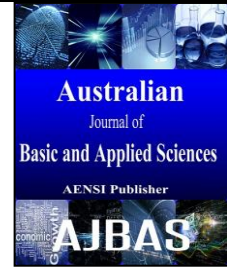




AUSTRALIAN JOURNAL OF BASIC AND APPLIED SCIENCES

ISSN: 1991-8178
EISSN: 2309-8414

DOI: 10.22587/ajbas.2017.11.14.10
Journal home page: www.ajbasweb.com



Height-Diameter Model for Managed Even-aged Stands of Ponderosa Pine for the Western United States Using Hierarchical Nonlinear Mixed-Effects Model.

Fabian C.C. Uzoh

Pacific Southwest Research Station, USDA Forest Service, 3644 Avtech Parkway, Redding, CA 96002, United States

Address For Correspondence:

Fabian C.C. Uzoh, Pacific Southwest Research Station, USDA Forest Service, 3644 Avtech Parkway, Redding, CA 96002, United States
E-mail: fuzoh@fs.fed.us; Tel: 1-530-226-2549; fax: 1-530-226-5091.

ARTICLE INFO

Article history:

Received 18 August 2017

Accepted 22 November 2017

Available online 6 December 2017

Keywords:

Height-Diameter Model; Hierarchical Nonlinear Mixed-Effects Model; Tree-growth modeling; Repeated Measures Analysis; variance components; *Pinus ponderosa*

ABSTRACT

Background: A regional individual-tree Height-Diameter (H-D) model was developed for ponderosa pine (*P. ponderosa* var. *ponderosa* and var. *scopulorum*), for the western United States. The dataset came from long-term permanent research plots in even-aged pure stands, both planted and natural on sites capable of the productivity estimated by Meyer [Meyer, W.H., 1938. Yield of Even-Aged Stands of Ponderosa Pine. USDA Technical Bulletin 630]. The database used a common study plan. The study plan divided the range of ponderosa pine trees into five provinces, with installations in Arizona (AZ), northern California (CA), Oregon (OR), Washington (WA), Montana (MT), and South Dakota (SD), USA. Regional H-D models exist for ponderosa pine. However, the study plans, the data collection and the analysis procedures used in developing the models differ. Consequently, comparisons of growth responses that may be due to geographic variation of the species are not possible. **Objective:** 1) develop a single regional H-D model for ponderosa pine trees that can provide useful guidelines for a variety of management objectives throughout the species range; 2) determine if inclusion of plot-specific auxiliary variables improves prediction accuracy of regional H-D models. **Methodology:** The dataset consisted of 305 plots and 29,449 trees. The study used 8208 trees that had both height and diameter measurements. The dataset has repeated measurements, which makes the measurements correlated. Additionally, the dataset has location effects and hierarchical structure; consequently, it violated the fundamental regression assumption of independent observations. Nonlinear Mixed-Effects Model (NLMM) approach provided an excellent solution. Specifically, NLMM allows parameter vectors to vary from plot to plot and variance components are simultaneously broken down into a) fixed-effect components that describe the shape of the typical growth curve over the entire population, b) random-effect components that enables the curve to capture the plot-specific and other unit-specific characteristics of the growth pattern. Further, NLMM allows the explicit incorporation of the hierarchical structure of the dataset. Finally, the inclusion of plot-specific auxiliary variables individualized the curve for each plot and location. **Conclusions:** The Chapman-Richards model form provided the best result. The results demonstrated that irrespective of geographic location, a regional H-D Model with variance components and plot-specific auxiliary variables that individualized the curve to capture the site-specific, plot-specific, and other unit-specific characteristics of the growth pattern decisively outperformed regional H-D Model with variance components but without plot-specific auxiliary variables. Consequently, the model developed herein should enhance comparisons of H-D relationships that may be due to geographic variation of the species throughout the range of ponderosa pine trees. Therefore, it would provide forest managers and researchers with useful guidelines for a variety of management objectives at the regional scale.

Open Access Journal

Published BY AENSI Publication

© 2017 AENSI Publisher All rights reserved

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

To Cite This Article: Uzoh, Fabian C.C., 2017. Height-Diameter Model for Managed Even-aged Stands of Ponderosa Pine for the Western United States Using Hierarchical Nonlinear Mixed-Effects Model. *Aust. J. Basic & Appl. Sci.*, 11(14): 69-87.

INTRODUCTION

Models describing the relationship between total tree height and diameter at breast height (dbh) outside bark are invaluable tools for forest management planning at any level by forest managers and researchers. Height and diameter are keystone variables in many growths and yield models because measurements of total tree height and diameter at breast height are required variables in the development and the application of growth and yield models (see, e.g., Curtis 1967; Stage 1973; Wykoff *et al.* 1982; Cortini *et al.* 2011; Johansson 2012). Height-Diameter (H-D) functions are essential for a variety of management objectives such as: 1) tree volume and biomass estimation (see, e.g., Meyer, 1940; Curtis, 1967; Osman *et al.*, 2013); 2) stand structural analysis and carbon budgeting models (see, Newton and Amponsah, 2007; Cortini *et al.*, 2011; Johansson 2012); 3) description of vertical stand structure and stand attributes dynamics (see, e.g., Stage, 1973; Dubrasich *et al.*, 1997); 4) evaluation of site productivity in uneven-aged and mixed-species stands (see, e.g., Stout and Shumway, 1982; Vanclay, 1994); and 5) assessment and modeling (see, Cortini *et al.*, 2011; Johansson 2012). Consequently, for reliable forest management planning, accurate measurement of these two variables are of utmost importance.

Tree height data is difficult, time-consuming, and expensive to obtain, especially in mountainous terrain and dense stands (see, e.g., Moore *et al.*, 1996; Hasenauer and Monserud, 1997; Uzoh and Oliver, 2006). As a result, rather than measure the heights of all the trees in a plot, inventory studies traditionally measure the diameter at breast height (dbh) of all the trees in the plot and height is measured on a sub-sample of trees. An H-D equation is developed using the trees that have both height and diameter measurements, and that equation is then used to predict height for all the trees in the plot (see, e.g., Curtis, 1967; Stage, 1973; Uzoh and Oliver, 2006). However, in cases where only the diameters were measured foresters use published regression equations to predict height values (see, e.g., Monleon, 2004; Saunders and Wagner 2008).

Traditionally, diameter was the only independent variable used in H-D models (see, e.g., Curtis, 1967; Wykoff *et al.*, 1982; Osman *et al.*, 2013). However, the ecological processes that govern the relationship between height and diameter growth vary within stands and between stands, they are neither uniform nor linear, and they are not constant over time (see, Curtis, 1967). These ecological processes are in turn, governed by factors such as environmental features and stand conditions (see, e.g., Curtis, 1967; Lopez, Sanchez, *et al.* 2003). As a result, many studies show that predictive accuracy of the base model can improve by the addition of plot level variables that can capture the special characteristics of each plot and location (see, e.g., Larsen and Hann, 1987; Huang *et al.*, 1992; Calama and Montero, 2004). Nevertheless, researchers such as Huang *et al.* (2009) and Mehtätalo *et al.* (2015) developed accurate individual tree regional H-D models without the inclusion of auxiliary variables.

H-D models are divided into two categories: local or simple H-D models, and generalized or regional H-D models. 1) Local or simple H-D models depend only on tree diameters and do not incorporate auxiliary variables or plot-specific predictors. Thus they are usually only applicable locally to the stand where the H-D data were collected (see, e.g., Curtis, 1967; Wykoff *et al.*, 1982; Robinson and Wykoff, 2004). 2a) a first group of generalized or regional H-D models depends only on tree diameters and does not incorporate auxiliary variables or plot-specific predictors (see, e.g., Lappi, 1997; Huang *et al.*, 2009). 2b) a second group of generalized or regional H-D models depends on tree diameters but also incorporates auxiliary variables or plot-specific predictors that can capture special characteristics of each plot. Such as the basal area and stand density (see, e.g., Larsen and Hann, 1987; Sharma and Parton, 2007), average of the top height trees (see, e.g., Calama and Montero, 2004; Sharma and Parton, 2007), and quadratic mean diameter (see, Zakrzewski and Bella, 1988). As a result, these models are applicable in similar plot conditions at the regional level.

Datasets used in most ecological and biological research are collected from multiple locations that are not similar to each other. In addition, these datasets have repeated measurements on the same subject, which are fundamental to most ecological and biological research (see, Gutzwiller and Riffell, 2007). These repeated measurements taken on the same subject over time are correlated. Consequently, multiple sources of heterogeneity are introduced and the fundamental regression assumptions that observations are independent are violated (see, Grégoire, 1987; Fortin *et al.*, 2007). Although the estimation of model parameters would remain consistent and unbiased (see, LeMay, 1990), estimation efficiency may decline and the standard errors of model parameters may be biased. As a result, hypothesis testing and statistical inference would be invalid (Grégoire, 1987; Grégoire, *et al.*, 1995; Fortin *et al.*, 2007). Consequently, H-D models developed with datasets from different locations with repeated measurements that ignore the issue of serial correlation and location effects are not adequate. Furthermore, H-D models characterized with hierarchical structured-datasets (e.g., trees nested within plots, plots nested within locations, and locations nested within province) that address the issues of serial correlation and location effects, but fail to address the issue of nesting are still not a complete solution. The reason is that the relationship between H and D between-stands is not the same as that within-stand across locations. Competition generally governs within-stand relationships, while stand age, site quality, stand density;

spatial distribution pattern, microsite conditions, and environmental factors governed between-stand relationships (see, e.g., Larsen and Hann, 1987; Zeide and Vanderschaaf, 2002).

A simple solution is to fit separate models by plot and each measurement regime or occasion. This approach is not practical because it is computationally prohibitive. Additionally, such H-D models would not allow prediction outside the modeling dataset. Importantly, when making predictions for new plots, such an approach leads to other problems like over-parameterization, if other auxiliary or plot-specific variables are included in the model (see, Mehtätalo *et al.*, 2015).

The Nonlinear mixed-effects model (NLMM) approach provided an excellent solution. NLMM eliminated the necessity to fit separate H-D models by plot for each measurement regime or occasion (see, Temesgen and Gadow, 2004). Specifically, NLMM approach allows parameter vectors to vary from plot to plot. Importantly, NLMM allow variance components to be simultaneously broken down into a) fixed-effect components that describe the shape of the typical growth curve over the entire population, b) the random-effect prediction that provides the plot-specific H-D relationship that captures special characteristics of each plot and location (see, e.g., Temesgen and Gadow, 2004; Meng and Huang, 2010; Crecente-Campo *et al.*, 2010). Additionally, NLMM also enable the direct modeling of the error structure imposed by the sampling scheme, thus accounts for the nonzero correlation (see, e.g., Littell *et al.*, 2006; Uzoh and Oliver 2008; Huang *et al.*, 2009; Meng and Huang, 2010). The result is a reduction in serial correlation and improvement in fit statistic and model performance. Furthermore, the error structure imposed by the sampling scheme is correctly modeled and partitioned (see, e.g., Littell *et al.*, 2006; Uzoh and Oliver 2008; Shawn *et al.*, 2010). In addition, the inclusion of auxiliary variables individualized the curve for each plot and location. As a result, made H-D models applicable in similar plot conditions at the regional level. The goal is always to strike a balance between overfitting and underfitting (see, Yang and Bozdogan, 2011).

In addition, the NLMM approach allows the explicit modeling of the hierarchical structure of the dataset (for details, see SAS User's Guide 2015, Example 70.7). This capability makes NLMMs exceptional in making predictions for new dataset (see, Trincado *et al.*, 2007). Nevertheless, bias may still be an issue in making predictions because biologically, NLMMs are asymptotically unbiased (see, Box, 1971; Fang and Bailey, 1998).

For decades, because of the importance of H-D relationships in forest management and planning, many researchers have aimed to develop better H-D relationships. As early as 1940, Meyer (1940) used the exponential model to express height curves. Thereafter, other researchers have employed diverse model forms to achieve the same goal. However, there is no unanimous agreement on which model form is the best, because data structure and modeling objectives dictate the choice of model form. For example, Yuancai and Parresol (2001), Sharma and Parton (2007), and Hassan *et al.* (2015) choose Bertalanffy-Richards (Bertalanffy, 1949, 1957; Richards, 1959). While Yang *et al.* (1978) and Peng *et al.* (2001) concluded that Weibull-type model forms provided better results than other model forms; and still other researchers such as Huang *et al.* (2000), Meng and Huang (2010), and Mehtätalo *et al.* (2015) choose the logistic function for modeling H-D relationships.

In the western United States, many regionally limited H-D models are available for ponderosa pine trees (Ritchie, 1999). However, comparisons of growth responses that may be due to geographic variation of the species are difficult, because the study plans, the data collection, and the analysis procedures used to develop these models differ. Most of the datasets used to develop these models originated, either totally or in part, from temporary plots, often using stem analysis techniques. Such data suffer from the same weaknesses as retrospective studies. The investigator is never certain that the response measured is the result of the stated condition (see, Uzoh and Oliver, 2008). As a result, forest managers need an H-D model that can provide useful guidelines and comprehensive strategies for meeting a variety of management objectives throughout the geographic range of ponderosa pine trees in the western United States. The opportunity to develop such an H-D model became possible through access to a unique database from permanent study plots located throughout the geographic range of ponderosa pine trees in the western United States (Fig. 1), and including stands of both planted and natural origin. The database consists of six levels-of-growing stock studies; all used a common study plan, supplemented by initial spacing and other permanent-plot thinning studies (see, Myers, 1967).

Therefore, the objectives of this study were (1) to overcome these weaknesses by presenting a single H-D model applicable throughout the geographic range of ponderosa pine stands in the western United States. Using only data from long-term permanent plots on sites capable of normal yields or yield from a fully stocked stand (see, Meyer 1938), and (2) to determine if inclusion of plot-specific auxiliary variables that account for the special characteristics of each plot and location improved prediction accuracy of generalized or regional H-D models. The second objective would be addressed by comparing the regional H-D models with and without plot-specific auxiliary variables using the model comparison dataset. The following groups of measured candidate explanatory (predictor) variables were considered for the H-D model with auxiliary variables:

1. Plot spatial information: latitude, longitude, elevation, slope and aspect.

2. Plot structure information: the plot basal area, the number of trees per hectare, the stand age, the average height of the top trees in the plot, the site index, the site density index, the quadratic mean diameter and the basal area per hectare in larger diameter.
3. Tree information: height, diameter, diameter periodic annual increment, and the tree basal area.

MATERIAL AND METHODS

Data description:

The foundation of the database is five levels-of-growing-stock studies established in the 1960s throughout the geographic range of ponderosa pine trees in the western United States by several western research stations of the USDA Forest Service. The dataset came from long-term permanent research plots in even-aged pure stands, both planted and natural on sites capable of the productivity estimated by Meyer (1938). The database used a common study plan. The study plan divided the range of ponderosa pine trees in the western United States into five provinces (Fig. 1) and specified five or six stand density levels replicated three times with installations in Arizona (AZ), northern California (CA), Oregon (OR), Washington (WA), Montana (MT), and South Dakota (SD) (see, Myers, 1967). The locations represents a wide range of ecological zones with contrasting soil, landforms, geology, and climate, and included two variety of ponderosa pine trees (*P. ponderosa* var. *ponderosa* and var. *scopulorum*) and a wide range of size classes. Previous studies have reported results from individual installations (Table I). Initial spacing and other permanent-plot thinning studies supplement the dataset. Stands were free or mostly free of competing shrubs that reduce the growth of young ponderosa pine, especially in central Oregon and California (see, e.g., Oliver, 1984; Oliver and Edminster, 1988; Oliver and Ryker, 1990; Oliver, 1990; Cochran and Barrett, 1999). Trees in all plots in the database were tagged to allow the collection of information on individual trees. The number of growing seasons between remeasurements was usually five. Eighty-two percent of the plots were observed for 20 years or more—at least four 5-years growth periods. Basic records for each plot included longitude, latitude, elevation, aspect, slope percent and plot size. Tree records at each remeasurement included diameter at breast height and total height on a sample of trees. In some studies, every tree height in the plot was measured; in others a systematic sample of tree heights was taken; yet in other studies height sample trees were randomly selected within 5-cm diameter classes across the range of tree sizes. Diameter and height measurements were repeated on the same trees every 5 years. Detailed description of each installation and data collection procedures can be found in the following publications: Barrett, 1983; Boldt and Van Deusen, 1974; Oliver 1979, 1997; Oliver and Edminster, 1988; Cochran and Barrett, 1995, 1999; Ronco *et al.*, 1985. The data resulting from this study came from 305 plots with a total of 29,449 trees, 152,314 observations, and 8208 trees had both height and diameter measurements. This analysis used only trees with both height and diameter measurements. Scatter plots of height versus diameter are in Fig. 2 and the scatter plot of height versus auxiliary variables used to develop the model are in Fig. 3. Distribution of plot in each province by stand origin are in Table 2, and the variables summary statistics are in Table 3.

Base Height-Diameter Model Development:

The dataset was randomly split into 10 parts, then nine of those 10 parts (90%) were used for model development and 1 part (10%) was kept for model comparison. Scattered-plot diagrams using the model development dataset were used in preliminary analysis to examine the relationship between H and D at the plot level without the inclusion of auxiliary variables or plot-specific predictors. Then, 18 model forms from previous studies (in the original or the modified form) were fitted to the dataset at the plot level using the model development dataset to determine the model forms that would best describe the H-D relationships using nonlinear regression analysis.

Goodness-of-fit criteria:

Model parameters and statistics were estimated using statistical analysis system software SAS (SAS Institute, 2014), using maximum likelihood estimation methodology to estimate parameters. The best linear unbiased predictor (BLUP) maximize the marginal likelihood function (see, Beal and Sheiner, 1982). Of the 18 model forms examined, 11 were selected (Gomperz, 1825; Pearl and Reed, 1920; Meyer, 1940; Weibull, 1951; Stoffels and van Soest, 1953; Chapman-Richards, 1959; Curtis, 1967; Stage, 1975; Larsen and Hann, 1987; and Wykoff *et al.*, 1982) (Table 4). The models were selected for further analysis based on: (1) Model fit statistics: - 2 times log-likelihood (-2LL), the Akaike Information Criterion (AIC), the Schwarz Bayesians information criterion (BIC), the significance of parameters, and the root mean square error (RMSE); (2) Visual analysis of the residual plots against the predicted values, the function plots; and (3) if convergence was met (see, e.g., Rawlings *et al.*, 1998; Hastie *et al.*, 2001; Burnham and Anderson, 2002; Littell *et al.*, 2006).



Fig. 1: Study area provinces used to develop the Individual tree Height-Diameter model for managed even-aged stands of ponderosa pine trees for the western United States using hierarchical nonlinear mixed-effects model. The Roman numerals show the division of the geographic area of ponderosa pine trees in the western United States into five provinces (from Myers, 1967).

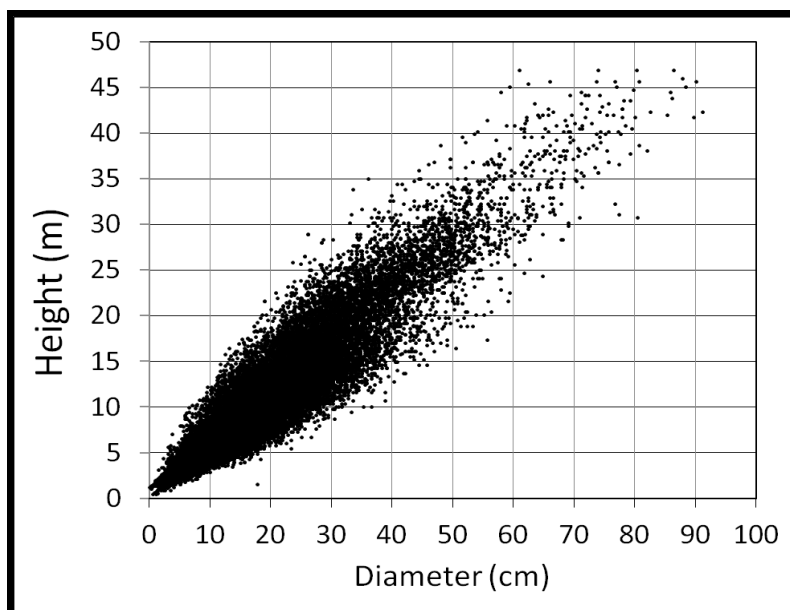


Fig. 2: Scatter plots of height versus diameter used to develop the Individual tree Height-Diameter model for managed even-aged stands of ponderosa pine trees for the western United States using hierarchical nonlinear mixed-effects model.

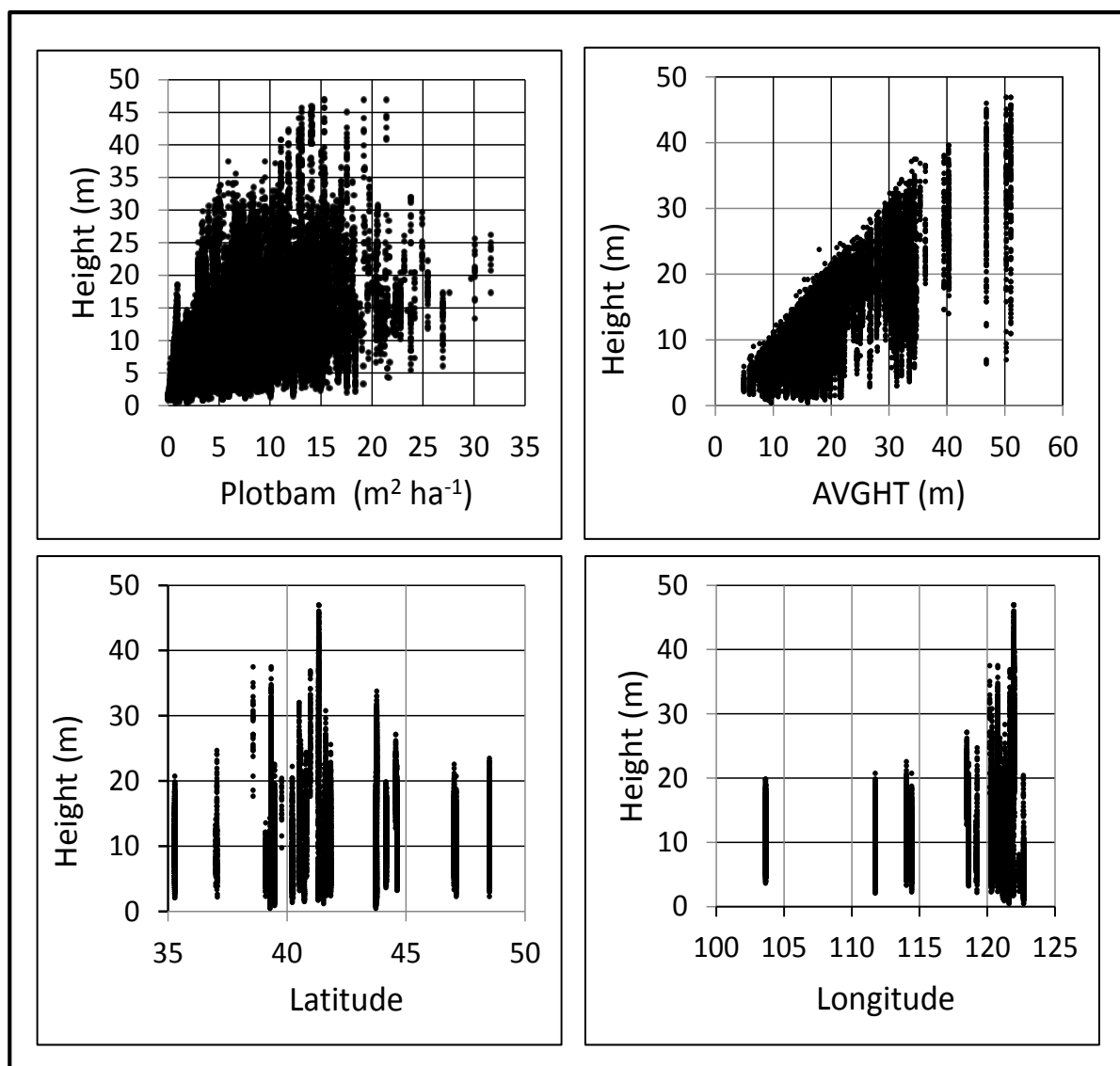


Fig. 3: Scatter plots of height versus auxiliary variables used to develop the Individual tree Height-Diameter model for managed even-aged stands of ponderosa pine trees for the western United States using hierarchical nonlinear mixed-effects model.

Table 1: Location and literature citations for five levels-of-growing-stock installations in ponderosa pine stands throughout the western United States.

Province	Installation Name	Geographic Location	Literature Citation
I	Elliot Ranch	West slope northern Sierra Nevada, CA	Oliver 1997
II	Lookout Mountain	East side of Cascade Range,	Cochran and Barrett 1999
III	Crawford Creek	Blue Mountain, OR	Cochran and Barrett 1995
IV	Black Hills	Black Hills, SD	Boldt and Van Deusen 1974
V	Taylor Woods	Coconino Plateau, AZ	Ronco <i>et al.</i> 1985

Table 2: Plot distribution in each province by stand origin and tree size for managed even-aged stands of ponderosa pine stands throughout the western United States.

Province		I	II	III	IV	V
Number of Plots						
Stand Origin						
	Natural	7	81	32	42	18
	Planted	92	18	15	0	0
Stand Size Class						
	Saplings	32	30	9	15	0
	Poles	58	61	38	27	18
	Sawtimber	9	8	0	0	0
	Total	99	99	47	42	18

Table 3: Summary statistics for variables used to develop the Individual tree Height-Diameter model for managed even-aged stands of ponderosa pine trees for the western United States using hierarchical nonlinear mixed-effects model.

Variables	Number of Trees	Mean	Minimum	Maximum
HTM (m)	8208	10.608	0.457	53.214
DBHCM (cm)	8208	18.89	1.02	98.04
Plot Variables	Number of Plots	Mean	Minimum	Maximum
SLOPE (percent)	305	6.16	0	42
ASPECT (degrees)	305	113.07	0	315
Stand age (years)	305	66.45	8	110
SIMM (m)	305	21.62	13.11	48.77
AVGHT (m)	305	18.7166	4.91947	51.0243
PLOTBAM (m ² ha ⁻¹)	305	0.003716	0.003716	52.37
SDIM (trees ha ⁻¹)	305	576.43	21.18	1,444.22
TPH (trees ha ⁻¹)	305	1,553.29	44.48	17,173.73
ELEVA (m)	305	1503.79	716.28	2266.19
LATITUDE (degrees)	305	42.22	35.282	48.52
LONGITUDE (degrees)	305	117.95	103.6	122.67

Table 4: Height-Diameter functions selected for evaluation.

Model Name	Model forms	AIC	RMSE
2. Curtis (1967)	$HT_{ijk} = 1.3 + \beta_1 * \exp(-\beta_2 * DBH_{ijk}^{\beta_3}) + e_{ijk}$	162464	6.0210
3. Gompertz (1825)	$HT_{ijk} = 1.3 + \beta_1 \exp(-\beta_2 \exp(-\beta_3 DBH_{ijk})) + e_{ijk}$	162813	6.0153
4. Weibull (1951)	$HT_{ijk} = 1.3 + \beta_1 (1 - \exp(-\beta_2 DBH_{ijk}^{\beta_3})) + e_{ijk}$	162835	6.3319
5. Chapman-Richards (1959)	$HT_{ijk} = 1.3 + \beta_1 (1 - \exp(-\beta_2 DBH_{ijk}^{\beta_3}))^{\beta_3} + e_{ijk}$	162836	6.0200
6. Stoffels and van Soest (1953)	$HT_{ijk} = 1.3 + \beta_1 DBH_{ijk}^{\beta_2} + e_{ijk}$	162841	6.0210
7. Curtis (1967)	$HT_{ijk} = 1.3 + \frac{\beta_1 DBH_{ijk}}{(1 + DBH_{ijk})^{\beta_2}} + e_{ijk}$	162841	6.0210
8. Stage (1975)	$HT_{ijk} = 1.3 + \beta_1 DBH_{ijk}^{\beta_2} + e_{ijk}$	162841	6.0210
9. Larsen and Hann (1987)	$HT_{ijk} = 1.3 + \exp(\beta_1 + \beta_2 * DBH_{ijk}^{\beta_3}) + e_{ijk}$	162851	9.4250
10. Meyer (1940)	$HT_{ijk} = 1.3 + \beta_1 (1 - \exp(-\beta_2 DBH_{ijk})) + e_{ijk}$	163147	12.738
11. Pearl and Reed (1920)	$HT_{ijk} = 1.3 + \frac{\beta_1 + u_{i1}}{1 + (\beta_2 + u_{i2}) \exp(-\beta_3 DBH_{ijk})} + e_{ijk}$	166301	6.6623
12. Wykoff <i>et al.</i> 1982	$HT_{ijk} = 1.3 + \exp\left(\beta_1 + \left(\frac{\beta_2}{DBH_{ijk} + 1}\right)\right) + e_{ijk}$	166755	6.7302

Nonlinear Mixed-Effects Model (NLMM) Development:

Following Davidian and Giltinan (1995 and 2003), Vonesh and Chinchillia (1997), NLMM are written as follows:

$$y_{ij} = f(x_{ij}, \beta, b_i) + \varepsilon_{ij} \quad (i=1, 2, \dots, m; j=1, 2, \dots, n_i) \quad (1)$$

where y_{ij} is the j^{th} observation on subject i , f is some Nonlinear function, m is the number of sampled subjects in the population, n_i is the number of observations on subject i , x_{ij} is a known vector of covariates, β is a (p x 1) vector of unknown fixed-parameters common to all subjects, b_i is a (q x 1) vector of unknown random parameters unique for the i^{th} subject, and ε_{ij} is the unknown within-subject random error. Detailed

information and procedures on nonlinear mixed-effects modelling (NLMM) logic and methodology are not provided here. For theoretical developments and a comprehensive overview, readers are referred to examples like Davidian and Giltinan (1995); Lappi (1997); Vonesh and Chinchillia (1997); Hall and Bailey (2001); Davidian and Giltinan (2003); Calama and Montero 2004; Littell *et al.* 2006.

The selected 11 model forms (Table 4) were then fitted to the model building dataset as “random coefficients” models where there is a random slope and intercept deviation for each plot (PROC NL MIXED, SAS Institute Inc. 2014), using maximum likelihood estimation methodology to estimate parameters. Serial correlation was accounted for using covariance structure. Ignoring or avoiding the implications of temporally correlated observations may introduce the possibility of spurious conclusions. Selecting the appropriate covariance structure is very important. In this analysis, the procedure outlined in Littell *et al.* (2006), and Gutzwiller and Riffell (2007) was used in selecting the appropriate covariance structure. The spatial power error structure (with multiple observations on individual trees autocorrelated in time) outperformed other spatial covariance structures such as Spherical, Exponential, Gaussian, and Matern (with multiple observations on an individual tree equally correlated irrespective of time).

Hierarchical Nonlinear Mixed-Effects Model:

In addition to serial correlation and location effects, the dataset has a hierarchical structure (i.e. the trees nested within the plots, the plots nested within the locations, and the locations nested within the province). As a result, the relationship between H and D between-stands is not the same as that within-stand across provinces. The fundamental feature of the hierarchical nonlinear-mixed effects model is the correct specification of the error structure imposed by the sampling scheme, because it explicitly separates the between- and the within-stand relationships (see, Monleon, 2004; Uzoh and Oliver, 2008; Crecente-Campo *et al.*, 2010). The hierarchical structure of the dataset was explicitly incorporated in the models using %NLINMIX macro (SAS Institute Inc., 2004), using maximum likelihood estimation methodology to estimate parameters (for details see, SAS User's Guide 2015, Example 70.7). Of the 11 models (Table 4), Equations 3 and 5 were selected based on goodness-of-fit criteria:

$$\text{Gomperz (1825)} \quad H_{ijk} = 1.3 + \beta_1 \exp(-\beta_2 \exp(-\beta_3 D_{ijk})) + h_l + e_{j(l)} + e_{ik(jl)} \quad (13)$$

$$\text{Chapman-Richards (1959)} \quad H_{ijk} = 1.3 + \beta_1 * (1 - \exp(-\beta_2 D_{ijk}))^{\beta_3} + h_l + e_{j(l)} + e_{ik(jl)} \quad (14)$$

where HT_{ijk} and DBH_{ijk} are the height and diameter measurement k on the i^{th} tree in the j^{th} plot (subject), $i = 1, 2, \dots, n_i$, $j = 1, 2, \dots, m$, n_i is the number of trees in plot j , m is the total number of plots in the population, β 's are the fixed-effects parameters common to all the plots; $\beta' s \sim N(0, \sigma_e^2)$; h_l is the random effect of the l -th location with h_l assumed to have an expected value of zero (0) and variance σ_L^2 ; $e_{j(l)}$ is a random error for plot j within locale l assumed to have an expected value of zero (0) and variance ($\sigma_{P(L)}^2$), and $e_{ik(jl)}$ is a random error for measurement k on tree i within plot j and locale l assumed to have an expected value of zero (0) and variance (σ^2) with the covariance between observations k and k' on the same tree separated by d years following an autoregressive process:

$$\text{Cov}(e_{ik(jl)}, e_{ik'(jl)}) = \begin{cases} \sigma^2 \rho^{|d|}, & \text{if } i=i', k \neq k', j=j', l=l' \\ \sigma^2, & \text{if } i=i', k=k', j=j', l=l' \\ 0, & \text{otherwise,} \end{cases} \quad (15)$$

where ρ is the spatial power (SP (POW)) serial correlation coefficient for errors across time (five years) on the same tree. The spatial power error structure outperformed other spatial covariance structures such as Spherical, Exponential, Gaussian, and Matern (with multiple observations on an individual tree equally correlated irrespective of time). The models were estimated using maximum likelihood (ML) approach. The Chapman-Richards model form was selected for further examination because it is biologically sound and superior in terms of model fit characteristics and less bias in estimating height.

Inclusion of Auxiliary Variables or Plot-Specific Parameters:

The second objective was to determine if inclusion of auxiliary variables improved prediction accuracy of generalized or regional H-D models. One of the fundamental means of capturing the within- and the between-

plot variations is by inclusion of auxiliary variables that account for the special characteristics of each plot and location. The auxiliary variables are easily measurable tree and plot characteristics. As a result, these plot special characteristics makes H-D models applicable in similar plot conditions at the regional level. Four groups of variables affect tree growth: tree size and vigor effects, site effects, competitive effects, and regional effects (see, e.g., Wyckoff, 1990; Calama and Mentero, 2004; Uzoh and Mori, 2012).

The following plot-special characteristics were tested for inclusion in the model:

1.	PLOTBAM	=	plot basal area per hectare ($m^2 ha^{-1}$)
2.	TPH	=	number of trees per ha (trees ha^{-1})
3.	PBATPH	=	plot basal area/number of trees per hectare
4.	StAGE	=	stand age
5.	BAINLD2M	=	basal area per hectare in larger diameter ($m^2 ha^{-1}$)
6.	AVGHT	=	average of the top height trees in the plot
7.	SIMM	=	site index (Meyer 1938)
8.	SDIM	=	stand density index (trees ha^{-1})
9.	ELEVAM	=	plot elevation in meters
10.	SLOPE	=	plot slope percent
11.	ASP	=	plot aspect
12.	SLCOSASP	=	SLOPE*cos(ASP)
13.	SLSINASP	=	SLOPE*sin(ASP)
14.	LAT	=	site latitude
15.	LONG	=	site longitude
16.	MEANDBHCM	=	plot mean diameter at breast height
17.	QMDM	=	plot quadratic mean diameter

Site effects covariate:

For a model to characterize tree growth potential adequately, it must include some empirical measures of site productivity potential (see, Spurr and Barnes, 1980). Traditionally, site index (SI) is a key variable used in H-D models (see, Biging, 1985). Biologically, site index (SI) is a legitimate empirical measure of site quality that captures the complex interplay between edaphic, climatic, and biotic factors of a site that reflects tree growth potential measured by the volume of wood produced (see, e.g., Daniel *et al.*, 1979; Spurr and Barnes, 1980). However, SI has many deficiencies (see, the discussion section).

As a result, other variables were tested; such as average of the top height trees in the plot (AVGHT), aspect, elevation, slope and their transformations. Stage (1976) transformation of slope and aspect, ((SL[COS(ASP)] and SL[SIN(ASP)])), were tested, but they were not statistically significant. Also, the Stage and Salas (2007) transformation of elevation, aspect, and slope was tested, but it did not work either; consequently, their proposed transformation was not used. The plots were located at elevations where ponderosa pine was best suited at that latitude. As a result, at lower latitudes, plots were located at higher elevations. Consequently, for Stage and Salas (2007) transformation of elevation, aspect, and slope to be successfully tested, this study would have been especially designed such that it had plots at several elevations on the same aspect on north and south slopes along a latitudinal transect. AVGHT was selected because it is biologically sound and superior to other site effects covariate tested and has less bias in estimating height. The site productivity potential covariate is:

$$H_{ijk} = 1.3 + \beta_1 * (1 - \exp(-\beta_2 D_{ijk}))^{\beta_3} * AVGHT_{ik}^{\beta_4} + h_l + e_{j(l)} + e_{ik(jl)} \quad (16)$$

where $AVGHT_{ik}$ is the average of the top height trees in the plot and other variables remain as previously defined.

Competitive effects covariate:

Tree growth potential is dictated by its competitive status relative to neighboring trees and the impact of forest management or disturbance (see, e.g., Wyckoff, 1990; Zeide and Vanderschaaf, 2002; Uzoh and Mori, 2012). The following variables were evaluated for competitive effects (CE) covariate:

1.	PLOTBAM	=	plot basal area per hectare ($m^2 ha^{-1}$)
2.	TPH	=	number of trees per ha (trees ha^{-1})
3.	PBATPH	=	plot basal area per hectare/number of trees per hectare
4.	BAINLD2M	=	basal area per hectare in larger diameter ($m^2 ha^{-1}$)
5.	SDIM	=	stand density index (trees ha^{-1})
6.	BAL/DBH	=	basal area per hectare in larger diameter divided by diameter at breast height of the subject tree

PLOTBAM was selected because it is biologically sound and superior to other site competition effects covariate tested and has less bias in estimating height. The competition effects (CE) covariate is:

$$H_{ijk} = 1.3 + \beta_1 * (1 - \exp(-\beta_2 D_{ijk}))^{\beta_3} * \text{AVGHT}_{ik}^{\beta_4} * \text{PLOTBAM}_{ik}^{\beta_5} + h_l + e_{j(l)} + e_{ik(jl)} \quad (17)$$

where PLOTBAM_{ik} is the plot basal per hectare and other variables remain as previously defined.

Regional effects covariate:

Geographic location may affect tree growth (see, e.g., Calama and Montero, 2004; Uzoh and Mori, 2012). Factors such as latitude (LAT), longitude (LONG), slope, aspect, and elevation effect tree growth. These factors generally do not have direct effects on tree growth, but they indirectly influence soil moisture, temperature, light, and other chemical and physical agents of the site (see, Uzoh and Oliver, 2008; Schippers *et al.*, 2015; Trouvé *et al.*, 2015). Elevation as a stand-alone variable was statistically significant. However, latitude was used instead because a high degree of multicollinearity existed between the two variables and latitude was more statistically significant than elevation. However, latitude alone could not represent a complete picture of geographic location effects. As a result, for a more complete picture of geographic location effects both latitude (LAT) and longitude (LONG) were include in the model and they were statistically significant. The regional effects covariates (REC) are:

$$H_{ijk} = 1.3 + \beta_1 * (1 - \exp(-\beta_2 D_{ijk}))^{\beta_3} * \text{AVGHT}_{ik}^{\beta_4} * \text{PLOTBAM}_{ik}^{\beta_5} + \beta_6 * \text{LAT}_{ik} + \beta_7 * \text{LONG}_{ik} + h_l + e_{j(l)} + e_{ik(jl)} \quad (18)$$

where LAT_{ik} and LONG_{ik} are the plot latitude and longitude and other variables remain as previously defined.

The variance increased with increasing predicted values $y \sim N(f(x), \sigma^2)$. Among the tested weights, weight of $1/f(x)**1.3$ was used because it gave the best fit to the models. The scatter plot of height against diameter and auxiliary variables are shown in Figures 2 and 3 respectively.

Model Comparison:

The second objective was to determine if inclusion of plot-specific auxiliary variables or plot-specific predictors improved prediction accuracy of generalized or regional H-D models. Independent data sets are often not available. As a result, the dataset was randomly divided into ten parts; then nine of those 10 parts (90%) were used for initial model development and 1 part (10%) was kept for model comparison between H-D model with and without auxiliary variables.

The residuals were calculated by subtracting the predicted height values from the observed height values using the model comparison dataset. The following lack-of-fit statistics were used to compare the two models: the H-D models with and without auxiliary variables, where Y_{ijk} and \hat{Y}_{ijk} are the actual and predicted height measurement k on the i^{th} tree in the j^{th} plot (subject), n_i is the number of trees in the j^{th} plot, m is the number of plots:

- 1) Mean Error (ME) = $\frac{1}{m} \sum_{j=1}^m \left[\frac{1}{n_i} \sum_{i=1}^{n_i} (Y_{ijk} - \hat{Y}_{ijk}) \right]$,
- 2) Mean Absolute Error (MAE) = $\frac{1}{m} \sum_{j=1}^m \left[\frac{1}{n_i} \sum_{i=1}^{n_i} |Y_{ijk} - \hat{Y}_{ijk}| \right]$, and
- 3) Residual Mean Squared Error (RMSE) = $\frac{1}{m} \sum_{j=1}^m \sqrt{\frac{1}{n_i} \sum_{i=1}^{n_i} (Y_{ijk} - \hat{Y}_{ijk})^2}$. If the result is a large negative value, the model is consistently over estimating. If the result is a large positive value, the model is consistently under estimating.

Results:

The Chapman-Richards (1959) model form was selected as the final model because it is biologically sound and superior in terms of model fit characteristics and less bias in estimating height.

Comparison between H-D models with and without plot-specific auxiliary variables:

To address the second objective of the study, auxiliary variables were added to the selected model form. The Chapman-Richards model form with AVGHT, PLOTBAM, LAT, and LONG resulted in the best overall predictive performance. The relative importance of an auxiliary variable is assessed by the change in the size of the standard error of prediction without the auxiliary variable in question (see, e.g., Uzoh and Oliver, 2006; Uzoh and Mori, 2012). Table 5 shows the ranking of the auxiliary variables based on this criterion. When tree diameter alone was used to estimate regional H-D model (equation 14) the RMSE was 9.5682. As shown in Table 5, the addition of AVGHT to the model reduced RMSE by 59%. The addition of PLOTBAM to the model reduced RMSE by 23%. A reduction of 23% was gained when LONG was added. The addition of LAT reduced RMSE by 19%. The final RMSE value with tree diameter and the four auxiliary variables reduced RMSE by over 60 percent (from 9.5682 to 3.638732).

The comparison shows that the residual mean squares error (RMSE) improved from 9.5682 to 3.638732, the mean absolute error (MAE) improved from 2.3133 to 1.45571, and the mean error (ME) improved from 1.8856 to -0.17578 (Table 6). The comparison by province results are presented in Table 7. Again, the H-D model with auxiliary variables outperformed the H-D model without auxiliary variables irrespective of geographic location. Figures 4 and 5 shows the scatter plots of standardized residuals against predicted tree height and standardized residuals histogram using the model building and model comparison datasets respectively. In both Figures, there was no evidence of apparent bias in the residuals, with 99.9% of residuals within ± 2 m.

The development of the final model used the entire dataset (model 18):

$$H_{ijk} = 1.3 + \beta_1 * (1 - \exp(-\beta_2 D_{ijk}))^{\beta_3} * \text{AVGHT}_{ik}^{\beta_4} * \text{PLOTBAM}_{ik}^{\beta_5} + \beta_6 * \text{LAT}_{ik} + \beta_7 * \text{LONG}_{ik} + h_l + e_{j(l)} + e_{ik(jl)} \quad (18)$$

The random effects were trees plots, and location. Where HT_{ijk} and D_{ijk} are the height and diameter measurement k on the i^{th} tree in the j^{th} plot (subjects), $i = 1, 2, \dots, n_i$, $j = 1, 2, \dots, m$, n_i is the number of trees in plot j , m is the total number of plots in the population; AVGHT_{ik} is the average of top height trees in the plot; PLOTBAM_{ik} is the plot basal area; LAT_{ik} and LONG_{ik} are the plot latitude and longitude; $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$, and β_7 are the fixed parameters to all plots; $\beta's \sim N(0, \sigma_e^2)$; h_l is the random effect of the l -th location with h_l assumed to have an expected value of zero (0) and variance σ_L^2 ; $e_{j(l)}$ is a random error for plot j within locale l assumed to have an expected value of zero (0) and variance ($\sigma_{P(L)}^2$); and $e_{ik(jl)}$ is a random error for measurement k on tree i within plot j and locale l assumed to have an expected value of zero (0) and variance (σ^2) with the covariance between observations k and k' on the same tree separated by d years following an autoregressive process:

$$\text{Cov}(e_{ik(jl)}, e_{ik'(jl)}) = \begin{cases} \sigma^2 \rho^{|d|}, & \text{if } i=i', k \neq k', j=j', l=l' \\ \sigma^2, & \text{if } i=i', k=k', j=j', l=l' \\ 0, & \text{otherwise,} \end{cases} \quad (15)$$

where ρ is the spatial power (SP (POW)) serial correlation coefficient for errors across time on the same tree.

As shown in Table 8, the variance components were partitioned correctly and their estimate are as follows: $\sigma_{Plot(Locale)}^2 = 0.9480$ (the variance associated with plot), $\sigma_{Locale}^2 = 2.0458$ (the variance associated with location), $\rho_{Serial\ correlation\ coefficient} = 0.9731$, and $\sigma_{Residual}^2 = 0.07166$. Figure 6 shows the function plots of the full model plus auxiliary variables. Prediction can be made as follows (Table 8):

$$HT_{ijk} = 1.3 + 5.0776 * ((1 - \exp(-0.01965 * DBH)) ** 1.093) * AVGHT ** 0.4059 * PLOTBAM ** 0.2772 + 0.1825 * LAT + (-0.05429 * LONG) \quad (19)$$

Residual plots were constructed for Eq. (18) for visual comparison (Figs. 4 and 5) using model building and comparison datasets respectively. These plots show that the predictive ability of the model is better for taller trees because NLMMs are asymptotically unbiased (see, e.g., Box, 1971; Fang and Bailey, 1998). Figure 6 shows the function plots of the full model plus auxiliary variables. The function plots exhibited sigmoid or S-shaped properties, which shows the three characteristics that should mark H-D relationships: (1) monotonic increment, (2) functional inflection point, and (3) asymptotic value (see, Yuancai and Parresol, 2001).

Table 5: Ranking of auxiliary variables based on change in the size of the residual mean squared error (RMSE) of prediction without the variable in question and the percent reduction in RMSE if the variable in question was included in the model using the model building dataset.

Variable	Value of RMSE of prediction without the variable in question	% Reduction in RMSE of prediction if the variable was included	Value of RMSE of prediction for the full model
AVGHT	7.4748	59%	3.042
LONG	4.1561	23%	
PLOTBAM	3.9437	23%	
LAT	3.769	19%	

Table 6: Comparison between Height-Diameter models with and without auxiliary variables using residual mean squares error (RMSE), mean error (ME), and absolute mean error (AME) for the Individual tree Height-Diameter model using the model comparison dataset of size 10% of the total dataset.

Model	n	RMSE	ME	MAE
Without auxiliary variables	3961	9.5682	1.8856	2.3133
With auxiliary variables	3961	3.638732	-0.17578	1.45571

Table 7: Comparison between Height-Diameter models with and without auxiliary variables using residual mean squares error (PMSE), mean error (ME), and absolute mean error (AME) by province for the Individual tree Height-Diameter model using the model comparison dataset of size 10% of the total dataset.

Province	n	Model with Auxiliary Variables			Model without Auxiliary Variables		
		Residual M S Error	Mean Error	Absolute Mean Bias	Residual M S Error	Mean Error	Absolute Mean Bias
I	793	4.051	-0.2664	1.4372	12.762	1.8626	2.3978
II	1444	4.1856	-0.5553	1.6108	8.6976	1.6711	2.1495
III	539	4.155	-0.0386	1.5967	12.406	2.7272	2.9352
IV	347	1.5034	0.4368	0.9556	9.8505	2.7333	2.8135
V	538	2.1182	0.6554	1.1597	4.6566	1.2249	1.775

Discussion:

The Chapman Richards model form outperformed the other model forms examined (Table 4). The model comparison results (Tables 6 and 7) demonstrated that irrespective of geographic location, a regional H-D Model with variance components and plot-specific auxiliary variables that individualized the curve to capture the site-specific, the plot-specific, or other unit-specific characteristics of the growth pattern decisively outperformed regional H-D Model with variance components but without plot-specific auxiliary variables. The H-D model with auxiliary variables mean biases and absolute mean biases were closer to zero than for the H-D models without auxiliary variables and the residual mean square errors value was reduced by over 60 percent (9.5682 to 3.638732). The results show that the H-D model predictions are closer to independent dataset than the H-D models without plot-specific auxiliary variables. This finding confirms previous findings (see, e.g., Larsen and Hann 1987; Calama and Montero 2004).

Table 8: Parameter estimates and variance components for the Individual Tree Height-Diameter model for managed even-aged stands of ponderosa pine trees throughout the western United States using hierarchical nonlinear mixed model analysis procedure using the entire dataset.

Parameters	Estimates	Standard Error	T-Value	P-values (2-tail)	Lower	Upper
β_1	5.0776	0.2180	23.29	0.0001	4.6502	5.5049
β_2	0.01965	0.000669	29.37	0.0001	0.01834	0.02096
β_3	1.0930	0.01106	98.87	0.0001	1.0713	1.1147
β_4	0.4059	0.009387	43.24	0.0001	0.3875	0.4243
β_5	0.2772	0.002881	96.23	0.0001	0.2716	0.2829
β_6	0.1825	0.07100	2.57	0.0102	0.04336	0.3217
β_7	-0.05429	0.02442	-2.22	0.0262	-0.1022	-0.00642
Variance components		Estimate	Std. Error	Z Value	Pr >Z	
$\sigma^2_{Plot(Locale)}$		0.9480	0.09361	10.13	0.0001	
σ^2_{Locale}		2.0458	0.4986	4.10	0.0001	
$\rho_{Serial\ correlation\ coefficient}$		0.9731	0.008217	118.43	0.0001	
$\sigma^2_{Residual}$		0.07166	0.000516	138.91	0.0001	

Consequently, the inclusion of plot-specific auxiliary variables individualized the curve for each plot and location. Therefore, made H-D models applicable in similar plot conditions at the regional level. More importantly, model comparison by geographic location results showed that inclusion of auxiliary variables reduced the residual error of prediction at each geographic location by more than half (Table 7). Without question, results show that diameter alone may not adequately capture the effects of all these factors. For example, the plot basal area (PLOTBAM, m^2 / ha) is an objective measure of the number of trees and their sizes that reflects the competitive stress experienced by a given tree within a plot better than stand level estimates of the variable. The effect is more evident where density varies within stand. The addition of PLOTBAM reduced the value of RMSE by 23%.

The same issue applies to the effect of geographic location. For example, trees growing in the west Cascades or on gentle, south-facing aspects on the western slope of the northern Sierra Nevada will grow faster than trees growing in the Black Hills of South Dakota and the Blue Mountains of Oregon or western Montana due to the availability of more moisture. Few if any sites east of the Sierra Nevada or Cascade Range can equal the better sites on the Pacific slope to the west due to moisture availability. In moisture limiting environments, one of the fundamental factors that dictates a plant's success or failure is the presence or absence of readily available soil moisture in the rooting medium (e.g., McMinn, 1952; Daubenmire, 1968; Schippers, *et al.* 2015; Trouvé *et al.*, 2015). LAT and LONG represents geographic location effects. The coefficient of LONG is negative because generally speaking, as you go eastward from the west, height growth decreases gradually, because ponderosa pine sites becomes drier with lower soil moisture. On the other hand, the coefficient of LAT is not negative because LAT has a more negative effect on diameter growth than on height growth. Factors such as slope, aspect, elevation, latitude, and longitude generally have indirect effects on tree growth, by influencing moisture, temperature, light, and other chemical and physical agents of site characteristics (see, Uzoh and Oliver, 2008; Schippers, *et al.* 2015; Trouvé *et al.*, 2015). LONG and LAT reduced the RMSE by 23% and 19% respectively.

Similarly, the site productivity potential effect variable, AVGHT, was superior to SI and other site variables tested in terms of fit characteristics and predictive ability; as a result, an excellent proxy for site index (SI). This finding confirms the previous findings (see, e.g., Sharma and Zhang, 2004; Newton and Amponsah, 2007). Traditionally, in western forests, forest growth has been modeled using the yield table approach based on extrapolating field measurements of tree growth according to empirically generated equations (see, Husch *et al.*, 1972). The fundamental assumption is that factors that limit tree growth are integrated in SI equations and they are constant over time. However, the yield table approach can fall short when long-term extrapolations are

examined or when stand characteristics are changed by the impacts of management or disturbance that may result in irregular stands (see, e.g., Aber *et al.*, 1982; Monserud, 1984).

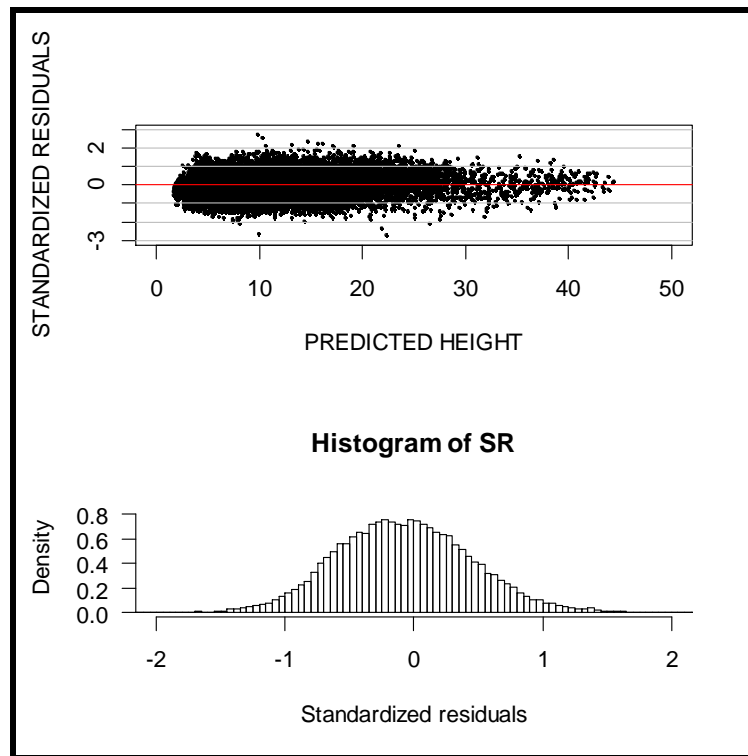


Fig. 4: Summary measures of goodness of fit for the Individual tree Height-Diameter model for managed even-aged stands of ponderosa pine trees throughout the western United States using hierarchical nonlinear mixed model with auxiliary variables using the model building dataset: scatter plot of standardized residuals against predicted tree height and standardized residuals histogram.

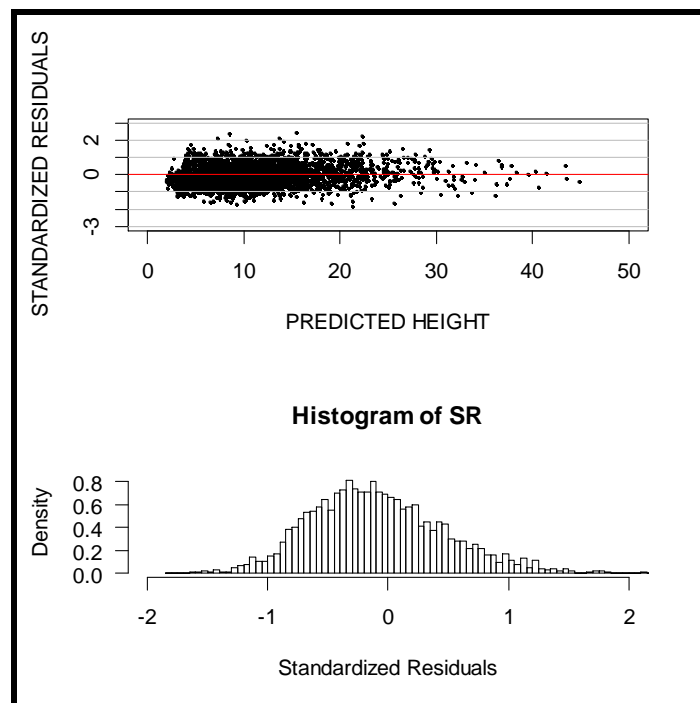


Fig. 5: Summary measures of goodness of fit for the Individual tree Height-Diameter model for managed even-aged stands of ponderosa pine trees throughout the western United States using hierarchical nonlinear mixed model with auxiliary variables using the model comparison dataset: scatter plot of standardized residuals against predicted tree height and standardized residuals histogram.

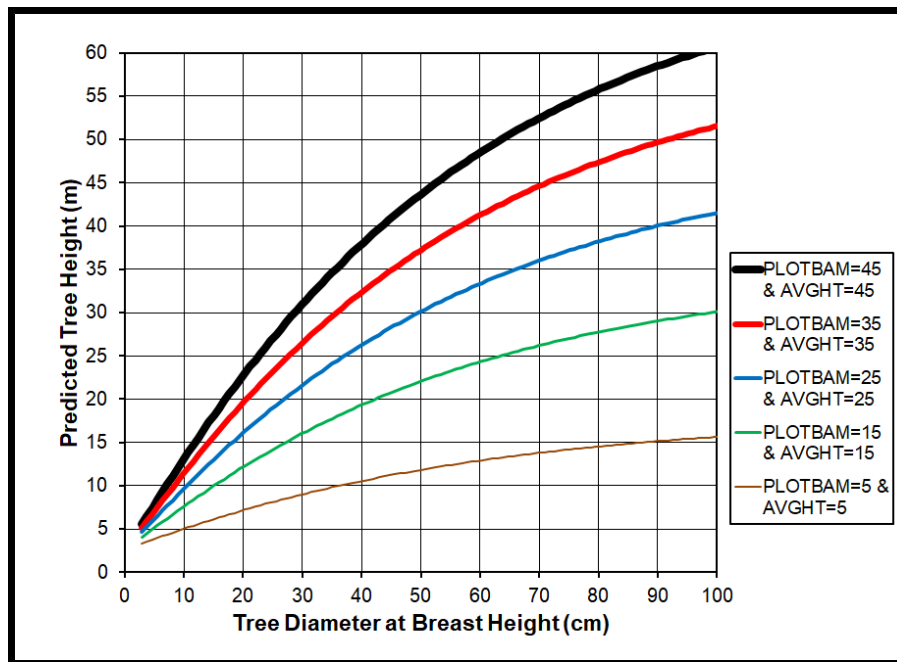


Fig. 6: Predicted tree heights (fixed responses) as a function of DBH for a given sets of measured covariates: plot basal area (PLOTBAM) at 5, 15, 15, 35, and 45 (m^2 / ha), latitude (LAT) at 42 and longitude (LONG) at = 115 and average of top height trees in the plot (AVGHT) at 5, 15, 25, 35, and 45 meters.

In addition, SI of the previous stand or an adjacent stand may not be a reliable choice. Specifically, SI for a plantation may not be the same as the previous stand because of genetic differences, changes in species composition, or treatments applied to enhance growth and development of young stands (see, Monserud, 1984). As a result, forest managers and researchers use average of the top height trees in the plot (AVGHT) because it is less affected by the impacts of forest management or disturbance (see, e.g., Sharma and Parton, 2007; Uzoh and Mori, 2012). The inclusion of AVGHT reduced RMSE by 59%. Again, diameter alone may not adequately capture this effect because it is a reflection of factors that limit growth, rather than the factors themselves. Fundamentally, auxiliary variables (such as competition, site, and geographic location effects) are plot special characteristics that modified tree growth while retaining the basic relationship between height and diameter growth. As a result, their addition improves the performance of H-D models and makes them applicable in similar biological and plot structural conditions at the regional level.

Conclusions:

The uniqueness of this study is that the datasets were from diverse ecological zones of ponderosa pine trees that represented a vast ecoregion of contrasting soil, landforms, geology, and climate, and included two varieties of ponderosa pine (*P. ponderosa* var. *ponderosa* and var. *scopulorum*). The statistical methodology used to analyze the dataset further demonstrated the uniqueness of this study. Specifically, repeated measurements are fundamental to most ecological and biological research and the datasets used in these studies come from diverse locations. However, analysis procedures used to analyze these datasets does not take into account the error structure imposed by the sampling schemes; consequently, this may result in spurious conclusions. This article discussed the proper methodology for analyzing such datasets. In addition, the hierarchical nonlinear mixed-effects model methodology used in this study correctly modeled the error structure imposed by the sampling scheme. Furthermore, the inclusion of both fixed-effect components that described the shape of the typical growth curve over the entire population, and the random-effect components that provided the plot-specific or other characteristics of H-D relationships improved model performance. Finally, the inclusion of the plot-specific auxiliary variables individualized the curves for each plot and location. As a result, made H-D models applicable in similar plot conditions at the regional level.

Consequently, the model developed herein should enhance comparisons of H-D relationships that may be due to geographic variation of the species throughout the range of ponderosa pine trees in the western United States. Therefore, it would provide forest managers and researchers with useful guidelines for a variety of management objectives at the regional scale.

ACKNOWLEDGMENTS

I am indebted to the following persons for establishing, maintaining, or providing the data from the permanent plot studies that made this endeavor possible: James Barrett, Patrick Cochran, and Douglas Maguire for Oregon and Washington; John Byrne and George Lightner for Montana; Wayne Shepperd and Lance Ashern for South Dakota; Frank Ronco and Carleton Edminster for Arizona; and, finally, Bill Oliver for California, who also assembled the dataset. Dr. Phil Weatherspoon's review greatly strengthened the clarity of this paper. My thanks to Dr. Nels Johnson for his statistical review. Finally, my thanks to the anonymous reviewers for their constructive comments and suggestions.

REFERENCES

- Aber, J.D., J.M. Melillo and C.A. Federer, 1982. Predicting the effects of rotation length, harvest intensity, and fertilization on fiber yield from northern hardwood forests in New England. *Forest Science*, 28: 31-45.
- Beal, S.L. and L.B. Sheiner, 1982. Estimating population kinetics. *CRC Critical Reviews in Biomedical Engineering*, 8: 195-222.
- Barrett, J.W., 1983. Growth of ponderosa pine poles thinned to different stocking levels in central Oregon. USDA Forest Service, Research Paper PNW-311.
- Bertalanffy, L.v., 1949. Problems of organic growth. *Nature*, 163: 156-158.
- Bertalanffy, L.v., 1957. Quantitative laws in metabolism and growth. *Quarterly Review of Biology*, 32: 217-231.
- Biging, G.S., 1985. Improved estimates of site index curves using a varying parameter model. *For. Sci.*, 31: 248-259.
- Boldt, C.E., and J.L. Van Deusen, 1974. Silviculture of ponderosa pine in the Black Hills: the status of our knowledge. USDA Forest Service Research Paper RM-124. p: 45.
- Box, M.J., 1971. Bias in nonlinear estimation. *Journal of the Royal Statistical Society Series B Statistical Methodological.*, 33(2): 171-201.
- Burnham, K.P., and D.R. Anderson, 2002. Model selection and multimodel inference: A practical information-theoretic approach, 2nd Edition. Springer, New York.
- Calama, R., and G. Montero, 2004. Interregional nonlinear height–diameter model with random coefficients for stone pine in Spain. *Canadian Journal of Forest Research*, 34: 150-163.
- Cochran, P.H., and J.W. Barrett, 1995. Growth and Mortality of Ponderosa Pine Poles Thinned to Various Densities in the Blue Mountains of Oregon. USDA Forest Service Research Paper PNW 483. US Department of Agriculture, Pacific Northwest Research Station, Portland, OR. USA.
- Cochran, P.H., and J.W. Barrett, 1999. Growth of ponderosa pine thinned to different stocking levels in central Oregon: 30-year results. USDA Forest Service Research Paper PNW 508. US Department of Agriculture, Pacific Northwest Research Station, Portland, OR. USA.
- Cortini, F, C.N., Filipescu, A., Groot, D.A. MacIsaac and T. Nunifu, 2011. Regional models of diameter as a function of individual tree attributes, climate and site characteristics for six major tree species in Alberta, Canada. *Forests*, 2: 814-831.
- Crecente-Campo, F., M. Tomé, P. Soares and U. Dieguez-Aranda, 2010. A generalized nonlinear mixed-effects height-diameter model for Eucalyptus globulus L. in northwestern Spain. *Forest Ecology Management*, 259(5): 943-952.
- Curtis, R.O., 1967. Height–diameter and height–diameter–age equations for second growth Douglas-fir. *Forest Science*, 13: 365-375.
- Daniel, T.W., J.A. Helms and F.S. Baker, 1979. Principles of Silviculture, second ed. McGraw-Hill, New York.
- Daubenmire, R., 1968. Structure and ecology of coniferous forest of the Northern Rocky Mountain. In: Taber, R.D. (Ed.), Proceedings of Coniferous Forest of Northern Rocky Mountain, University of Montana Foundation, Missoula USA. pp: 25-41.
- David, H.C., S.P. Netto, J.E. Arce, A.P.D. Corte, A.M. Filho and A. Behling, 2015. Updating of Dominant Height Growth Modeling and site Index of Pinus taeda L. in southern Brazil., 9(2): 115-125.
- Davidian, M. and D.M. Giltinan, 1995. *Nonlinear models for repeated measurement data*. Chapman & Hall, New York, p: 360.
- Davidian, M. and D.M. Giltinan, 2003. Nonlinear models for repeated measurement data: an overview and update. *J. Agricultural, Biological, and Environmental Statistics*, 8: 387-419.
- Dubrasich, M.E., D.W. Hann and J.C. Tappeiner, 1997. Methods for evaluating crown area profile of forest stands. *Canadian Journal of Forest Research*, 27: 385-392. doi: 10:1139/x96-206.
- Fang, Z., and R.L. Bailey, 1998. Height–diameter models for tropical forest on Hainan Island in southern China. *Forest Ecology and Management*, 110: 315-327.

Fortin, M., G. Daigle, C.H. Ung, J. Bégin and L. Archambault, 2007. A variance–covariance structure to take into account repeated measurements and heteroscedasticity in growth modeling. *European Journal of Forest Research*, 126: 573-585.

Gomperz, B., 1825. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions Royal Society London. Biological Science*, 115: 513-585.

Grégoire, T.G., O. Schabenberger and J.P. Barrett, 1995. Linear modeling of irregularly spaced, unbalanced, longitudinal data from permanent plot measurement. *Canadian Journal Forest Research*, 25(1): 137-156. doi:10.1139/x95-017.

Grégoire, T.G., O. Schabenberger, and J.P. Barrett, 1995. Linear modeling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. *Canadian Journal Forest Research*, 25: 137-156.

Gutzwiller, K.J. and S.K. Riffell, 2007. Using statistical models to study temporal dynamics of animal-landscape relations. In: Bissonette, J.A., Storch, I. (Eds.), *Temporal Dimensions of Landscape Ecology: Wildlife Responses to Variable Resources*. Springer-Verlag, New York.

Hall, D.B. and R. L. Bailey, 2001. Modeling and prediction of forest growth variables based on multilevel nonlinear mixed models. *Forest Science*, 47(3): 311-321.

Hasenauer, H. and R.A. Monserud, 1997. Biased prediction for tree height increment models developed from smoothed 'data'. *Ecology Modeling*, 98: 13-22.

Hastie, T., R. Tibshirani and J. Friedman, 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York.

Huang, S., S.J. Titus and D.P. Wiens, 1992. Comparison of nonlinear height–diameter functions for major Alberta tree species. *Canadian Journal Forest Research*, 22: 1297-1304.

Huang, S., D. Price, and S.J. Titus, 2000. Development of ecoregion-based height–diameter models for white spruce in boreal forests. *Forest Ecology and Management*, 129: 125-141.

Huang, S., S.X. Meng and Y. Yang, 2009. Using nonlinear mixed model technique to determine the optimal tree height prediction model for black spruce. *Modern Applied Science*, 3: 3-18.

Husch, B., C.I. Miller and T.W. Beers, 1972. *Forest Mensuration*, third ed. Ronald Press Co., New York, NY, p: 402.

Johansson, T., 2012. Site index curves for young hybrid larch growing on former farmland in Sweden. *Forests*, 3: 723-735.

Lappi, J., 1997. A longitudinal analysis of height/diameter curves, *Forest science*, 43(4): 555-570.

Larsen, D.R., and D.W. Hann, 1987. Height-diameter Equations for Seventeen Tree Species in Southwest Oregon. *Forest Research Laboratory, Oregon State University, Corvallis, USA. Research Bulletin 49*.

LeMay, V.M., 1990. MSLS: a linear least squares technique for fitting a simultaneous system of equations with a generalized error structure. *Canadian Journal Forest Research*, 20(12): 1830-1839. doi:10.1139/x90-246.

Littell, R.C., G.A. Milliken, W.W. Stroup, R.D. Wolfinger, and O. Schabenberger, 2006. *SAS for Mixed Models*. Second Ed. SAS Institute, Cary, NC.

Lopez, S.C.A., J.G. Varela, F.C. Dorado, A.R. Alboreca, R.R. Soalleiro, J.G. Alvarez Gonzalez and F.S. Rodriguez, 2003. A height–diameter model for *Pinus radiata* D. Don in Galicia (Northwest Spain). *Annals Forest Science*, 60: 237-245.

McMinn, R.G., 1952. The role of soil drought in the distribution of vegetation in the northern Rocky Mountains. *Ecology*, 33: 1-15.

Mehtätalo, L., S. de-Miguel, and T.G. Gregoire, 2015. Modeling height-diameter curves for prediction. *Canadian Journal of Forest Research*, 45(7): 826-837, 10.1139/cjfr-2015-0054.

Meng, S. X., and S. Huang, 2010. Incorporating correlated error structures into mixed forest growth models: prediction and inference implications. *Canadian Journal of Forest Research*, 40(5): 977-990.

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

Meyer, W.H., 1938. Yield of Even-Aged Stands of Ponderosa Pine. *US Department of Agriculture Technical Bulletin 630*.

Myers, C.A., 1967. Growing stock levels in even-aged ponderosa pine. *US Forest Service Research Paper RM 33*. US Department of Agriculture, Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colorado. USA.

Monleon, V.J., 2004. A hierarchical model for tree height prediction. *Proceedings of the 2003 Meeting of the American Statistical Association, Section on Statistics & the Environment, Alexandria, VA*. pp: 2865-2869.

Monserud, R.A., 1984. Problems with site index: an opinionated review. In: Bockheim, J.G. (Ed.), *Proceedings — Forest Land Classification Symposium*. Department of Soil Science. University of Wisconsin, Madison, WI, pp: 167-180.

Moore, J.A., L. Zhang and D. Stuck, 1996. Height-Diameter Equations for Ten Tree Species in the Inland Northwest. *WJAF* 11(4).

Newton, P.F. and I.G. Amponsah, 2007. Comparative evaluation of five height–diameter models developed for black spruce and jack pine stand-types in terms of goodness-of-fit, lack-of-fit and predictive ability. *Forest Ecology and Management*, 247: 149-166.

Oliver, W.W., 1984. Brush reduces growth of thinned ponderosa pine in northern California. US Forest Service Research Paper PSW-172. US Department of Agriculture, Pacific Southwest Research Station, Albany, CA. USA.

Oliver, W.W. and C.B. Edminster, 1988. Growth of ponderosa pine thinned to different stocking levels in the western United States. In *Proceedings — Future forests of the mountain Western*, September 29 - October 3, 1986, Missoula, Montana. Compiler by Schmidt, W.C. USDA Forest Service General Technical Report INT-243: 153-159.

Oliver, W.W., 1997. Twenty-five-year growth and mortality of planted ponderosa pine repeatedly thinned to different stand densities in northern California. *Western Journal Applied Forest*, 12(4): 122-130.

Oliver, W.W. and R.A. Ryker, 1990. *Pinus ponderosa Dougl. Ex Laws. Ponderosa Pine*. In: Burns, R.M., Honkala, B.H. (Technical Coordinators), *Silvics of North America Volume 1, Conifers*. USDA Forest Service Agricultural Hand Book. 654.

Osman, E.H., E.A. Idris and E.M. Ibrahim, 2013. Height-Diameter prediction models for some utilitarian natural tree species. *Journal of forest Products and industries*, 2(2): 31-39 ISSN: 2325.

Pearl, R. and L.J. Reed, 1920. On the rate of growth of the population of the united states since 1790 and its mathematical representation. *Proceedings National Academic Science U.S.A.* 6: 275-288. doi:10.1073/pnas.6.6.275.

Peng, C.H., L. Zhang, S. Huang, X. Zhou, J. Parton and M. Woods, 2001. Developing Ecoregion-based Height-Diameter Models for Jack Pine and Black Spruce in Ontario. *Forest Research Report 159*. Ministry of Natural Resources, Ontario Forest Research Institute, Ontario, Canada. pp: 10.

Rawlings, J.O., S.G. Pantula, and D.A. Dickey, 1998. *Applied regression analysis: A research tool*, 2nd Edition. Springer-Verlag, New York.

Richards, F.J., 1959. A flexible growth function for empirical use. *Journal of Experimental Botany*, 10: 290-300.

Ritchie, M.W., 1999. A compendium of forest growth and yield simulators for the Pacific Coast States. US Forest Service, General Technical Report PSW 174. US Department of Agriculture, Pacific Southwest Research Station, Albany, CA.

Robinson, A.P., and W.R. Wykoff, 2004. Imputing missing height measures using a mixed-effects modeling strategy. *Canadian Journal Forest Research*, 34: 2492-2500.

Ronco, Jr., F., C.B. Edminster and D.P. Trujillo, 1985. Growth of ponderosa pine thinned to different stocking levels in northern Arizona. Res. Pap. RM-262. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station, p: 15.

SAS Software, 2014. Version 9.4, SAS Institute Inc., Cary, NC. USA.

SAS User's Guide, 2015. SAS Institute Inc., Cary, NC. USA. Example 70.7 Overdispersion Hierarchical Nonlinear Mixed-Effects Model in SAS/STAT(R) 13.2 User's Guide. Available at:

http://support.sas.com/documentation/cdl/en/statug/67523/HTML/default/viewer.htm#statug_nlmixed_examples07.htm

Schippers, P., F. Sterck, M. Vlam and P.A. Zuidema, 2015. Tree growth variation in the tropical forest: understanding effects of temperature, rainfall and CO₂. *Global change Biology*, 21(7): 2749-2761. <http://dx.doi.org/10.1111/gcb.12877>.

Sharma, M. and J. Parton, 2007. Height-diameter equations for boreal tree species in Ontario using a mixed-effects modeling approach. *Forest Ecology Management*, 249: 187-198.

Spurr, S.H. and B.V. Barnes, 1980. *Forest Ecology*, third ed. Wiley, New York, 687.

Stage, A.R., 1973. *Prognosis Model for Stand Development*. Research Paper INT-137. USDA Forest Service, Intermountain Forest and Range Experiment Station, Ogden, UT. USA. pp: 32.

Stage, A.R., 1975. Prediction of height increment for models of forest growth. Research Paper INT-164. USDA Forest Service, Intermountain Forest and Range Experiment Station, Ogden, UT. USA.

Stage, A.R., 1976. An expression for the effect of slope, aspect and habitat type on tree growth. *Forest Science*, 22(4): 457-460.

Stage, A.R., C. Salas, 2007. Interactions of Elevation, Aspect, and Slope in Models of Forest Species Composition and Productivity. *For. Sci.*, 53: 486-492.

Stoffels, A., and J. van Soest, 1953. The main problems in sample plots. *Ned. Boschb. Tijdschr*, 25: 190-199.

Stout, B.B. and D.L. Shumway, 1982. Site quality estimation using height and diameter. *Forest Science*, 28: 639-645.

Temesgen, H. and K.V. Gadow, 2004. Generalized height–diameter models—an application for major tree species in complex stands of interior British Columbia. *European Journal Forest Research*, 123: 45-51.

Trincado, G., C.L. VanderSchaaf and H.E. Burkhart, 2007. Regional mixed-effects height–diameter models for loblolly pine (*Pinus taeda* L.) plantations. *European Journal of Forest Research*, 126: 253-262.

Trouvé, R., J.D. Bontemps, I. Seynave, C. Collet and F. Lebourgeois, 2015. Stand density, tree social status and water stress influence allocation in height and diameter growth of *Quercus petraea* (Liebl.). *Tree physiology*, 35(10): 1035-1046. <http://dx.doi.org/10.1093/treephys/tpv067>.

Uzoh, F.C.C. and W.W. Oliver, 2006. Individual tree height increment model for managed even-aged stands of ponderosa pine throughout the western United States using linear mixed effects models. *Forest Ecology Management*, 221: 147-154.

Uzoh, F.C.C. and W.W. Oliver, 2008. Individual tree diameter increment model for managed even-aged stands of ponderosa pine throughout the western United States using a multilevel linear mixed effects model. *Forest Ecology Management*, 256(3): 438-445.

Uzoh, F.C.C. and S.R. Mori, 2012. Applying survival analysis to managed even-aged stands of ponderosa pine for assessment of tree mortality in the western United States. *Forest Ecology Management*, 285: 101-122.

Vanclay, J.K., 1994. *Modeling Forest Growth and Yield: Applications to Mixed Tropical Forests*. CAB International, Wallingford.

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

Weibull, W., 1951. A statistical distribution function of wide applicability. *Journal Applied Mechanics Transactions ASME*, 18(3): 293-297.

Wyckoff, W.R., N.L. Crookston and A.R. Stage, 1982. User's guide to the stand prognosis model. US Forest Service General Technical Report INT 133. US Department of Agriculture, Intermountain Forest and Range Experiment Station, Ogden, UT. USA.

Wyckoff, W.R., 1990. A basal area increment model for individual conifers in the northern Rocky Mountains. *For. Sci.*, 36(4): 1077-1104.

Wyckoff, W.R., N.L. Crookston and A.R. Stage, 1982. User's guide to the stand prognosis model. US Forest Service General Technical Report INT 133. US Department of Agriculture, Intermountain Forest and Range Experiment Station, Ogden, UT. USA.

Wyckoff, W.R., 1990. A basal area increment model for individual conifers in the northern Rocky Mountains. *For. Sci.*, 36(4): 1077-1104.

Yuancai, L. and B.R. Parresol, 2001. Remarks on Height–Diameter Modelling. USDA Forest Service, Research Note SRS-10, Southern Research Station, Asheville, NC, USA. pp: 5.

Zakrzewski, W.T. and I.E. Bella, 1988. Two new height models for volume estimation of lodgepole pine stands. *Canadian Journal Forest Research*, 18: 195-201.

Zeide, B. and C. Vanderschaaf, 2002. The effect of density on the height–diameter relationship. In: Outcalt, K.W. (ed). *Proceedings of the 11th Biennial Southern Silvicultural Research Conference*. 2001 March 20–22, Knoxville, TN. USA. USDA Forest Service, General Technical Report SRS–48, Asheville, NC, pp: 463-466.