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

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), and 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 enable 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.
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, and they are not constant over time. 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 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 5 years. Eighty-two percent of the plots were observed for 20 years or more—at least four 5-year 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 (Gompertz, 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.
Table 1: Location and literature citations for five levels-of-growing-stock installations in ponderosa pine stands throughout the western United States.

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

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.

<table>
<thead>
<tr>
<th>Province</th>
<th>Number of Plots</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stand Origin</td>
<td>Natural</td>
<td>7</td>
<td>81</td>
<td>32</td>
<td>42</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Planted</td>
<td>92</td>
<td>18</td>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Stand Size Class</td>
<td>Saplings</td>
<td>32</td>
<td>30</td>
<td>9</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Poles</td>
<td>58</td>
<td>61</td>
<td>38</td>
<td>27</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Sawtimber</td>
<td>9</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>99</td>
<td>99</td>
<td>47</td>
<td>42</td>
<td>18</td>
</tr>
</tbody>
</table>
\[
\begin{align*}
\text{Table 3:} & \text{ Summary statistics for variables used to develop the Individual tree Height-Diameter model for managed even-aged stands of ponderosa pine trees for the western United States using hierarchical nonlinear mixed-effects model.} \\
\hline 
\text{Variables} & \text{Number of Trees} & \text{Mean} & \text{Minimum} & \text{Maximum} \\
\hline 
\text{HTM (m)} & 8208 & 10.608 & 0.457 & 53.214 \\
\text{DBHCM (cm)} & 8208 & 18.89 & 1.02 & 98.04 \\
\hline 
\text{Plot Variables} & \text{Number of Plots} & \text{Mean} & \text{Minimum} & \text{Maximum} \\
\hline 
\text{SLOPE (percent)} & 305 & 6.16 & 0 & 42 \\
\text{ASPECT (degrees)} & 305 & 113.07 & 0 & 315 \\
\text{Stand age (years)} & 305 & 66.45 & 8 & 110 \\
\text{SIMM (m)} & 305 & 21.62 & 13.11 & 48.77 \\
\text{AVGHT (m)} & 305 & 18.7166 & 4.91947 & 51.0243 \\
\text{PLOTBAM (m² ha⁻¹)} & 305 & 0.003716 & 0.003716 & 52.37 \\
\text{SDIM (trees ha⁻¹)} & 305 & 576.43 & 21.18 & 1,444.22 \\
\text{TPH (trees ha⁻¹)} & 305 & 1,553.29 & 44.48 & 17,173.73 \\
\text{ELEVA (m)} & 305 & 1503.79 & 716.28 & 2266.19 \\
\text{LATITUDE (degrees)} & 305 & 42.22 & 35.282 & 48.52 \\
\text{LONGITUDE (degrees)} & 305 & 117.95 & 103.6 & 122.67 \\
\hline 
\end{align*}
\]

\[
\begin{align*}
\text{Table 4:} & \text{ Height-Diameter functions selected for evaluation.} \\
\hline 
\text{Model Name} & \text{Model forms} & \text{AIC} & \text{RMSE} \\
\hline 
2. Curtis (1967) & \text{HT}_{ij} = 1.3 + \beta_1 \cdot \exp(-\beta_2 \cdot \text{DBH}_{ijk}) + \epsilon_{ijk} & 162464 & 6.0210 \\
3. Gomperz (1825) & \text{HT}_{ij} = 1.3 + \beta_1 \cdot \exp(-\beta_2 \cdot \text{DBH}_{ijk}^{-\beta_3}) + \epsilon_{ijk} & 162813 & 6.0153 \\
4. Weibull (1951) & \text{HT}_{ij} = 1.3 + \beta_1 \cdot (1 - \exp(-\beta_2 \cdot \text{DBH}_{ijk}^{-\beta_3})) + \epsilon_{ijk} & 162835 & 6.3319 \\
5. Chapman-Richards (1959) & \text{HT}_{ij} = 1.3 + \beta_1 \cdot (1 - \exp(-\beta_2 \cdot \text{DBH}_{ijk}^{-\beta_3})) + \epsilon_{ijk} & 162836 & 6.0200 \\
6. Stoffels and van Soest (1953) & \text{HT}_{ij} = 1.3 + \beta_1 \cdot \text{DBH}_{ijk}^{-\beta_2} + \epsilon_{ijk} & 162841 & 6.0210 \\
7. Curtis (1967) & \text{HT}_{ij} = 1.3 + \frac{\beta_1 \cdot \text{DBH}_{ijk}^{-\beta_2}}{(1 + \text{DBH}_{ijk}^{-\beta_2})^{-\beta_3}} + \epsilon_{ijk} & 162841 & 6.0210 \\
8. Stage (1975) & \text{HT}_{ij} = 1.3 + \beta_1 \cdot \text{DBH}_{ijk}^{-\beta_2} + \epsilon_{ijk} & 162841 & 6.0210 \\
9. Larsen and Hamb (1987) & \text{HT}_{ij} = 1.3 + \exp(\beta_1 + \beta_2 \cdot \text{DBH}_{ijk}^{-\beta_3}) + \epsilon_{ijk} & 162851 & 9.4250 \\
10. Meyer (1940) & \text{HT}_{ij} = 1.3 + \beta_1 \cdot (1 - \exp(-\beta_2 \cdot \text{DBH}_{ijk}^{-\beta_3})) + \epsilon_{ijk} & 163147 & 12.738 \\
11. Pear and Reed (1920) & \text{HT}_{ij} = 1.3 + \frac{\beta_1 + \text{u}_{ij}}{1 + (\beta_2 + \text{u}_{ij}) \cdot \exp(-\beta_3 \cdot \text{DBH}_{ijk}^{-\beta_4})} + \epsilon_{ijk} & 166301 & 6.6623 \\
12. Wykoff et al. 1982 & \text{HT}_{ij} = 1.3 + \exp \left( \beta_1 + \frac{\beta_2}{\text{DBH}_{ijk} + 1} \right) + \epsilon_{ijk} & 166755 & 6.7802 \\
\hline 
\end{align*}
\]

**Nonlinear Mixed-Effects Model (NLMM) Development:**

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

\[
y_{ij} = f(x_{ij}, \beta, b_j) + \epsilon_{ij} \quad (i=1,2,...,m; \ j=1,2,...,n_j)
\]

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, \( \beta \) is a \((p \times 1)\) vector of unknown fixed-parameters common to all subjects, \( b_j \) is a \((q \times 1)\) vector of unknown random parameters unique for the \( j \)th subject, and \( \epsilon_{ij} \) is the unknown within-subject random error. Detailed
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., 2014), 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:

\[
H_{ijk} = 1.3 + \beta_1 \exp(-\beta_2 D_{ijk}) + h_I + e_{j(l)} + e_{ik(j)}
\]  
(13)

Chapman-Richards (1959) \[
H_{ijk} = 1.3 + \beta_1 (1 - \exp(-\beta_2 D_{ijk}))^{\beta_3} + h_I + e_{j(l)} + e_{ik(j)}
\]  
(14)

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

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

where \(\rho\) 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., Wykoff, 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, ((SI[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:

$$H_{ijk} = 1.3 + \beta_1 * (1 - \exp(-\beta_2 D_{ijk})^{\beta_3} * AVGHT_{ik}^{\beta_4} + h_i + e_{j(i)} + e_{ik(j)})$$ (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., Wykoff, 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:

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

where $\text{PLOTBAM}_{ik}$ 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:

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

where $\text{LAT}_{ik}$ and $\text{LONG}_{ik}$ are the plot latitude and longitude and other variables remain as previously defined.

The variance increased with increasing predicted values $y \sim \mathcal{N}(f(x), \sigma^2)$. Among the tested weights, weight of $1/(x^{**1.3})$ 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) = \[ \frac{1}{m} \sum_{j=1}^{m} \left( \frac{1}{n_i} \sum_{i=1}^{n_i} |Y_{ijk} - \hat{Y}_{ijk}| \right), \]

3) Residual Mean Squared Error (RMSE) = \[ \frac{1}{m} \sum_{j=1}^{m} \left( \frac{1}{n_i} \sum_{i=1}^{n_i} (Y_{ijk} - \hat{Y}_{ijk})^2 \right). \] 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 show 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):

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

The random effects were trees plots, and location. Where \(H_{ijk}\) and \(D_{ijk}\) are the height and diameter measurement \(k\) on the \(i^{th}\) tree in the \(j^{th}\) plot (subjects), \(i = 1, 2, \ldots n_i, j = 1, 2, \ldots m, n_i\) is the number of trees in plot \(j, m\) is the total number of plots in the population; \(\text{AVGHT}_{ik}\) is the average of top height trees in the plot; \(\text{PLOTBAM}_{ik}\) is the plot basal area; \(\text{LAT}_{ik}\) and \(\text{LONG}_{ik}\) are the plot latitude and longitude; \(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \text{ and } \beta_7\) are the fixed parameters to all plots; \(\beta' s \sim N(0, \sigma^2_e)\); \(h_l\) is the random effect of the \(l^{th}\) location with \(h_l\) assumed to have an expected value of zero (0) and variance \(\sigma^2_h\); \(e_{j(l)}\) is a random error for plot \(j\) within locale \(l\) assumed to have an expected value of zero (0) and variance \((\sigma^2_{P(L)})\); 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 \(\sigma^2\) with the covariance between observations \(k\) and \(k'\) on the same tree separated by \(d\) years following an autoregressive process:

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

where \(\rho\) 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: $\sigma^2_{\text{Plot}(\text{Locale})} = 0.9480$ (the variance associated with plot), $\sigma^2_{\text{Locale}} = 2.0458$ (the variance associated with location), $\rho = 0.9731$, and $\sigma^2_{\text{Residual}} = 0.07166$. Figure 6 shows the function plots of the full model plus auxiliary variables. Prediction can be made as follows (Table 8):

$$HT_{ijk} = 1.3 + 5.0776^* ((1 - \exp(-0.01965 * DBH))^{**1.093}) * \text{AVGHT}^{**0.4059} *$$

$$PLOTBAM^{**0.2772} + 0.1825^* \text{LAT} + (-0.05429^* \text{LONG})$$

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.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value of RMSE of prediction without the variable in question</th>
<th>% Reduction in RMSE of prediction if the variable was included</th>
<th>Value of RMSE of prediction for the full model</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVGHT</td>
<td>7.4748</td>
<td>59%</td>
<td>3.042</td>
</tr>
<tr>
<td>LONG</td>
<td>4.1561</td>
<td>23%</td>
<td></td>
</tr>
<tr>
<td>PLOTBAM</td>
<td>3.9437</td>
<td>23%</td>
<td></td>
</tr>
<tr>
<td>LAT</td>
<td>3.769</td>
<td>19%</td>
<td></td>
</tr>
</tbody>
</table>

**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.

<table>
<thead>
<tr>
<th>Model</th>
<th>n</th>
<th>RMSE</th>
<th>ME</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without auxiliary variables</td>
<td>3961</td>
<td>9.5682</td>
<td>1.8856</td>
<td>2.3133</td>
</tr>
<tr>
<td>With auxiliary variables</td>
<td>3961</td>
<td>3.638732</td>
<td>-0.17578</td>
<td>1.45571</td>
</tr>
</tbody>
</table>

**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.

<table>
<thead>
<tr>
<th>Province</th>
<th>n</th>
<th>Residual M S Error</th>
<th>Mean Error</th>
<th>Absolute Mean Bias</th>
<th>Residual M S Error</th>
<th>Mean Error</th>
<th>Absolute Mean Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>793</td>
<td>4.051</td>
<td>-0.2664</td>
<td>1.4372</td>
<td>12.762</td>
<td>1.8626</td>
<td>2.3978</td>
</tr>
<tr>
<td>II</td>
<td>1444</td>
<td>4.1856</td>
<td>-0.5553</td>
<td>1.6108</td>
<td>8.6976</td>
<td>1.6711</td>
<td>2.1495</td>
</tr>
<tr>
<td>III</td>
<td>539</td>
<td>4.155</td>
<td>-0.0386</td>
<td>1.5967</td>
<td>12.406</td>
<td>2.7272</td>
<td>2.9352</td>
</tr>
<tr>
<td>IV</td>
<td>347</td>
<td>1.5034</td>
<td>0.4368</td>
<td>0.9556</td>
<td>9.8505</td>
<td>2.7333</td>
<td>2.8135</td>
</tr>
<tr>
<td>V</td>
<td>538</td>
<td>2.1182</td>
<td>0.6554</td>
<td>1.1597</td>
<td>4.6566</td>
<td>1.2249</td>
<td>1.775</td>
</tr>
</tbody>
</table>

**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.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimates</th>
<th>Standard Error</th>
<th>T-Value</th>
<th>P-values (2-tail)</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>5.0776</td>
<td>0.2180</td>
<td>23.29</td>
<td>0.0001</td>
<td>4.6502</td>
<td>5.6049</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.01965</td>
<td>0.000669</td>
<td>29.37</td>
<td>0.0001</td>
<td>0.01834</td>
<td>0.02096</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>1.0930</td>
<td>0.01106</td>
<td>98.87</td>
<td>0.0001</td>
<td>1.0713</td>
<td>1.1147</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.4059</td>
<td>0.009387</td>
<td>43.24</td>
<td>0.0001</td>
<td>0.3875</td>
<td>0.4243</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>0.2772</td>
<td>0.002881</td>
<td>96.23</td>
<td>0.0001</td>
<td>0.2716</td>
<td>0.2829</td>
</tr>
<tr>
<td>$\beta_6$</td>
<td>0.1825</td>
<td>0.07100</td>
<td>2.57</td>
<td>0.0102</td>
<td>0.04336</td>
<td>0.3217</td>
</tr>
<tr>
<td>$\beta_7$</td>
<td>-0.05429</td>
<td>0.02442</td>
<td>-2.22</td>
<td>0.0262</td>
<td>-0.1022</td>
<td>-0.00642</td>
</tr>
</tbody>
</table>

Variance components

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Z Value</th>
<th>Pr &gt; Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_{\text{Plot (Locale)}}$</td>
<td>0.9480</td>
<td>0.09361</td>
<td>10.13</td>
<td>0.0001</td>
</tr>
<tr>
<td>$\sigma^2_{\text{Locale}}$</td>
<td>2.0458</td>
<td>0.4986</td>
<td>4.10</td>
<td>0.0001</td>
</tr>
<tr>
<td>$\rho_{\text{Serial correlation coefficient}}$</td>
<td>0.9731</td>
<td>0.008217</td>
<td>118.43</td>
<td>0.0001</td>
</tr>
<tr>
<td>$\sigma^2_{\text{Residual}}$</td>
<td>0.07166</td>
<td>0.000516</td>
<td>138.91</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

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 Ampomah, 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 Cochrane, 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


