INTRODUCTION

The Weibull function was introduced by Bailey and Dell (1973) to model diameter distributions in forest stands. The advantages of the Weibull include its flexibility to fit shapes commonly found in both uneven-aged and even-aged stands and the ease of computing probabilities (or proportions of trees in diameter classes) without the need for numerical integration.

Cao (2004) developed a new approach, termed CDF Regression, that outperformed the existing methods in estimating Weibull parameters of a stand from known density and dominant height. The objective of this study was to evaluate four methods to predict the parameters of Weibull functions that modeled diameter distributions of a future stand, whose attributes need to be predicted from current stand attributes.

DATA

Data were from the Southwide Seed Source Study, which involved 15 loblolly pine (Pinus taeda L.) seed sources planted at 13 locations across 10 Southern States (Wells and Wakeley 1966). Seedlings were planted at a 6 foot x 6 foot spacing. Each 0.04-acre plot consisted of 49 trees measured 4 times at ages 10, 15 or 16, 20 or 22, and 25 or 27. A subset (100 plots) of the original data was randomly selected as the fit data set, to be used for fitting the models. Furthermore, only one growing period was randomly chosen from each plot. The fit data set therefore contained 100 growth periods. Another 100 plots were randomly selected from the remaining original data to form a validation data set. All 3 growing periods from these plots were used to evaluate the methods, resulting in a total of 300 observations in the evaluation data set.

METHODS

The Weibull probability density function (pdf), used in this study to characterize diameter distribution, has the following form:

\[ f(x) = \frac{c}{b} \left( \frac{x - a}{b} \right)^{c-1} \exp \left[ - \left( \frac{x - a}{b} \right)^{c} \right] \]  

(1)

Data were from the Southwide Seed Source Study, which involved 15 loblolly pine (Pinus taeda L.) seed sources planted at 13 locations across 10 Southern States (Wells and Wakeley 1966). Seedlings were planted at a 6 foot x 6 foot spacing. Each 0.04-acre plot consisted of 49 trees measured 4 times at ages 10, 15 or 16, 20 or 22, and 25 or 27. A subset (100 plots) of the original data was randomly selected as the fit data set, to be used for fitting the models. Furthermore, only one growing period was randomly chosen from each plot. The fit data set therefore contained 100 growth periods. Another 100 plots were randomly selected from the remaining original data to form a validation data set. All 3 growing periods from these plots were used to evaluate the methods, resulting in a total of 300 observations in the evaluation data set. Table 1 shows summary statistics for stand attributes at the end of each growth period for the fit and validation datasets.

Table 1—Means (and standard deviations) of stand attributes at the end of each growth period for the fit and validation data

<table>
<thead>
<tr>
<th>Age</th>
<th>n</th>
<th>Dominant height</th>
<th>Trees per acre</th>
<th>Basal area sq.ft./acre</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>30</td>
<td>42 (4)</td>
<td>778 (212)</td>
<td>146 (37)</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>55 (3)</td>
<td>537 (127)</td>
<td>138 (18)</td>
</tr>
<tr>
<td>20</td>
<td>33</td>
<td>58 (7)</td>
<td>484 (117)</td>
<td>148 (31)</td>
</tr>
<tr>
<td>25</td>
<td>20</td>
<td>61 (6)</td>
<td>436 (171)</td>
<td>139 (47)</td>
</tr>
<tr>
<td>27</td>
<td>13</td>
<td>63 (10)</td>
<td>327 (90)</td>
<td>129 (38)</td>
</tr>
</tbody>
</table>

- - - - - - - - - - - - - Validation data - - - - - - - - - - - - -

<table>
<thead>
<tr>
<th>Age</th>
<th>n</th>
<th>Dominant height</th>
<th>Trees per acre</th>
<th>Basal area sq.ft./acre</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>86</td>
<td>43 (6)</td>
<td>702 (187)</td>
<td>134 (27)</td>
</tr>
<tr>
<td>16</td>
<td>14</td>
<td>53 (4)</td>
<td>515 (79)</td>
<td>130 (14)</td>
</tr>
<tr>
<td>20</td>
<td>95</td>
<td>56 (7)</td>
<td>526 (131)</td>
<td>153 (34)</td>
</tr>
<tr>
<td>22</td>
<td>5</td>
<td>46 (7)</td>
<td>385 (48)</td>
<td>83 (36)</td>
</tr>
<tr>
<td>25</td>
<td>72</td>
<td>65 (8)</td>
<td>489 (133)</td>
<td>176 (41)</td>
</tr>
<tr>
<td>27</td>
<td>28</td>
<td>67 (10)</td>
<td>322 (81)</td>
<td>141 (30)</td>
</tr>
</tbody>
</table>

where \( a, b, \) and \( c \) are the location, scale, and shape parameters, respectively, and \( x \) is tree diameter at breast height (d.b.h.).

The annual growth model was employed in this study to project stand attributes, because it was found by Ochi and Cao (2003) to provide better stand predictions than compatible growth models. The stand attributes at time \( t+1 \) were projected from time \( t \) as follows:

\[ \hat{N}_{t+1} = \hat{N}_{t}^{b_{1}} / \hat{A}_{t+1} \]  

(2)

\[ \hat{B}_{t+1} = \hat{B}_{t}^{b_{2}} / \hat{A}_{t+1} \]  

(3)

\[ \hat{D}_{t+1} = \hat{D}_{t}^{b_{3}} / \hat{A}_{t+1} \]  

(4)

\[ \hat{S}_{t+1} = \hat{S}_{t}^{b_{4}} / \hat{A}_{t+1} \]  

(5)
where $A =$ stand age in years, $\hat{N} =$ the predicted dominant height (average height of the dominants and codominants), $\hat{N}_2 =$ predicted number of trees per acre, $\hat{D} =$ the predicted stand basal area in square feet per acre, $\hat{D}_{\text{min}} =$ the minimum diameter in inches, $SD =$ the predicted standard deviation of diameter, and the subscripts place the attributes at time $t$ or $(t+1)$. In addition, the b’s are regression coefficients.

The future 93rd diameter percentile ($\hat{D}_{93}$) at the end of the growth period (age $A_t$) was predicted from the future quadratic mean diameter ($\hat{D}_{Q2}$) and relative spacing ($RS_2$):

$$\hat{D}_{93} = \hat{D}_{Q2}[1 + \exp(b_1 + b_2RS_2)] \tag{6}$$

where $\hat{D}_{Q2} = \sqrt{\hat{D}_2 / 0.005454}, RS_2 = \sqrt{43560 / \hat{N}_2 / \hat{H}_2}$, and $\hat{H}_2, \hat{N}_2$, and $\hat{D}_2$ are dominant height, number of trees, and basal area per acre at age $A_t$, respectively.

Borders (1989) method was used to simultaneously estimate regression parameters of equations to predict $N_2, B_2, D_{\text{min}}$, $SD_2$, and $D_{93}$. The following methods to obtain the Weibull parameters were evaluated.

**Method 1—Moment Estimation**

The Weibull location parameter ($a$) must be smaller than the predicted minimum diameter in the stand ($D_{\text{min}}$). We set $a = 0.5\hat{D}_{\text{min}}$, since Frazier (1981) found that this gave best results in terms of goodness-of-fit. The other Weibull parameters, $b$ and $c$, were recovered from the first two moments of the diameter distribution (Cao and others 1982):

$$b = (\hat{D}_2 - a) / \Gamma_1 \tag{7}$$

$$\hat{D}_{Q2}^2 + a^2 - 2a\hat{D}_2 - b^2\Gamma_2 = 0 \tag{8}$$

where $\hat{D}_2 = \sqrt{\hat{D}_{Q2}^2 - SD_2^2}$ and $\Gamma_1 = \Gamma(1 + i/c)$.

**Method 2—Hybrid**

The Weibull location parameter was also computed from $a = 0.5\hat{D}_{\text{min}}$. In the hybrid method, developed by Baldwin and Feduccia (1987), the $b$ and $c$ parameters were recovered from a moment (the quadratic mean diameter or $\hat{D}_{Q2}$) and a percentile (the 93rd diameter percentile or $\hat{D}_{93}$):

$$b = (\hat{D}_{93} - a) / 2.659261/c \tag{9}$$

$$0 = a2 - \hat{D}_{Q2} + 2a(\hat{D}_{93} - a)\Gamma_1 / 2.659261c + a(\hat{D}_{93} - a)2\Gamma_2 / 2.659261c \tag{10}$$

where $2.65926 = -\ln(1 - 0.93)$.

**Method 3—CDF Regression**

Similar to the previous two methods, the Weibull location parameter was set at $a = 0.5\hat{D}_{\text{min}}$. The scale and shape parameters were computed from:

$$b = -a\Gamma_1 + \sqrt{a^2(\Gamma_1^2 - \Gamma_2)} + D_{Q2}^2\Gamma_2 / \Gamma_2 \tag{11}$$

$c = \exp(b_1 + b_2RS_2 + b_3\hat{H}_2) \tag{12}$

As in the procedure developed by Cao (2004), this method iteratively searched for the coefficients $b_i$’s in equation (12) to minimize the following function:

$$\sum_{i=1}^{p} \sum_{j=1}^{n} (F_{ij} - \hat{F}_{ij})^2/n_i \tag{13}$$

where $F_{ij} =$ observed cumulative probability of tree $j$ in the $i^{th}$ plot, $\hat{F}_{ij} = 1 - \exp \left( - \left( (x_{ij} - a) / b \right)^c \right) =$ value of the cumulative distribution function (cdf) of the Weibull distribution evaluated at $x_{ij}, x_i =$ d.b.h. of tree $j$ in the $i^{th}$ plot, $n_i =$ number of trees in the $i^{th}$ plot, and $p =$ number of plots. This approach, similar to the procedure developed by Cao (2004), fit a regression equation for $c$, but the objective was to minimize the sum of squares of error with respect to the cdf (hence the name CDF Regression), rather than to $c$.

**Method 4—Cumulative TPA Regression**

This method is similar to method 3 (CDF Regression), except that equation (12) was replaced with:

$$c = \exp(b_1 + b_2RS_2 + b_3\hat{H}_2 + b_4\ln(\hat{N}_2)) \tag{14}$$

The coefficients $b_i$’s in equation (14) were iteratively searched to minimize the following function:

$$\sum_{i=1}^{p} \sum_{j=1}^{n} (T_{ij} - \hat{T}_{ij})^2/n_i \tag{15}$$

where $T_{ij}$ and $\hat{T}_{ij}$ are observed and predicted cumulative number of trees per acre up to diameter $x_i$.

Again, this approach was similar to fitting a regression equation for $c$, but the goal was to minimize the sum of squares of error with respect to the cumulative trees per acre (hence the name Cumulative TPA Regression), rather than to $c$.

**EVALUATION**

The following goodness-of-fit statistics were computed separately for each method and for each plot of the fit data and each plot-age combination of the validation data.

1. The chi-square statistic, and
2. Reynolds and others (1988) error indices for the $j^{th}$ plot-age combination, based on number of trees per acre in each diameter class ($EI_{N,j}$) and basal area per acre in each diameter class ($EI_{B,j}$):

$$EI_{N,j} = \sum |m_{ik} - \hat{m}_{ik}| \tag{16}$$

$$EI_{B,j} = \sum |b_{ik} - \hat{b}_{ik}| \tag{17}$$

where $m_{ik}$ and $\hat{m}_{ik}$ are the observed and predicted number of trees per acre in diameter class $k$ for the $j^{th}$ plot-age combination, and $b_{ik}$ and $\hat{b}_{ik}$ are the observed and predicted basal area per acre in diameter class $k$ for the $j^{th}$ plot-age combination. The sum includes all diameter classes in the $j^{th}$ plot-age combination.
RESULTS AND DISCUSSION
Table 2 displays the summaries of the chi-square statistic and the error indices for both datasets. The CDF Regression method produced the lowest goodness-of-fit statistics for both fit and validation data in all but one case. The lone exception is its second rank in terms of $EI_{\text{TPA}}$ for the fit dataset. On the other hand, the rankings of the rest of the methods were not clearcut, with each method received ranks varied from 2 to 4.

The same predicted variables ($\hat{N}_2$, $\hat{D}_{\text{min}}$, and $\hat{D}_{q}$) were employed in all methods. This means that all methods used the same location parameter and were constrained to produce the same total basal area per acre. The CDF Regression method performed well because the entire distribution (rather than just the average diameter or the 93rd diameter percentile) contributed to the fitting criterion.

Because total number of trees had to be predicted, error in the total number would in turn cause error in predicting number of trees in each diameter class. The Cumulative TPA Regression method was designed to optimize the Weibull parameter prediction by working directly with number of trees rather than proportions as did the CDF Regression. It was somewhat surprising to find that the CDF Regression was still the better method.

Cao (2004) concluded that the CDF Regression was best for predicting Weibull parameters to characterize current stands with known stand attributes. In this study, the CDF Regression was also found to be the most appropriate method for a future stand where its attributes had to be predicted from current attributes.

<table>
<thead>
<tr>
<th>Method</th>
<th>Chi-square statistic</th>
<th>Error index based on TPA</th>
<th>Error index based on BA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fit data</td>
<td>Validation data</td>
<td>Fit data</td>
</tr>
<tr>
<td>Moment Estimation</td>
<td>15(29)</td>
<td>30(88)</td>
<td>247 (108)</td>
</tr>
<tr>
<td>Hybrid</td>
<td>8 (6)</td>
<td>11 (25)</td>
<td>281 (132)</td>
</tr>
<tr>
<td>CDF Regression</td>
<td>8 (5)</td>
<td>9 (18)</td>
<td>250 (114)</td>
</tr>
<tr>
<td>CTPA Regression</td>
<td>6 (5)</td>
<td>7 (6)</td>
<td>250 (114)</td>
</tr>
</tbody>
</table>

*The smaller the statistic value, the better the fit.
*For each statistic, an underlined italic number indicates the best among four methods.

LITERATURE CITED