ForestFit: An R package for modeling tree diameter distributions

Mahdi Teimouri\textsuperscript{a}, Jeffrey W. Doser\textsuperscript{b} and Andrew O. Finley\textsuperscript{b}

\textsuperscript{a}Department of Statistics, Faculty of Science and Engineering, Gonbad Kavous University, Gonbad Kavous, Iran
\textsuperscript{b}Department of Forestry, Michigan State University, East Lansing, MI, USA

\begin{abstract}
Modeling the diameter distribution of trees in forest stands is a common forestry task that supports key biologically and economically relevant management decisions. The choice of model used to represent the diameter distribution and how to estimate its parameters has received much attention in the forestry literature; however, accessible software that facilitates comprehensive comparison of the myriad modeling approaches is not available. To this end, we developed an R package called ForestFit that provides a limited set of functions to fit a large variety of popular diameter distributions, and height-diameter models. ForestFit offers several methods for estimating the parameters of two- and three-parameter Weibull distributions, which are listed in Table 1 along with relevant sources for further details regarding each estimation method.

\end{abstract}

1. Introduction

The diameter distribution of trees in a stand is a relatively simple measure used to inform forest management decisions that have both biological and economical relevance [1]. Determining models and associated parameter estimation techniques to characterize an entire stand’s diameter distribution based upon a sample of trees has been an important topic in forestry over the last 50 years [12, 21]. Using an appropriate model to describe the diameter distribution enables an evaluation of the stand growth as well as potential estimates of stand volume. Numerous probability distributions have been proposed and used to model the diameter distribution, the most common including the two-parameter Weibull distribution, three-parameter Weibull distribution [1], Johnson’s SB (JSB) distribution [7], and finite mixture distributions [37]. Forest biometricians have developed a suite of methods for estimating the parameters of these models, including simple methods such as the method of moments and method of percentiles [34], maximum likelihood techniques [11], and Bayesian methods [12]. Another important tree and stand characterization used by foresters is the diameter to height relationship. Because diameter is an inexpensive field measurement relative to acquiring tree height, the forestry literature is rich with models to predict height based on diameter [32]. Because of the extensive suite of techniques in the forestry literature to model diameter distributions and height-diameter relationships, there is a need for freely available software that facilitates streamlined model parameter estimation and comparison. To this end, we develop an open-source R package called ForestFit that provides a limited set of functions to fit a large variety of popular diameter distributions and height-diameter models. Importantly, the functions are designed to simplify comparison among models and parameter estimation techniques using a variety of model fit criteria, without requiring an in-depth knowledge of the statistical techniques used in the estimation process.

In this paper, we describe the functionality of the ForestFit package and illustrate its features using the analysis of mixed ponderosa pine (\textit{Pinus ponderosa}) forest plots located in the Blue Mountains near Burns, Oregon, USA [17].

2. ForestFit components

In this section we provide an overview of the features available in the ForestFit package along with relevant citations that provide a more rigorous statistical treatment.

2.1. Estimating parameters of the Weibull distribution

The Weibull distribution is perhaps the most conspicuous distribution used for modeling diameter distributions due to its ability to represent common shapes in diameter distribution data [1, 9, 10, 21, 23, 28]. ForestFit offers several methods for estimating the parameters of two- and three-parameter Weibull distributions, which are listed in Table 1 along with relevant sources for further details regarding each estimation method.

2.2. Bayesian analysis

Computational improvements over the last 30 years have led to the increase in popularity of Bayesian techniques for estimating model parameters for diameter distributions [12]. In ForestFit we provide functionality for estimating the parameters of the three-parameter Weibull and four-parameter JSB distributions using the Bayesian paradigm. The JSB distribution is becoming increasingly popular for modeling diameter distributions in forestry [7, 8, 41] and is commonly applied in other fields, e.g., hydrology [5] and climatology [20].

With the exception of the location parameter, ForestFit uses the same algorithm as [12] for estimating the shape and
scale parameters of the three-parameter Weibull distribution.

2.3. Modeling diameter distributions using grouped data

Data collection protocols in forestry often lead to diameter observations that are recorded in groups or classes rather than as individual tree measurements. More specifically, suppose a data set has been classified into $m$ separate groups with lower bounds $r_0, r_1, \ldots, r_m$ where $r_0 = \min \{x_1, \ldots, x_n\}$ and $r_m = \max \{x_1, \ldots, x_n\}$, with $x_i$ being the individual diameter for tree $i$ out of a total $n$ trees in the sample. We define $f_i$, for $i = 1, \ldots, m$, as the frequency or number of trees in the $i$-th group. A schematic of grouped data is provided in Table 2.

Table 2
A schematic of grouped data.

| Class number | Class         | Frequency |
|--------------|---------------|-----------|
| 1            | $(r_0 - r_1)$ | $f_1$     |
| 2            | $(r_1 - r_2)$ | $f_2$     |
| 3            | $(r_2 - r_3)$ | $f_3$     |
| \vdots       | \vdots        | \vdots    |
| $m - 1$      | $(r_{m-2} - r_{m-1})$ | $f_{m-1}$ |
| $m$          | $(r_{m-1} - r_m)$ | $f_m$     |

ForestFit provides estimation of diameter distribution parameters for the three-parameter Birnbaum-Saunders (BS) distribution, the three-parameter generalized exponential (GE) distribution, and the three-parameter Weibull distribution using grouped data through the use of the vectors $r = (r_0, r_1, \ldots, r_m)$ and $f = (f_1, f_2, \ldots, f_m)$.

2.4. Modeling diameter distributions using finite mixture models

Certain forest stands (e.g., uneven aged or disturbed stands) exhibit multimodal diameter distributions that are not adequately represented by a single probability distribution. In such settings, forest biometricians and statisticians often use finite mixture distributions to characterize complex and multimodal diameter distributions [39, 18, 37, 30, 19]. The probability density function (pdf) of a $K$-component mixture model has the form

$$g(x|\Theta) = \sum_{k=1}^{K} \omega_k f(x|\theta_k), \quad (1)$$

where $\Theta = (\mathbf{\alpha}', \mathbf{\theta}_1', \ldots, \mathbf{\theta}_K')'$ in which $\mathbf{\theta}_k = (\alpha_k, \beta_k)'$ is the parameter vector of the $k$-th component with pdf $f(\cdot|\theta_k)$ and $\mathbf{\alpha} = (\alpha_1, \ldots, \alpha_K)'$ is the vector of weight parameters. The weight parameters $\omega_k$s are non-negative and sum to one, i.e., $\sum_{k=1}^{K} \omega_k = 1$. ForestFit enables $f(\cdot|\theta_k)$ to take the following forms for individual (i.e., ungrouped) data: Burr XII, Chen, Fisher, Fréchet, gamma, Gompertz, log-logistic, log-normal, Lomax, skew-normal, and two-parameter Weibull (pdfs are given in Appendix A). For grouped data, the finite mixture distribution can be composed of gamma, log-normal, skew-normal, and two-parameter Weibull distributions.

Finite mixture distributions are increasing in popularity in all areas of environmental and ecological statistics, see, e.g., [13]. To facilitate testing and simulation using the common mixture models implemented in ForestFit, the package includes functions to compute the probability density function, cumulative density function, quantile function, and simulate random variables from the finite mixture models composed of the aforementioned probability distributions. Further, ForestFit provides specific functions to compute the probability density function, cumulative density function, quantile function, and simulate random variables from the gamma shape mixture model that has received increased attention in forestry for characterizing diameter distributions [33].

2.5. Estimating the parameters of height-diameter models

In addition to modeling diameter distributions, characterizing the height-diameter relationship is a common task in forestry [31]. A variety of mathematical relationships have been proposed to model this relationship, the most common of which are implemented in ForestFit and listed with associated references in Table 3.

3. Illustrations

Here we illustrate key features available in ForestFit. All models and estimation methods assume the data points (e.g., diameter of an individual tree) are independent and identically distributed. For illustration, we use a set of 0.08...
Table 3
Common mathematical relationships used to depict the height-diameter relationship in a forest stand. \( H \) is the height of a tree, \( D \) is the diameter of the tree, and \( \beta_1, \beta_2, \beta_3 \) are estimated parameters.

| Model          | Reference | Formula                                                                 |
|---------------|-----------|--------------------------------------------------------------------------|
| Chapman-Richards | [26]     | \( H = 1.3 + \beta_1 + \frac{\beta_2}{D + \beta_3} \)                |
| Gompertz      | [35]     | \( H = 1.3 + \beta_1 e^{-\beta_2 D^{3/2}} \)                          |
| Hossfeld IV   | [38]     | \( H = 1.3 + \frac{\beta_1}{D^p + \beta_2} \)                        |
| Korf          | [6]      | \( H = 1.3 + \beta_1 e^{-\beta_2 D^{3/2}} \)                          |
| logistic      | [22]     | \( H = 1.3 + \frac{\beta_1}{1 + \beta_2 e^{-\beta_3 D^{3/2}}} \)     |
| Prodan        | [24]     | \( H = 1.3 + \frac{\beta_1}{D^p + \beta_2 + \beta_3} \)              |
| Ratkowsky     | [25]     | \( H = 1.3 + \beta_1 e^{-\beta_2 D^{3/2}} \)                          |
| Sibbesen      | [15]     | \( H = 1.3 + \beta_1 D^{\beta_2} D^{\beta_3} \)                       |
| Weibull       | [36]     | \( H = 1.3 + \beta_1 \left(1 - e^{-\beta_2 D^{\beta_3}}\right) \)    |

Table 4
Parameter estimation methods for the Weibull distribution and associated code used in the method argument of the \texttt{fitWeibull} function. MM refers to modified moment, ML refers to maximum likelihood, and MML refers to modified maximum likelihood.

| Method               | Code    |
|----------------------|---------|
| L-moment             | \texttt{lm}     |
| ML (2 parameters)    | \texttt{ml}    |
| Logarithmic moments  | \texttt{mlm}   |
| Method of moments    | \texttt{moment}|
| Percentiles          | \texttt{pm}    |
| Rank correlation     | \texttt{rank}  |
| Least squares        | \texttt{reg}   |
| U-statistics         | \texttt{ustat} |
| Weighted ML          | \texttt{wml}   |
| ML (3 parameters)    | \texttt{mle}   |
| MM Type 1            | \texttt{ml1}   |
| MM Type 2            | \texttt{ml2}   |
| MM Type 3            | \texttt{ml3}   |
| MML type 1           | \texttt{mml1}  |
| MML type 2           | \texttt{mml2}  |
| MML type 3           | \texttt{mml3}  |
| MML type 4           | \texttt{mml4}  |
| Maximum product spacing | \texttt{mps} |
| T-L moment           | \texttt{tlm}   |

The code below calls \texttt{fitWeibull} to fit the three parameter Weibull distribution using maximum product spacing (\texttt{mps}) to the plot 72 data, with the function output printed below.

```
library(ForestFit)
data(DBH)
d <- DBH[DBH[, 1] == 55, 10]
h <- DBH[DBH[, 1] == 55, 11]
d72 <- d[is.na(d)]
h55 <- h[!is.na(h)]

fitWeibull(data, location, method, starts)
```

where \texttt{data} is a vector of data observations, \texttt{location} indicates whether to use a three-parameter Weibull distribution (\texttt{TRUE}) or a two-parameter Weibull distribution (\texttt{FALSE}), \texttt{method} denotes the estimation method to use, and \texttt{starts} is a vector of starting values for the parameters to be estimated. Available estimation methods and their code in ForestFit are shown in Table 4.

The first element of the output list, \texttt{estimate}, contains the point estimates for the model parameters. The second element of the list, \texttt{measures}, contains a series of goodness-of-fit measures including Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (CAIC), Bayesian Information Criterion (BIC), Hannan-Quinn information criterion (HQIC), Anderson-Darling (AD), Cramér-von Misses (CMV), Kolmogorov-Smirnov (KS), and log-likelihood (log.likelihood).

| AIC  | CAIC | BIC  | HQIC |
|------|------|------|------|
| 238.7653 | 239.6542 | 243.0672 | 240.1676 |
| AD    | CMV  | KS   | log.likelihood |
| 0.9055053 | 0.1239423 | 0.1441135 | -116.3826 |

The code below calls \texttt{fitWeibull} to fit the three parameter Weibull distribution using maximum product spacing (\texttt{mps}) to the plot 72 data, with the function output printed below.
These measures allow for comparison between different models and estimation techniques to determine the best fitting model according to the desired criteria.

### 3.2. Bayesian analysis

Bayesian inference for diameter distributions modeled using the JSB and Weibull distributions is provided via the `fitbayesJSB` and `fitbayesWeibull` functions, respectively. The `fitbayesJSB` function has the form

\[
\text{fitbayesJSB}(\text{data}, \text{n.burn} = 8000, \text{n.simul} = 10000)
\]

where `data` is a vector of observations, `n.burn` is an integer representing the number of burn-in iterations (i.e., the number of iterations the Gibbs sampler takes to reach convergence), and `n.simul` is the total number of Gibbs sampler iterations. By default, `fitbayesJSB` uses 10,000 total iterations and 8,000 burn-in iterations. `fitbayesWeibull` has the same arguments. Below we call both functions using the diameters of all live trees in plot 72.

\[
\text{fitbayesJSB}(d72, \text{n.burn}=8000, \text{n.simul}=10000)
\]

\[
\text{estimate}
\begin{array}{cccc}
\delta & \gamma & \lambda & \xi \\
[1,] & 0.6897204 & 0.4077035 & 43.51213 & 11.68084
\end{array}
\]

\[
\text{measures}
\begin{array}{ccc}
\text{AD} & \text{CVM} & \text{KS} \\
[1,] & 0.5059267 & 0.06804388 & 0.1235231
\end{array}
\]

\[
\text{log.likelihood} \quad [1,] \quad -113.4759
\]

\[
\text{fitbayesWeibull}(d72, \text{n.burn}=8000, \text{n.simul}=10000)
\]

\[
\text{estimate}
\begin{array}{ccc}
\alpha & \beta & \mu \\
[1,] & 1.632042 & 21.18533 & 10.28903
\end{array}
\]

\[
\text{measures}
\begin{array}{ccc}
\text{AD} & \text{CVM} & \text{KS} \\
[1,] & 0.7635954 & 0.11028456 & 0.1378961
\end{array}
\]

\[
\text{log.likelihood} \quad [1,] \quad -116.7509
\]

We see that for plot 72, all goodness-of-fit measures indicate the JSB model is a better fit for the data (Figure 1).

### 3.3. Modeling diameter distributions using grouped data

Suppose diameter observations are given in `m` separate groups as shown in Table 2. ForestFit provides the function `fitgrouped` for estimating the parameters of the three-parameter BS and Weibull distributions using grouped data. This function takes the form

\[
\text{fitgrouped}(r, f, \text{family}, \text{method1}, \text{starts}, \text{method2})
\]

where `r` is a length `m + 1` vector of group boundaries as shown in Table 2, with the first element of `r` being the lower bound of the first group and all other `m` values representing the upper bound of the `m` groups. `f` is a vector of length `m` of the group frequencies. `family` represents the distribution used in the model, taking values `‘weibull’` or `‘birnbaum-saunders’`. `method1` is a string denoting the method of estimation. Here we enable three methods of estimation: approximated maximum likelihood (`‘aml’`), EM algorithm (`‘em’`), and maximum likelihood (`‘ml’`). `starts` is a numeric vector containing the starting values for the shape, scale, and location parameters, respectively. Lastly, `method2` indicates the optimization method of the log-likelihood, taking values `‘BFGS’`, `‘CG’`, `‘L-BFGS-B’`, `‘Nelder-Mead’`, and `‘SANN’`. If using the EM algorithm, the `stats` and `method2` arguments do not need to be specified. Below we manually group the diameter observations from plot 57 into six groups, and use the function `fitgrouped` on this grouped data set.

\[
m <- 6
\]

\[
r <- \text{seq}(\text{min}(d57), \text{max}(d57), \text{length}=m+1)
\]

\[
D <- \text{data.frame}(\text{table}(\text{cut}(d57, r)))
\]

\[
\text{fitgrouped}(r = r, f = D$Freq, \text{family} = "\text{birnbaum-saunders}", \text{method1} = "\text{em}"
\]

\[
\text{estimate}
\begin{array}{ccc}
\alpha & \beta & \mu \\
[1,] & 0.6234071 & 8.660411 & 8.453387
\end{array}
\]
Figure 2: Diameter distribution of trees in plot 57 classified into six groups. The estimated pdf of the BS distribution is superimposed.

3.4. Modeling diameter distributions of grouped data using finite mixture models

The `fitmixturegrouped` function provides an interface for fitting a wide-range of finite mixture distributions to grouped data and takes the form

```r
fitmixturegrouped(family = "skew-normal", r = r, f = f, K = K, initial = TRUE, starts = c(omega, alpha, beta, lambda))
```

where \( r \) and \( f \) are vectors of the class boundaries and frequencies as previously described in Section 3.3. \( K \) is a single numeric value denoting the number of component pdfs to include in the finite mixture model. All parameters in finite mixture models are estimated using the EM algorithm. The family argument is a character string denoting the distribution used in the mixture model. Currently available distributions are the gamma (‘gamma’), log-normal (‘log-normal’), skew-normal (‘skew-normal’), and Weibull (‘weibull’) families. The initial argument denotes whether or not the user wants to specify starting values for the EM algorithm. By default the argument is set to FALSE and thus the user does not need to specify a value for the starts argument. If initial=TRUE, the user must specify a sequence of initial values taking the form \( \text{starts} = (\alpha_1, \alpha_2, \beta_1, \beta_2) \), where the three vectors are of length \( K \) with elements \( \alpha_i \) being the weight of the \( i \)-th component, and \( \alpha_i \) and \( \beta_i \) are the associated parameters of the \( i \)-th component. If family = ‘skew-normal’, the user must specify starting values for the weights and three parameters for each component distribution (see Appendix A). In the code that follows, we illustrate the `fitmixturegrouped` function by fitting a two-component skew-normal mixture model to plot 51’s grouped data.

```r
m <- 8; K <- 2
r <- seq(min(d51), max(d51), length=m+1)
D <- data.frame(table(cut(d51, r)))
f <- D$Freq
omega <- c(0.5, 0.5); alpha <- c(10, 40)
beta <- c(2, 2); lambda <- c(2, -2)
fitmixturegrouped(family = "skew-normal", r = r, f = f, K = K, initial = TRUE, starts = c(omega, alpha, beta, lambda))
```

Figure 2 displays the estimated BS pdf to plot 57’s diameter distribution data.

Similar to previous functions using models with a single pdf, the output is a list where the first element (i.e., `estimate`) contains the estimated parameters of the first and second components of the model, and the second element (i.e., `measures`) contains the goodness-of-fit measures. Figure 3 displays the estimated two-component skew-normal pdf.

3.5. Modeling diameter distributions using finite mixture models fitted to ungrouped data

As a companion to the grouped data `fitmixturegrouped` function described in Section 3.4, the `fitmixture` function fits finite mixture models to individual diameter distribution data, and takes the form

```r
fitmixture(data, family, K, initial = FALSE, starts)
```

where `data` is a vector of individual diameter measurements and `family` denotes the family of probability distributions to use. Available families of pdfs are the BS (‘birnbaum-saunders’), Burr XI (‘burrxi’), Chen (‘chen’), Mahdi Teimouri, Jeffrey W. Doser, and Andrew O. Finley: Preprint submitted to Elsevier
Figure 3: Histogram of grouped diameter distribution data in plot 51. Superimposed is the estimated pdf of the two-component skew-normal mixture model.

Figure 4: Diameter distribution of plot 51. Superimposed is the estimated pdf of the two-component log-normal mixture model.

Fisher (’f’), Fréchet (’Frechet’), gamma (’gamma’), generalized exponential (’ge’), Gompertz (’gompertz’), log-normal (’log-normal’), log-logistic (’log-logistic’), Lomax (’lomax’), skew-normal (’skew-normal’), and Weibull (’weibull’) families. The remaining arguments are analogous to those used in the fitmixturegrouped function. Below, the fitmixture function is called to fit a two-component log-normal mixture model to plot 51’s diameter data.

```r
fitmixture(d51, "log-normal", 2)
```

$estimate

weight alpha beta
[1,] 0.6522847 2.618974 0.2998769
[2,] 0.3477153 3.668380 0.1461719

$measures

AIC CAIC BIC HQIC
[1,] 419.7617 420.9381 429.9769 423.7317
[2,] 5.790067 1.107157 0.349687 0.204.8808

$cluster

[1] 1 1 1 1 2 2 2 2 2 2 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1
[2] 1 1 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[4] 2 1 1 1 2 2 2 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

Interpretation of the first and second components of the output list is analogous to the interpretation provided in Section 3.4. The third list element (cluster) specifies the component of the mixture distribution from which each individual diameter measurement arises. Figure 4 displays the fitted two-component log-normal pdf to the diameter distribution data of plot 51.

3.6. Distribution functions for finite mixture models

Analogous with R’s standard functions for working with common probability distributions, we provide commands for computing the density, distribution function, quantile function, and random generation from the finite mixture models used in ForestFit. The functions take the following forms

```r
dmixture(x, family, K, param)
pmixture(x, family, K, param)
qmixture(p, family, K, param)
rmixture(n, family, K, param)
```

where x is a vector of observations, n is the number of realizations to be simulated from the mixture model, p is a numeric value that satisfies 0 < p < 1, family is one of the families introduced in Section 2.4, K is the number of components, and param is the mixture model parameter vector that takes the same form as the starts argument of the fitmixturegrouped function described previously in Section 3.4. As an illustration, the code below generates 500 realizations from a three-component BS mixture model and displays the resulting distribution in Figure 5:

```r
n <- 500; K <- 3
weight <- c(0.4, 0.3, 0.3)
alpha <- c(0.1, 0.25, 0.5)
beta <- c(0.5, 2)
```
3.7. The gamma shape mixture (GSM) model

The gamma shape mixture model (GSM) introduced by [33] has received considerable attention in forestry for characterizing diameter distributions. Because of its popularity, we have implemented specific functions in ForestFit for estimating the parameters of the GSM distribution, computing the density function, computing the distribution function, and generating realizations for the GSM distribution. The function calls for the aforementioned tasks are

\[
\begin{align*}
\text{param} & \leftarrow c(\text{weight, alpha, beta}) \\
X & \leftarrow \text{rmixture}(n, \text{"birnbaum-saunders"}, K, \text{param}) \\
XX & \leftarrow \text{seq}(0, \text{max}(X), 0.01) \\
pdf & \leftarrow \text{dmixture}(XX, \text{"birnbaum-saunders"}, K, \text{param}) \\
\text{hist}(X, \text{freq=FALSE, breaks=120, col='gray', las = 1, main = '"}, \text{xlim = c(0, 11))} \\
\text{lines}(XX, pdf, \text{col='blue'})
\end{align*}
\]

Figure 5: Histogram of 500 random realizations from a three-component BS mixture model.

\[
\begin{align*}
\text{n} & \leftarrow 500 \\
K & \leftarrow 10 \\
\text{omega} & \leftarrow \text{rep}(0.1, 10) \\
\text{beta} & \leftarrow 0.25 \\
X & \leftarrow \text{rgsm}(n, \text{omega}, \text{beta}) \\
XX & \leftarrow \text{seq}(0, \text{max}(X), 0.01) \\
\text{out} & \leftarrow \text{fitgsm}(X, K) \\
\text{pdf} & \leftarrow \text{dgsm}(XX, \text{out}$omega, \text{out}$beta) \\
\text{hist}(X, \text{freq = FALSE, breaks = 30, col = 'gray'}, \text{main = '"}, \text{xlim = c(0, 80))} \\
\text{lines}(XX, pdf, \text{col = 'blue'})
\end{align*}
\]

Figure 6: Histogram of 500 realizations from a ten-component GSM model.

\[
\begin{align*}
\text{out} \\
\text{beta} & \left[1\right] 0.2071086 \\
\text{omega} & \left[1\right] 5.553682e-02 2.005843e-01 1.948963e-01 \\
\text{beta} & \left[4\right] 9.702007e-03 2.543254e-02 1.802041e-01 \\
\text{beta} & \left[7\right] 3.084612e-01 2.515575e-02 2.694728e-05 \\
\text{beta} & \left[10\right] 8.429742e-11 \\
\text{measures} \\
\text{AIC CAIC BIC HQIC} & \left[1\right] 4061.953 4065.655 4184.176 4109.913 \\
\text{AD CVM KS} & \left[1\right] 4061.953 4065.655 4184.176 4109.913
\end{align*}
\]

whether to compute the upper tail (lower.tail = FALSE) or lower tail of the distribution (lower.tail = TRUE). As an illustration, we generate 500 realizations from a ten-component GSM model and display the resulting distribution in Figure 6.

where data is a vector of observations, n is the number of realizations simulated from the GSM model, k is the number of components, omega is a vector of mixing parameters, beta is the rate parameter, log indicates whether to compute the density function (log = FALSE) or the log-density function (log = TRUE), log.p indicates whether to compute the the distribution function (log.p = FALSE) or the log-transformed distribution function (log.p = TRUE), and lower.tail indicates
3.8. Modeling height-diameter relationships

The general function for fitting a growth curve to paired observations of height and diameter is

```
fitgrowth(h, d, model, starts)
```

where `h` and `d` are vectors of height and diameter observations, respectively. The `model` argument denotes the model to fit with options: Chapman-Richards (‘chapman-richards’), Gompertz (‘gompertz’), Hossfeld IV (‘hossfeldiv’), Korf (‘korf’), logistic (‘logistic’), Prodan (‘prodan’), Ratkowsky (‘ratkowsky’), Sibbesen (‘sibbesen’), and Weibull (‘weibull’). Here we fit a Weibull growth curve to the height-diameter relationship of plot 55.

```
starts <- c(18, 0.0005, 2)
fıtgrowth(h55, d55, “weibull”, starts=starts)
```

The output list consists of five parts:

1. summary: estimated parameters, standard error of the estimators, t-value and corresponding p-value.
2. residuals: residuals of the fitted growth curve.
3. ‘var-cov’: variance-covariance matrix of the model coefficient estimators.
4. ‘residual Std. Error’: residual standard error.
5. A scatterplot of the height-diameter relationship with the fitted growth curve (Figure 7)

4. Future development

We are currently working on software functionality that includes: 1) a broader set of goodness-of-fit measures for...
the Bayesian models; 2) a function that automates “optimal” model/method selection that searches among all methods in Table 4 with selection based on a set of user identified goodness-of-fit measures; 3) expanded set of mixture models, e.g., [33]; 4) addition of random effects that accommodate grouped distribution and height-diameter data, e.g., stands, species, functional types, etc.

5. Conclusion

A common task in forestry is modeling the diameter distribution of trees in a given forest stand. In this work, we have developed and introduced an R package called ForestFit that provides numerous models and estimation techniques for modeling diameter distributions from both individual tree data and grouped data, as well as providing additional functionality for simulating finite mixture distributions and fitting growth curves to height-diameter data. While this package was developed in the context of forestry, the models we fit and simulate have numerous applications throughout other ecological and environmental fields, such as botany, fisheries, and hydrology. For example, the JSB distribution is used in hydrology to model rain drop size distribution [5] and finite mixture distributions are broadly applied in environmental and ecological statistics (for numerous examples, see [13]), suggesting ForestFit can have wide use across ecological and environmental fields. ForestFit is freely available on CRAN at https://cran.r-project.org/web/packages/ForestFit/index.html.

A. Probability density functions used in ForestFit

- **log-normal** $f(x, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right)$,
- **Lomax** $f(x, \theta) = a(1/1+x)^{\alpha+1}$,
- **skew-normal** $f(x, \theta) = 2\phi\left(\frac{x-a}{\beta}\right)\Phi\left(\frac{\lambda-x\cdot a}{\beta}\right)$,
- **Weibull** $f(x, \theta) = \frac{a}{\beta} \left(1 + \frac{x-a}{\beta}\right)^{-\alpha-1} \exp\left(-\left(\frac{x-a}{\beta}\right)^{\alpha}\right)$,

where $\theta$ is the family parameter vector; $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density function and distribution function of the standard normal distribution, respectively. Notably, the three-parameter BS, gamma, GE, and Weibull distributions simplify to the two-parameter BS, gamma, GE, and Weibull distributions, respectively, when $\mu = 0$.

References

[1] Bailey, R.L., Dell, T. 1973. Quantifying diameter distributions with the weibull function. Forest Science 19, 97–104. doi:10.1093/forestscience/19.2.97.
[2] Cohen, C.A., Whitten, B., 1982. Modified maximum likelihood and modified moment estimators for the three-parameter weibull distribution. Communications in Statistics-Theory and Methods 11, 2631–2656. doi:10.1080/03610928208828412.
[3] Cousineau, D., 2009. Nearly unbiased estimators for the three-parameter weibull distribution with greater efficiency than the iterative likelihood method. British Journal of Mathematical and Statistical Psychology 62, 167–191. doi:10.1348/000711007X270843.
[4] Cran, G., 1988. Moment estimators for the 3-parameter weibull distribution. IEEE Transactions on Reliability 37, 360–363. doi:10.1109/24.9839.
[5] Cugenero, K., De Michele, C., 2015. Johnson sb as general functional form for raindrop size distribution. Water Resources Research 51, 6276–6289. doi:10.1002/2014WR016484.
[6] Flewelling, J.W., Jong, R.D., 1994. Considerations in simultaneous curve fitting for repeated height–diameter measurements. Canadian Journal of Forest Research 24, 1408–1414. doi:10.1139/x94-181.
[7] Fonseca, T.F., Marques, C.P., Parresol, B.R., 2009. Describing maritime pine diameter distributions with johnson’s sb distribution using a new all-parameter recovery approach. Forest Science 55, 367–373. doi:10.1093/forestscience/55.4.367.
[8] George, F., Ramachandran, K., 2011. Estimation of parameters of johnson’s system of distributions. Journal of Modern Applied Statistical Methods 10, 9. doi:10.22237/jmasm/1320120480.
[9] Gorgoso, J., González, J.A., Rojo, A., Grandas-Arias, J., 2007. Modelling diameter distributions of betula alba l. stands in northwestern spain with the two-parameter weibull function. Forest Systems 16, 113–123. doi:10.5424/forestscience/2007162-01002.
[10] Gorgoso-Varela, J.J., Rojo-Alboreca, A., 2014. Use of gumbel and weibull functions to model extreme values of diameter distributions in forest stands. Annals of forest science 71, 741–750. doi:10.1007/s13595-014-0369-1.
[11] Gove, J.H., Fairweather, S.E. 1989. Maximum-likelihood estimation of weibull function parameters using a general interactive optimizer and grouped data. Forest Ecology and Management 28, 61–69. doi:10.1016/S0378-1127(89)80074-1.
[12] Green, E.J., Roesch Jr, F.A., Smith, A.F., Strawderman, W.E., 1994. Bayesian estimation for the three-parameter weibull distribution with tree diameter data. Biometrics , 254–269 doi:10.2307/2531217.
[13] Hoskins, J.R.M., 1990. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. Journal of the
Liu, F., Li, F., Zhang, L., Jin, X., 2014. Modeling diameter distribution of mixed-species forest stands. Scandinavian Journal of Forest Research 29, 653–661. doi:10.1080/02827581.2014.960891.

Teimouri, M., Naderajah, S., 2012. A simple estimator for the Weibull frequency distribution function and percentile-based estimation methods for the Weibull distribution. Statistics 47, 93–109. doi:10.1080/02331882.2011.559657.

Zhang, L., Xie, M., Tang, L., 2008. On weighted least squares estimation for the parameters of Weibull distribution, in: Recent Advances in Reliability and Quality in Design. Springer, pp. 57–84. doi:10.1007/978-1-84800-113-8_3.