Evaluation of Methods to Predict Weibull Parameters for Characterizing Diameter Distributions

Krishna Prasad Poudel and Quang V. Cao

Abstract: Compared with other distribution functions, the Weibull distribution has been more widely used in describing diameter distributions because of its flexibility and relative simplicity. Parameters of the Weibull distribution are generally predicted either by the parameter prediction method or by the parameter recovery method. The coefficients of the regression equations for predicting Weibull parameters, moments, or percentiles are often estimated by use of different approaches such as ordinary least squares, seemingly unrelated regression (SUR), or cumulative distribution function regression (CDFR). However, there is no strong rationale for preferring one method over the other. We developed and evaluated different methods of predicting parameters of Weibull distribution to characterize diameter distribution using data from the Southwide Seed Source Study. The SUR and the CDFR approaches were applied to 10 different parameter prediction and parameter recovery methods. A modified CDFR approach was developed by modifying the CDFR technique such that the CDF is computed using information from diameter classes instead of individual trees as in the CDFR approach. These methods were evaluated based on four goodness-of-fit statistics (Anderson-Darling, Kolmogorov-Smirnov, negative log-likelihood, and error index). The CDFR approach provided better results than the SUR approach for all methods. The modified CDFR approach consistently provided better results than the SUR approach and was superior to the CDFR approach in all evaluation statistics but the Anderson-Darling statistic. FOR. SCI. 59(2): 243–252.

Keywords: CDF regression, parameter prediction, parameter recovery

GROWTH-AND-YIELD INFORMATION IS ESSENTIAL for forest management planning. The approaches commonly used in growth-and-yield modeling can be classified into three broad categories: whole-stand models, size-class models, and individual-tree models (Burkhart 1979). Whole-stand models require few details to simulate growth but provide rather general information about the future stand. They predict future yields as a function of stand-level attributes, such as stand age, site index, and stand density. Individual-tree models use a tree as the basic growth unit and therefore provide detailed information about stand dynamics. Depending on whether or not tree location coordinates are required, they are further divided into two classes, distance-dependent and distance-independent. Size-class models are a compromise between whole-stand models and individual-tree models. When the class size is infinitely large and only one class exists, then the method is the whole-stand approach. When the class width is infinitely small and each tree is considered a single class, then the method is the individual-tree approach (Vanclay 1991).

Diameter distribution models are size-class models that provide information for each diameter class: number of trees, basal area, and volume per unit area. Diameter distribution is usually characterized by estimating the parameters of some theoretical distributions (Kangas and Maltamo 2000). Many different probability density functions (PDFs) such as log-normal, exponential, gamma, beta, Weibull, and Johnson's S_B have been used to describe diameter distributions. Clutter and Bennett (1965) applied the beta distribution to describe diameter distributions on data from old-field slash pine plantations to develop variable density stand tables. The beta distribution was used in many yield models for slash pine (Bennett and Clutter 1968, Bennett et al. 1978), loblolly pine (Lenhart and Clutter 1971, Lenhart 1972, Burkhart and Strub 1974), birch, and pedunculate oak (Gorgoso-Varela et al. 2008).

Since Bailey and Dell (1973) introduced the use of the Weibull function in forestry, it has been favored by many researchers because it can fit a variety of shapes. Another advantage is that its cumulative distribution function (CDF) exists in closed form, allowing easy calculation of proportion of trees in each diameter class. The Weibull function has been widely used to model diameter distribution of loblolly pine (Smalley and Bailey 1974, Feduccia et al. 1979, Matney and Sullivan 1982, Clutter et al. 1984, Baldwin and Feduccia 1987), longleaf pine (Lohrey and Bailey 1977, Jiang and Brooks 2009), loblolly and slash pine (Brooks et al. 1992), Scots pine, Austrian pine, and *Pinus halepensis* (Palahi et al. 2006), birch (Gorgoso-Varela et al. 2007), and black poplar (Andrasev et al. 2009).

Initially, the Weibull parameters were directly predicted from regression as functions of stand attributes, such as age and measures of site quality and stand density. Recent methods have been developed to recover the Weibull

Krishna Prasad Poudel (krishna.poudel@oregonstate.edu), Oregon State University. Quang V. Cao (qcao@lsu.edu), Louisiana State University, Baton Rouge, LA.

Table 1. Summary statistics of the stand- and tree-level variables for the fit and validation data from 10 bootstrap samples.

Variable	n	Mean	SD	Minimum	Maximum
Fit data					
Dominant height (m)	2,000	14.8	4.2	4.2	24.1
No. of trees/ha	2,000	1509	522	915	2929
Basal area (m²/ha)	2,000	31.0	9.5	4.5	61.8
Tree diameter (cm)	50,419	15.3	5.1	1.0	36.1
Validation data					
Dominant height (m)	2,000	14.3	4.2	4.2	24.1
No. of trees/ha	2,000	1757	581	915	2990
Basal area (m²/ha)	2,000	33.0	9.2	4.3	61.8
Tree diameter (cm)	57,590	14.7	4.7	1.0	36.1

parameters from diameter moments and/or percentiles, which are first predicted by use of regression equations from stand-level variables. These two approaches are called parameter prediction and parameter recovery, respectively (Siipilehto et al. 2007). Coefficients of regression equations for predicting the Weibull parameters (in the parameter prediction method) and moments or percentiles (in the parameter recovery method) are often estimated by use of ordinary least squares or seemingly unrelated regression.

Cao (2004) obtained the regression coefficients in the parameter prediction method by minimizing the sum of squared differences between the observed and predicted cumulative probability. He termed this new approach the CDF regression (CDFR) method, which produced better goodness-of-fit statistics than other methods. The CDFR technique was also found by Newton and Amponsah (2005) and Cao and McCarty (2006) to yield the best goodness-of-fit statistics among the methods tested. Nord-Larsen and Cao (2006) applied this technique to even-aged beech with satisfactory results. Jiang and Brooks (2009), however, found that the parameter recovery method by Bailey et al. (1989) provided better results than the CDFR method for young longleaf pine plantations.

A successful diameter-distribution model requires good prediction of its parameters. The objective of this study was to develop new methods for predicting parameters of the Weibull PDF for characterizing diameter distributions. The CDFR method used by Cao (2004) for the parameter prediction method was modified and extended to various parameter recovery methods. These new methods were then evaluated against the established methods based on the abilities of the resulting Weibull PDFs to approximate diameter distributions of loblolly pine plantations.

Data

Data from the Southwide Seed Source Study, which involves 15 loblolly pine (*Pinus taeda* L.) seed sources planted at 13 locations across 10 southern states (Wells and Wakeley 1966), were used. Seedlings were planted at 1.8-m × 1.8-m spacing. Each plot is of size 0.0164 ha, and trees were measured at ages of 10, 15, 20, and 25 years.

The bootstrapping method was used to select 10 different random samples from the original data set. To be included in the samples, a plot had to have at least 15 surviving trees at measurement age. Each sample consisted of a fit data set, which comprised 50 randomly selected plots for each age group (200 plots in total), and a validation data set, which

included another 50 randomly selected plots. The validation data consisted of all four measurements for each plot, resulting in 200 plot-age combinations in each sample. The fit data were used for development of predicting equations, whereas the validation data were used to evaluate the methods. Table 1 presents the summary statistics of the stand- and tree-level variables for the fit and validation data for all 10 samples.

Methods

The Weibull PDF used in this study is of the following form:

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

where a, b, and c are the location, scale, and shape parameters of the Weibull distribution, respectively, and x is tree dbh. In defining the diameter distribution, the location parameter a relates to the smallest possible dbh in the stand, the scaling parameter b describes the spread of the distribution, and the parameter c determines the shape of the distribution.

The following general form of regression equation, adopted from Cao (2004), was used to predict the Weibull parameters or diameter moments or percentiles:

$$y = \exp[b_1 + b_2 \, RS + b_3 \ln(N) + b_4 \ln(H) + b_5/A] + \varepsilon,$$

$$(2)$$

where y represents Weibull parameters, diameter percentiles, or moments (mean or variance), $RS = (\sqrt{10,000/N})/H$ is the relative spacing (ratio of the average distance between trees to the average height of the dominants and codominants), N is number of trees per ha, H is dominant height (average height of the dominants and codominants) in meters, A is stand age in years, $\ln(\cdot)$ is the natural logarithm, b_ks are regression coefficients, and ε is random error.

Methods for Predicting Weibull Parameters

The 10 methods evaluated in this study belonged to four broad categories: parameter prediction, moment-based parameter recovery, percentile-based parameter recovery, and hybrid methods (Table 2). Method 1 is a parameter prediction method. This method was first introduced by Clutter and Bennett (1965) to predict parameters of the beta distributions for old-field slash pine plantations. Three Weibull parameters that were obtained from the fit data via

Table 2.   Summary of methods for predicting Weibull parameters included in this study.

Methods	Equations[a]
Parameter prediction	
Method 1: Weibull parameters	a, b, or $c = \exp[b_1 + b_2\,RS + b_3\ln(N) + b_4\ln(H) + b_5/\text{age}] + \varepsilon$
Moment-based parameter recovery	
Method 2: $\bar{D}$ and D_{var}	$b = \dfrac{\hat{\bar{D}} - a}{G_1}$; c is obtained from $b^2(G_2 - G_1^2) - \hat{D}_{\text{var}} = 0$
Method 3: D_q and D_{var}	$b = -aG_1/G_2 + [(a/G_2)^2(G_1^2 - G_2) + \hat{D}_q^2/G_2]^{0.5}$; c is obtained from $b^2(G_2 - G_1^2) - \hat{D}_{\text{var}} = 0$
Percentile-based parameter recovery	
Method 4: D_{31} and D_{63}	$c = \dfrac{\ln\!\left(\dfrac{\ln(1 - 0.63)}{\ln(1 - 0.31)}\right)}{\ln(\hat{D}_{63} - a) - \ln(\hat{D}_{31} - a)}$; $b = \dfrac{\hat{D}_{63} - a}{[-\ln(1 - 0.63)]^{1/c}}$
Method 5: D_{50} and D_{95}	$c = \dfrac{\ln\!\left(\dfrac{\ln(1 - 0.95)}{\ln(1 - 0.50)}\right)}{\ln(\hat{D}_{95} - a) - \ln(\hat{D}_{50} - a)}$; $b = \dfrac{\hat{D}_{50} - a}{[-\ln(1 - 0.50)]^{1/c}}$
Hybrid methods	
Method 6: D_q, D_{25}, and D_{95}	$c = \dfrac{\ln\!\left(\dfrac{\ln(1 - 0.95)}{\ln(1 - 0.25)}\right)}{\ln(\hat{D}_{0.95} - a) - \ln(\hat{D}_{0.25} - a)}$; $b = -aG_1/G_2 + [(a/G_2)^2(G_1^2 - G_2) + \hat{D}_q^2/G_2]^{0.5}$
Method 7: D_q, D_{31}, and D_{63}	$c = \dfrac{\ln\!\left(\dfrac{\ln(1 - 0.63)}{\ln(1 - 0.31)}\right)}{\ln(\hat{D}_{0.63} - a) - \ln(\hat{D}_{0.31} - a)}$; $b = -aG_1/G_2 + [(a/G_2)^2(G_1^2 - G_2) + \hat{D}_q^2/G_2]^{0.5}$
Method 8: $\bar{D}$ and D_{95}	$b = \dfrac{\hat{D}_{95} - a}{[-\ln(1 - 0.95)]^{1/c}}$; c is obtained from $a + b\Gamma(1 + 1/c) - \hat{\bar{D}} = 0$
Method 9: D_q and D_{95}	$b = \dfrac{\hat{D}_{95} - a}{[-\ln(1 - 0.95)]^{1/c}}$; c is obtained from $b^2 G_2 + 2abG_1 + a^2 - \hat{D}_q^2 = 0$
Method 10: D_q, D_{25}, D_{50}, and D_{95}	$a = (n^{1/3}\hat{D}_0 - \hat{D}_{50})/(n^{1/3} - 1)$; $c = \dfrac{\ln\!\left(\dfrac{\ln(1 - 0.95)}{\ln(1 - 0.25)}\right)}{\ln(\hat{D}_{95} - a) - \ln(\hat{D}_{25} - a)}$; $b = -aG_1/G_2 + [(a/G_2)^2(G_1^2 - G_2) + \hat{D}_q^2/G_2]^{0.5}$

[a] Notation: a, b, and c are Weibull location, shape, and scale parameters, respectively; $RS = (\sqrt{10{,}000/N})/H$ is the relative spacing (ratio of the average distance between trees to the average height of the dominants and codominants); N is number of trees per ha; H is dominant height (average height of the dominants and codominants) in m; A is stand age in years; $\ln(.)$ is the natural logarithm; b_ks are regression coefficients; $\hat{D}_p$ is predicted value of the pth diameter percentile; G_i is $\Gamma(1 + i/c)$, where $\Gamma(\cdot)$ is the complete gamma function; $\hat{\bar{D}}$, predicted mean diameter; $\hat{D}_{\text{var}}$ = predicted diameter variance; D_q = quadratic mean diameter; and n = number of trees in the plot.

maximum likelihood estimation were then used as dependent variables in Equation 2.

Methods 2 and 3 are moment-based parameter recovery methods. The Weibull location parameter was computed from the predicted minimum diameter ($\hat{D}_0$) in the stand as follows:

$$a = 0.5\hat{D}_0. \qquad (3)$$

This method of computing the Weibull location parameter was used for all subsequent methods, except for method 10. For the rest of the article, $\hat{D}_p$ denotes the predicted pth diameter percentile.

Equation 2 was used to predict minimum diameter (D_0), average diameter ($\bar{D}$), quadratic mean diameter (D_q), and diameter variance (D_{var}). In method 2, predicted average diameter and diameter variance were used to recover the Weibull shape and scale parameters.

Similarly, these Weibull parameters were recovered from quadratic mean diameter and diameter variance in method 3.

Methods 4 and 5, the percentile-based parameter recovery methods, used different combinations of diameter percentiles in recovering the Weibull parameters. Method 4 used $\hat{D}_{31}$ and $\hat{D}_{63}$ and method 5 used $\hat{D}_{50}$ and $\hat{D}_{95}$ to recover the Weibull shape and scale parameters.

Methods that use both moments and percentiles for recovering Weibull parameters are defined as hybrid methods by Cao (2004). Methods 6–10 in this study belonged to this category. In method 6, the Weibull scale and shape parameters were recovered from $\hat{D}_{25}$, $\hat{D}_{95}$, and the predicted quadratic mean diameter ($\hat{D}_q$). Method 7 is similar to method 6, except that $\hat{D}_{31}$ and $\hat{D}_{63}$ replaced $\hat{D}_{25}$ and $\hat{D}_{95}$. Methods 8 and 9 involve $\hat{D}_{95}$ and either the predicted average diameter ($\hat{D}$, method 8), or $\hat{D}_q$ (method 9). Method 10, adopted from Bailey et al. (1989), is similar to method 6, except that the Weibull location parameter was computed as a function of $\hat{D}_0$, $\hat{D}_{50}$, and the number of trees in the plot.

Three Approaches of Model Fitting

Investigated in this study were three different approaches for obtaining estimates for the regression coefficients, b_ks, in general Equation 2. The approaches are described below.

The SUR Approach

Because the error terms are correlated among the equations used to predict Weibull parameters, moments, or percentiles, the seemingly unrelated regression (SUR) approach was used to simultaneously estimate the b_ks in the system of equations with SAS MODEL procedure, option SUR (SAS Institute, Inc. 2009). The dependent variables were the three Weibull parameters for method 1 and the diameter moments and/or percentiles for the rest of the methods.

The CDFR Approach

The CDFR approach was originally developed by Cao (2004). In this approach, the b_ks were obtained by minimizing the sum of squared differences between observed and predicted cumulative probability:

$$\text{minimize } \sum_{i=1}^{p} \frac{\sum_{j=1}^{n_i}(F_{ij} - \hat{F}_{ij})^2}{n_i} \qquad (4)$$

where $F_{ij} = (j - 0.5)/n_i$ is observed cumulative probability of the tree j in the ith plot-age combination, j is rank (from smallest to largest) of that tree in terms of dbh in the ith plot-age combination, n_i is the number of trees in the ith plot-age combination, $\hat{F}_{ij} = 1 - \exp\{-[(x_{ij} - a)/b]^c\}$, value of the Weibull CDF evaluated at x_{ij}, x_{ij} is dbh of tree j in the ith plot-age combination, and p is the number of plot-age combinations.

Note that $F_{ij} = (j - 0.5)/n_i$ is the Hazen's plotting position used for graphing empirical PDFs (Cunnane 1978). This method results in a CDF value less than 1 for the observed maximum diameter in a plot. In contrast, $F_{ij} = j/n_i$ as defined by Cao (2004) leads to a maximum CDF value of 1 for each plot. The former formula was used in this study because the predicted Weibull CDF would approach but never attain the value of 1.

The minimum diameter was predicted separately from the system of equations because including the Weibull location parameter in the system of equations, for many plots, resulted in too low a value for the predicted location parameters compared with the observed minimum diameters. The location parameter was computed from $\hat{D}_0$ for models 1–9 and from $\hat{D}_0$,

$\hat{D}_{50}$, and the number of trees in the plot for model 10. SAS MODEL procedure (SAS Institute, Inc. 2009) was used to fit the CDF regression. In some of the methods (2, 3, 8, and 9), the Weibull shape parameter, c, was solved iteratively during the fitting process by use of the secant method (Press et al. 1992).

Modified CDFR Approach

The modified CDFR approach is similar to the CDFR approach, except that diameter class information, instead of individual trees, as in the CDFR approach, is used to compute the CDF. The modified CDF regression was then fitted with SAS MODEL procedure (SAS Institute, Inc. 2009). Diameter classes having 2-cm width were used and the b_ks were obtained by minimizing the following function:

$$\text{minimize } \sum_{i=1}^{p} \frac{\sum_{j=1}^{m_i-1}(F_{ij} - \hat{F}_{ij})^2}{m_i - 1} \qquad (5)$$

where $F_{ij} = (\sum_{k=1}^{j} n_{ik})/n_i$ is observed cumulative probability of the kth diameter class in the ith plot-age combination, n_{ik} is number of trees in the kth diameter class in the ith plot-age combination, m_i is total number of diameter classes in the ith plot-age combination, $\hat{F}_{ij} = 1 - \exp\{-[(x_{ij} + 1 - a)/b]^c\}$, value of the Weibull CDF evaluated at the upper bound of the kth diameter class, and x_{ik} is the midpoint of the kth diameter class in the ith plot-age combination.

From 5, note that the sum for diameter classes only goes from 1 to $m_i - 1$ because we ignored the maximum diameter class, whose CDF value at its upper bound is 1. The maximum dbh class was not necessary in this case because the theoretical Weibull distribution would approach 1 anyway. Preliminary analysis showed that this approach was better than the one that included the maximum diameter class. A sample SAS program outlining the fitting procedure for methods 2 and 10 using the modified CDFR approach is presented in the Appendix.

Model Evaluation

The following four goodness-of-fit statistics were computed for each method to evaluate the methods. The method producing the lowest values for each of the evaluation statistics is the best method.

Anderson-Darling (AD) statistic (Anderson and Darling 1954):

$$AD_i = -n_i - \sum_{j=1}^{n_i}(2j - 1)[\ln(u_j) + \ln(1 - u_{n_i-j+1})]/n_i, \qquad (6)$$

where $u_j = F(x)_j = 1 - \exp\{-[(x_j - a)/b]^c\}$, n_i is number of trees in the ith plot-age combination, and x_js are dbh, sorted in ascending order for each plot-age combination ($x_1 \leq x_2 \ldots \leq x_{n_i}$).

Kolmogorov-Smirnov (KS) statistic (Massey 1951):

$$KS_i = \max\{\max_{1 \leq i \leq n_i}[(j/n_i) - u_j],$$

$$\max_{1 \leq i \leq n_i}[(u_j - (j - 1))/n_i]\}, \qquad (7)$$

where n_i and u_j were previously defined in Equation 6.

Table 3. Means (and SD) of the goodness-of-fit statistics produced by different methods based on three approaches.

Method	AD	KS	mLogL	EI	Rank
SUR					
1	3.343 (3.860)	0.2374 (0.092)	73.970 (20.976)	2,361.88 (830.58)	28.000
2	2.883 (3.529)	0.2283 (0.086)	72.994 (20.263)	2,304.06 (795.49)	9.943
3	3.003 (3.578)	0.2315 (0.089)	73.104 (20.309)	2,323.59 (815.32)	14.446
4	9.907 (7.928)	0.4092 (0.114)	81.579 (22.621)	2,420.84 (802.78)	
5	18.040 (10.610)	0.5373 (0.104)	90.042 (25.125)	2,590.28 (855.29)	
6	3.022 (4.048)	0.2277 (0.087)	73.171 (20.945)	2,311.86 (811.06)	12.553
7	3.052 (4.181)	0.2279 (0.087)	73.318 (21.321)	2,312.71 (815.38)	13.848
8	3.211 (4.412)	0.2298 (0.088)	73.336 (21.426)	2,320.48 (824.68)	16.577
9	3.236 (4.408)	0.2309 (0.090)	73.375 (21.382)	2,328.30 (831.78)	18.115
10	2.892 (3.489)	0.2312 (0.085)	73.252 (20.177)	2,321.01 (814.80)	14.154
CDFR					
1	2.461 (2.679)	0.2224 (0.079)	73.400 (19.770)	2,293.92 (782.35)	6.085
2	2.456 (2.636)	0.2227 (0.079)	73.398 (19.767)	2,294.61 (782.95)	6.198
3	2.449 (2.593)	0.2228 (0.078)	73.467 (19.790)	2,294.58 (782.11)	6.585
4	2.462 (2.737)	0.2221 (0.079)	73.361 (19.782)	2,292.76 (781.92)	5.610
5	2.465 (2.726)	0.2222 (0.079)	73.367 (19.810)	2,293.74 (783.93)	5.795
6	2.523 (2.593)	0.2243 (0.078)	73.709 (19.956)	2,299.67 (783.26)	9.692
7	2.485 (2.846)	0.2240 (0.079)	73.502 (19.927)	2,284.57 (775.33)	6.852
8	2.453 (2.637)	0.2228 (0.079)	73.422 (19.731)	2,293.63 (780.70)	6.258
9	2.460 (2.629)	0.2230 (0.079)	73.462 (19.747)	2,294.48 (782.02)	6.732
10	2.490 (2.724)	0.2229 (0.078)	73.319 (19.725)	2,302.22 (793.88)	6.682
Modified CDFR					
1	2.527 (3.144)	0.2210 (0.080)	72.908 (19.866)	2,279.16 (773.73)	1.724
2	2.528 (3.216)	0.2201 (0.080)	72.858 (19.898)	2,278.73 (778.24)	1.050
3	2.526 (3.207)	0.2201 (0.080)	72.858 (19.898)	2,278.52 (778.16)	1.000
4	2.546 (3.332)	0.2195 (0.080)	72.892 (20.061)	2,280.95 (782.36)	1.351
5	2.555 (3.277)	0.2210 (0.080)	72.909 (19.978)	2,281.39 (779.90)	2.139
6	2.546 (2.959)	0.2221 (0.079)	73.028 (19.941)	2,288.65 (782.86)	3.832
7	2.537 (3.016)	0.2210 (0.079)	72.902 (19.957)	2,286.87 (786.77)	2.424
8	2.542 (3.260)	0.2203 (0.080)	72.895 (19.927)	2,279.67 (779.16)	1.529
9	2.555 (3.267)	0.2207 (0.080)	72.900 (20.002)	2,282.18 (782.98)	2.029
10	2.586 (3.366)	0.2218 (0.079)	72.961 (19.854)	2,288.90 (789.26)	3.646

Overall rank for each of the 28 methods is the relative rank based on the sum of the relative ranks from the four goodness-of-fit statistics. SUR methods 4 and 5 were not included in the ranking because their statistics were much worse than those from the remaining methods.

Negative log-likelihood (mLogL) statistic:

$$\text{mLogL} = \sum_{j=1}^{n_i} \left[\ln(b) - \ln(c) + (1-c) \ln\left(\frac{x_{ij} - a}{b}\right) \right.$$
$$\left. + \left(\frac{x_{ij} - a}{b}\right)^c \right], \quad (8)$$

where n_i was previously defined in Equation 6, x_{ij} is the dbh of tree j in the ith plot-age combination, and mLogL is the negative value of the log-likelihood function of the Weibull distribution.

Error index (EI; Reynolds et al. 1988):

$$EI_i = \sum_{k=1}^{m_i} |n_{ik} - \hat{n}_{ik}|, \quad (9)$$

where n_{ik} and $\hat{n}_{ik}$ are observed and predicted number of trees per ha in diameter class k, respectively, and m_i is number of diameter classes for the ith plot-age combination.

Ranking of Methods

The traditional standard or ordinal ranks for m methods are 1, 2, …, m. They show the order of the methods, but fail to depict the exact positions of the methods compared with one another. In this study, we propose a new method of ranking, in which each method is given a relative rank, computed to display the relative position of the method. The relative rank of method i is defined as

$$R_i = 1 + \frac{(m-1)(S_i - S_{\min})}{S_{\max} - S_{\min}} \quad (10)$$

where R_i is the relative rank of method i ($i = 1, 2, …, m$), S_i is the goodness-of-fit statistic produced by method i, $S_{\min}$ is the minimum value of S_i, and $S_{\max}$ is the maximum value of S_i.

In this ranking system, the best and the worst methods have relative ranks of 1 and m, respectively. Ranks of the remaining methods are expressed as real numbers between 1 and m. Because the magnitude and not only the order of the S_is is taken into consideration, the new ranking system should provide more information than the traditional ordinal ranks. For example, relative ranks of 1, 1.2, 4.7, 4.9, and 5 in the case of five methods suggest that the methods fall into two groups; the first group (1 and 1.2) and the second group (4.7, 4.9, and 5) were separated by a large gap.

Results and Discussion

All 10 methods were performed using three different approaches (SUR, CDFR, and modified CDFR), resulting in a total of 30 methods. The full model used to predict the

Weibull parameters, diameter moments, and diameter percentiles were in the form of Equation 2. The final models were obtained based on the backward elimination procedure by manually removing insignificant variables (at the 5% level of significance). We preferred the backward elimination approach because variables tend to perform well in groups that might be missed by the forward and stepwise approaches. Means and SDs for four goodness-of-fit statistics along with their overall relative ranks produced by the 30 methods are shown in Table 3. Methods 4 and 5 of the SUR approach produced evaluation statistics that were much higher (worse) than those from the rest of methods and were dropped from the ranking. As a result, the evaluation was based on statistics from the remaining 28 methods. The overall ranks ranged from 1.000 to 3.832 for the modified CDFR approach, from 5.610 to 9.692 for the CDFR approach, and from 9.943 to 28.000 for the SUR approach. Method 3 of the modified CDFR approach ranked as the best method among all 30 methods evaluated in this study, whereas method 5 based on the SUR approach was the poorest performer. The radar plot based on the relative ranks of 30 methods (Figure 1) shows three distinct groups. The modified CDFR approach was best, followed by the CDFR approach. The SUR approach was a distant third.

SUR Versus the Two CDFR Approaches

The CDFR approach yielded lower AD, KS, and EI values than did the SUR approach. There was, however, some overlap in the mLogL statistic between the SUR and CDFR approaches. On the other hand, the modified CDFR approach consistently produced lower values for all four evaluation statistics than did the SUR approach. These results indicate that the CDFR and the modified CDFR approaches were clearly superior to the SUR approach. On average, the modified CDFR approach reduced the AD, KS, EI, and mLogL statistics by 51.61, 20.91, 3.83, and 3.26%, respectively, compared with the SUR approach. The SUR approach did not perform as well as the other two approaches because it is based on point estimates, either the Weibull parameters, diameter moments, or percentiles. The CDFR and modified CDFR approaches, conversely, aimed to optimize the entire distribution by minimizing the squared difference between the observed and predicted cumulative probabilities.

CDFR Versus Modified CDFR Approaches

Figure 1 shows that the modified CDFR approach performed better in terms of the KS, EI, and mLogL statistics
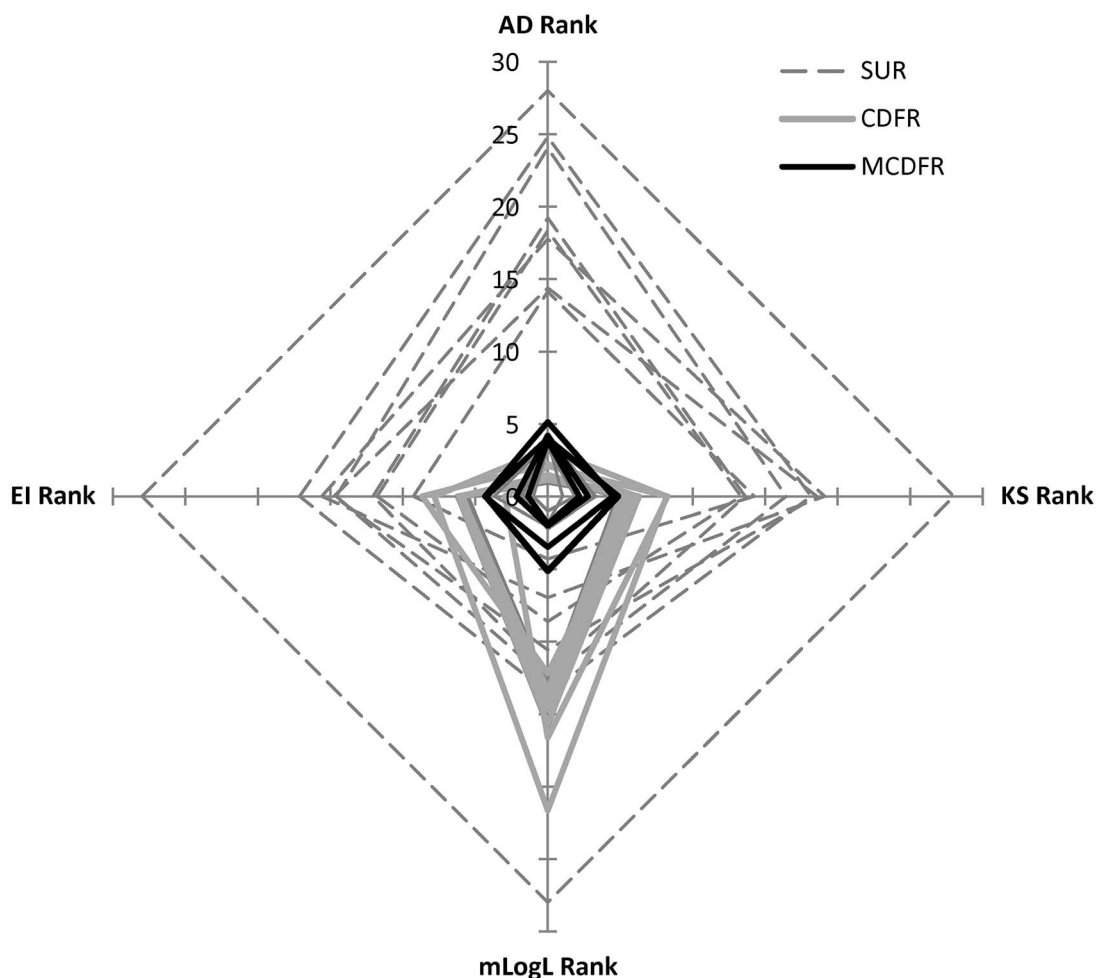

Figure 1.   Overall comparison of the SUR, CDFR, and modified CDFR (MCDFR) approaches for the 28 methods. The method resulting in the smallest area inside the box represents the best method. SUR methods 4 and 5 were not included in the ranking because their statistics were much worse than those from the remaining methods.

but consistently produced higher AD values than the CDFR approach. The reason for the increased AD values for the modified CDFR approach is not clear. The EI statistics, on the other hand, are based on diameter-class frequency and therefore should be more suitable for the modified CDFR approach, which is also based on diameter-class information.

Overparameterization

Three methods are overparameterized in the sense that they require three percentiles to estimate the Weibull shape and scale parameters: D_q, D_{31}, and D_{63} for method 7; and D_q, D_{25}, and D_{95} for methods 6 and 10. In addition, method 10 also estimates the Weibull location parameter from three attributes: $\hat{D}_0$, $\hat{D}_{50}$, and the number of trees in the plot. The performances of these methods were not consistent: good for one sample and poor for the next one. Overall, these overparameterized methods did not produce results that were as good as results for simpler methods. Furthermore, prediction of multiple diameter percentiles in these three methods might require the imposition of constraints, i.e., predicted percentiles cannot illogically cross one another.

Best and Worst Methods for the Modified CDFR Approach

Among the three approaches, the modified CDFR approach was the best for estimating the coefficients of the regression equations for predicting Weibull parameters (in the parameter prediction method) and moments or percentiles (in the parameter recovery methods). Figure 2 shows the relative ranks computed just for the 10 methods using the modified CDFR approach. The overall highest-ranked method was method 3 (D_q and D_{var}), followed closely by method 2 ($\bar{D}$ and D_{var}), with a relative rank of 1.195. The worst method was method 10 (D_q, D_{25}, D_{50}, and D_{95}), which ranked last in terms of two evaluation statistics and was second to last for the other two.

Summary and Conclusions

Whereas both the CDFR and modified CDFR approaches performed better than the SUR approach for all methods, the modified CDFR was superior to the CDFR approach for all evaluation statistics, except for the AD statistic. The results are consistent with the findings by Cao (2004), who evaluated method 1 of the CDFR approach against other methods of the SUR approach. The poor performance of the SUR approach may be because its objective is to optimize the point estimates (Weibull parameters, diameter moments and/or percentiles) instead of the distribution itself as in the CDFR and the modified CDFR approaches. For a given method, the overall ranking from the modified CDFR approach was better than that from the SUR and CDFR approaches.

If a choice is to be made among these three approaches, we would recommend the modified CDFR approach over the SUR and CDFR approaches, based on the findings of this study. On the other hand, method 2 for the SUR approach, method 4 for the CDFR approach, and methods 2 and 3 for the modified CDFR approach produced best results for each respective approach.
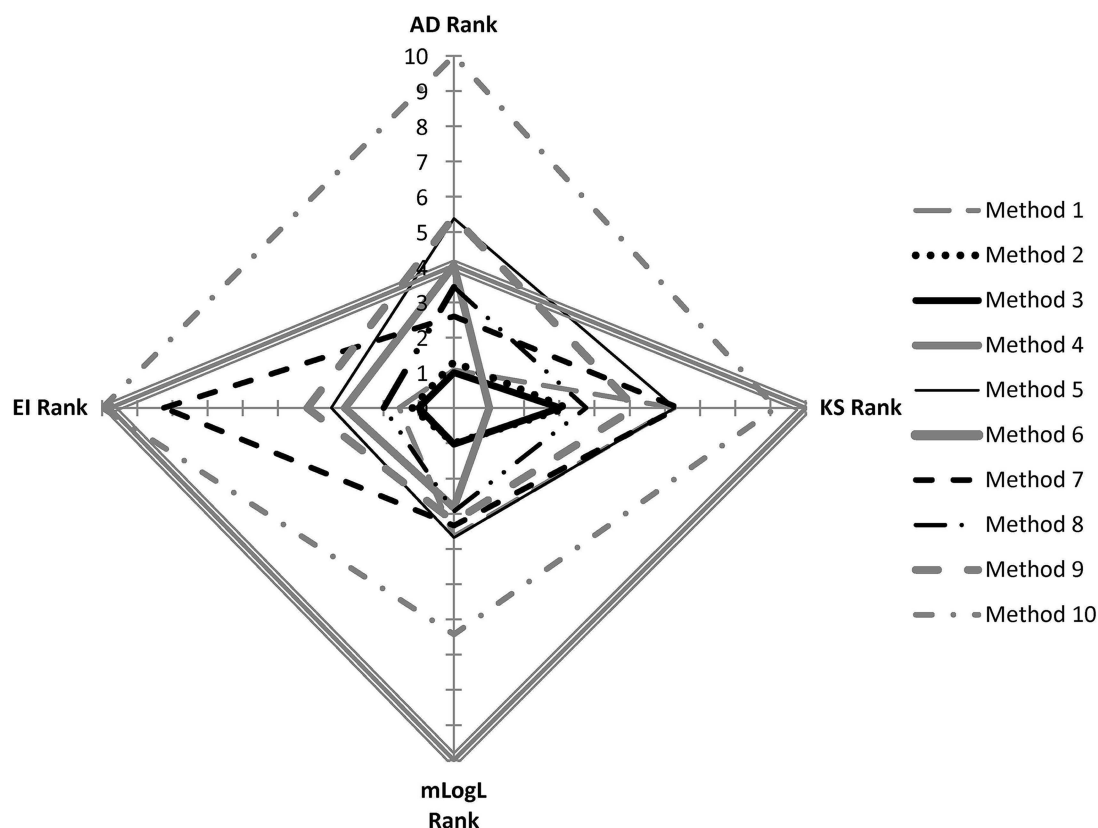

Figure 2. Relative ranks of 10 methods based on the modified CDFR approach. The method resulting in the smallest area inside the box represents the best method.

Results from the 10 bootstrap samples revealed that the relative ranks for the methods varied from sample to sample. Overall, there are sufficient reasons to believe that the SUR approach is not as effective as the other two approaches. However, performance differences between the CDFR and the modified CDFR approaches might vary with different data sets. Research using data sets from loblolly pine as well as from other species should be conducted to further verify the findings of this study.

Literature Cited

ANDERSON, T.W., AND D.A. DARLING. 1954. A test of goodness of fit. *J. Am. Stat. Assoc.* 49:765–769.

ANDRASEV, S., M. BOBINAC, AND S. ORLOVIC. 2009. Diameter structure models of black poplar selected clones in the section Aigeiros (Duby) obtained by the Weibull distribution. *Sumarski List* 133:589–603.

BAILEY, R.L., AND T.R. DELL. 1973. Quantifying diameter distributions with the Weibull function. *For. Sci.* 19:97–104.

BAILEY, R.L., T.M. BURGAN, AND E.J. JOKELA. 1989. Fertilized mid-rotation-aged slash pine plantations—Stand structure and yield prediction models. *South. J. Appl. For.* 13:76–80.

BALDWIN, V.C. JR., AND D.P. FEDUCCIA. 1987. *Loblolly pine growth and yield prediction for managed West Gulf plantations.* USDA For. Serv., Res. Paper SO-236. 27 p.

BENNETT, F.A., AND J.L. CLUTTER. 1968. *Multiple-product yield estimates for unthinned slash pine plantations—Pulpwood, saw timber, gum.* USDA For. Serv., Res. Paper SE-35. 21 p.

BENNETT, F.A., F.T. LLOYD, B.F. SWINDEL, AND E.W. WHITEHORNE. 1978. *Yields of veneer and associated products from unthinned, old-field plantations of slash pine in the north Florida and South Georgia flatwoods.* USDA For. Serv., Res. Paper SE-176. 80 p.

BROOKS, J.R., B.E. BORDERS, AND R.L. BAILEY. 1992. Predicting diameter distributions for site- prepared loblolly and slash pine plantations. *South. J. Appl. For.* 16:130–133.

BURKHART, H.E. 1979. Growth and yield of southern pines—State of the art. In *Proc. of the Southern forest economists workshop*, Chapel Hill, NC. 11 p.

BURKHART, H.E., AND M.R. STRUB. 1974. A model for simulation of planted loblolly pine stands. P. 128–135 in *Growth models for tree and stand simulation*, Fries, J. (ed.). Royal College of Forestry, Stockholm.

CAO, Q.V. 2004. Predicting parameters of a Weibull function for modeling diameter distribution. *For. Sci.* 50:682–685.

CAO, Q.V., AND S.M. MCCARTY. 2006. New methods for estimating parameters of Weibull functions to characterize future diameter distributions in forest stands. P. 338–340 in *Proc. of the Southern Silviculture Research Conference.* USDA For. Serv., Gen. Tech. Rep. SRS-92.

CLUTTER, J.L., AND F.A. BENNETT. 1965. *Diameter distributions in old-field slash pine plantations*. GA For. Res. Council Rep. 13. 9 p.

CLUTTER, J.L., W.R. HARMS, G.H. BRISTER, AND J.W. RHENNEY. 1984. *Stand structure and yields of site-prepared loblolly pine plantations in the lower coastal plain of the Carolinas, Georgia, and North Florida*. USDA For. Serv., Gen. Tech. Rep. SE-27. 173 p.

CUNNANE, C. 1978. Unbiased plotting-positions—A review. *J. Hydrol.* 37:205–222.

FEDUCCIA, D.P., T.R. DELL, W.F. MANN JR., AND B.H. POLMER. 1979. *Yields of unthinned loblolly pine plantations on cutover sites in the West Gulf region*. USDA For. Serv., Res. Paper SO-148. 88 p.

GORGOSO-VARELA, J.J., J.G. ÁLVAREZ-GONZÁLEZ, A. ROJO, AND J.A. GRANDAS-ARIAS. 2007. Modeling diameter distributions of *Betula alba* L. stands in northwest Spain with the two-parameter Weibull function. *Invest. Agrar. Sist. Recur. For.* 16:113–123.

GORGOSO-VARELA, J.J., A. ROJO-ALBORECA, E. AFIF-KHOURI, AND M. BARRIO-ANTA. 2008. Modeling diameter distributions of birch (*Betula alba* L.) and pedunculate oak (*Quercus robur* L.) stands in northwest Spain with the beta distribution. *Invest. Agrar. Sist. Recur. For.* 17:271–281.

JIANG, L.C., AND J.R. BROOKS. 2009. Predicting diameter distributions for young longleaf pine plantations in Southwest Georgia. *South. J. Appl. For.* 33:25–28.

KANGAS, A., AND M. MALTAMO. 2000. Percentile based basal area diameter distribution models for Scots pine, Norway spruce and birch species. *Silva Fenn.* 34:371–380.

LENHART, J.D. 1972. *Cubic volume yields for unthinned old-field loblolly pine plantations in the Interior West Gulf Coastal Plain.* Stephen F. Austin State Univ., Texas For. Serv., Paper 14. 46 p.

LENHART, J.D., AND J.L. CLUTTER. 1971. *Cubic-foot yield tables for old-field loblolly pine plantations in the Georgia Piedmont*. GA For. Res. Council Rep. 22—Ser. 3. 12 p.

LOHREY, R.E., AND R.L. BAILEY. 1977. *Yield tables and stand structure for unthinned long leaf pine plantations in Louisiana and Texas.* USDA For. Serv., Res. Paper SO-133. 55 p.

MASSEY, F.G. JR. 1951. The Kolmogorov-Smirnov test for goodness of fit. *J. Am. Stat. Assoc.* 46:68–78.

MATNEY, T.G., AND A.D. SULLIVAN. 1982. Compatible stand and stock tables for thinned and unthinned loblolly pine stands. *For. Sci.* 28:161–171.

NEWTON, P.F., AND I.G. AMPONSAH. 2005. Evaluation of Weibull-based parameter prediction equation systems for black spruce and jack pine stand types within the context of developing structural stand density management diagrams. *Can. J. For. Res*. 35:2996–3010.

NORD-LARSEN, T., AND Q.V. CAO. 2006. A diameter distribution model for even-aged beech in Denmark. *For. Ecol. Manage*. 231:218–225.

PALAHI, M., T. PUKKALA, AND A. TRASOBARES. 2006. Modelling the diameter distribution of *Pinus sylvestris, Pinus nigra* and *Pinus halepensis* forest stands in Catalonia using the truncated Weibull function. *Forestry* 79:553–562.

PRESS, W.H., S.A. TEUKOLSKY, W.T. VETTERLING, AND B.P. FLANNERY. 1992. *Numerical recipes in Fortran 77,* 2nd ed. Cambridge University Press, Cambridge, UK. 933 p.

REYNOLDS, M.R., T.E. BURK, AND W.C. HUANG. 1988. Goodness-of-fit tests and model selection procedures for diameter distribution models. *For. Sci.* 34:373–399.

SAS INSTITUTE, INC. 2009. *Base SAS® 9.2 procedures guide*. SAS Institute Inc., Cary, NC.

SIIPILEHTO, J., S. SARKKOLA, AND L. MEHTÄTALO. 2007. Comparing regression estimation techniques when predicting diameter distributions of Scots pine on drained peatlands. *Silva Fenn.* 41:333–349.

SMALLEY, G.W., AND R.L. BAILEY. 1974. *Yield tables and stand structure for loblolly pine plantations in Tennessee, Alabama, and Georgia highlands*. USDA For. Serv., Res. Paper SO-96. 81 p.

VANCLAY, J.K. 1991. *Modelling the growth and yield of tropical moist forests*. Queensland Forest Series, Brisbane, Australia. 151 p.

WELLS, O.O., AND P.C. WAKELEY. 1966. Geographic variation in survival, growth, and fusiform- rust infection of planted loblolly pine. *For. Sci. Monogr.* 11. 40 p.

Appendix: Sample SAS Program

```
data one;
input plot age n hd tph d dclass;
            *   age = stand age,
            *   n = number of trees per plot;
            *   hd = dominant height of the stand;
            *   tph = number of trees per hectare;
            *   d = upper bound of the diameter class;
            *   F = empirical cumulative probability up to diameter d;
   hdlog = log(hd);
   tphlog = log(tph);
   rs = (sqrt(10000/tph))/hd;
   a = 0.5*(exp(7.36711 - 2.73286*rs - 0.57797*tphlog - 7.41606/age));

data one; set one; by plot age;
   if (1-cdf)> 0.00001;                     *   Delete the maximum diameter class;

*******************************************;
*   Method 2 - The Modified CDFR Approach  *;
*******************************************;

proc model itprint;
   parms b11=5.2 b12=-1.5 b13=-0.4 b14=0.2
         b21=-5.9 b23=0.1 b24=2.5;
   pdq = exp(b11 + b12*rs + b13*tphlog + b14*hdlog);
   pdvar = exp(b21 + b23*tphlog + b24*hdlog);

   c = 1.1;                                 *   Starting value for c;
   c1 = c;
   g1 = gamma(1 + 1/c);
   g2 = gamma(1 + 2/c);
   b = ((-a*g1)/g2) + ((((a/g2)**2)*(g2*g2-g1) + (pdq*pdq/g2))**0.5);
   fOld = b*b * (g2 - g1**2) - pdvar;
   c = c1 + 1;
   g1 = gamma(1 + 1/c);
   g2 = gamma(1 + 2/c);
   b = ((-a*g1)/g2) + ((((a/g2)**2)*(g2*g2-g1) + (pdq*pdq/g2))**0.5);
   fNew = b*b * (g2 - g1**2) - pdvar;
   do j = 1 to 20 while (fOld*fNew > 0);
     c1 = c;
     c = c1 + 1;
     g1 = gamma(1 + 1/c);
     g2 = gamma(1 + 2/c);
     b = ((-a*g1)/g2) + ((((a/g2)**2)*(g2*g2-g1) + (pdq*pdq/g2))**0.5);
     fNew = b*b * (g2 - g1**2) - pdvar;
     end;                                   *   At this step, fOld and fNew
                                            *   are of opposite signs
                                            *   Solution is in the interval (c1, c);
   inc = c - c1;
   do while (abs(fNew) > 1e-8);             *   Secant method to solve fNew = 0;
                                            *   1e-8 = 0.00000001 is the tolerance;
     inc = -fNew * inc / (fNew - fOld);
     c = c + inc;
     fOld = fNew;
     g1 = gamma(1 + 1/c);
     g2 = gamma(1 + 2/c);
     b = ((-a*g1)/g2) + ((((a/g2)**2)*(g2*g2-g1) + (pdq*pdq/g2))**0.5);
     fNew =b*b * (g2 - g1**2) - pdvar;
     end;
```

```
   if d > a
     then fhat = 1 - exp(-(((d-a)/b)**c));
     else fhat = 0;
   cdf = fhat;
   _weight_ = 1/nclass;
   fit cdf;
   title 'Modified CDFR Approach - Method 2';

*********************************************;
*   Method 10 - The Modified CDFR Approach  *;
*********************************************;

proc model itprint;
   parms b11=6.4 b12=-2.3 b15=0.1
         b31=7.0 b32=-2.7 b34=0.2 b35=0.4
         b41=7.8 b42=-2.6 b43=-0.5 b44=0.2;

   d0 = exp(8.828995 - 4.91421*rs - 0.7847*tphlog);
   d50 = exp(7.017923 - 2.7597*rs - 0.51166*tphlog);
   d25 = exp(b11 + b12*rs + b15/age);
   d95 = exp(b31 + b32*rs + b34*hdlog + b35/age);
   dq = exp(b41 + b42*rs + b43*tphlog + b44*hdlog);

   a = (((n**(1/3))*d0) - d50)/((n**(1/3) - 1));
   c = (log(log(0.05)/log(0.75)) / (log(d95-a) - log(d25-a)));
   g1 = gamma(1 + 1/c);
   g2 = gamma(1 + 2/c);
   b = ((-a*g1)/g2) + ((((a/g2)**2)*(g2*g2-g1) + (dq*dq/g2))**0.5);

   if d > a
     then cdf = 1 - exp(-(((d-a)/b)**c));
     else cdf = 0;
   fit cdf;
   _weight_ = 1/nclass;

   title 'Modified CDFR Approach - Method 10';

run; quit;
```