downward or upward by the calculated ratio estimates, such that the averages of the residuals (i.e., mean biases) are zero on both the population and subject-specific levels. To avoid a digression from the main objectives of this study, we did not demonstrate the correction procedures in the present study. Interested readers can find more details elsewhere (e.g., Huang 2008).

3.3 Distributions of biases from the base and expanded models

To determine the possible reasons for the lower accuracy from the expanded models, the frequency distributions of subject-specific biases from the base and expanded models were examined. Table 5 lists the summary statistics for the bias ($\bar{e}_i$), Bias%$_i$ and RMSE$_i$ calculated plot-by-plot. The calculations were done for 71 plots of the validation data, and as a comparison, for 93 plots of the modelling data as well. The plots where the Bias%$_i$ exceeded ±2.5% of the observed mean were identified and listed separately, where the ±2.5% threshold was chosen by dividing the one-sided 5% significance level commonly used in statistical inferences into two-sides.

The frequency distributions of the biases, Bias%s and RMSEs were examined. Here we only show the distributions of the Bias%s from the base model (9) and the expanded model (11) (Figure 4), as the distribution from the expanded model (10) was similar to that of (11), and the distributions of the other statistics showed similar patterns to those of Figure 4. Actual values and additional information related to the distributions are available, but here we only list those (Table 6) relevant to the validation data.

The contrasts between the base and expanded models are obvious. For instance, when (9) was used, 18 out of 71 plots of the validation data produced biases exceeded ±2.5% of the observed mean (Tables 5-6). When (10) and (11) were used, 23 and 27 plots produced biases exceeded ±2.5% of the observed mean. The results on the modelling data were similar: 15 out of 93 plots had biases exceeded ±2.5% when (9) was used, and 30 and 37 plots had biases exceeded ±2.5% when (10) and (11) were used. The frequencies of poor predictions from the expanded models are greater than those from the base model.

The poorer performance of the expanded models can also be seen in Figure 4, where the biases are centered around zero but those from the expanded models are spread out more than those from the base model, suggesting larger prediction variations and lower accuracies from the expanded models. On the validation data, the standard deviations of the biases for (10) and (11) are 0.612 and 0.664, respectively (Table 5). Both are larger than that (0.399) for the base model (9). The δ value (0.159) of the base model is less than half of the δ values (0.386 and 0.447) of the expanded models, suggesting that the base model is more than twice as accurate as the expanded models. The same conclusion can also be reached on the modelling data (Table 5).

It was generally taken for granted that the predictive accuracy of a base model would be improved if additional tree and stand level variables were incorporated into the model. This is true in the traditional least squares regression, and it may also be true in the biased PA predictions using only the fixed parameters under the NLMM framework. However, in terms of subject-specific predictions that the NLMM technique was developed for, we found that the addition of top height in (10), and top height and basal area in (11) both produced lower accuracy than the base model (9). In fact, judging from the results in Tables 4-5, the overall accuracy (δ) of the predictions is reduced as additional variables are added into the base model. This is true on the validation data, as well as on the modelling data.

The possible reason that caused the base model to perform better in local predictions than the expanded models likely originated from the inclusion of subject-specific random parameters in the base model. Under the NLMM framework, the random parameters in the base model already allowed the plot level variations associated with different site-specific factors to be accounted for without actually requiring that they be identified or measured. This renders in a statistical sense (as oppose to biological) that the addition of other plot level variables such as top height and stand density, which also represent the same differences in site factors that the random parameters already represent, unnecessary.

For the purpose of evaluating the biological implications of top height and basal area on black spruce height growth under the NLMM framework, it is still useful to look at the significance of these variables when entered into the base model. However, for subject-specific local predictions, the results of this study show that the addition of other variables beyond the base height-diameter model is unwarranted. Apparently, the random parameters in an aptly specified base model can take into account all or most of the other known and unknown factors on a given site, thereby eliminating the need to measure or to include other variables when subject-specific local predictions are made.

4. Conclusions

We examined four base height-diameter models for black spruce under the NLMM framework. We found that the Chapman-Richards base model was the best. We compared this base model with the expanded models that included top height and/or basal area as the additional predictors through the evaluation of the overall biases and the frequency distributions of the biases from individual plots. We found that the base model produced the most accurate subject-specific height predictions. For black spruce, optimal, subject-specific tree height predictions can be achieved using a simple height-diameter model fitted by the NLMM technique without including any other additional variables. We also found that the frequency distributions of the biases from individual plots where the main interest of a NLMM lies are more powerful in revealing the adequacy of the fitted NLMM than many other statistics. Such frequency distributions should be routinely examined in any future study involving NLMMs.

References

Alberta Forest Service. (2000). Permanent sample plots field procedures manual. Forest Management Branch, Alberta

Sustainable Resource Development, Edmonton, Alberta, Pub. FMOPC 83-03, 106 p.

Arabatzis, A.A., & Burkhart, H.E. (1992). An evaluation of sampling methods and model forms for estimating height–diameter relationships in loblolly pine plantations. *For. Sci.* 38: 192–198.

Beal, S. L., & Sheiner, L.B. (1982). Estimating population kinetics. *Critical Reviews in Biomedical Engineering* 8: 195-222.

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

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

Chiswell, K., & Monahan, J. F. (2004). Dealing with model bias in a nonlinear mixed model. In *Proc. of the 2004 American Statistical Association*, Alexandria, VA, pp. 2419-2426.

Cochran, W.G. (1977). *Sampling techniques* (3rd ed.). Wiley, New York, 428 p.

Curtis, R.O. (1967). Height-diameter and height-diameter-age equations for second growth Douglas-fir. *For. Sci.* 13: 365-375.

Davidian, M., & Giltinan, D.M. (1995). *Nonlinear models for repeated measurement data*. Chapman & Hall, New York, 360 p.

Davidian, M., & Giltinan, D.M. (2003). Nonlinear models for repeated measurement data: an overview and update. *J. Agricultural, Biological, and Environmental Statistics* 8: 387–419.

Eerikäinen, K. (2003). Predicting the height–diameter pattern of planted *Pinus kesiya* stands in Zambia and Zimbabwe. *For. Ecol. Manage.* 175: 355–366.

Fang, Z., & Bailey, R.L. (2001). Nonlinear mixed effects modeling for slash pine dominant height growth following intensive silvicultural treatments. *For. Sci.* 47: 287–300.

Fitzmaurice, G.M., Laird, N.M., & Ware, J.H. (2004). *Applied longitudinal analysis*. Wiley, New York, 506 p.

Funatogawa, T., & Funatogawa, I. (2007). The Bayesian bias correction method of the first-order approximation of nonlinear mixed-effects models for population pharmacokinetics. *Jour. of Biopharmaceutical Statistics* 17: 381-392.

Hartley, H.O., & Ross, A. (1954). Unbiased ratio estimators. *Nature* 174: 270-271.

Hu, F.B., Goldberg, J., Hedeker, D., Flay, B.R., & Pentz, M.A. (1998). Comparison of population- averaged and subject-specific approaches for analyzing repeated binary outcomes. *American Journal of Epidemiology* 147: 694-703.

Huang, S. (1999). Ecoregion-based individual tree height-diameter models for lodgepole pine in Alberta. *Western J. Applied For.* 14: 186-193.

Huang, S. (2008). A generalized procedure for bias-free predictions at population and local levels. A discussion paper, Pub-Feb-22-2008, Forest Management Branch, ASRD, Edmonton, Alberta, 61 p.

Huang, S., & Titus, S.J. (1994). An age-independent individual tree height prediction model for boreal spruce-aspen stands in Alberta. *Can. J. For. Res.* 24: 1295–1301.

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

Huang, S., Yang, Y., & Wang, Y. (2002). A critical look at procedures for validating growth and yield models. In *Modelling Forest Systems*. CAB Publishing, Wallingford, UK, pp. 243-250.

Kjellsson, M.C., Jönsson, S., & Karlsson, M.O. (2004). The back-step method – method for obtaining unbiased population parameter estimates for ordered categorical data. *The AAPS Journal* 6: Article 19 (http://www.aapsj.org).

Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., & Schabenberger, O. (2006). *SAS system for mixed models* (2nd ed.), SAS Institute Inc., Cary, NC, 814 p.

Lindstrom, M.J., & Bates, D.M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics* 46: 673–687.

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

Nicholas, N.S., Gregoire, T.G., & Zedaker, S.M. (1991). The reliability of tree crown position classification. *Can. J. For. Res.* 21: 698-701.

Picard, R.R., & Cook, R.D. (1984). Cross-validation of regression models. *J. Am. Stat. Assoc.* 79: 575-583.

Schabenberger, O. (1994). Nonlinear mixed effects growth models for repeated measures in ecology. In *Proc. of ASA, Section on Stats. & Environment*, Toronto, Canada, pp. 156-161.

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

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

Vonesh, E.F., & Chinchilli, V.M. (1997). *Linear and nonlinear models for the analysis of repeated measurements*. Marcel Dekker, New York, 560 p.

Wykoff, W.R., Crookston, N.L., & Stage, A.R. (1982). User's guide to the stand prognosis model. USDA For. Serv. Gen. Tech. Rep. INT-133. 122 p.

Yang, Y., Monserud, R.A., & Huang, S. (2004). An evaluation of diagnostic tests and their roles in validating forest biometric models. *Can. J. For. Res.* 34: 619-629.

Young, M.L., Preisser, J.S., Qaqish, B.F., & Wolfson, M. (2007). Comparison of subject-specific and population averaged models for count data from cluster-unit intervention trials. *Statistical Methods in Medical Research* 16: 167-184.

Table 1. Summary statistics for black spruce model fitting and validation data

Data type	Level	Variable	N	Mean	Min	Max	SD
Model fitting	Tree	DBH (cm)	917	14.040	1.300	39.400	5.292
		H (m)	917	13.092	2.300	26.200	3.997
	Plot	BA (m^2/ha)	93	14.418	0.643	54.608	12.030
		TopH (m)	93	15.759	5.500	23.550	3.986
Validation	Tree	DBH (cm)	786	12.291	1.400	35.800	5.301
		H (m)	786	10.673	2.000	27.400	4.277
	Plot	BA (m^2/ha)	71	15.896	0.515	45.344	12.269
		TopH (m)	71	15.350	7.067	27.400	4.605

Note: DBH is tree diameter at breast height (1.3 m above ground), H is total tree height, BA is basal area per hectare, TopH is top height (i.e., average height of the 100 largest DBH trees per ha), N is sample size (i.e., number of trees at tree level, or number of plots at plot level). Min, max, and SD are minimum, maximum, and standard deviation, respectively.

Table 2. Parameter estimates and fit statistics for the base and expanded height-diameter models

Parameter	Base models				Expanded models	
	(5)	(6)	(7)	(9)	(10)	(11)
β_1	3.2418	1.6320	33.0854	25.4440	9.1617	8.9909
β_2	-11.5278	0.7401	3.8144	0.04734	0.07668	0.07127
β_3			-1.1826	1.0748	1.2266	1.1617
β_4					0.6745	0.6752
β_5						0.03388
$\sigma_{b_1}^2$	0.04409	0.3556	1053.65	10.7688	0.8135	1.1145
$\sigma_{b_1 b_2}$	-0.4310	-0.05916	18.8578	0.2834	0.08974	0.1277
$\sigma_{b_2}^2$	6.2356	0.01046	0.3426	0.02546	0.03450	0.02895
σ^2	1.2736	1.3557	1.2704	1.2931	1.2864	1.2923
P	6	6	7	7	8	9
m	93	93	93	93	93	93
-2LL	3094.6	3104.2	3079.9	3075.5	2959.3	2948.7
AIC	3106.6	3116.2	3093.9	3089.5	2975.3	2966.7
BIC	3121.8	3131.4	3111.6	3107.2	2995.5	2989.5
Ranking	[5]	[6]	[4]	[3]	[2]	[1]

Note: β_1-β_5 are fixed parameters, $\sigma_{b_1}^2$ and $\sigma_{b_2}^2$ are the variances for random parameters b_1 and b_2, respectively, $\sigma_{b_1 b_2}$ is the covariance between b_1 and b_2, σ^2 is the residual (error) variance, P is the total number of parameters (includes fixed parameters, variance-covariance components of the random parameters, plus the residual variance component), m is the number of subjects (plots), -2LL is -2 times log-likelihood of the model, AIC is Akaike information criterion, and BIC is the Schwarz's Bayesian information criterion (AIC and BIC are defined in eq. (27)), and ranking refers to the rank of the model by the AIC and BIC values (smaller is better).

Table 3. Height predictions for an example plot with three height-diameter (H-DBH) observations

Model	Observed		Predicted						δ	Bias%
	DBH	H	H_fix	e_fix	der_b$_1$	der_b$_2$	H_ss	e_ss		
(9)	7.9	7.8	8.577	-0.777	0.286	-8.475	7.411	0.389	0.748	-3.41
	10.7	8.2	10.738	-2.538	0.371	-8.709	9.384	-1.184		
	20.3	14.5	16.455	-1.955	0.596	-7.306	14.745	-0.245		
	Mean	10.167	11.923	-1.756			10.513	-0.347		
	SD	3.758	4.071	0.897			3.796	0.792		
(10)	7.9	7.8	8.497	-0.697	0.380	-5.678	7.589	0.211	1.111	-5.68
	10.7	8.2	10.597	-2.397	0.491	-5.394	9.730	-1.530		
	20.3	14.5	15.467	-0.967	0.748	-3.355	14.914	-0.414		
	Mean	10.167	11.520	-1.354			10.744	-0.577		
	SD	3.758	3.575	0.913			3.767	0.882		
(11)	7.9	7.8	8.454	-0.654	0.376	-6.029	7.731	0.069	1.156	-6.23
	10.7	8.2	10.479	-2.279	0.482	-5.767	9.804	-1.604		
	20.3	14.5	15.244	-0.744	0.732	-3.741	14.862	-0.362		
	Mean	10.167	11.392	-1.226			10.799	-0.633		
	SD	3.758	3.486	0.913			3.668	0.869		

Note: H_fix is height prediction from fixed parameters, H_ss is height prediction from fixed and random parameters, e_fix and e_ss are associated residuals, der_b$_1$ and der_b$_2$ are derivatives with respect to b$_1$ and b$_2$, SD is the standard deviation, δ is defined in eq. (28), and Bias% = 100 × mean bias / mean H.

Table 4. Summary statistics for the overall height prediction biases

Data	Parameters used	Model	N	Bias				δ	Bias%
				Mean	Min	Max	SD		
Validation	Fixed & random	(9)	786	-0.023	-3.342	4.198	0.870	0.757	-0.216
		(10)	786	-0.045	-3.799	4.183	0.892	0.798	-0.422
		(11)	786	-0.068	-3.793	4.177	0.903	0.820	-0.637
	Fixed only	(9)	786	-0.804	-8.584	7.450	1.990	4.607	-7.533
		(10)	786	-0.223	-4.910	4.490	1.221	1.541	-2.089
		(11)	786	-0.350	-5.261	4.312	1.197	1.555	-3.279
Modelling	Fixed & random	(9)	917	-0.004	-5.247	4.140	1.074	1.153	-0.031
		(10)	917	-0.002	-5.703	4.408	1.091	1.190	-0.015
		(11)	917	-0.002	-5.643	4.403	1.097	1.203	-0.015
	Fixed only	(9)	917	0.451	-4.559	7.462	1.838	3.582	3.445
		(10)	917	0.284	-6.109	4.706	1.395	2.027	2.169
		(11)	917	0.109	-5.883	5.007	1.336	1.797	0.833

Note: N is sample size, SD is standard deviation, δ is defined in eq. (28), and Bias% = 100 × mean bias / mean H (equals to 10.673 for the validation data, and 13.092 for the modelling data, see Table 1).

Table 5. A summary of plot-based height prediction biases from the base and expanded models.

Model	Type	Variable	Validation data				Modelling data			
			Freq	Mean	SD	δ	Freq	Mean	SD	δ
(9)	All	$\overline{e}_i$	71	-0.013	0.399	0.159	93	-0.011	0.270	0.073
		$\overline{e}_i\%$	71	-0.615	3.304		93	-0.404	1.961	
		$RMSE_i$	71	0.913	0.459		93	0.924	0.572	
	$|\overline{e}_i\%| > 2.5$	$\overline{e}_i$	18	-0.097	0.759		15	-0.001	0.598	
		$\overline{e}_i\%$	18	-2.088	6.236		15	-1.317	4.275	
		$RMSE_i$	18	1.105	0.581		15	0.729	0.329	
(10)	All	$\overline{e}_i$	71	-0.105	0.612	0.386	93	-0.054	0.442	0.198
		$\overline{e}_i\%$	71	-1.148	5.240		93	-0.286	2.952	
		$RMSE_i$	71	0.990	0.563		93	1.002	0.605	
	$|\overline{e}_i\%| > 2.5$	$\overline{e}_i$	23	-0.277	1.038		30	-0.185	0.735	
		$\overline{e}_i\%$	23	-3.200	8.811		30	-1.161	4.897	
		$RMSE_i$	23	1.265	0.712		30	1.111	0.667	
(11)	All	$\overline{e}_i$	71	-0.076	0.664	0.447	93	0.005	0.472	0.223
		$\overline{e}_i\%$	71	-1.039	5.514		93	0.105	3.170	
		$RMSE_i$	71	1.024	0.562		93	1.018	0.601	
	$|\overline{e}_i\%| > 2.5$	$\overline{e}_i$	27	-0.154	1.060		37	-0.028	0.726	
		$\overline{e}_i\%$	27	-2.223	8.767		37	-0.095	4.860	
		$RMSE_i$	27	1.247	0.672		37	1.066	0.626	

Note: $\overline{e}_i$, $\overline{e}_i\%$ and $RMSE_i$ are plot-specific bias, Bias% and root mean square error defined in eqs. (22), (24) and (25), respectively, freq refers to the number of plots, SD is standard deviation, and δ is an overall measure of accuracy defined in eq. (28).

Table 6. Frequency distributions of Bias% from the base and expanded models on the validation data

Class	Range		Mid-point	Base model		Expanded models			
	From	To		(9)		(10)		(11)	
				Mean	Freq	Mean	Freq	Mean	Freq
1	-∞	-10.5		-19.93	1	-23.85	2	-19.06	3
2	-10.5	-9.5	-10			-9.64	1	-10.42	1
3	-9.5	-8.5	-9			-8.89	2	-9.26	1
4	-8.5	-7.5	-8			-7.58	1	-8.27	1
5	-7.5	-6.5	-7	-7.08	3				
6	-6.5	-5.5	-6			-5.88	3	-6.04	2
7	-5.5	-4.5	-5	-5.17	2			-5.00	2
8	-4.5	-3.5	-4	-3.68	1	-3.97	2	-4.05	2
9	-3.5	-2.5	-3	-3.13	3	-3.21	2	-2.82	3
11	-2.5	-1.5	-2	-2.03	3	-1.91	8	-2.00	7
12	-1.5	-0.5	-1	-0.87	17	-0.97	12	-1.01	14
13	-0.5	0.5	0	-0.13	19	-0.01	13	-0.05	14
14	0.5	1.5	1	1.16	13	0.91	10	0.94	3
15	1.5	2.5	2	2.15	1	1.98	5	2.04	6
16	2.5	3.5	3	2.92	5	2.93	5	2.98	2
17	3.5	4.5	4	3.72	2	3.78	3	4.05	3
18	4.5	5.5	5	4.96	1	4.93	1	5.06	3
19	5.5	6.5	6					5.55	2
20	6.5	7.5	7					6.77	1
21	7.5	8.5	8						
22	8.5	9.5	9						
23	9.5	10.5	10			10.22	1		
24	10.5	∞						12.61	1
Total					71		71		71
No. of plots where |Bias%| > 2.5					18		23		27
% of plots where |Bias%| > 2.5					25.4		32.4		38.0

Note: mean and freq refer to the mean bias and frequency (number of plots) associated with the defined class. The frequencies are displayed in Figures 4(a)-4(b).
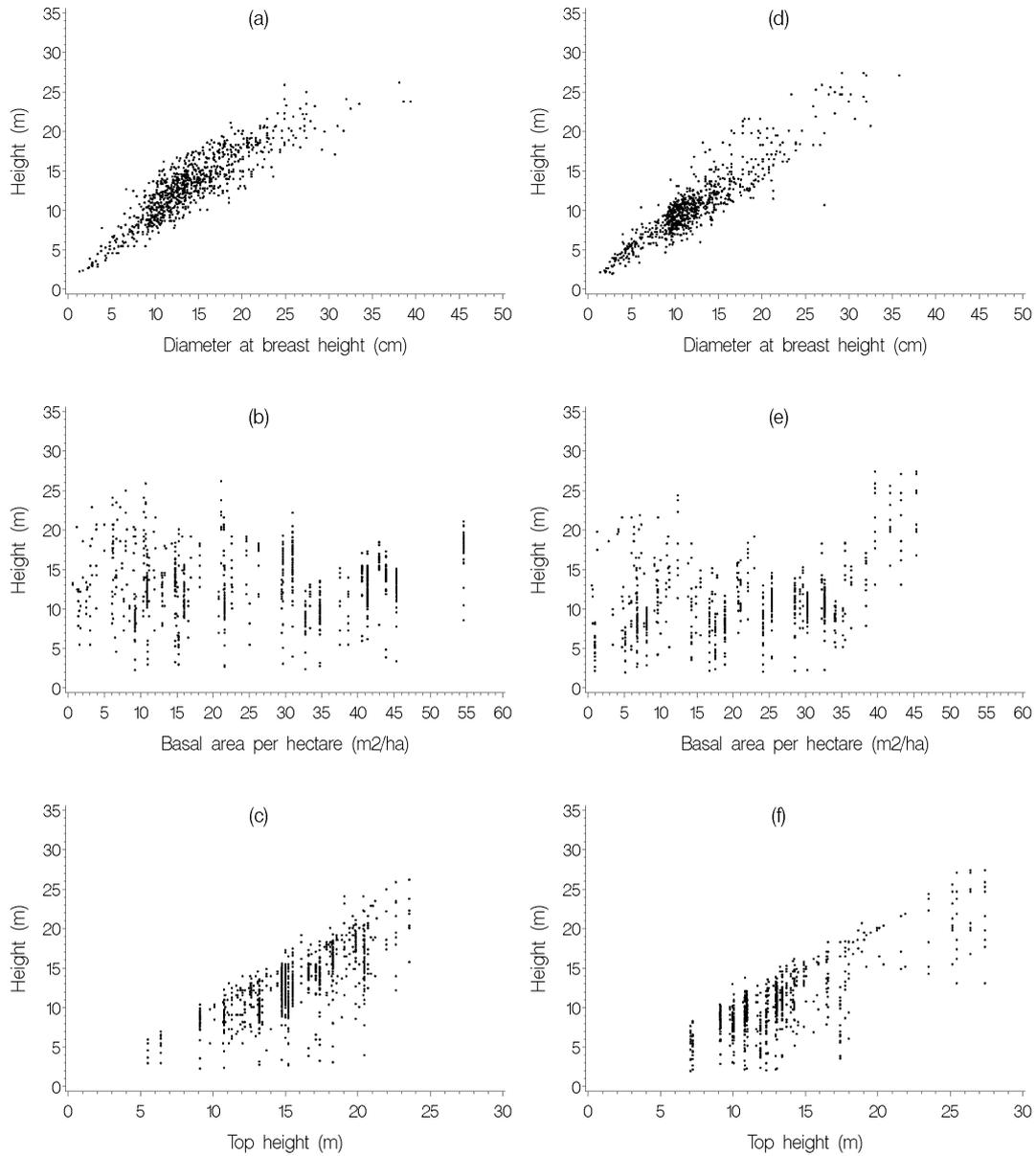
Figure 1. Model fitting (a, b, c) and validation data (d, e, f). Summary statistics are listed in Table 1.
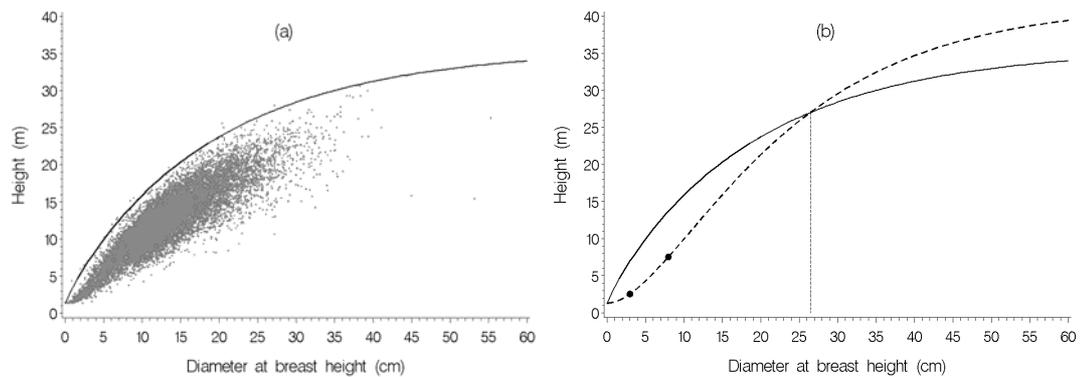
Figure 2. Maximum height-diameter relationship (solid line in a and b) for black spruce. The relationship is defined by eq. (26). The localized curve (dashed line) from the two data points in (b) gives unrealistic predictions when extrapolated beyond a DBH of 26.5 (cm). Therefore, for DBH > 26.5, height predictions are constrained by the maximum relationship.
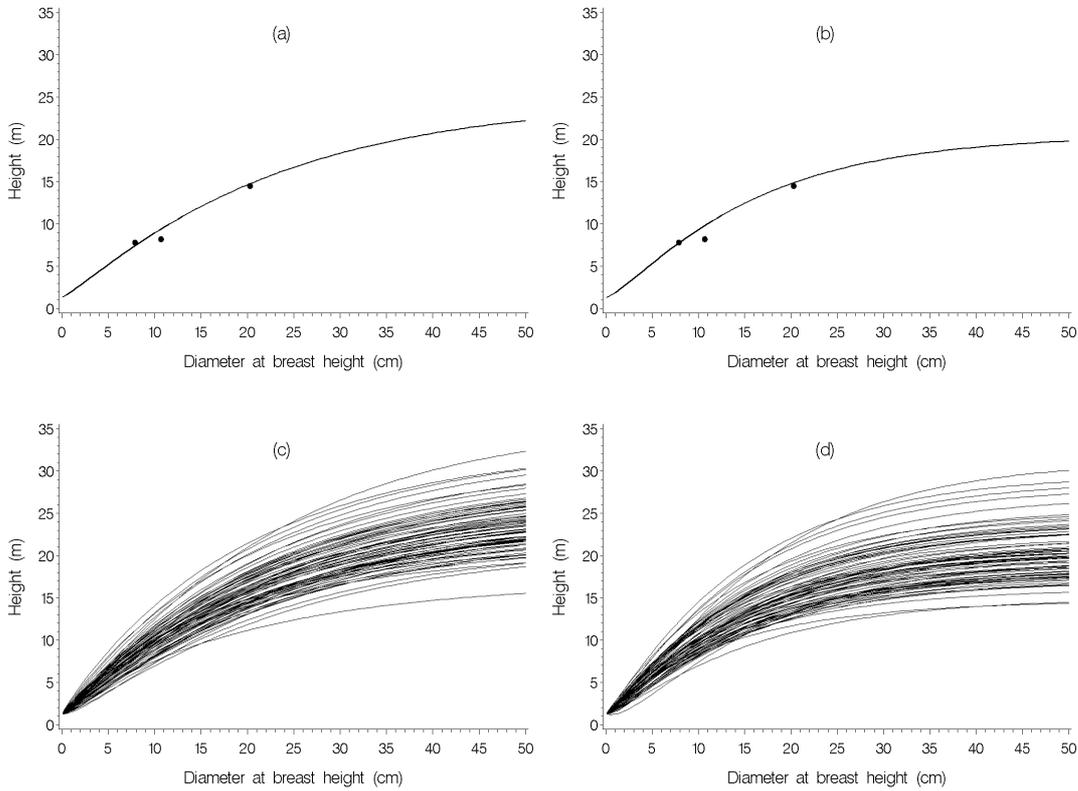
Figure 3. Height prediction curves from the base model (9) (a, c) and expanded model (11) (b, d), for an example plot (a, b) and for all 71 plots (c, d). Actual values for the example plot are listed in Table 3.
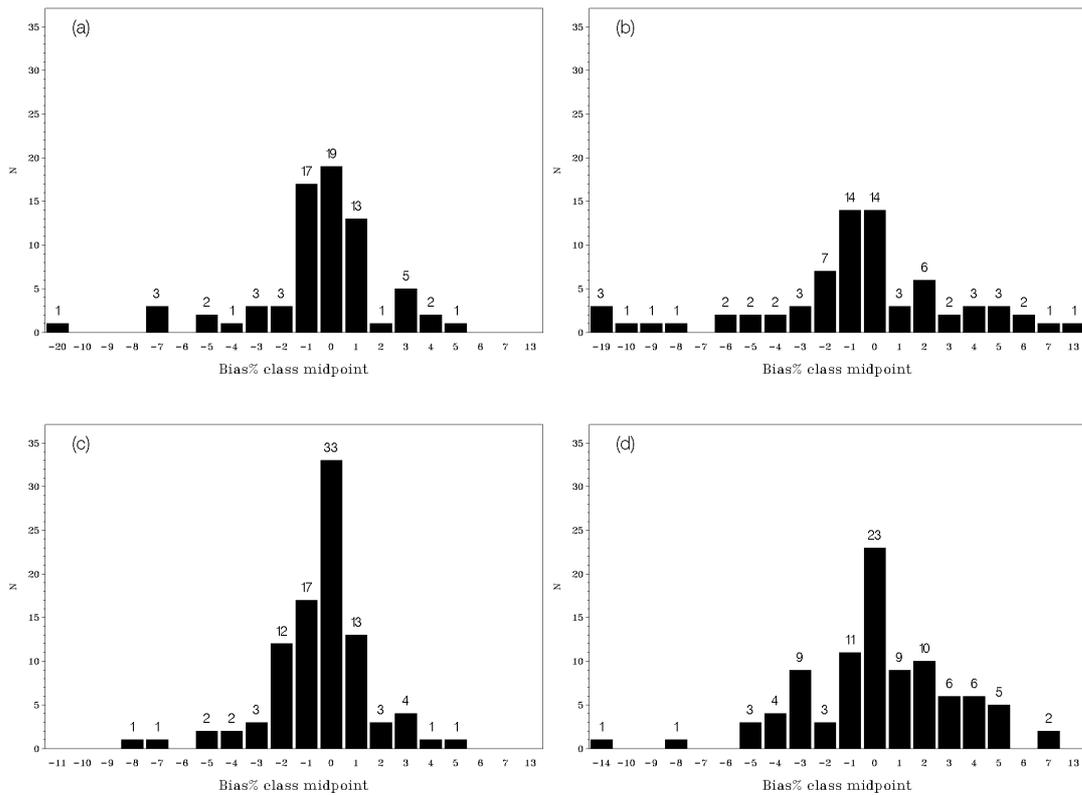
Figure 4. Frequency distributions of Bias% from models (9) (a, c) and (11) (b, d) on the validation data (a, b) and modelling data (c, d). Actual values related to (a) and (b) are given in Table 6.