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

Article *in* Forest Ecology and Management · July 2008

DOI: 10.1016/j.foreco.2008.04.046

Fabian C.C. Uzoh
US Forest Service

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Forest Ecology and Management 221 (2006) 147–154
Forest Ecology and Management
www.elsevier.com/locate/foreco

Individual tree height increment model for managed even-aged stands of ponderosa pine throughout the western United States using linear mixed effects models[☆]

Fabian C.C. Uzoh [*], William W. Oliver [1]

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

Received 17 December 2004; received in revised form 13 September 2005; accepted 14 September 2005

Abstract

A height increment model is developed and evaluated for individual trees of ponderosa pine throughout the species range in western United States. The data set used in this study came from long-term permanent research plots in even-aged, pure stands both planted and of natural origin. The data base consists of six levels-of-growing stock studies supplemented by initial spacing and other permanent-plot thinning studies for a total of 310 plots, 34,263 trees and 122,082 observations. Regression analysis is the most commonly used statistical method in forest modeling. However, research studies with repeated measurements are common in forestry and other biological disciplines. We choose the mixed models instead of the regression analysis approach because it allows for proper treatment of error terms in a repeated measures analysis. The model is well behaved and possessed desirable statistical properties. Our goal is to present a single height increment model applicable throughout the geographic range of ponderosa pine in the United States and by using only data from long-term permanent plots on sites capable of the productivity estimated by Meyer [Meyer, W.H., 1938. Yield of Even-aged Stands of Ponderosa Pine. US Department of Agriculture Technical Bull. 630].
Published by Elsevier B.V.

Keywords: Height growth; Tree-growth modeling; Repeated measures analysis; *Pinus ponderosa*

1. Introduction

Height increment data for modeling is usually obtained either by remeasured heights on standing trees or by stem analysis (stem analysis is the procedure used for determining past growth by directly measuring the accumulated stem increments of height and diameter). Of these two methods stem analysis is the most accurate, although very expensive. Remeasured heights are particularly difficult in mountainous terrain and in dense stands. Additionally, in boreal and many temperate forests height growth can be slow and height measurement error can be large relative to the increment (Hasenauer and Monserud, 1997).

Consequently, rather than measure heights of all trees, growth and yield modelers develop heuristic functions of diameter. Diameter at breast height (dbh) is obtained on all trees and heights are measured on a sub-sample. A heuristic function predicting height from diameter is developed using trees that have both height and diameter measurements. The height-diameter equation developed is then used to predict height for all trees (Meyer, 1940; Curtis, 1967; Curtis et al., 1981; Wykoff et al., 1982; Larsen and Hann, 1987; Burkhart et al., 1992). Historically, these equations have included only diameter as the independent variable (Curtis, 1967). More recently, however, other independent variables have been included such as, stand basal area, basal area in trees of larger diameter, site index, slope, aspect, elevation, crown competition factor, and soil and precipitation factors (Larsen and Hann, 1987; Hasenauer and Monserud, 1997; Uzoh, 2001).

Many regional growth and yield models are available for ponderosa pine (*Pinus ponderosa* Dougl.) (Ritchie, 1999). Since data collection and analysis procedures differ among these models, comparisons of growth responses that may be due to geographic variation of the species are not possible. Also,

* Corresponding author. Tel.: +1 530 226 2549; fax: +1 530 226 5091.
E-mail addresses: fuzoh@fs.fed.us (Fabian C.C. Uzoh), boliver@fs.fed.us (W.W. Oliver).
[1] Tel.: +1 530 226 2535; fax: +1 530 226 5091.

0378-1127/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.foreco.2005.09.012

Table 1
Location and literature citations for five levels-of-growing-stock installations in ponderosa pine in 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, OR	Cochran and Barrett (1999a)
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 et al. (1985)

data bases are often compiled all or in part from temporary plots in unmanaged stands 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. Therefore, forest managers need a height increment model that can provide useful guidelines for a variety of management objectives throughout the geographic range of ponderosa pine in the United States. We had the opportunity to develop such a height increment equation through access to a unique data base of long-term studies of levels-of-growing-stock, initial spacing and other permanent-plot thinning studies throughout the range of ponderosa pine in the United states. The objective of this study is to overcome these weaknesses by presenting a height increment model applicable throughout the geographic range of ponderosa pine in the United States and by using only data from long-term permanent plots on sites capable of normal yields (Meyer, 1938). An individual-tree/distance-independent height increment model for managed even-aged stands of ponderosa pine is presented using linear mixed effects models.

2. Methods

2.1. The data base

The foundation of the data base is six levels-of-growing-stock studies established throughout the western United States in the 1960s. All used a common study plan that specified five or six stand density levels replicated three times (Myers, 1967). Results from individual installations have been reported previously (Table 1). These data were supplemented with initial spacing and other permanent-plot thinning studies. Individual tree data were from plots in planted stands or stands of natural origin and included a wide range of size classes (Tables 2 and 3). Stands were free or mostly free of competing shrubs that reduce growth of young ponderosa pine, especially in central Oregon and California (Oliver, 1984, 1990; Cochran and Barrett, 1999b). Trees in all plots in the data base were tagged allowing the collection of information on individual trees. The number of growing seasons between remeasurements was usually five, but most plots were observed for a much longer period. Eighty-two percent of the plots were observed for 20 years or more—four 5-years growth periods. Basic records for each plot included latitude, elevation, aspect, slope percent and plot size. Tree records at each remeasurement included species, 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 2 in. diameter classes across the range of tree sizes. Height measurements were repeated on the same trees ensuring that the 5-year height increment is given by the difference between the two successive observations of height.

3. Data analysis

3.1. Theoretical formulation

Regardless of mathematical form, two approaches to model development are commonly used (Wykoff, 1990):

- Predict maximum potential increment as a function of site quality and tree maturity, then adjust potential with a modifier function that reflects intertree competition (Monserud, 1975; Alder, 1979; Hahn and Leary, 1979; Lear and Holdaway,

Table 2
Distribution of plots in each province by stand origin and tree size used to develop the height increment model for managed even-aged stands of ponderosa pine throughout the western United States

	Province				
	I	II	III	IV	V
Number of plots					
Stand origin					
Natural	11	71	33	42	18
Planted	95	26	14	0	0
Stand size class					
Saplings	31	10	0	0	0
Plots	64	83	47	42	18
Sawtimber	11	4	0	0	0

Table 3
Summary statistics for the data used to develop the height increment model for managed even-aged stands of ponderosa pine throughout western United States

Variable	No. of trees observed	Mean	S.D.	Minimum	Maximum
HT (m)	34263	10.797	4.542	0.034	53.214
dbh (cm)	34263	18.167	8.801	0.254	98.044
SI (m)	310	21.619	6.859	13.106	48.768
Age (year)	310	58.759	21.416	4.000	110.000
ELEVA (m)	310	41.789	4.303	35.280	48.500
BAL (m^2/ha)	34263	13.930	9.570	0	74.249
Slope (per)	310	6.466	7.264	0	42.000
Aspect (rad)	310	116.951	100.942	0	360.000
SDI (trees/ha)	34263	473.026	226.967	0	1444.220

1979; Holdaway, 1984; Shifley and Brand, 1984; Shifley, 1987).
- Develop a composite model that incorporates tree, stand and site characteristics in a single equation (Cole and Stage, 1972; Wykoff, 1990; Uzoh, 2001).

Martin and Ek (1984) described these respective choices as "semi-empirical" and "empirical." In practical terms, the differences in approach are mostly semantic (Wykoff, 1990). Wykoff and Monserud (1988) maintained that, if the relationships within a model are based on generally accepted principles of tree growth (Assmann, 1970), either approach can produce acceptable behavior. In the first case, competition and vigor are used to explain deviations from an age and site dependent potential; in the second case, similar effects are used to explain deviation about a mean growth rate that has been corrected for other tree, site and stand effects (Wykoff, 1990; Uzoh, 2001). The second (composite) approach was chosen to avoid the estimation problem associated with the first approach because while the potential approach is a useful construct for purposes of organizing model structure, it is extremely difficult to observe (Wykoff, 1990).

Spatial information from mapped tree locations was not available; therefore, our model is distance-independent (Munro, 1974). The choice of variables was restricted to site, stand and tree attributes that could be reliably obtained from stand inventories normally used in the various regions (e.g., USDA Forest Service, 1978).

3.2. The equation

Growth of the individual trees was potentially affected by three groups of variables: (1) tree size and vigor effects, (2) site effects and (3) competitive effects. The combination of some of these predictor variables and the transformation of others were initially tested for predicting 5-year periodic annual height increment using mixed models analysis procedure. The variable selection process involves a series of steps beginning with an initial data exploration that involves plotting the data and examining correlation statistics to identify those variables that may be useful in the model.

3.3. Tree size effects

We started by defining the relationship between increment and size and accounting for the two different sizes of experimental units: a spatial unit which is an individual tree and a set of temporal units which are the repeated measurements on individual trees. The following equation was used:

$$E[\ln(\text{PAIHT})] = \beta_0 + \beta_1 \ln(\text{dbh}) + \beta_2(\text{dbh})^2 + h_l + \varepsilon_{j(l)} + \varepsilon_{ik(jl)}, \qquad (1)$$

where E[ln(PAIHT)] is the expected value of the natural logarithm of 5-year periodic annual height increment in meters; ln(dbh), the value of the natural logarithm of initial diameter in cm; $(\text{dbh})^2$, the value of the square of initial diameter; β_0, β_1, β_2 are regression coefficients; h_l, the fixed effect of the lth location; $\varepsilon_{j(l)}$, the random error for plot j within locale l assumed to have expected value of zero (0) and constant variance (σ_l^2) and $\varepsilon_{ik(jl)}$ is a random error for measurement k on tree i within plot j and locale l assumed to have 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:

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

where ρ is the serial correlation coefficient for errors across time on the same tree.

When the resulting predictive function is plotted against dbh, the resulting function is a skewed unimodal shape with a maximum between 20 and 30 cm (Fig. 1). Additionally, the intercept term, β_0, can be expanded to include other tree and site effects that modify height increment while still retaining the basic relationship between tree size and growth (Uzoh, 2001).

3.4. Site effects

For a model to adequately characterize tree growth, it must include some measure of site productivity. Latitude, longitude, aspect, slope, elevation and site index were initially tested for site effects variables. The site quality indicator term is represented as:

$$\text{SITE} = \beta_3 \text{SI}, \qquad (2)$$

where SI is site index (m) (Meyer, 1938).

Meyer's site index was chosen because the data came from the widest geographic range of any site index system presently available. Provincial site index systems may more accurately portray sites within that province but they are widely divergent (Dunning and Reineke, 1933; Minor, 1964; Boldt and Van Deusen, 1974; Barrett, 1978; Powers and Oliver, 1978). Because, the main objective of this study is to provide forest managers with a single model that can provide useful guidelines for a variety of management objectives throughout the geographic range of ponderosa pine, we had to use a range-wide site index system.

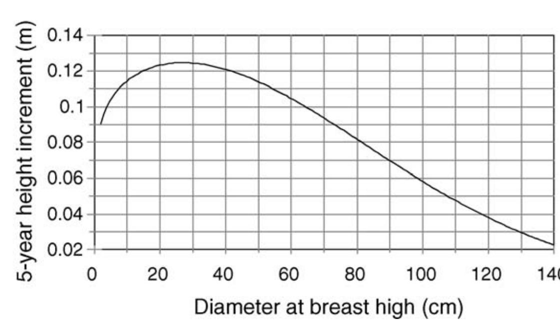

Fig. 1. Plot of 5-year periodic annual height increment by initial diameter at breast height using the coefficients derived for Eq. (1).

In addition to SI, other geoclimatic variation (OGV) may remain because of the range of variation in species characteristics and stand conditions in the study area, which extended from the Black Hills of South Dakota to the Pacific Coast. As a result, slope (SL), aspect (ASP) and elevation (ELEVA) terms can help in refining the overall site effect. These factors generally have no direct effect on tree growth, but act indirectly by influencing moisture, temperature, light and other chemical and physical agents of the site. Slope and aspect are included using Stage's (1976) transformation. The combined effects of slope, aspect and elevation are represented by OGV:

$$\mathrm{OGV} = \beta_4 \mathrm{SL}[\cos(\mathrm{ASP})] + \beta_5 \mathrm{ELEVA}, \tag{3}$$

where SL is the slope angle in percent, ASP the aspect in radians and ELEVA is elevation in meters.

3.5. Competitive effects

Finally, the increment attained by an individual tree is also dependent on its competitive status relative to neighboring trees and the impact of management. Basal area in larger trees (BAL), stand density index (SDI), stand basal area and basal area in larger trees divided by dbh of the subject tree were initially tested for competitive effects variables. The competitive effects (CE) term is represented by SDI (Reineke, 1933) for overall stand density and BAL for the individual tree's competitive position. SDI has a distinct advantage over stand basal area as a measure of stand density because it is less influenced by age and site quality. BAL has been used often as a tree-position variable in equations for predicting growth because it describes a tree's position in relation to all trees measured in a plot or stand (Ritchie and Hann, 1985; Wykoff, 1986, 1990; Dolph, 1988; Uzoh et al., 1998; Uzoh, 2001). The competitive effects (CE) term is represented by SDI and BAL:

$$\mathrm{CE} = \beta_6 \mathrm{SDI} + \beta_7 \mathrm{BAL}, \tag{4}$$

where SDI is stand density index (m) and BAL is the basal area in larger trees (m^2/ha) divided by the dbh of the subject tree.

3.5.1. Model selection

Many different tools can be used in evaluating competing models to determine which model is most appropriate. Most of these criteria are based on the presumption that you want to create a model that minimizes the unexplained variability (the mean squared error of prediction) with the fewest number of variables possible. Of the potential models, the one selected was chosen on the basis of the following criteria:

- The covariance structure was chosen among the two candidates of autoregressive errors and compound symmetry based on a maximum likelihood estimation of the fixed effects and random effects and choosing the structure that produced the smallest Akaike's information criterion (AIC) (Rawlings et al., 1998; Hastie et al., 2001; Burnham and Anderson, 2002).
- Restricted maximum likelihood (REML) was used to fit different fixed effects models. Then AIC was used to assess model fitness.
- Residual plots were examined to check on normality assumptions and the Spearman rank correlation coefficient was calculated for examining the stability of the variance across the range of independent variables (Carroll and Ruppert, 1988).

3.5.2. Repeated measures analysis

Selecting an appropriate covariance model is important in repeated measures analyses. If an important correlation is ignored by using a model that is too simple, the risk of Type I error rates is increased for fixed effects tests. If the model is too complex, power and efficiency is sacrificed. This decision process can be assisted by using the goodness of model fit criteria (AIC).

In this study based on the value of AIC, the autoregression covariance structure (with multiple observations on individual trees autocorrelated in time) outperformed other covariance structures such as compound symmetry (with multiple observations on an individual tree equally correlated irrespective of time). The covariance structure with the smallest value of the criteria is considered most desirable. Because of the repeated sampling from individual trees, it seemed natural to consider an autoregressive process to describe that error structure. Other covariance structures were also considered and the final selection was made using the value of AIC.

The three random terms (a variance component for plots within locales and for trees within plots along with the correlation between successive measurements on individual trees) are represented, which provided the best relationship between 5-year periodic annual height growth increment of individual ponderosa pine trees and (1) tree size and vigor effects, (2) site effects and (3) competitive effects.

4. Model testing and validation

Shugart (1984) defined model validation as ‘‘procedures, in which a model is tested on its agreement with a set of observations that are independent of those observations used to structure the model and estimate its parameters’’. There are many types of validation methods available; some are qualitative and others are quantitative (Holmes, 1983; Sargent, 1999). The use of statistical tests in model validation has drawn much debate since the work of Wright (1972) because of the varying criteria for the ‘‘value’’ of models and the methods of determining it (Mayer et al., 1994; Morehead, 1996). Because each model is unique, no single validation technique or method is widely applied. For selecting the most suitable regression model, it is generally advisable to use some measure of lack of fit in combination with one or more test statistics (Kozak and Kozak, 2003). Therefore, it is important to know that model validation is not designed to prove that a model is correct (Popper, 1963), but rather to show that model predictions are close enough to independent data and that decisions made based on the model are defensible (Yang et al., 2004).

There are four procedures commonly used in model validation: (1) a comparison of predictions and coefficients with physical theory; (2) a comparison of results with those obtained by theory and simulation; (3) the use of new data; (4) the use of data splitting or cross validation (Snee, 1977). Since a new data set is often not available, data splitting is regarded as an acceptable alternative by most practitioners provided that the data set is large enough (Yang et al., 2004).

The dataset was randomly split into 10 parts and 90% was used for initial model development and 10% was used for model validation. The final model was developed using the entire dataset. Using the validation dataset, a mixed model analysis and a regression model analysis of the actual height increments versus estimated height increments were then fit using the values of the natural logarithm. The two lacks of fit statistics, prediction mean squares error and mean bias were selected because they are both meaningful and diagnostically useful in determining the predictive power of models. Mean bias indicates trends in lack of fit, and prediction mean squares error indicates the extent of the spread of the residuals about the mean.

5. Results

The following equation provided the best fit:

$$E[\ln(\mathrm{PAIH})] = b_0 + b_1\ln(\mathrm{dbh}) + b_2(\mathrm{dbh})^2 + b_3\mathrm{SIM} + b_4\mathrm{SL}[\cos(\mathrm{ASP})] + b_5\mathrm{ELEVA} + b_6\mathrm{SDI} + b_7\mathrm{BAL} + \widehat{h}_l + e_{j(l)} + e_{ik(jl)}, \quad (5)$$

the random effects were trees, plots and locations. Where $E[\ln(\mathrm{PAIH})]$ is the expected value of the natural logarithm of 5-year periodic annual height increment (m), ln(dbh) the value of the natural logarithm of initial dbh (cm), $(\mathrm{dbh})^2$ the square of initial dbh (cm), SIM the Meyer's site index (m), SL the average slope for the stand (percent), ASP the average aspect for the stand (radians), ELEVA the elevation for the stand (m), SDI the stand density index (trees/ha), BAL is the basal area in larger trees (m^2/ha) divided by the dbh of the subject tree (see Table 4). b_0, b_1, b_2, b_3, b_4, b_5, b_6 and b_7 are regression coefficients, $\widehat{h}_l$ the fixed effect of the lth location, $e_{j(l)}$ the random error for plot j within locale l assumed to have expected value of zero (0) and constant variance (σ_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 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:

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

where ρ is the serial correlation coefficient for errors across time on the same tree.

Table 4
Parameter estimates for the height increment model for managed even-aged stands of ponderosa pine throughout western United States using the SAS system MIXED model analysis procedure

Parameter	95% Confidence interval			
	Estimate	Standard error	*T*-statistic	*P* (2-TAIL)
Intercept	−3.2377	0.09869	−34.27	0.0001
ln dbh (cm)	0.4807	0.00702	68.49	0.0001
dbh^2 (cm)	−0.00021	0.00007	−28.63	0.0001
SI M (m)	0.04605	0.00213	21.67	0.0001
SL[cos(ASP)]	0.01158	0.00221	5.25	0.0001
ELEVA (m)	−0.00016	0.00005	−3.21	0.0015
SDI (trees/ha)	−0.00037	0.00002	−15.95	0.0001
BAL (m^2/ha)	−0.00352	0.00031	−11.52	0.0001

Standard error of estimate = 0.53705 (m).

From the foregoing analysis, the height increment model developed [Eq. (5)] displayed a unimodal, positively skewed shape that is typical of tree growth processes (Fig. 2) (Assmann, 1970; Wykoff, 1990; Uzoh, 2001). The function plots (Fig. 2) patterns are logical and in line with expectations. A value of 1000 was assigned for elevation (m), a value of 25 was assigned for both slope and aspect, a value of 120 was assigned for SDI and a value of 10 was assigned for BAL. The coefficients from Eq. (5) are presented in Table 4. The model of 5-year height increment using a logarithmic transformation indicate that function plots are comparable with those found in other height increment studies where the same transformation was applied (Stage, 1975; Hasenauer and Monserud, 1997; Uzoh, 2001).

The logarithmic-bias correction to the intercept term (Flewelling and Pienaar, 1981) was estimated by adding half of the mean squared error to the intercept term (Baskerville, 1972). Flewelling and Pienaar (1981) suggested that for degrees of freedom >30 and $S^2 < 0.5$, the multiplicative correction of $e^{-S^2/2}$ is usually adequate. Since the residual mean squared error estimate is <0.5 and the sample size is >30, Baskerville's correction should be a close approximation to the true logarithmic bias for the equation presented (Eq. (5)). Therefore, Baskerville's method was used for the logarithmic bias correction.

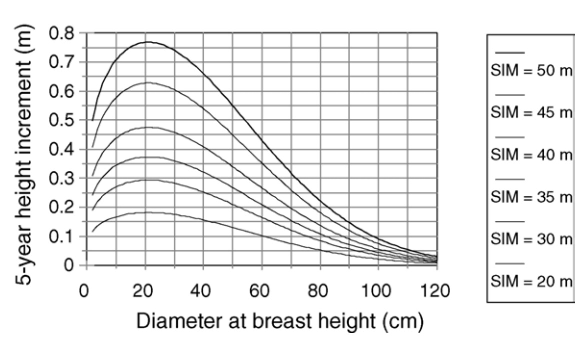

Fig. 2. Plot of 5-year periodic annual height increment by initial diameter at breast height and site index (SI) for managed even-aged stands of ponderosa pine throughout the western United States using the coefficients in Table 4.

Table 5
Prediction mean squares error (MS error, m), mean bias (m) and absolute mean bias (a mean bias, m) by mixed model and regression model using validation data set of size 10% of the total data set

Model	*n*	MS error	Mean bias	A mean bias
Mixed	9250	0.2663	−0.1880	0.3889
Regression	9250	0.4955	−0.2495	0.4061

The validation results are presented in Tables 5 and 6.

6. Discussion

The relative importance of a variable is assessed by the change in size of the standard error of prediction without the variable in question. Table 7 shows the ranking of the variables based on this criterion.

Site index (Meyer, 1938) (SIM) had more effect on height growth than any other variable (Table 7). It is important to realize, however, that other factors were combined under the variable SIM. The data were scattered over a vast geographic area of contrasting soils and climate, and included two varieties of ponderosa pine (*P. ponderosa* var. *ponderosa* and var. *scopulorum*). Some of the genetic differences may have affected PAI height. Nevertheless, what we called SIM seemed to perform credibly in integrating and explaining these complex differences. A contributing reason for the good performance of site index may have been that stockability was not a problem. All data were from sites capable of the productivity estimated by Meyer (1938).

The natural logarithm of initial diameter (ln(dbh)) had the second greatest effect on height growth followed by the square of initial diameter $(\mathrm{dbh})^2$ (Table 7). The size of initial diameter is an indication of a tree's competitive status within a plot or stand, and thus, an expression of tree vigor. The square of initial diameter $(\mathrm{dbh})^2$ had the third greatest effect on diameter growth. The inclusion of $(\mathrm{dbh})^2$ gave Eq. (5) its asymptotic approach to zero for large heights, removing the need for imposition of an arbitrary maximum height (Fig. 2).

Within the fundamental constraint of site quality, tree position significantly influenced height growth. BAL (basal area in larger trees (m^2/ha)/(dbh of the subject tree)) had the fourth-greatest influence on height growth (Table 7). The increment attained by an individual tree is dependent on its competitive status relative to neighboring trees. Consequently, the coefficient of BAL is negative, indicating a competition modifier that would reduce height growth rates relative to a tree's competitive status. Therefore, the largest diameter tree in a plot would have a BAL value of zero, while the smallest diameter tree in the plot would have a BAL value near that of the plot's total basal area. As BAL decreases, the predicted increment increases. The more open-grown the tree, the less it is influenced by competitors because the measure of relative size is tied to stand density. As a result, dominance is less of a factor in increment predictions in sparsely stocked stands (Wykoff, 1990; Uzoh et al., 1998; Uzoh, 2001).

Table 7
Ranking of variables based on change in the size of the standard error of prediction without the variable in question

Variable	Value of standard error of prediction without the variable in question	Value of standard error of prediction for the full model
SIMM	0.7225	6088
ln dbh	0.6461	
dbh^2	0.6322	
BAL (m^2/ha)	0.6160	
SL[cos(ASP)]	0.6136	
ELEVA	0.6121	
SDI	0.6097	

Overall stand density as measured by SDI had the seventh-greatest influence on height growth (Table 7). The importance of SDI in the model suggests that height growth of all trees in a stand is affected by stand density—trees with the largest diameters as well as those with the smallest diameters. This relationship is in accordance with that reported for the two levels-of-growing-stock installations in Oregon (Cochran and Barrett, 1995, 1999a).

After thinning from below in dense stands, BAL is unchanged, but predicted increment increases because SDI is lower. In rare instances when growth after thinning from above is modeled, predicted response may be overestimated, at least initially. Both BAL and SDI are reduced, causing a predicted growth increase greater than that for thinning from below. The response might be delayed until tree crowns and root systems of subordinate trees expand to take advantage of the added space. In general, however, the effects of BAL and SDI are biologically rational and simple in concept yet they can accommodate extensive variation in stand structure and site conditions (Wykoff, 1990; Uzoh et al., 1998).

Table 6
Prediction mean squares error (m), mean bias (m) and absolute mean bias (m) by province for mixed model and regression model using validation data set of size 10% of the total data set

Province	*n*	Mixed model validation statistics			Regression validation statistics		
		Prediction MS error	Mean bias	Absolute mean bias	Prediction MS error	Mean bias	Absolute mean bias
I	1475	0.2794	−0.2085	0.3887	0.4915	−0.2485	0.3894
II	2572	0.3414	−0.2420	0.4684	0.5705	−0.2728	0.4764
III	2216	0.2512	−0.1472	0.3922	0.5001	−0.2286	0.4176
IV	1726	0.2317	−0.1219	0.3647	0.4857	−0.2086	0.3772
V	2254	0.1438	−0.1742	0.3032	0.3743	−0.2573	0.3343

Stage's (1976) transformation of slope (SL) and aspect (ASP) (SL[cos(ASP)]) had the fifth greatest effect on height growth (Table 7). The transformation of slope and aspect has two important properties, it is circular and optima exist with respect to both slope and aspect.

Elevation had the sixth greatest effect on height growth (Table 7). Elevation is an important variable because average temperatures decline as elevations increase. Tree growth is sensitive to growing season temperature, which as the model location coefficient describes this climatic trend. Consequently, the coefficient of elevation is negative.

PROC MIXED will not calculate the coeffecients of variable with very high degrees of multicollinearity. Consequently, latitude, which would seem to be an obvious location variable was dropped in favor of elevation because a high degree of multicollinearity existed between the two variables and because elevation had more effect on height growth than did latitude. The two variables were confounded because ponderosa pine is found at increasing elevations as latitudes decrease. Age and longitude were dropped because they were highly correlated with the more important variable, site index.

The validation results of this study decisively show that the mixed model analysis outperformed the regression analysis model (Tables 5 and 6). The mixed model analysis mean biases and absolute mean biases were closer to zero than for the regression model analysis and the mixed model analysis prediction mean squares errors were smaller. The results show that the mixed model predictions are closer to the independent data set than is the regression model prediction. Our finding confirms the previous findings (Biging, 1985; Gregoire et al., 1995; Keselman et al., 1999; Hall and Bailey, 2000; Littell et al., 2000; Kowalchuk and Keselman, 2001). What makes mixed models unique is the ability of the models to include both fixed regression parameters (fixed effects) that describe the shape of the typical growth curve over the entire population, and random regression coefficients (random effects) that individualize the curve to capture site-specific, tree-specific, or other unit-specific characteristics of the growth pattern (Hall and Bailey, 2000).

Acknowledgments

We are 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. Dr. James Baldwin's constructive review of the earlier draft greatly strengthened this paper. We also thank the anonymous reviewers for their constructive comments and suggestions.

References

Alder, D., 1979. A distance independent tree model for exotic conifer plantations in East Africa. For. Sci. 25 (1), 59–71.

Assmann, E., 1970. The Principles of Forest Yield Study. Pergamon Press, New York.

Barrett, J.W., 1978. Height growth and site index curves for managed even-aged stands of ponderosa pine in the Pacific Northwest. US Forest Service Research Paper PNW 232. US Department of Agriculture, Pacific Northwest Research Station, Portland, OR.

Baskerville, G.L., 1972. Use of logarithmic regression in the estimation of plant biomass. Can. J. For. Res. 2, 49–53.

Biging, G.S., 1985. Improved estimates of site index curves using a varying-parameter model. For. Sci. 31 (1), 248–259.

Boldt, C.E., Van Deusen, J.L., 1974. Silviculture of ponderosa pine in the Black Hills: the status of our knowledge. US Forest Service Research Paper RM 124. US Department of Agriculture, Rocky Mountain Forest and Range Experiment Station, Fort Collins, CO.

Burkhart, H.E., Parker, R.C., Strub, M.R., Oderwald, R.G., 1992. Yield of Old-field Loblolly Pine Plantations. School of Forestry and Wildlife Resources, Virginia Polytechnic Institute and State University, Blacksburg. Publ. FWS-3-72.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, second ed. Springer, New York.

Carroll, R.J., Ruppert, D., 1988. Transformation and Weighting in Regression. Chapman and Hall, New York.

Cochran, P.H., Barrett, J.W., 1995. Growth and mortality of ponderosa pine poles thinned to various densities in the Blue Mountains of Oregon. US Forest Service Research Paper PNW 483. US Department of Agriculture, Pacific Northwest Research Station, Portland, OR.

Cochran, P.H., Barrett, J.W., 1999a. Growth of ponderosa pine thinned to different stocking levels in central Oregon: 30-year results. US Forest Service Research Paper PNW 508. US Department of Agriculture, Pacific Northwest Research Station, Portland, OR.

Cochran, P.H., Barrett, J.W., 1999b. Thirty-five-year growth of ponderosa pine saplings in response to thinning and understory removal. US Forest Service Research Paper PNW 512. US Department of Agriculture, Pacific Northwest Research Station, Portland, OR.

Cole, D.M., Stage, A.R., 1972. Estimating future diameter of lodgepole pine. US Forest Service Research Paper INT 131. US Department of Agriculture, Intermountain Forest and Range Experiment Station, Ogden, UT.

Curtis, R., 1967. Height–diameter and height–diameter–age equations for second-growth Douglas-fir. For. Sci. 13 (4), 365–375.

Curtis, R.O., Clendenen, G.W., DeMars, D.J., 1981. A new stand simulator for coast Douglas-fir: DFSIM user's guide. US Forest Service General Technical Report PNW 128. US Department of Agriculture, Pacific Northwest Research Station, Portland, OR.

Dolph, K.L., 1988. Prediction of periodic basal area increment for young-growth mixed onifers in the Sierra Nevada. US Forest Service Research Paper PSW 190. US Department of Agriculture, Pacific Southwest Forest and Range Experiment Station, Berkeley, CA.

Dunning, D., Reineke, L.H., 1933. Preliminary Yield Tables for Second-Growth Stands in the California Pine Region. US Department of Agriculture, Technical Bull. 354.

Flewelling, J.W., Pienaar, L.V., 1981. Multiplicative regression with lognormal error. For. Sci. 27 (2), 281–289.

Gregoire, T.G., Schabenberger, O., Barrett, J.P., 1995. Linear modeling of irregularly spaced, unbalanced, longitudinal data from permanent plot measurements. Can. J. For. Res. 25, 137–156.

Hahn, J.T., Leary, R.S., 1979. Potential diameter growth function. A generalized forest growth projection system applied to the Lake States. US Department of Agriculture, General Technical Report NC-49.

Hall, D.B., Bailey, R.L., 2000. Predicting dominant height growth in forest plantations based on multilevel nonlinear mixed models. Univ. of Georgia Stat. Dep. Tech. Rep. #2000–5.

Hasenauer, H., Monserud, R.A., 1997. Biased prediction for tree height increment models developed from smoothed 'data'. Ecol. Model. 98, 13–22.

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

Holdaway, M.R., 1984. Modeling the effect of competition on tree diameter growth as applied in STEMS. US Department of Agriculture, General Technical Report NC-94.

Holmes, W.M., 1983. Confidence building in simulation models as a practical process. In: Proceedings of the 1983 Summer Computer Simulation Conference, 11–13 July 1983, Vancouver, B.C. The Society for Modeling and Simulation International, San Diego, California, USA.

Keselman, H.J., Algina, J., Kowalchuk, R.K., Wolfinger, R.D., 1999. A comparison of recent approaches to the analysis of repeated measurements. Br. J. Math. Stat. Psychol. 52, 63–78.

Kowalchuk, R.K., Keselman, H.J., 2001. Mixed model pairwise multiple comparisons of repeated measures means. Psychol. Methods 6, 282–296.

Kozak, A., Kozak, R., 2003. Does cross validation provide additional information in the evaluation of regression models? Can. J. For. Res. 33, 976–987.

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

Lear, R.A., Holdaway, M.A., 1979. Modifier function. A Generalized Forest Growth Projection System Applied to the Lake States. US Department of Agriculture, General Technical Report NC-49.

Littell, R.C., Pendergast, J., Natajan, R., 2000. Modeling covariance structure in the analysis of repeated measures data. Stat. Med. 19, 1793–1819.

Martin, G.L., Ek, A.R., 1984. A comparison of competition measures and growth models for predicting plantation red pine diameter and height growth. For. Sci. 30 (3), 731–743.

Mayer, D.G., Stuart, M.A., Swain, A.J., 1994. Regression of real-world data on model output: an appropriate overall test of validity. Agric. Syst. 45, 93–104.

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 Bull. 630.

Minor, C.O., 1964. Site-index curves for young-growth ponderosa pine in northern Arizona. US Forest Service Research Note RM 37. US Department of Agriculture, Rocky Mountain Forest and Range Experiment Station, Fort Collins, CO.

Monserud, R.A., 1975. Methodology for simulating Wisconsin northern hardwood stand dynamics. Ph.D. Thesis, Univ. Wisconsin-Madison, USA. 173 p. (Univ. Microfilms 76-2496).

Morehead, L.A., 1996. Determining the factors influential in the validation of computer- based problem solving systems. Ph.D. Dissertation, Portland State University, Portland, Oregon, USA. (University Microfilms UMI No. 9628864)

Munro, D.D., 1974. Forest growth models—a prognosis. In: Fries, J. (Ed.), Proceedings—Growth Models for Tree and Stand Simulation. Royal Coll. Stockholm For. Res. Note 30.

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, CO.

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 Forest and Range Experiment Station, Berkeley, CA.

Oliver, W.W., 1997. Twenty-five-year growth and mortality of planted ponderosa pine repeatedly thinned to different stand densities in northern California. West. J. Appl. For. 12 (4), 122–130.

Oliver, W.W., 1990. Spacing and shrub competition influence 20-year development of planted ponderosa pine. West. J. Appl. For. 5 (3), 79–82.

Popper, K.R., 1963. Conjectures and Refutations. Routledge and Kegan Paul, London.

Powers, R.F., Oliver, W.W., 1978. Site classification of ponderosa pine stands under stocking control in California. US Forest Service Research Paper PSW 128. US Department of Agriculture, Pacific Southwest Forest and Range Experiment Station, Berkeley, CA.

Rawlings, J.O., Pantula, S.G., Dickey, D.A., 1998. Applied Regression Analysis: A Research Tool, second ed. Springer-Verlag, New York.

Reineke, L.H., 1933. Perfecting a stand-density index for even-aged forests. J. Agric. Res. 46, 627–638.

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.

Ritchie, M.W., Hann, D.W., 1985. Equations for predicting basal area increment in Douglas-fir and grand fir. In: For. Res. Lab. Res. Bull. 51, Oregon State University, Corvallis, OR, USA.

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

Sargent, R.G., 1999. Validation and verification of simulation models. In: Farrington, P.A., Nembhard, H.B., Sturrock, D.T., Evans, G.W. (Eds.), Proceedings of the 1999 Winter Simulation Conference, 5–8 December 1999, Phoenix, Ariz, Institute of Electrical and Electronics Engineers, New York, pp. 39–48.

Shifley, S.R., 1987. A generalized system of models forecasting central states tree growth. US Forest Service Research Paper NC-279.

Shifley, S.R., Brand, G.J., 1984. Chapman–Richards growth function constrained for maximum tree size. For. Sci. 30 (4), 1066–1070.

Shugart, H.H., 1984. A Theory of Forest Dynamics. Springer-Verlag, New York.

Snee, R.D., 1977. Validation of regression models: methods and examples. Technometrics 19, 415–428.

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

Stage, A.R., 1976. An expression for the effect of slope, aspect and habitat type on tree growth. For. Sci. 22 (4), 457–460.

US Forest Service, 1978. Field instruments: stand examination—forest inventory. In: US Forest Service Handbook 2404, US Department of Agriculture, Rocky Mountain Forest and Range Experiment Station, Northern Region, Missoula, MT.

Uzoh, F.C.C., 2001. A height increment equation for young ponderosa pine plantations using precipitation and soil factors. For. Ecol. Manage. 142 (1–3), 191–201.

Uzoh, F.C.C., Dolph, K.L., Anstead, J.R., 1998. Periodic annual diameter increment after overstory removal in mixed conifer stands. US Forest Service Research Paper PSW 238. US Department of Agriculture, Pacific Southwest Research Station, Albany, CA.

Wright, R.D., 1972. Validating dynamic models: an evaluation of tests of predictive powers. In: Proceedings of 1972 Summer Computer Simulation Conference, 13–16 June 1972, San Diego, Calif. Simulation Councils, La Jolla, California, USA.

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

Wykoff, W.R., 1986. Supplement to user's guide for the stand prognosis model—version 5.0. US Forest Service General Technical Report INT 208. US Department of Agriculture, Intermountain Forest and Range Experiment Station, Ogden, UT.

Wykoff, W.R., Monserud, R.A., 1988. Representing site quality in increment models: a comparison of methods. In: Ek, A.R., Shifley, S.A., Burk, T.E. (Eds.), Forest Growth Modeling and Prediction. US Forest Service General Technical Report NC-120.

Wykoff, W.R., Crookston, N.L., Stage, A.R., 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.

Yang, Y., Monserud, R.A., Huang, S., 2004. An evaluation of diagnostic tests and their roles in validating forest biometric models. Can. J. For. Res. 34, 619–629.