FITTING HEIGHT-DIAMETER
CURVES

by

Stephen A.Y. Omule

RR84008-HQ

RESEARCH REPORT

Internal Reports of the Ministry of Forests Research Program
Fitting Height-Diameter Curves

by

Stephen A.Y. Omule
Research Silviculturist
Growth and Yield Studies
Research Branch

May, 1984

Revised July, 1984

Province of
British Columbia
Ministry of
Forests
This publication is RR84008-HQ

Copies of this report may be obtained, depending on supply, from:

Research Branch
B.C. Ministry of Forests
1450 Government Street
Victoria, B.C.
V8W 3E7

Note:

This report may not be cited without the prior written approval of the Director of Research, B.C. Ministry of Forests.

Citation:

ABSTRACT

The method of piecewise linear regression and the use of the $S_{BB}$ distribution in fitting height-diameter curves are described. The $S_{BB}$ distribution is particularly recommended because the fitting procedure is simple and the resulting equations are usually very precise and have little bias. As well, the equation allows for the generation of height-diameter frequencies, and may produce curves that are consistent over time. A Statistical Analysis System (SAS) program has been developed within the Research Branch to fit the $S_{BB}$ distribution and is available for use by all branch personnel.
TABLE OF CONTENTS

The Problem ........................................................................................................ 1
Fitting Procedure ............................................................................................... 1
    Piecewise Linear Regression ........................................................................ 3
    $S_{BB}$ Distribution ................................................................................. 5
Discussion .......................................................................................................... 13
Conclusion ......................................................................................................... 18
References ......................................................................................................... 19

TABLES

1 Height-diameter sample data set (EP 418, Plot 1, Meas 6, Age 29) ......................... 2
2 Analysis of variance table for piecewise regression (equation 6) ............................. 6
3 Analysis of variance table for a single regression (equation 1) .............................. 9
4 Coefficients for $S_{BB}$ height-diameter curves .............................................. 16

FIGURES

1A Piecewise linear regression (EP 418, Plot 1, Meas 6, Age 29) ......................... 7
1B Piecewise linear regression residuals .......................................................... 8
2A Single linear regression (EP 418, Plot 1, Meas 6, Age 29) ............................. 10
2B Single regression residuals ........................................................................... 11
3A $S_{BB}$ regression equation (EP 418, Plot 1, Meas 6, Age 29) ...................... 14
3B $S_{BB}$ regression residuals ........................................................................... 15
4 $S_{BB}$ regression over time EP 418, Plot 1 .................................................... 17
THE PROBLEM

A height-diameter curve is a function that assigns a height to any given tree diameter. The form of the curve usually varies with tree age, species, site, and management. The curves are particularly useful in estimating tree volumes, done when a standard tree volume table is entered with a measured diameter at breast height (dbh) and an estimated (or, sometimes measured) total height. Estimates of tree volume, however, can be biased and/or imprecise if unrepresentative height-diameter curves are used to estimate tree heights.

Given a sample of height and diameter measurements, the problem is to fit the "best" equation (or system of equations) to the data such that the height estimates obtained are consistent and logical for the range of diameters and, for remeasurement data, over time. After working on this problem with several remeasurement data sets of height and diameter, I have found the use of piecewise linear regression (Cunia 1974) and the $S_{BB}$ distribution (Schreuder and Hafley 1977) to be very successful. Fitting height-diameter curves by piecewise linear regression and the $S_{BB}$ distribution are described in this report, together with an example that uses forest service data.

FITTING PROCEDURES

To a sample of diameter ($x$) and height ($y$) values (shown in Table 1) we wish to fit an equation (or system of equations) for predicting unmeasured tree heights. One approach is to fit a series of linear regression models, restricting each one to a certain interval of $x$-values: a piecewise linear regression. Another approach is to fit a single suitable model to the entire data set: an $S_{BB}$ distribution. Each method is described in the following sections.

Piecewise Linear Regression

The method described uses dummy variables and assumes that the $x$-values of the intersection points of the linear regression models are known. Only the case of fitting a set of two regressions is considered. Details of the method can be found in Cunia (1974).
TABLE 1. Height-diameter sample data set (EP 418, Plot 1, Meas 6, Age 29)

<table>
<thead>
<tr>
<th>Tree no.</th>
<th>Diameter (cm)</th>
<th>Height (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>211</td>
<td>8.1</td>
<td>8.8</td>
</tr>
<tr>
<td>221</td>
<td>1.8</td>
<td>2.1</td>
</tr>
<tr>
<td>229</td>
<td>5.1</td>
<td>4.9</td>
</tr>
<tr>
<td>230</td>
<td>23.9</td>
<td>17.4</td>
</tr>
<tr>
<td>240</td>
<td>18.0</td>
<td>13.7</td>
</tr>
<tr>
<td>245</td>
<td>6.1</td>
<td>7.3</td>
</tr>
<tr>
<td>251</td>
<td>19.3</td>
<td>16.2</td>
</tr>
<tr>
<td>255</td>
<td>8.1</td>
<td>9.1</td>
</tr>
<tr>
<td>260</td>
<td>12.4</td>
<td>11.9</td>
</tr>
<tr>
<td>264</td>
<td>5.3</td>
<td>5.5</td>
</tr>
<tr>
<td>271</td>
<td>9.1</td>
<td>9.4</td>
</tr>
<tr>
<td>278</td>
<td>14.0</td>
<td>13.7</td>
</tr>
<tr>
<td>280</td>
<td>8.6</td>
<td>9.4</td>
</tr>
<tr>
<td>287</td>
<td>18.3</td>
<td>12.8</td>
</tr>
<tr>
<td>290</td>
<td>4.1</td>
<td>4.9</td>
</tr>
<tr>
<td>298</td>
<td>3.6</td>
<td>4.9</td>
</tr>
<tr>
<td>300</td>
<td>10.7</td>
<td>10.4</td>
</tr>
<tr>
<td>305</td>
<td>9.9</td>
<td>10.7</td>
</tr>
<tr>
<td>312</td>
<td>4.6</td>
<td>4.9</td>
</tr>
<tr>
<td>314</td>
<td>2.5</td>
<td>3.0</td>
</tr>
<tr>
<td>315</td>
<td>27.7</td>
<td>18.6</td>
</tr>
<tr>
<td>320</td>
<td>18.5</td>
<td>16.2</td>
</tr>
<tr>
<td>329</td>
<td>21.6</td>
<td>17.4</td>
</tr>
<tr>
<td>340</td>
<td>19.8</td>
<td>16.8</td>
</tr>
<tr>
<td>342</td>
<td>16.8</td>
<td>16.2</td>
</tr>
<tr>
<td>346</td>
<td>3.6</td>
<td>3.7</td>
</tr>
<tr>
<td>347</td>
<td>4.1</td>
<td>4.9</td>
</tr>
<tr>
<td>348</td>
<td>4.1</td>
<td>4.6</td>
</tr>
<tr>
<td>350</td>
<td>11.7</td>
<td>12.2</td>
</tr>
<tr>
<td>356</td>
<td>19.0</td>
<td>16.8</td>
</tr>
<tr>
<td>360</td>
<td>12.2</td>
<td>12.5</td>
</tr>
<tr>
<td>363</td>
<td>4.3</td>
<td>4.9</td>
</tr>
<tr>
<td>364</td>
<td>5.1</td>
<td>5.2</td>
</tr>
<tr>
<td>365</td>
<td>15.0</td>
<td>13.7</td>
</tr>
<tr>
<td>367</td>
<td>3.0</td>
<td>3.7</td>
</tr>
<tr>
<td>370</td>
<td>25.1</td>
<td>19.2</td>
</tr>
<tr>
<td>380</td>
<td>8.6</td>
<td>9.8</td>
</tr>
<tr>
<td>389</td>
<td>18.0</td>
<td>15.8</td>
</tr>
<tr>
<td>390</td>
<td>11.2</td>
<td>11.0</td>
</tr>
<tr>
<td>391</td>
<td>2.8</td>
<td>3.7</td>
</tr>
<tr>
<td>398</td>
<td>16.8</td>
<td>15.5</td>
</tr>
<tr>
<td>399</td>
<td>11.4</td>
<td>12.2</td>
</tr>
<tr>
<td>420</td>
<td>27.2</td>
<td>18.6</td>
</tr>
<tr>
<td>421</td>
<td>5.1</td>
<td>5.2</td>
</tr>
<tr>
<td>430</td>
<td>13.5</td>
<td>14.9</td>
</tr>
<tr>
<td>438</td>
<td>2.5</td>
<td>3.4</td>
</tr>
<tr>
<td>440</td>
<td>9.7</td>
<td>10.4</td>
</tr>
<tr>
<td>446</td>
<td>2.0</td>
<td>2.7</td>
</tr>
<tr>
<td>458</td>
<td>17.0</td>
<td>14.6</td>
</tr>
<tr>
<td>461</td>
<td>15.2</td>
<td>13.4</td>
</tr>
<tr>
<td>469</td>
<td>17.5</td>
<td>16.2</td>
</tr>
<tr>
<td>470</td>
<td>9.7</td>
<td>9.4</td>
</tr>
</tbody>
</table>
A widely used model for fitting height-diameter data is:

\[
\ln H = a + b/D
\]  

where \( H \) = total height and \( D \) = diameter at breast height (Curtis 1967).

One problem with this model (1), however, is that it may not be continuously unbiased, tending to over- and underestimate heights at the smaller and larger diameters, respectively. To correct this, the model can be fitted separately to smaller and larger diameter trees, giving a piecewise linear function of the form:

\[
y = a_1 + b_1x \quad \text{for } x \leq x_0
\]

\[
= a_2 + b_2x \quad \text{for } x \geq x_0
\]

where \( y = \ln H, x = 1/D, \) and \( x_0 \) is the intersection point of the two linear functions. Since a continuous regression function is required, i.e. the function must be single-valued at \( x = x_0 \), then the following condition must be met:

\[
a_1 + b_1x_0 = a_2 + b_2x_0
\]

To accomplish this two dummy variables must be defined as:

\[
x_{11} = 1 \text{ if } x \leq x_0, \quad 0 \text{ otherwise}
\]

and \( x_{21} = 1 \text{ if } x > x_0, \quad 0 \text{ otherwise.} \)

The piecewise function (2) can then be written as:

\[
y = a_1x_{11} + b_1x_{12} + a_2x_{21} + b_2x_{22}
\]

where

\[
x_{12} = x_{11}x, \text{ and } x_{22} = x_{21}x
\]
Since the function is to be continuous at $x_0$, then (setting $x_{11} = x_{21} = 1$ and $x_{12} = x_{22} = x_0$):

$$a_1 + b_1 x_0 = a_2 + b_2 x_0$$

or

$$a_2 = a_1 + x_0 (b_1 - b_2)$$

(4)

Plugging (4) into (3) and rearranging the terms yields the model:

$$y = a_1 (x_{11} + x_{21}) + b_1 (x_{12} + x_{21} x_0) + b_2 (x_{22} - x_{21} x_0)$$

(5)

which can then be fitted in the usual manner, provided $x_0$ is known. If we are fitting model (1), for instance, and assuming $x_0 = 10$, then the piecewise linear function to be fitted is:

$$y = a'_1 x_1 + b'_1 x_2 + b'_2 x_3$$

(6)

where

$$y = \ln H$$

$$x_1 = x_{11} + x_{21} = 1 \text{ (by definition of } x_{11} \text{ and } x_{21})$$

$$x_2 = x_{11}/D + x_{21}/10$$

$$x_3 = x_{21}/D - x_{21}/10$$

and $a'_1$, $b'_1$, and $b'_2$ are the usual regression coefficients.

The sample data in Table 1 are used to illustrate the fitting procedure. With model (6) used in the Statistical Analysis System (SAS), the $D = \text{diameter and } H = \text{height data are read and the transformation } y = \ln H, x_1 = 1, x_2 = x_{11}/D + x_{21}/10,$ and $x_3 = x_{21}/D - x_{21}/10$ is done. Then, with either the GLM or REG procedure, a multiple linear regression analysis with three independent variables and no intercept is performed. Table 2 presents the
analysis of variance table; Figure 1A, the plots of the regression equation; and Figure 1B, the residuals. For purposes of comparison, a single regression of the form (1) was fitted to the same data set and the results are shown in Table 3 and Figure 2 (A and B).

Figures 1 and 2 indicate that for this data set there is considerable reduction in the estimation error and bias when model (6) is used instead of (1). It is worth noting, however, that model (6) tends to underestimate at the intersection point \( x_0 = D = 10 \text{ cm} \), mainly because of the "kinky" nature of the joint. More complex smoothing techniques, such as the use of spline functions (of which piecewise regression is a special case), could be applied to improve the regression function.

Another problem with the piecewise method is that, for remeasurement data, the height-diameter curves tend to cross each other at different ages; that is, the curves often do not form a consistent and logical progression over time. Use of age as another independent variable could be considered.

The \( S_{BB} \) distribution method described below seems to circumvent these problems.

\( S_{BB} \) Distribution

The \( S_{BB} \) distribution (Johnson 1949b) is a bivariate extension of the marginal \( S_B \) distribution (Johnson 1949a). Schreuder and Hafley (1977) introduced the \( S_{BB} \) distribution to height-diameter curve fitting, defining the two variates as \( y_1 = (D - \varepsilon_1)/\lambda_1 \) and \( y_2 = (H - \varepsilon_2)/\lambda_2 \), where \( \varepsilon_1 \) and \( \varepsilon_2 \) are the smallest values and \( \lambda_1 \) and \( \lambda_2 \) are the range of diameter (D) and height (H), respectively, in the population. They then expressed the median regression of H on D as:

\[
H = \lambda_2 \theta \left[ \left( \frac{\varepsilon_1 + \lambda_1 - D}{D - \varepsilon_1} \right)^\phi + 2 \right]^{-1} + \varepsilon_2
\]

(7)

where

\[
\theta = \exp \left\{ (\rho y_1 - y_2)/\delta_2 \right\}
\]

and \( \phi = \rho \delta_1/\delta_2 \) (\( \phi > 0 \)).
TABLE 2. Analysis of variance table for piecewise regression (equation 6)

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>Sums of squares</th>
<th>mean square</th>
<th>F-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3</td>
<td>232.795</td>
<td>77.5982</td>
<td>4423.974</td>
</tr>
<tr>
<td>Error</td>
<td>49</td>
<td>0.859</td>
<td>0.0175</td>
<td></td>
</tr>
<tr>
<td>Uncorr. Total</td>
<td>52</td>
<td>233.654</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Estimated model: $\ln \hat{H} = 2.607556x_1 - 5.249851x_2 - 12.816299x_3$

where $H = \text{total height (m)}$ and $x_i$ are as defined in text.
FIGURE 1A. Piecewise linear regression E.P. 418, Plot 1, Meas 6 (Age 29).
FIGURE 1B. Piecewise linear regression residuals.
TABLE 3. Analysis of variance table for a single regression (equation 1)

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>Sums of squares</th>
<th>mean square</th>
<th>F-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>29.755462</td>
<td>29.755462</td>
<td>867.891</td>
</tr>
<tr>
<td>Error</td>
<td>50</td>
<td>1.714239</td>
<td>0.034285</td>
<td></td>
</tr>
<tr>
<td>Uncorr. Total</td>
<td>51</td>
<td>31.469701</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Estimated model: \( \ln \bar{H} = 2.884938 - 6.1442/D \)

where \( \bar{H} \) = total height (m) and \( D \) = diameter at breast height (cm)
FIGURE 2A. Single linear regression E.P. 418, Plot 1, Meas 6 (Age 29).
Parameters of the distribution are $\delta_i$ and $\gamma_i$ ($i = 1, 2$). Note that equation (7) has eight parameters to estimate, namely: $\epsilon_1$, $\epsilon_2$, $\lambda_1$, $\lambda_2$, $\gamma_1$, $\gamma_2$, $\delta_1$, $\delta_2$. However, for the height-diameter relationship, Schreuder and Hafley (1977) suggest that it is simplest to specify the values of $\epsilon_1$, and $\epsilon_2$ as 0 cm and 1.3 m (breast height), respectively; and to set values of $\lambda_1$ and $\lambda_2$, which must be larger than the largest respective observations in the data set. This leaves four parameters to estimate: $\gamma_1$, $\gamma_2$, $\delta_1$, and $\delta_2$. Schreuder and Hafley gave the following maximum likelihood formulae for estimating these parameters:

\begin{align*}
\hat{\gamma}_i &= \frac{-\bar{f}_i}{S_i} \quad (8) \\
\text{and} \quad \hat{\delta}_i &= \frac{1}{S_i} \quad (9)
\end{align*}

where

\begin{align*}
\bar{f}_i &= \frac{\sum_{j=1}^{n} f_{ij}}{n} \\
S_i^2 &= \frac{1}{n} \sum_{j=1}^{n} \left( f_{ij} - \bar{f}_i \right)^2 \\
\end{align*}

\begin{align*}
f_{ij} &= \ln \left( \frac{y_{ij}}{(1-y_{ij})} \right) \\
y_{1j} &= \frac{(D_j - \epsilon_1)}{\lambda_1} \\
y_{2j} &= \frac{(H_j - \epsilon_2)}{\lambda_2} \\
j &= 1, 2, \ldots, n; \quad i = 1, 2 \\
\text{and} \quad n = \text{number of height-diameter pairs in the data.}
\end{align*}

The correlation between $H$ and $D$ is then obtained as:

\begin{align*}
\hat{\rho} &= \frac{\sum_{j=1}^{n} z_{1j} z_{2j}}{n}
\end{align*}
where

\[ z_{ij} = \hat{\gamma}_1 + \hat{\delta}_1 \ln \left\{ y_{1j} / (1 - y_{1j}) \right\}, \quad i = 1, 2; \quad j = 1, 2, \ldots, n. \]

Thus, fitting model (7) involves: 1) specifying values for \( \varepsilon_1, \varepsilon_2, \lambda_1, \) and \( \lambda_2 \) and 2) estimating the parameters \( \gamma_1, \gamma_2, \delta_1, \) and \( \delta_2 \) using equations (8) and (9). The regression so obtained begins at (0 cm, 1.3 m) and terminates at (0 + \( \lambda_1 \) cm, 1.3 + \( \lambda_2 \) m).

To obtain percentile limits for the median regression the reader is referred to Schreuder and Hafley (1977).

Equation (7) was fitted to the sample data set in Table 1 with the use of SAS. The resulting curves and parameter estimates are given in Figure 3 (A and B) and Table 4, respectively. The SAS program which was developed is available for use. Also given in Table 4 are the sums of simple, absolute, and squared deviations from the regression and similar results for the same trees at other ages. We notice from inspection of the residual plots that this regression is more precise than either (6) or (1), for this data set. It was also demonstrated (Figure 4) that the equation (7) can produce a consistent and logical progression of height-diameter curves over time.

**DISCUSSION**

We have considered two approaches to fitting height-diameter curves: the piecewise linear regression and the \( S_{BB} \) distribution, both of which were found to be more desirable than single curve fitting of the forms suggested by Curtis (1967). However, fitting a piecewise linear function requires more data than fitting a single function, because data are needed to fit each individual segment (Winkler and Hayes 1975). If there is little data, this procedure degenerates to a simple "connect-the-dots" exercise, which, although it might provide the "best fit" in a least-squares sense for the small set of available data, is of minimal value in making predictions. Fitting the \( S_{BB} \) distribution by the method of maximum likelihood also requires a large number of observations. For a suitable data set, however, it performs well for
FIGURE 3A. $S_{BB}$ regression equation E.P. 418, Plot 1, Meas 6 (Age 29).
FIGURE 3B. $S_{BB}$ regression residuals.
TABLE 4. Coefficients for $S_{BB}$ height-diameter curves

Equation: $H = (T*L2/(((E1+L1-D)/(D-E1))**P)+T)) + E2$

Where $H =$ Total height in m; $D=$dbh in cm

$T, P, L1, L2, E1, \text{and } E2$ are equation coefficients

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14</td>
<td>29</td>
<td>0.05</td>
<td>1.131</td>
<td>0.0</td>
<td>1.37</td>
<td>17.75</td>
<td>8.5</td>
<td>.9659 0.266 9.204 5.293</td>
</tr>
<tr>
<td>1</td>
<td>17</td>
<td>68</td>
<td>0.03</td>
<td>1.120</td>
<td>0.0</td>
<td>1.37</td>
<td>20.05</td>
<td>11.6</td>
<td>.9487 2.485 26.433 23.823</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>74</td>
<td>0.05</td>
<td>1.151</td>
<td>0.0</td>
<td>1.37</td>
<td>26.71</td>
<td>13.4</td>
<td>.9673 2.940 30.061 25.521</td>
</tr>
<tr>
<td>1</td>
<td>22</td>
<td>80</td>
<td>0.11</td>
<td>1.202</td>
<td>0.0</td>
<td>1.37</td>
<td>29.51</td>
<td>15.2</td>
<td>.9800 3.133 36.236 30.334</td>
</tr>
<tr>
<td>1</td>
<td>26</td>
<td>68</td>
<td>0.15</td>
<td>1.236</td>
<td>0.0</td>
<td>1.37</td>
<td>35.32</td>
<td>18.0</td>
<td>.9742 5.496 44.555 59.034</td>
</tr>
<tr>
<td>1</td>
<td>29</td>
<td>52</td>
<td>0.20</td>
<td>1.334</td>
<td>0.0</td>
<td>1.37</td>
<td>39.22</td>
<td>19.2</td>
<td>.9876 1.279 27.279 27.893</td>
</tr>
<tr>
<td>1</td>
<td>34</td>
<td>50</td>
<td>0.19</td>
<td>1.431</td>
<td>0.0</td>
<td>1.37</td>
<td>46.00</td>
<td>22.9</td>
<td>.9894 2.950 31.554 30.908</td>
</tr>
<tr>
<td>1</td>
<td>37</td>
<td>50</td>
<td>0.21</td>
<td>1.314</td>
<td>0.0</td>
<td>1.37</td>
<td>47.71</td>
<td>26.2</td>
<td>.9822 2.405 44.101 63.973</td>
</tr>
<tr>
<td>1</td>
<td>41</td>
<td>43</td>
<td>0.34</td>
<td>1.370</td>
<td>0.0</td>
<td>1.37</td>
<td>52.62</td>
<td>27.1</td>
<td>.9875 0.435 33.533 52.619</td>
</tr>
</tbody>
</table>

$T = \theta, P = \emptyset, E1 = \varepsilon_1, E2 = \varepsilon_2, L1 = \lambda_1, L2 = \lambda_2$
FIGURE 4. $S_{BB}$ regression over time E.P. 418, Plot 1.
fitting height-diameter curves, and has a further advantage in that it allows the generation of bivariate frequencies for diameter and height. One problem in fitting the $S_{BB}$ is the specification of the parameters $\lambda_1$ and $\lambda_2$; the ranges of diameter and height, respectively. The method of iteratively searching for $\lambda_1$ (Schreuder and Hafley 1977) may be used, although the quality of fit is apparently unaffected by the choice of $\lambda_1$ so long as they are consistent with the data to be fitted.

Proper choice of model (i.e. an accurate model and one that reflects the underlying process being modelled) may improve the fit so that the use of piecewise regression becomes unnecessary. In particular, non-linear models, such as the modified Weibull (Kozak and Yang 1978), have been suggested to model height-diameter relationships. However, in the iterative techniques used in estimating parameters of such models, one is not always assured of a solution in a reasonable number of iterations. In addition, there is the problem of estimating initial values of the parameters for iteration.

CONCLUSION

The $S_{BB}$ distribution offers an excellent method for fitting height-diameter curves. The fitting procedure is simple, its resulting equations are usually very precise and have little bias, it allows for the generation of height-diameter frequencies, and it is fairly consistent over time. The piecewise regression can also be used, particularly when there is reason to believe that separate equations for certain ranges of diameter may be more appropriate than one equation.

REFERENCES


