Incorporating Additional Tree and Environmental Variables in a Lodgepole Pine Stem Profile Model

John C. Byrne, *USDA Forest Service, Intermountain Research Station, Moscow, ID 83843.*

ABSTRACT. *A new variable-form segmented stem profile model is developed for lodgepole pine* (Pinus contorta) *trees from the northern Rocky Mountains of the United States. I improved estimates of stem diameter by predicting two of the model coefficients with linear equations using a measure of tree form, defined as a ratio of dbh and total height. Additional improvements were obtained by fitting this model to individual national forest data sets. Other tree and environmental variables tested but found of little use in improving stem profile estimates were crown ratio, habitat series, elevation, slope percent, and aspect. West. J. Appl. For. 8(3): 86–90.*

Inventory, timber sale preparation, and growth and yield estimation require accurate and compatible methods for estimating individual tree dimensions (volume, stem diameter, and length), especially for an important commercial species like lodgepole pine *(Pinus contorta)*. A complete system of equations for estimation of individual tree dimensions can be obtained from a stem profile function. Stem profile models commonly require only *D* (i.e., diameter outside bark at 4.5 ft above the ground) and total height (H) measurements as inputs. Other tree measurements, including crown ratio (CR) and various measures of form, have been used as predictor variables to improve estimates from a stem profile equation (Valenti and Cao 1986, Newnham 1988). The use of commonly measured environmental variables (i.e., geographic location, slope percent, aspect, elevation, and habitat type), used as indicators of site quality by the Prognosis Model (Wykoff et al. 1982), have not been explored to any great degree. Larson (1963) and Newberry and Burkhart (1986) noted that site quality has some relationship to tree taper and form.

This paper describes the development of a stem profile model for lodgepole pine. Relationships among tree (crown ratio, form) and environmental (geographic location, habitat type, slope percent, aspect, and elevation) variables and lodgepole pine stem profile are explored and are incorporated into the stem profile equation when found to be significant.

Methods

Data

Data came from felled lodgepole pine trees from nine national forests (NF) in Idaho, Montana, and Wyoming (USDA Forest Service 1973, 1984b). Within each national forest, stands within habitat type groupings (grouped by productivity potential) were sampled until lodgepole pine trees with a range of *D*, *H*, and CR were found.

Sample trees were felled and then cut at predetermined points along the stem. At each cut, four measurements were taken: diameter inside bark (dib), to the nearest 0.1 in., of the minor axis of the stem; dib of the major axis (i.e., 90° from the minor axis); average double bark thickness; and the height (to the nearest 0.1 ft) from the ground to the cut. Cuts were made at the stump (from 0.1 to 1.0 ft above ground), at *D*, and then, starting at the stump, at approximately 8.25 ft intervals up to a 5.6 in. dib. The next bucking points were at 5.0 in., 4.0 in., and 2.5 in. dib. Cuts were also made at 1/2 and 3/4 the total height. *D* was measured (with a diameter tape) before, and total height after, the tree was felled. Crown ratio (CR), where branches in the lower crown were ocularly rearranged to achieve a balanced crown, was estimated to the nearest 10%. In addition, national forest, habitat type, aspect (by 45° classes; e.g., NE = 22.5° - 67.5°), and elevation (in feet above mean sea level) at the stand level and slope percent (by 10% classes) at the tree level were recorded. Trees with dead or forked tops were not selected.

A total of 3,624 trees were available. I systematically split the data by putting every third tree into a validation data set for testing, with the remaining data in a developmental data set for fitting the stem profile models. Each data subset had a similar representation of data as the full data set. See Tables 1 and 2 for summaries of the data.

Stem Profile Model

The well-known Max and Burkhart (1976) stem profile model was used as the base model (see Appendix 1). It is a segmented model that joins quadratic equations for three parts of the tree (i.e., the butt, the middle, and the top). The three

Table 1. Summary of full data set for pertinent tree and environmental variables by national forest.

	National forest									
Variable/ Statistic	Beaver-head	Idaho Panhandle	Flathead	Gallatin	Kootenai	Lolo	Nez Perce	Targhee	Challis	All data
DBH (in.)										
Average	7.4	8.9	8.7	8.3	8.1	8.2	10.0	9.5	8.5	8.3
Range	0.9–19.8	0.5–22.3	1.3–17.6	1.7–21.7	0.8–21.0	1.0–18.3	7.0–18.9	5.0–24.1	5.0–14.6	0.5–24.1
Total ht (ft)										
Average	48.1	53.3	57.1	55.4	59.1	60.7	78.3	61.8	51.3	56.7
Range	8.3–98.8	7.2–95.3	11.6–98.7	11.7–104.5	10.4–110.4	8.3–112.1	47.5–107.1	34.5–96.4	32.8–84.8	7.2–112.1
Crown ratio (%)										
Range	10–90	10–90	10–90	10–90	10–90	10–90	20–80	10–90	20–80	10–90
Elev. (ft)										
Range	6,440–8,080	4,400–6,160	3,200–6,340	6,600–8,100	2,700–5,880	3,200–6,560	3,800–6,400	5,200–8,000	6,600–8,700	2,700–8,700
Slope (%)										
Range	0–40	0–60	0–50	0–30	0–40	0–70	0–50	0–30	0–30, 50,70	0–70
Aspect[1]										
Classes represented	All	All except N, NE, NW	All	All	All except level	All	All except level	All except NE	All except SE, S, SW	All
Habitat type										
Series[2] represented	PSME ABLA	ABLA TSME	PSME ABGR THPL ABLA	PSME PIEN ABLA	PSME PIEN ABLA	PSME ABGR THPL ABLA	ABGR ABLA	PSME ABLA PICO	ABLA PICO	All

[1] All = all codes represented (i.e., N, NE, E, SE, S, SW, W, NW, level).
[2] Habitat series as defined in Pfister et al. 1977, Steele et al. 1981, Steele et al. 1983, Cooper et al. 1987.

equations are made to be continuous for the tree bole at two join points. Of the six coefficients in the model, the join point coefficients (a_1 and a_2) signify changes in the stem profile. The a_1 join point occurs near where butt swell ends, and the a_2 join point occurs within the crown. This model is conditioned such that $d = 0$ when $h = H$.

Incorporating Additional Variables in Stem Profile Models

I used the technique described by Burkhart and Walton (1985) and Valenti and Cao (1986) to incorporate additional variables into the stem profile model. They decided on a suitable equation form for predicting a stem profile coefficient from an additional variable and then directly substituted this equation into the profile model and fit it to the entire developmental data set. The following additional variables were used: national forest (as an indicator of geographic location), habitat series, elevation, slope percent, aspect, and crown ratio. I also used D and H. Transformations of the continuous variables, along with interactions between aspect and slope, are utilized because of their importance in predicting diameter and height growth (Stage 1976, Wykoff et al. 1982). SAS software (SAS Institute Inc. 1987) is used for all analyses.

Important additional variables were determined using the approach of Valenti and Cao (1986) for *CR* incorporation. First, for each of the continuous variables (i.e., elevation, D, H, D/H [i.e., $(D/12)/H$], and two aspect-slope combinations), the developmental data were sorted in ascending order and then split into 10 groups of approximately equal numbers of observations. For the other variables, I used logical groups (i.e., national forest, habitat series, 10% crown ratio groups, 10% slope percentage groups, and 45° aspect groups). Then, the model was fit to each of the groups defined for a particular variable. This procedure was conducted separately for both outside bark (ob) and inside

Table 2. Number of trees and height/diameter observations in the data sets by national forest.

	Number of trees			Height/diameter observations[1]		
National forest	Full data set	Developmental data	Validation data	Full data set	Developmental data	Validation data
Beaverhead	890	594	296	8,024	5,343	2,681
Idaho Panhandle	249	166	83	2,535	1,687	848
Flathead	356	238	118	3,603	2,395	1,208
Gallatin	573	382	191	5,727	3,912	1,815
Kootenai	297	198	99	2,905	1,913	992
Lolo	862	575	287	8,429	5,632	2,797
Nez Perce	211	141	70	2,787	1,862	925
Targhee	121	81	40	1,041	696	345
Challis	65	44	21	484	325	159
Totals	3,624	2,419	1,205	35,535	23,765	11,770

[1] Does not include total height observations.

bark (ib) data. Substantial reductions in Sum of Squared Error (SSE), as tested by an F-statistic for a reduced (without additional variable) vs. complete (with additional variable) model (Ott 1977), indicate variables that may be useful in improving the model.

The important continuous variables from the above analysis were used as predictors of the stem profile coefficients. Because of the difficulty of using the important discrete variables from the above analysis in the model, I reserved them for later in the analysis. The model was fit to the full developmental data set with one of the six coefficients being substituted by a simple equation in one of the additional variables. Three equation forms were used (Valenti and Cao 1986):

Linear:	$P_i = f_1 + f_2(X)$	(1)
Logarithmic:	$P_i = f_1 + f_2(\ln X)$	(2)
Hyperbolic:	$P_i = f_1 + f_2(1/X)$	(3)

where

P = a stem profile model parameter,

f_1, f_2 = regression coefficients to be predicted,

X = an additional variable.

For each coefficient in the model (including the join point coefficients), each of these three equation forms with each of the important additional variables was directly substituted into the model and then parameters were estimated using the developmental data. New models, derived by incorporating the best individual equation form-variable combinations for different coefficients, were also fit to the data. For important discrete variables, the model developed with important continuous variables above was fit to each class of the discrete variables. SSE analysis, as previously described, was used to determine the best models that incorporate additional variables.

Comparison of the Stem Profile Models

The best models fit to the developmental data were evaluated further using the validation data and residual analysis. These models were used to predict d for each observation in the validation data set even though d^2 is the dependent variable in each model. This was done because d is a very commonly used variable and therefore errors in d were more easily interpreted. Three test statistics were calculated from the residuals (observed value - predicted value), i.e., average residual ($\bar{R}$), standard deviation of the residuals (s), and percent variation explained, *PVE* (Byrne and Reed 1986). Average residuals and standard deviation of residuals closest to zero and *PVE* closest to one indicated the best prediction. The best system was selected by comparing these statistics for the best models from the developmental data. In addition, these residual statistics were calculated using the developmental data to verify that similar conclusions are reached for both data sets. Also, to determine whether the models fit each part of the bole equally well, test statistics were calculated with the validation data for percentages of total height (10% classes).

To detect any weaknesses in the model, residuals from the validation data were graphed against the important additional variables (even those not used in the model), for the ib data. Residuals from measurements taken at three points on the tree bole were graphed. These three points were midclear bole (midway between the base of the stem and crown base), crown base, and midcrown (midway between the crown base and the tree top). For actual application, I determined final coefficients by fitting the best model to the entire data set (developmental and validation data combined).

Results

SSE analysis indicated that all of the additional variables tested were statistically significant at the 0.01 probability level. The greatest reduction in SSE was obtained with the *D/H* ratio (9% reduction), for both ob and ib data. Variables of lesser value, in order of SSE reduction were *D*, elevation, NF, habitat series, and crown ratio for ob data; and elevation, NF, habitat series, crown ratio, and *D* for ib data. Slope percent, aspect, transformations of slope and aspect, and *H* had even lower SSE reduction so were not used as additional predictors.

The continuous variables, *D/H*, *D*, elevation, and CR, were then used to predict the coefficients of the stem profile model Incorporating *D*, elevation, and CR for ib data, and elevation and CR for ob data, produced only a small reduction in SSE. Greatest reductions in SSE were obtained when both b_1 and b_2 were estimated, with the linear equation form, using *D/H* or *D* as an additional predictor (i.e., *D/H* and *D* models, respectively) for ob data and *D/H* as an additional predictor for ib data.

To determine any additional improvement possible using the best discrete variables, NF and habitat series, the *D/H* and *D* models were fit to each NF (i.e., *D/H*-NF and *D*-NF models, respectively) and habitat series (i.e., *D/H*-HS and *D*-HS models, respectively). All of these new models had reduced SSE and were statistically significant, but using NF produced a greater reduction than habitat series. The *D/H*-NF and *D*-NF models for ob data, and the *D/H*-NF model for ib data appear to be the best new models developed from the developmental data using the available additional variables. The *D/H* and *D* models were retained for further analysis to verify the usefulness of fitting to individual NFs.

Using the validation data, the three test statistics, ($\bar{R}$), s, and *PVE* (see Table 3), indicated that the *D/H*-NF model is clearly the best for both ob and ib data. Test statistics based on the developmental data are very similar for both ob and ib. Not surprisingly, the ib models are less variable than the corresponding ob models, due to the irregularity of bark thickness.

Residuals calculated at 10% height intervals for the tree stem from the developmental data indicated that the *D/H*-NF model was the best model along the entire stem for the ob data. The other models had average residuals farther from 0 and larger standard deviations than the *D/H*-NF model for a majority of percentage height classes. For the ib data, the *D/H*-NF model had average residuals closer to 0 for most classes and lower standard deviations for all classes than the *D/H* model. These results, as well as the F-statistic test for SSE, verify the gain in predictive ability by fitting to individual NFs (*D/H*-NF model) instead of pooling the data. As might be expected, prediction error tends to be higher at the base of a tree due to the variability of the butt swell.

Table 3. Residual[1] statistics for validation data set.

Model	Outside bark data			Inside bark data		
	Average residual	Standard deviation residual	PVE	Average residual	Standard deviation residual	PVE
D/H	0.024	0.570	0.970	-0.015	0.548	0.971
D/H-NF	0.020	0.535	0.974	-0.017	0.516	0.974
D	-0.032	0.586	0.968			
D-NF	-0.036	0.568	0.970			

[1] Residual = observed *d* - predicted *d*, in inches.

Graphs of residuals from the *D/H*-NF model against *D/H* produced a fan-shaped pattern, with residuals increasing as *D/H* increased at all three locations on the tree. These graphs also indicated some variation in the magnitude of the residuals from the top to bottom of the tree. A function with *D/H* and *h/H* as independent variables was developed as a weighting factor to eliminate this pattern of unequal residuals. Reparameterization of the *D/H*-NF model using this weighting factor corrected the patterned residuals, but residual analysis showed no change in the predictions from the model. Therefore, the original *D/H*-NF model coefficients were retained for use.

Residuals were also graphed for the other important variables (i.e., national forest, habitat series, elevation, and crown ratio). No patterns were observed for national forest and elevation. The model slightly overestimated *d* for the Engelmann spruce (*Picea engelmannii*) and mountain hemlock (*Tsuga mertensiana*) habitat series, especially at the crown base and midcrown locations. For crown ratio, the midclear bole and crown base residuals looked good, but for the midcrown residuals, there was a slight overestimation associated with high crown ratio, from 60%-90% CR.

Discussion

The advantage of incorporating *D/H* into the stem profile model has been shown. *D/H* can be considered an indicator of tree form. Trees with the same *D* but different heights will generally have somewhat different forms. The differing tree forms caused by the range of *D* and *H* combinations will be represented by using the *D/H* variable to predict the b_1 and b_2 coefficients in the stem profile model. This technique allows for incorporation of some measure of tree form without requiring the use of an additional stem diameter, which is time-consuming and difficult to measure.

Using separate coefficients for each NF seems to help in describing the differing environmental conditions that contribute to varying stem profiles for lodgepole pine in the Northern Rockies. Different environmental and climatic stresses affect trees on the NFs considered in this study. For example, the Idaho Panhandle NF (in northern Idaho) has a moister, maritime-influenced climate whereas the Gallatin NF (in southwestern Montana) has a drier, continental climate. Each NF is represented by a complex combination of environmental factors, including vegetation patterns, climate, landforms, elevation, etc. Some combination of environmental variables might do a better job in helping to predict varying stem profile, but the available data reflects the usefulness of using NF alone as a surrogate for such a combination of factors.

The patterns of overestimation seen for some habitat series and high crown ratios mainly reflect the limited occurrence of these characteristics in the forest. The Engelmann spruce and mountain hemlock habitat series are only minor components of the land base in this geographic area (Byrne et al. 1988). As for the bias in heavily crowned trees, most trees normally do not maintain such large crowns when growing in forest stands and therefore were not heavily sampled in this study. Such high crown ratios indicate more open-grown trees.

Model Application

Data from the developmental and validation data sets were combined, and the *D/H*-NF model fit to the entire data set. Table 4 for ob data and Table 5 for ib data show the coefficients for each NF, as well as coefficients for the entire data set (all NF data sets

Table 4. Coefficients for the D/H-NF model, outside bark data.

National forest	Coefficient							
	b_3	b_4	a_1	a_2	f_1	f_2	f_3	f_4
Beaverhead	104.397948	-2.183345	0.070168	0.782064	-3.427675	-54.912103	1.168917	52.424091
Idaho Panhandle	134.441969	-1.739870	0.060289	0.786951	-3.577941	-15.133016	1.362265	25.289147
Flathead	180.118890	-2.144045	0.060662	0.807914	-3.328044	-83.304301	1.077652	76.189012
Gallatin	114.054765	-2.156758	0.067748	0.781523	-3.518426	-37.028555	1.182589	42.128033
Kootenai	137.493492	-1.579915	0.064608	0.774465	-2.527736	-65.128102	0.710065	60.071600
Lolo	114.917322	-1.346061	0.069318	0.760924	-2.339391	-54.762665	0.622919	52.394400
Nez Perce	177.267918	-2.526703	0.060030	0.834903	-4.406124	-61.646518	1.617829	66.349579
Targhee	81.804828	-1.975942	0.065700	0.731487	-3.402352	-24.313446	1.293340	27.412999
Challis	43.244194	-1.666913	0.094450	0.772752	-1.588609	-140.653850	-0.260453	128.277526
Full data set	156.425019	-1.766880	0.059515	0.780459	-3.228737	-33.504041	1.157767	35.376538

Table 5. Coefficients for the D/H-NF model, inside bark data.

National forest	Coefficient							
	b_3	b_4	a_1	a_2	f_1	f_2	f_3	f_4
Beaverhead	90.568523	-1.962273	0.069688	0.764338	-2.988761	-53.327745	1.002217	49.374054
Idaho Panhandle	107.913802	-1.578220	0.060584	0.767100	-3.092332	-20.968552	1.129189	28.216212
Flathead	156.963535	-1.988964	0.056600	0.783103	-3.044257	-70.718097	0.982754	67.609824
Gallatin	92.741639	-1.989368	0.070212	0.762345	-3.095119	-40.068917	1.017405	41.960071
Kootenai	127.445303	-1.432999	0.061698	0.759498	-2.341331	-53.634131	0.673612	51.480968
Lolo	99.173326	-1.352238	0.066329	0.752011	-2.230199	-56.580619	0.605297	52.622730
Nez Perce	139.184999	-2.323381	0.061520	0.813900	-3.817534	-68.943455	1.363008	69.380872
Targhee	65.102088	-1.918440	0.068208	0.725536	-3.223871	-26.078065	1.248376	27.408256
Challis	35.329978	-1.568828	0.097382	0.749788	-1.289302	-140.725558	-0.285214	122.967481
Full data set	123.460573	-1.650902	0.060639	0.761115	-2.862620	-36.847296	1.003935	36.499678

combined). To use the model, first estimate the b_1 and b_2 coefficients of the stem profile model using the following equations:

$$b_1 = f_1 + f_2\,[(D/12)/H] \quad (4)$$

$$b_2 = f_3 + f_4\,[(D/12)/H] \quad (5)$$

using the coefficients for the nearest NF from Tables 4 or 5. Substitute the calculated b_1 and b_2 coefficients along with b_3, b_4, a_1, and a_2 into the appropriate equation and solve. See Byrne and Reed (1986) for compatible equations to calculate total cubic-foot volume, merchantable height, and merchantable cubic-foot volumes to either a height or diameter limit. Calculate board-foot or scaled cubic-foot volumes by applying an appropriate board-foot or cubic-foot rule to the dimensions of logs predicted using the stem profile equation. Use discretion when applying the coefficients outside the range of data considered in this study.

Literature Cited

BURKHART, H.E., and S.B. WALTON. 1985. Incorporating crown ratio into taper equations for loblolly pine trees. For. Sci. 31:478–484.

BYRNE, J.C., and D.D. REED. 1986. Complex compatible taper and volume estimation systems for red and loblolly pine. For. Sci. 32:423–443.

BYRNE, J.C., A.R. STAGE, and D.L. RENNER. 1988. Distribution of permanent plots to evaluate silvicultural treatments in the Inland Empire. USDA For. Serv. Res. Note INT-386. 7 p.

COOPER, S.V., K.E. NEIMAN, R. STEELE, and D.W. ROBERTS. 1987. Forest habitat types of northern Idaho: A second approximation. USDA For. Serv. Gen. Tech. Rep. INT-236. 135 p.

LARSON, P. R. 1963. Stem form development of forest trees. For. Sci. Monogr. 5. 42 p.

MAX, T.A., and H.E. BURKHART. 1976. Segmented polynomial regression applied to taper equations. For. Sci. 22:283–289.

NEWBERRY, J.D., and H.E. BURKHART. 1986. Variable-form stem profile models for loblolly pine. Can. J. For. Res. 16:109–114.

NEWNHAM, R.M. 1988. A variable-form taper function. Petawawa Nat. Forestry Inst. Infor. Rep. PI-X-83. Forestry Canada, Chalk River, Ontario. 33 p.

OTT, L. 1977. An introduction to statistical methods and data analysis. Duxbury Press, Belmont, CA. 730 p.

PFISTER, R.D., B.L. KOVALCHIK, S.F. ARNO, and R.C. PRESBY. 1977. Forest habitat types of Montana. USDA For. Serv. Gen. Tech. Rep. INT-34. 174 p.

SAS INSTITUTE, INC. 1987. SAS/STAT Guide for Personal Computers, Version 6 Edition. Cary, NC: SAS Institute Inc. 1028 p.

STAGE, A.R. 1976. An expression of the effect of aspect, slope, and habitat type on tree growth. For. Sci. 22(3):457-460.

STEELE, R., R.D. PFISTER, R.A. RYKER, and J.A. KITTAMS. 1981. Forest habitat types of central Idaho. USDA For. Serv. Gen. Tech. Rep. INT-114. 138 p.

STEELE, R., S.V. COOPER, D.M. ONDOV, D.W. ROBERTS, and R.D. PFISTER. 1983. Forest habitat types of eastern Idaho–western Wyoming. USDA For. Serv. Gen. Tech. Rep. INT-244. 122 p.

USDA FOREST SERVICE. 1973. Net volume field procedures. Div. of Timber Manage., Intermountain Region, USDA For. Serv. 37 p.

USDA FOREST SERVICE. 1984. Timber Management Staff interoffice memo on lodgepole pine tree volume equations. File 2440, May 4, 1980. Region 1, Missoula, MT.

VALENTI, M.A., and Q. V. CAO. 1986. Use of crown ratio to improve loblolly pine taper equations. Can. J. For. Res. 16:1141–1145.

WYKOFF, W. R., N. L. CROOKSTON, and A. R. STAGE. 1982. User's guide to the stand prognosis model. USDA For. Serv. Gen. Tech. Rep. INT-133. 112 p.

APPENDIX 1

The Max and Burkhart (1976) stem profile model:

$$d^2 = D^2[b_1(\frac{h}{H} - 1) + b_2((\frac{h}{H})^2 - 1) + b_3(a_1 - \frac{h}{H})^2 I_1 + b_4(a_2 - \frac{h}{H})^2 I_2] \quad (6)$$

where

$a_1, a_2, b_1, b_2, b_3, b_4$ = regression coefficients

d = stem diameter (in inches) at height h

h = height above the ground (in feet) to d

D = diameter (outside bark) at 4.5 ft above ground, in inches

H = total tree height, in feet

$$I_i = \begin{cases} 1, & \text{if } \frac{h}{H} \le a_i \\ 0, & \text{if } \frac{h}{H} > a_i; i = 1,2 \end{cases}$$