Forest Ecology and Management 479 (2021) 118610

Contents lists available at ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco

Modeling dominant height using stand and water balance variables for loblolly pine in the Western Gulf, US

Anil Koirala*, Cristian R. Montes, Bronson P. Bullock

Warnell School of Forestry and Natural Resources, University of Georgia, 180 E Green Street, Athens, GA, USA

ARTICLE INFO

Keywords:
Dominant height
Evapotranspiration
Site index
Water deficit

ABSTRACT

Site-index models based on dominant height and age are commonly accepted measures of site productivity in forest plantations. Therefore, reliable estimates of dominant height over time are critical, as they provide the foundations for site productivity estimates used in growth and yield models. Forest site productivity depends on the combination of many environmental factors, whereas tree height is only a general proxy indicator of these factors interacting over time, and as such, subject to large errors. Hence, selection of appropriate environmental variables along with field-based data is expected to result in more consistent estimates. Although previous studies have incorporated climatic variables into site-index models, there is still a lack of understanding among the forest research community with respect to the selection of suitable climatic variables. Models including climatic water balance components such as water-deficit and excess-water are scarce. These components are especially important to address the influence of drought on plant growth. The main objective of this study is to evaluate the change in dominant height estimates of forest stands with respect to changes in water balance components. High-resolution actual evapotranspiration (AET) and potential evapotranspiration (PET) were combined to determine water excess/deficit of sites. These were incorporated into a loblolly pine polymorphic site index model using a generalized algebraic difference approach and maximum likelihood calibration methods that accounted for long-term uncertainty in the results. Our model further improves the precision in dominant height estimates for loblolly pine in the southeast US.

1. Introduction

Since the mid-twentieth century, the US forest sector has witnessed extraordinary increase in the pine plantation acreage, especially in the South. In the 1950s, only about 900,000 ha of pine plantation existed in the region, it has dramatically increased more than twenty times in just 50 years (Fox et al., 2004). Loblolly pine (*Pinus taeda L.*) is one of the most important planted tree species in the South and covers more than 14 million hectares with a projected area of about 22 million hectares by 2040 (Restrepo et al., 2019; Sabatia and Burkhart, 2014). This increase resulted in an increase in growth and yield studies conducted for loblolly pine in the region. The surge in growth modeling studies has produced some commendable results regarding plantation, management and silvicultural requirements of this popular tree species (Borders et al., 2014, 2004; Harrison and Borders, 1996; Zhao et al., 2019).

The height and diameter growth of a tree depend on the overall productivity of the site, which, in turn, depends on several factors such as climate, water, soil, topography and others. In forestry, the common and widely accepted indirect method of assessing site productivity is site index (SI), which is based on the relationship between tree height and their age at a given site. Although several geocentric site productivity indicators based on climate, soil, or topography are extensively used in agricultural sector, it is not always considered practical in forestry because of high cost in data collection and higher variability among forest sites. Therefore, the geocentric approach based on one or more easily measured tree or stand variables gained popularity in forestry (Skovsgaard and Vanclay, 2013). In the US context, site index is generally calculated by averaging the total heights of dominant and codominant trees in well-stocked, even-aged stands at a specified index (base) age (e.g. 25 or 50 years) (Graves, 1907). The universally accepted definition of dominant (top) height for determining site index does not exist. For example, the number of dominant trees are not important in the US definition, but in Europe, the 100 thickest trees per hectare or the 20% thickest trees in the stand are generally used (Sharma et al., 2002).

A dominant height curve developed from the site index equation gives the average dominant height of the stand over the ages of data

* Corresponding author.
E-mail addresses: anilk@uga.edu (A. Koirala), crmontes@uga.edu (C.R. Montes), bronsonbullock@uga.edu (B.P. Bullock).

https://doi.org/10.1016/j.foreco.2020.118610
Received 1 June 2020; Received in revised form 8 September 2020; Accepted 9 September 2020
0378-1127/

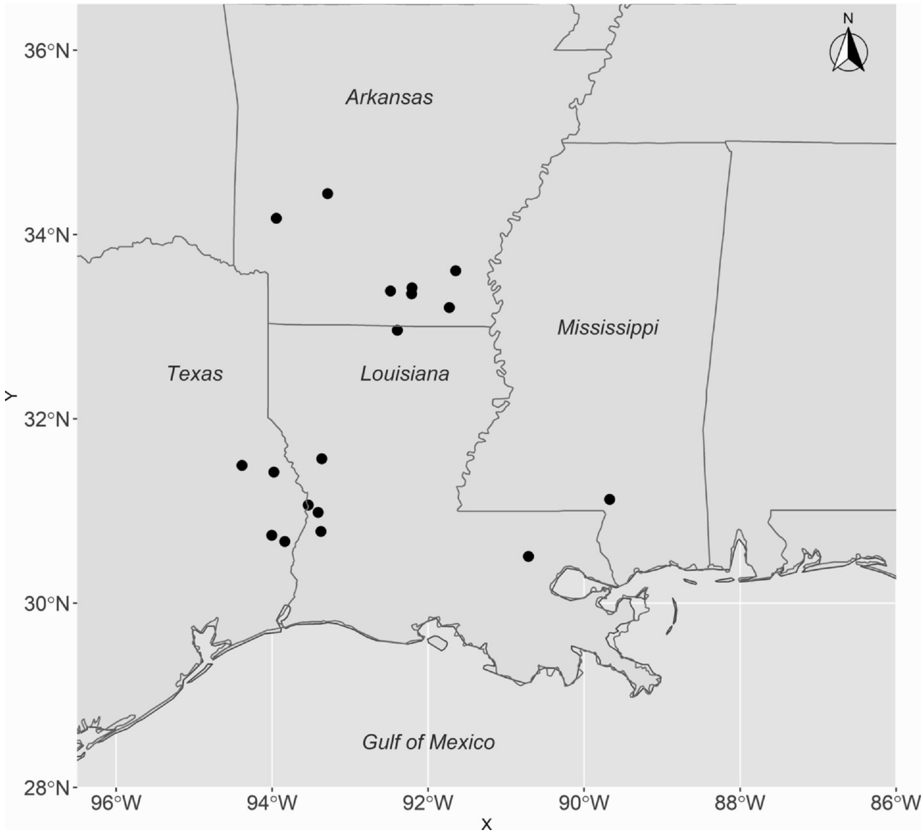

Fig. 1. Map of study area showing the 18 Western Gulf Culture Density study installations with black dots across Arkansas, Louisiana, Mississippi and Texas.

utilized. This also assumes an average climatic condition during the growth period. Curves developed for dominant height equations can be either polymorphic or anamorphic and are commonly derived from three different techniques: guide curve, parameter prediction and differential equation methods (Clutter et al., 1983). Bailey and Clutter (1974) proposed an algebraic difference approach (ADA) to derive dominant height equations based on the differential equation principle. This type of equation had only one site-specific parameter and the resulting dominant height curves were anamorphic in nature. Cieszewski and Bailey (2000) expanded the ADA and made it more flexible by allowing more than one site-specific parameter in the model, keeping the equation as base age invariant. Base-age invariant means the models are not sensitive to the base-age selected and any formulation of models lead to the same parameter estimates independent of the base-age selected. Height curves produced this way can be either anamorphic or polymorphic in nature. Since its development, this approach, also known as the generalized algebraic difference approach (GADA), has been extensively used to model dominant-height and age of forest stands (Carvalho and Parresol, 2005; Diéguez-Aranda et al., 2006; Özçelik et al., 2019; Weiskittel et al., 2009).

Researchers have previously attempted to incorporate climatic and biophysical variables into site index equations (Monserud et al., 2006; Sabatia and Burkhart, 2014; Wang et al., 2004). However, such models have hardly been used by industrial foresters or implemented in growth and yield simulators commonly used in the South East US (Peay, 2019). A possible reason for that is a practice of incorporating all possible climatic variables in a model, which will not only compromise in model bias but also increase the model complexity. On the other hand, a more parsimonious model will grant unbiasedness at the cost of generality.

Loblolly pine growth is affected by several environmental factors and water availability is one of the important factors among them (Lambers et al., 2008; Landsberg et al., 2003). Continuous availability of water to the plants sustains the photosynthesis process, which plays a vital role in energy production, plant growth and wood formation. During droughts, because of stress, plants close stomata to offset water loss from water deficit (Gonzalez-Benecke et al., 2010). This causes slower growth and impacts important plant metabolisms. The climatological water balance in hydrology consists of an interesting relationship between soil water holding capacity, precipitation, actual evapotranspiration (AET) and potential evapotranspiration (PET) (Thornthwaite et al., 1957). Water deficit (WD) occurs when the evapotranspiration from the ground surface and plants is higher than precipitation in the given period, it is calculated as the difference between PET and AET. Excess water (EW) occurs when the amount of precipitation is higher than that of evapotranspiration; it is the surplus water that was not absorbed by the soil when soil moisture is at field capacity. The available water (AW) is the water that is available to use by the plants for their photosynthetic and other activities. Several important climatic and biophysiological factors such as precipitation, solar radiation, soil properties, plant water use capacity and others are used to calculate these water balance variables. Therefore, we believe that incorporating water balance variables in the dominant height/site index models can provide a better prediction of site productivity. The main objectives of this study were to evaluate the change in dominant height estimates of forest stands with respect to change in water balance components and to develop base-age invariant polymorphic site index and dominant height models that can directly incorporate these climatic variables and improve precision in dominant height estimates in the southeast US.

2. Materials and methods

2.1. Study area

Data utilized in this study came from the Western Gulf Culture Density study (WGCDS) managed by the Plantation Management

Table 1
Western Gulf Culture Density study soil groups based on drainage class and depth to subsurface restrictive layer.

WGCDS Soil Group	Drainage Class	Depth to Subsurface Restrictive Layer (m)
A	Poorly to somewhat poorly	< 0.508
B	Poorly to somewhat poorly	< 0.508
C	Moderately well to well	< 0.508
D	Moderately well to well	< 0.508

Research Cooperative at the University of Georgia. Installations examining planting density, cultural treatment intensity, and thinning of loblolly pine were established at 18 locations in Arkansas, Louisiana, Texas, and Mississippi during the 2000/2001, 2001/2002, and 2002/2003 dormant seasons (Fig. 1). Each installation consisted 16 plots, of which six plots included thinning as a treatment. For this study, only data from the 10 unthinned plots per installation were taken into consideration. These 10 plots represent a unique combination of five planting densities (494, 1112, 1730, 2347, and 2965 trees per hectare TPH) × two silvicultural intensities (intensive and maximum). Treatment plot size ranges from 0.09 to 0.22 ha.

Data from 176 plots out of 180 potential unthinned plots were used in this research; four plots were excluded due to excessive mortality. The study represented major soil types prevalent in the Western Gulf region of the United States; these soil groups were based on drainage class and depth to a subsurface restrictive layer (Table 1). All sites had previously been forested, thus sites were mechanically site prepared according to the soil group and planted with loblolly pine. Each site was planted with the best open-pollinated family for that location as determined by the industrial cooperators. Each planting spot was double planted to ensure good survival, with one of the seedlings cut at ground line in the fall following planting as needed to leave one seedling per planting spot. Plots were fertilized according to the two levels of cultural intensities, intensive and maximum (Kane et al., 2011). In addition, insecticide was applied for tip moth (*Rhyacionia frustrana*) control in the first and second growing seasons.

2.2. Field measurements

All trees were measured for diameter at breast height (dbh) at each measurement while total height (HT) was measured for all trees only up to age 6. After that, total heights were measured on all trees for the plots with 494 TPH, every other tree for the 1112 TPH and every third tree for the higher planting densities. After the 8th growing season, height to the base of the live crown was measured on all trees that were measured for total height. Only field measured trees were used in dominant height modeling in this research. Following standard Plantation Management Research Cooperative's guidelines, the dominant and co-dominant trees were classified as live, non-defective trees, with dbh greater than average dbh of all trees on the plot at each age. This practice has also been adopted in other studies in the region (Gallagher et al., 2019; Sabatia and Burkhart, 2014; Sharma et al., 2002; Wang et al., 2020). Individual tree, outside bark cubic meter volume was calculated by modifying Pienaar et al. (1987) cubic feet volume equation. The sum of all individual tree volumes in a plot was then divided by area of the plot to get per hectare volume in the stand. The age range of the loblolly pine stands for this study was from 2 to 18 years. More details on all field measurement is presented in Gallagher et al. (2019).

2.3. Water balance data

The DAYMET climatic information from Oak Ridge National Laboratory was used to obtain monthly minimum and maximum temperature, precipitation and solar radiation data. This data consists of a 1 km^2 grids with extrapolation of daily weather data from various weather stations in the US. The data was processed to obtain average annual values for 2000 to 2018. The extent of all raster datasets was fixed with a common projection system. After that, these data were masked for our study area of the Western Gulf region. The potential evapotranspiration was then calculated using Hargreaves and Samani (1982) temperature-based equation as it is considered as a one of standard PET calculation methods:

$$E_{TC} = 0.0023 * R_A * (T + 17.8) * T_R^{0.50}$$

where

R_A = mean radiation (millimeters/day), which is a function of the latitude
T_R = temperature difference = mean daily maximum temperature – mean daily minimum temperature (Celsius)
T = mean air temperature (Celsius)

When the amount of precipitation was lower than the evapotranspiration, the water deficit value was calculated by subtracting precipitation by potential evapotranspiration. Likewise, when the amount of precipitation was higher than the potential evapotranspiration, the excess water value was calculated by subtracting potential evapotranspiration from precipitation.

$$WD = \sum_{i=1}^{n} [PET_i > R_i](PET_i - R_i)$$

$$EW = \sum_{i=1}^{n} [PET_i < R_i](R_i - PET_i)$$

where, WD and EW are already explained above, PET is the potential evapotranspiration and R is the rainfall or precipitation. The raster layers of final water balance dataset which were incorporated in the dominant height models are presented in Fig. 2.

2.4. Modeling approach

Four commonly used forms of dominant height equations with three parameters were used as base models in this study: the Schumacher function (Schumacher, 1939), Chapman-Richards function (Chapman, 1961; Richards, 1959), Lundqvist-Korf function (Lundqvist, 1957) and Hossfeld IV function (Cieszewski, 2003), respectively. The Schumacher function is widely used in growth and yield studies throughout the world. Several authors have successfully modified this function into ADA and GADA type site index models. The Chapman-Richards function, also known as the Bertalanffy-Richards growth model, is another popular exponential type function, which is frequently used in site-index modeling (Chapman, 1961; Diéguez-Aranda et al., 2006; Richards, 1959). The Lundqvist-Korf function is a generalization of Schumacher function and is considered very suitable for height modeling and preferred by foresters from central European countries (Lundqvist, 1957). The Hossfeld IV is another commonly used site index model, which has a fractional type function (Cieszewski, 2003). All four models used in this study had three parameters, among which two parameters were set as site-specific parameters during GADA formulation. These four base growth functions for dominant height estimation are presented as models 1 to 4 and labeled *M1*, *M2*, *M3* and *M4*,

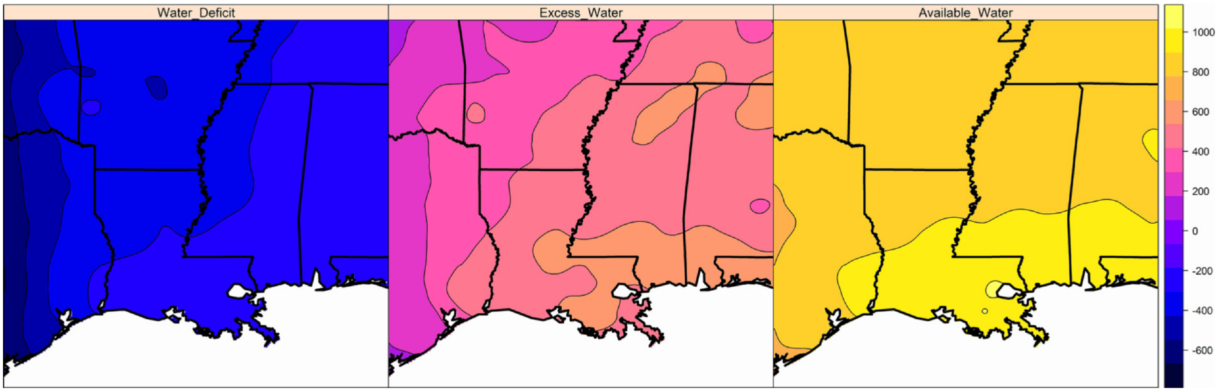

Fig. 2. Map showing water deficit (WD), excess water (EW) and available water (AW) in millimeters for Western Gulf region using 1 km^2 grid.

respectively.

$$\text{Schumacher function:}\quad HD = \exp(a + bA^{c}) \tag{M1}$$

$$\text{Chapman-Richards function:}\quad HD = a\{1 - \exp(bA)\}^{c} \tag{M2}$$

$$\text{Lundqvist-Korf function:}\quad HD = a\exp\left(\frac{b}{A^{c}}\right) \tag{M3}$$

$$\text{Hossfeld IV function:}\quad HD = \frac{a}{1 + bA^{-c}} \tag{M4}$$

where *HD* is the dominant height of the even-aged forest stand at age ***A***, and *a, b* and *c* are the parameters to be estimated.

2.4.1. Dominant height projection

The base age invariant site index models were first conceptualized by Bailey and Clutter (1974), which was later termed as the algebraic difference approach or ADA models. Cieszewski and Bailey (2000) extended the ADA model into a more generalized form by allowing more than one site specific parameter to the base model, making it more dynamic and termed GADA. In GADA models, any number of parameters can be considered as site-specific parameters with the introduction of an unobserved site index usually referred to as unobservable site variable χ. This variable χ is presumed as a function of the factors associated with forest sites and affects the tree height growth (Krumland and Eng, 2005). During the initial phase of modeling, the exact value and form of χ is unknown; it is only used during the later stages of model development. Therefore, its value is predicted simultaneously with the global parameters. The GADA formulations of our four base models along with the site-specific parameters are presented in Table 2. All four models had two site-specific parameters in order to account for the variability among the sites.

2.4.2. Model fitting and validation

The five-fold cross validation approach of model fitting and validation was used in this study. This approach involves randomly dividing the set of observations into *K* groups or folds (5 groups in our case), of approximately equal size. The first group of the data is considered as a validation data set, and the remaining folds i.e. (*K-1*) are used for model fitting. The parameter estimates for the dominant height models were obtained from the fit data, and later those estimates were used to predict the dominant height growth for the validation data set.

The four site index models were compared based on five model fitting and validation statistics, Akaike information criterion (AIC) and root mean square error (RMSE) as fitting statistics and mean difference (MD) between observed and predicted dominant heights, mean absolute difference (MAD), and fit index (FI) as five-fold cross validation statistics.

$$AIC = -2loglik + 2k \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}} \tag{2}$$

$$MD = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)}{n} \tag{3}$$

$$MAD = \frac{\sum_{i=1}^{n}|Y_i - \hat{Y}_i|}{n} \tag{4}$$

$$FI = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2} \tag{5}$$

where, k is the number of parameters in the model, Y_i is the observed dominant height, $\hat{Y}_i$ is the predicted dominant height and $\bar{Y}_i$ is the mean value of Y_i.

2.4.3. Incorporation of water balance variables in the model

In order to incorporate relevant climatic water balance variables in our best model from GADA procedure, the LASSO (least absolute shrinkage and selection operator) regression approach was used. LASSO is a type of penalized regression (Tibshirani, 2011, 1996), which has gained interests in recent years among statistics and machine learning

Table 2
Table showing the formulation of GADA models for all four base models. Note that each model has two site-specific parameters.

Base Models	Site-specific Parameters	GADA Model Forms
Schumacher: $HD = \exp(a + bA^{c})$	$a = b_1 + \chi$ $b = b_2 + b_3\chi$	$HD = exp(b_1 + \chi_0 + (b_2 + b_3\chi_0)\ A^{b_4})$
Chapman-Richards: $HD = a\{1 - \exp(bA)\}^{c}$	$a = \exp(\chi)$ $c = b_2 + b_3/\chi$	$HD = exp(\chi_0) * (1 - exp(b_1A)^{(b_2 + b_3/\chi_0)}$
Korf: $HD = a\exp\left(\frac{b}{A^{c}}\right)$	$a = \exp(\chi)$ $b = b_1\chi$	$HD = exp(\chi_0)exp\left(\frac{b_1\chi_0}{A^{b_2}}\right)$
Hossfeld IV: $HD = \frac{a}{1 + bA^{-c}}$	$a = b_1 + \chi$ $b = b_2/\chi$	$HD = \frac{b_1 + \chi_0}{1 + b_2\chi_0^{-1}A^{b_3}}$

communities in the wake of the high dimensional nature of current data sets (Zhao and Yu, 2006). In the context of linear and generalized linear models, it is currently one of the most popular methods of model selection (Nardi and Rinaldo, 2011). Penalized regression methods attempt to achieve two fundamental goals of regression, predicting accurately and selecting the relevant variables, simultaneously. LASSO has two major advantages over classical variable selection methods such as subset selection, and forward or backward selection. First, it performs continuous selection of the variables, which makes it more stable than other procedures. Second, it is more computationally feasible for larger data sets like those in this study.

The three major variables associated with water balance components excess water, water deficit and available water, along with tree age were used as response variables in the LASSO procedures. As four non-linear models were used in this study, we needed to linearize the best GADA model to run LASSO and incorporate climatic variables. The "glmnet" R language package was used for the LASSO computation (Friedman et al., 2010).

3. Results

This paper utilized data from Western Gulf culture density study managed by the University of Georgia. A total of 176 intensively managed unthinned loblolly pine stands were considered for our analysis. Table 3 contains the summary statistics for stand level variables for both fitting and validation datasets that were randomly divided by the five-fold cross validation process. The fitting datasets had an average dominant height of 9.5 m, stand density of 1155 TPH, basal area of 13.6 m^2 per hectare, and volume of 52.6 m^3 per hectare. Likewise, the validation dataset had an average dominant height of 9.7 m dominant height, stand density of 1049 trees per hectare, basal area of 13.2 m^2 per hectare, and volume of 52.5 m^3. The observed dominant heights at each measurement age for both datasets are presented in Fig. 3.

3.1. Model performance and selection

Four dominant height models to measure the site productivity of loblolly pine in the Western Gulf region were fitted using base-age invariant dynamic equations. Parameterization of all four models were carried out employing the maximum likelihood estimation framework to find optimal values for each parameter. The parameter estimates and their standard errors for the global parameters in all four models along with the fit and validation statistics are presented in Table 4. Parameters for each model were estimated from fit datasets while the validation statistics were calculated from validation datasets. In terms of fit statistics, the RMSE values for all models were lower around 1 m; the lowest value was for model M3, the Lundqvist-Korf model. The AIC value of M2 Chapman-Richards model was lowest among all four. The lower value of AIC in Chapman-Richards model might be the result of higher number of parameters in the model than the Lundqvist-Korf.

Table 3
Summary statistics of stand characteristics for fitting and validation datasets from all unthinned plots of the WGCD study.

Dataset	N	Variable	Mean	Minimum	Maximum	SD
Fitting	1342	Stand Age (years)	7.8	2	18	4.5
		Dominant height (m)	9.5	1.7	25.8	5.4
		Volume (m^3/ha)	52.6	0.02	530.8	59.5
		Stand density (trees/ha)	1155	74	2990	763
		Basal Area (m^2/ha)	13.6	0.003	78	11.04
Validation	336	Stand Age (years)	8.02	2	18	4.4
		Dominant height (m)	9.7	1.6	23.9	5.2
		Volume (m^3/ha)	52.5	0.01	424.6	63.5
		Stand density (trees/ha)	1049	124	2965	679
		Basal Area (ft^2/ha)	13.2	0.02	65.2	11.2

Regarding validation statistics, the mean difference between observed and predicted values was lowest in model M3, the Lundqvist-Korf model. Similarly, the mean absolute difference was lowest for M3. All models were able to explain more than 90 percent of the variation in height predictions. Among them, the highest fit index value of 0.981 (~98%) was for model M3, Lundqvist-Korf model. However, models M1 and M4 also performed well. The Chapman-Richards model had the poorest fit with highest RMSE, mean differences and the lowest fit index values. Models M1, and M4 performed consistently well in validation statistics. However, model M1 had higher AIC values than other models. Likewise, model M4 had comparatively larger error and bias (mean difference) than M3. Therefore, based on the fit and validation statistics, model M3 was selected to be the best model. This model also had the lowest number of parameters among evaluated models. Despite its selection as the best model based on validation statistics, model M3 was further compared with other three models in terms of the distribution of residuals.

Evaluation of residuals for all models showed that models M1 and M4 exhibited some degree of bias and heteroscedasticity (Fig. 4). The Chapman-Richards M2 model clearly demonstrated higher levels of bias and non-constant variance of the residuals with increasing height. The residuals of model M3 were symmetrically distributed and clustered around zero. Model M3 outperformed other models based on residual plot analysis.

After examination of residuals, the prediction capabilities of all four models were evaluated. Guide curves for site index value of 25 m at base age 25 years and planting density of 1730 trees per hectare with intensive culture were constructed to examine extrapolative behavior of all four models through age 35 (Fig. 5). Models M1 and M4 showed similar prediction of dominant heights up to base age 25 years. After that, M4 underpredicted dominant height compared to other models. The Chapman-Richards model (M2) performed well in terms of prediction through age 25. However, model M3 visually performed better than other models in terms of prediction capability.

The predicted dominant height curves for three site index values (SI 20, 25, 30 m) at base age 25 years were constructed for all models (Fig. 6). These curves were overlaid on the entire observed dominant height datasets. The site index plots revealed that model M3 adequately represented observed values while predicting dominant heights with less bias. Models M1 and M4 overpredicted dominant heights at younger ages on sites with higher site indices. Models M1, M2 and M4 were compared against model M3 as the model for comparison based on M3 having the best fit statistics (Fig. 7). The comparison plot for a site index value of 25 m at base age 25 years showed that prediction line from model M2 was closer to M3 through base age. Based on validation statistics as well as graphical analysis, model M3, the Lundqvist-Korf dominant height function was selected as the best model for the given dataset in the Western Gulf region of the United States. The final model proposed for dominant height prediction and site index based on entire dataset is given as:

$$HD = \exp(\chi_0)\exp\left(\frac{-1.031 * \chi_0}{Age^{0.323}}\right)$$

where $\chi_0 = \frac{\ln(H_0)}{1 + \frac{-1.031}{t_0^{0.323}}}$

The model including the base age t_0 (i.e. 25 year in our case) and site index H_0 at base age is give as:

$$HD(t, t_0, H_0) = \exp\left(\frac{ln(H_0)}{1 + \frac{-1.031}{t_0^{0.323}}}\right) * \exp\left(\frac{-1.031 * ln(H_0)}{1 + \frac{-1.031}{t_0^{0.323}}}\right)$$

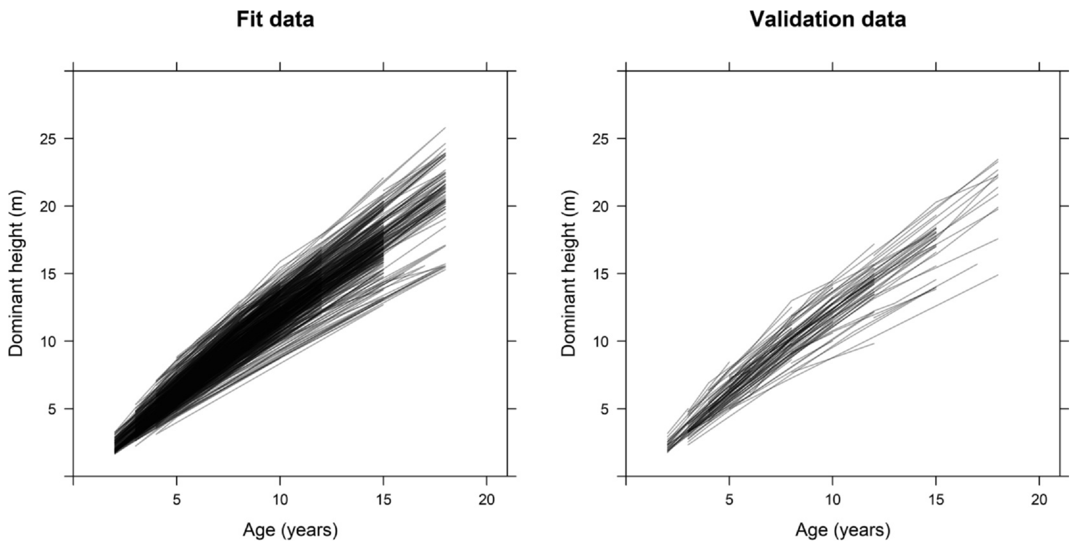

Fig. 3. Dominant height growth trajectories for fitting and validation datasets. Each line represents a plot from the field measurements.

3.2. Model with climatic variables

As model M3, the Lundqvist-Korf model, was regarded as the best model for dominant height prediction for our dataset, we incorporated climatic water balance variables in the model. The LASSO variable selection procedure was utilized to select among three water balance variables, water deficit (WD), excess water (EW) and available water (AW). This procedure was developed specifically for linear models, therefore, the Lundqvist-Korf model, in its non-linear form, needed to be linearized. The linearization process of the Lundqvist-Korf model in order to incorporate climatic variables follows:

$H = \beta_0 exp\left(\frac{\beta_1}{A^m}\right)$,where H is dominant height and A is age taking natural logarithm (ln) on both sides, inspired by Bailey and Clutter (1974),

$$\ln(H) = \ln(\beta_0) + \beta_1\left(\frac{1}{A}\right)^m$$

taking derivative with respect to age,

$$\frac{1}{H}\frac{dH}{dA} = -\beta_1 m\left(\frac{1}{A}\right)^{m+1}$$

taking natural logarithm (ln) on both sides again:

$$\ln\left(\frac{1}{H}\frac{dH}{dA}\right) = \ln(-\beta_1 m) + (m+1)\ln\left(\frac{1}{A}\right)$$

The derivative term $\frac{dH}{dA}$ can be approximated by using the mid points of the difference equation, i.e. $dH \approx H_2 - H_1$; $H = \frac{H_2 + H_1}{2}$, where H_2 and H_1 are heights at age 2 (A_2) and age 1 (A_1); and $dA \approx A_2 - A_1$; $A = \frac{A_2 + A_1}{2}$.

Now, replacing these values into the previous equation:

$$\ln\left[\frac{2}{H_2 + H_1}\left(\frac{H_2 - H_1}{A_2 - A_1}\right)\right] = \gamma_0 + \gamma_1 \ln\left(\frac{2}{A_2 + A_1}\right)$$

where $\gamma_0 = \ln(-\beta_1 m)$ and $\gamma_1 = m + 1$

The value of m can be easily calculated using the presented coefficient (γ_1). Hence, the final linearized version of Lundqvist-Korf equation, including water balance variables is given as:

$$\ln(H) = b_0 + b_1\ln(WD) + b_2\ln(EW) + b_3\ln(AW) + b_4\left(\frac{1}{A'}\right) \quad \text{(M5)}$$

where WD = water deficit, EW = excess water, AW = available water, and $A' = Age^m$

The linearized Lundqvist-Korf model (M5) was then evaluated for variable selection with the LASSO procedure. Water deficit and available water were selected as defining predictors of site productivity in the Western Gulf region by this process. Water deficit negatively impacted site productivity with a higher impact on the further western portion of the region near east Texas. Available water, on the other hand, had a positive impact on productivity. The dominant height projection from the new equation (M5) had lower AIC and RMSE values

Table 4
Parameter estimates from the entire dataset, and their corresponding goodness-of-fit statistics from the fivefold evaluation for GADA models.

Model	Parameter	Estimate	Std. Error	Fit statistics		Five-fold validation statistics		
				RMSE	AIC	MD	MAD	FI
M1	b0	−3.809	0.006	0.902	−1607.7	0.051	0.506	0.977
	b1	−1194.9	0.043					
	b2	142.2	0.005					
	b3	−0.389	0.002					
M2	b0	−0.003	0.089	1.317	−2346.4	−0.309	0.847	0.935
	b1	−1.41	0.023					
	b2	15.176	1.001					
M3	b0	−1.031	0.033	0.783	−2290.9	0.025	0.497	0.981
	b1	0.323	0.117					
M4	b0	38.099	2.305	0.818	−1243.4	−0.047	0.578	0.979
	b1	19.167	1.914					
	b2	−1.295	0.077					

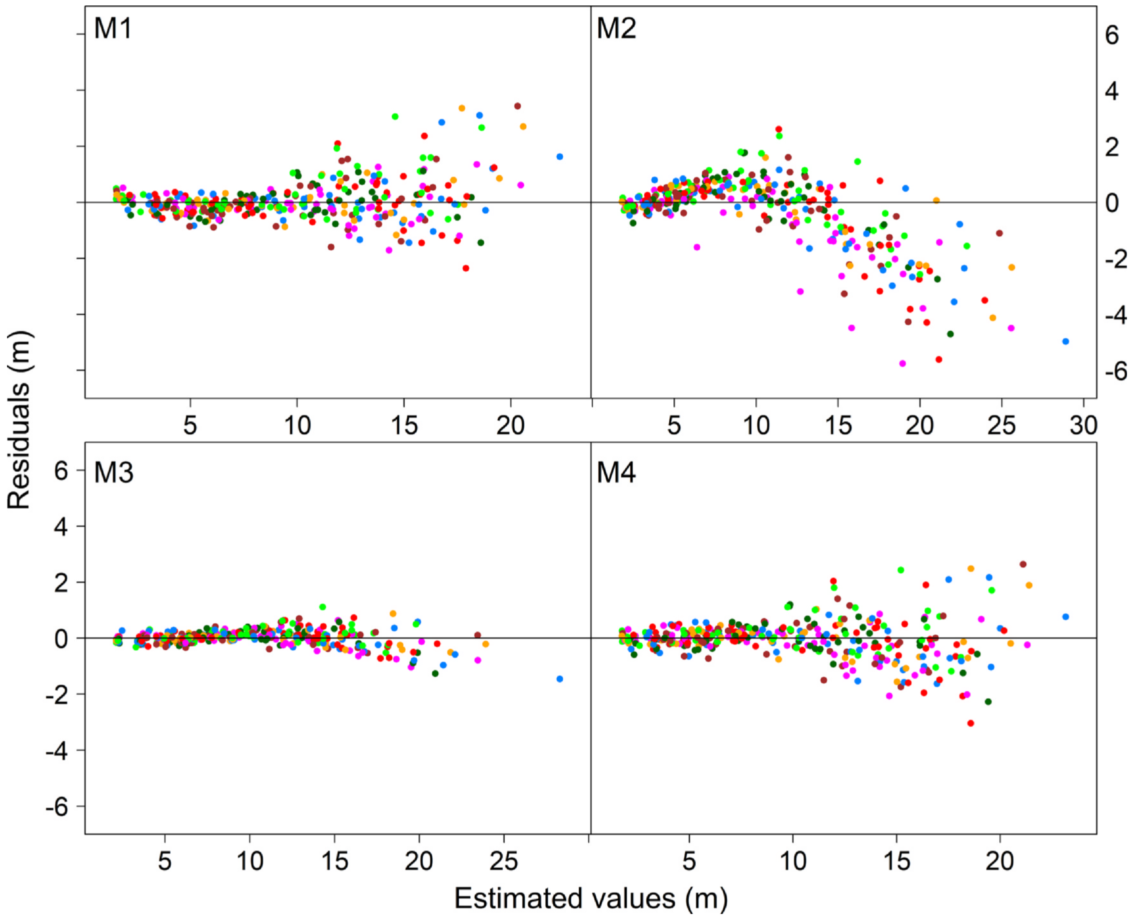

Fig. 4. Plot of residuals against predicted dominant height (m) for four site index GADA models. The color dots indicate each measurement plot from the study. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

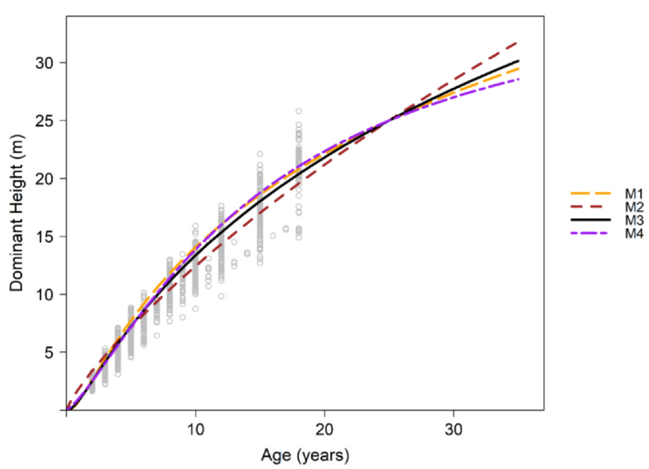

Fig. 5. Guide curves constructed from all four dominant height models overlaid with the observed dominant heights for the entire WGCD loblolly pine plantation dataset.

for the fitting dataset compared to the best GADA model (i.e. M3). Model M5 had a slightly higher MD value than M3, but it had a lower MAD value. The fit index value for M5 was very similar to that of M3 (Table 5). The site index map created from spatial regression using model M5 showed that the Western Gulf region had site index (or productivity) value up to 22 m (Fig. 8). In other words, some parts of the Western Gulf region, especially on the eastern side near the Mississippi river had the highest predicted loblolly pine productivity. The western side of the region had relatively lower productivity for loblolly pine height growth. Overall, based on site index map (Fig. 8) and the water deficit map (Fig. 3), we concluded that the areas with higher water deficit had lower site index values for loblolly pine.

4. Discussion

The main objective of this study was to develop a suitable site index model to predict dominant height growth (or site productivity) of loblolly pine plantations in the Western Gulf region by incorporating water balance variables into a traditional site index modeling framework to aid forest land managers. Site index is considered as a reliable proxy for the site productivity and many silvicultural and economic decisions pertaining to plantation forestry are based on site index values. Site index determination is also regarded as one of the first steps in whole stand growth and yield modeling process. Loblolly pine is the most important planted tree species in the US South in terms of planting acreage, growing stock, harvesting intensity and economic impact (Oswalt et al., 2019). Researchers in the past have used various site index model forms to predict dominant height of loblolly pine stands in the US South based solely on tree measurement data (Amateis and Burkhart, 1985; Cao, 1993; Cao et al., 1997; Diéguez-Aranda et al., 2006; McDill and Amateis, 1992). Recently, site index models combining both tree measurement data and biophysical information have also been applied to loblolly pine plantations (Sabatia and Burkhart, 2014; Subedi and Fox, 2016). This study modeled site index in two separate modeling steps. In the first step, similar to traditional site index modeling, only tree measurements data were fitted on four base-age invariant dynamic models. In these cases, we assumed that the

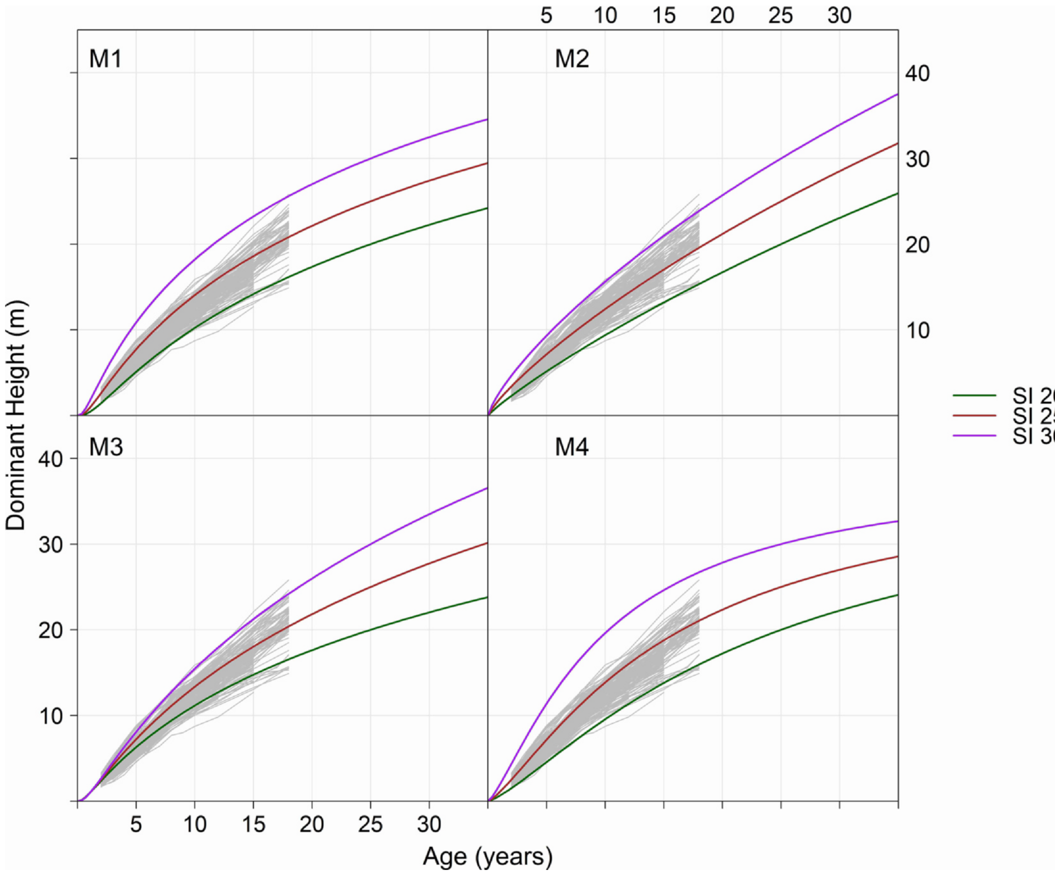

Fig. 6. Dominant height curves constructed from four models for three commonly used site index values (20 m, 25 m and 30 m) overlaid with the observed dominant height values shown in gray on the background.

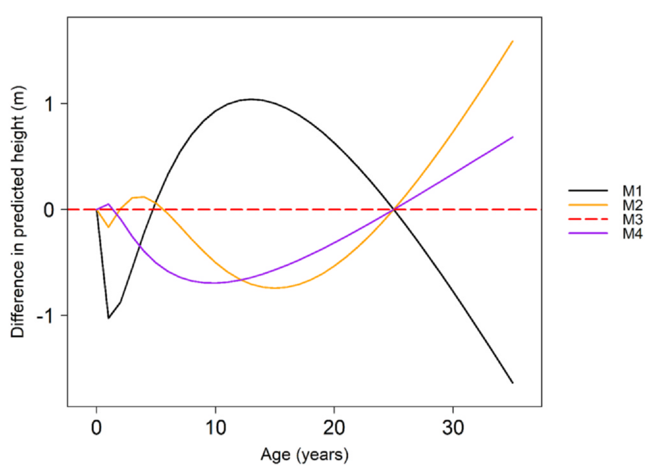

Fig. 7. Comparison between three models M1, M2 and M4 against model M3 as the model for comparison for a site index value of 25 m at base-age 25 years.

Table 5
Fitting and validation statistics for two forms of the Lundqvist-Korf model (M3 and M5).

Model	RMSE	AIC	MD	MAD	FI
M3	0.783	−2290.9	0.025	0.497	0.981
M5	0.615	−2457.2	0.141	0.288	0.980

variability related to sites could be addressed by unobservable site-specific parameters in the models. The second step involved incorporating climatic variables in the best model selected from first step. Using this approach, either of the model (with or without climatic variables) can be used to predict dominant heights depending on data availability for the region.

In this study the Lundqvist-Korf model derived using the GADA method outperformed other candidate models. This result differs from other studies carried out in the western part of our study region (i.e. east Texas) (Coble and Lee, 2010; Trim et al., 2019). Though it should be noted that neither of these studies had considered the Lundqvist-Korf model for evaluation. Trim et al. (2019) found that the Chapman-Richards and McDill-Amateis models were the best models for their study area. Our results also confirmed the lowest AIC value for the Chapman-Richards model. However, it was outperformed by the Lundqvist-Korf model in terms of validation statistics, residual analysis and prediction. The fit index for all models were greater than 93%, which implies the high predictive capabilities of all models investigated. Like previous studies, both Lundqvist-Korf and Chapman-Richards GADA models performed better in height prediction before the base age 25 (Fig. 5). With the use of advance modeling techniques such as Bayesian methods and machine learning, Lundqvist-Korf and Chapman-Richards models could be further investigated in future.

Site index assessment based solely on tree measurement data has always been a concern for silviculturalists and forest ecologists. Accurate assessments of site index can only be achieved by understanding complex interaction between soil properties, climate, silvicultural treatment, and tree physiology. Early studies in the 1960s and 70s had attempted to develop soil property based site index for several tree species in the South. Nevertheless, foresters recognized the huge variability associated with soil properties even over small areas. Inclusion of climatic variables in site index models has gained popularity with readily available climatic data from weather stations and

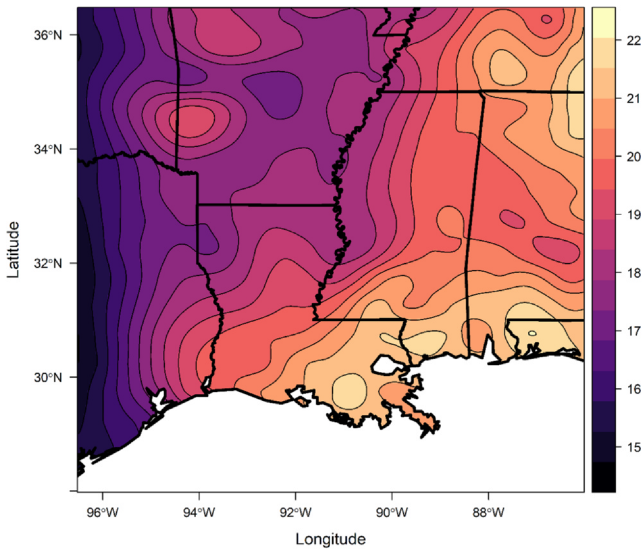

Fig. 8. Final site index map (in meters) for the Western Gulf study area using spatial regression and water balance variables.

national databases. This study used water balance components as climatic variables in dominant height and site index modeling. These variables are based on daily temperature, precipitation and solar radiation information. The inclusion of these variables into the height growth models allowed more accurate estimates of loblolly pine dominant heights in the Western Gulf. Water deficit was the defining climatic variable for dominant height growth and had a negative impact on the height growth estimates. Lower precipitation typical in the Western Gulf region increased water deficit and drought level, which in turn impacts the site productivity. Our results showed that areas closer to the Mississippi river basin had comparatively higher site index values, matching the understanding that areas near the Mississippi basin are more fertile.

5. Conclusion

This study indicated two forms of the Lundqvist-Korf model as the best dominant height models for loblolly pine plantations in the Western Gulf region. The modified Lundqvist-Korf model with GADA formulation was considered the best model for predicting dominant height growth out of field measurement data. The results indicated that the performance of the Lundqvist-Korf model to predict dominant height can be further improved through a reparameterization that includes a simplified indicator of water stress, the water deficit. The inclusion of this variable resulted in similar estimates as the empirical Lundqvist-Korf model without climatic variable. This implies that the statistical and biological properties of the model were not compromised during the process. However, model with climatic variable performed better in terms of fit and validation statistics than the one without climatic variable. Forest managers have the flexibility to use either of the models based on their data availability. A dominant height model and its corresponding site index estimator that includes a landscape level auxiliary variable is a step forward for loblolly pine site productivity modeling in the Western Gulf region of the US South. The variable selection methodology used in this study provides a new way to evaluate the appropriateness of adding auxiliary information into a model, provided it can be linearized.

CRediT authorship contribution statement

Anil Koirala: Conceptualization, Methodology, Formal analysis, Validation, Writing - original draft. **Cristian R. Montes:** Funding acquisition, Conceptualization, Methodology, Writing - review & editing. **Bronson P. Bullock:** Visualization, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the Plantation Management Research Cooperative (PMRC) for data access, and the PMRC field crew for their relentless work in data collection and management of the Western Gulf Culture Density study sites. This work was supported by the USDA National Institute of Food and Agriculture, McIntire Stennis project 1023340. The authors would also like to thank Texas A & M University and collaborators for study establishment and management through 2008.

References

Amateis, R.L., Burkhart, H.E., 1985. Site index curves for loblolly pine plantations on cutover site-prepared lands. South. J. Appl. For. 9, 166–169.
Bailey, R.L., Clutter, J.L., 1974. Base-age invariant polymorphic site curves. For. Sci. 20, 155–159. https://doi.org/10.1093/forestscience/20.2.155.
Borders, B., Harrison, W.M., Zhang, Y., Shiver, B.D., Clutter, M., Cieszewski, C.J., Daniels, R.F., 2004. Growth and yield models for second rotation loblolly pine plantations in the Piedmont/Upper Coastal Plain and Lower Coastal Plain of the southeastern U.S, University of Georgia, PMRC Technical Report 2004-4, 63p.
Borders, B., Zhao, D., Wang, M., Kane, M.B., 2014. Growth and yield models for second/third rotation loblolly pine plantations in the Piedmont/Upper Coastal Plain and Lower Coastal Plain of the southeastern U.S., rsity of Georgia, PMRC Technical Report 2014-1, 49p.
Cao, Q.V., 1993. Estimating coefficients of base-age-invariant site index equations. Can. J. For. Res. 23, 2343–2347.
Cao, Q.V., Baldwin, V.C., Lohrey, R.E., 1997. Site index curves for direct-seeded loblolly and longleaf pines in Louisiana. South. J. Appl. For. 21, 134–138.
Carvalho, J.P., Parresol, B.R., 2005. A site model for Pyrenean oak (Quercus pyrenaica) stands using a dynamic algebraic difference equation. Can. J. For. Res. 35, 93–99. https://doi.org/10.1139/x04-155.
Chapman, D.G., 1961. Statistical problems in population dynamics. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. University of California California, pp. 153–186.
Cieszewski, C.J., 2003. Developing a well-behaved dynamic site equation using a modified Hossfeld IV Function Y 3 = (axm)/(c + xm–1), a simplified mixed-model and scant subalpine fir data. For. Sci. 49, 539–554.
Cieszewski, C.J., Bailey, R.L., 2000. Generalized algebraic difference approach: theory based derivation of dynamic site equations with polymorphism and variable asymptotes. For. Sci. 46, 116–126. https://doi.org/10.1093/forestscience/46.1.116.
Clutter, J.L., Fortson, J.C., Pienaar, L.V., Brister, G.H., Bailey, R.L., 1983. Timber Management: A Quantitaive Approach. John Wiley & Sons Inc, New York, NY.
Coble, D.W., Lee, Y., 2010. Self-referencing site index equations for unmanaged loblolly and slash pine plantations in east texas. In: Stanturf, J.A. (Ed.), Proceedings of the 14th Biennial Southern Silvicultural Research Conference. USDA Forest Service Southern Research Station, pp. 349–353.
Diéguez-Aranda, U., Burkhart, H.E., Amateis, R.L., 2006. Dynamic site model for loblolly pine (Pinus taeda L.) plantations in the United States. For. Sci. 52, 262–272.
Fox, T.R., Jokela, E.J., Allen, H.L., 2004. The evolution of pine plantation silviculture in the Southern United States. In: Rauscher, H.M., Johnsen, K. (Eds.), Southern Forest Science: Past, Present, and Future. Forest Service, Southern Research Station, Asheville, NC, pp. 63–81.
Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33, 22.
Gallagher, D.A., Montes, C.R., Bullock, B.P., Kane, M.B., 2019. Two-step regression process for whole stand loblolly pine survival projection and quantifying uncertainty. For. Sci. 65, 265–276. https://doi.org/10.1093/forsci/fxy055.
Gonzalez-Benecke, C.A., Martin, T.A., Clark, A., Peter, G.F., 2010. Water availability and genetic effects on wood properties of loblolly pine (Pinus taeda). Can. J. For. Res. 40, 2265–2277. https://doi.org/10.1139/X10-162.
Graves, H.S., 1907. Forest Mensuration, first. ed. John Wiley & Sons Inc, New York, NY.
Hargreaves, G.H., Samani, Z.A., 1982. Estimating potential evapotranspiration. J. Irrig. Drain. Eng. 108, 225–230.
Harrison, W.M., Borders, B., 1996. Yield prediction and growth projection for site-prepared loblolly pine plantations in the Carolinas, Georgia, Alabama and Florida, University of Georgia, PMRC Technical Report 1996-1, 59p.
Kane, M.B., Zhao, D., Rheney, J., Chappell, N., Messina, M., 2011. Western Gulf culture density study: Results through age 8, University of Georgia, PMRC Technical Report 2011-1, 52p.

Krumland, B., Eng, H., 2005. Site index systems for major young-growth forest and woodland species in northern California., California Forestry Report No. 4. Sacramento, CA.
Lambers, H., Chapin III, F.S., Pons, T.L., 2008. Plant Physiological Ecology, second. ed. Springer Science & Business Media, New York, NY.
Landsberg, J.J., Waring, R.H., Coops, N.C., 2003. Performance of the forest productivity model 3-PG applied to a wide range of forest types. For. Ecol. Manage. 172, 199–214. https://doi.org/10.1016/S0378-1127(01)00804-0.
Lundqvist, B., 1957. On the height growth in cultivated stands of pine and spruce in Northern Sweden. Medd Fran Statens Skogforsk 47, 1–64.
McDill, M.E., Amateis, R.L., 1992. Measuring forest site quality using the parameters of a dimensionally compatible height growth function. For. Sci. 38, 409–429. https://doi.org/10.1093/forestscience/38.2.409.
Monserud, R.A., Huang, S., Yang, Y., 2006. Predicting lodgepole pine site index from climatic parameters in Alberta. For. Chron. 82, 562–571. https://doi.org/10.5558/tfc82562-4.
Nardi, Y., Rinaldo, A., 2011. Autoregressive process modeling via the lasso procedure. J. Multivar. Anal. 102, 528–549.
Oswalt, S.N., Smith, W.B., Miles, P.D., Pugh, S.A., 2019. Forest resources of the United States, 2017, US Forest Service, Gen. Tech. Report WO-97, 223 p. https://doi.org/10.5962/bhl.title.101492.
Özçelik, R., Cao, Q.V., Gómez-García, E., Crecente-Campo, F., Eler, Ü., 2019. Modeling dominant height growth of cedar (Cedrus libani A. Rich) stands in Turkey. For. Sci. 65, 725–733. https://doi.org/10.1093/forsci/fxz038.
Peay, W.S., 2019. Evaluating growth and yield modeling of loblolly pine (Pinus taeda L.) in the southeastern United States and assessing geographic bounds on model usage. MS Thesis. University of Georgia.
Pienaar, L. V., Burgan, T., Rheney, J.W., 1987. Stem volume, taper and weight equations for site-prepared loblolly pine plantations, University of Georgia, PMRC Technical Report No. 1987-1.
Restrepo, H.I., Bullock, B.P., Montes, C.R., 2019. Growth and yield drivers of loblolly pine in the southeastern U.S.: a meta-analysis. For. Ecol. Manage. 435, 205–218. https://doi.org/10.1016/j.foreco.2018.12.007.
Richards, F.J., 1959. A flexible growth function for empirical use. J. Exp. Bot. 10, 290–301.
Sabatia, C.O., Burkhart, H.E., 2014. Predicting site index of plantation loblolly pine from biophysical variables. For. Ecol. Manage. 326, 142–156. https://doi.org/10.1016/j.foreco.2014.04.019.
Schumacher, F.X., 1939. A new growth curve and its application to timber yield studies. J. For. 37, 819–820.
Sharma, M., Amateis, R.L., Burkhart, H.E., 2002. Top height definition and its effect on site index determination in thinned and unthinned loblolly pine plantations. For. Ecol. Manage. 168, 163–175. https://doi.org/10.1016/S0378-1127(01)00737-X.
Skovsgaard, J.P., Vanclay, J.K., 2013. Forest site productivity: a review of spatial and temporal variability in natural site conditions. For. An Int. J. For. Res. 86, 305–315. https://doi.org/10.1093/forestry/cpt010.
Subedi, S., Fox, T., 2016. Predicting loblolly pine site index from soil properties using partial least-squares regression. For. Sci. 62, 449–456. https://doi.org/10.5849/forsci.15-127.
Thornthwaite, C.W., Mather, J.R., Carter, D.B., 1957. Instructions and tables for computing potential evapotranspiration and the water balance. Publ. Climatol. 10, 185–311.
Tibshirani, R., 2011. Regression shrinkage and selection via the lasso: a retrospective. J. R. Stat. Soc. B 73, 273–282.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. B 58, 267–288.
Trim, K.R., Coble, D.W., Weng, Y., Stovall, J.P., Hung, I.-K., 2019. A new site index model for intensively managed loblolly pine (Pinus taeda) plantations in the West Gulf Coastal Plain. For. Sci. 1–12. https://doi.org/10.1093/forsci/fxz050.
Wang, G.G., Huang, S., Monserud, R.A., Klos, R.J., 2004. Lodgepole pine site index in relation to synoptic measures of climate, soil moisture and soil nutrients. For. Chron. 80, 678–686.
Wang, M., Montes, C.R., Bullock, B.P., Zhao, D., 2020. An empirical examination of dominant height projection accuracy using difference equation models. For. Sci. 66, 267–274. https://doi.org/10.1093/forsci/fxz079.
Weiskittel, A.R., Hann, D.W., Hibbs, D.E., Lam, T.Y., Bluhm, A.A., 2009. Modeling top height growth of red alder plantations. For. Ecol. Manage. 258, 323–331. https://doi.org/10.1016/j.foreco.2009.04.029.
Zhao, D., Bullock, B.P., Montes, C.R., Wang, M., Greene, D., Sutter, L., 2019. Loblolly pine outperforms slash pine in the southeastern United States – a long-term experimental comparison study. For. Ecol. Manage. 450, 15. https://doi.org/10.1016/j.foreco.2019.117532.
Zhao, P., Yu, B., 2006. On model selection consistency of Lasso. J. Mach. Learn. Res. 7, 2541–2563.