Forestry An International Journal of Forest Research

Institute of Chartered Foresters

Forestry 2016; **89**, 215–229, doi:10.1093/forestry/cpw004
Advance Access publication 10 February 2016

Using quadratic mean diameter and relative spacing index to enhance height–diameter and crown ratio models fitted to longitudinal data

Pradip Saud[1]*, Thomas B. Lynch[1], Anup K. C.[1] and James M. Guldin[2]

[1]*Department of Natural Resource Ecology and Management, Oklahoma State University, Ag Hall Room 008C, Stillwater, OK 74078-6013, USA*
[2]*US Forest Service, Southern Research Station, Hot Spring, AR 71902, USA*

*Corresponding author. Tel: +1 4057445447; Fax: +4057443530; E-mail: pradipsaud@hotmail.com

Received 24 July 2015

The inclusion of quadratic mean diameter (QMD) and relative spacing index (RSI) substantially improved the predictive capacity of height–diameter at breast height (d.b.h.) and crown ratio models (CR), respectively. Data were obtained from 208 permanent plots established in western Arkansas and eastern Oklahoma during 1985–1987 and remeasured for the sixth time (2012–2014). Existing height–d.b.h. and CR estimation models for naturally occurring shortleaf pine forests (*Pinus echinata* Mill.) were updated and modified for improved performance. Additionally, eight height–d.b.h. relationship models that use only d.b.h. (fundamental local models) were modified using covariates. The model performance was evaluated using fit statistics [root mean square error (RMSE), Fit index and Akaike information criteria (AIC)]. The results showed that the best model form which was an extended non-linear model with autoregressive first order AR (1) structure and power variance function performed better than extended mixed-effects models and predicted well as an ordinary least squares non-linear model. The autocorrelation within individual trees was larger for the height–d.b.h. relationship than for CR estimation. The addition of QMD to mean dominant height (H_D) greatly improved height–d.b.h. relationship with a reduction of 8 per cent in RMSE, compared with the use of basal area per hectare. Similarly, multiplying a fundamental local model by using QMD raised to a parametric power reduced RMSE by 16 per cent, improved Fit index by 12 per cent and decreased the AIC value by 7 per cent. D.b.h., H_D and RSI best explained the crown ratio relationship with an improved Fit index by 6.7 per cent compared with alternative non-linear models without RSI. The logistic model for CR also provided prediction accuracy similar to that of a commonly used non-linear model. A non-linear model with an application of remedial measures to enhance adherence to modelling assumptions can provide better parameter estimates than mixed-effects modelling approach.

Keywords: mixed-effects model, autocorrelation, height–d.b.h. relationship, crown ratio, quadratic mean diameter, relative spacing index

Introduction

Two fundamental mensurational quantities in forest inventory, i.e. height and diameter, are frequently used to characterize forest productivity in forest growth and yield models. Diameter measurement at breast height (d.b.h.) is less costly and requires less effort than the total height measurement. Therefore, height–d.b.h. relationship models have been used in the growth and yield models to predict ‘missing’ tree heights (Lynch and Murphy, 1995), to predict the future tree heights (Lynch and Murphy, 1995; Lynch *et al.*, 1999) and also to impute height to estimate volume production (Garber *et al.*, 2009). The height–d.b.h. relationship is non-linear from the biological perspective with a curve that is asymptotic to a maximum possible total height at upper ranges of diameter. Therefore, various models that were non-linear with respect to parameters have been proposed to model the height–d.b.h. relationship of different tree species. The height–d.b.h. has sometimes been modelled using only the d.b.h. as a single independent variable (Meyer, 1940; Richards, 1959; Burkhart and Strub, 1974; Stage, 1975; Bates and Watts, 1980; Wykoff *et al.*, 1982; Ratkowsky and Giles, 1990; Schmidt *et al.*, 2011; VanderSchaaf, 2014; Sharma and Breidenbach, 2015). Here, we term this type of model as ‘fundamental local model’ because these models are: primarily developed at local or at a regional level, specific to a tree species and site, developed when stand-level covariate or competition index is difficult or inconvenient to obtain, and can be easily extended to incorporate additional covariates.

Height–d.b.h. relationships are also modelled using plot or stand-level covariates to demonstrate the influence of stand density or effect of competition (Lynch *et al.*, 1999; Sharma and Parton, 2007; Temesgen *et al.*, 2007, 2008; Budhathoki *et al.*, 2008; Arcangeli *et al.*, 2014; Temesgen *et al.*, 2014; Sharma and

Breidenbach, 2015). Therefore, fundamental local models are expected to have large error variance (Huang *et al*., 1992; Fang and Bailey, 1998; Sharma and Breidenbach, 2015) compared with the models that use additional covariates to d.b.h. such as stand structures (dominant height) (Lappi, 1997; Lynch *et al*., 1999; Sharma and Breidenbach, 2015), relative dimensions at the stand level [ratio of trees per hectare to basal area per hectare (BAH)] (Sharma and Parton, 2007), competition index [basal area in larger trees (BAL)] (Temesgen *et al*., 2007) and stand density (BAH) (Sharma and Yin Zang, 2004; Budhathoki *et al*., 2008).

Individual tree height (H_i) and d.b.h. (D_i) can be used as independent variables with other plot or stand-level attributes to model another important tree characteristic, the 'crown ratio' (CR) of an individual tree which is the ratio of live crown length to the total height. The crown ratio is an important measure of tree vigour that reflects competition experienced by an individual tree because stand density over the period reduces the crown length (Smith *et al*., 1992; Hynynen, 1995; Temesgen *et al*., 2005). The distance between trees determines the crown shape and size, which is related to the crown length, total height and diameter increment attained by an individual tree (Smith *et al*., 1992; Monserud and Sterba, 1996). In many growth and yield models, CR is used for improved prediction of forest attributes. For example, height increment (Daniels and Burkhart, 1975), basal area increment (Wykoff, 1990; Monserud and Sterba, 1996; Leites *et al*., 2009), taper and volume of individual trees (Valenti and Cao, 1986; Jiang *et al*., 2007; Jiang and Liu, 2011) and survival of an individual tree (Saud *et al*., 2016). It is also useful in estimating crown biomass for energy production (Tahvanainen and Forss, 2006). The crown ratio has been modelled using either a variety of non-linear functions (Holdaway, 1986; Dyer and Burkhart, 1987; Lynch *et al*., 1999) or the logistic function (Hasenauer and Monserud, 1996; Temesgen *et al*., 2005) which is also a non-linear function. Both approaches utilize model forms that restrict crown ratio predictions to the feasible range of 0–1.

However, the height–d.b.h. relationship can vary over time due to differences in stand age, productivity and competition (Lappi, 1997; Peng *et al*., 2001; Sharma and Parton, 2007; Budhathoki *et al*., 2008) and also differences in a geographical region (Calama and Montero, 2004; Arcangeli *et al*., 2014). This also applies in the case of *CR* modelling. Such variation could be reduced by using distance-independent variables as covariates. Relative dimensions (ratios) are distance-independent indices, which measure the hierarchical position of the subject tree within plot, e.g. ratio of D_i to max d.b.h.; ratio of quadratic mean diameter (QMD) to d.b.h. (RAQD), BAL and crown competition factor in larger trees (CCFL). These dimensionless ratios when used as covariates help to assure better prediction and tend to make the fundamental relationships among tree components more stable (Burkhart and Tomé, 2012, p. 202). Ducey (2009) and Zhao *et al*. (2012) demonstrated that the relative spacing index (RSI) accounted for the effects of space between individual trees on crown ratio more effectively than other covariates. Interestingly, in addition to RSI, Ducey (2009) also suggested the inverse of RAQD as an important variable in height prediction and CR estimation. Temesgen *et al*. (2005, 2007, 2008, 2014) suggested that distance-independent variables including BAL and CCFL also improve fits for tree height–d.b.h. and crown relationships.

Height–d.b.h. relationship and CR models are often developed using mensurational records of permanent plots from repeated measurements. The tree attributes such as height, d.b.h. and crown length measured at different time intervals are autocorrelated and also exhibit heterogeneous errors either at a tree or stand level. As a result, the use of non-linear ordinary least square (OLS) estimation is often not reliable, because an assumption of random samples and independent observations is violated and the presence of autocorrelation does not conform to the assumptions of OLS. Therefore, mixed-effects modelling as an alternative to OLS for repeated measurements and grouped data has been widely used in forestry growth and yield models (Lappi, 1991; Lynch *et al*., 2005, 2012; Budhathoki *et al*., 2008; Temesgen *et al*., 2014). This approach helps to address a possible source of subject-specific variation that the OLS approach does not consider because the fixed-effect parameters represent population average responses, while random effects parameters represent response specific to each sampling unit (Lappi, 1991; Lynch *et al*., 2005). One advantage of the mixed-effects model is that random effects can be calibrated for an unsampled location (new data that were not part of the original estimation data) to improve the predictive accuracy for the resulting calibrated mixed-effects model (Lappi, 1991; Peng *et al*., 2001; Lynch *et al*., 2005, 2012; Sharma and Parton, 2007; VanderSchaaf, 2014).

Mixed-effects models or hierarchical mixed-effects models easily account for spatial autocorrelation by using a plot-specific or group-specific random effects, but not so for temporal autocorrelation within observations. However, shorter study time period (lag) has been a limiting factor in the analysis of longitudinal data in growth modelling that sometimes does not allow us to specify appropriate autocorrelation structures. Patterns of heterogeneous errors can often be associated with the covariates (Pinheiro and Bates, 2000). Many investigators have used graphical methods to investigate the assumption of constant error variance for the d.b.h.–height relationship model (Fang and Bailey, 1998; Peng *et al*., 2001; Sharma and Parton, 2007; Budhathoki *et al*., 2008). But perhaps few have done a formal statistical test on this issue for non-linear model forms (Temesgen *et al*., 2007; Lynch *et al*., 2012). Others have used weighted regression or logarithmic transformations that tend to stabilize variance (Huang *et al*., 1992; Fang and Bailey, 1998; Temesgen *et al*., 2007, 2014).

Shortleaf pine (*Pinus echinata* Mill.) forests contain standing volume in the southern US second only to loblolly pine (*Pinus taeda* L.) among the four southern pines (Lawson, 1990). However, relatively few quantitative studies of the height–d.b.h. relationship, CR or other aspects of growth and yield of natural stands of shortleaf pine have been published compared with other southern pines in the US (Budhathoki *et al*., 2008). Graney and Burkhart (1973) provided a polymorphic system of site index curves to estimate dominant stand height for shortleaf pine using non-linear ordinary least square (OLS) method. The relationship between height–d.b.h. for naturally occurring even-aged stands of shortleaf pine was fitted using seemingly unrelated regression by Lynch and Murphy (1995), non-linear OLS by Lynch *et al*. (1999) and mixed-effects estimation by Budhathoki *et al*. (2008). The studies of Lynch *et al*. (1999) for height–d.b.h. relationship and CR estimation and of Budhathoki *et al*. (2008) for height–d.b.h. relationship used only the first two and three measurements, respectively, of the Oklahoma State University (OSU) and USDA Forest Service Southern Research Station (USFS) naturally occurring shortleaf pine growth study.

Although, Budhathoki *et al*. (2008) updated height–d.b.h. relationship by adding BAH as an independent variable and by fitting a

mixed-effects model, they did not fit a crown ratio model. But now six measurements of shortleaf plot data spanning a 25-year period are available that allow us to update existing models and incorporate the most recent developments in modelling the height–d.b.h. and crown relationships. This also provides an opportunity to test other model forms and independent covariates that may provide improved fits to the data from other studies as well. Therefore, the current study aims to update and improve the existing height prediction and CR estimation models suitable for practical application by resolving the issues of autocorrelation and heteroscedasticity of errors in repeated measurements. The specific aims are: to modify and improve the model performance by introducing plot or stand-level covariates and to evaluate the performance of different sets of OLS non-linear models, mixed-effects models and their extended forms while correcting for autocorrelation and stabilizing heterogeneity of errors. In addition to this, we will evaluate available fundamental local models for the height–d.b.h. prediction that do not use dominant height and modify them with plot or stand-level covariates to minimize the prediction error. Moreover, also we will test homoscedasticity assumptions independently for each model. It is expected that the best model for both the height prediction and the crown ratio estimation will be useful for estimating volume, biomass and other tree attributes of the natural stand of shortleaf pine.

Materials and methods

Data

In 1985–1987, the Department of Forestry (now part of the Department of Natural Resource Ecology and Management) at Oklahoma State University (OSU) and the US Forest Service (USFS) at Monticello, Arkansas, collaboratively established growth and yield plots in even-aged natural shortleaf pine stands. The plots were established as permanent plots in the Ozark and Ouachita National Forests in western Arkansas and southeastern Oklahoma. Prior to the establishment of this study in 1985, the major sources of data for shortleaf growth and yield were from fully stocked plots or unmanaged shortleaf pine stands (Lynch *et al*., 1999). These plots were designed to represent a range of ages, basal area levels and site qualities, which were designated as design variables so that plots were thinned to specific residual densities at their establishment (for a detailed description, see Lynch *et al*., 1999).

The measurement plots were circular with a radius of 17.4 m (57.2 feet) and area of 0.0809 ha (0.2 acres). The measurement plots were surrounded by a buffer strip 10 m (33 feet) wide that received the same silvicultural treatments as plots at establishment. At plot establishment, woody understory vegetation stems with d.b.h. >2.54 cm in were controlled using herbicide. Based on the definitions given by Avery and Burkhart (2002, p. 163), individual tree crowns were classified as dominant, codominant, intermediate or suppressed trees. The average height of dominants and codominant was used as the dominant height in this study and was also used to determine site index for each plot. The ring counts of increment cores for the dominants and codominant were averaged to obtain plot age (PAG).

The total sample consisted of 208 plots. These plots have been remeasured in every 4–6 years, with the latest (sixth) measurement made during the period from 2012 to 2014. At each measurement, D_i of all trees from plot were measured, but H_i in metres and crown length (height to base of live crown) in metres were recorded for selected subsample trees from each plot to represent the range of tree diameters and crown classes of dominant, codominant and intermediate trees on the plot. At least two trees per 2.5 cm d.b.h. class were selected where available. At plot establishment, a procedure was followed which attempted to achieve an approximately even distribution of height measurement trees within diameter classes and to spread the measurements approximately evenly on the plot. This was based on an initial d.b.h. tally. The initial goal was to have at least two and no more than five trees per d.b.h. class. The numbers of height subsample trees within plots were increased subsequent to the first measurement in order to continue a good representation of samples within plot d.b.h. classes (e.g. if due to growth or mortality, there were no longer two trees per d.b.h. class). Once selected, all height measurement trees were remeasured in later remeasurements unless the tree died before the sixth measurement.

Ice-damaged trees occurred on a total of 101 plots in the year 2001, just before the fourth measurement (for a detail information see Stevenson *et al*., 2016). It was expected that the plots with significant numbers of ice-damaged trees could influence the growth characteristics of individual trees on these plots. Therefore, in the model development process, plots with more than 30 per cent of ice-damaged trees were excluded. Any individual trees having ice damage was also removed from the model development process. Trees having forks or other significant defects were excluded from the model development dataset. Many plots were re-thinned to their original basal area levels just after the third measurement, while some plots were left unthinned. An average thinned (removed) basal area was 6.94 $m^2\ ha^{-1}$ with a range of 0.69–19.36 $m^2\ ha^{-1}$. A variable that exhibits simple thinning effect 'THINHA' = (Thinned basal area per hectare/(years since thinning) was formulated assuming that thinning effect decreases over the time. The mean and standard deviation (SD) of all variables from the first measurement to the last (sixth) measurement are shown in Table 1. The data consisted of total 14 028 observations, and the summary statistics of the variables used in modelling height prediction and crown ratio estimation are presented in Table 2. The H_i ranged from 3.048 to 38.100 m with a mean of 20.433 m and SD of 6.222 m and CR ranged from 0.055 to 0.80 with a mean of 0.373 and SD of 0.094 (Table 2). The data used for modelling height–d.b.h. relationship and CR estimation are shown in Figure 1a,b.

Height prediction model with dominant height

To predict individual tree height, Lynch and Murphy (1995) developed a compatible height prediction model which was used in the shortleaf pine growth prediction system described in Lynch *et al*. (1999). Equation (1) was used to predict an individual shortleaf pine tree height. Budhathoki *et al*. (2008) accounted for competition effects of other trees on individual tree growth by including the variable BAH [equation (2)].

$$(H_i - c) = b_0(H_D - c)^{b_1} \exp(b_2 D_i^{b_3}) \quad (1)$$

$$(H_i - c) = b_0(H_D - c)^{b_1} \exp(b_2 D_i^{b_3} + b_4 \mathrm{BAH}) \quad (2)$$

where H_D is the average plot dominant and codominant height, c the breast height (1.371 m) at which d.b.h. is measured and b_0, b_1, b_2, b_3 and b_4 are parameters to be estimated.

The existing models need to be tested for thinning effect since post-thinning measurements are now available. The simple effect of thinning '*THINHA*' on height prediction model [equation (1)] was found not significant in the model. Therefore, we tested inclusion of covariates that represent stand density which might reflect competition levels as well as tree relative position in the stand. Both covariates, QMD and RAQD, provided a substantial improvement in the fit statistics and showed identical performance, but we preferred QMD in equation (3) to avoid correlated covariates in the model, which will be discussed below. A variable that is a function of an inverse of QMD, the ratio of the number of trees per hectare (TPH) to BAH (TPH/BAH), was also used by Sharma and Parton (2007) in the extended model of the Chapman–Richards function for boreal tree species in Ontario, Canada. The model used by Sharma and Parton (2007) was also modified by using QMD variable, instead TPH/BAH in equation (4) for testing on our

Table 1 Descriptive statistics (mean with SD in parentheses) of stand level and tree variables recorded for six measurement times of naturally occurring even-aged shortleaf pine stand

Variables	Measurements					
	1st ($n = 2682$)	2nd ($n = 3017$)	3rd ($n = 3215$)	4th ($n = 1750$)	5th ($n = 1677$)	6th ($n = 1687$)
CR	0.373 (0.094)	0.365 (0.096)	0.372 (0.096)	0.376 (0.089)	0.374 (0.092)	0.364 (0.084)
D (cm)	28.991 (10.974)	26.426 (10.388)	28.333 (10.5)	31.195 (10.003)	33.276 (10.33)	35.021 (10.68)
H (m)	19.864 (6.222)	18.684 (6.34)	19.894 (6.087)	20.777 (5.396)	21.698 (5.215)	22.877 (5.221)
H_D (m)	20.433 (5.596)	19.364 (5.704)	20.625 (5.356)	21.17 (4.83)	22.131 (4.56)	23.293 (4.477)
PAG (years)	62.465 (22.114)	56.585 (20.207)	61.537 (20.258)	66.486 (20.742)	72.251 (20.737)	77.828 (20.756)
BAH ($m^2\ ha^{-1}$)	24.191 (8.653)	23.135 (7.566)	26.156 (8.352)	23.609 (8.693)	26.027 (9.147)	28.063 (9.565)
QMD (cm)	28.997 (9.594)	26.442 (8.942)	28.468 (8.816)	31.118 (8.644)	33.256 (8.861)	35.036 (9.107)
RSI	0.276 (0.101)	0.269 (0.092)	0.256 (0.091)	0.293 (0.112)	0.284 (0.109)	0.274 (0.107)
RAQD	1.063 (0.323)	1.073 (0.347)	1.078 (0.359)	1.043 (0.269)	1.044 (0.268)	1.045 (0.269)
BAHG ($m^2\ ha^{-1}$) years^{-1}	0.458 (0.274)	0.491 (0.296)	0.505 (0.303)	0.409 (0.241)	0.41 (0.23)	0.404 (0.214)

n, total number of observations; CR, crown ratio; D, diameter at breast height (cm); H, individual tree height; H_D, average plot dominant and co-dominant height (m); PAG, plot age (years); BAH, stand basal area per hectare ($m^2\ ha^{-1}$); QMD, quadratic mean diameter (m); RSI, relative spacing index; RAQD, ratio of QMD to D; and BHAG, ratio of BAH to PAG.

Table 2 Summary statistics of the variables used to model height prediction and crown estimation of the naturally occurring even-aged shortleaf pine stand ($n = 14\,028$)

Variables	Mean	SD	Minimum	Maximum
CR	0.373	0.094	0.055	0.800
D (cm)	28.991	10.974	2.794	67.564
H (m)	19.864	6.222	3.048	38.100
H_D (m)	20.433	5.596	6.706	36.019
PAG (years)	62.465	22.114	18.000	119.000
BAH ($m^2\ ha^{-1}$)	24.191	8.653	2.035	48.684
QMD (cm)	28.997	9.594	7.887	58.258
RSI	0.276	0.101	0.126	0.841
RAQD	1.063	0.323	0.386	7.012
BAHG ($m^2\ ha^{-1}$) years^{-1}	0.458	0.274	0.025	1.338

shortleaf pine dataset.

$$(H_i - c) = b_0(H_D - c)^{b_1} \exp(b_2 D_i^{b_3} \mathrm{QMD}) \tag{3}$$

$$(H_i - c) = b_0(H_D)^{b_1}(1 - \exp^{-b_2 \mathrm{QMD}^{b_3} D_i})^{b_4} \tag{4}$$

After testing several models using OLS, equation (3) was selected for fitting with the mixed-effects model approach because it provided root mean square error (RMSE) extremely close to the smallest RMSE from equation (4), but contains only four parameters, making it somewhat simpler in form than equation (4). All parameters were tested for possible inclusion of plot-level random effects assuming the same plot effect holds for all remeasurements of the same plot. Plot level random effects associated with b_0 (asymptotic height), b_1 (slope) and b_2 (curvature) were significant. However, due to sizeable differences in the Akaike information criteria (AIC) values, the random effect associated with the b_3, the parameter multiplicative to exponent of D_i in equation (3), was selected for use in the final model. This resulted in the mixed-effects model [equation (5)]

$$(H_{ij} - c) = b_0(H_{Dj} - c)^{b_1} \exp\{b_2 D_{ij}^{(b_3+u_j)} \mathrm{QMD}_j\} + \varepsilon_{ij} \tag{5}$$

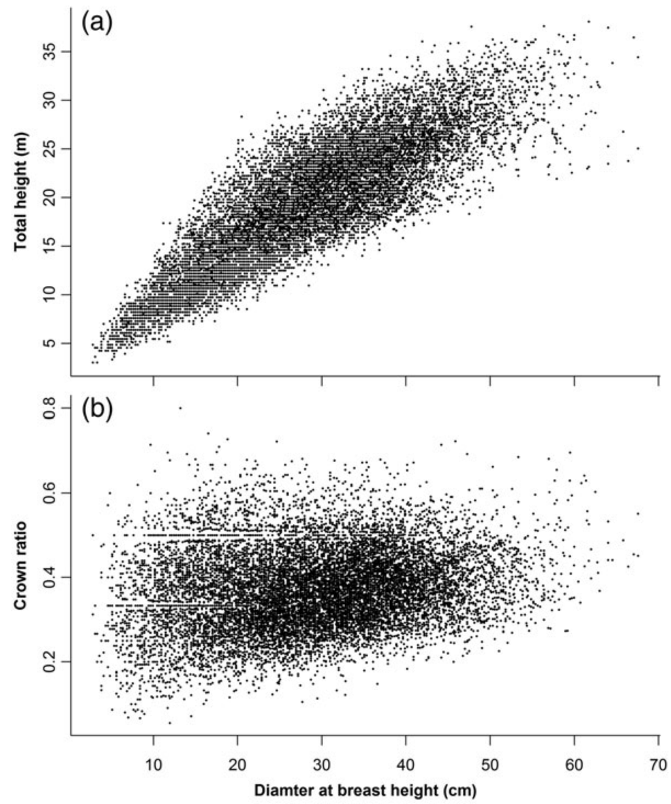

Figure 1 Scatter plot of whole data showing the distribution of height (a) and the crown ratio (b) of shortleaf pine along the diameter range.

Table 3 Fundamental local models for height prediction with d.b.h. only, and with modified height prediction model with QMD

Equation	Common height model	Equation	Modified model	Source
12	$H_i - c = b_0(1 - e^{-(b1Di)})$	20	$H_i - c = b_0(1 - e^{-(b1Di)})QMD^{b2}$	Meyer (1940)
13	$H_i - c = b_0(1 - e^{-b1Di})^{b2}$	21	$H_i - c = b_0(1 - e^{-b1Di})^{b2}QMD^{b3}$	Richards (1959)
14	$H_i - c = b_0 e^{b1/Di}$	22	$H_i - c = b_0 e^{b1/Di}QMD^{b2}$	Burkhart and Strub (1974)
15	$H_i - c = b_0 D_i^{b1}$	23	$H_i - c = b_0 D_i^{b1}QMD^{b2}$	Stage (1975)
16	$H_i - c = b_0 D_i(b_1 + D_i)$	24	$H_i - c = b_0 D_i(b_1 + D_i)QMD^{b2}$	Bates and Watts (1980)
17	$H_i - c = e^{b0 + (b1/(Di + 1))}$	25	$H_i - c = e^{b0 + (b1/(Di + 1))}QMD^{b2}$	Wykoff *et al.* (1982)
18	$H_i - c = b_0 e^{b1/(Di + b2)}$	26	$H_i - c = b_0 e^{b1/(Di + b2)}QMD^{b3}$	Ratkowsky and Giles (1990)
19	$H_i - c = (D_i/(b_1 + b_2 D_i))^{b3}$	27	$H_i - c = (D_i/(b_1 + b_2 D_i))^{b3}\ QMD^{b4}$	Schmidt *et al.* (2011), Sharma and Breidenbach (2015)

where H_{ij} is total height of tree i in plot j; i the attribute of an individual tree; j the attribute of the plot; b_0, b_1, b_2 and b_3 are fixed effects parameters; u_j the random effect associated with parameter b_3 and specific to plot j and ε_{ij} within-plot error (random error for tree i in plot j).

Since each tree from the same plot was repeatedly measured six times, it is possible that autocorrelation within individual observations and non-constant error variance exists. Therefore, to resolve the issues of heteroscedasticity and autocorrelation, we modelled a power variance function and first order autoregressive AR (1) structure, and a combination of both in the best non-linear and mixed-effects models (Pinheiro and Bates, 2000, p. 391; Cryer and Chan, 2008, p. 66). AR (1) structure was selected because the data have relatively fewer lags (5) and AR (1) requires the estimation of only one parameter. The best selected non-linear models [equation (3)] and mixed-effects model [equation (5)] were referred to as 'base model' forms. Hereafter, the extended model forms with additional power variance function and AR (1) were termed as 'extended non-linear models' (ENMs) and 'extended mixed-effects models' (EMEMs). The assumption of autocorrelation within an individual tree was considered for ENMs, but this assumption was not compatible while modelling mixed-effects models. It may be due to the differences in a hierarchy of the group at which the errors are correlated, and random effects are associated. So, we assumed autocorrelation within plot for mixed-effects models. The resulted ENMs are as:

$$\text{Model 3} + \text{power variance function} \tag{6}$$

$$\text{var}(\varepsilon_{ij}) = \sigma^2 |\upsilon_{ij}|^{2\delta}$$

error variance (ε_{ij}) was modelled with one covariate. υ_{ij} is the covariate and δ the power parameter. H_D was selected as covariate for modelling heterogenouses errors because of smaller AIC value and large likelihood ratio (LR) statistics.

$$\text{Model 3} + \text{AR (1)} \tag{7}$$

AR (1) models the correlated errors (ε_{it}) as:

$$\varepsilon_{it} = \phi_1 \varepsilon_{it-1} + \omega_{it} \quad \text{and} \quad \omega_{it} \sim iid\, N(0, \sigma_\omega^2)$$

where i is the individual tree and t the measurement (lag) and ϕ the autocorrelation between lags. The correlation 'ρ' between residuals of an observation pair declines exponentially with the number of periods (t) apart, i.e. $\rho = \phi^t$.

$$\text{Model 3} + \text{power variance function} + \text{AR (1)} \tag{8}$$

The resulted EMEMs are as:

$$\text{Model 5} + \text{power variance function} \tag{9}$$

$$\text{Model 5} + \text{AR (1)} \tag{10}$$

$$\text{Mode 5} + \text{power variance function} + \text{AR (1)} \tag{11}$$

Height prediction model without dominant height

A variety of height prediction models with only the single independent variable 'd.b.h.' are commonly used in practice when dominant height is not readily available [in Table 3, equations (12–19)]. These eight equations [equations (20–27)] were fitted to predict the shortleaf pine tree height and tested for modification by including QMD, which is typically available from forest inventory data. Other stand-level variables including TPH, BAH and RAQD were also tested. We found the inclusion of QMD into these modified 'local' models [equations (20–27)] substantially improved model performance. But, these models were neither tested as a mixed-effects model nor as extended models.

The individual tree crown ratio model used by Lynch *et al.* (1999) for shortleaf pine was also modified and tested. This crown ratio function was developed by Dyer and Burkhart (1987) for planted loblolly pine tree data and also used by Hynynen (1995) for Scots pine stands. The base equation (28) used to predict the current individual tree shortleaf pine crown ratio together with an alternative equation (29) is shown as:

$$CR_i = 1 - \exp\left[-\left(b_0 + \frac{b_1}{PAG}\right)\left(\frac{D_i}{H_i}\right)^{b_2}\right] \tag{28}$$

$$CR_i = 1 - \exp\left[-\left(b_0 + \frac{b_1}{H_i}\right)\left(\frac{D_i}{H_D}\right)^{b_2}\right] \tag{29}$$

where CR_i is the crown ratio of tree i; b_0, b_1 and b_2 are parameters to be estimated; and other variables are as defined above.

The effect of thinning was not included in crown ratio estimation models because although the effect was found significant in equation (28), but it did not markedly reduce the mean square error. Equation (28) was modified to equation (29) by replacing PAG with H_D. Both equations were further modified to equations (30 and 31) by adding BAH. Equation (30) was then modified by using RSI in equation (32). RSI was calculated as $=\left(\sqrt{10000/TPH}\right)/H_D$.

$$CR_i = 1 - \exp\left[-\left(b_0 + \frac{b_1}{PAG} + BAH^{b_2}\right)\left(\frac{D_i}{H_i}\right)^{b_3}\right] \tag{30}$$

$$\mathrm{CR}_i = 1 - \exp\left[-\left(b_0 + \frac{b_1}{H_i}\right)\left(\mathrm{BAH}^{b_2} + \frac{D_i}{H_D}\right)^{b_3}\right] \quad (31)$$

$$\mathrm{CR}_i = 1 - \exp\left[-\left(b_0 + \frac{b_1}{H_i}\right)\left(\mathrm{RSI}^{b_2} + \frac{D_i}{H_D}\right)^{b_3}\right] \quad (32)$$

The logistic function approach to crown ratio estimation was also tested. The following model [equation (33)] proved to be a good alternative to the exponential model approach given above.

$$\mathrm{CR}_i = \left[1 + \exp\left(-\left(b_0 + b_1 H_i + b_2 \frac{D_i}{H_D} + b_3 \mathrm{BAH}\right)\right)\right]^{-1} \quad (33)$$

The mixed-effects approach was also used to fit crown ratio estimation models. Based on performance, equations (31 and 32) were modelled with the mixed-effects approach. It was found that the random effect associated with the parameter b_0 for both models as shown below performed better than the other alternatives:

$$\mathrm{CR}_{ij} = 1 - \exp\left[-\left((b_0 + \mu_j) + \frac{b_1}{H_i}\right)\left(\mathrm{BAH}_j^{b_2} + \frac{D_i}{H_{D_j}}\right)^{b_3}\right] + \varepsilon_{ij} \quad (34)$$

$$\mathrm{CR}_{ij} = 1 - \exp\left[-\left((b_0 + \mu_j) + \frac{b_1}{H_i}\right)\left(\mathrm{RSI}_j^{b_2} + \frac{D_i}{H_{D_j}}\right)^{b_3}\right] + \varepsilon_{ij} \quad (35)$$

where CR_{ij} is the crown ratio of tree *i* in plot *j*.

The crown ratio estimation model with better fit index and AIC values: equation (26) non-linear model; and equation (35) mixed-effects model were selected as base models for modelling heterogeneous errors and autocorrelation structures. The resulted ENMs are: equations (36–38); and EMEMs models are: equations (39–41).

$$\text{Model 6} + \text{power variance function} \quad (36)$$

Individual d.b.h. was found a better performing covariate than alternative in reducing AIC value for modelling heterogenous errors.

$$\text{Model 26} + \text{AR (1)} \quad (37)$$

$$\text{Mode 26} + \text{power variance function} + \text{AR (1)} \quad (38)$$

$$\text{Model 35} + \text{power variance function} \quad (39)$$

$$\text{Model 35} + \text{AR (1)} \quad (40)$$

$$\text{Mode 35} + \text{power variance function} + \text{AR (1)} \quad (41)$$

Statistical analysis

All non-linear models, mixed-effects models and extended form of models were fitted in R (R Development Core Team, 2012) using the 'nls', 'nlme' and 'gnls' functions, respectively (Pinheiro *et al.*, 2014). Models were compared using the Fit index, RMSE and AIC. Likelihood ratio (LR) statistics was also used to compare the ENMs with the base non-linear model [equation (3)] and, EMEMs with base mixed-effects model [equation (5)]. The Fit index for non-linear models was calculated based on equation (42) and RMSE was calculated based on equation (43). Fit indices and RMSEs for mixed models and the extended models were calculated based on the actual height or crown ratio predictions using only parameter estimates of the fixed covariates while setting random effects equal to zero.

$$\text{Fit index} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (42)$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_i^n (y_i - \hat{y}_i)^2}{n - p}} \quad (43)$$

where y_i is the observed value for *i*th observation, $\hat{y}_i$ the predicted value by a model, $\bar{y}_i$ the mean observed value and *p* the number of parameters estimated by a model. For ease in comparison, Fit index was interpreted as the percentage (multiplied by 100).

The Goldfeld–Quandt test was used to test an assumption of homoscedasticity of the error variance for both height prediction and crown ratio estimation model (Judge *et al.*, 1988, pp. 371–372). The dataset was divided into three parts ordered from the smallest to the largest value of the independent variable (d.b.h.), and the middle (1/3rd) of the data were excluded. The Goldfeld–Quandt test compares the ratio of the residual sum of squares of the model from the upper range (3/8ths of total data) observations to the model from lower range (3/8ths of total data) observation. Standardized residual plots were plotted against the fitted (predicted) values, and against d.b.h. values. The standardized residuals were also plotted against the mid-range of the design variables (plot basal area, site index and plot age), but they are not shown. Standardized residual plots of all height prediction models were similar to each other, and the same was the case for CR models. Therefore, only standardized residual plots from the base model and the best model for both height prediction and crown ratio estimation are presented.

Results

The patterns of height and crown changes over the time for each measurement can be observed in Figure 2. Some of the changes could be due to thinning from below after the third measurement and removal of many trees due to ice storm damage at the fourth (Figure 2a). Due to the study design, large trees in older age classes were present even at the first measurement since the study included a balanced range of age classes at that time. During later measurements, trees in the younger age classes grew in height, resulting in an increase in the mean height for the study as a whole (Table 2). The mean crown ratio appears to be relatively constant over time, and possibly average increases in total height are balanced by the crown recession over time as might be expected (Figure 2b and Table 2).

Height prediction model with dominant height

Table 4 displays the fit statistics (RMSE, AIC and Fit Index) of height–d.b.h. relationship models: non-linear model [equations (1–4)], mixed-effects model [equation (5)], ENMs [equations (6–8)] and EMEMs [equations (9–11)]. The parameter estimates of all models were significantly different from zero, but estimates of the selected models are shown in Table 5. The fit statistics of equations (3 and 4) were similar and better than the alternative models [equations (1 and 2)]. However, the AIC value of equation (4) was slightly smaller. The smaller AIC value might be associated with the likelihood estimation function that involves the number of parameters in a model, i.e. equation (3) has four parameters; and equation (4) has five parameters. The inclusion of QMD as stand-level covariate in the equation (3) showed the reduction in

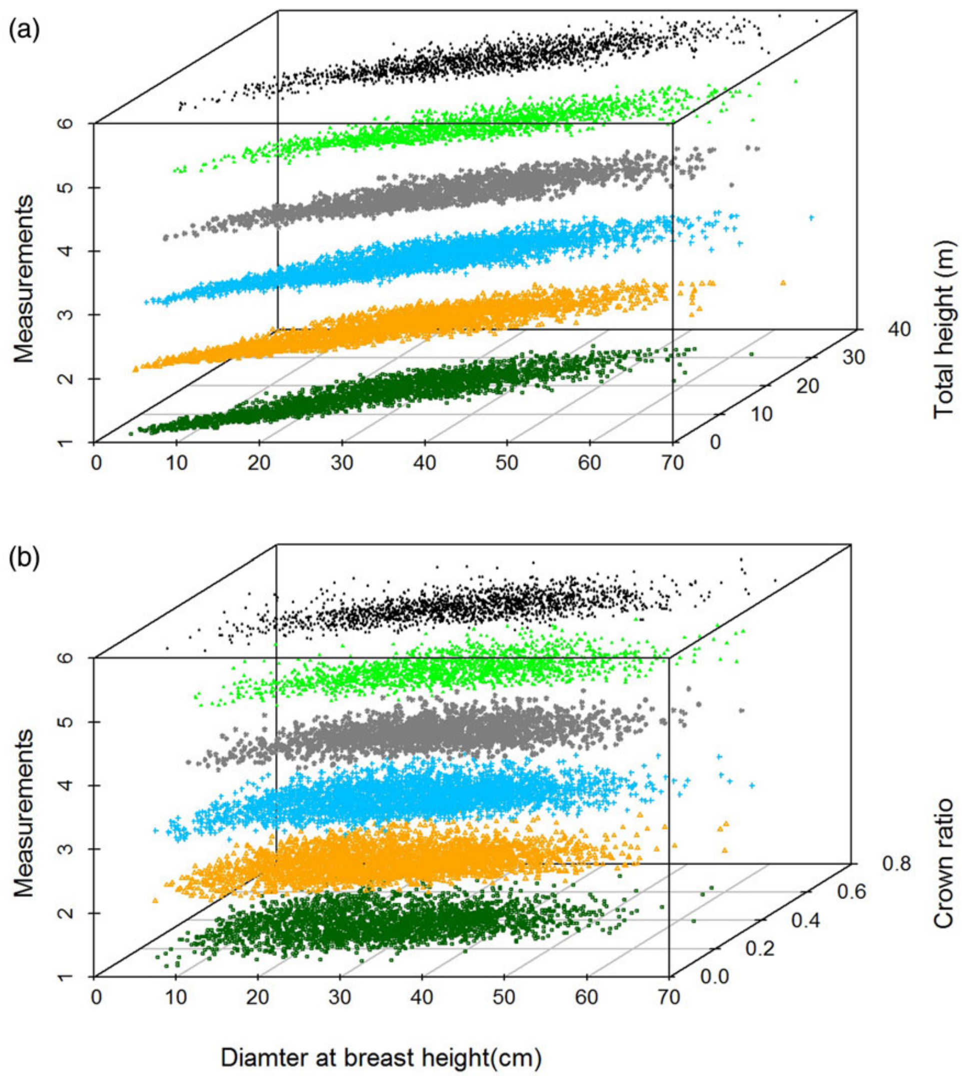

Figure 2 Three-dimensional scatter plot of each measurement: the distribution of height (a) and the crown ratio (b) of shortleaf pine along the diameter range. *X*-axis: d.b.h., *y*-axis: total height, crown ratio and *z*-axis: measurement time.

RMSE by 8 per cent compared with a model without QMD [equation (1)] and also compared with a model with BAH as the covariate [equation (2)] (Table 5). This suggested that the selection and the position of the stand-level covariate also affect the performance of a model (Table 4). Further, the mixed-effects model [equation (5)] showed similar Fit index (95.84) and RMSE (1.27) to equation (3) when the random component was assumed to be zero and only the fixed effects were used to make height predictions (Table 4). The SD $\hat{\sigma}(\mu_j) = 1.19319$ of the random component associated with the parameter b_2 was significant (*P*-value <0.0001) with the confidence interval of [1.208723, 1.80569]. The AIC value of equation (5) was lower than of equation (3), but this includes random effects parameters in equation (5) that are usually not available for prediction unless calibration can obtain them.

The LR statistics suggested the ENMs and EMEMs were significantly different from their base model form [equations (3) and (5), respectively] (Table 4). The ENMs and EMEMs provided similar RMSE and Fit indices, but equation (8) provided smaller AIC value and large significant LR statistics (Table 4). This suggested that modelling both variance function and autocorrelation structures in a non-linear model performed better than just modelling variance function [equation (6)] and other EMEMs (Table 4). In EMEMs, the similar AIC value and LR statistics indicated that EMEM equation (9) could be a better alternative model to equation (11) for height prediction, but ignoring autocorrelation could lead to an underestimate of standard errors if account is made only for heteroscedasticity.

Both extended model forms (ENMs and EMEMs) showed that relatively small power parameter estimates were needed to stabilize the issue of heteroscedasticity [equations (6, 8, 9 and 11)] and large autocorrelations within individual tree heights were observed [equations (7 and 8)], but moderate autocorrelation was observed between the tree heights within plot [equations (10 and 11)] (Table 4). The parameter estimates of equations (5, 8 and 11) were very similar (Table 5). Interestingly, the EMEM [equation (11)] showed the greater reduction in the SD of the random

effect associated with plot than the mixed-effects model [equation (5)] (Table 5).

The Goldfeld–Quandt test did not indicate violations of the assumption of homoscedasticity of error variance for any of equations (1–4). For example, Goldfeld–Quandt variance ratio was 0.80 for equation (3) which was less than tabulated $F_{(5257,\ 5257)} = 2.04$ at $\alpha = 0.05$ level. The standardized residuals did not show any systematic pattern to indicate a violation of an assumption of homogeneity of variance. So, the standardized residuals of the better performing models: non-linear model [equation (3)], mixed-effects model [equation (5)] and the ENM [equation (8)] are shown in Figure 3. The residual distribution pattern was slightly different between non-linear models and mixed-effects models. The non-linear model showed some dip near the lower fitted values (Figure 3a), while mixed-effects model showed a more compact distribution of standardized residuals (Figure 3b). The residual distribution of all ENMs was similar to that shown in Figure 3c for equation (8). Although extended models had better AIC values, residual distribution patterns were not different from their base models [equations (3 and 5)]. The standardized residuals plotted against d.b.h. (Figure 3d–f) also showed similar patterns and trends as discussed for fitted values. Mixed-effects models showed variation at the lower diameter range compared with the non-linear models and ENMs. The standardized residuals plotted against the range of design variables of shortleaf pine growth study also showed that the median residuals were almost centred to zero with a minimum bias, but these graphs are not shown.

Table 4 Fit statistics [RMSE, Fit index (per cent) and AIC], power parameter (δ), and autocorrelation (φ) and likelihood ratio statistics (LR) of the different forms of non-linear height–d.b.h. relationship

Models	RMSE	Fit index (%)	AIC	δ	φ	LR
Equation (1)	1.38	95.11	48 776			
Equation (2)	1.37	95.16	48 624			
Equation (3)	1.27	95.85	46 453			
Equation (4)	1.26	95.92	46 230			
Equation (5)	1.27	95.84	45 157			
Equation (6)	1.27	95.85	46 203	0.323		251.9
Equation (7)	1.27	95.85	40 305		0.803	9912.6
Equation (8)	1.27	95.84	36 067	0.398	0.810	10389.9
Equation (9)	1.27	95.83	44 790	0.383		369.2
Equation (10)	1.27	95.83	45 140		0.037	19.1
Equation (11)	1.27	95.83	44 771	0.3836	0.039	390.5

OLS non-linear [equations (1–4)], mixed-effects model [equation (5)], extended non-linear models (ENM) [equations (6–8)] and EMEMs [equations (9–11)].

Height prediction models without dominant height

The parameter estimates of the fundamental local height prediction models with d.b.h. as the only independent variable were significant (P-value <0.0001) (Table 6). The fit indices and RMSE for fundamental local models [equations (12–19)] were very similar, but equations (14, 15 and 17) did not perform as well as alternative models (Table 6). An average Fit index and RMSE was 73.85 and 3.18 m, respectively, for the fundamental local models. Fit statistics improved significantly across modified local models when a height prediction model function was multiplied by QMD raised to a power [equations (20–27)] (Table 6). The stand-level covariate QMD with a parameter in the power position increased values of fit indices and decreased RMSE substantially for both two parameters and three parameters d.b.h.–height prediction model functions. On average, model RMSE was decreased by 17.31 per cent (i.e. 0.55 m) and the Fit index was increased by 11 per cent (i.e. 8.24). The greatest reduction in RMSE (0.64 m) was observed in equation (22) that corresponds to base equation (14), and the lowest reduction in RMSE (0.48 m) was observed in equation (23) that corresponds to base equation (15). The AIC value of each modified local models was reduced by an average of 7.34 per cent (i.e. 5304) compared with the same equation form without QMD. Substantial reductions in AIC value and RMSE for these equation indicate that QMD as independent variable plays an important role in minimizing model error compared with models that use the only d.b.h. as the independent variable. For the fundamental local models [equations (12–19)] and modified local models with QMD [equations (20–27)], the Goldfeld–Quandt test failed to reject homogeneity of variance at $\alpha = 0.05$ of significance because all estimated variance ratios were less than $F = 2.04$.

Crown ratio estimation

The crown ratio estimation model with average dominant height [equation (29)] instead plot age [equation (22)] had better fit statistics and a better AIC value (Table 7) than other alternatives. Similar improvements in the model performance resulted from adding stand basal area when equation (28) was modified to equation (30), and equation (29) modified to equation (31) (Table 7). Crown ratio estimation models with PAG and BAH [equation (30)] did not

Table 5 Parameter estimates and standard error in the parentheses of the selected height prediction models

Parameters	Equation (1)	Equation (3)	Equation (5)	Equation (8)	Equation (11)
b_0	2.00052 (0.02780)	1.37232 (0.01111)	1.41103 (0.01583)	1.42302 (0.01659)	1.39325 (0.01473)
b_1	0.82570 (0.00346)	0.94119 (0.00271)	0.93335 (0.00369)	0.93107 (0.00398)	0.93888 (0.00355)
b_2	−9.34600 (0.41130)	−1.58186 (0.06269)	−1.36397 (0.05693)	−1.38491 (0.07173)	−1.25512 (0.04758)
b_3	−1.16510 (0.02090)	−1.69434 (0.01684)	−1.64451 (0.01739)	−1.64367 (0.02327)	−1.61206 (0.01638)
b_4			1.41103 (0.01583)	1.42302 (0.01659)	1.39325 (0.01473)
$\hat{\sigma}(\mu_j)$			1.19456		0.38801
$\hat{\sigma}(\varepsilon_{ij})$			0.03067	0.40902	0.03061

OLS non-linear [equations (1–3)], mixed-effects model [equation (5)], ENM [equation (8)] and EMEM [equation (11)].

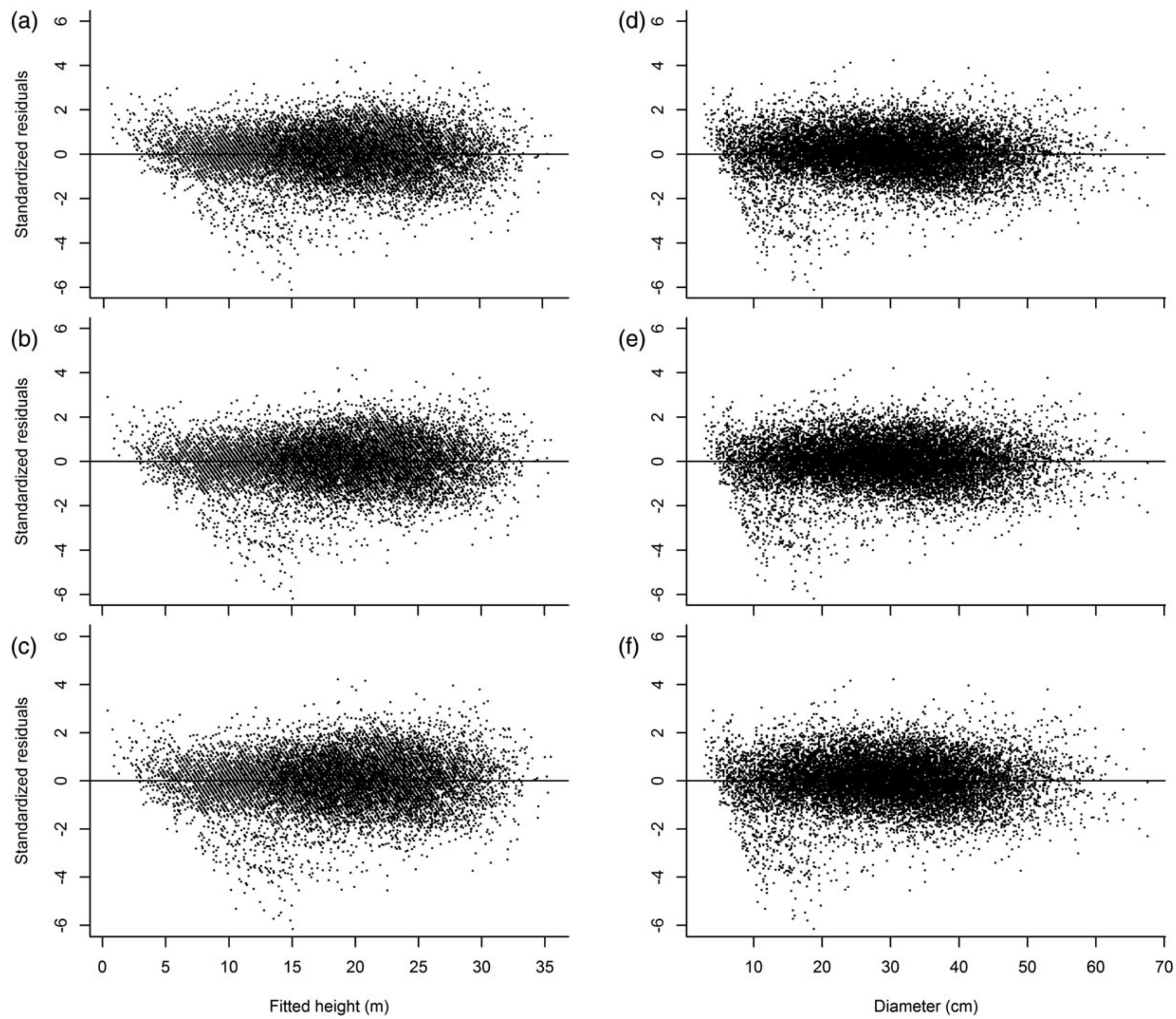

Figure 3 Scatter plot of standardized residuals *vs* fitted values (left panel) and *vs* d.b.h. (right panel) for the total height prediction models: (a) OLS non-linear [equation (3)], (b) mixed-effects model [equation (5)] and (c) ENM with AR (1) and power variance function [equation (8)].

improve model performance as much as the inclusion of H_D [equation (29)]. The crown ratio estimation model that included RSI [equation (32)] had better fit statistics and AIC value than all other alternative models. Fitting the logistic function to this dataset suggested that equation (33) can be used as an alternative crown ratio estimation model because it had similar fit statistics and a similar AIC value compared with other slightly better alternative models. The parameter estimates of the crown ratio estimation models were significant (P-value <0.0001), but only estimates of the selected models are shown (Table 8).

The mixed-effects model for crown ratio estimation [equation (35)] with RSI had a slightly better Fit index and RMSE than the model with BAH [equation (34)], although it had a similar AIC value (Table 7). However, the fit statistics (RMSE and Fit indices) of mixed-effects models setting random effects to zero were not smaller than those of the OLS non-linear models [equations (31 and 32)] (Table 7). The SD of the error ($\hat{\sigma}(\varepsilon_{ij})$) of equation (35) was slightly smaller (0.01658) with a 95 per cent confidence interval of [0.01478, 0.01859] than of equation (34), but estimate of the SD of the random component ($\hat{\sigma}(\mu_j)$) of equation (35) was identical with equation (34).

The ENMs [equations (36–38)] and EMEMs [equation (39–41)] showed reduced AIC values compared with the base models: equations (32 and 35), respectively. The large differences in the LR statistics and smaller AIC values indicated that ENMs performed better than EMEMs (Table 7). Both ENMs and EMEMs provided negative power parameter estimate (Table 7). However, the ENM with a variance function [equation (36)] showed Fit index better than others, and close to the non-linear base model [equation (32)] but not with a greatly reduced AIC value (Table 7).

Table 6 Parameter estimates and fit statistics (RMSE, Fit index and AIC) for height prediction models with d.b.h. [equations (12–19)], and along with QMD [Equations (20–27)]

Equation (no.)	Parameters estimates				RMSE	Fit index (%)	AIC
	b_0	b_1	b_2	b_3			
12	41.2084	0.0215			3.16	74.27	72 061
13	38.8776	0.0247	1.0551		3.16	74.28	72 058
14	37.3973	−18.5592			3.25	72.64	72 921
15	1.6609	0.7209			3.19	73.76	72 336
16	67.0409	72.9095			3.16	74.24	72 076
17	3.6529	−20.2320			3.23	73.05	72 708
18	51.5498	−39.2751	10.7639		3.16	74.28	72 055
19	1.4364	0.0324	1.1802		3.16	74.28	72 055
20	3.5866	0.0716	0.5393		2.62	82.24	66 863
21	3.5567	0.0741	1.0387	0.5405	2.62	82.24	66 864
22	4.1877	−8.2663	0.5365		2.62	82.26	66 847
23	1.0548	0.3493	0.5055		2.71	81.09	67 738
24	4.48945	13.7148	0.5188		2.63	82.08	66 983
25	1.4838	−9.2260	0.5279		2.62	82.26	66 845
26	4.2827	−8.6416	0.4166	0.5327	2.62	82.26	66 847
27	0.7076	0.8722	10.6304	0.5329	2.62	82.26	66 847

Table 7 Fit statistics [RMSE, Fit index (per cent), and AIC], variance function (δ), and autocorrelation (φ) and likelihood ratio statistics (LR) of the different form of crown ratio (CR) estimation model

Models	RMSE	Fit index (%)	AIC	δ	φ	LR
Equation (28)	0.07225	41.3	−33 919			
Equation (29)	0.07169	42.17	−34 128			
Equation (30)	0.07141	42.56	−34 223			
Equation (31)	0.07071	43.67	−34 496			
Equation (32)	0.07052	44.07	−34 588			
Equation (33)	0.07106	43.18	−34 374			
Equation (34)	0.07117	42.99	−35 467			
Equation (35)	0.07088	43.44	−35 475			
Equation (36)	0.7053	44.02	−34 978	−0.246		390.74
Equation (37)	0.07073	43.69	−37 878		0.548	3291.8
Equation (38)	0.07095	43.35	−38 397	−0.364	0.55	3636
Equation (39)	0.07089	43.44	−35 726	−0.282		498.2
Equation (40)	0.07081	43.57	−35 703		0.129	230
Equation (41)	0.07079	43.6	−35 467	−0.277	0.127	720.7

OLS non-linear [equations (28–31)], logistic model [equation (32)], mixed-effects models [equations (34 and 35)], extended non-linear models (ENM) [equations (36–38)] and EMEMs [equations (39–41)].

The ENMs with AR (1): equations (37 and 38) showed that observed CR of an individual tree in repeated measurements were moderately correlated (0.55), but EMEMs equations (40 and 41) showed CR were very weakly correlated (0.12) within a plot (Table 7). Interestingly, ENMs equation (38) provided the smallest AIC value and comparable Fit index with equation (35) (Table 7) but had greatly changed parameter estimates (Table 8). It indicated that a non-linear model with both power variance function and AR (1) structure could be used in the repeated measurements, as a substitute modelling approach to the mixed-effects modelling approach.

The Goldfeld–Quandt test also failed to reject homogeneity of variance for the crown ratio estimation models [equations (28–32)] at the $\alpha = 0.05$ level of significance because the estimated variance ratios were less than $F = 2.08$. For example, the ratios of variance were 1.09 and 1.08 for equations (21 and 32), respectively. The standardized residual plot also indicated homogeneity of variance. The residual distribution patterns were fairly similar within the ENMs and the EMEMs with small improvements over the base model. The residual distribution pattern, Figure 4a of non-linear model [equation (32)] and Figure 4b of the mixed-effects model [equation (35)], was similar, although the latter was slightly compact and elongated to right, indicating more constant error variance. The residual pattern of the ENM [equation (38)] showed some errors at the lower fitted values but more compact error at fitted middle values than other models (Figure 4c). Similarly, the standardized residuals plotted against d.b.h. showed similar trends among the model forms (Figure 4d–f) and suggested large prediction errors at the lower diameter range. The standardized residuals from equation (32) plotted against the design variables also showed that the median residuals are almost centred at zero with a minimum bias over the range of design variables, but these graphs are not shown.

Discussion

The inclusion of QMD improved the fit statistics of the modified equation for height prediction [equation (3)] compared with the models presented by Budhathoki *et al.* (2008) [equation (2)] and Lynch *et al.* (1999) [equation (1)] (Table 4). It was found that the height prediction bias of 0.02 m by equation (1) was reduced to

Table 8 Parameter estimates standard error in the parentheses of the selected CR estimation models

Parameters	Equation (28)	Equation (31)	Equation (32)	Equation (35)	Equation (36)	Equation (38)	Equation (41)
b_0	0.26736 (0.00217)	0.17518 (0.00477)	0.16527 (0.00293)	0.149905 (0.000352)	0.16644 (0.00499)	0.26438 (0.00375)	0.15291 (0.00403)
b_1	3.24279 (0.08715)	1.23511 (0.04487)	1.19436 (0.03247)	0.971799 (0.0377)	1.258 (0.05096)	1.56297 (0.054)	1.02018 (0.04693)
b_2	0.98527 (0.0105)	−0.2342 (0.02597)	0.40621 (0.03076)	0.152942 (0.03284)	0.44583 (0.02861)	2.38424 (0.14721)	0.19512 (0.03973)
b_3		1.0726 (0.02635)	1.07876 (0.02026)	1.18193 (0.02442)	1.06131 (0.02732)	0.74909 (0.01253)	1.14689 (0.02515)
b_4				0.06721	0.16644 (0.00499)	0.26438 (0.00375)	0.16404
$\hat{\sigma}(\mu_j)$				0.01658			0.01654
$\hat{\sigma}(\varepsilon_{ij})$				0.06721	0.15588	0.19222	0.01654

OLS non-linear [equations (28, 31 and 32)], mixed-effects model [equation (35)], ENM [equations (36 and 38)] and EMEM [equation (41)].

0.005 m by equation (3) which supported the conclusion that equation (3) has better prediction capability. It was also observed that the stand competition variable 'QMD' in equation (3) can be substituted by relative measure variable 'RAQD' to obtain similar parameter estimates, standard error and RMSE. The parameter estimates and fit statistics of the mixed-effects model [equation (5)] were similar to an OLS non-linear model [equation (3)]. Prediction from a non-linear mixed model with the random effect parameter set to zero were not as good as the OLS results, as expected (Tables 4 and 7). Garber *et al*. (2009) demonstrated that imputing height from the fixed estimates of OLS non-linear had less bias in volume prediction (4 m^3 ha^{-1}) than a mixed-effects model (20 m^3 ha^{-1}) when height was imputed for Douglas fir, as it was discussed by Temesgen *et al*. (2008) while predicting height for Douglas-fir. The smaller AIC value in the mixed-effects model [e.g. equation (5)] than OLS non-linear models [e.g. equations (1–3)] might be attributed to the inclusion of the random parameters in the mixed-effects model. However, unless calibration data are available, the random parameters may not be of practical help for most prediction problems. On the other hand, several authors have shown that prediction using mixed models can be attractive when calibration data are available and have indicated, mixed model prediction may not be better than OLS unless calibration data are available (Lynch *et al*., 2005, 2012; VanderSchaaf, 2014).

The AR (1) structure suggested autocorrelation of residuals within individual observation was large for height prediction and was moderate for CR estimation. The large autocorrelation can be perceived as a wider residual distribution with increasing diameter as in Figure 3, and moderate autocorrelation can be perceived as a circular distribution of residuals as in Figure 4. Ducey (2009) also observed the circular pattern of residual distribution for *CR* modelling of *Pinus strobus*. However, the autocorrelation remediation at the larger lags is not reliable because it is based on fewer residual pairs. The EMEMs did not greatly improve model fits compared with the ENMs for both height–d.b.h. relationship and CR estimation. So, the AIC values of the EMEMs were not smaller than ENMs (Tables 4 and 7). The smaller AIC value in the ENMs could be a result of more successful modelling of autocorrelation structure within individual tree than within plot or grouped data as in the mixed model. Because of this, AIC value of the EMEMs of both height–d.b.h. relationship and CR estimation was not greatly reduced compared with the ENMs (Tables 4 and 7). Although LR statistics suggested EMEM with AR (1) structure was significantly different from the base mixed-effects model (Tables 4 and 7), both models had similar SD of random effect and error variance. Perhaps, the mixed-effects models each group (plots) with its random effect and the assigned AR (1) (within stand) is formally identical to a random effect model which has both among-group variance (σ^2_μ) and within-group variance (σ^2_w) that corresponds to correlation parameters(ρ, σ^2). Therefore, it is possible that ENM with autocorrelation structure and variance function can perform better than the EMEMs.

The small, positive power parameter in both ENMs and EMEMs of height–d.b.h. relationship indicated small amount of heterogeneous error due to influence of H_D. The negative power parameter associated with d.b.h. could be due to the pattern of the standardized residuals distributed along the lower diameter that showed greater variability. But the fact that large standardized residuals occurred in this study (Figures 3 and 4) may be because the data

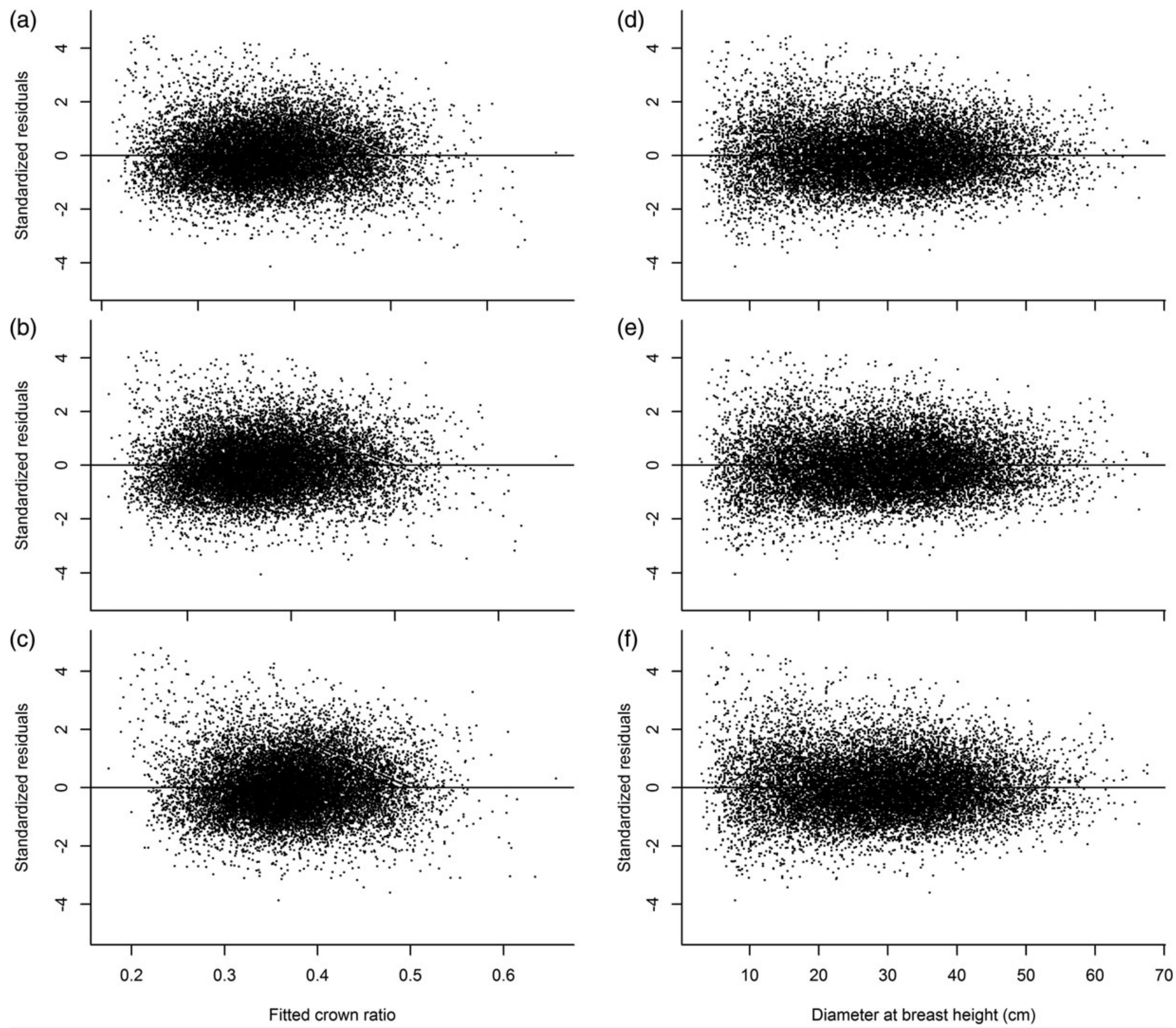

Figure 4 Scatter plot of standardized residuals *vs* fitted values (left panel) and *vs* d.b.h. (right panel) for the crown ratio estimation models: (a) OLS non-linear [equation (32)], (b) mixed-effects model [equation (35)] and (c) extended non-linear model with AR (1) and power variance function [equation (38)].

were from naturally occurring stands as observed in Budhathoki *et al*. (2008) for shortleaf pine and in Sharma and Parton (2007) for boreal tree species in Ontario, Canada. Height growth in naturally occurring stands is not as uniform as that in plantations, such as described, for example, by Buford (1991) for loblolly pine. It could also partially due to measurement errors since the accuracy of commonly used height measuring devices is likely to be in 0.5 m if not more, and there can also be variation in the accuracy obtained by individuals in the field measurement crew.

The issues of heteroscedastic residual distribution (non-constant variance) when present can sometimes be substantially resolved, if a weighted OLS non-linear model is used. After testing different weight functions, the weight 1/d.b.h. was found beneficial in improving model performance and addressing non-constant error variance as used by Huang *et al*. (1992), Fang and Bailey (1998) and Temesgen *et al*. (2007, 2014). For example, if weight as 1/d.b.h. was used in equation (3), the residuals were slightly more compact and constant than shown in Figure 3, but some were still beyond ± 4 at lower diameter range. The standard error of parameter estimates was not considerably different from those of non-weighted equation (3). Because the weighted parameter estimates did not have variances that were substantially different from the unweighted parameter estimates and due to the results of the Goldfeld–Quant test, we used the unweighted parameter estimates. Of course, mathematically the sum of squares between predicted and actual heights cannot be reduced below

that obtained with ordinary unweighted OLS by weighting a model that has the same form and the same independent variables.

Modified local models [equations (20–27)] showed substantially improved and similar model performance except for equation (23) (Table 7). The parameter of QMD used in the 'power' position was more effective than in the 'linear' position. The other variables of stand competition such as BAH and TPH did not show any improvement, but the inclusion of relative measures 'RAQD' provided RMSE similar to the RMSE with the inclusion of 'QMD' in some modified local models [equations (17, 20 and 21)]. However, RAQD provided larger RMSE for other modified local models than it was observed with QMD. Therefore, the evidence suggests that QMD can be a better surrogate covariate in improving model performance when the information on H_D is not available. Temesgen *et al*. (2007) also suggested that the use of BAH and BAL in the fundamental local model and that can reduce RMSE up to 15 per cent. The height–d.b.h. prediction model has been often modified using H_D to reduce RMSE of a model; however, in practice H_D may not be easily available for natural stands.

Equations (28–33) fitted with OLS provided a choice of alternative models for CR estimation. The model [equation (32)] with covariate H_D and RSI performed better than a model with PAG and BAH [equation (31)], and H_D and PAG [equation (30)] in addressing CR variability. It was expected because the distance between the trees reflects the crown competition and also thinning changes abruptly CR of an individual tree (Smith *et al*., 1992; Hynynen, 1995). So, RSI has been used a measure of relative competition index in modelling CR (Ducey, 2009; Zhao *et al*., 2012). It was observed that the prediction bias was very small for all CR models, but it was negative for equations (28–31), while it was positive for equation (32). It suggested that model with RSI tended to overestimate than other alternative models. Researchers have also used crown competition factor (CCF) as a covariate in association with BAL to model CR (Hasenauer and Monserud, 1996; Temesgen *et al*., 2007).

The small amount of variation explained (44 per cent) by equation (32) in CR estimation may be due to the inherent variability of naturally occurring forests compared with plantations (Figure 4a–c). The proportion of variation explained by the CR models of Dyer and Burkhart (1987) for loblolly pine plantations was greater (60 per cent) than the proportion of variation in CR explained by the natural stand models of Hasenauer and Monserud (1996) in Austrian natural forest stands (49 per cent–17 per cent). The significance of equations (29 and 31) is that the model can be applied when PAG is not available. Hynynen (1995) also found H_D as an important independent covariate in CR modelling of Scot pine stands. The addition of BAH as a linear term to the ratio of D and H_D [equation (25)] improved fit statistics more than combining it linearly to individual tree height. Smith *et al*. (1992) and Monserud and Sterba (1996) indicated that using BAH as an independent variable can improve prediction of an individual tree crown length. This leads to the expectation that BAH could also improve the fit of CR models.

The logistic function model for CR estimation [equation (29)] performed similarly to the model of equation (31). The logistic function restricts predicted crown ratio bounds within a 0–1 interval. As in the d.b.h.–height models, the parameter estimates of OLS non-linear models for CR estimation were also better than mixed-effects models (Table 7). The mixed-effects models [equations (27 and 28)] of CR estimation also performed more poorly than OLS non-linear CR estimation [equations (31 and 32)], but the AIC values were smaller for mixed-effects models (Table 5).

Conclusions

The modified height–d.b.h. relationship [equation (3)] and crown ratio relationship model [equation (32)] provided better accuracy than existing models for estimating the height and crown ratio of natural even-aged stands of shortleaf pine. The inclusion of QMD as a measure of stand competition rather than BAH as an independent variable helped to improve the understanding of the height–d.b.h. relationship. Also, an inclusion of QMD enhanced the precision of height prediction model forms that do not utilize dominant height as a covariate. The RSI and H_D improved the relationship of crown ratio with height and d.b.h. compared with using PAG in CR estimations model. It was found that inclusion of RSI instead of made small but definite improvements in CR estimation and also that the logistic function can be used as a comparable choice to an alternative non-linear model for CR prediction.

By alleviating heterogeneous error at stand level and adjusting autocorrelation at the individual tree level in repeated measurements, a quality non-linear model with minimum information loss can be obtained. The parameter estimates of such a model are preferred as an alternative to the mixed-effects modelling approach in predicting missing height and crown ratios. In repeated measurements, the autocorrelation within individual observation is larger while predicting the height of that individual tree than estimating its crown ratio. The small, but positive weight of power parameter remediated heterogeneous errors and improved model performance for height prediction. Mixed models provided similar fit statistics when predictions were based only on the fixed effects parameters compared with those of non-linear models fitted by OLS. However, the mixed-effects models may provide improved predictions when calibration data are available.

Parameter estimates of the ENM, equation (8) for height prediction and equation (38) for CR estimation, can be incorporated in the Shortleaf Pine Stand Simulator (Huebschmann *et al*., 1998) which can be used to develop information for practical forest management decision-making, i.e. estimation of the total stand volume and biomass production, for naturally occurring even-aged shortleaf pine forests. The relationships between d.b.h., height and crown ratio could have significant implications in inventories for biomass and carbon estimation of natural stands of shortleaf pine in the southern US. To the extent that these formulations are a novel approach in the forestry literature, they could be considered for application in other forest types in addition to well-known existing equation forms.

Acknowledgements

This research was supported by Department of Natural Resource Ecology and Management) at Oklahoma State University (OSU) and USDA Forest Service Southern Research Station (USFS). This article has been approved for publication by the Oklahoma Agricultural Experiment Station and supported by project OKL0-2843. The authors would like to thank all crew members who contributed in collecting data since permanent plot establishment (1985–1987) to the current measurement (2012–2014).

Conflict of interest statement

None declared.

Funding

Grant number 11-C11330124-042, 25 Year Remeasurement of Natural Even-aged Shortleaf Pine Growth and Yield Study, Source: USDA Forest Service.

References

Arcangeli, C., Klopf, M., Hale, S.E., Jenkins, T.A.R. and Hasenauer, H. 2014 The uniform height curve method for height–diameter modeling: an application to Sitka spruce in Britain. *Forestry* **87**, 177–186.

Avery, T.E. and Burkhart, H.E. 2002 *Forest Measurements*, 5th edn. McGraw-Hill. 456 pp.

Bates, D.M. and Watts, D.G. 1980 Relative curvature measures of nonlinearity. *J. R. Stat. Soc. Ser. A* **42**, 1–25.

Budhathoki, C.B., Lynch, T.B. and Guldin, J.M. 2008 A mixed-effects model for the dbh–height relationship of shortleaf pine (*Pinus echinata* Mill.). *South. J. Appl. For*. **32**, 5–11.

Buford, M.A. 1991 Performance of four yield models for predicting stand dynamics of a 30-year-old loblolly pine (*Pinus taeda* L.) spacing study. *For. Ecol. Manage*. **46**, 23–38.

Burkhart, H.E. and Strub, M.R. 1974 A model for simulation of planted loblolly pine stands. In: *Growth Models for Tree and Stand Simulation*. Royal College of Forestry Stockholm, p. 379.

Burkhart, H.E. and Tomé, M. 2012 *Modeling Forest Trees and Stands*. Springer Science & Business Media.

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

Cryer, J.D. and Chan, K.-S. 2008 *Time Series Analysis: With Applications in R*. Springer.

Daniels, R.F. and Burkhart, H.E. 1975 Simulation of individual tree growth and stand development in managed loblolly pine plantations. Virginia Poly. Inst. and State Univ. Pub. FWS-5-75, p. 69.

Ducey, M.J. 2009 Predicting Crown Size and Shape from Simple Stand Variables. *J. Sustain. For*. **28**, 5–21.

Dyer, M.E. and Burkhart, H.E. 1987 Compatible crown ratio and crown height models. *Can. J. For. Res*. **17**, 572–574.

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

Graney, D.L. and Burkhart, H.E. 1973. *Polymorphic site index curves for shortleaf pine the Ouachita mountains*. USDA Forest Service Res. Pap. SO-85. p. 12.

Garber, S.M., Temesgen, H., Monleon, V.J. and Hann, D.W. 2009 Effects of height imputation strategies on stand volume estimation. *Can. J. For. Res*. **39**, 681–690.

Hasenauer, H. and Monserud, R.A. 1996 A crown ratio model for Austrian forests. *For. Ecol. Manage*. **84**, 49–60.

Holdaway, M.R. 1986 Modeling tree crown ratio. *For. Chron*. **62**, 451–455.

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

Huebschmann, M.M., Lynch, T.B. and Murphy, P.A. 1998 *Shortleaf Pine Stand Simulator: An Even-aged Natural Shortleaf Pine Growth-and-Yield Model: User's Manual*. Oklahoma Agricultural Experiment Station, Division of Agricultural Sciences and Natural Resources, Oklahoma State University.

Hynynen, J. 1995 Predicting tree crown ratio for unthinned and thinned Scots pine stands. *Can. J. For. Res*. **25**, 57–62.

Jiang, L.-C. and Liu, R.-L. 2011 Segmented taper equations with crown ratio and stand density for Dahurian Larch (*Larix gmelinii*) in Northeastern China. *J. For. Res*. **22**, 347–352.

Jiang, L., Brooks, J.R. and Hobbs, G.R. 2007 Using crown ratio in yellow-poplar compatible taper and volume equations. *North J. Appl. For*. **24**, 271–275.

Judge, G.G., Hill, R.C., Griffiths, W., Lutkepohl, H. and Lee, T.-C. 1988 *Introduction to the Theory and Practice of Econometrics*. Wiley, NY.

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

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

Lawson, E.R. 1990 *Pinus echninata* Mill. Shortleaf Pine. In *Silvics of North America*. R.M. Burns, B.H. Honkala and technical coordinators (eds), Vol. 1. Conifers. Service Handbook 654. U.S. Department of Agriculture, Washington, DC, 316–326.

Leites, L.P., Robinson, A.P. and Crookston, N.L. 2009 Accuracy and equivalence testing of crown ratio models and assessment of their impact on diameter growth and basal area increment predictions of two variants of the Forest Vegetation Simulator. *Can. J. For. Res*. **39**, 655–665.

Lynch, T.B. and Murphy, P.A. 1995 A compatible height prediction and projection system for individual trees in natural, even-aged shortleaf pine stands. *For. Sci*. **41**, 194–209.

Lynch, T.B., Hitch, K.L., Huebschmann, M.M. and Murphy, P.A. 1999 An individual-tree growth and yield prediction system for even-aged natural shortleaf pine forests. *South J. Appl. For*. **23**, 203–211.

Lynch, T.B., Holley, A.G. and Stevenson, D.J. 2005 A random-parameter height–Dbh model for cherrybark oak. *South J. Appl. For*. **29**, 22–26.

Lynch, T.B., Budhathoki, C. and Wittwer, R.F. 2012 Relationships between height, diameter, and crown for eastern cottonwood (*Populus Deltoides*) in a great plains riparian ecosystem. *West J. Appl. For*. **27**, 176–186.

Meyer, H.A. 1940 A mathematical expression for height curves. *J. For*. **38**, 415–420.

Monserud, R.A. and Sterba, H. 1996 A basal area increment model for individual trees growing in even- and uneven-aged forest stands in Austria. *For. Ecol. Manage*. **80**, 57–80.

Peng, C., Zhang, L. and Liu, J. 2001 Developing and validating nonlinear height–diameter models for major tree species of Ontario's boreal forests. *North J. Appl. For*. **18**, 87–94.

Pinheiro, J. and Bates, D. 2000 *Mixed-Effects Models in S and S-PLUS*. Springer Science & Business Media.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. and R Core Team 2014. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-117, http://CRAN.R-project.org/package=nlme.

Ratkowsky, D.A. and Giles, D.E. 1990 *Handbook of Nonlinear Regression Models*. Marcel Dekker.

R Development Core Team. 2012 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Richards, F.J. 1959 A Flexible growth function for empirical use. *J. Exp. Bot*. **10**, 290–301.

Saud, P., Lynch, T.B. and Guldin, J.M. 2016 Twenty five years long survival analysis of an individual shortleaf pine (*Pinus Echinata Mill*.) trees. In *Proceedings of the 18th biennial southern silvicultural research conference. e–Gen. Tech. Rep. SRS–212* Schweitzer, C.J., Clatterbuck, W.K. and Oswalt, C.M. (eds). Department of Agriculture, Forest Service, Southern Research Station, Asheville, NC, U.S., pp. 555–557.

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

Sharma, R.P. and Breidenbach, J. 2015 Modeling height–diameter relationships for Norway spruce, Scots pine, and downy birch using Norwegian national forest inventory data. *For. Sci. Technol*. **11**, 44–53.

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

Sharma, M. and Yin Zhang, S. 2004 Height–diameter models using stand characteristics for *Pinus banksiana* and *Picea mariana*. *Scand J. For. Res*. **19**, 442–451.

Smith, W.R., Farrar, R.M. Jr., Murphy, P.A., Yeiser, J.L., Meldahl, R.S. and Kush, J.S. 1992 Crown and basal area relationships of open-grown southern pines for modeling competition and growth. *Can. J. For. Res*. **22**, 341–347.

Stage, A.R. 1975 *Prediction of Height Increment for Models of Forest Growth*.

Intermountain Forest and Range Experiment Station, Forest Service, U.S. Dept. of Agriculture, Ogden, Utah. p. 32.

Stevenson, D.J., Lynch, T.B., Saud, P., Heineman, R., Holeman, R., Wilson, D. *et al*. 2016. Winter in the Ouchitas—A Severe Winter Storm Signature in *Pinus echinata* in the Ouachita Mountains of Oklahoma and Arkansas, USA. *Dendrochronologia* **37**, 107–115.

Tahvanainen, T. and Forss, E. 2008 Individual tree models for the crown biomass distribution of Scots pine, Norway spruce and birch in Finland. *For. Ecol. Manage*. **255**, 455–467.

Temesgen, H., LeMay, V. and Mitchell, S.J. 2005 Tree crown ratio models for multi-species and multi-layered stands of southeastern British Columbia. *For. Chron*. **81**, 133–141.

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

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

Temesgen, H., Zhang, C.H. and Zhao, X.H. 2014 Modelling tree height–diameter relationships in multi-species and multi-layered forests: a large observational study from Northeast China. *For. Ecol. Manage*. **316**, 78–89.

Valenti, M.A. and Cao, Q.V. 1986 Use of crown ratio to improve loblolly pine taper equations. *Can. J. For. Res*. **16**, 1141–1145.

VanderSchaaf, C.L. 2014 Mixed-effects height–diameter models for ten conifers in the inland Northwest, USA. *South For. J. For. Sci*. **76**, 1–9.

Wykoff, W.R. 1990 A basal area increment model for individual conifers in the northern Rocky Mountains. *For. Sci*. **36**, 1077–1104.

Wykoff, W.R., Crookston, N.L., Stage, A.R. and Forest, I. 1982 *User's Guide to the Stand Prognosis Model*.

Zhao, D., Kane, M. and Borders, B.E. 2012 Crown ratio and relative spacing relationships for loblolly pine plantations. *Open J. For*. **2**, 101.