This file was created by scanning the printed publication.
Text errors identified by the software have been corrected;
however, some errors may remain.

A HIERARCHICAL LINEAR MODEL FOR TREE HEIGHT PREDICTION

Vicente J. Monleon
Forest Inventory and Analysis, PNW Research Station, USDA Forest Service, Portland, Oregon 97205
and **Department of Statistics, Oregon State University, Corvallis, Oregon 97331**

KEY WORDS: Best Linear Unbiased Prediction, Mixed Effects Models, Multicollinearity, Forestry

Abstract
Measuring tree height is a time-consuming process. Often, tree diameter is measured and height is estimated from a published regression model. Trees used to develop these models are clustered into stands, but this structure is ignored and independence is assumed. In this study, hierarchical linear models that account explicitly for the clustered structure of the data are compared with model forms currently used in forestry. The data consist of 1433 Douglas-firs from 99 Oregon stands measured in 2000, and an independent evaluation dataset of similar size measured in 2001. Overall model performance improved substantially if the stand random effect could be predicted: root mean squared error (RMSE) decreased from 4.91 m (current models) to less than 3.73 m (hierarchical model, 1 tree sampled). However, if the random effect could not be estimated, the improvement was small (RMSE 4.45 m). The within-stand relationship between height and diameter was different from that between stands. As a result, the random and fixed components of the model are confounded. A mixed model that did not account for this problem performed worse than the model that assumed an independent data structure.

Introduction
Tree height is measured by triangulation, a time consuming process. In most forestry applications, only tree diameter is actually measured. Height is predicted from published regression equations.

Trees grow, and are sampled, in clusters or 'stands'. A stand is a group of trees that occupy a small area and share a common environment and history. Most current models ignore the clustered structure of the data and are fitted under the assumption of independent observations. The main argument supporting this practice is that ordinary least squares estimates are unbiased. Since users are often only interested in point estimates, trying to incorporate a more reasonable variance structure may not be necessary. Besides the possible loss of efficiency, there are several problems with this approach:

First, in most applications, there are two levels of information: only the diameters of the trees in the stand are known; or the diameter is known for all trees, and both the height and diameter are known for a small sample. In the later case, foresters either ignore the additional information and use the regional, published models; or fit a stand-specific regression to the available data. A mixed model approach that predicts a random stand effect may result in a more efficient prediction.

Second, models that ignore the stand effect assume that the relationship between tree height and diameter within-stand is the same as that between-stands. This is unlikely, since the ecological processes that control those relationships should be different. The within-stand relationship between height and diameter may be controlled by competition between individual trees, while the between-stand relationship may be controlled by the overall stand age and environment. Hierarchical models allow for the explicit separation of the between- and within-stand relationships, and thus for a correct specification of the model.

The objectives of this study are to develop a method to predict tree height when the height and diameter of a sample of trees from the stand are known, as well as when only the diameters are known; and to study the performance of the proposed models using an extensive validation dataset. An important requirement is that, as with most current models, users should be able to predict height using a hand-held calculator or spreadsheet.

Data
Douglas-fir tree height and diameter were obtained from the national forest inventory conducted by the Forest Inventory and Analysis program of the USDA Forest Service. Trees were sampled in a spatially balanced sample of plots across all non-federal land in Western Oregon during 2000 and 2001. The sample locations for the two years were drawn independently. The diameter and height of all Douglas-fir trees in each plot were measured using standard techniques. Plot size was 0.067 ha for trees with diameter between 12.7 and 61 cm, and 0.4 ha for trees with diameter greater than 61 cm. In 2000, a total of 99 stands and 1433 trees were measured. The median number of trees per stand was 11 (range from 1 to 62). In 2001, 97 stands and 1589 trees were measured. The median number of trees per stand was 13 (range from 1 to 67). Figure 1 shows the diameter and height data for both years.

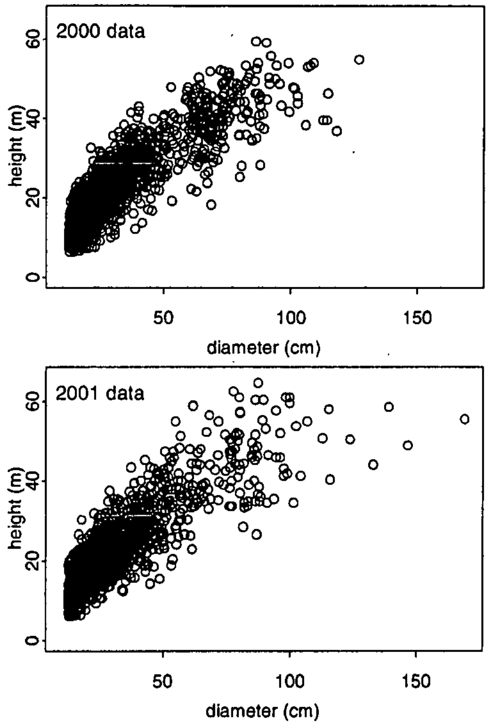

Figure 1. *Scatterplot of tree height vs. tree diameter. The 2000 dataset was used to fit the models, and the 2001 dataset to evaluate their performance.*

Models

Five different models were considered (Table 1). The Chapman-Richards (e.g. Garman *et al.* 1995) and Exponential (e.g. Hanus *et al.* 1999) models are frequently used. The OLS polynomial is a simple approximation to the non-linear models. These three models ignore the clustered structure of the data. The random intercept model includes a variance component to account for the stand effect. The hierarchical model attempts to separate the within-stand height diameter relationship from the between-stand relationship.

Model parameters were estimated with the 2000 data using standard software (S-Plus). Model performance was evaluated using the 2001 data.

Predicting tree height when only the diameter of trees in a stand are known

If only the diameters are known, the random intercept for the stand cannot be estimated. Therefore, only the part of the random intercept and hierarchical models that involves fixed parameters – the marginal expectation – can be used for prediction.

The overall predictive performance of the two nonlinear and OLS polynomial models was almost identical (Table 2). The estimated functions were undistinguishable, except for the large-diameter trees, where the data was very sparse – there are only 11 trees with diameter greater than 100 cm (Fig. 2).

The random intercept model performed substantially worse than the current standards and showed a significant bias, with a tendency to underpredict tree height (Fig. 2). This difference between the OLS models and the mixed model indicates that the between-stand and within-stand height-diameter relationship are not the same. Models that ignore the clustered structure of the data closely follow the between-stand relationship of height vs. diameter. The random intercept model accounts for part of the between-stand relationship with the stand-specific intercept – the random and fixed effects are confounded. Therefore, the fixed part of this model more closely follows the within-stand relationship

Table 1. *Models considered in this study.*

Name	Model form
Chapman-Richards	$H_{ij} = \beta_0\left(1 - e^{\beta_1 D_{ij}}\right)^{\beta_2} + \varepsilon_{ij}$
Exponential	$H_{ij} = \exp\left(\beta_0 + \beta_1 D_{ij}^{\beta_2}\right) + \varepsilon_{ij}$
OLS polynomial	$H_{ij} = \beta_0 + \beta_1 D_{ij} + \beta_2 D_{ij}^2 + \beta_3 D_{ij}^3 + \varepsilon_{ij}$
Random intercept	$H_{ij} = \beta_0 + \beta_1 D_{ij} + \beta_2 D_{ij}^2 + \beta_3 D_{ij}^3 + b_i + \varepsilon_{ij}$
Hierarchical	$H_{ij} = \beta_0 + \beta_1 \overline{D}_{i.} + \beta_2 \overline{D}_{i.}^2 + \beta_3 \overline{D}_{i.}^3 +$
	$\alpha_1 (D_{ij} - \overline{D}_{i.}) + \alpha_2 \left(D_{ij} - \overline{D}_{i.}\right)^2 + \alpha_3 \left(D_{ij} - \overline{D}_{i.}\right)^3 + b_i + \varepsilon_{ij}$

where H_{ij} and D_{ij} are the height and diameter of the j-th tree in the i-th plot, respectively; $\overline{D}_{i.}$ is the average tree diameter for plot i; β's and α's are fixed parameters; $b_i \sim N(0,\sigma_b^2)$, $\varepsilon_{ij} \sim N(0,\sigma_e^2)$, b_i and ε_{ij} independent.

Table 2. *Model performance. Only the fixed part of the random intercept and hierarchical models was used. The models were fitted to the 2000 data, and evaluated with the 2001 data.*

Model	RMSE (m)
Chapman-Richards	4.95
Exponential	4.96
OLS polynomial	4.94
Random intercept	5.60
Hierarchical	4.62

between height and diameter. As expected, the between-stand relationship should be less steep than the within-stand relationship (Fig. 2).

The hierarchical model allows separating the between- and within-stand relationship of height vs. diameter, and performed best (Table 2). However, the gain was only 0.33 m or 7%. The moderate improvement may be due to the relatively low number of trees per stand, or to the fact that the range in tree diameter for some stands was very wide. The

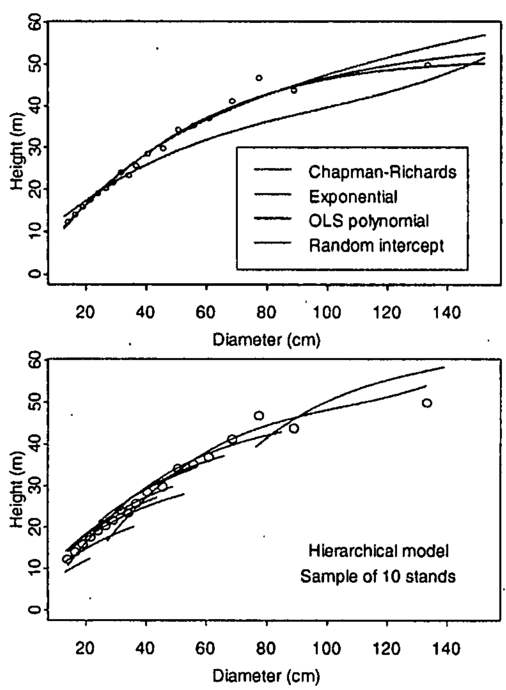

Figure 2. *Predictive functions – fixed effects only. For clarity, the points represent the average tree height by narrow diameter intervals (2001 season data). There are only 11 trees with diameter greater than 100 cm.*

slopes of the curves tend to be flatter, reflecting the within-stand relationship but, in contrast with the random intercept model, the predictions were not biased.

Predicting tree height when the diameter and height of a sample of trees are known

The random intercept in the random intercept and hierarchical models can be predicted using Best Linear Unbiased Prediction (BLUP, Goldberg 1962). If a sample of m trees with height H_i and diameter D_i, $i = 1,\ldots m$, is taken from a new stand, the random intercept for a stand can be estimated by:

$$\hat{b}_{new} = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_e^2 + m\hat{\sigma}_b^2}\sum_{i=1}^{m}\left(H_i - \mathbf{D}_i'\hat{\boldsymbol{\beta}}\right) \qquad (1)$$

The height of another tree from the same stand can be estimated by:

$$\hat{H}_{new} = \hat{b}_{new} + \mathbf{D}'_{new}\hat{\boldsymbol{\beta}}, \qquad (2)$$

where H_i and D_i are the height and vector of covariates of the trees measured in the new plot; $\hat{\sigma}_b^2$, $\hat{\sigma}_e^2$, and $\hat{\boldsymbol{\beta}}$ the estimated parameters from the 2000 season data; and $\hat{H}_{new}$ the predicted height of a tree with vector of covariates D_{new}. The estimated parameters are set to their restricted maximum likelihood (variance components) or maximum likelihood (fixed effects) estimates. This is not unreasonable, since the sample size was so large that the estimated variability of the parameters estimates was negligible.

To assess the performance of the models in this context, only stands in which at least 5 trees were measured were selected. There were 81 such stands (1402 trees) from the 2000 season data, and 74 stands (1538 trees) from the 2001 season data. The estimated variance components and RMSE for the fixed part of the models, calculated with the 2001 data, are reported in Table 3. The prediction accuracy of the BLUP was evaluated when the height and diameter of a sample of 1 to 4 trees per stand is known, as follows:

1. A random sample of $m=1,\ldots,4$ trees from a stand was chosen and $\hat{b}_{new}$ predicted (eq. 1)
2. The height of the remaining trees in the stand, $\hat{H}_{new}$, was predicted (eq. 2), and the residual calculated.

Table 3. *Estimated variance components and RMSE of the marginal model (stands with at least 5 trees)*

Model	$\hat{\sigma}_b$	$\hat{\sigma}_e$	RMSE(m)
Chapman-Richards		4.97	4.91
Exponential		4.99	4.92
OLS-polynomial		4.97	4.91
Random intercept	5.05	2.82	5.52
Hierarchical	3.00	2.78	4.45

3. The process was repeated for all stands, to obtain the overall RMSE. It was iterated 100 times. The average RMSE is reported.

The predicted performance of the models improved substantially, even when only a few trees per stand were available to predict the random intercept (Table 4). With only 1 tree, the RMSE of the hierarchical model was 1.18 m (24%) smaller than that from the current models. The performance improved as the sample size increased, with a 35% decrease in the RMSE for a sample of 4 trees.

The predictive performance of the random intercept model was as good as that of the hierarchical model, especially as the sample size increased. The discrepancy between the performance of the random intercept model when only the fixed part was used (Table 2) and when both the fixed and random part were used (Table 4) may be explained by the confounding between the random and fixed effects. In the random intercept model, some of the between-stand variability was incorporated into the random stand effect. This results in a problem similar to that of multicollinearity in multiple linear regression, but here both fixed and random effects are involved. It is well known that multicollinearity is not a problem for prediction, as long as the subjects to be predicted come from the same population as those used to develop the models (Rawlings *et al.* 1998).

Predictive performance at the stand level

In most applications, foresters are interested in predicting tree height in a particular stand. To evaluate the performance of the models at the stand level, and compare the performance of the BLUP estimates with a stand specific equation, all the stands with at least 30 trees measured were selected (17 stands). Then, for each stand,

1. The RMSE for the tree height predictions using the Chapman-Richards and OLS polynomial model was calculated.
2. A random sample of 15 trees from the stand was selected.

Table 4. *Performance of the BLUP, as a function of the number of trees used to estimate the random intercept.*

	Sample size			
	1	2	3	4
Model	RMSE (m)			
Random intercept	3.91	3.51	3.34	3.23
Hierarchical	3.73	3.49	3.33	3.22

3. A third order polynomial regression was fitted to those trees, the height of the remaining trees in the stand predicted, and the RMSE calculated.
4. The intercept for the random intercept and hierarchical models was predicted with the 15-tree sample, and the RMSE for the remaining trees calculated.
5. The random intercept was also predicted, but using only the first 5 trees in the sample, and the RMSE for the remaining trees calculated.
6. The process was iterated 100 times for each stand

Model performance was best when the best linear unbiased predictor was used, even if a stand specific regression model was available, and even when a small sample of trees was used to predict the random intercept (Table 5). An interesting aspect of the BLUP was its robustness. The average stand-specific RMSE was 3.36, but this average was heavily influenced by two stands (RMSE 7.26 and 8.19, stands 5 and 15). If those stands were excluded, the average RMSE was 2.78 m, much smaller but still worse than that of the BLUP. Stands 5 and 15 had very influential observations, and the predictive performance changed greatly depending on whether those observations were included in the sample used to develop the regression. Since the BLUP pools information from the overall population, it was not affected as much by the idiosyncrasies of the sample selected.

Conclusions

Compared with common methods, tree height prediction can improve substantially if a small sample of tree heights and diameters from the stand is available, and relatively simple linear mixed models are used. The performance of these models is better than that of stand-specific models fitted with a much larger sample.

If a sample of tree heights and diameters is not available, the marginal expectation from a random intercept model performs worse than models that ignore the clustered structure of the data. A hierarchical approach that separates the within- and

Table 5. *Model performance at the stand level. CR, OLS, RI and HIER denote the Chapman-Richards, OLS polynomial, random intercept, and hierarchical models, respectively. The stand-specific results were obtained from fitting a polynomial regression to a random sample of 15 trees from the stand, and predicting the remaining trees. The BLUP results were obtained after selecting a sample of m=15 or m=5 trees/stand to predict the random intercept.*

	RMSE (m)								
						BLUP			
	Marginal model				Stand-	Random intercept		Hierarchical	
Stand	CR	OLS	RI	HIER	specific	m=15	m=5	m=15	m=5
1	2.44	2.45	2.45	2.45	2.76	2.02	2.19	2.07	2.21
2	2.32	2.33	1.88	1.81	2.80	1.64	1.74	1.65	1.72
3	3.50	3.51	5.23	0.94	1.53	1.52	1.87	1.01	1.09
4	2.78	2.79	4.34	1.51	2.32	1.46	1.48	1.26	1.36
5	5.88	5.88	6.00	6.56	7.26	4.47	4.76	4.66	4.90
6	8.32	8.30	8.41	8.24	4.49	2.75	2.85	2.70	2.90
7	2.67	2.69	2.83	2.41	2.95	2.40	2.46	2.38	2.39
8	5.06	5.08	4.56	4.22	3.23	2.40	2.58	2.34	2.53
9	4.84	4.84	3.82	6.22	2.78	2.12	2.21	2.20	2.40
10	2.39	2.40	3.54	2.25	2.42	1.71	1.78	1.60	1.71
11	3.83	3.81	3.40	4.37	1.86	1.51	1.58	1.48	1.60
12	2.11	2.13	2.45	1.27	1.63	1.30	1.36	1.17	1.20
13	3.88	3.88	3.85	4.09	4.13	3.87	3.89	3.95	3.95
14	4.03	4.04	2.84	3.70	3.17	2.95	3.10	2.66	2.76
15	5.39	5.41	5.44	3.72	8.19	2.25	2.42	1.99	2.13
16	3.41	3.44	3.31	1.88	2.32	2.21	2.33	1.95	2.04
17	3.74	3.74	2.76	3.93	3.34	2.93	3.01	2.68	2.75
Average	3.92	3.92	3.95	3.50	3.36	2.32	2.45	2.22	2.33

between-stand relationship between height and diameter performs best.

Regional inventories provide a very large, spatially balanced dataset of randomly selected plots across the entire range of tree species. They can be used advantageously to develop simple hierarchical mixed models that are applicable widely.

Acknowledgements

Many people from the Forest Inventory and Analysis Program of the USDA Forest Service have been involved in the design of the inventory and collection of the data. Thanks go to our data collection and compilation staff. Special thanks go to many landowners who allowed field crews on their lands. J. Alegria, D. Azuma, T. Barrett, J. Fried and D. Wilson contributed very useful comments and reviews.

References

Garman, S.L., Acker, S.A., Ohmann, J.L., Spies, T.A. 1995. Asymptotic height-diameter equations for twenty-four species in Western Oregon. Research Contribution 10, Forest Research Laboratory, Oregon State University, Corvallis.

Goldberg, A.S. 1962. Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association* 57:369-375.

Hanus, M.L., Hann, D.W., Marshall, D.D. 1999. Predicting height for undamaged and damaged trees in Southwest Oregon. Research Contribution 27, Forest Research Laboratory, Oregon State University, Corvallis.

Rawlings, J.O., Pantula, S.G., Dickey, D.A. 1998. Applied Regression Analysis, 2nd ed. Springer-Verlag, New York. 657p.