Research Article
doi: 10.1590/01047760202228013112

CERNE

Modelling dominant height growth including a rainfall effect using the algebraic difference approach

Morries Chauke [1✤], Henry Mwambi [1], Heyns Kotze [2]

[1] School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Scottsville, Pietermaritzburg, South Africa
[2] Mondi South Africa, South Africa

FOREST MANAGEMENT

ABSTRACT

Background: Estimating forest productivity is critical for effective management and site assessment. The dominant height is used to calculate the Site Index (SI), which is commonly used to assess forest productivity. In this study, an algebraic difference approach was used to develop a dominant height model incorporating the rainfall effect for *Eucalyptus grandis* x *Eucalyptus urophylla* (*E. grandis* x *E. urophylla*). The dataset consists of 75 permanent sample plots ranging in age from 0.5 to 11 years, as well as 7 rainfall stations spread across plantations in Coastal Zululand, South Africa. Using fixed and mixed-effects in the predictor function, twelve candidate models were derived from the Bertalanffy-Richards, Lundqvist-Korf, McDill-Amateis, and Hossfeld growth functions. A continuous-time autoregressive error structure was used to account for serial autocorrelation in the longitudinal unbalanced data. Model fit statistics and graphical methods were used to evaluate the candidate models.

Results: The addition of the rainfall effect increased model precision by 37%. The mixed-effects formulation produced 18% more precision when compared to similar models with all parameters fixed. Due to their compatibility with expected biological behaviour and good performance on validation data, mixed-effects models based on Lundqvist-Korf and McDill-Amateis functions were chosen as the final models.

Conclusion: Unlike similar models that do not take rainfall into account, these models can capture the effects of severe rainfall conditions such as drought and can thus be used in short-rotation pulp forests with fluctuating rainfall.

Keywords: Site index, algebraic difference approach, polymorphism, fixed-effects, mixed-effects, climatic effects.

HIGHLIGHTS

The nonlinear mixed-effects technique was used to forecast dominant height over time for *Eucalyptus grandis* x *Eucalyptus urophylla* in Coastal Zululand, South Africa, taking into account changes in rainfall.
We looked into and contrasted nonlinear fixed and mixed-effects modeling methodologies. In comparison to the strictly fixed effects model, the final nonlinear mixed-effects models were 18% more precise.
The addition of the rainfall effect increased model precision by 37%.
The models developed in this study can be utilized in short-rotation pulp forests with varying rainfall to capture the effects of extreme rainfall conditions such as drought.

CHAUKE, M.; MWAMBI, H.; KOTZE, H. Modelling dominant height growth including a rainfall effect using the algebraic difference approach. CERNE, v.28, e-103112, doi: 10.1590/01047760202228013112

✤Corresponding author e-mail: morries.chauke@mondigroup.com
Submitted: 06/06/2022 Accepted: 16/08/2022

CERNE ISSN 0104-7760
UNIVERSIDADE FEDERAL DE LAVRAS

INTRODUCTION

The *E. grandis* x *E. urophylla* hybrid species is widely planted across plantations in Coastal Zululand, South Africa, primarily for pulpwood production. This hybrid species is not only productive, but also combines the characteristics of the parent species, *E. grandis* and *E. urophylla*, for good survival and disease tolerance as reported in Stanger et al. (2011).

Accurate estimation of forest productivity is important for efficient management and site assessment, thus a variety of methods for estimating productivity have been developed (Burkhart and Tomé, 2012; Clutter et al., 1983; Goelz and Burk, 1992; Sharma et al., 2015). Ideally, forest productivity should be measured directly in terms of stand volume yield, as is done with other agricultural crops. However, stand volume can be influenced by effects of management such as planting density and stand history, as well as the length of stand rotation (Anta and Diéguez-Aranda, 2005; Davis et al., 2001). Based on the hypothesis that height correlates well with stand volume growth, stand height has been used as indicator of site productivity (Baur, 1877). In particular, site index defined as the average height of a specific number of thickest trees in a given unit area (also referred to as dominant height) at reference age is commonly used to evaluate forest productivity because is not influenced by density (Carmean and Lenthall, 1989; Clutter et al., 1983; Skovsgaard and Vanclay, 2008).

The dominant height modelling can take two forms: static and dynamic site index equations. Static site equations have the general form $y = f(t,S,\beta)$, where y is the height at age t, β is a vector of parameters, and S is a site index at fixed base age (Cieszewski, 2011). Even though static models can be useful for prediction they cannot directly solve for the dominant height given a reference measurement. While iterative methods can be used to solve this problem, convergence cannot be guaranteed (Parresol and Vissage, 1998).

Dynamic site index equations have the general form $y_2=f(t_2,t_1,y_1,\beta)$, where y_2 and y_1 are the function values at t_2 and t_1, respectively, and, β is as previously defined. Bailey and Clutter (1974) presented a technique for dynamic equation derivation that is known in forestry as the algebraic difference approach (ADA). This method was referred to by Clutter et al. (1983) as "difference equations" approach and then as "algebraic difference equations (ADE)" by Borders et al. (1984). The ADA technique consists of solving one of the base model parameters with its initial conditional solution. Depending on which parameter is substituted, models derived with ADA are either anamorphic (proportional curves) with multiple asymptotes or polymorphic with a single asymptote (Cieszewski, 2001; Cieszewski, 2002). It is believed that anamorphic curves do not adequately represent the dominant height-age relationship and thus the use of polymorphic models is preferred (Goelz and Burk, 1992; Monserud, 1984; Parresol and Vissage, 1998). Another desirable attribute in site index models is multiple asymptotes (Cieszewski, 2002; Cieszewski and Bailey, 2000). Cieszewski and Bailey (2000) introduced a generalization of the ADA, the generalized algebraic difference approach (GADA). The main advantage of GADA is that it allows more than one parameter to be site-specific (Cieszewski, 2001; Cieszewski and Bailey, 2000) and includes the ability to simulate polymorphism and multiple asymptotes (Cieszewski, 2002; Cieszewski and Bailey, 2000; Cieszewski et al., 2007).

These models only account for projection age (t_2), previous age (t_1), and the previous dominant height (y_1) measurement in their current form. While these models typically guarantee simplicity and robustness, they do not consider the influence of other variables that are important in determining dominant height growth. A change in these unaccounted-for variables could impair the explanatory power of these models. In this sense, in order to achieve adequate accuracy in site quality estimation, site productivity variables such as dominant height must consider as many factors affecting dominant height as possible. When assessing productivity, a variety of site-specific climatic, physiographic, and soil characteristics are frequently considered, and this consideration has been shown to improve growth and yield projections (Weiskittel et al., 2011).

Numerous efforts have been explored to incorporate these characteristics in growth equations: Hunter and Gibson (1984) presented an effort to use multiple regression to link site index to environmental variables like soil and climate, and their model revealed that rainfall has a significant impact on site index. Woollons et al. (1997) used ADA to investigate the effect of meteorological factors and soil type on dominant height and basal area. They found no significant improvement in dominant height growth. Snowdon et al. (1998) explored introducing climatic-based indices into projection models and found an improvement in fit for dominant height and other growth features. yılmaz et al. (2015) conducted analysis of variance to evaluate site productivity and relationships between site index and ecological parameters, and they found significant associations between site index and average sand, silt, clay, and field capacity in physiological and absolute soil depth. Despite the fact that this study found that incorporating environmental factors into models had little effect, Wang et al. (2007) used the algebraic difference equation approach to develop a non-linear mixed model for *Eucalyptus globulus*. Bravo-Oviedo et al. (2008) reported improvement in applicability of an inter-regional model by integrating climate and soil variables with GADA and reported that height/site index models that directly incorporate climatic variables using a mixed-effects technique and found that integrating climatic variables enhanced the fit statistics for the stand height model.

A study by González-García et al. (2015) revealed that by incorporating the site variability into growth model analysis, the model becomes more responsive to changes in the environment, improving model accuracy.

According to Scolforo et al. (2017), the treatment of soil water availability as a covariate in dominant height growth models, enhanced the ability to explain site quality for clonal eucalypt stands in Brazil. A recent study by Koirala et al. (2020) evaluated the effect of water balance components to dominant height using a GADA and found that accounting for these effects improves the precision in dominant height estimates.

Despite significant reductions in growth of this hybrid species as a result of emerging drought events in South Africa, no known efforts have been made to assess the climatic effect on growth. Although applying the findings of these and other recognized studies may be appealing, differences in variables and measurement methods may mean that those studies do not directly apply to South Africa. As a result, a study that investigates the effect of climate on growth models in South Africa is required. It should be noted, however, that rainfall is the only climatic variable available in this study.

The objective of this study is to develop a dominant height-age model that adequately explains the growth pattern of *E. grandis* x *E. urophylla* across its distribution area in Coastal Zululand, South Africa using ADA. Subsequently, test whether or not the performance of the model can be improved under different rainfall conditions by incorporating the rainfall effect in the model. Individual random variation can be explained reasonably by mixed effects commonly associated with repeated measurement data (Adame et al., 2008). Other established studies that employ the nonlinear mixed effects modeling approach include (Bailey and Clutter, 1974; Huang et al., 2009; Ni and Zhang, 2007; Nothdurft et al., 2006; Ou et al., 2016; Sharma and Parton, 2007; Sharma et al., 2018; Temesgen et al., 2008). Cieszewski and Strub (2018), Socha et al. (2021), and Sprengel et al. (2022) compared nonlinear fixed-effects and nonlinear mixed-effects modeling approaches and highlighted the behaviour of the two approaches. This study investigates both fixed and mixed-effects models. Key references on the construction and application of the nonlinear mixed-effects model include (Bates and Pinheiro, 1994; Crecente-Campo et al., 2010; Cudeck and Harring, 2007; Dorado et al., 2006; Harring and Liu, 2016; Huang et al., 2009; Huang et al., 2011; Lindstrom and Bates, 1990; Luwanda and Mwambi, 2016).

MATERIAL AND METHODS

Study area and data

The work presented in this paper uses data from Mondi plantations in Coastal Zululand (South Africa, Figure 1), with an altitude range of 12-107m. Data for developing the dominant height models were obtained from 75 permanent sample plots (PSPs), representing the variability of the stand age and site quality of *E. grandis* x *E. urophylla* in this study area. The PSPs are rectangular in shape, consist of around 400 m^2 of plot area. There are two spacings represented (3 x 2 m and 3 x 2.5 m), resulting in 1667 and 1333 trees per hectare respectively. In this study, PSPs were established between 1994 and 2015 and the first measurements were taken 0.5-11 years after planting. The dominant height of a PSP is defined as the average height of the 20% thickest trees in a PSP. Figure 2 illustrates the longitudinal profiles for the repeated measurements for PSPs. The profiles plot shows that, dominant height increases nonlinearly over time, with wide ranges among PSPs suggesting evidence of PSP-to-PSP heterogeneity. Site index of Eucalypts in Mondi is defined as dominant height at reference age (8 years). Clear-felling usually occurs between 7 and 12 years after the PSPs are planted. According to the consensus, a dominant height is measured as long as a PSP remains standing during the measurement season, so some PSPs in Figure 2 exceed the reference age. As a measure of the rainfall effect, the mean monthly precipitation (*mmp*) between measurements was used. In total, 1024 measurements were obtained for this study.

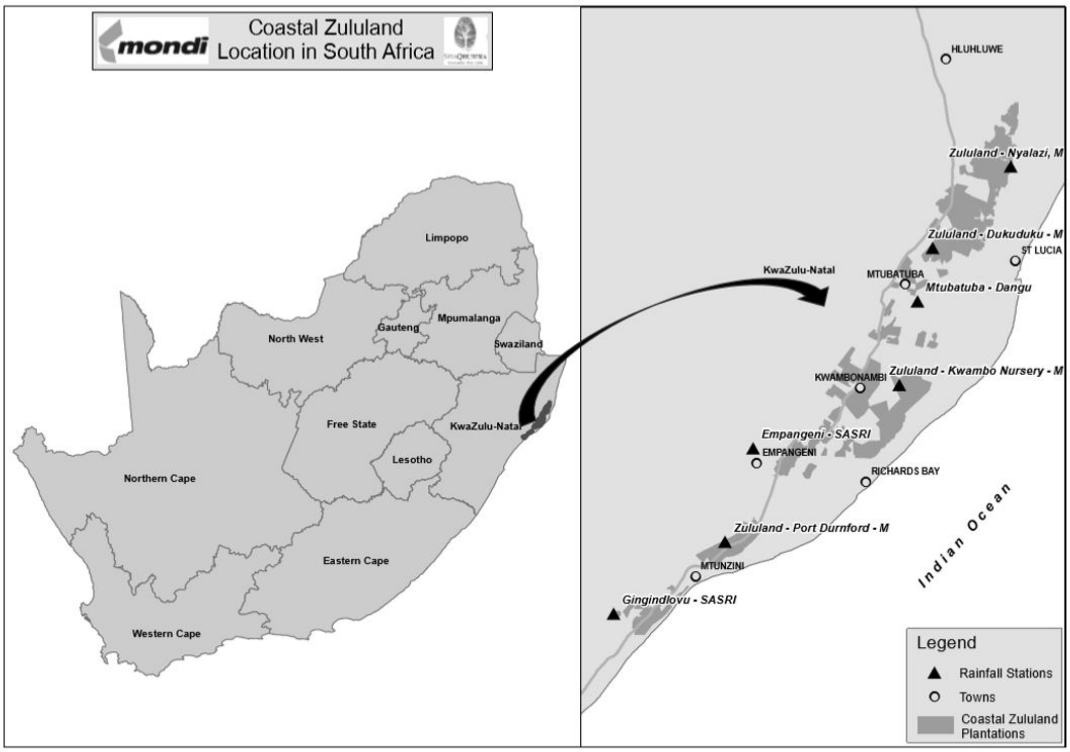

Figure 1. Map showing the studied area of *E. grandis* x *E. urophylla* and rainfall stations in Coastal Zululand.

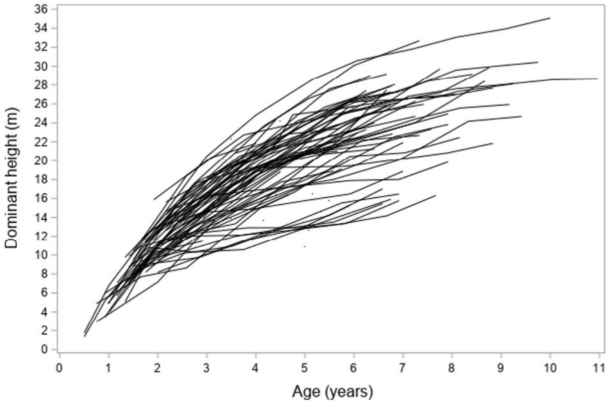

Figure 2. *E. grandis* x *E. urophylla* PSP dominant height profiles over time in Coastal Zululand.

The summary of each PSP also includes other (not used in this study) important stand measures, such as trees per hectare (TPH), quadratic mean DBH (DBHQ, cm), basal area (BA, m^2/ha), minimum diameter (Dmin, cm), site index (SI, m), and mean annual increment (MAI, m^3/ha/yr). Additionally, 652 repeat forest inventory plots within the same study area were used for model validation. Summary statistics for model variables and other important variables are presented in Table 1.

Climate data and rainfall station allocation to PSPs

The study area is classified as a subtropical climatic zone, with a mean annual temperature of approximately 22°C (Gardner et al., 2007) and a wide range of mean annual precipitation (MAP) (from 800 mm on dry sites to more than 1300 mm on wet sites).

Data on rainfall was collected from seven rainfall stations in the study area. PSPs were assigned to the rainfall stations closest to them. As seen in Table 2, the number of PSPs differed among the rainfall stations. The long term (1989-2020) annual rainfall patterns are presented in Figure 3. The drought line in the South African context is defined as 75% of the long-term mean annual precipitation (LTM) (Baudoin et al., 2017). Mean annual precipitation below 75% of the LTM is classified as drought.

Table 1. Descriptive statistics for the dominant height data used for model fitting and validation.

Variable (units)	Fitting data				Validation data			
	Mean	Minimum	Maximum	CV* (%)	Mean	Minimum	Maximum	CV* (%)
Plant year	-	1994	2015	-	-	1996	2017	-
Measurement year	-	1995	2019	-	-	2004	2020	-
Model stand variables								
Age (years)	4.5	0.5	11.0	48.4	6.0	3.0	12.8	24.0
Hd (m)	18.1	1.3	35.1	37.7	21.8	8.4	34.2	21.7
Other stand variables								
Trees per hectare	1206	534	1568	15.9	1297	934	1773	11.9
DBHQ (cm)	13.3	2.0	24.4	29.1	15.1	8.0	23.6	16.7
Basal area (m^2/ha)	29.0	7.4	51.5	28.9	23.1	6.4	51.5	30.5
Dmin (cm)	6.6	2.2	15.1	34.5	4.3	3.0	11.0	30.4
Site index (m)	25.7	15.2	34.0	15.0	25.1	12.2	33.6	16.7
MAI (m^3/ha/yr)	28.8	3.3	64.9	47.8	29.0	1.3	70.6	45.0
Rainfall effect								
Mmp (mm)	93	49	135	33	90	41	132	24

Notes: Hd=dominant height, Min=minimum, Max=maximum, and CV=coefficient of determination.

Table 2. Descriptive statistics for the 7 rainfall stations in Coastal Zululand for the period (1989 – 2020).

Rainfall Station Name	Annual Rainfall (mm) *mmp* (mm)	Mean	Min	Max	CV (%)	Number of PSPs
	mmp (mm)	954	473	1492	25	14
Gingindlovu - SASRI	76	917	513	1398	23	7
Mtubatuba - Dangu	71	854	469	1518	27	3
Zululand - Dukuduku, M	75	903	529	1431	25	11
Zululand - Kwambo Nursery, M	110	1314	660	2356	29	21
Zululand - Nyalazi, M	82	983	419	1688	28	7
Zululand - Port Dunford, M	110	1320	573	1909	23	12

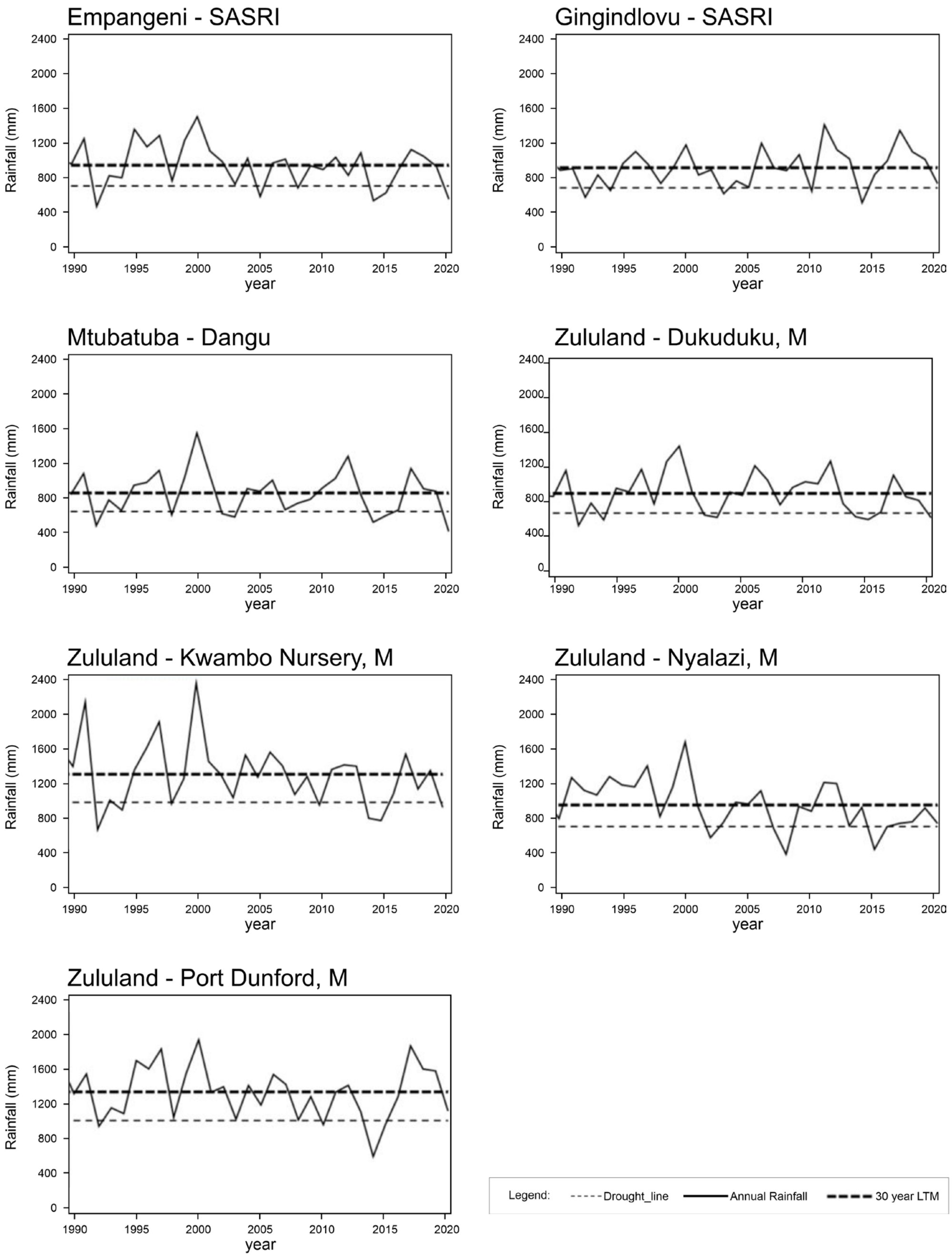

Figure 3. Annual rainfall patterns for the seven rainfall stations in Coastal Zululand.

Stand dominant height base models

When deciding on a modeling approach, we considered its applicability to temporal series data, regardless of interval length and thus a projection form. The following base growth functions were used to generate candidate models for describing dominant height: Bertalanffy-Richards (Richards, 1959), Lundqvist-Korf (Korf, 1939), McDill-Amateis (McDill and Amateis, 1992), and Hossfeld I (Kiviste et al., 2002). These functions are widely used in forest research and they are reported to be flexible in modelling stand growth attributes, including dominant height (Rojo Alboreca et al., 2017; Sánchez-González et al., 2005; Tahar et al., 2012). The ADA technique was used in this study to demonstrate the concept of incorporating rainfall into the model in the context of South Africa. As previously stated, the ADA method requires the selection of one parameter to isolate, while the rest are estimated statistically. When deciding which parameter to isolate, the desire for polymorphic site index curves and the strength of the rainfall-parameter relationship were considered. That is, among the parameters that produce polymorphic site index curves when isolated, the one with the weakest relationship to rainfall was considered site specific. In summary, an ADA technique, in which only one parameter isolated to achieve polymorphic site index with common asymptote, was selected.

The effect of rainfall on dominant height projections

Each parameter was expressed as a function of rainfall in order to determine the mathematical structure appropriate to incorporate rainfall. Parameter estimates were plotted against the mean monthly precipitation for each PSP to elucidate this dependence. Based on these plots, a linear function was found to be adequate to describe the relationship between the parameters and rainfall. Table 3 presents the base growth functions and ADA candidate models derived from them.

Model Parameterization

In general, the fixed-effects candidate (M1-M8) models can be expressed as [1].

$$y_{ij}=f(v_{ij},\beta)+e_{ij} \quad [1]$$

where y_{ij} is the j^{th} dominant height of the i^{th} PSP, f is a nonlinear function, $\beta=[\beta_0, \beta_1, \beta_2]$` are the parameters to be estimated, $v_{ij}=(y_{i1},x_{ij},x_{i1})$ and $v_{ij}=(y_{i1},x_{ij},x_{i1},mmp)$ are explanatory variables for models M1-M4 and M5-M8, respectively; y_{i1} and x_{i1} are the dominant height measured and the age of the measurement for the i^{th} PSP, respectively; e_{ij} are error terms assumed to be independent and identically distributed with mean 0.

On the other hand, the nonlinear mixed-effects models (M9-M12) can be defined as a hierarchical model as follows [2].

$$y_{ij}=f(\phi_{ij},v_{ij})+e_{ij} \quad [2]$$

where f is a general nonlinear function of an individual specific parameter vector ϕ_{ij}, a predictor vector v_{ij}, and ε_{ij} is a normally distributed within- PSP error term. The parameter vector ϕ_{ij} is modelled as [3].

$$\phi_{ij}=A_{ij}\beta+B_{ij}u_i,\ u_i\sim N(0,D) \quad [3]$$

where β is a (p×1) vector of fixed parameters common to all PSPs, u_i is a (q×1) vector of random-effects specific to PSP i, D is the variance-covariance matrix of the random effects, A_{ij} and B_{ij} are design matrices for the fixed and random effects, respectively. It is assumed that the within-group errors e_{ij} are independent and normally distributed with mean zero and variance σ^2.

The longitudinal data from PSPs were transformed to generate a structure that considers all possible forward growth intervals among dominant height-age pairs for each PSP. There has been a recommendation from previous studies Goelz and Burk (1992), Huang (1999), and Rojo Alboreca et al. (2017) for data structures that consider all possible growth intervals to ensure the most consistent and stable results. Regardless of the method used, the dominant height growth must not be assumed to be immune to autocorrelation between measurements. Inferences on parameter estimates are unreliable if the error terms are correlated ($cov(e_{ij},e_{ij-1})\neq 0$) and this correlation is not accounted for. This may affect the standard error, resulting in inflated or deflated (depending on the sign of the correlation) t statistics. This enhances the risk of making Type I or Type II errors. Thus, to obtain more efficient estimates of errors, the autocorrelation must be accounted for (Goelz and Burk, 1992).

The error terms may be expanded to allow for the first-order autocorrelation as follows [4].

$$e_{ij}=\rho e_{ij-1}+\varepsilon_{ij} \quad [4]$$

where ρ represents the autocorrelation between the residuals from estimating y_{ij} and y_{ij-1}, $\varepsilon_{ij}\sim i.i.d.N(0,\sigma^2)$.

The Durbin-Watson test by Durbin and Watson (1971) is generally employed to detect autocorrelation in error terms. The Durbin-Watson test statistic is defined as [5].

$$d=\frac{\sum(\hat{e}_{ij}-\hat{e}_{ij-1})^2}{\sum \hat{e}_{ij}^2}\simeq 2(1-\rho) \quad [5]$$

According to Holt and Refenes (1998), as $\hat{\rho}$ takes values in the range [-1,1], the Durbin-Watson can be interpreted as follows: d≃4: $\hat{\rho}$=-1, strong negative autocorrelation; d≃2: $\hat{\rho}$=0, no autocorrelation; d≃0: $\hat{\rho}$=1, strong positive autocorrelation.

This implies that d values close to 2 are evidence of the absence of autocorrelation. Values away from d=2 indicate the presence of the first-order autocorrelation. The model fitting was performed in SAS, Version 9.4 (SAS Institute Inc, 2013). Accordingly, the ETS MODEL procedure, which includes the Durbin-Watson test in its current edition, was used to fit the fixed-effects models (M1-M8). The NLMIXED procedure, on the other hand, was utilized for fitting the mixed-effects models (M9-M12). Unlike the MODEL procedure, the NLMIXED procedure does not incorporate the Durbin-Watson test. The test can be done by including zlag1 (residual.y) in the model formulation.

Table 3. Candidate equations tested for modelling dominant height.

Base function	Free parameter (F)	Solution for F with initial values (x_{i1}, y_{i1})	ADA	Model	Model type
Bertalanffy-Richards $y_{ij} = \beta_0\{1-exp(-\beta_1 x_{ij})\}^{\beta_2} + \varepsilon_{ij}$	$\beta_2 = F$	$F_0 = \frac{ln\left(\frac{y_{i1}}{\beta_0}\right)}{\ln(1-\exp(-\beta_1 x_{i1}))}$	$y_{ij} = \beta_0 \left(\frac{y_{i1}}{\beta_0}\right)^{\frac{\ln(1-\exp(-\beta_1 x_{ij}))}{\ln(1-\exp(-\beta_1 x_{i1}))}} + e_{ij}$	M1	
Lundqvist-Korf $y_{ij} = \beta_0 exp(-\beta_1 x_{ij}{}^{-\beta_2}) + \varepsilon_{ij}$	$\beta_1 = F$	$F_0 = -ln\left(\frac{y_{i1}}{\beta_0}\right) x_{i1}{}^{\beta_2}$	$y_{ij} = \beta_0 \left(\frac{y_{i1}}{\beta_0}\right)^{\left(\frac{x_{i1}}{x_{ij}}\right)^{\beta_2}} + \varepsilon_{ij}$	M2	
McDill-Amateis $y_{ij} = \frac{\beta_0}{1+\beta_2/x_{ij}{}^{\beta_1}} + \varepsilon_{ij}$	$\beta_2 = F$	$F_0 = \left(\frac{\beta_0}{y_{i1}} - 1\right) x_{i1}{}^{\beta_1}$	$y_{ij} = \frac{\beta_0}{1+\left(\frac{\beta_0}{y_{i1}} - 1\right)\left(\frac{x_{i1}}{x_{ij}}\right)^{\beta_1}} + e_{ij}$	M3	Fixed-effects with no rainfall effect
Hossfeld I $y_{ij} = \frac{x_{ij}^2}{\beta_0 + \beta_1 x_{ij} + \beta_2 x_{ij}^2} + \varepsilon_{ij}$	$\beta_1 = F$	$F_0 = \frac{x_{i1}}{y_{i1}} - \frac{\beta_0}{x_{i1}} - \beta_2 x_{i1}$	$y_{ij} = \frac{x_{ij}^2}{\beta_0 + x_{ij}\left(\frac{x_{i1}}{y_{i1}} - \frac{\beta_0}{x_{i1}} - \beta_2 (x_{ij} - x_{i1})\right)} + e_{ij}$	M4	
Bertalanffy-Richards $y_{ij} = \beta_0\{1-exp(-\beta_1 x_{ij})\}^{\beta_2} + \varepsilon_{ij}$	$\beta_2 = F$	$F_0 = \frac{ln\left(\frac{y_{i1}}{\beta_0}\right)}{\ln(1-\exp(-\beta_1 x_{i1}))}$	$y_{ij} = \beta_0^* \left(\frac{y_{i1}}{\beta_0}\right)^{\frac{ln(1-exp(-\beta_1^* x_{ij}))}{ln(1-exp(-\beta_1^* x_{i1}))}} + e_{ij}$	M5	
Lundqvist-Korf $y_{ij} = \beta_0 exp(-\beta_1 x_{ij}{}^{-\beta_2}) + \varepsilon_{ij}$	$\beta_1 = F$	$F_0 = -ln\left(\frac{y_{i1}}{\beta_0}\right) x_{i1}{}^{\beta_2}$	$y_{ij} = \beta_0^* \left(\frac{y_{i1}}{\beta_0^*}\right)^{\left(\frac{x_{i1}}{x_{ij}}\right)^{\beta_2^*}} + e_{ij}$	M6	
McDill-Amateis $y_{ij} = \frac{\beta_0}{1+\beta_2/x_{ij}{}^{\beta_1}} + \varepsilon_{ij}$	$\beta_2 = F$	$F_0 = \left(\frac{\beta_0}{y_{i1}} - 1\right) x_{i1}{}^{\beta_1}$	$y_{ij} = \frac{\beta_0^*}{1+\left(\frac{\beta_0^*}{y_{i1}} - 1\right)\left(\frac{x_{i1}}{x_{ij}}\right)^{\beta_1^*}} + e_{ij}$	M7	Fixed-effects with rainfall effect
Hossfeld I $y_{ij} = \frac{x_{ij}^2}{\beta_0 + \beta_1 x_{ij} + \beta_2 x_{ij}^2} + \varepsilon_{ij}$	$\beta_1 = F$	$F_0 = \frac{x_{i1}}{y_{i1}} - \frac{\beta_0}{x_{i1}} - \beta_2 x_{i1}$	$y_{ij} = \frac{x_{ij}^2}{\beta_0^* + x_{ij}\left(\frac{x_{i1}}{y_{i1}} - \frac{\beta_0^*}{x_{i1}} - \beta_1^* (x_{ij} - x_{i1})\right)} + e_{ij}$	M8	
Bertalanffy-Richards $y_{ij} = \beta_0\{1-exp(-\beta_1 x_{ij})\}^{\beta_2} + \varepsilon_{ij}$	$\beta_2 = F$	$F_0 = \frac{ln\left(\frac{y_{i1}}{\beta_0}\right)}{\ln(1-\exp(-\beta_1 x_{i1}))}$	$y_{ij} = \beta_0^u \left(\frac{y_{i1}}{\beta_0^u}\right)^{\frac{\ln(1-\exp(-\beta_1^u x_{ij}))}{\ln(1-\exp(-\beta_1^u x_{i1}))}} + e_{ij}$	M9	
Lundqvist-Korf $y_{ij} = \beta_0 exp(-\beta_1 x_{ij}{}^{-\beta_2}) + \varepsilon_{ij}$	$\beta_1 = F$	$F_0 = -ln\left(\frac{y_{i1}}{\beta_0}\right) x_{i1}{}^{\beta_2}$	$y_{ij} = \beta_{0i}^u \left(\frac{y_{i1}}{\beta_{0i}^u}\right)^{\left(\frac{x_{i1}}{x_{ij}}\right)^{\beta_2^u}} + e_{ij}$	M10	
McDill-Amateis $y_{ij} = \frac{\beta_0}{1+\beta_2/x_{ij}{}^{\beta_1}} + \varepsilon_{ij}$	$\beta_2 = F$	$F_0 = \left(\frac{\beta_0}{y_{i1}} - 1\right) x_{i1}{}^{\beta_1}$	$y_{ij} = \frac{\beta_0^u}{1+\left(\frac{\beta_0^u}{y_{i1}} - 1\right)\left(\frac{x_{i1}}{x_{ij}}\right)^{\beta_1^u}} + e_{ij}$	M11	Mixed-effects with rainfall effect
Hossfeld I $y_{ij} = \frac{x_{ij}^2}{\beta_0 + \beta_1 x_{ij} + \beta_2 x_{ij}^2} + \varepsilon_{ij}$	$\beta_1 = F$	$F_0 = \frac{x_{i1}}{y_{i1}} - \frac{\beta_0}{x_{i1}} - \beta_2 x_{i1}$	$y_{ij} = \frac{x_{ij}^2}{\beta_0^u + x_{ij}\left(\frac{x_{i1}}{y_{i1}} - \frac{\beta_0^u}{x_{i1}} - \beta_1^u (x_{ij} - x_{i1})\right)} + e_{ij}$	M12	

y_{ij} and y_{i1} are dominant height observations of the i^{th} PSP at age x_{ij} and age x_{i1}, respectively; β_0, β_1, and β_2 are model parameters, common to all PSPs; $\beta_p^* = \beta_p + \tau_p\ mmp$ (p=0,1,2) are expansions of the fixed-effects parameters with a linear rainfall effect (τ_p); $\beta_p^u = \beta_p + u_p + \tau_p\ mmp$ (p=0,1,2) are PSP-specific parameters with rainfall effect; e_{ij} is an error term; Models M1-M4 are fixed-effects models without the rainfall effect; models M5-M8 are fixed-effects with the rainfall effect included; and models M9-M12 are mixed-effects with rainfall effect included.

Model Evaluation

The candidate models were evaluated graphically and using quantitative statistical measures. The graphical analysis consisted of visually inspecting 1) dominant height/site index curves for biological plausibility, and 2) plots of residuals against predicted values for possible systematic discrepancies. The quantitative statistical measures consisted of three statistical criteria: root mean square error (*RMSE*), the adjusted coefficient of determination (R^2_{adj}), and Akaike's information criterion (AIC). These criteria are defined as follows [6], [7], and [8].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \hat{y}_{ij})^2} \quad [6]$$

$$R^2_{adj} = 1 - \frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \hat{y}_{ij})^2}{\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2} \quad [7]$$

$$AIC = 2p - 2\log(L), \quad [8]$$

where y_{ij} and $\hat{y}_{ij}$ are observed and predicted values of dominant height for PSP i with n_i observations, $\bar{y}$ is the mean of all y_{ij}, $n = \sum_{i=1}^{m} n_i$, p is the number of model parameters, L is the value of the likelihood function for a model. Fitted models with the largest value of R^2_{adj} and the smallest values of AIC, and *RMSE* are preferred.

In order to assess the prediction quality of the models, the models were applied to an independent dataset. Studies by Huang (1999) and Krisnawati et al. (2009) also emphasize that the data used to estimate the model parameters may not be used to evaluate prediction quality. The statistical criteria (*RMSE*, R^2_{adj}, AIC) were also employed as validation statistics calculated from an independent dataset (validation data) to help assess the quality of the prediction. The following statistical criteria were also considered over and above the stated criteria: the mean residual (*MRES*), measuring systematic deviation from predictions, the absolute mean residual (*AMRES*), which ignores the sign of error in prediction, the mean absolute percentage error (*MAPE*), and the Schwarz Bayesian information criterion (BIC). These validation statistics were defined as follows [9], [10], [11], and [12].

$$MRES = \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \hat{y}_{ij}) \quad [9]$$

$$AMRES = \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i}|y_{ij} - \hat{y}_{ij}| \quad [10]$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i}|y_{ij} - \hat{y}_{ij}| \quad [11]$$

$$BIC = p\log(n) - 2\log(L), \quad [12]$$

In general, smaller values of *MRES*, *AMRES*, *MAPE*, and BIC indicate better performance.

RESULTS AND DISCUSSION

The Nonlinear least-squares approach

In Figure 4, residual plots are shown for Models M1-M8. Based on the plot of residuals against predicted values, it appears that residuals are not randomly distributed above and below the reference line of zero at both tail ends of the models without the rainfall effect (M1-M4). This indicates that the assumption of constant variance may not been adhered to completely. Similarly, the Q-Q plots for models M1-M4 indicate departure from the normality assumption at both tail ends. The residuals against predicted values are evenly spread over and below zero in fixed-effects models which include the rainfall effect (M5-M8), in accordance with the constant variance assumption. Furthermore, the Q-Q plots indicate no significant departures from normality, except for a few observations at the lower tail. In general, there was no substantial departure from model assumptions for the models that included rainfall as a covariate.

In order to determine whether or not the correlated error term was required for each candidate function, the ADA models (M1 - M4) were fitted without including the correlated error first. The fit statistics are presented in Table 4. As can be seen from the R^2_{adj}, all fitted models explain a significant amount of variability in the data. However, the Durbin-Watson statistics were low [0.48, 0.51], suggesting a positive autocorrelation for the residual terms.

Table 4. Fit statistics for models M1-M4 without accounting for correlated error structure.

Model	MSE	Root MSE	R^2_{adj}	AIC	Durbin Watson
M1	7.3735	2.7154	0.8015	5671.52	0.4806
M2	6.317	2.5134	0.83	5490.28	0.4911
M3	6.7167	2.5917	0.8192	5562.18	0.4872
M4	6.4917	2.5479	0.8253	5522.24	0.5059

The Durbin-Watson statistics are close to 2 [2.09, 2.10] after including autocorrelation. According to Table 5, by using first-order autocorrelation in the error terms, we were not only able to eliminate the autocorrelation issue, but also improved the fit statistics substantially.

Table 5. Fit statistics for models M1-M4 with correlated error structure accounted for.

Model	MSE	Root MSE	R^2_{adj}	AIC	Durbin Watson
M1	3.0617	1.7498	0.9176	4642.44	2.0884
M2	2.6087	1.6151	0.9298	4454.76	2.1016
M3	2.7347	1.6537	0.9264	4510.04	2.0995
M4	2.6784	1.6366	0.9279	4485.66	2.109

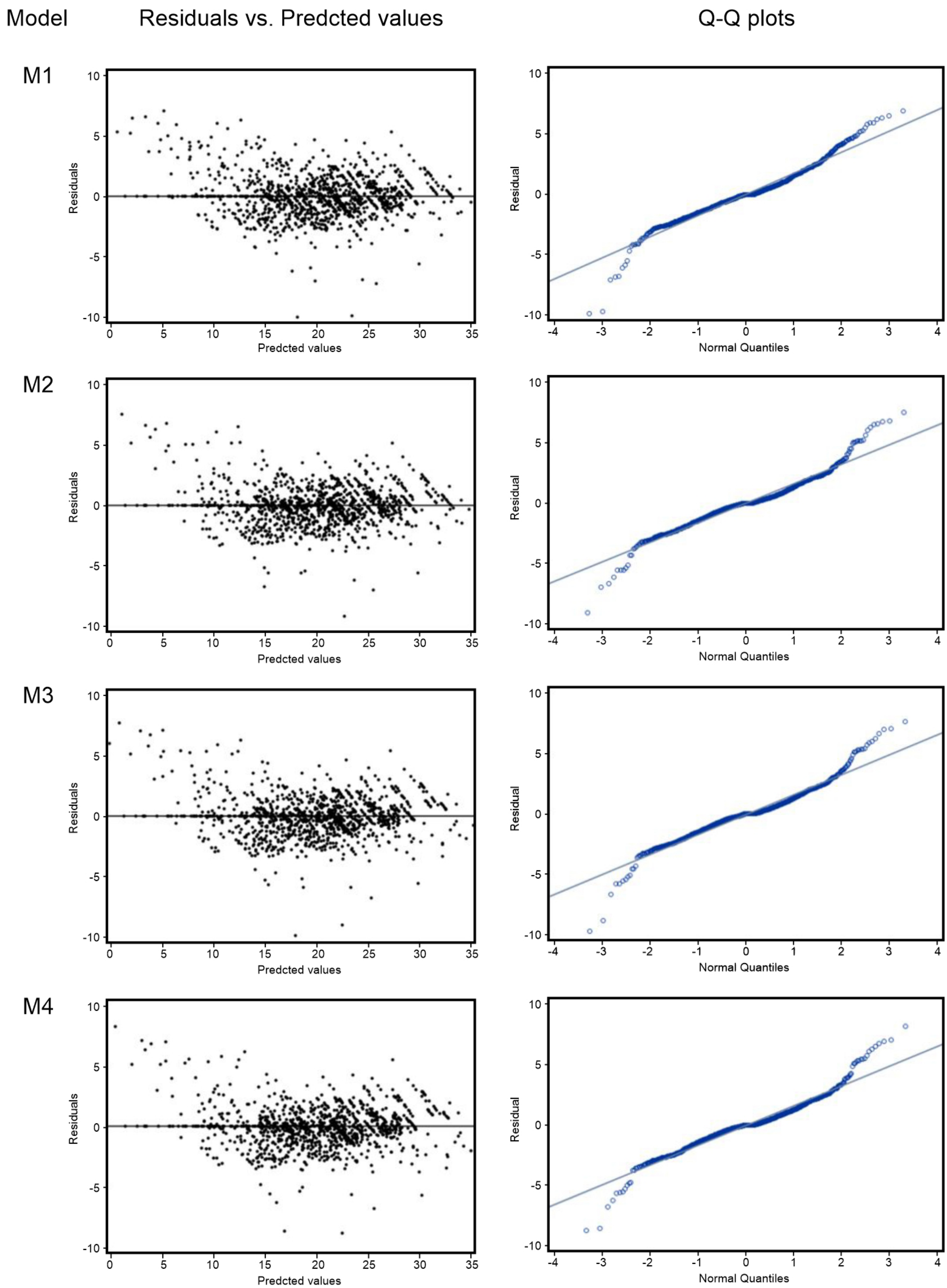

Figure 4. Scatter plot of residuals vs predicted values (left) and Q-Q normality plots (right) for Models M1-M8.

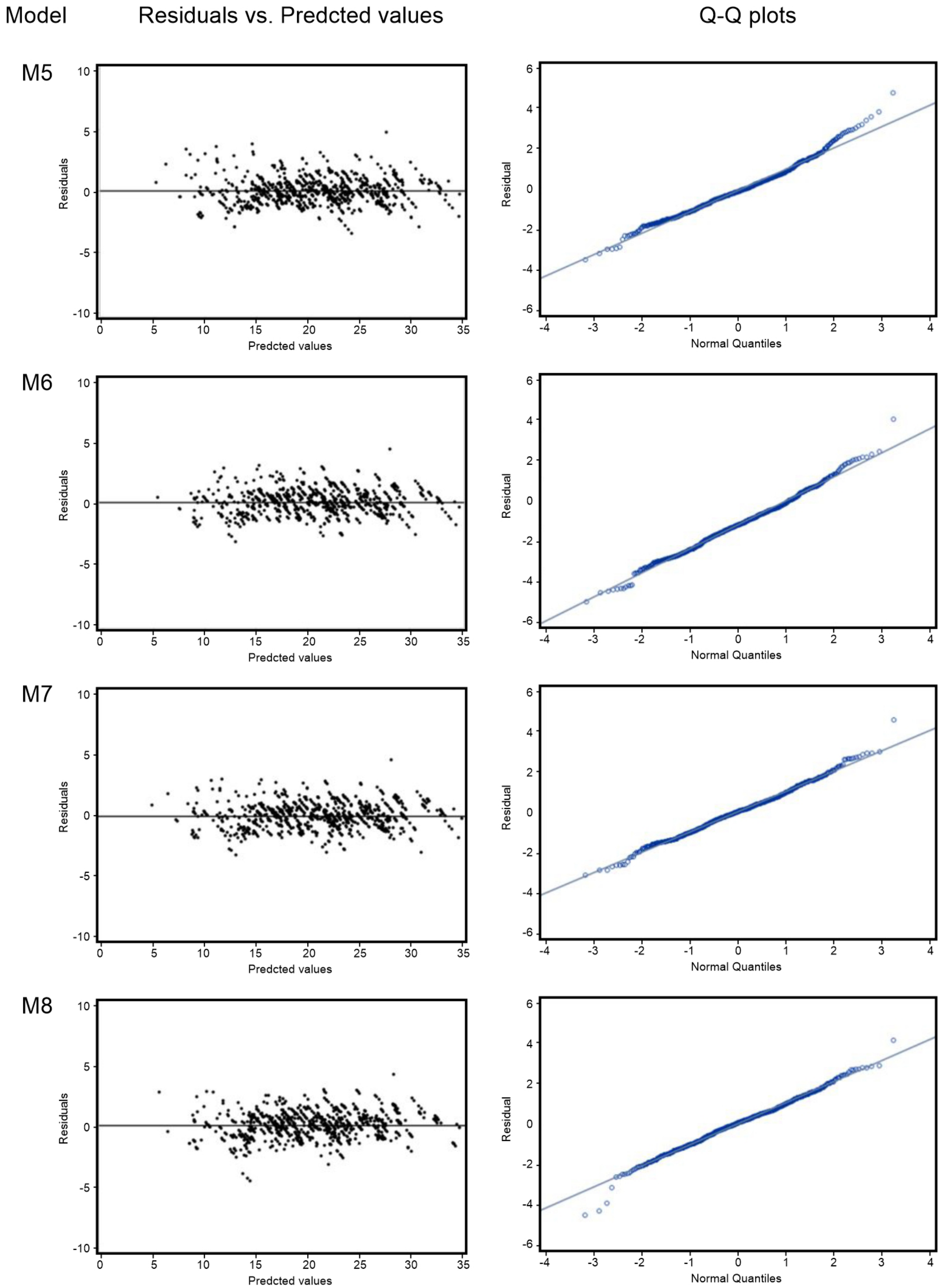

Figure 4 continuation. Scatter plot of residuals vs predicted values (left) and Q-Q normality plots (right) for Models M1-M8.

The parameter estimates and fit statistics for the fixed-effects models with and without the rainfall effect are summarized in Table 6. All parameter estimates are statistically significant at the 5% level for the fixed-effects models without the rainfall effect (M1-M4). Models M5-M8 are fixed-effects that incorporate rainfall effect in both model parameters, the results show that the rainfall effect is only significant in one of the two model parameters. The rainfall effect can be positive or negative and it is different for each model. *RMSE* values for models (M5-M8) are smaller than those of models (M1-M4), suggesting that the inclusion of rainfall effect generally improves model precision. By averaging the values in Table 6 for models M1-M4 and then for models M5-M8, it can be shown that the improvement in precision by the inclusion of rainfall effect is approximately 37% ((1.66-1.04)/1.66). Despite the small differences, Lundqvist-Korf functions produced smaller *RMSE* values than the other three functions. The R^2_{adj} values are at least 0.9 for all eight models, indicating that each model explains at least 90% of the total variability in the data. Generally, the R^2_{adj} values for models (M5-M8) are better than those for models (M1-M4), indicating that the rainfall effect improves the percentage of total variability explained. In comparison to the other three functions, the Lundqvist-Korf function produced marginally higher values of R^2_{adj}. In models (M5-M8), AIC values are lower than those in models M1-M4, which indicates a better goodness of fit through the inclusion of rainfall. Similar to what was seen with R^2_{adj} , slightly lower values of AIC were obtained with the Lindqvist-Korf function than with the other two functions. According to the fit statistics presented thus far, the fixed-effects formulation of the Lundqvist-Korf function that incorporates the rainfall effect (model M6) is the best fit among the eight fixed-effects candidate models.

Table 6. Parameter estimates and fit statistics for the dominant height growth models M1-M8.

Model	Parameter	Estimate	Std Err	P-value	*RMSE*	R2adj	AIC
M1	b_0	52.7344	2.2230	<.0001	1.740	0.918	4642.4
	b_1	-0.0447	0.0040	<.0001			
	ρ	0.7686	0.0093	<.0001			
M2	b_0	78.0573	3.7305	<.0001	1.615	0.930	4454.8
	b_2	0.3803	0.0114	<.0001			
	ρ	0.7762	0.0091	<.0001			
M3	b_0	42.1776	0.7371	<.0001	1.654	0.926	4510.0
	b_1	1.0218	0.0117	<.0001			
	ρ	0.7809	0.0090	<.0001			
M4	b_0	0.0261	0.0030	<.0001	1.637	0.928	4485.7
	b_2	0.0248	0.0003	<.0001			
	ρ	0.7815	0.0091	<.0001			
M5	b_0	36.9920	3.9101	<.0001	1.075	0.965	2440.6
	b_1	0.0410	0.0145	0.0048			
	b_3	-0.0015	0.0002	<.0001			
	b_4	0.0481	0.0374	0.1994			
	ρ	1.1234	0.0192	<.0001			
M6	b_0	110.5158	17.3146	<.0001	1.009	0.969	2337.5
	b_2	0.0760	0.0269	0.0048			
	b_3	-0.2067	0.1464	0.1585			
	b_4	0.0031	0.0004	<.0001			
	ρ	1.1049	0.0180	<.0001			
M7	b_0	48.7329	5.1480	<.0001	1.023	0.968	2359.4
	b_1	0.2447	0.0510	<.0001			
	b_3	0.0085	0.0006	<.0001			
	b_4	-0.0379	0.0485	0.4356			
	ρ	1.1031	0.0176	<.0001			
M8	b_0	0.0096	0.0038	<.0001	1.063	0.966	2422.9
	b_2	0.0427	0.0016	<.0001			
	b_3	-0.0001	0.0000	<.0001			
	b_4	-0.0002	0.0000	0.7416			
	ρ	1.0960	0.0182	<.0001			

The nonlinear mixed-effects approach

The plots of residuals for the nonlinear mixed-effects candidate models M9-M12 are presented in Figure 5. It can be seen that the residuals for model M10 are not randomly distributed above and below zero, and thus not adhering to the assumption of constant variance while the rest of the candidate models appear to adhere to this assumption. Looking at the Q-Q plots, again model M10 appears to not adhere to the assumption of normality, while the rest of the models satisfactorily adhere to this assumption.

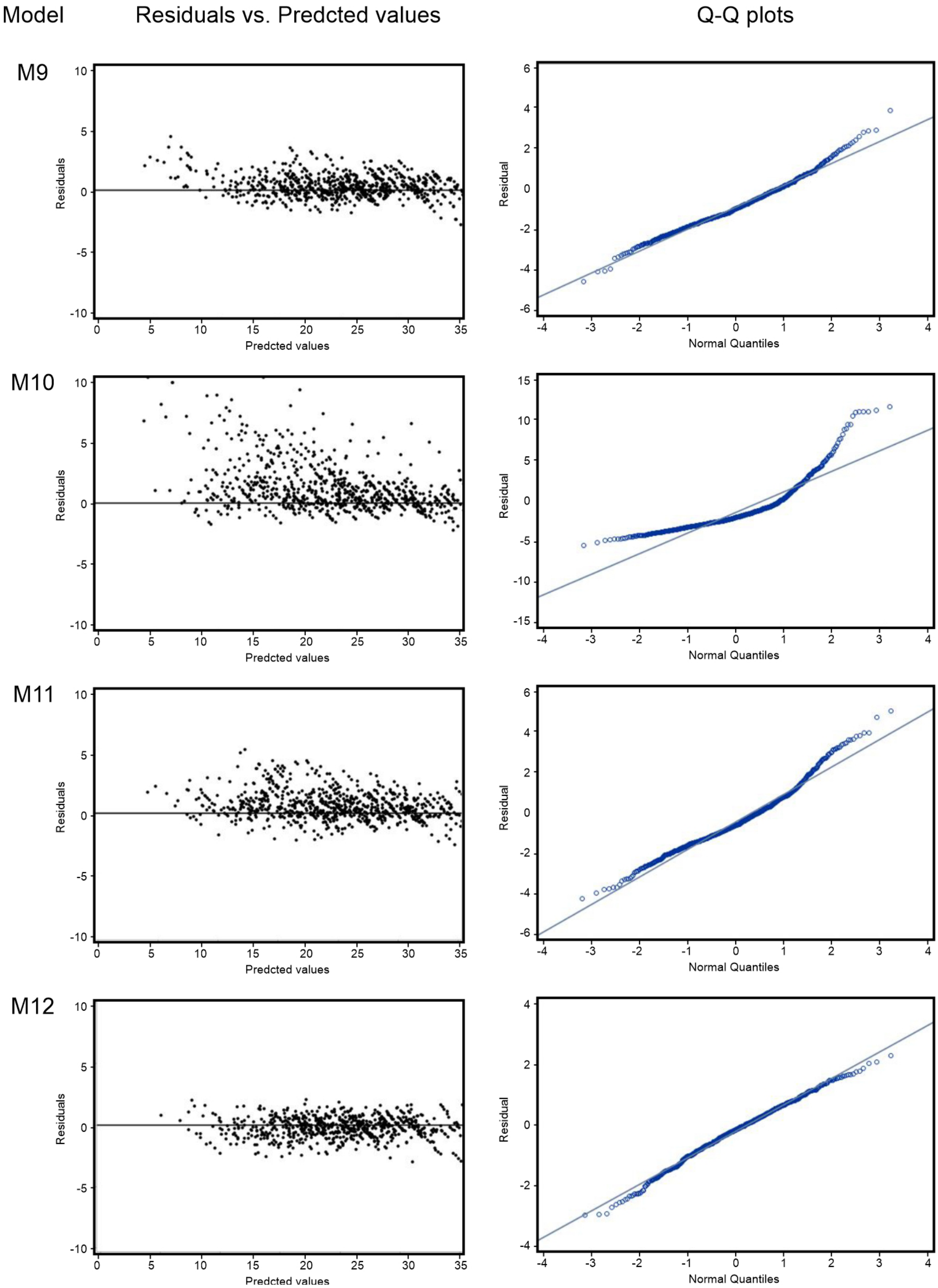

Figure 5. Scatter plot of residuals vs predicted values (left) and Q-Q normality plots (right) for Models M9-M12.

The parameter estimates and fit statistics for the mixed-effects models (M9-M12) are presented in Table 7. All fitted models (M9-M12) explain a significant amount of variability in the data. For all four models, the variances (σ_i^2,*i*=0,1,2) associated with the random effects were significant (p-values < 0.05), indicating that model parameters varied between PSPs. It can be seen that nonlinear mixed-effects with the rainfall effect included have approximately 18 % better precision in terms of *RMSE* ((1.04-0.86)/1.04) than similar models that consider all parameters as fixed by averaging the *RMSE* values for fixed-effects models (M5-M8) and mixed-effects models (M9-M12). The R^2_{adj} values are are at least 0.89, indicating that each model explains at least 89% of the variability in the data. The Bertalanffy-Richards formulation (M9) produced the lowest value of *RMSE*, indicating better precision than the other three mixed-effects models. Additionally, Hossfeld model formulation (M12) had the lowest value of AIC, making it a contender for consideration as the best fit among the candidate models.

Table 7. Parameter estimates and fit statistics for the dominant height growth models M9-M12.

Model	Parameter	Estimate	Standard	Pr > |t|	95% Confidence Limits		*RMSE*	R^2 adj	AIC
M9	b_0	30.951	1.352	<.0001	28.250	33.652	0.787	0.984	2276.9
	b_1	0.108	0.013	<.0001	0.082	0.133			
	b_3	-0.003	0.000	<.0001	-0.004	-0.003			
	b_4	0.017	0.013	0.2043	-0.009	0.043			
	σ_0^2	51.202	14.507	0.0008	22.230	80.175			
	σ_{01}	0.242	0.152	0.1169	-0.062	0.546			
	σ_1^2	0.007	0.002	0.0021	0.003	0.012			
	σ_ε^2	0.620	0.033	<.0001	0.553	0.686			
	ρ	0.012	0.005	0.0159	0.002	0.022			
M10	b_0	5.097	3.301	0.1274	-1.496	11.689	0.870	0.896	2402.4
	b_2	0.549	0.040	<.0001	0.469	0.630			
	b_3	0.566	0.045	<.0001	0.477	0.655			
	b_4	0.000	0.000	0.3209	-0.001	0.000			
	σ_0^2	651.070	247.690	0.0107	156.400	1145.740			
	σ_{02}	-2.833	1.097	0.0121	-5.024	-0.642			
	σ_2^2	0.021	0.008	0.0066	0.006	0.036			
	σ_ε^2	0.757	0.041	<.0001	0.676	0.839			
	ρ	0.025	0.005	<.0001	0.015	0.035			
M9	b_0	37.977	3.431	<.0001	31.124	44.830	0.908	0.970	2471.5
	b_1	0.038	0.077	0.6244	-0.116	0.192			
	b_3	0.011	0.001	<.0001	0.010	0.013			
	b_4	0.014	0.031	0.6576	-0.049	0.077			
	σ_0^2	168.330	66.297	0.0135	35.927	300.730			
	σ_{01}	-2.196	1.103	0.0507	-4.399	0.007			
	σ_1^2	0.093	0.028	0.0015	0.037	0.149			
	σ_ε^2	0.825	0.045	<.0001	0.736	0.914			
	ρ	0.033	0.005	<.0001	0.023	0.044			
M12	b_0	-0.0436	0.0237	0.0704	-0.0909	0.0037	0.871	0.978	2222.7
	b_2	0.0503	0.0025	<.0001	0.0453	0.0553			
	b_3	-0.0003	0.0000	<.0001	-0.0003	-0.0002			
	b_4	0.0003	0.0002	0.2008	-0.0002	0.0007			
	σ_0^2	0.0128	0.0040	0.002	0.0048	0.0207			
	σ_{02}	0.0004	0.0002	0.0849	-0.0001	0.0008			
	σ_2^2	0.0001	0.0000	<.0001	0.0000	0.0001			
	σ_ε^2	0.7583	0.0430	<.0001	0.6725	0.8442			
	ρ	0.0179	0.0054	0.0015	0.0071	0.0287			

To evaluate the biological behaviour of the candidate models, the growth curves of each candidate model were examined. Growth curves were derived from the site index classes 15, 25, and 35 m, covering the site index range observed in Table 1. For models M5 – M12, *mmp* values of 49, 90, and 135 mm were used for site index values of 15, 25, and 35 m respectively. These predictions were derived by setting y_0 and x_0 in each candidate model to be the site index and reference age, where the reference age for determining the site index of a stand was based on the age range with the lowest error, an approach used in similar previous studies by González-García et al. (2015) and Tahar et al. (2012), and based on the results, the reference age was at 8 years (also in line with current practice at Mondi Forests), respectively. The results are presented in Figure 6. To interpret the results in Figure 6, one must remember that each column represents the same class of models (i.e., models M1-M4 are fixed-effects derived from the four mathematical functions without the rainfall effect, models M5-M8 are fixed-effects of the four mathematical functions with the rainfall effect accounted for in the parameters, and models M9-M12 are mixed-effects derived from the four mathematical functions with the rainfall effect accounted for). The rows in Figure 6 show the three different model forms derived from the same function. Models M1, M5, and M9, for example, are fixed-effects without rainfall, fixed-effects with rainfall, and mixed-effects with rainfall, respectively, and are based on the Bertalanffy-Richards Richards. The results indicate that the models without the rainfall effect produced unrealistically high values of dominant height at early stages for all four mathematical functions. Incorporating the rainfall effect into the fixed-effects models improved the model fit significantly, and particularly corrected the unrealistic dominant height estimates at an early stage. However, the Bertalanffy-Richards formulation still produced unrealistically high values early on. Including the rainfall effect in mixed-effects models worked even better, except for Bertalanffy-Richards which failed to produce sensible values for high site indices at an early stage. At an early age, the Hossfeld formulation (M12) fails to distinguish between the average and high site index. There are indications in Figure 6 that models with rainfall are more accurate. Specifically, mixed-effects models appear to perform better than those with all parameters fixed. At this stage of the analysis, there is not enough evidence to declare one of the twelve candidate models as the best, but it appears from Figure 6 that mixed-effects derived from the Bertalanffy-Richards is the worst candidate. Additionally, model M10, which is a mixed-effects formulation with rainfall included from the Lundqvist-Korf function, produced polymorphic site index curves that were desirable, despite incompatibility with distributional assumptions on residuals, making it a viable candidate for the best fit among the twelve models considered.

Model validation

In this study, each model was used to predict dominant heights based on the validation data (repeat forest inventory). The validation statistics are presented in Table 8. It can be observed that the prediction errors of the candidate models are minimal, with the largest values of *MRES*, *AMRES*, and *MAPE* within 0.65 m, 1.39 m, and 6.6%, respectively. At least 93% of the variability in the validation data was explained by all candidate models (R^2_{adj} values of at least 0.93). The AIC and BIC values indicate that models without rainfall effects are not as good at predicting dominant height values, particularly at an early stage. Although this does not necessarily imply that all models that accounted for rainfall effect worked better, as seen in Model M9, which accounts for rainfall effect in Bertalanffy-Richards parameters, which exhibited a strange behaviour in Figure 6. This model (M9) performed the lowest on validation data based on the selection criteria. According to the selection criteria, model M11, a mixed-effects model derived from the McDill and Amateis functions and including rainfall effects in its parameters, produced the best results, followed by model M10, a model derived from the Lundqvist-Korf function, which produced the best overall results in Figure 6.

According to the statistics provided thus far, Model 11 and Model 10, mixed-effects formulations of McDill-Amateis and Lundqvist-Korf functions, respectively, consistently outperformed the other models. It is acknowledged that the distributional assumptions on residuals were not compatible with model M10, however this model showed good behaviour on observation data and performance on validation data cannot be overlooked.

Table 8. Validation statistics of the twelve candidate dominant height models.

Model	Equation	*MRES* (m)	*AMRES* (m)	*MAPE* (%)	*RMSE* (m)	R^2 adj	AIC	BIC
M1	CR	-0.453	1.353	6.255	1.752	0.931	1234.2	1254.2
M2	LK	-0.104	1.242	5.617	1.617	0.942	1059.1	1079.1
M3	MA	-0.113	1.262	5.739	1.643	0.94	1093.2	1113.2
M4	HF	-0.003	1.256	5.68	1.645	0.94	1095.5	1115.4
M5	CR	-0.639	1.343	6.249	1.763	0.93	1251.5	1281.5
M6	LK	-0.294	1.153	5.233	1.528	0.948	938.1	968.1
M7	MA	-0.299	1.164	5.315	1.535	0.947	948.5	978.5
M8	HF	-0.368	1.185	5.425	1.557	0.946	979.7	1009.6
M9	CR	-0.396	1.388	6.5	1.809	0.927	1313.6	1358.6
M10	LK	-0.122	1.139	5.188	1.498	0.95	901.9	946.9
M11	MA	0.006	1.128	5.07	1.491	0.95	891.3	936.2
M12	HF	0.006	1.159	5.149	1.537	0.947	957.3	1002.3

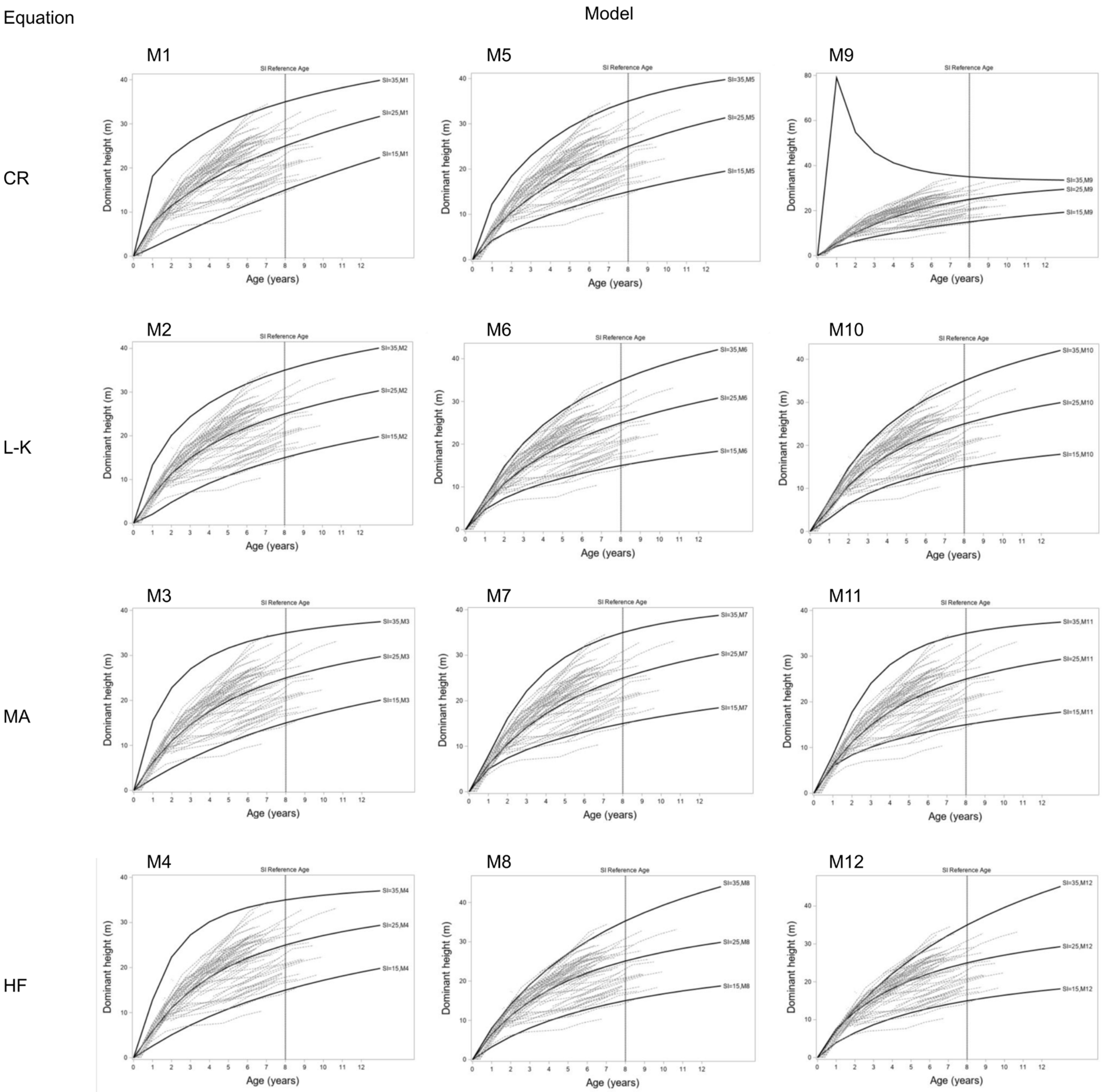

Figure 6. Dominant height curves for site index classes 15, 25, and 35 m at reference age of 8 years for models M1-M12.

Summary and model application

In this study, an ADA by Bailey and Clutter (1974) was used to model the dominant height for *E. grandis* x *E. urophylla* plantations in Coastal Zululand, South Africa. The rainfall effect was subsequently incorporated using methods from existing studies to examine whether improved model performance is possible under different rainfall conditions. A total of twelve candidate models were studied and fitted to the data obtained from permanent sample plots in this area. The candidate models were based on Bertalanffy-Richards, Lundqvist-Korf, McDill-Amateis, and Hossfeld I growth functions. All models accounted for serial correlation by using the autoregressive error term. The twelve candidate models were fitted sequentially as: fixed-effects without a rainfall effect (M1-M4), fixed-effects with a rainfall effect (M5-M8), and mixed-effects with a rainfall effect (M9-M12).

A linear addition of the rainfall effect was made to the model parameters to account for height growth changes caused by rainfall variability. Including rainfall effect improved the fit statistics, and this was expected, since rainfall is one of the primary drivers of growth. In this study, we had information on rainfall and no other climatic variable(s).

Based on the results, the inclusion of the rainfall effect improved the dominant height models by approximately 37%. The nonlinear mixed-effects with the rainfall effect included have on average 18% better precision than similar models that consider all parameters to be fixed. This is consistent with the findings from a study by Wang et al. (2007) that including the random effects improved the dominant height model when compared to a similar model with all parameters considered as fixed. Cieszewski and Strub (2018), Socha et al. (2021), and Sprengel et al. (2022) compared nonlinear fixed-effects and nonlinear mixed-effects modeling approaches found better behaviour with nonlinear effects with GADA. The distributional assumptions on residuals were not compatible with the nonlinear mixed-effects formulation of Lundqvist-Korf function. On the other hand, the nonlinear mixed-effects of Bertalanffy-Richards function produced the best fit statistics and adhered to the model assumptions. However, a visual inspection of site index trajectories revealed uncharacteristic behaviour for model M9, particularly at early ages of high site index class, suggesting that this model should not be used. The rest of the models without the rainfall effect (M2-M4) also produced unrealistic values at young ages for a high site class, a behaviour that was significantly improved by incorporating the rainfall effect.

Although the mixed-effects model M10 violated the model assumptions, it produced desirable site index curves for the site index range, showing good behaviour.

Validation was done using repeat forest inventory data for all candidates. In interpreting the validation statistics, it should be acknowledged that repeat forest inventories utilize the same stands, but not necessarily the same trees, unlike PSPs. Validation statistics were comparable for the models, but the mixed-effects of McDill-Amateis formulation (Model M11) produced the best statistics overall. Including the rainfall effect in the models resulted in differences in growth trajectories, showing that growth trajectory is affected by rainfall. This result is consistent with a study by Bravo-Oviedo et al. (2008), who reported that differences in growth trajectory depend partially on climate and soil conditions at specific sites.

The evidence produced in model behaviour and performance on validation data leads to two final models M10 and M11, the mixed-effects formulations of Lundqvist-Korf and McDill Amateis, respectively. Without loss of generality we consider the case final conditional models M10 and M11 with PSP specific random effects assumed to be at the expected value of zero. Then the two fitted models are respectively

$$y = (5.097 + 0.566 * mmp)\left(\frac{y_0}{5.097 + 0.566 * mmp}\right)^{\left(\frac{x_0}{x}\right)^{0.549-0.0003*mmp}} \quad [10]$$

and

$$y = \frac{37.977 + 0.014mmp}{1 + \left(\frac{37.977 + 0.014mmp}{y_0} - 1\right)\left(\frac{x_0}{x}\right)^{0.038+0.011mmp}} \quad [11]$$

In order to demonstrate the application of the selected models, recall that Figure 3 shows rainfall patterns for various rainfall stations. When looking at the patterns closely, it can be observed that a pronounced drought was experienced in the 2014-2015 period. By comparing the forest inventories before drought (2014-2016) and after drought (2016), we can demonstrate the behaviour of the selected models. For this comparison, 40 repeat inventories were available. We will compare these two models to their counterparts without the rainfall effect in order to highlight the effects of rainfall (i.e. M2 vs. M10, M3 vs. M11). The results are presented in Table 9. It can be seen that prediction errors of the two final models (M10, M11) are minimal (over-prediction within 0.2 m using the *MRES*). When comparing these models to their counterparts without the rainfall effect (M2, M3), all statistics presented indicate that including the rainfall effect increases projection accuracy significantly (at least 80 percent improvement in terms of observed mean error).

Additionally, the utility of the case model can be proved by projecting the behaviour of a PSP. Figure 7 depicts a comparison of Models M3 and M11 in their projection accuracy on the PSP measurements from 2013 to 2021. The interval between the measurements is approximately one year. The 2013 measurement was used as (x_0 ,y_0) for projection in 2014 to 2021 using the model without the rainfall effect (M3). Projections were also done using model M11, which accounts for the *mmp* between measurement intervals. To account for the *mmp* between measurements, the recent previous measurement was used as reference, for example, the 2014 measurement was used as (x_0 ,y_0) for the 2015 projection.

In this example, the coefficient of variation in *mmp* is 26.4%, which is representative of the variability in this area (Table 2). The results show that projecting from the first observations without taking the rainfall effect into account can result in overly optimistic values. A projection that account for rainfall between measurements, on the other hand, are generally accurate and reflect the PSP's behavior. The observed final (2021) dominant height measurement is 22.3 m; however, using the final model without rainfall effect (M3), the projected final dominant height measurement is 29.8 m, a 33.4% overestimation. The projection that includes the rainfall effect, on the other hand, is 22.4 m, a 0.5% overestimation. This shows that, despite the high variability in the rainfall data, accounting for rainfall effects produces estimates that are closer to reality.

Table 9. Statistics for a selected models applied to inventory data projected from 2013-2014 to 2016.

Model	Equation	*MRES* (m)	*AMRES* (m)	*MAPE* (%)	*RMSE* (m)	R^2 adj	AIC	BIC
Lundqvist-Korf without and with rainfall								
M2	LK	-1.05	1.282	5.799	1.518	0.946	41.4	48.2
M10	LK	-0.201	0.963	4.202	1.179	0.962	31.1	46.3
McDill-Amateis without and with rainfall								
M3	MA	-0.968	1.22	5.674	1.46	0.95	38.3	45
M11	MA	-0.101	0.858	3.743	1.039	0.971	21.1	36.3

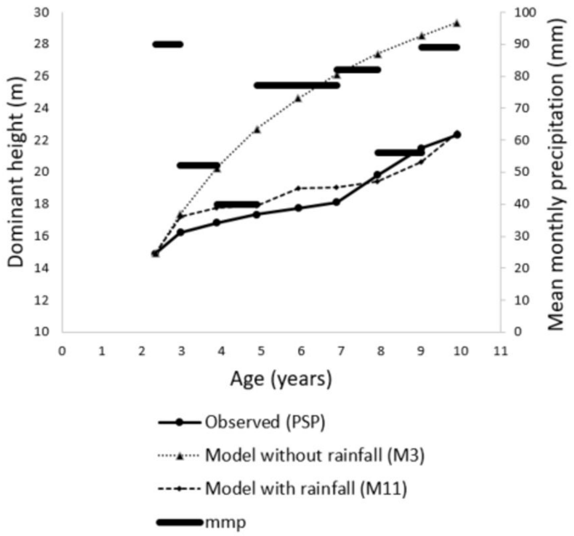

Figure 7. Projections from a model without and with rainfall effect on observed PSP dominant height data.

CONCLUSION

In this study, polymorphic ADA models for projecting the dominant height of *E. grandis* x *E. urophylla* in Coastal Zululand, South Africa, were developed. The data consists of PSPs from different sites, and high correlations were observed for measurements from the same PSP as well as differences between PSPs, indicating that nonlinear mixed-effects models were more appropriate. The final models chosen were nonlinear mixed-effects models derived from Lundqvist-Korf and McDill-Amateis base functions because they represent expected biological behaviour and performed well on validation data.

Rainfall is variable in this study area (weighted CV =26.9%). The reported improvement in precision due to incorporating the rainfall effect in the model is approximately 37%. Despite the high variability in rainfall, the model application (Figure 7) demonstrates that the rainfall incorporation in the model produces estimates that are close to reality. Similar models that do not account for the effect of rainfall, on the other hand, can produce unrealistic estimates. The models that incorporate rainfall are therefore desirable as they can be used for estimates, particularly in short-rotation pulp forests where rainfall fluctuates. These models were fitted and tested on a single hybrid species on a zone of climate characterized by hot and humid summers, and cool to mild winters. These models may not be directly transferable to other species or regions. Therefore, extrapolation to another domain should be done with caution. There are some limitations in this study: Other factors such as temperature, soil water deficit, soil fertility, etc. may become available in the future and could be incorporated in the model using the methodology presented. The explanatory variables are assumed to be measured without error in this study. Future research should include extending the models to account for measurement error in explanatory variables. Finally, South Africa would benefit from a polymorphic model with multiple asymptotes as this hybrid species continues to expand to different areas and sites. Therefore, areas for future research extending the current polymorphic ADA model to represent multiple asymptotes, that is GADA.

ACKNOWLEDGMENTS

The authors would like to thank Mondi South Africa (Pty) Ltd for providing funding and research data and the South African Sugarcane Research Institute (SASRI) for assisting with rainfall data.

AUTHORSHIP CONTRIBUTION

Project Idea: MC, HK, HM
Database: MC, HK
Processing: MC, HM
Analysis: MC, HK, HM
Writing: MC, HK, HM
Review: MC, HK, HM

REFERENCES

ADAME, P., DEL RÍO, M., CAÑELLAS, A mixed nonlinear height–diameter model for pyrenean oak (*Quercus pyrenaica* Willd). Forest Ecology and Management, v. 256, n. 1-2, p. 88-98, 2008.

ANTA, M., DIÉGUEZ-ARANDA, U. Site quality of pedunculate oak (*Quercus robur* L.) stands in Galicia (northwest Spain). European Journal of Forest Research, v. 124, p. 19-28, 2005.

BAILEY, R.L., CLUTTER, J.L. Base-age invariant polymorphic site curves. Forest Science, 20, n.2, p. 155-159, 1974.

BATES, D.M., PINHEIRO, J. Model Building for Nonlinear Mixed Effects Models. University of Wisconsin, Department of Biostatistics, 1994.

BAUDOIN, M.-A., VOGEL, C., NORTJE, K., NAIK, M. Living with drought in South Africa: lessons learnt from the recent El Niño drought period. International Journal of Disaster Risk Reduction, v. 23, p. 128-137, 2017.

BAUR, FERTRAGS- ODER ZUWACHSTAFELN FÜR DIE FICHTE. IN: F. BAUR (Editor), Die Fichte in Bezug auf Ertrag, Zuwachs und Form: Unter Zugrundlegunf der an der K. Württ. forstlichen Versuchsanstalt angestellten Untersuchungen. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1-58, 1877.

BORDERS, B., BAILEY, R., WARE, K. Slash pine site index from a polymorphic model by joining (splining) nonpolynomial segments with an algebraic difference method. Forest Science, v. 30, n. 2, p. 411-423, 1984.

BRAVO-OVIEDO, A., TOME, M., BRAVO, F., MONTERO, G., DEL RIO, M. Dominant height growth equations including site attributes in the generalized algebraic difference approach. Canadian Journal of Forest Research, v. 38, n. 9, p. 2348-2358, 2008.

BURKHART, H.E., TOMÉ, M. Modeling forest trees and stands. Springer Science & Business Media, 2012.
CARMEAN, W.H., LENTHALL, D.J. Height-growth and site-index curves for jack pine in north central Ontario. Canadian Journal of Forest Research, v. 19, n. 2, p. 215-224, 1989.

CIESZEWSKI, C.J. Three methods of deriving advanced dynamic site equations demonstrated on inland Douglas-fir site curves. Canadian Journal of Forest Research, v. 31, n. 1, p. 165-173, 2001.

CIESZEWSKI, C.J. Comparing Fixed- and Variable-Base-Age Site Equations Having Single Versus Multiple Asymptotes. Forest Science, v. 48, n.1, p. 7-23, 2002.

CIESZEWSKI, C.J., BAILEY, L. Generalized Algebraic Difference Approach: Theory Based Derivation of Dynamic Site Equations with Polymorphism and Variable Asymptotes. Forest Science, v. 46, n. 1, p.116-126, 2000.

CIESZEWSKI, C.J., STRUB, M. Comparing properties of self-referencing models based on Nonlinear-Fixed-Effects versus Nonlinear-Mixed-Effects modeling approaches. Mathematical and Computational Forestry and Natural-Resource Sciences, v. 10, n. 2, p. 46-57, 2018.

Cieszewski, C.J., Strub, M. and Zasada, M. New dynamic site equation that fits best the Schwappach data for Scots pine (*Pinus sylvestris* L.) in Central Europe. Forest Ecology and Management, v. 243, n. 1, p. 83-93, 2007.

CLUTTER, J.L., FORTSON, J.C., PIENAAR, L.V., BRISTER, G.H., BAILEY, R.L. Timber management: A quantitative approach. John Wiley & Sons, N.Y., p.334, 1983.

CRECENTE-CAMPO, F., TOMÉ, M., SOARES, P., DIÉGUEZ-ARANDA, U. A generalized nonlinear mixed-effects height–diameter model for *Eucalyptus globulus* L. in northwestern Spain. Forest Ecology and Management, v.259, n. 5, p. 943-952, 2010.

CUDECK, R., HARRING, J.R. Analysis of nonlinear patterns of change with random coefficient models. Annu. Rev. Psychol, v.58, p. 615-637, 2007.

DAVIS, L., JOHNSON, K., BETTINGER, P. ,HOWARD, T. Forest Management: To Sustain Ecological, Economic, and Social Values, 2001.

DORADO, F.C., DIÉGUEZ-ARANDA, U., ANTA, M.B., RODRÍGUEZ, M.S.,VON GADOW, K. A generalized height–diameter model including random components for radiata pine plantations in northwestern Spain. Forest Ecology and Management, v. 229, n. 1-3, p. 202-213, 2006.

DURBIN, J., WATSON, G.S. Testing for serial correlation in least squares regression. III. Biometrika, v. 58, n. 1, p. 1-19, 1971.

GARDNER, R.A., LITTLE, K.M., ARBUTHNOT, A. Wood and fibre productivity potential of promising new eucalypt species for coastal Zululand, South Africa. Australian Forestry, v. 70, n. 1, p. 37-47, 2007.

GOELZ, J., BURK, T. Development of a well-behaved site index equation: jack pine in north central Ontario. Canadian Journal of Forest Research, v. 22, n. 6, p. 776-784, 1992.

GONZÁLEZ-GARCÍA, M., HEVIA, A., MAJADA, J., DE ANTA, R.C., BARRIO-ANTA, M. Dynamic growth and yield model including environmental factors for *Eucalyptus nitens* (Deane & Maiden) Maiden short rotation woody crops in Northwest Spain. New Forests, v. 46, n. 3, p. 387-407, 2015.

HARRING, J.R.,LIU, J. A comparison of estimation methods for nonlinear mixed-effects models under model misspecification and data sparseness: A simulation study. Journal of Modern Applied Statistical Methods, v. 15, n. 1, p. 27, 2016.

HOLT, W., REFENES, P. The Durbin-Watson test for neural regression models, Risk Measurement, Econometrics and Neural Networks. Springer, pp. 57-68, 1998.

HUANG, S. Development of compatible height and site index models for young and mature stands within an ecosystem-based management framework. Empirical and process-based models for forest tree and stand growth simulation: 61-98, 1999.

HUANG, S., MENG, S.X.,YANG, Y. Assessing the goodness of fit of forest models estimated by nonlinear mixed-model methods. Canadian Journal of Forest Research, v. 39, n. 12, p. 2418-2436, 2009.

HUANG, S., YANG, Y., MENG, S. X. Developing Forest Models from Longitudinal Data: A Case Study Assessing the Need to Account for Correlated and/or Heterogeneous Error Structures under a Nonlinear Mixed Model Framework. Journal of Forest Planning, v. 16, n. special, p. 121-131, 2011.

HUNTER, I.R., GIBSON, A.R. Predicting Pinus Radiata Site Index From Environmental Variables, v. 14, n. 1, p. 53-64, 1984.

KIVISTE, A., ÁLVAREZ GONZÁLEZ, A., ROJO ALBORECA, A., RUIZ GONZÁLEZ, A. D. Funciones de crecimiento de aplicación en el ámbito forestal. INIA, p.190, 2002.

KOIRALA, A., MONTES, C., BULLOCK, B. Modeling dominant height using stand and water balance variables for loblolly pine in the Western Gulf, US. Forest Ecology and Management, v. 479, n. 1, p. 118610, 2021.

KORF, V. A mathematical definition of stand volume growth law. Lesnická práce, v. 18, p. 337-339, 1939.

KRISNAWATI, H., WANG, Y., ADES, P.K., WILD, I.W. Dominant Height and Site Index Models for Acacia Mangium Willd. Plantation. Indonesian Journal of Forestry Research, v. 6, n. 2, p. 148-165, 2009.

LINDSTROM, M.L., BATES, D. M. Nonlinear mixed effects models for repeated measures data. Biometrics, v. 46, n. 3, p. 673-87, 1990.

LUWANDA, A.G., M WAMBI, H.G. A Nonlinear Mixed-Effects Model for multivariate longitudinal data with dropout with application to HIV disease dynamics. Journal of Agricultural, Biological, and Environmental Statistics, v. 21, n. 2, p. 277-294., 2016.

MCDILL, M.E., AMATEIS, R.L. Measuring forest site quality using the parameters of a dimensionally compatible height growth function. Forest Science, v. 38, n. 2, p. 409-429, 1992.

MONSERUD, R.A. Height Growth and Site Index Curves for Inland Douglas-fir Based on Stem Analysis Data and Forest Habitat Type. Forest Science, v. 30, n. 4, p. 943-965, 1984

NI, C., ZHANG, L. An analysis and comparison of estimation methods for self-referencing equations. Canadian Journal of Forest Research, v.37, n. 8, p. 1472-1484, 2007.

NOTHDURFT, A., KUBLIN, E., LAPPI, J. A non-linear hierarchical mixed model to describe tree height growth. European Journal of Forest Research, v. 125, n. 3, p. 281-289, 2006.

OU, G., WANG, J., XU, H., CHEN, K., ZHENG, H. ZHANG, B., SUN, X. XU, T., XIAO, Y. Incorporating topographic factors in nonlinear mixed-effects models for aboveground biomass of natural Simao pine in Yunnan, China. Journal of Forestry Research, v. 27, n. 1, p. 119-131, 2016.

PARRESOL, B.R., VISSAGE, J.S. White pine site index for the southern forest survey/; Bernard R. Parresol, John S. Vissage. Research paper SRS; 10, 1998.

RICHARDS, F. A flexible growth function for empirical use. Journal of experimental Botany, v. 10, n. 2, p. 290-301, 1959.

ROJO ALBORECA, A., CABANILLAS SALDAÑA, A.M., BARRIO ANTA, M., NOTIVOL PAÍNO, E., GORGOSO VARELA, J.J. Site index curves for natural Aleppo pine forests in the central Ebro valley (Spain), v. 23, n. 1, p. 143-159, 2017.

SÁNCHEZ-GONZÁLEZ, M., TOMÉ, M., MONTERO, G. Modelling height and diameter growth of dominant cork oak trees in Spain. Annals of Forest Science, v. 62, n. 7, p. 633-643, 2005.

SAS Institute Inc, 2013. SAS® 9.4 Statements: Reference: Cary, NC: SAS Institute Inc.

SCOLFORO, H.F., SCOLFORO, J.R.S., STAPE, J.L., MCTAGUE, J.P., BURKHART, H., MCCARTER, J., NETO, F.C., LOOS, R.A., SARTORIO, R.C. Incorporating rainfall data to better plan eucalyptus clones deployment in eastern Brazil. Forest Ecology and Management, v.391, p.145-153, 2017.

SHARMA, M., PARTON, J. Height–diameter equations for boreal tree species in Ontario using a mixed-effects modeling approach. Forest Ecology and Management, v. 249, n. 3, p. 187-198, 2007.

SHARMA, M., SUBEDI, N., TER-MIKAELIAN, M., PARTON, J. Modeling climatic effects on stand height/site index of plantation-grown jack pine and black spruce trees. Forest Science, v.61, n. 1, p. 25-34, 2015.

SHARMA, R.P., VACEK, Z., VACEK, S. Generalized nonlinear mixed-effects individual tree crown ratio models for Norway spruce and European beech. Forests, v. 9, n. 9, p. 555, 2018.

SKOVSGAARD, J.P., VANCLAY, J. Forest site productivity: A review of the evolution of dendrometric concepts for even-aged stands. Forestry, v. 81, n. 1, p. 13-31, 2008.

SNOWDON, P., BENSON, M.L., WOOLLONS, R.C. Incorporation of climatic indices into models of growth of Pinus radiata in a spacing experiment. New Forests, v. 16, n. 2, p. 101-123, 1998.

SOCHA, J., TYMIŃSKA-CZABAŃSKA, L., BRONISZ, K., ZIĘBA, S., HAWRYŁO, P. Regional height growth models for Scots pine in Poland. Scientific Reports, v. 11, p. 1-14, 2021.

SPRENGEL, L., SPIECKER, H., SHUIRONG, W. Two subject specific modelling approaches to construct base-age invariant polymorphic site index curves with varying asymptotes. Silva Fennica, v. 56, n. 1, 2022.

STANGER, T., GALLOWAY, G., RETIEF, E.C. Final results from a trial to test the effect of plot size on Eucalyptus hybrid clonal ranking in coastal Zululand, South Africa. Southern Forests: a Journal of Forest Science, v. 73, n. 3-4, p. 131-135, 2011.

TAHAR, S., PALAHI, M., GARCHI, S., BONET, J.A., AMMARI, Y., PIQUE, M. Modeling dominant height growth in planted *Pinus pinea* stands in Northwest of Tunisia. International Journal of Forestry Research, 2012.

TEMESGEN, H., MONLEON, V., HANN, D. Analysis and comparison of nonlinear tree height prediction strategies for Douglas-fir forests. Canadian Journal of Forest Research, v. 38, n. 3, p. 553-565, 2008.

WANG, Y., LEMAY, V.M., BAKER, T.G. Modelling and prediction of dominant height and site index of *Eucalyptus globulus* plantations using a nonlinear mixed-effects model approach. Canadian Journal of Forest Research, v.37, n. 8, p. 1390-1403, 2007.

WEISKITTEL, A.R., HANN, D.W., KERSHAW, J.A., JR., VANCLAY, J.K. Forest growth and yield modeling. Wiley-Blackwell, Chicester, UK. 2011.

WOOLLONS, R.C., SNOWDON, P., MITCHELL, N.D. Augmenting empirical stand projection equations with edaphic and climatic variables. Forest Ecology and Management, v. 98, n.3, p. 267-275, 1997.

YILMAZ, M., USTA, A., ÖZTÜRK, İ. Relationships between site indices and ecological factors for black alder stands in the Turkish Eastern black sea region. Fresenius Environmental Bulletin, v.24, p. 1507-1515, 2015.