CrossMark

Use of upper stem diameters in a polynomial taper equation for New Zealand radiata pine: an evaluation

Charles O. Sabatia[1,2]

Abstract

Background: Polynomial taper models are the tree profile equation types that are most commonly used to describe stem profiles of New Zealand radiata pine (*Pinus radiata* D.Don). Among these equations are model forms that include an upper stem diameter measurement as a predictor. Such taper functions may be more costly to use given the need to measure the additional stem diameter. Therefore, it is important to have an insightful understanding of the contribution of the additional diameter measurement to taper model prediction bias and precision to better inform decisions on whether to develop advanced technology for, or invest in, tree upper stem diameter measurements(s) for taper equation use.

Methods: Prediction precision and bias for diameter and volume were evaluated for a regular polynomial taper model with dbh and total height as tree-level predictors (DH model) and eight taper models that included an upper stem diameter as an additional tree-level predictor (DH+ models). Two sets of radiata pine stem sectional data from 66 stands across New Zealand were used in the investigation. Seventy percent of the trees in the larger one of the two sets were used for model fitting. The rest of the trees in the larger data set, and those in the remaining data set, were used for model validation.

Results: A DH+ taper equation that included an upper stem diameter measured at half the distance between breast height and the tree tip exhibited the lowest prediction error for stem diameters. Most of the reductions in diameter prediction error, from use of this equation, occurred in the upper half of the tree bole. Tree-level volume predictions from both DH and DH+ taper models were unbiased across the range of tree sizes investigated, but DH + models exhibited greater precision in volume prediction.

Conclusions: Including an outside bark upper stem diameter, measured at half the distance between breast height and the tree tip, in a polynomial taper equation for New Zealand radiata pine may result in a considerable improvement in tree volume prediction precision especially for total aboveground volume. Use of a diameter measured at 6 m may not result in prediction precision that is different from that of a DH taper equation.

Keywords: *Pinus radiata*, Stem taper, Tree profile equations, Tree volume estimation

Correspondence: charles.sabatia@msstate.edu
[1]Forest Science, Scion, 49 Sala Street, Private Bag 3020, Rotorua 3046, New Zealand
[2]Present Address: Department of Forestry, Mississippi State University, P.O. Box 9681, Mississippi State, MS 39762, USA

Background

Polynomial stem taper equations are the type of tree profile equation that are most commonly used to compute stem volume estimates that support forest management decisions in New Zealand. The most widely used version of these equations is the compatible polynomial taper equation (Goulding and Murray 1976; Gordon 1983; Katz et al. 1984; Hayward 1987). A compatible polynomial taper equation is constructed to include algebraic constraints that ensure that the estimate of total stem volume from the analytical integration of the taper equation is equal to the estimate from an associated tree volume equation (Diéguez-Aranda et al. 2006). Regular polynomial taper models, without volume compatibility constraints, have also been fitted for some of the tree species grown in New Zealand (e.g. Gordon et al. 1995, 1999). Such taper models are not constrained to guarantee volume predictions that equal those from the associated tree volume equation and hence are free from possible effects of volume compatibility constraints on taper equation predictive ability.

Including upper stem diameter(s) as additional taper model predictor(s) is a widely used method of accounting for inter-tree stem profile differences in stem taper functions. Various approaches of incorporating the diameter(s) in taper models have been proposed and demonstrated, which include use of upper stem diameters in a taper model equation system (e.g. Kilkki et al. 1978; Kilkki and Varmola 1979), use of the diameter(s) as direct predictors or as algebraic constraints (e.g. Czaplewski and McClure 1988; Kozak 1998; Cao 2009; Sabatia and Burkhart 2015) or use of the diameter(s) to compute the empirical Bayes estimate of the random effects taper equation parameter(s) of the subject tree (e.g. Sharma and Parton 2009; Cao and Wang 2011; Arias-Rodil et al. 2015; Sabatia and Burkhart 2015). Results from these studies suggest that the best upper stem diameter to include in a taper model probably differs by species, by taper model form and/or by region of the world. Most published studies have pointed to an upper stem diameter measured at approximately 50 % of total height, but others have pointed to diameters measured at as low as 38 % of total height and others at as high as 60 % of total height.

In taper models for New Zealand radiata pine (*Pinus radiata* D.Don), use of upper stem diameters was first demonstrated by Gordon and Budianto (1999) who developed a polynomial taper model that included an upper stem diameter, measured at 6 m above the ground (approximately 16 % of the average height of a mature New Zealand radiata pine tree), as an algebraic constraint predictor. The authors did not evaluate the role of the upper stem diameter in this equation nor did they evaluate other upper stem diameter measurements.

Therefore, potential users of the '3-point volume and taper equation for radiata pine' of Gordon and Budianto (1999) do not have a basis of deciding whether use of this model, as opposed to a dbh-and-total-height-only model (hereinafter referred to as the DH model), is beneficial. In addition, it is important to identify the upper stem diameter that is best to use in this model type to provide informed recommendations for users and add to the knowledge on the question of whether or not the most effective upper stem diameter to include in a taper model may differ by model type. In a recent study, Sabatia and Burkhart (2015) investigated the use of upper stem diameters in a segmented Max and Burkhart (1976) taper equation for New Zealand radiata pine. This study used a model form that is not currently used in New Zealand forestry, and hence, the research related more to the general scientific interest of taper equation researchers than it did to New Zealand users of taper equations. The current study seeks to follow up on the findings of Sabatia and Burkhart (2015) using a taper model that is one of those currently used in New Zealand forestry.
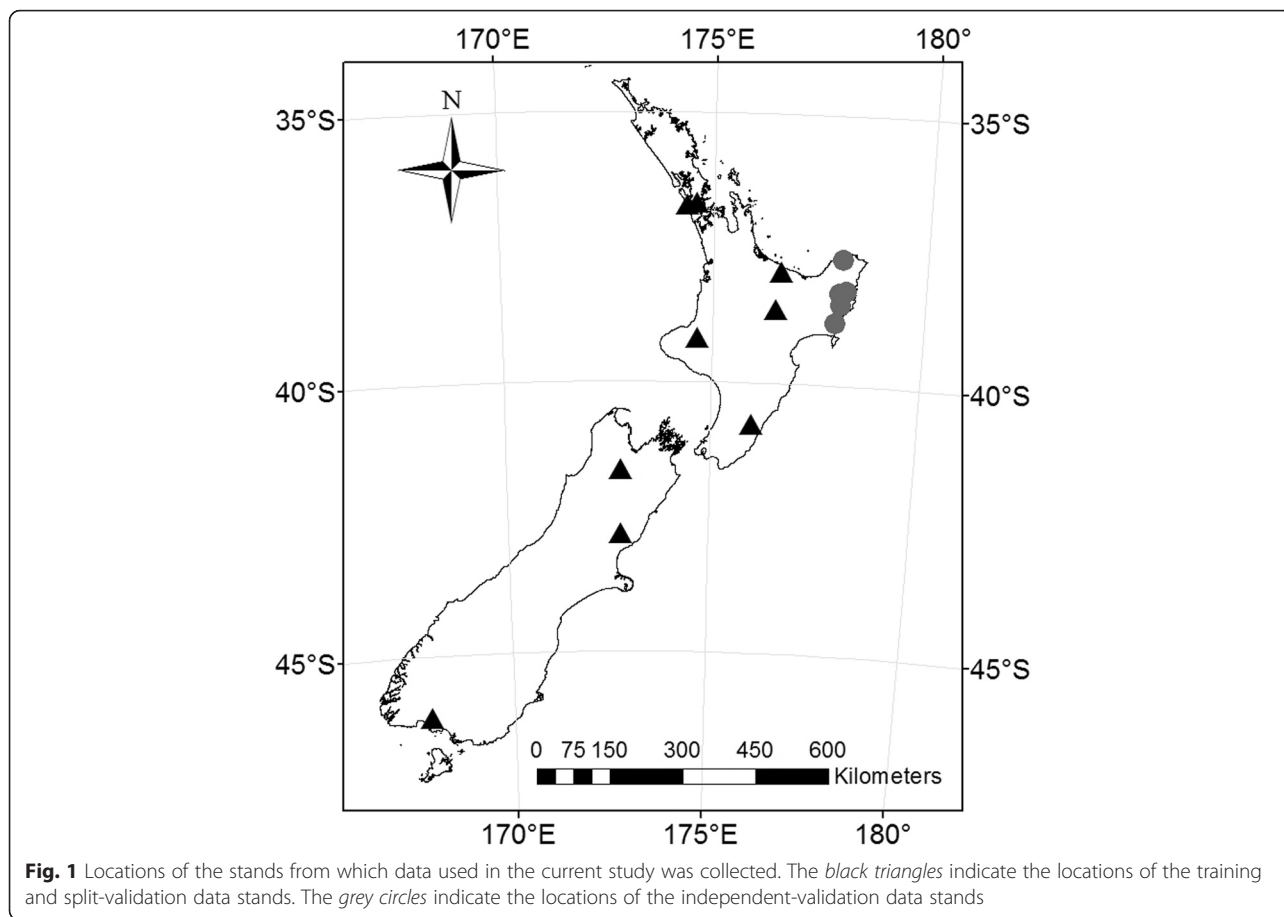
The objectives of the current study were to identify the most effective upper stem diameter for inclusion in a regular, no-volume-compatibility-constraint, polynomial taper model for radiata pine in New Zealand and to determine the effect, on tree volume prediction precision and bias, of including the upper stem diameter as a predictor in the taper model.

Methods

Data

Two sets of data were used in the study. The first set consisted of data from 792 trees of different levels of genetic improvement from 61 radiata pine stands located in nine forest estates spread across New Zealand (Fig. 1). The ages of these stands ranged from 9 to 39 years and stand density ranged from 150 to 700 trees ha^{-1}. A more detailed description of these data can be found in Gordon and Budianto (1999). The second set consisted of data from 107 genetically improved trees from five stands in the eastern part of the North Island of New Zealand that were all 23 to 25 years old (Fig. 1). Stand density in these stands ranged between 230 and 280 trees ha^{-1}. The two data sets were the same ones used in Sabatia and Burkhart (2015) and were provided by the New Zealand Forest Research Institute Limited (Scion).

Measurements on each tree in the data sets included dbh (breast height = 1.4 m above the ground), total height, outside bark diameter and bark thickness measurements at select points along the tree bole. The points were 0.15, 0.7, 1.4 (breast height) and 3 m above the ground and thereafter every 3 m up to the point where the remaining distance to the tip was less than

Fig. 1 Locations of the stands from which data used in the current study was collected. The *black triangles* indicate the locations of the training and split-validation data stands. The *grey circles* indicate the locations of the independent-validation data stands

3 m. Diameter measurements were taken on felled trees. Total aboveground outside bark volume was calculated for each tree by applying Smalian's formula to each of the bolt sections between stump height and the last diameter measurement point up in the stem, then applying volume of a cylinder formula to the stump (because groundline diameters were not measured) and volume of a cone formula to the tip portion and, lastly, summing up the bolt volumes. Merchantable volume to 14-cm outside bark top diameter (pulpwood top diameter limit) and to 22-cm outside bark top diameter (sawtimber top diameter limit) were calculated for trees that had an above-stump merchantable height of at least 3.8 m by applying Smalian's formula to each of the bolt sections between stump height and the top diameter limit and then summing up the bolt volumes. Where height at 14- or at 22-cm top diameter was not actually measured during data collection, it was computed by quadratic interpolation using measured heights at three closest diameters whose range included 14 or 22 cm, whichever was applicable. The specified diameter limits and the merchantable height requirement of at least 3.8 m were based on the current New Zealand domestic and export

markets small end diameter and log length limits for radiata pine (Laurie Forestry 2016).

Out of the 792 trees in the first data set, 554 trees (70 %) selected by proportional allocation stratified random sampling, using stand as stratum, were used in model fitting. The remaining 238 trees were used for model validation on data from the same population as the model fitting data set. Data from these 554 trees will hereinafter be referred to as the 'fitting' data and those from the 238 trees as the 'split-validation' data. The 107 trees in the second data set were used to carry out model validation on independent data from a population that may be different from the one used in model fitting. Data from these trees will hereinafter be referred to as the 'independent-validation' data. Summary statistics for dbh and total heights of the trees in the fitting, split-validation and independent-validation data sets are given in Table 1. Profile plots of the trees in these data sets are shown in Fig. 2.

The taper models

Taper models investigated in the current study were based on the polynomial taper model form that was introduced

Table 1 Summary statistics of the dbh and tree total height of the trees in the training, split-validation (Split-val) and independent-validation (Ind-val) data

Statistic	dbh (cm)			Total height (m)		
	Training data	Split-val data	Ind-val data	Training data	Split-val data	Ind-val data
Minimum	15.9	19.9	36.6	19.4	15.4	25.6
Median	47.0	46.2	54.3	35.3	35.1	37.1
Maximum	76.1	76.8	77.8	49.9	49.7	48.3

for *Eucalyptus saligna* Sm. in New Zealand forestry by Gordon et al. (1995). The form of this model is

$$\mathrm{dob} = \sqrt{\mathrm{dbh}^2\left(\beta_1 z^{\gamma_1} + \beta_2 z^{\gamma_2} + \beta_3 z^{\gamma_3}\right)} + \varepsilon \qquad (1)$$

where $z = \frac{H-h}{H}$, dob is the outside bark diameter at height h above the ground, H is the total height of the tree, β_1 to β_3 and γ_1 to γ_3 are model parameters that must satisfy the condition $\gamma_1 < \gamma_2 < \gamma_3$ and ε is an independent random error with 0 expectation.

Gordon and Budianto (1999) specified Eq. 1 into a model form, for radiata pine, that included an upper stem diameter term as an additional tree-level predictor (hereinafter referred to as the DH+ model) by expressing some of the parameters of Eq. 1 as functions of dbh, upper stem diameter at 6-m height and total height. The DH+ model of Gordon and Budianto (1999) will hereinafter be referred to as the GB99 model. In this model, the Eq. 1 parameters β_1 and β_2 were simultaneously constrained to guarantee that a tree's taper curve passes through the observed dbh and also through the diameter at 6 m above the ground. Additionally, parameters β_3, γ_1 and γ_3 were, respectively, expressed as functions of taper rate for the stem section between breast height and 6 m, taper rate for the stem section between 6 m and total height and as a function of form quotient at 6 m. The functional forms for the affected parameters were

$$\beta_1 = \frac{1 - \frac{z_{bh}^{\gamma_2}}{z_6^{\gamma_2}}\left(\frac{D_6^2}{\mathrm{dbh}^2} - \beta_3 z_6^{\gamma_3}\right) - \beta_3 z_{bh}^{\gamma_3}}{z_{bh}^{\gamma_1} - \frac{z_{bh}^{\gamma_2} z_6^{\gamma_1}}{z_6^{\gamma_2}}} \qquad (2)$$

$$\beta_2 = \frac{\frac{D_6^2}{\mathrm{dbh}^2} - \beta_1 z_6^{\gamma_1} - \beta_3 z_6^{\gamma_3}}{z_6^{\gamma_2}} \qquad (3)$$

$$\beta_3 = \beta_{30} + \beta_{31}\frac{\mathrm{dbh} - D_6}{6 - 1.4} \qquad (4)$$

$$\gamma_1 = \gamma_{10} + \gamma_{11}\frac{D_6}{H - 6} \qquad (5)$$

$$\gamma_3 = \gamma_{31} H \frac{D_6}{\mathrm{dbh}} \qquad (6)$$

In Eqs. 2 to 6, $z_{bh} = \frac{H-1.4}{H}$ and $z_6 = \frac{H-6}{H}$, D_6 is the measured diameter at 6-m height, and H is the total height. The rest of the terms are as defined for Eq. 1. According to the authors of the GB99 model, the roles of Eqs. 4, 5 and 6 in the taper model are to make it respond to changes in tree size. The GB99 model was one of the DH+ models evaluated in the current study.

The second DH+ model evaluated here, hereinafter referred to as M3, was of the same form as GB99 except that any upper stem diameter (D_{us}) could be used in the constraint Eqs. 2 and 3 in place of the specific upper stem diameter D_6. Consequently, Eq. 1 parameters β_1 and β_2 were expressed in M3 as
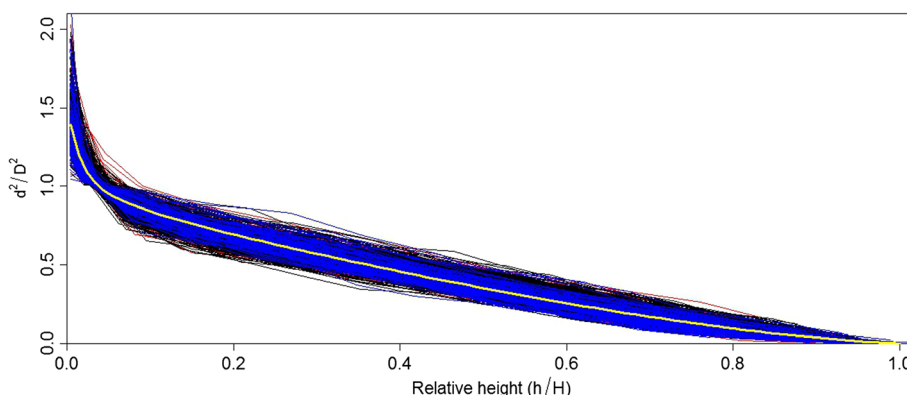

Fig. 2 Stem profile plots of the 899 trees used in the current study. Profiles of the trees in the fitting, split-validation and independent-validation data sets are shown in *black*, *red* and *blue lines*, respectively. The *yellow line* shows the BASE model-predicted tree profile for one of the trees (randomly selected), whose dbh and total height were 47.5 cm and 36.8 m, respectively

$$\beta_1 = \frac{1 - \frac{z_{bh}^{\gamma_2}}{z_{us}^{\gamma_2}}\left(\frac{D_{us}^2}{\text{dbh}^2} - \beta_3 z_{us}^{\gamma_3}\right) - \beta_3 z_{bh}^{\gamma_3}}{z_{bh}^{\gamma_1} - \frac{z_{bh}^{\gamma_2} z_{us}^{\gamma_1}}{z_{us}^{\gamma_2}}} \qquad (7)$$

$$\beta_2 = \frac{\frac{D_{us}^2}{\text{dbh}^2} - \beta_1 z_{us}^{\gamma_1} - \beta_3 z_{us}^{\gamma_3}}{z_{us}^{\gamma_2}} \qquad (8)$$

where $\frac{z_{us} = H - h_{DUS}}{H}$ is the value of z at height h_{DUS}, h_{DUS} is the measurement point for D_{us} and the rest of the terms are as defined for Eqs. 2 and 3. The rest of the Eq. 1 parameters in M3 were treated as global.

Model M3 was a more parsimonious generalized version of GB99. Upper stem diameter measurements (D_{us}) from seven different points along the stem (h_{DUS}) were used in M3 resulting in seven different versions of this model, which were evaluated to identify the most effective D_{us} to use in this model form. The seven h_{DUS} points were (1) a point at 50 % of the height above dbh and (2) a point at 20 % of the total height of the tree and then a point every 10 percentage points up to 70 % of the total height of the tree. When h_{DUS} occurred where diameter was not actually measured during data collection, D_{us} for the h_{DUS} was computed by quadratic interpolation involving a closest measured diameter below and a closest measured diameter above the h_{DUS} plus one other measured diameter that was next closest to the h_{DUS}.

The DH model evaluated in the current study, hereinafter referred to as the BASE model, was a version of the GB99 model in which the parameters β_1 and β_2 were expressed as

$$\beta_1 = \frac{1 - \beta_2 z_{bh}^{\gamma_2} - \beta_3 z_{bh}^{\gamma_3}}{z_{bh}^{\gamma_1}} \qquad (9)$$

and

$$\beta_2 = \beta_{21}\frac{\text{dbh}}{H - 1.4} \qquad (10)$$

with the rest of the Eq. 1 parameters in this model being treated as global. Equation 9 is a constraint that guarantees that the BASE model taper curve passes through the tree's dbh. Equation 10 plays a role similar to that played by Eq. 4 in the GB99 model. It expresses BASE model parameter β_2 as a function of the tree's above-dbh taper rate. I chose to express β_2, instead of β_3 as in GB99, as a function of taper rate because working with β_2 in the BASE model resulted in a better model fit. An intercept is not included in Eq. 10 because it was found to be insignificant. The purpose of the BASE model in the current study was to provide a baseline for evaluating the effect of including an upper stem diameter measurement, in the DH+ models GB99 and M3, on taper model prediction precision and bias.

Model fitting

The BASE, GB99 and the seven versions of the M3 model were first fitted separately to the fitting data by ordinary nonlinear least squares regression using the NLIN procedure in SAS/STAT® software Version 9.3 (SAS Institute Inc. 2002 - 2004). Residuals from these regressions were recovered and used to develop weight functions that were subsequently used in weighted nonlinear least squares regression to compute final parameter estimates. The weight function was developed by fitting the following equation to the residuals and tree dimension data:

$$e_{ij}^2 = \text{dbh}^{\alpha_0} \times \exp(\alpha_1 I_1 L_1 + \alpha_2 I_2 L_2 + \alpha_3 I_3 L_3) + \epsilon_{e_{ij}} \qquad (11)$$

where e_{ij} is the jth residual on the ith tree, L_1 is the distance from the ground level, L_2 is the upper distance from dbh, L_3 is the upper distance from the upper stem diameter measurement point, α_0 to α_3 are parameters to be estimated by nonlinear regression, $\epsilon_{e_{ij}}$ is the random error with 0 expectation and I_1 to I_3 are dummy variables for the part of the tree bole defined as follows:

$I_1 = 1$ for the bole part below breast height and 0 otherwise
$I_2 = 1$ for the bole part between dbh and the upper stem diameter measurement point and 0 otherwise
$I_3 = 1$ for the bole part above the upper stem diameter measurement point and 0 otherwise

For the BASE model, which did not include an upper stem diameter, the weight function was developed from the equation

$$e_{ij}^2 = \text{dbh}^{\alpha_0} \times \exp(\alpha_1 I_1 L_1 + \alpha_2 I_2 L_2) + \epsilon_{e_{ij}} \qquad (12)$$

where L_1 is the distance from the ground level, L_2 is the upper distance from dbh and I_1 and I_2 are dummy variables for the part of the tree bole as follows:

$I_1 = 1$ for the bole part below breast height and 0 otherwise
$I_2 = 1$ for the bole part above dbh and 0 otherwise

and the rest are as defined for Eq. 11.

Equations 11 and 12 are based on the reasoning that error variance will depend on tree size (represented by dbh) and distance from the ground or distance from the constraint point. Constraint points (breast height and height at the upper stem diameter measurement point) were considered as the other starting points, aside from the ground level, because error variance at these points is 0 (because the models investigated are conditioned to

predict the observed dbh and the observed upper stem diameter).

After identifying suitable weight functions based on Eqs. 11 and 12, the parameters of the BASE, GB99 and the seven versions of M3 were refitted by weighted nonlinear least squares using the model procedure in SAS/ETS® software Version 9.3 (SAS Institute Inc. 2002 - 2004). In this regression analysis, residual correlations due to multiple dob measurements on a single tree were modelled using a modified first-order continuous autoregressive correlation structure (CAR (1)) according to the procedure in Zimmerman and Núñez-Antón (2001). Accordingly, the model error component of the jth observation on the ith tree (ε_{ij}) was expanded as

$$e_{ij} = \left(\rho_0 + \rho_1 I_1 + ... + \rho_n I_n\right)^{h_{ij}-h_{ij-1}} e_{ij-1} + \varepsilon_{ij} \qquad (13)$$

where e_{ij}, for j as the second or greater measurement, is the jth ordinary residual on the ith tree; ρ_0, ρ_1, to ρ_n are the CAR (1) correlation parameters for the residuals for different portions of the tree bole; $h_{ij} - h_{ij-1}$ is the distance between the jth and the jth – 1 dob in metres; and I_1 to I_n are dummy variables for the part of the tree bole ($I_n = 1$ for residuals in the given part of the bole and 0 otherwise). In Eq. 13, the CAR (1) correlation parameter has been expanded to take into consideration the possibility of residual correlations in different parts of the tree bole exhibiting different signs (positive or negative) and magnitudes due to the effect of dbh and upper stem diameter constraints on the taper model (Lappi 2006). A decision on how many residual correlation parts to divide the tree boles into was based on examination of the residual versus lag residual Pearson correlations for 10 % point bole portion ranges (0 to 10 % of total height, 10 to 20 %,..., and 90 to 100 %). Based on the signs and magnitudes of the correlations, tree boles were divided into two to four residual correlation sections which were distinguished using the dummy variables in Eq. 13. Across all models investigated in the current study, residuals around the constraint points were negatively correlated. Those in the lower 10 % of the bole were weakly correlated.

Model evaluation
The fitted models were evaluated for fit on the fitting data set and for prediction performance on the split-validation and independent-validation data sets. All models were evaluated for diameter fit and prediction performance, but only the BASE, GB99 and the best one of the seven M3 models were evaluated for volume prediction performance. Model-predicted volume was calculated from the model-predicted diameters using Smalian's formula as was done for the observed tree volumes (see

the 'Data' section of the current paper for the description of the observed tree volume calculations).

Model evaluation for diameter fit and prediction performance was based on the global statistics mean bias (MB), mean absolute bias (MAB) and root mean square error (RMSE). Evaluation was also done based on a stem-portion-level measure of fit and prediction bias that will be hereinafter referred to as percent mean stem portion bias (MSPB%). MB, MAB and RMSE were, respectively, computed as

$$MB_{dob} = \frac{1}{N} \sum_{i=1}^{N} \frac{\sum_{j=1}^{n_i} \left(\widehat{dob}_{ij}-dob_{ij}\right)}{n_i} \qquad (14)$$

$$MAB_{dob} = \frac{1}{N} \sum_{i=1}^{N} \frac{\sum_{j=1}^{n_i} \left|\widehat{dob}_{ij}-dob_{ij}\right|}{n_i} \qquad (15)$$

$$RMSE_{dob} = \sqrt{\frac{\sum_{i=1}^{N}\sum_{j=1}^{n_i} \left(dob_{ij}-\widehat{dob}_{ij}\right)^2}{(n_1 + n_2 + ... + n_N)-p}} \qquad (16)$$

In Eqs. 14, 15 and 16, $\widehat{dob}_{ij}$ and dob_{ij} are, respectively, the jth predicted and jth measured diameter on the ith tree, n_i is the number of diameter measurements on the ith tree, N is the total number of trees and p is the number of taper model parameters. The subscript dob on the statistic abbreviation indicates that it is a diameter prediction statistic. To calculate MSPB%, the bole of every tree was divided into ten relative height portions 0 to 10 %, 10 to 20 %,..., and 90 to 100 % (note that a tree bole is continuous and hence the endpoint of one relative height potion is the beginning of the next one). The MSPB% for each portion was then computed as

$$MSPB\% = \left(\frac{\frac{1}{N} \sum_{i=1}^{N} \frac{\sum_{j=1}^{n_i} \left(\widehat{dob}_{ij}-dob_{ij}\right)}{n_i}}{}\right) \times {}^{100}\!\big/\overline{dob} \qquad (17)$$

where $\widehat{dob}_{ij}$ and dob_{ij} are, respectively, the jth predicted and jth measured diameter in the portion of interest on the ith tree, n_i is the number of diameter measurements in the portion of interest on the ith tree, N is the total number of trees and $\overline{dob}$ is the mean dob for the stem portion of interest across all the N trees. A 95 % confidence interval of MSPB% was also computed using the standard error of the mean computed in Eq. 17.

Evaluation for volume prediction performance was based on the statistics MB, MAB and RMSE and also on plots of tree-level prediction errors versus tree dbh. The statistics were, respectively, computed as

$$MB_v = \frac{1}{N} \sum_{i=1}^{N} (\hat{v}_i - v_i) \qquad (18)$$

$$MAB_v = \frac{1}{N} \sum_{i=1}^{N} |\hat{v}_i - v_i| \qquad (19)$$

$$RMSE_v = \sqrt{\frac{\sum_{j=1}^{N} (v_i - \hat{v}_i)^2}{N - p}} \qquad (20)$$

where $\hat{v}_i$ and v_i are, respectively, the ith predicted and ith observed outside bark total aboveground, pulpwood or sawtimber volume; N is the total number of trees; and p is the number of taper model parameters. The subscript v on the statistic abbreviations distinguishes the statistics as volume prediction statistics.

Diameter and volume prediction performance ranking of the models was based on the MAB and RMSE statistics. The MB statistic was not used in the ranking process because it does not capture large positive and negative deviations, which cancel out when deviations get summed during MB calculation.

Results

The diameter fit statistics for the seven versions of M3 taper models are given in Table 2. The model with D_{us} measured at half the distance between breast height and the tip of the tree had the lowest RMSE. This model will hereinafter be referred to as the M3_HABH model.

The parameter estimates for the BASE, M3_HABH and GB99 models are given in Table 3. The fit and prediction statistics for these models are given in Table 4. A tree profile predicted using the BASE model, for one tree randomly selected from the fitting data set, whose dbh and total height were 47.5 cm and 36.8 m, respectively, is shown in Fig. 2. Generally, the inclusion of upper stem diameters improved model fit and reduced taper model prediction error, with the diameter measured at half the

Table 2 Fit statistics for diameter prediction by the seven versions of M3 taper models fitted on the 'fitting' data set

h_{DUS}	Fit statistic		
	MB	MAB (%)	RMSE
20 %	−0.331	3.758	5.404
30 %	−0.249	3.434	4.995
40 %	−0.077	3.243	4.731
50 %	0.208	3.108	4.543
50 % HABH	0.263	*3.093*	*4.523*
60 %	1.201	3.404	4.982
70 %	0.583	3.406	4.978

h_{DUS} stands for the measurement point for the upper stem diameter D_{us} as a proportion of the total height of the tree. 50 % HABH stands for 50 % of the height above breast height. The lowest MAB and RMSE are given in italics. The percentages are based on a global mean outside bark diameter of 31.65 cm

Table 3 Parameter estimates for the BASE, GB99 and M3_HABH taper models

Parameter	Estimate		
	BASE model	Model GB99	Model M3_HABH
β_1	T-S	T-S	T-S
β_2	–	T-S	T-S
β_{21}	0.0929 (0.0130)	–	–
β_3	0.4125 (0.0189)	–	0.5502 (0.0053)
β_{30}	–	0.8213 (0.0351)	–
β_{31}	–	−0.2040 (0.0199)	–
γ_1	1.4420 (0.0061)	–	1.7149 (0.0440)
γ_{10}	–	0.9377 (0.0237)	–
γ_{11}	–	0.3522 (0.0169)	–
γ_2	16.6627 (2.4157)	14.5371 (0.6206)	2.1955 (0.1190)
γ_3	64.2933 (3.0251)	–	34.3438 (0.4982)
γ_{31}	–	1.3431 (0.0351)	–

The standard errors of the parameter estimates are given in parentheses. For the BASE model, parameter β_1 is constrained to dbh, and for models M3_HABH and GB99, parameter β_1 is constrained to dbh and β_2 to the upper stem diameter (diameter at 6 m for the GB99 model). The estimates for the constrained parameters are tree-specific (T-S). The parameter estimates given were all statistically significant ($p \leq 0.0001$). Residual correlation parameters are not included because they are not important for model prediction purposes

distance between breast height and the tip of the tree being more effective than that measured at 6 m above the ground (Table 4).

The along-tree-bole diameter prediction bias trends for the BASE, GB99 and M3_HABH models are compared in Figs. 3, 4a and b for the fitting, split-validation and independent-validation data sets, respectively. For trees in the fitting and split-validation data sets, the BASE model generally underestimated diameters for the bole section between 30 and 80 % of total height and overestimated those in the section above 80 % of total height. For those trees in the independent-validation data set, the model generally overestimated diameters above 50 % of total height. The along-tree-bole bias trend for the GB99 model was not very different from that exhibited by the BASE model. On the other hand, the M3_HABH model generally had smaller diameter prediction biases in both the 30 to 80 % and the above 80 % stem portions for trees in the fitting and those in the split-validation data set (Figs. 3 and 4a) and in the above 60 % stem portion for trees in the independent-validation data set (Fig. 4b).

The prediction error statistics for individual tree volumes are given in Table 4. The greater diameter prediction precisions of the GB99 and M3_HABH models relative to the BASE model, as expected, translated into greater precision for stem volume prediction. The improvement in volume prediction precision, due to inclusion of upper stem diameters in the taper model, was generally larger compared to the corresponding

Table 4 Diameter fit/prediction and volume prediction error statistics for the BASE, GB99 and M3_HABH taper models

Dimension	Model	Fitting data			Split-validation data			Independent-validation data		
		MB_{dob} (%)	MAB_{dob} (%)	$RMSE_{dob}$ (%)	$MB_{dob/v}$ (%)	$MAB_{dob/v}$ (%)	$RMSE_{dob/v}$ (%)	$MB_{dob/v}$ (%)	$MAB_{dob/v}$ (%)	$RMSE_{dob/v}$ (%)
Stem diameters	BASE	0.143	4.254	5.802	0.696	4.313	5.788	1.603	5.060	7.052
	GB99	−0.429	3.561	5.208	−0.174	3.529	5.110	1.005	4.291	6.136
	M3_HABH	0.263	*3.093*	*4.523*	0.466	*3.142*	*4.488*	0.535	*3.872*	*5.615*
Total aboveground volume	BASE				1.133	5.515	7.720	1.371	5.322	7.408
	GB99				−1.316	3.768	5.247	0.525	4.683	6.372
	M3_HABH				0.768	*2.729*	*3.824*	−0.816	*2.810*	*3.937*
Volume to 14-cm top diameter (pulpwood volume)	BASE				0.784	5.328	7.553	0.707	5.092	6.989
	GB99				−1.586	3.671	5.153	−0.003	4.430	6.056
	M3_HABH				0.616	*2.675*	*3.709*	−1.232	*2.863*	*4.140*
Volume to 22-cm top diameter (sawtimber volume)	BASE				0.884	5.123	7.402	0.129	4.718	6.514
	GB99				−1.616	3.378	4.974	−0.472	4.038	5.697
	M3_HABH				0.763	*2.861*	*3.982*	−1.592	*2.994*	*4.457*

The percentages are based on the following overall means: (a) fitting data stem diameter means = 31.65 cm; (b) split-validation stem diameter means = 31.38 cm and tree volume mean = 2.528 m³ for total aboveground volume, 2.439 m³ for pulpwood volume and 2.310 m³ for sawtimber volume; and (c) independent-validation stem diameter means = 36.67 cm and volume mean = 3.649 m³ for total aboveground volume, 3.556 m³ for pulpwood volume and 3.429 m³ for sawtimber volume. The lowest MAB and RMSE values are shown in italics

improvement in diameter prediction precision. Considering the better performing upper stem diameter model M3_HABH, gains in stem volume prediction precision were highest for total aboveground volume and lowest for sawtimber volume.

The across-tree-size volume prediction errors for the BASE, GB99 and M3_HABH taper models are plotted in Fig. 5 for total aboveground volume of trees in the split-validation data set. Across-tree-size prediction errors for sawtimber and pulpwood volumes in this data set exhibited trends that were similar to those in Fig. 5 (residual plots not shown). Volume (total aboveground, sawtimber and pulpwood) prediction errors for trees in the independent-validation data set also exhibited similar trends (residual plots not shown). For volume prediction, the three taper models were generally unbiased across tree size.

Discussion

For the polynomial taper model investigated in the current study, the diameter measurement at half the distance between breast height and the tip of the tree was the best diameter to include in a DH+ model. This finding is different from the one by Sabatia and Burkhart (2015) who, with the same data used in the current study, concluded that the diameter measurement at 60 % of total height was the best to include in a Max and Burkhart (1976) segmented taper model that had been constrained to dbh and an upper stem diameter. Nonetheless, the finding in the current study concurs
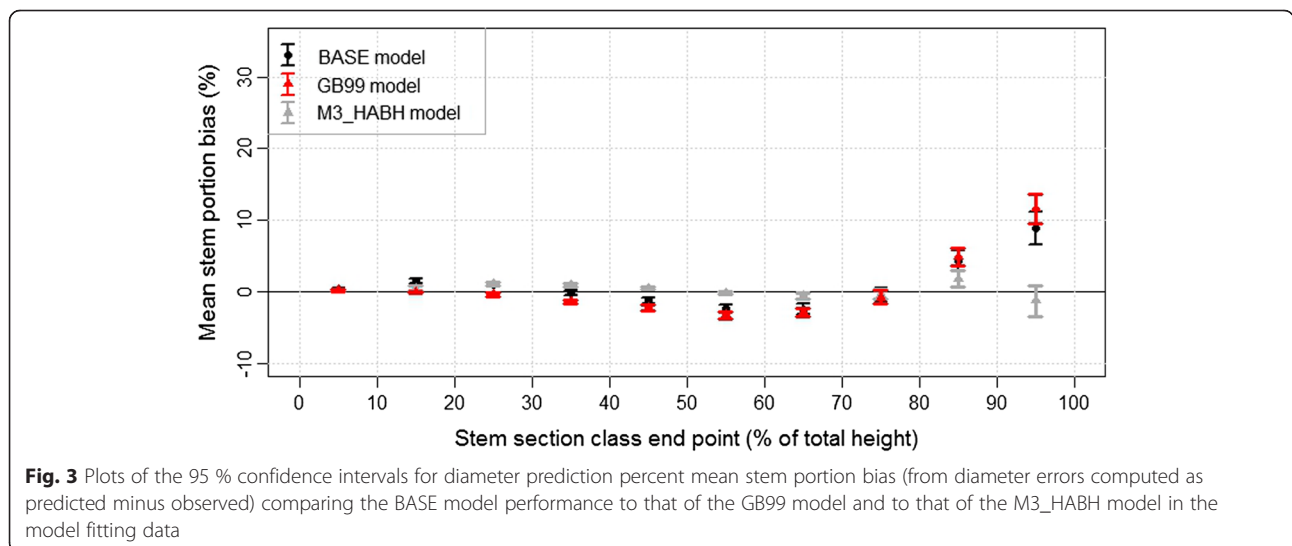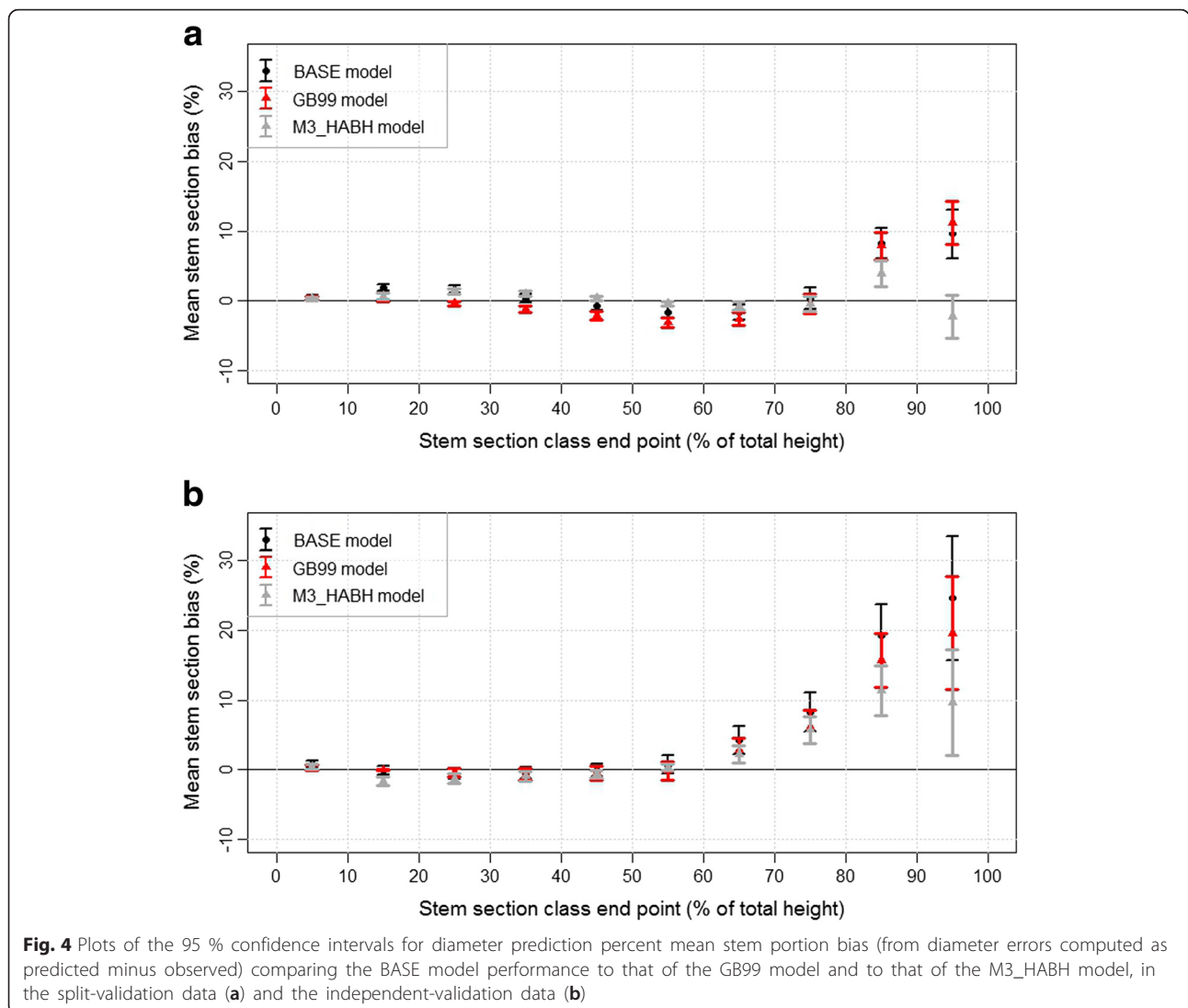

Fig. 3 Plots of the 95 % confidence intervals for diameter prediction percent mean stem portion bias (from diameter errors computed as predicted minus observed) comparing the BASE model performance to that of the GB99 model and to that of the M3_HABH model in the model fitting data

Fig. 4 Plots of the 95 % confidence intervals for diameter prediction percent mean stem portion bias (from diameter errors computed as predicted minus observed) comparing the BASE model performance to that of the GB99 model and to that of the M3_HABH model, in the split-validation data (**a**) and the independent-validation data (**b**)

with that by Cao (2009) who concluded that a diameter at half the distance between breast height and the tip of the tree was the best to include in a modified version of the Max and Burkhart (1976) segmented taper model that had also constrained to dbh and an upper stem diameter. In general, the finding in the current study points to a region in the neighbourhood of 50 % of total height as being the best point to measure an upper stem diameter to be included in a DH+ taper model, similar to what some previous studies have found (e.g. Kozak 1998; Arias-Rodil et al. 2015). The differences in findings between studies may be due to differences in taper model form.

Despite a difference in conclusion on the best upper stem diameter to include in a DH+ taper model, between the current study and Sabatia and Burkhart (2015), the decrease in diameter prediction RMSE from use of the best

upper stem diameter (between 1.3 and 1.5 percentage point decrease reported in the current study and between 1 and 3 percentage point decrease reported in the previous study) was not very different. These improvements in diameter prediction precisions were also not very different from those observed in previous studies that have incorporated upper stem diameters in taper models (e.g. Kozak 1998; Sharma and Parton 2009; Yang et al. 2009; Gómez-García et al. 2013 (see Table 5 for details of diameter prediction RMSEs that were reported in these studies)). The improvements in stem diameter prediction precision have been observed to be generally small and probably not of significant impact (Kozak 1998; Sabatia and Burkhart 2015). The findings on stem volume prediction errors (Table 4 and Fig. 5), however, indicate that upper stem diameter measurements could significantly improve the reliability of volumes predicted by polynomial taper
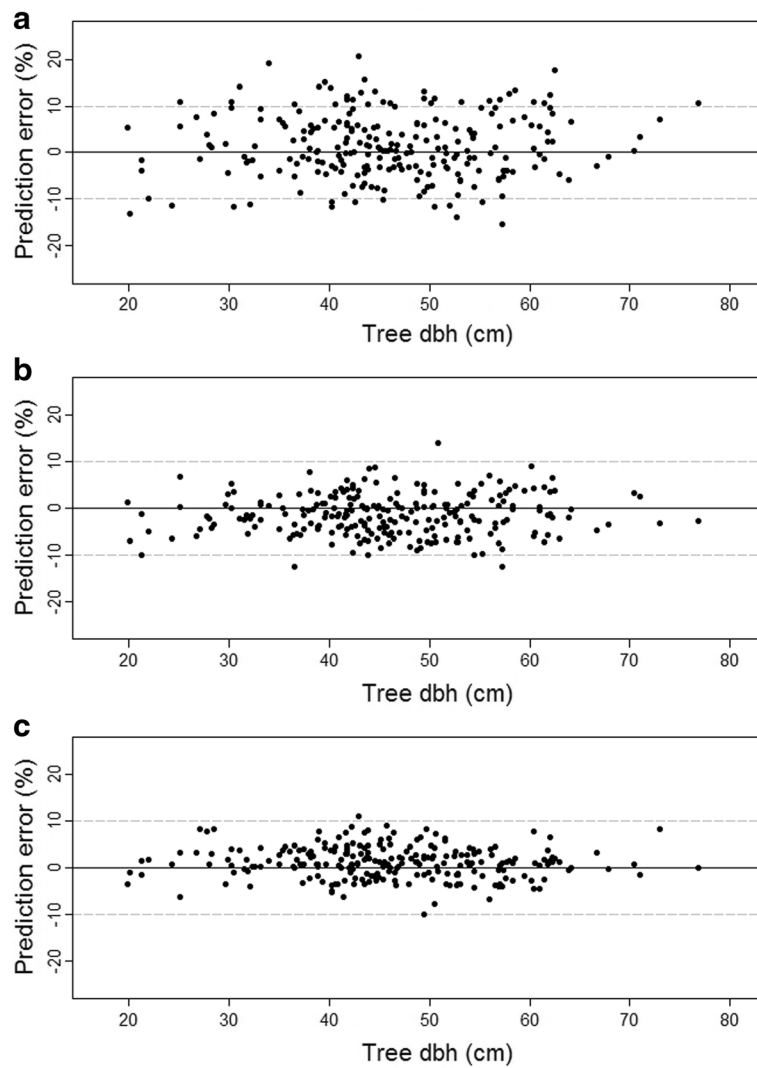
Fig. 5 Taper model prediction errors (predicted minus observed) for total aboveground volume calculated from diameters predicted by the BASE (**a**), GB99 (**b**) and M3_HABH (**c**) stem taper models for trees in the split-validation data set

Table 5 With- and without-upper-stem-diameter diameter prediction RMSEs for taper models investigated in previous studies

Author and year	RMSE units	RMSE	
		Model without upper stem diameter	Model with upper stem diameter
Kozak (1998)[a]	%	12.56	9.91
Kozak (1998)[a]	%	10.80	8.45
Sharma and Parton (2009)[a]	$\frac{dob}{dbh}$	0.787	0.5909
Sharma and Parton (2009)[a]	$\frac{dob}{dbh}$	0.4915	0.3764
Yang et al. (2009)	cm	0.7681	0.6708
Gómez-García et al. (2013)	cm	1.55	1.43

The studies included in this table are those where a range of upper stem diameters were investigated and RMSE reported. The 'with-upper-stem-diameter' model included in the table was the best performing upper stem diameter in the cited study
[a]For these authors, the second citation of the same author provides information on a second species that was investigated in the same study

models. The cumulative effect of small gains in diameter prediction precision across a tree likely results in considerable gains in the precision of tree-level volume predictions.

Similar to the results with the segmented taper model in Sabatia and Burkhart (2015), use of the best upper stem diameters in the polynomial model in the current study mainly improved the performance of the taper model in the upper half of the tree bole (Figs. 3 and 4). Consequently, volumes of tree boles that are longer than 50 % of the tree's total height will be predicted more precisely by a DH+ taper model than they would be predicted by a DH model. On the other hand, tree boles that do not include, or only include a small part of, the stem portion above 50 % of total height are unlikely to be predicted differently by a DH+ taper model compared to how they would be predicted by a DH one. This explains the observation that the decrease in volume prediction error, due to use of the M3_HABH taper model, was largest for total aboveground volume and smallest for sawtimber volume (Table 4). Sawtimber top diameter limits were, on average, at a relative height of 60 % for trees in the split-validation data set and at a relative height of 67 % for those in the independent-validation data set. Pulpwood top diameter limits were, on average, at a relative height of 78 and 81 % for trees in the split-validation and for those in the independent-validation data sets, respectively.

For prediction of diameters up to 60 % of total height, the BASE model performed better on trees in the independent-validation data set than it did on those in the split-validation data set (Fig. 4a versus Fig. 4b). This observation was contrary to expectation as a model would be expected to perform better on data that is more closely related to the fitting data set. The possible explanation for this is that trees in the independent-validation data set had profiles that were more clustered around the mean profile of the fitting data set where the BASE model, which was essentially a global model, is expected to perform better (see Fig. 2). Profiles of trees in the independent-validation are likely to exhibit this pattern because the trees were from stands that were less dispersed geographically (Fig. 1), covered a narrower range of ages (23 to 25 years compared to 9 to 39 years for trees in the fitting and split-validation data sets) and covered a narrow range of stand densities (230 to 280 trees ha^{-1} compared to 150 to 700 trees ha^{-1} in the fitting and split-validation data sets). The difference in BASE model performance for the lower two thirds of the tree bole, between the split-validation and the independent-validation data sets, can also be seen in the sawtimber RMSEs and magnitude of change in these RMSEs when the M3_HABH taper model is used instead of the BASE model (Table 4). Split-validation data sawtimber RMSE was larger, and it reduced by a greater magnitude (approximately 3.5 percentage points compared to close to 2.0 percentage points for the independent-validation data set) when the M3_HABH taper model was used. Whether or not a data set, on which one wishes to apply a taper model, falls in the portion of the model fitting data where a global model performs best is in most cases unknown to a user. Thus, use of upper stem diameters in polynomial taper models could make the models robust to effects of unknown differences between model fitting and independent model application data sets.

Conclusions

If use of an upper stem diameter in a no-volume-compatibility-constraint polynomial taper equation for New Zealand radiata pine is desired, a diameter measured at half the distance between breast height and the tip of the tree would be the most effective upper stem diameter to use. Consequently, the model of choice would be the version of the Gordon and Budianto (1999) model in which a tree's taper curve is constrained to pass through this upper stem diameter, i.e. M3_HABH. The original Gordon and Budianto (1999) model, GB99, which included an upper stem diameter measured at 6 m above the ground among its predictor variables, may not give predictions that are different from those of a polynomial model of the same form but without upper stem diameters among the predictor variables.

With use of the version of the Gordon and Budianto (1999) model that includes a diameter at half the distance between breast height and the tip of the tree (M3_HABH), considerable gains in stem volume prediction precision may be realized especially for total aboveground volume. It should, however, be noted that in such an application, gains in diameter and/or volume prediction precision may be lower if the measurement precision for the upper stem diameters is low. Upper stem diameters used in the current study were measured without error or were estimated (through quadratic interpolation) with a relatively high precision. It should also be noted that the reported volume prediction errors are based on outside bark diameters. Outside bark to inside bark diameter ratios may be used to adjust predictions from the reported taper equations if inside bark volumes are desired.

The study also showed that taper equation form may affect the decision on which upper stem diameter is best to include in the equation as an additional predictor. The diameter at half the height between breast height and the tip of the tree found to be the best in the current study was slightly different from the 60 % of total height found in an earlier study with a segmented model.

Author's contributions

The author conceived the study, carried out all the analyses and wrote the manuscript.

Author's information

The author is currently an Assistant Professor of Forest Biometrics at Mississippi State University, Mississippi State, MS, USA. Previously, he was a Scientist for Forest Modelling and Biometrics at Scion in Rotorua, New Zealand.

Competing interests

The author declares that he has no competing interests.

References

Arias-Rodil, M., Diéguez-Aranda, U., Puerta, F. R, López-Sánchez, C. A., Líbano, E. C., Obregón, A. C., et al. (2015). Modelling and localizing a stem taper function for *Pinus radiata* in Spain. *Canadian Journal of Forest Research, 45*(6), 647–658. doi:10.1139/cjfr-2014-0276.

Cao, Q. V. (2009). Calibrating a segmented taper equation with two diameter measurements. *Southern Journal of Applied Forestry, 33*(2), 58–61.

Cao, Q. V., & Wang, J. (2011). Calibrating fixed- and mixed-effects taper equations. *Forest Ecology and Management, 262*(4), 671–673. doi:10.1016/j.foreco.2011.04.039.

Czaplewski, R. L., & McClure, J. P. (1988). Conditioning a segmented stem profile model for two diameter measurements. *Forest Science, 34*(2), 512–522.

Diéguez-Aranda, U., Castedo-Dorado, F., Álvarez-González, J. G., & Rojo, A. (2006). Compatible taper function for Scots pine plantations in northwestern Spain. *Canadian Journal of Forest Research, 36*(5), 1190–1205. doi:10.1139/x06-008.

Gómez-García, E., Crecente-Campo, F., & Diéguez-Aranda, U. (2013). Selection of mixed-effects parameters in a variable–exponent taper equation for birch trees in northwestern Spain. *Annals of Forest Science, 70*(7), 707–715.

Gordon, A. D. (1983). Comparison of compatible polynomial taper equations. *New Zealand Journal of Forestry Science, 13*(2), 146–155.

Gordon, A. D., & Budianto, M. (1999). *A 3-point stem volume and taper equation for radiata pine*. Rotorua: Forest and Farm Plantation Management Cooperative. Report No. 66.

Gordon, A. D., Lundgren, C., & Hay, E. (1995). Development of a composite taper equation to predict over-and under-bark diameter and volume of *Eucalyptus saligna* in New Zealand. *New Zealand Journal of Forestry Science, 25*(3), 318–327.

Gordon, A. D., Lundgren, C., & Hay, E. (1999). Composite taper equations to predict over-and under-bark diameter and volume of *Eucalyptus pilularis, E. globoidea,* and *E. muelleriana* in New Zealand. *New Zealand Journal of Forestry Science, 29*(2), 311–317.

Goulding, C., & Murray, J. (1976). Polynomial taper equations that are compatible with tree volume equations. *New Zealand Journal of Forestry Science, 5*(3), 313–322.

Hayward, W. J. (1987). Volume and taper of *Eucalyptus regnans* grown in the central North Island of New Zealand. *New Zealand Journal of Forestry Science, 17*(1), 109–120.

Katz, A., Dunningham, A. G., & Gordon, A. D. (1984). *A Compatible volume and taper equation for New Zealand Pinus radiata D. Don grown under the direct sawlog regime*. Rotorua: New Zealand Forest Service. FRI Bulletin No. 67.

Kilkki, P., Saramäki, M., & Varmola, M. (1978). A simultaneous equation model to determine taper curve. *Silva Fennica, 12*(2), 120–125.

Kilkki, P., & Varmola, M. (1979). A nonlinear simultaneous equation model to determine taper curve. *Silva Fennica, 13*(4), 293–303.

Kozak, A. (1998). Effects of upper stem measurements on the predictive ability of a variable-exponent taper equation. *Canadian Journal of Forest Research, 28*(7), 1078–1083.

Lappi, J. (2006). A multivariate, nonparametric stem-curve prediction method. *Canadian Journal of Forest Research, 36*(4), 1017–1027. doi:10.1139/x05-305.

Laurie Forestry (2016). Marketing services. http://www.laurieforestry.co.nz/Forestry-Marketing. Accessed 2 July 2016.

Max, T. A., & Burkhart, H. E. (1976). Segmented polynomial regression applied to taper equations. *Forest Science, 22*(3), 283–289.

Sabatia, C. O., & Burkhart, H. E. (2015). On the use of upper stem diameters to localize a segmented taper equation to new trees. *Forest Science, 61*(3), 411–423. doi:10.5849/forsci.14-039.

SAS Institute Inc. (2002 - 2004). SAS 9.4 Help and Documentation. SAS Institute Inc.: Cary, NC.

Sharma, M., & Parton, J. (2009). Modeling stand density effects on taper for jack pine and black spruce plantations using dimensional analysis. *Forest Science, 55*(3), 268–282.

Yang, Y., Huang, S., Trincado, G., & Meng, S. X. (2009). Nonlinear mixed-effects modeling of variable-exponent taper equations for lodgepole pine in Alberta, Canada. *European Journal of Forest Research, 128*(4), 415–429.

Zimmerman, D. L., & Núñez-Antón, V. (2001). Parametric modelling of growth curve data: an overview. *Test, 10*(1), 1–73.