A 3-POINT DERIVATION
OF DOMINANT TREE HEIGHT EQUATIONS

Don C. Bragg[1]

Abstract.—This paper describes a new approach for deriving height-diameter (H-D) equations from limited information and a few assumptions about tree height. Only three data points are required to fit this model, which can be based on virtually any nonlinear function. These points are the height of a tree at diameter at breast height (d.b.h.), the predicted height of a 10-inch d.b.h. tree from an existing H-D model, and the height at species maximum d.b.h., estimated from a linear regression of big trees. Dominant sweetgum (*Liquidambar styraciflua* L.) from the Arkansas region and yellow-poplar (*Liriodendron tulipifera* L.) from across the southeastern United States were used to estimate height at species maximum d.b.h. A composite of these field-measured heights and site index trees from the U.S. Forest Service's Forest Inventory and Analysis (FIA) database were used to compare the 3-point equations (fit to the Chapman-Richards model) with the Forest Vegetation Simulator (FVS) default H-D models. Because of the limited range of diameters in the FIA site trees, the Chapman-Richards equations developed from site trees under-predicted large tree heights for both species. For the sweetgum, the 3-point equation was virtually identical to the FVS default model. However, the 3-point equation noticeably improved dominant height predictions for yellow-poplar.

INTRODUCTION

Simulation models are increasingly used to project forest dynamics over large spatial and temporal scales. The Forest Vegetation Simulator (FVS), for example, now encompasses virtually the entire continental United States, as well as parts of Alaska and Canada (Dixon 2009). Most of these models incorporate several mathematical relationships to define the nature of simulated trees and stands, usually in the form of regression equations derived from field-sampled data. One of the best examples of these models can be found in height-diameter (H-D) models, which use any of a number of functions to describe the trend between total tree height and (typically) diameter at breast height (d.b.h.).

Numerous H-D models have been developed for North American tree species (for example, Huang and others 1992, 2000; Lootens and others 2007; Zhang 1997). However, it is unusual for H-D models to be derived across the size range of a given taxon, primarily because few data sets include very large trees. Centuries of lumbering, land clearing, and catastrophic disturbance have greatly reduced the number of big trees, making their presence rare in most inventories. Some have sought to improve upon existing allometric relationships by expanding the range of data via the incorporation of champion-sized trees (for instance, Bragg 2008b, Shifley and Brand 1984). This approach has some limitations, including sensitivity to the individual chosen as the champion tree and the possibility that the specimen selected may not adequately represent the maximal expression of height and diameter, especially for open-grown champions (Bragg 2008b).

---

[1]Research Forester, Southern Research Station, U.S. Forest Service, P.O. Box 3516 UAM, Monticello, AR 71656. To contact, call (870) 367-3464 or email at dbragg@fs.fed.us.

Regional H-D models are often used in forest simulators because of their relatively simple and consistent implementation. However, these regional H-D models can suffer from inadequacies in their derivation. One possible option is to use some form of local adjustment for stand conditions (such as mixed-effects modeling), although this approach would require additional processing of the original data, and without more field sampling is not capable of extending the range of data. In this paper, I present a new approach to creating H-D equations for sweetgum (*Liquidambar styraciflua* L.) and yellow-poplar (*Liriodendron tulipifera* L.) from limited field data and existing models.

STUDY AREAS

Field data on tree heights for this study were gathered from two distinct sources. Big tree heights were taken with the sine height measurement technique (Bragg 2008a) by the author and other members of the Eastern Native Tree Society (ENTS) on several visits to different forests. For sweetgum, this sampling occurred in the Arkansas area, and included trees from Overton Park (Memphis, TN), the Levi Wilcoxon Demonstration Forest (Ashley County, AR), Hyatt's Woods (Drew County, AR), Noxubee National Wildlife Refuge (near Starkville, MS), and Big Oak Tree State Park (Mississippi County, MO). The more geographically diverse yellow-poplar sample came from stands in the Bankhead National Forest, AL; Cohutta Wilderness, GA; Ocmulgee Flats, GA; Panther Creek, GA; Sosbee Cove, GA; Sugar Cove, GA; Blanton Forest, KY; Bullen Creek, MS; Western North Carolina Nature Center, NC; Shelton Laurel, NC; the Great Smoky Mountains National Park, NC and TN; Tamassee Knob, SC; Lee Branch, SC; Meeman-Shelby State Forest Park and Overton Park near Memphis, TN; Savage Gulf, TN; and Fall Creek Falls State Resort Park, TN.

Site index trees from the U.S. Forest Service's Forest Inventory and Analysis (FIA) program were combined with the ENTS height data to help validate the models' predictions. For sweetgum, this sample consisted of 298 trees from the FIA plots across the state of Arkansas. For yellow-poplar, 2,860 FIA site index trees from Alabama, Georgia, Mississippi, North Carolina, South Carolina, and Tennessee were used. These publically available data were downloaded from the FIA website (http://fiatools.fs.fed.us/fiadb-downloads/datamart.html).

METHODS

The use of limited numbers of data points to fit tree attributes is not new—for example, Zeide (1999) developed a system to predict height growth in southern pines using only two points. His design assumed simple two-parameter equations, while many H-D models use at least three. Hence, for this mathematical reason, as well as the biological behavior of most trees, I will pursue a 3-point design.

NONLINEAR MODEL FITTING APPROACH

The 3-point H-D derivation approach makes several assumptions about tree height and the nonlinear nature of its relationship to d.b.h. Under ideal circumstances, only three data pairs are needed to fit simple nonlinear equations—one to establish the beginning point, one to express the most common portion of the data range, and a final one to delimit the maximum value of an asymptotic H-D function (Fig. 1).

Fortunately, the first point to establish the beginning of the range can be considered a universal constant—by definition, a tree that has just obtained a d.b.h. has a height of 4.5 ft, regardless of species, site quality, stand density, and so on. Thus, this data pair can be defined as (>0.0, 4.5).
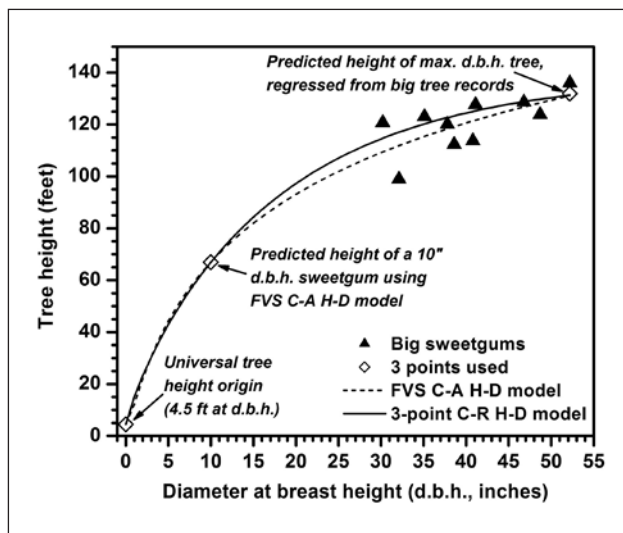
Figure 1.—A diagram of how the 3-point dominant H-D modeling approach worked for sweetgum from the Arkansas area. The three points (open diamonds) are chosen as the universal tree height origin, the predicted height of a tree at 10 inches d.b.h. (using a reasonable model), and the height of a tree of the "biggest" d.b.h., predicted from a simple linear regression of big tree records.

The second data point should be taken from where a preponderance of tree diameters can be found. Across most of the Central Hardwoods Region, centuries of logging and other disturbances have greatly reduced the number of very large stems, so the overwhelming majority of trees have relatively small diameters. Since this work focuses on species with considerable potential to grow large (in other words, no large shrubs), a tree of modest diameter (say, 10-inch d.b.h.) should provide a reasonable approximation of this central tendency. But what height should be used for this 10-in. d.b.h. stem? It would be possible to take samples either from the field or a publically available database (such as FIA), but it is easier to use an existing H-D model to predict its height. For trees of this size, most H-D models were derived from an abundance of trees. It is unlikely that there will be substantial differences between most regression equations at modest diameters—this conclusion is apparent in other research comparing different H-D functions (for example, Huang and others 1992, 2000).

The final information needed for the 3-point approach is a reliable estimate of the maximum height. Because virtually every H-D equation is designed to be monotonically increasing, by definition tree height will be maximized at the species' maximum diameter. Though there is no biologically maximal tree diameter as there appears to be for height (for example, Ryan and Yoder 1997, Koch and others 2004, Domec and others 2008), there is a general frontier that can be considered to be an upper diameter limit. For this paper, I assumed that the maximum diameters were 52.2 inches for sweetgum (from a sample of 10 very large individuals) and 87.2 inches for yellow-poplar (from a sample of 72 very large individuals). These data were then used to fit linear ordinary least squares regression equations to produce a maximum tree height estimate for the actual (measured) maximum diameters.

I followed the same approach for developing the 3-point regression equations for both species. The minimum pair tree data followed the universal height starting point, while the moderate pair (10-inch d.b.h.) height was predicted using the Curtis-Arney equation of the Forest Vegetation Simulator (FVS) (Forest Vegetation Simulator Staff 2009) and the maximum pair was generated as described in the previous paragraph. Three sets of data pairs (Table 1) were then fit with nonlinear regression to the Chapman-Richards function:

$$HT = 4.5 + b_1(1-e^{-b_2 \cdot d.b.h.})^{b_3} \qquad [1]$$

where HT = total tree height and b_1-b_3 are species-specific coefficients. The Chapman-Richards function is popular among H-D modelers because of its biologically interpretable parameters and simple yet flexible

Table 1.—Data pairs (d.b.h. = x, height = y) used to fit 3-point regression models to Chapman-Richards function. See text for definitions of these data pairs.

Species	At minimum		At moderate		At maximum	
	d.b.h.	height	d.b.h.	height	d.b.h.	height
	in.	ft.	in.	ft.	in.	ft.
Sweetgum	0.0001	4.5	10.0	66.9	52.2	131.9
Yellow-poplar	0.0001	4.5	10.0	69.6	87.2	171.2

form (Kershaw and others 2008). The FVS default (Curtis-Arney) equation (Forest Vegetation Simulator Staff 2009) is similar in some aspects:

$$HT = 4.5 + b_1 e^{(-b_2 \cdot d.b.h.^{b_3})} \tag{2}$$

but the parameters are not as interpretable and this particular equation does not generate an asymptote for most species within biologically reasonable heights. Note that the Chapman-Richards and 3-point models were designed to predict dominant tree heights, not the average height of the entire population. However, it is not unusual for researchers to exclude suppressed trees or those with broken tops (for example, Lootens and others 2007). Presumably, site index trees were chosen from dominant and co-dominant specimens in the stands where they were found—this approach was definitely true of the data in the ENTS big tree collection.

COMPARISON OF MODELS

The effectiveness of the 3-point approach was evaluated using a comparison between the three different regression models: the 3-point Chapman-Richards derived model, the FVS default height equation (Curtis-Arney), and a Chapman-Richards model fit only with the FIA site trees (CRSI). Goodness-of-fit comparisons are made using the following measures—fit index (R^2), root mean square error (RMSE), bias, and corrected Akaike Information Criterion (AIC_c):

$$R^2 = 1 - \sum_{i=1}^{n}(HT_i - \widehat{HT}_i)^2 / \sum_{i=1}^{n}(HT_i - \overline{HT})^2 \tag{3}$$

$$RMSE = \sqrt{\sum_{i=1}^{n}(HT_i - \widehat{HT}_i)^2/(n-p)} \tag{4}$$

$$bias = \sum_{i=1}^{n}(\widehat{HT}_i - HT_i)^2/n \tag{5}$$

$$AIC_c = 2p + n[1n(\hat{\sigma}^2)] + \frac{2p(p+1)}{n-p-1} \tag{6}$$

In these equations, height (HT) and predicted height ($\widehat{HT}$) are for the ith tree, n is the total number of observations, p is the number of function parameters, and $\hat{\sigma}^2 = \sum_{i=1}^{n}\hat{\epsilon}_i^2/n$, where $\hat{\epsilon}_i$ are the estimated residuals from the fitted model. In addition to these test statistics, visual comparisons in model performance have been made.

RESULTS

Nonlinear regression coefficients for the three different models of sweetgum and yellow-poplar can be found in Table 2. Of the models tested, the Chapman-Richards equation fit specifically to the FIA data (CRSI) fit the best for both sweetgum and yellow-poplar (Table 3). All sweetgum models did a reasonable job of projecting the data, with fit indices exceeding 0.52 in each case and only limited amounts of error and modest negative bias appearing in all of the models. However, using the ΔAICc criterion of 2 as the threshold for separating the models with the substantial evidence as the best approximating model from those with less support (Burnham and Anderson 2002), it appears that the 3-point model is just as good as the CRSI for sweetgum Table 3).

Figure 2 shows the sweetgum data (a) from the Arkansas FIA database and a comparison (b) of the FVS default H-D equation with the FIA site index-derived Chapman-Richards equation. While there is little apparent difference between these models up to about 25 inches d.b.h., there are growing differences as tree size increases. The 3-point and FVS H-D equations differ even less across the entire range of tree diameters considered, especially in the context of the large tree data (Fig. 2c).

Table 2.—Nonlinear regression coefficients for the three models.

Model by species	b_1	b_2	b_3	n
Sweetgum				
FVS default (Curtis-Arney)[a]	290.90548	3.62395	-0.37201	?
Chapman-Richards, SI only	107.017	0.10970	1.29264	298
3-point Chapman-Richards	134.359	0.05198	0.84997	3
Yellow-poplar				
FVS default (Curtis-Arney)[a]	625.76966	3.87321	-0.23349	?
Chapman-Richards, SI only	130.620	0.07882	1.05417	2,860
3-point Chapman-Richards	159.821	-0.10301	2.03246	3

[a] Coefficients from Forest Vegetation Simulator Staff (2009) and rounded off to 5 decimal places.

Table 3.—Goodness-of-fit statistics by species for comparisons among the FVS default height models, Chapman-Richards models fit to the FIA site index (SI) trees only, and the 3-point Chapman-Richards models, all using the SI and Eastern Native Tree Society height data for validation. All models have three parameters.

Model by species	R^2	RMSE	bias	AIC_c	ΔAIC_c*
Sweetgum (304 trees, average height = 76.6 ft)					
FVS default (Curtis-Arney)	0.527	10.936	-1.900	1457.5	13.0
Chapman-Richards, SI only	0.547	10.705	-0.122	1444.5	0.0
3-point Chapman-Richards	0.544	10.734	-0.798	1446.1	1.7
Yellow-poplar (2,932 trees, average height = 88.2 ft)					
FVS default (Curtis-Arney)	0.446	14.993	-7.620	15880.3	972.8
Chapman-Richards, SI only	0.602	12.701	-0.771	14907.5	0.0
3-point Chapman-Richards	0.446	14.986	4.730	15877.6	970.1

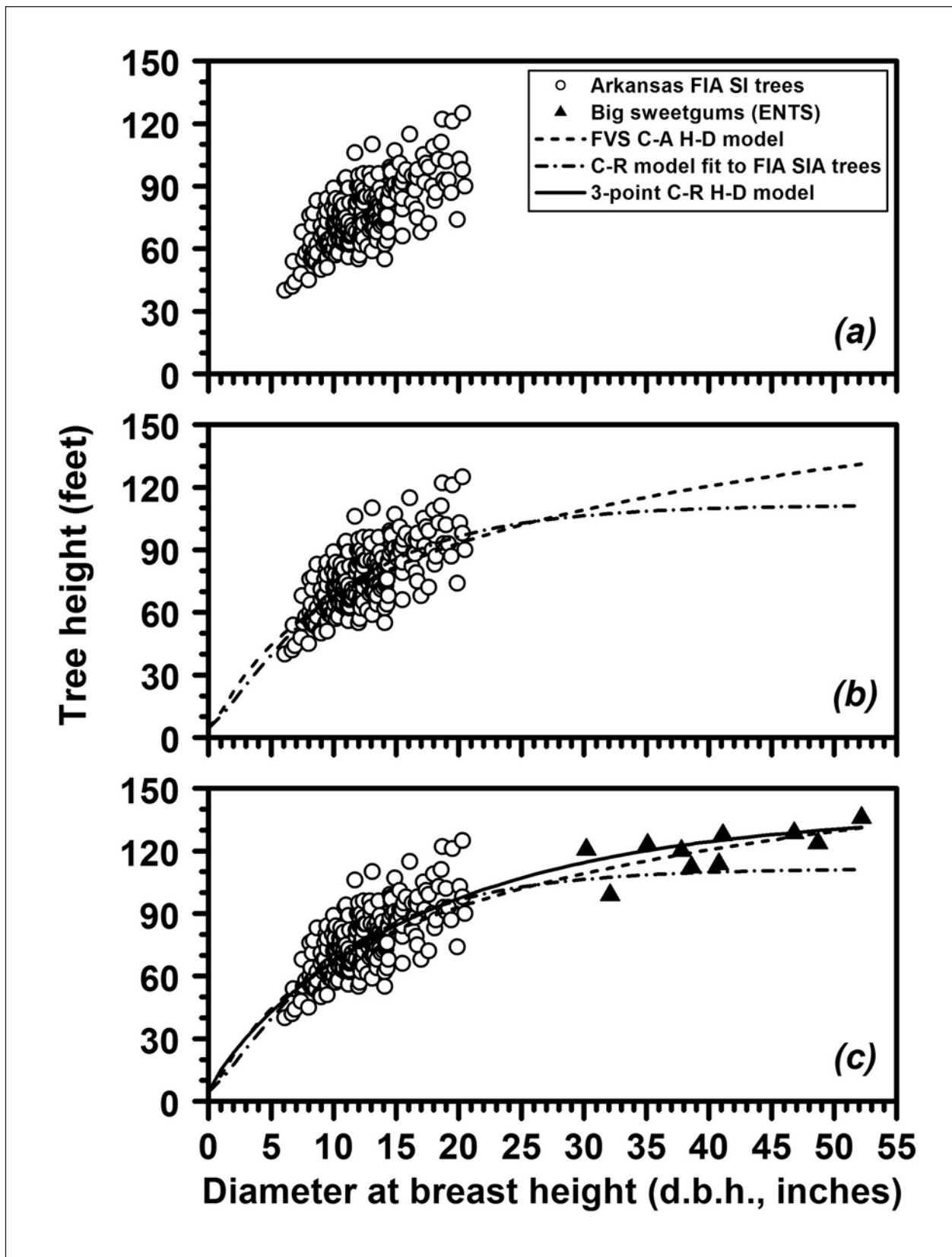* ΔAIC_c = current AIC_c – minimum AIC_c

Figure 2.—Validation of the 3-point method for Arkansas-area sweetgum height prediction, using the independent FIA site index tree data for the state (a). The FVS default H-D model and a H-D equation developed from the Chapman-Richards function fit to the FIA site index data differed little (b) until sweetgums exceeded 25 inches d.b.h. The 3-point and FVS default H-D models were virtually indistinguishable across much of the size range (c).

Yellow-poplar had substantially better goodness-of-fit measures with the CRSI than with the 3-point or FVS default (Curtis-Arney) H-D equations (Table 3). Both of these models had fit indices of 0.446 (compared to 0.602 for the CRSI model) and significantly greater bias and AICc scores (there are only very minor differences between the 3-point and FVS default equations). Figure 3 illustrates the differences among these models—unlike sweetgum, noticeable departures from the yellow-poplar 3-point model appeared for small-, medium-, and large-diameter trees with both of the other competing models. The test statistics (Table 3) are highly influenced by the large number of small- to moderate-sized yellow-poplars, so the better fit of the 3-point model across the large-diameter trees was not enough to substantially improve the goodness-of-fit outcomes.
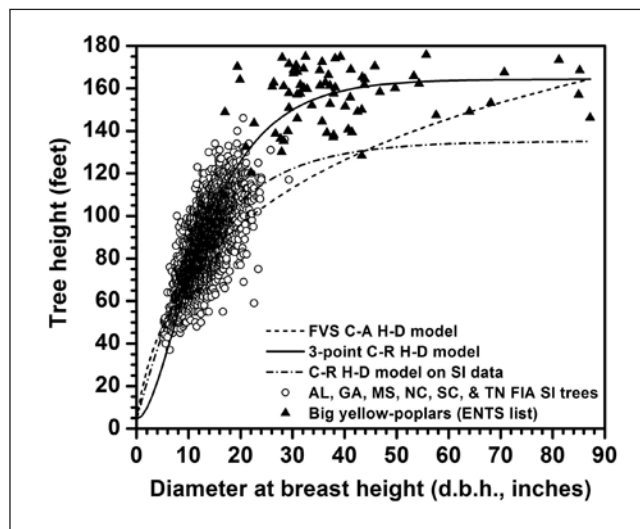

Figure 3.—H-D models developed for yellow-poplar across the southeastern U.S. differed considerably more than those for the Arkansas sweetgum. The 3-point model did a noticeably better job of fitting large-diameter trees.

Figure 4 highlights the differences in outcomes based on model type compared to the validation data. Unlike the sweetgum models from the Arkansas region, which showed very little difference, the yellow-poplar models varied considerably in their predictions. In general, the 3-point model predicted noticeably shorter trees at small diameters (up to 10-in. d.b.h.) and substantially greater heights for moderate- to large-sized yellow-poplars. The CRSI and FVS default H-D models for yellow-poplar differed less up to about 50-in. d.b.h., after which considerable differences in height were predicted between these models. Interestingly, the 3-point and FVS default models predicted virtually the same height for the tree of maximum d.b.h. with both species.
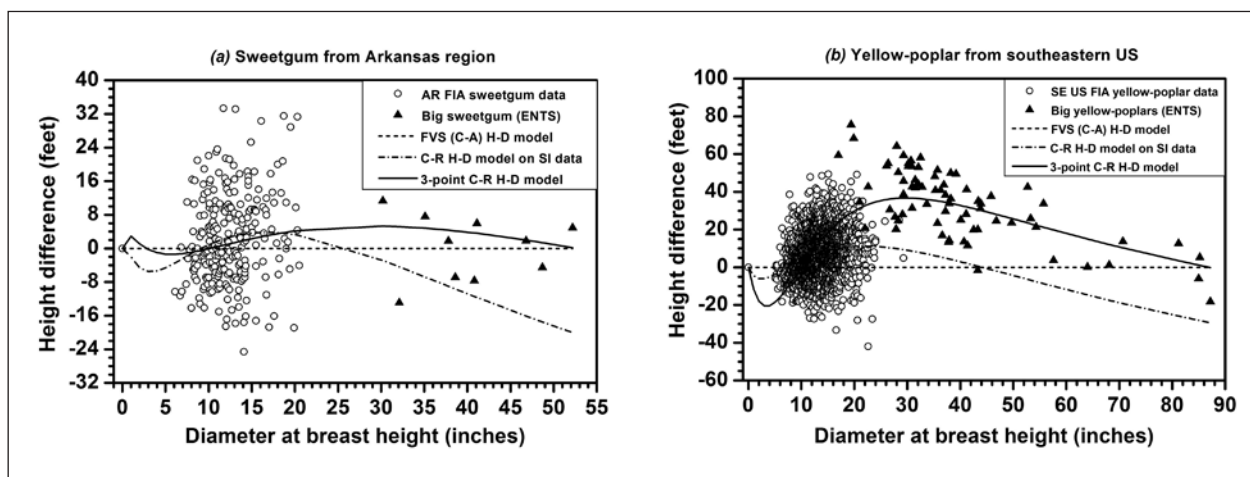

Figure 4.—Residual analysis for Arkansas-area sweetgum (a) and southeastern United States yellow-poplar (b). All comparisons are made relative to the published FVS default model (dashed line). Note the remarkable (and likely coincidental) agreement between the 3-point and FVS default models at the maximum diameter simulated for both species.

DISCUSSION

MODEL PERFORMANCE

It is not surprising that the CRSI models had the best fit to the validation data of the three compared—it was the only one using all of this information. This result applies especially to the yellow-poplar model because of the very large size (n = 2,932 trees) of the data set. However, conventional goodness-of-fit statistics are not necessarily the best measure of predictive success when models are fit to this type of data.

Most H-D data sets have a large quantity of noise—there simply are too many factors influencing actual tree height to account for them all in a single, simple equation. For example, the 95-percent confidence intervals for the equations Lootens and others (2007) fit to their data are quite wide, suggesting that any of several model forms could have been used and would have been capable of predicting height from diameter just about as reliably. Others have used additional factors (for example, dominant canopy height [Kershaw and others 2008] or local spatial patterns [Zhang and others 2004]) or more complicated model-fitting approaches (for example, mixed-effects [Trincado and others 2007, Budhathoki and others 2008] or geographically weighted regression [Zhang and others 2004]) to improve the fit of their equations.

However, relatively small gains in predictive ability, even if significant, are probably less meaningful than an accurate portrayal of the relationship between height and diameter across the full range of tree sizes (see also Lei and Parresol 2001). For instance, many researchers are concerned about the behavior of their models at the extremes of the diameter range (for example, Lei and Parresol 2001, Lootens and others 2007, Bragg 2008b), knowing that many such equations are often applied beyond the range of the data for which they were fit. In this study, the CRSI equation for yellow-poplar matched the validation data much better than either the 3-point or the FVS default equations, but because of the limited data range of the CRSI model, it had a maximum tree height asymptote of about 131 feet. Many yellow-poplar exceed this height across the southeastern United States.

This prediction highlights one of the biggest challenges to using a data set where most of the samples are in the smaller end of the possible tree size range. These data unduly influence model fit and may fail to reflect the true relationship of height and diameter as the trees reach larger size classes. Bragg (2008a) proposed a different model extension technique using champion tree data and a range of data points from the original H-D model to address this issue. Because of how Bragg's (2008a) approach is applied, however, it is more sensitive to a single big tree height than is this 3-point method.

FUTURE APPLICATIONS

The 3-point approach for dominant tree H-D equations is still in the earliest phases of development. These preliminary results nevertheless suggest that this approach offers a biologically and statistically reasonable alternative to other forms of model design. In an era of increasingly limited resources and a growing dependence on simulation models, the creation of accurate predictive systems with limited development costs will be important.

ACKNOWLEDGMENTS

LITERATURE CITED

Bragg, D.C. 2008a. **An improved height measurement technique tested on mature southern pines.** Southern Journal of Applied Forestry. 32(1): 38-43.

Bragg, D.C. 2008b. **Practical extension of a Lake States tree height model.** Northern Journal of Applied Forestry. 25(4): 186-194.

Budhathoki, C.B.; Lynch, T.B.; Guldin, J.M. 2008. **A mixed-effects model for the dbh-height relationship of shortleaf pine (*Pinus echinata* Mill.).** Southern Journal of Applied Forestry. 32(1): 5-11.

Burnham, K.P.; Anderson, D.R. 2002. **Model selection and multimodel inference: a practical information-theoretic approach.** 2nd ed. New York, NY: Springer-Verlag. 488 p.

Dixon, G.E., comp. 2009. **Essential FVS: a user's guide to the forest vegetation simulator.** Internal Rep., revised 2009 June 3. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Forest Management Service Center. 220 p.

Domec, J.-C.; Lachenbruch, B.; Meinzer, F.C.; Woodruff, D.R.; Warren, J.M.; McCulloh, K.A. 2008. **Maximum height in a conifer is associated with conflicting requirements for xylem design.** Proceedings of the National Academy of Sciences. 105(33): 12069-12074.

Forest Vegetation Simulator Staff. 2009. **Forest Vegetation Simulator southern (SN) variant overview.** Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Forest Management Service Center. 63 p.

Huang, S.; Price, D.; Titus, S.J. 2000. **Development of ecoregion-based height-diameter models for white spruce in boreal forests.** Forest Ecology and Management. 129(1-3): 125-141.

Huang, S.; Titus, S.J.; Wiens, D.P. 1992. **Comparison of nonlinear height-diameter functions for major Alberta tree species.** Canadian Journal of Forest Research. 22(9): 1297-1304.

Kershaw, J.A.; Morrissey, R.C.; Jacobs, D.F.; Seifert, J.R.; McCarter, J.B. 2008. **Dominant height-based height-diameter equations for trees in southern Indiana.** In: Jacobs, D.F.; Michler, C.H., eds. Proceedings, 16th central hardwood forest conference; 2008 April 8-9; West Lafayette, IN. Gen. Tech. Rep. NRS-P-24. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station: 341-355.

Koch, G.W.; Sillett, S.C.; Jennings, G.M.; Davis, S.D. 2004. **The limits to tree height.** Nature. 428: 851-854.

Lei, Y.; Parresol, B.R. 2001. **Remarks on height-diameter modeling.** Res. Note SRS-10. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 5 p.

Lootens, J.R.; Larsen, D.R.; Shifley, S.R. 2007. **Height-diameter equations for 12 upland species in the Missouri Ozark highlands.** Northern Journal of Applied Forestry. 24(2): 149-152.

Ryan, M.G.; Yoder, B.J. 1997. **Hydraulic limits to tree height and tree growth.** BioScience. 47(4): 235-242.

Shifley, S.R.; Brand, G.J. 1984. **Chapman-Richards growth function constrained for maximum tree size.** Forest Science. 30(4): 1066-1070.

Trincado, G.; VanderSchaaf, C.L.; Burkhart, H.E. 2007. **Regional mixed-effects height-diameter models for loblolly pine (*Pinus taeda* L.) plantations.** European Journal of Forest Research. 126(2): 253-262.

Zeide, B. 1999. **Pattern of height growth for southern pine species.** Forest Ecology and Management. 118:183-196.

Zhang, L. 1997. **Cross-validation of nonlinear growth functions for modeling tree height-diameter relationships.** Annals of Botany. 79(3): 251-257.

Zhang, L.; Bi, Huiquan; Cheng, P.; Davis, C.J. 2004. **Modeling spatial variation in tree diameter-height relationships.** Forest Ecology and Management. 189(1-3): 317-329.