A Comparison of Pine Height Models for the Crossett Experimental Forest

D. Bragg

Southern Research Station, USDA Forest Service, P.O. Box 3516 UAM, Monticello, AR 71656

Correspondence: dbragg@fs.fed.us

Abstract

Many models to predict tree height from diameter have been developed, but not all are equally useful. This study compared a set of height-diameter models for loblolly (*Pinus taeda*) and shortleaf (*Pinus echinata*) pines from Ashley County, Arkansas. Almost 560 trees ranging in diameter at breast height (DBH) from 0.3 cm (both species) to 91.9 cm (for shortleaf) or 108.2 cm (for loblolly) were chosen for measurement. Height equations were then fit to four different functions (Chapman-Richards, modified logistic, exponential, and Curtis-Arney) with weighted nonlinear least squares regression using DBH as the only predictor. Models were evaluated using a series of goodness-of-fit measures, including fit index (R^2), root mean square error (RMSE), bias, and corrected Akaike information criterion (AICc). All of the models fit the data very well, with 96 to 98% of the variation explained for loblolly pine, and 96 to 97% explained for shortleaf pine. Similarly, few differences were apparent in RMSE, bias, and AICc, although it was clear that the Curtis-Arney function fit both pine species slightly less well across the upper range of the diameters. Only subtle differences appeared in curve shape for small- to moderate-sized pines, with increasing departures predicted above 75 cm DBH. Given their overall similarity in performance, the modified logistic function was the preferred height-diameter model because of its more intuitive allometry at the upper extreme of pine size, especially when compared to the original FVS height dubbing equation. A unified height-diameter model capable of predicting total tree height for either pine taxa was also developed with a modified logistic function.

Introduction

Tree height is one of the most important measures used to describe forests, as it directly relates to the competitive interactions between plants, fiber yield, stand structural complexity, and habitat suitability for many organisms. As valuable as this information is, this metric is often neglected because the determination of total tree height is a time-consuming process prone to error if improperly done. As an example, those conducting large-scale forest inventories often choose to predict tree height as a function of a much easier to assess attribute (bole diameter) rather than measuring it directly (Bechtold et al. 1998, Barrett 2006). Modeling height may not optimally fit any given tree, but over the course of a large inventory, it often proves an economic balance of measurement efficiency and accuracy (Barrett 2006).

Unfortunately, our need to reliably measure tree height often exceeds our capacity to accurately forecast this variable. It is not because we lack the tools to do so—there are many models to predict tree height from diameter. Rather, choosing the appropriate model using the best measurement technique has not been done for most species in most locations. This is true even for commercially important taxa at major research locations. For instance, we have no local height equations for loblolly (*Pinus taeda*) or shortleaf (*Pinus echinata*) pine on the Crossett Experimental Forest in Ashley County, Arkansas, even though scientists have studied these taxa there since the mid 1930s.

Local height equations are particularly valuable in that they are derived using specimens found in the immediate study area. Thus, these equations better reflect nuances in tree allometry attributable to local conditions. Theoretically, such a model is preferable to ones developed for individual states, or even the entire southeastern United States (e.g., Bechtold et al. 1998, FVS Staff 2008). Hence, this study was initiated with the objective of developing the most accurate height-diameter model possible given a sample of loblolly and shortleaf pines taken from the Crossett Experimental Forest and surrounding areas.

Materials and Methods

Study areas

The Crossett Experimental Forest (CEF) covers 680 ha in the extreme southern portion of Ashley County, 11 km south of the city of Crossett. The CEF landscape is dominated by upland forests of loblolly and shortleaf pine, with a minor and varying hardwood component. Most of the soils on the CEF are silt loams, and are considered to be of good quality for growing pine, with loblolly site index of 25 to 30 m (50 year base age) (Gill et al. 1979). Virtually all of the pines on the CEF are of natural origin (i.e., naturally regenerated from local seed sources). Pine seedlings, saplings, and small poles are abundant across the experimental forest. However, most of the pine overstory on the CEF is mature (> 40 years old), with extensive areas of even-aged and uneven-aged stands (Baker and Bishop 1986). In certain locations, sawtimber-sized individuals exceeding 75 cm in diameter can be found, although silvicultural practices usually mean this is an upper size limit.

Because of the maximum pine size threshold imposed by decades of intensive management over most of the CEF, a small number of exceptionally large loblolly (25 trees) and shortleaf (19 trees) pine were sampled on the nearby Levi Wilcoxon Demonstration Forest (LWDF). The LWDF is an unmanaged old-growth pine-dominated stand owned by Plum Creek Timber Company located roughly 16 km to the northeast of the CEF in Ashley County (Bragg 2004). The LWDF occurs on comparable landforms, has a similar range of site qualities as seen on the upland forests of the CEF, and (because of its age) has substantially larger specimens of both loblolly and shortleaf pine than the CEF.

Sample tree selection and measurement

Most of the CEF sample of pines across the full range of diameters at breast height (DBH, or the stem diameter at 1.37 m above the ground surface) were collected by systematically locating four 0.13-ha circular plots in randomly selected compartments. A number of additional small diameter pines were sampled along the roads on the CEF to ensure these size classes were not underrepresented. As stated earlier, a few dozen trees were measured on the LWDF to supplement the CEF loblolly and shortleaf samples.

Pines less than 3 cm DBH had their DBH measured to the nearest millimeter using a hand caliper, and larger pines had their DBH measured (to the mm) with a steel diameter tape. For the 415 loblolly pines examined, DBH ranged from 0.3 cm to 108.2 cm, and of the 143 shortleaf pines sampled, DBH varied from 0.3 cm to 91.9 cm (Figure 1). Without the LWDF additions, the maximum CEF loblolly and shortleaf pine diameters would have been 78.0 cm and 80.8 cm, respectively.

Total tree height was measured using one of two approaches. For pines up to about 10 m tall, a telescoping pole was used to estimate height to the nearest 3 cm. Taller trees were measured using a TruPulse™ 200 laser hypsometer following the sine method of height determination. The sine method is more accurate and precise than the tangent-based approach incorporated in the factory-default TruPulse height routine because it directly measures the crown of a tree, rather than approximating it with angles and a baseline distance (Blozan 2006, Bragg 2008).

With the sine method and the TruPulse hypsometer, pine height can be reliably estimated to the nearest 15 cm for very large trees (the accuracy of the tangent method with this hypsometer is probably between ± 1 to 3 m, and can exceed 5 m for some trees).

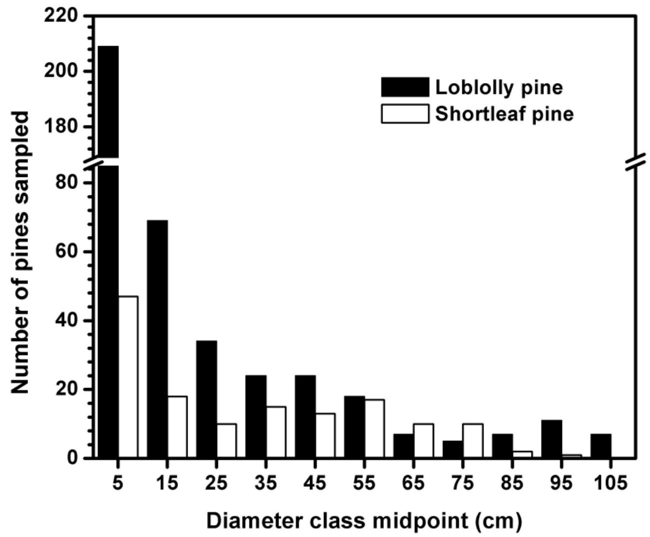

Figure 1. Diameter class distribution of loblolly and shortleaf pines selected for height-diameter model fitting.

Height model selection and statistical comparisons

Many height-diameter models exist—for instance, Huang et al. (2000) evaluated 27 different functions for stands of white spruce (*Picea glauca*) in the boreal forests of Canada. Rather than testing the scores possible, a dozen commonly used height-diameter models were fit using weighted nonlinear least squares regression with DBH as the only independent variable (and the inverse of DBH as the weight of the loss function). These were compared graphically for their fit to the data, and the four best performers (the Chapman-Richards, modified logistic, exponential, and Curtis-Arney functions) were retained for further comparison.

The Chapman-Richards function is as follows:

$$HT = 1.37 + b_1\left(1 - e^{-b_2 DBH}\right)^{b_3} \quad (1)$$

where predicted pine height (*HT*, in m) is a function of *DBH* and a set of species-specific coefficients (b_1, b_2, b_3, …, b_n). The modified logistic equation:

$$HT = 1.37 + \frac{b_1}{1 + \left(1/b_2\right)DBH^{b_3}} \quad (2)$$

and the Curtis-Arney (also known as the Korf/ Lundqvist) function:

$$HT = 1.37 + b_1\left(e^{b_2\left(DBH^{b_3}\right)}\right) \quad (3)$$

also used the same predictor and same number of coefficients. Finally, the exponential function:

$$HT = 1.37 + b_1\left(e^{b_2+b_3\left(b_4/\left(DBH+b_5\right)\right)}\right) \quad (4)$$

applied five coefficients but the same variable as the previous models.

As a final comparison, the height dubbing function of the Southern Variant of the Forest Vegetation Simulator (FVS) was used directly from its source (FVS Staff 2008). Unless provided by the user, the FVS height dubbing function is used to calculate height for every tree processed by the model. Hence, the coefficients given in the Southern Variant description (which covers Arkansas) are assumed applicable without modification to the pine sampled in this paper. Significant departures of the FVS height dubbing model from expectations are important, as this model is extensively applied across the region.

Models were evaluated using a series of goodness-of-fit measures, including fit index (R^2), root mean square error (RMSE), bias, and corrected Akaike information criterion (AICc). The fit index used in the statistical analysis package (Statistica, version 8.0) is a nonlinear analog to conventional R^2 used in linear regression (i.e., sum of squares residual (*SSR*) divided by the total sum of squares (*SST*)). RMSE equals:

$$RMSE = \sqrt{\sum_{i=1}^{n}\left(HT_i - \hat{H}T_i\right)^2 \Big/ \left(n-p\right)} \quad (5)$$

where HT_i is the height of the i^{th} pine, $\hat{H}T_i$ is the predicted height of that pine, n is the total number of observations, and p is the number of function parameters. Bias was determined by:

$$Bias = \sum_{i=1}^{n}\left(\hat{H}T_i - HT_i\right)\Big/ n \quad (6)$$

where bias is negative if the predicted height is less than the actual (measured) height. AICc is a measure that allows for the comparison of multiple models with differing numbers of parameters:

$$AICc = 2p + n\left(\ln\left(\hat{\sigma}^2\right)\right) + \frac{2p(p+1)}{n-p-1} \quad (7)$$

and $\hat{\sigma}^2 = \sum \hat{\varepsilon}_i^2 \Big/ n$. This version of the AIC is preferable because it has a second order correction for limited sample sizes (Burnham and Anderson 2002). Smaller AICc values indicate better models.

Results and Discussion

Evaluating model fit

The functional forms in this paper did a good job of matching the overall trends in tree size. All of the models explained between 96 and 98% of the variation in both pine species (Table 1). Similarly, few differences were apparent in RMSE, bias, and AICc. For all but the Curtis-Arney equation, RMSE averaged around 1.85 m for loblolly pine and 2.43 m for shortleaf, suggesting that departures between predicted and actual heights were limited (even the Curtis-Arney differed by only 2 m). These models showed little evidence of bias in their fit, regardless of species.

While there were subtle differences in the AICc values for all functional forms within each species, only the Curtis-Arney departed noticeably from the others. The Chapman-Richards, modified logistic, and exponential functions were within 4% of the others' AICc scores for both pine species, and 4.8 to 20.4%, respectively, with the Curtis-Arney equation.

Figures 2 and 3 show how each model form fit the actual loblolly and shortleaf pine data, respectively. Importantly, each of the functions tracked the relationship between height and diameter in both pine species well, including the rapid increase in height with diameter at small DBH, followed by a slowing trend as the trees reached moderate (30 to 40 cm DBH) size.

Table 1. Sample size and goodness-of-fit measures by height-diameter model for pines from the CEF and LWDF.

Height-diameter model	n	R^2	RMSE	Bias	AICc	ΔAICc	%AICc
			Loblolly pine				
Chapman-Richards	415	97.56	1.83	0.004	505.73	0.00	100.0
Modified logistic	415	97.49	1.86	0.010	515.91	10.18	102.0
Exponential	415	97.44	1.88	-0.065	525.01	19.28	103.8
Curtis-Arney	415	96.86	2.07	0.018	608.87	103.14	120.4
			Shortleaf pine				
Chapman-Richards	143	96.53	2.43	0.006	256.76	1.95	100.8
Modified logistic	143	96.47	2.45	0.010	259.20	4.39	101.7
Exponential	143	96.58	2.41	-0.022	254.81	0.00	100.0
Curtis-Arney	143	96.27	2.52	-0.080	267.01	12.20	104.8

ΔAICc = model AICc – minimum AICc; %AICc = (model AICc/minimum AICc) x 100

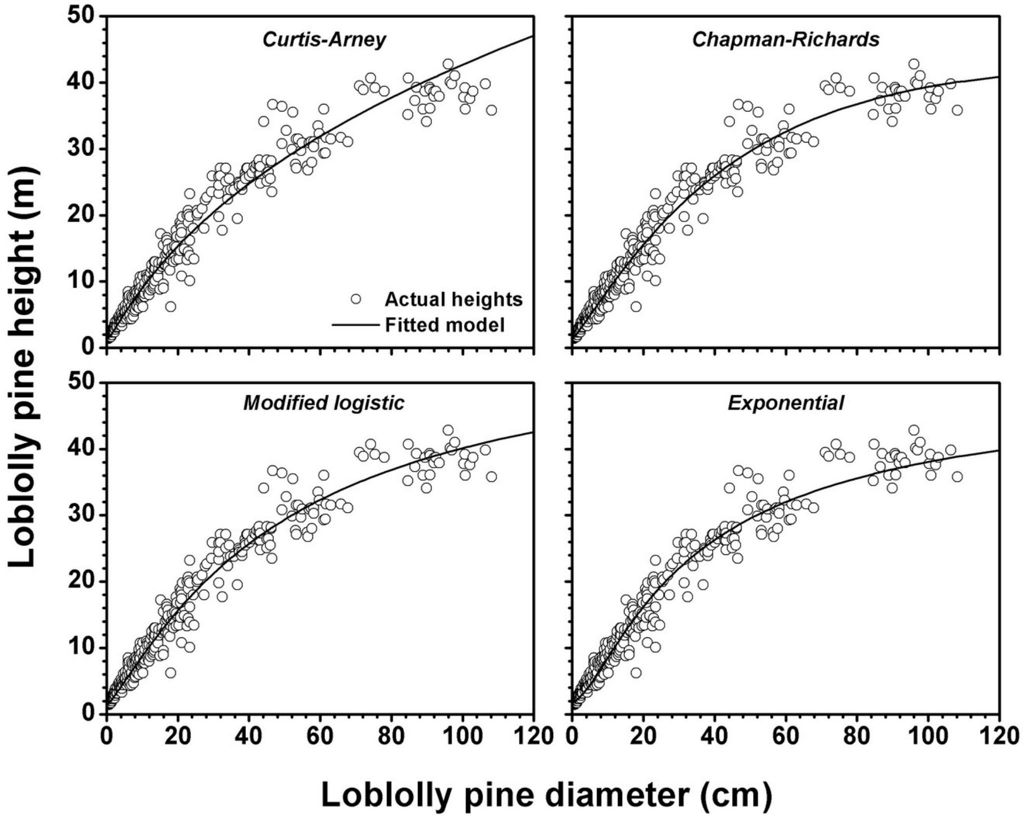

Figure 2. Fit of different models to the loblolly pine data used to derive the equations.

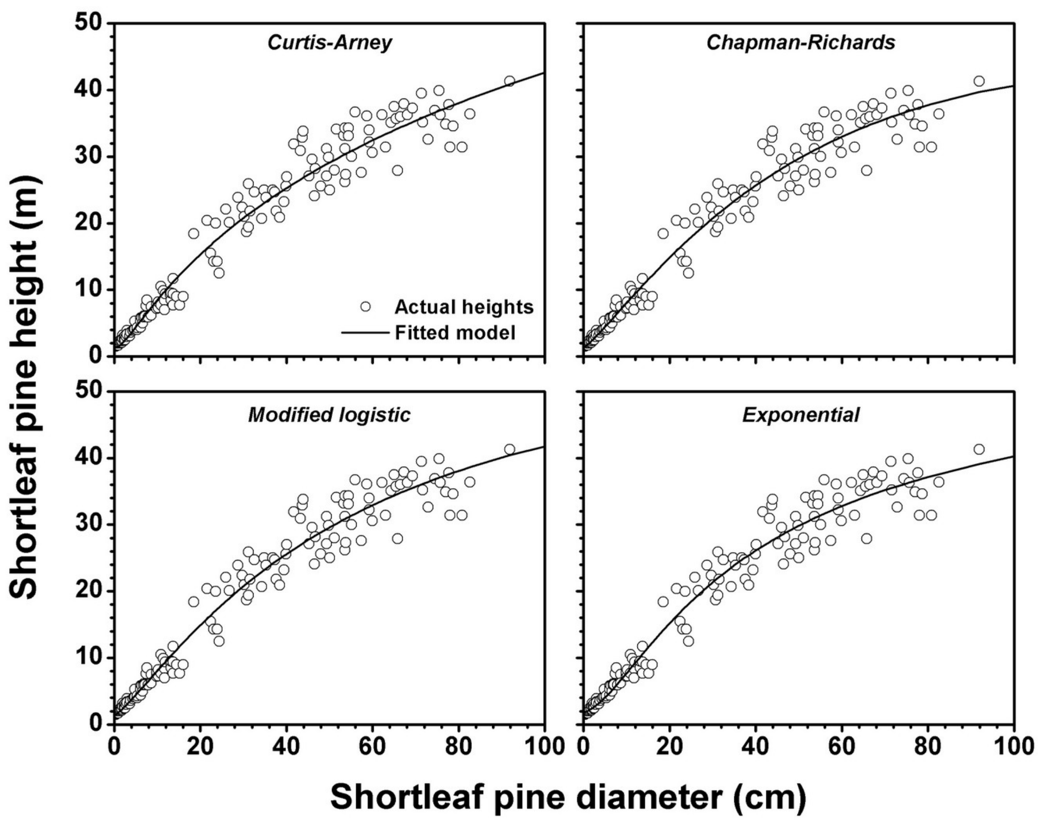

Figure 3. Fit of different models to the shortleaf pine data used to derive the equations.

Model comparisons

Further comparisons can be made by including the default equation from the Forest Vegetation Simulator (FVS—see Table 2 for coefficients of both species for all equation forms). The FVS height equations, which use the Curtis-Arney form, are provided for evaluation only—it would be inappropriate to compare their fit to those models developed in this paper because of differences in the data used.

It was clear that the Curtis-Arney function fit both pine species most poorly across the range of the diameters (Table 1). Figures 2 and 3 indicate that the most prominent departures created by using the Curtis-Arney function occur in the largest size classes. The Curtis-Arney notably over-predicts the height of large diameter pines, a trend especially noticeable with loblolly pine (Figure 2). This tendency is even more apparent when extended towards the upper size limits of both species (Figures 4 and 5). For instance, the Curtis-Arney equation predicts an almost 60 m tall loblolly pine at 200 cm DBH, an improbable height for this species in Arkansas.

The champion-sized loblolly pines in Figure 4 are intended to provide context for predictions beyond the original data range. Note that the only one of these trees measured with the same sine-based method used in this paper is the current national champion, located in the Congaree Swamp National Park in South Carolina. The other, more local champions were probably measured with either the tangent or similar triangle methods, both of which can be much less reliable (Blozan 2006, Bragg 2008). Unfortunately, at least three of these champion trees are now dead and cannot be remeasured with the sine method to verify their heights. If we assume that the heights reported for these trees are reasonable, it can be inferred from Figure 4 that most of the equations would do a reasonable job of predicting very large loblolly pines.

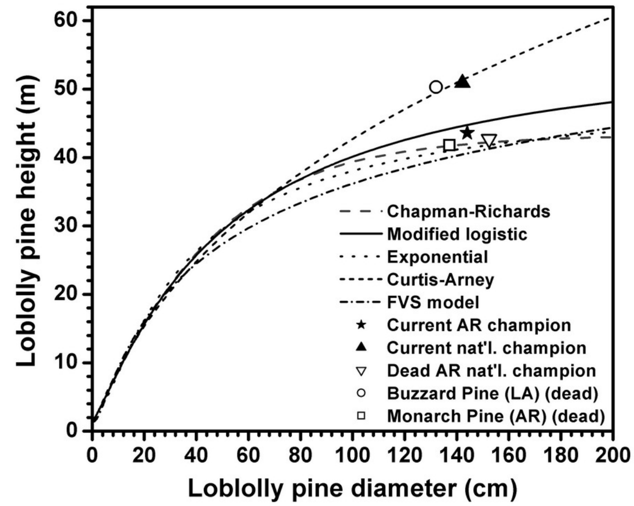

Figure 4. The height predictions of the loblolly pine models derived for this paper, the FVS dubbing equation, and a handful of champion-sized pines. Loblolly pine height predictions are extended out to 200 cm DBH (a reasonable upper size possibility for this species, at least from historical records) to show the results of extrapolations beyond the range of the original data.

The Curtis-Arney function is probably too high in its height predictions, although it would do a better job of fitting the current national champion. However, the national champion loblolly pine is growing in a very favorable site (a fertile bottomland), which is not representative of conditions on the CEF.

It is apparent in Figures 4 and 5 that only subtle differences in curve shape, and hence, height prediction, are realized in small- to moderate-sized pines, regardless of the model used. Even the FVS height dubbing equation does remarkably well up to about 40 cm DBH for both loblolly and shortleaf pine.

Table 2. Model coefficients by pine species from individuals measured on the CEF and LWDF.

Species	Coefficient	Chapman-Richards	Modified logistic	Exponential	Curtis-Arney	FVS dubbing [a]
Loblolly pine						
	b_1	41.9641	55.9834	2.2595	499.0730	243.8606
	b_2	0.0247	0.0103	3.0866	-7.0057	4.2846
	b_3	1.1496	-1.1703	-10.6490	-0.2246	-0.4713
	b_4	--	--	3.0016	--	--
	b_5	--	--	6.5158	--	--
Shortleaf pine						
	b_1	44.3850	59.8416	5.0109	195.5000	444.0922
	b_2	0.0235	0.0076	2.4111	-7.0638	4.1188
	b_3	1.2117	-1.2175	-10.7870	-0.3287	-0.3062
	b_4	--	--	3.6285	--	--
	b_5	--	--	7.9802	--	--

[a] Coefficients given for loblolly and shortleaf pine height equations (Curtis-Arney functions) taken from FVS Staff (2008).

However, total tree heights for larger shortleaf pine on the CEF would be significantly under-predicted using the current FVS height model, as would the biggest of the loblolly pines.

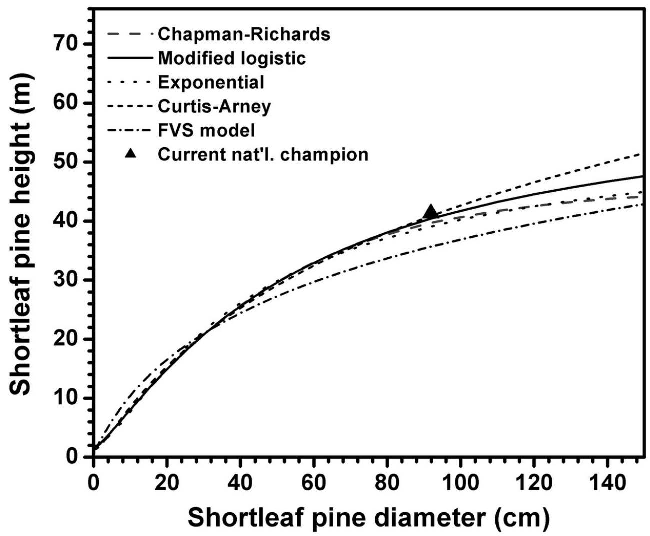

Figure 5. The height predictions of the shortleaf pine models derived for this paper, the FVS dubbing equation, and a single champion-sized pine (the current national champion shortleaf pine, found on the LWDF). Shortleaf pine height predictions are extended out to 150 cm DBH (a reasonable upper size possibility for this species, at least from historical records) to show the results of extrapolations beyond the range of the original data.

Model recommendation for the CEF

Given their overall similarity in performance, the modified logistic function was the preferred height-diameter model because of its more intuitive allometry at the upper extreme of pine size. This recommendation is made in part of how much the other equations (with the exception of the Curtis-Arney, which has already been rejected because of its behavior with large diameter pines) flatten in their height projections over 100 cm DBH. Even though the modified logistic equation's ΔAICc value (Table 1) is generally interpreted as providing only limited support for the equivalence of this model and the exponential and Chapman-Richards equations, the differences were not drastic. More importantly, the modified logistic function allows for some height increment in these big trees without being too aggressive. Thus, it is capable of capturing the likely allometric patterns of very large trees without significant departures from the more conservative height-diameter models at small to moderate diameters (Figures 6 and 7).

A comparison of the modified logistic and FVS models show that there are definite advantages in using a local height equation. For loblolly pine, the difference between the two differed little until moderate-sized diameters are reached, after which the FVS equation noticeably under-predicts loblolly height. Shortleaf pine behaved somewhat differently, with the FVS model slightly over-predicting heights for some small diameter pines (Figure 7) and under-predicting heights for moderate to large shortleaf.

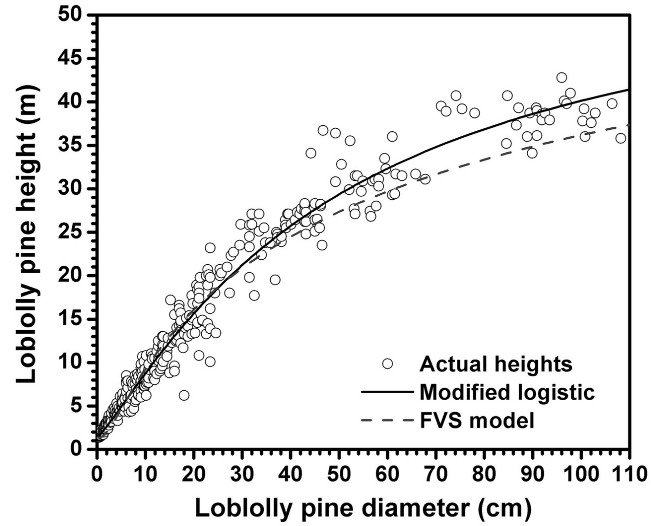

Figure 6. Comparison of recommended modified logistic and FVS height-diameter models on loblolly pine from the CEF and LWDF.

There would likely be little impact of the slight overestimate for small shortleaf, but at the largest size classes of both pine species, the FVS model would under-predict heights by about 5 m. Given that the CEF is primarily managed for sawtimber, such a departure could have significant ramifications when the current FVS model is applied.

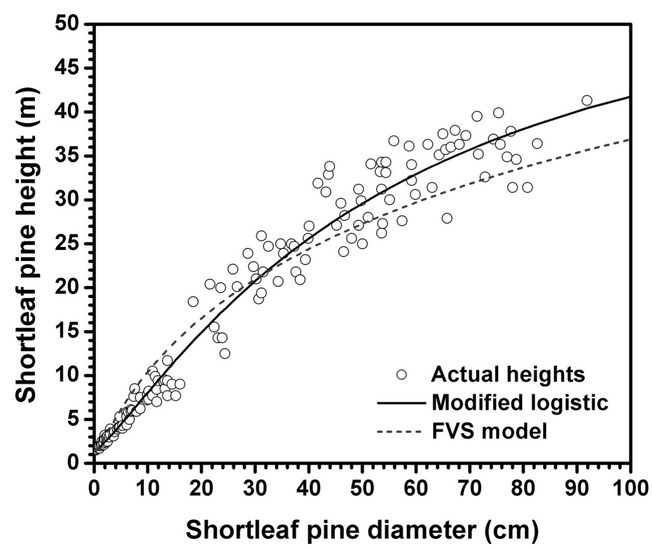

Figure 7. Comparison of recommended modified logistic and FVS height-diameter models on shortleaf pine from the CEF and LWDF.

A unified pine model

A cursory examination of the data for both loblolly and shortleaf pine suggested that there were few differences in the height-diameter allometry between these species for the CEF—so few, in fact, it is possible to derive a "unified" height-diameter model to project either species. The following modified logistic equation was fit to all 558 pines:

$$HT = 1.37 + \frac{57.4042}{1 + 103.9933 DBH^{-1.1760}} \quad (8)$$

and explained over 97% of the variation (Figure 8).

A unified model, though slightly biased and not as precise as one developed for each species (Table 3), does have a number of key advantages. For instance, distinguishing between small stature loblolly and shortleaf pine can often prove difficult in the CEF area, especially when the young twigs cannot be examined. A generic pine model makes it less critical that species are known exactly in order to predict their height.

It is also appropriate to use equation (8) to assist in stand structure reconstructions from historical inventories that may not be adequately differentiated—General Land Office surveyors in Arkansas, for example, did not separate pines into loblolly or shortleaf, but rather called any member of the genus *Pinus* "pine." In this example, the uncertainty in taxonomic classification cannot be corrected. The use of this generic model should provide more appropriate estimates of pine height, regardless of species.

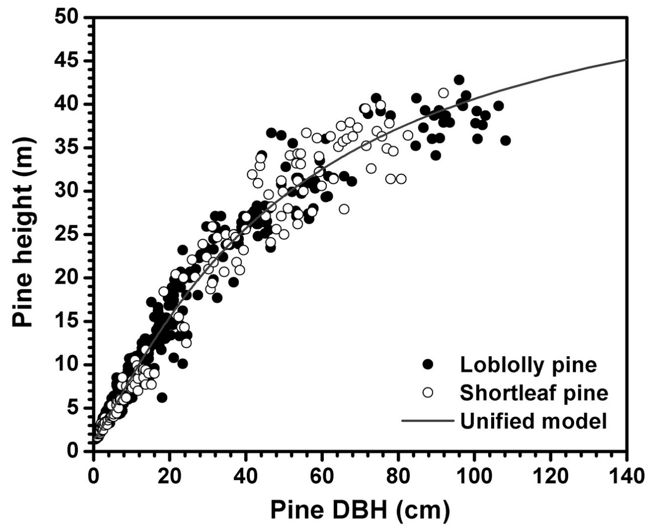

Figure 8. Height-diameter equation fit to all pine data sampled in the CEF and LWDF.

Table 3. Comparison of model height predictions using the modified logistic regression models developed specifically for loblolly, shortleaf, and both pines combined.

	Predicted height (m)					
DBH (cm)	Loblolly model	Unified model	Difference	Shortleaf model	Unified model	Difference
5	4.91	4.81	-0.10	4.44	4.81	0.37
25	18.58	18.45	-0.13	17.96	18.45	0.49
45	27.62	27.67	0.05	27.68	27.67	-0.01
65	33.61	33.85	0.24	34.35	33.85	-0.50
85	37.77	38.17	0.40	39.07	38.17	-0.89
105	40.79	41.33	0.54	42.52	41.33	-1.19
125	43.07	43.71	0.64	45.14	43.71	-1.42
145	44.84	45.57	0.73	47.17	45.57	-1.60

Conclusions

A local set of height-diameter equations is helpful when examining the patterns of tree allometry, especially for an area in which extensive scientific work is being conducted. This avoids the vagaries of models developed for other regions while allowing for the unique attributes of growth patterns in a specific location to be expressed. The data from this study confirm that a local height equation yielded a meaningful improvement in prediction accuracy when compared to the generic model incorporated in the FVS simulator.

This study also showed that if the bounds of the field data were not violated, there are many equations capable of expressing the relationship between pine height and diameter on the CEF.

The recommendation of the modified logistic function was made not because it was the absolute best fit of the data, but rather it fit the data comparably well and it seemed to do a more reasonable job of projecting pine height beyond the upper range of the diameters sampled. Such a trait is desirable, because even though it is statistically inappropriate to extend models beyond the range of data from which they were derived, users will almost inevitably do so—or may do so unwittingly, if the height equation is incorporated in a larger model system. Hence, it is logical to use a model form that behaves reasonably for any conceivable diameter that may occur.

Acknowledgments

I would like to thank Mike Chain and Kirby Sneed (both of the USDA Forest Service) for their assistance in measuring the pines and entering the data. Thanks also to Conner Fristoe and Plum Creek Timber Company for access to the LWDF. Nancy Koerth, Mike Shelton, Curtis VanderSchaaf, and three anonymous reviewers provided helpful comments on this manuscript.

Literature Cited

Baker JB and LM Bishop. 1986. Crossett demonstration forest guide. USDA Forest Service General Report R8-GR 6. 55 p.

Barrett TM. 2006. Optimizing efficiency of height modeling for extensive forest inventories. Canadian Journal of Forest Research 36:2259-2269.

Bechtold WA, SJ Zarnoch, and WG Burkman. 1998. Comparisons of modeled height predictions to ocular height estimates. Southern Journal of Applied Forestry 22(4):216-221.

Blozan W. 2006. Tree measuring guidelines of the Eastern Native Tree Society. Bulletin of the Eastern Native Tree Society 1(1):3-10.

Bragg DC. 2004. Composition, structure, and dynamics of a pine-hardwood old-growth remnant in southern Arkansas. Journal of the Torrey Botanical Society 131(4):320-336.

Bragg DC. 2008. An improved tree height measurement technique tested on mature southern pines. Southern Journal of Applied Forestry 32(1):38-43.

Burnham KP and DR Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. 2nd ed. New York: Springer-Verlag. 488 p.

Forest Vegetation Simulator Staff (FVS Staff) (Forest Management Service Center, USDA Forest Service, Fort Collins, CO). 2008. Southern (SN) Variant overview of the Forest Vegetation Simulator. Open file report. Fort Collins (CO): USDA Forest Service. 62 p. <www.fs.fed.us/fmsc/ftp/fvs/docs/overviews/snvar.pdf> Accessed on 5 Mar 2008.

Gill HV, DC Avery, FC Larance, and CL Fultz. 1979. Soil survey of Ashley County, Arkansas. USDA Soil Conservation Service and USDA Forest Service. 92 p.

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