

# Characterizing wood fiber and particle length with a mixture distribution and a segmented distribution

Quang V. Cao\* and Qinglin Wu

School of Renewable Natural Resources, Louisiana State University Agricultural Center, Baton Rouge, LA, USA

\*Corresponding author.

School of Renewable Natural Resources, Louisiana State University Agricultural Center, Baton Rouge, LA 70803, USA  
Phone: +1-225-5784218, Fax: +1-225-5784227  
E-mail: qciao@lsu.edu

## Abstract

The length data from 12 samples of wood fibers and particles were described using lognormal and Weibull distributions. While both distributions fitted the middle range of the data well, the lognormal distribution provided a closer fit for short fibers and particles and the Weibull distribution was more appropriate for long ones. A mixture of the lognormal and Weibull distributions was developed using a variable weight to allow the new distribution to take the lognormal form for short fibers and gradually change to the Weibull form for long fibers. In the segmented distribution approach, a left segment of the lognormal distribution was joined to a right segment from the Weibull form. The Anderson-Darling goodness-of-fit test at the 5% level failed to reject the hypothesis that the mixture distribution and the segmented distribution fitted the data. Q-Q plots showed that both the mixture and segmented distributions provided an excellent fit to the fiber and particle length data, combining the best features of the lognormal and the Weibull distributions. These two new distributions are therefore better alternatives than the single lognormal and Weibull distributions for this data set.

**Keywords:** goodness-of-fit tests; lognormal; maximum likelihood estimation; Q-Q plots; Weibull.

## Introduction

The dimension and morphology of wood fibers and particles have an important effect on the physical and mechanical properties of wood-based products such as paper, paper board, insulation board, medium-density fiberboard (MDF), particleboard, hardboard, and wood-fiber polymer composites (Takahashi et al. 1979; Mark and Gillis 1983; Eckert et al. 1997; Lee et al. 2001; Huber et al. 2003). Significant variability in fiber length exists for wood fibers from the same tree, from different tree species, and from different manufacturing processes. For example, fibers from softwoods are normally longer than those from hardwoods. In the same tree, the length of individual fibers may vary, depending on many factors, such as distance from the ground, distance from the pith,

early wood or late wood, heartwood or sapwood, etc. (Mark and Gillis 1983). Moreover, composite manufacturing processes (e.g., MDF versus particleboard) can produce fibers of various lengths, depending on the processing conditions and wood species. Accurate description of the length distribution for wood fibers and particles can lead to improved quality control measures during product manufacturing and better prediction of the final product properties (Lu et al. 2007).

Probability density functions have been used to characterize the distributions of fiber length, starting with the normal distribution (Tasman 1972). Subsequent distributions included the lognormal distribution (Yan 1975; Dodson 1992; Kropholler and Sampson 2001), and the Erlang family of distributions (Mark and Gillis 1983). Recently, Lu et al. (2007) discovered that both the lognormal and Weibull distributions fit the middle range of fiber length data well; the lognormal provided a closer fit for short fibers, whereas the Weibull was more appropriate for long fibers.

Since a single probability density function cannot adequately describe the fiber length distribution over its entire range, the next logical step should be to consider a mixture of the lognormal and Weibull distributions for modeling the length of wood fibers and particles. Mixtures of distributions have generally been employed to model the distributions of populations that consist of more than one component. For example, wood samples from increment cores contain fibers and other cells, called “fines.” A mixture of two truncated normal distributions (Mörlling et al. 2003) or two lognormal distributions (Svensson et al. 2006) has been used to characterize the fine and fiber length distributions. Some other examples of the distributions used in a mixture include the normal distribution in bioclimatic modeling (Gavin and Hu 2005) or in classifying ecological habitats (Zhang et al. 2004), Poisson or binomial distributions in estimating species richness (Mao and Colwell 2005), and Weibull distributions in modeling tree diameter (Liu et al. 2002).

An alternative approach to form a flexible, continuous distribution is to join multiple segments of distributions together. Different types of segmented distributions have been successfully applied to describe diameter distributions of forest trees (Cao and Burkhart 1984; Borders et al. 1987; Borders and Patterson 1990).

The objective of this study was to derive a mixture distribution and a segmented distribution such that the resulting distributions fitted both short and long ends of the fiber and particle data.

## Data

Data for this study came from fibers and particles obtained from eight different sources (i.e., manufactur-

**Table 1** Types of sample source and range of fiber length measurements.

Package designation	Source	Type	Usage in wood composites	Length range (mm)
A1	A	Fiber	MDF core	0.19–5.29
A2	A	Fiber	MDF face	0.15–6.44
B1	B	Fiber	MDF core	0.12–5.65
B2	B	Fiber	MDF face	0.02–5.52
B3	B	Fiber	MDF face/core	0.14–4.02
C1	C	Fiber	MDF face/core	0.11–4.57
D1	D	Fiber	MDF face/core	0.16–5.36
D2	D	Fiber	MDF face/core	0.17–5.95
E1	E	Fiber	MDF face/core	0.18–4.80
F1	F	Particle	Particleboard core	0.89–15.95
G1	G	Particle	Particleboard core	0.54–17.94
H1	H	Particle	Flakeboard core	0.89–16.05

MDF, medium-density fiberboard.

ers). Each source provided from one to three 100-g packages for a total of 12 packages, labeled A1–H1 (Table 1). Among the samples, nine packages were MDF fibers, two were core particles for particleboard, and one was core particles for flakeboard. These fibers and particles were produced by thermomechanical grinding (MDF process), hammer milling (particleboard process), and flaking (flakeboard process). Fibers or particles from each package were oven-dried at 80°C for 24 h and sealed in plastic bags to keep the moisture content between 3% and 5%.

A random sample of 500 fibers or particles from each package was measured using an imaging system that consisted of a Leica MZFIII microscope (Leica Microsystems GmbH, Wetzlar, Germany), a CCD digital camera (Diagnostic Instruments, Sterling Heights, MI, USA), a v-Lux 1000 optical lighter (Volpi MFG. USH Co., Auburn, NY, USA), an RT SP402-115 power supply (Diagnostic Instruments, Sterling Heights, MI, USA), and a computer. Spot Advance imaging software (Diagnostic Instruments) was used to record images, and Image-Pro Plus software (Media Cybernetics, Inc., Silver Spring, MD, USA) was utilized to measure individual fiber lengths in millimeters (Lu et al. 2007).

### Mixture of the lognormal and Weibull distributions

Let  $x$  be the fiber or particle length in mm, and  $f_1(x)$  be the lognormal probability density function (pdf) with parameters  $\mu$  and  $\sigma$ :

$$f_1(x) = \left( \frac{1}{x\sigma\sqrt{2\pi}} \right) \exp \left[ -\frac{1}{2} \left( \frac{\ln(x) - \mu}{\sigma} \right)^2 \right], x \geq 0. \quad (1)$$

The corresponding cumulative distribution function (cdf) is:

$$F_1(x) = \Phi \left( \frac{\ln(x) - \mu}{\sigma} \right), \quad (2)$$

where  $\Phi(z)$  is the cdf of the standard normal distribution.

The Weibull distribution with two parameters  $b$  and  $c$  has the following pdf:

$$f_2(x) = \left( \frac{cx^{c-1}}{b^c} \right) \exp \left[ -\left( \frac{x}{b} \right)^c \right], x \geq 0, \quad (3)$$

and its corresponding cdf:

$$F_2(x) = 1 - \exp \left[ -\left( \frac{x}{b} \right)^c \right]. \quad (4)$$

Figure 1 presents graphs of the lognormal and Weibull distributions that fit the observed data for package H1.

A mixture of the lognormal and Weibull distributions is defined in this study as having the following cdf:

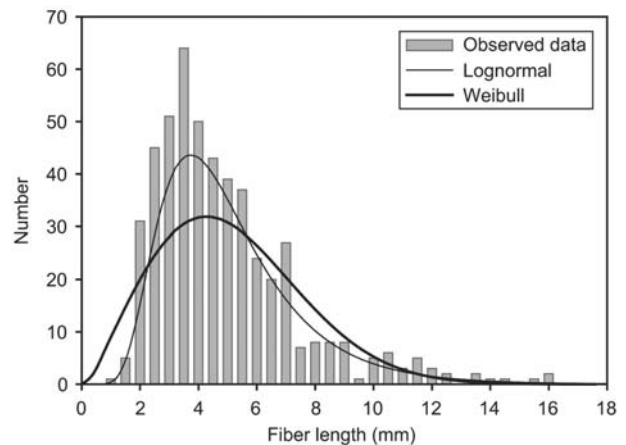
$$F(x) = [1 - w(x)]F_1(x) + w(x)F_2(x), \quad (5)$$

where  $0 \leq w(x) \leq 1$  is a weight function.

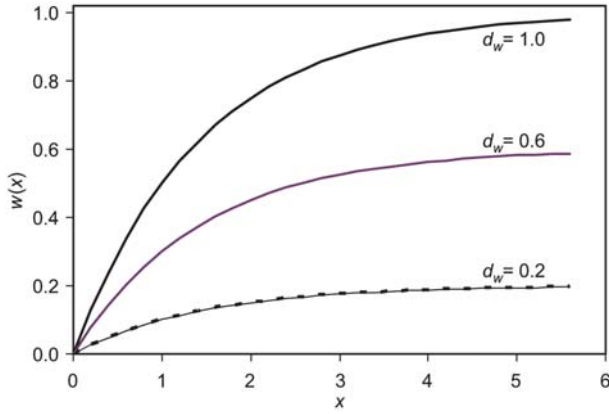
If  $w(x) = 0$  when  $x = 0$ , the mixture distribution resembles the lognormal distribution for short fibers or particles. For long ones,  $w(x)$  approaches  $d_w$  (where  $0 \leq d_w \leq 1$ ), i.e., the mixture distribution is a weighted average of the lognormal and Weibull distributions. For the new distribution to gradually change from the lognormal to the Weibull form,  $w(x)$  has to gradually change its value from 0 to  $d_w$ . We selected the following function, modified from a simple exponential cdf, to describe  $w(x)$ :

$$w(x) = d_w [1 - \exp(-b_w x)]. \quad (6)$$

We speculate that the sample median of  $x$ , or  $x_{\text{med}}$ , is where the change from the lognormal to the Weibull form is half complete. This results in:



**Figure 1** Graphs of the lognormal and Weibull distributions that fit the observed data for package H1.



**Figure 2** Graphs of  $w(x)$  versus  $x$  for different values of  $d_w$ , where  $w(x) = d_w[1 - \exp(-b_w x)]$ ,  $b_w = [\ln(2)]/x_{med}$ , and  $x_{med}$  is the median of  $x$ .

$$w(x_{med}) = d_w/2. \quad (7)$$

Solving Eq. (7) yields:

$$b_w = [\ln(2)]/x_{med}. \quad (8)$$

Figure 2 shows the shapes of  $w(x)$  for different values of  $d_w$ .

The pdf of the mixture distribution can be obtained by taking the derivative of the mixture cdf function [Eq. (5)] with respect to  $x$ :

$$f(x) = [1 - w(x)]f_1(x) + w(x)f_2(x) + w'(x)[F_2(x) - F_1(x)], \quad (9)$$

where  $w'(x) = b_w d_w \exp(-b_w x)$ . Care must be taken to ensure that  $f(x) \geq 0$  for all  $x$ .

The mixture of two distributions has a total of five parameters,  $\mu$ ,  $\sigma$ ,  $b$ ,  $c$ , and  $d_w$ , which can be estimated for each package by maximizing the following log-likelihood function:

$$\ln(L) = \sum_{i=1}^n \ln[f(x_i)], \quad (10)$$

where:  $x_i$  is the length in mm of the  $i$ th fiber or particle,  $i = 1, 2, \dots, n$ ;  $n = 500$  is the sample size of each package; and  $\ln(\cdot)$  is the natural logarithm.

The parameters were estimated using SAS procedure NLIN (SAS Institute 2004), which employs an algorithm similar to that used in the SAS program published by Cao (2004).

### Segmented distribution

A segmented distribution is obtained by joining a left segment from the lognormal [ $f_1(x)$ ] and a right segment from the Weibull [ $f_2(x)$ ] together at a join point  $t$ . It is reasonable to assume that the change from the lognormal to the Weibull form takes place at the sample median length, i.e.,  $t$  is fixed at  $x_{med}$ . The segmented distribution is defined here as having the following pdf:

$$f(x) = \begin{cases} f_1(x)/\beta, & 0 \leq x \leq t \\ \alpha f_2(x)/\beta, & x > t \end{cases}, \quad (11)$$

where  $\alpha = f_1(t)/f_2(t)$ , so that  $f(x)$  is continuous at the join point  $t$ , and  $\beta = \int_0^t f_1(x)dx + \int_t^\infty \alpha f_2(x)dx = F_1(t) + \alpha[1 - F_2(t)]$ ;  $\beta$  is used to scale  $f(x)$  such that  $f(x)$  integrates to 1.

The corresponding cdf is then:

$$F(x) = \begin{cases} F_1(x)/\beta, & 0 \leq x \leq t \\ \{F_1(t) + \alpha[F_2(x) - F_2(t)]\}/\beta, & x > t \end{cases}. \quad (12)$$

The four parameters ( $\mu$ ,  $\sigma$ ,  $b$ , and  $c$ ) of the segmented distribution were estimated for each package by maximizing the following log-likelihood function:

$$\ln(L) = \sum_{x_i \leq t} \ln\left(\frac{f_1(x_i)}{\beta}\right) + \sum_{x_i > t} \ln\left(\frac{\alpha f_2(x_i)}{\beta}\right). \quad (13)$$

SAS procedure NLIN (SAS Institute 2004) was also used to estimate the above parameters.

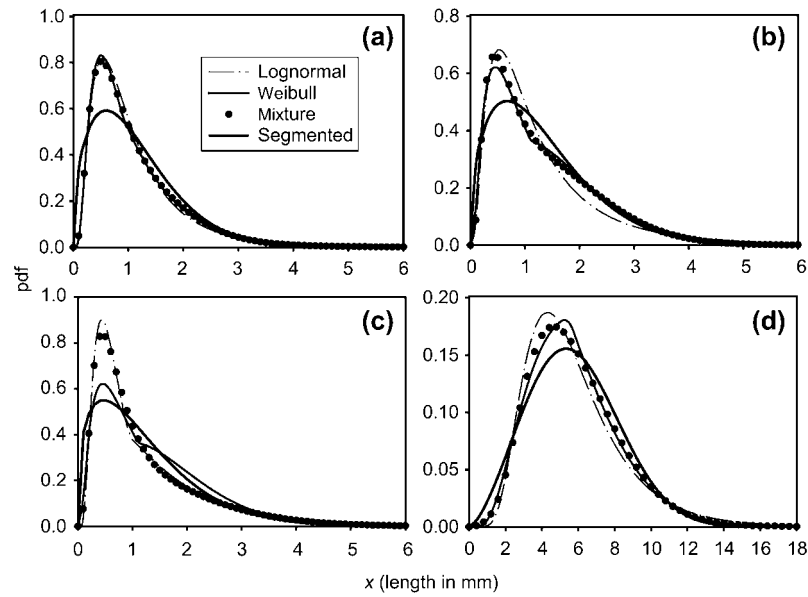
## Results and discussion

The estimate of  $d_w$  for the mixture distribution dictates the shape of the right tail. For five out of 12 packages,  $d_w = 1$ , indicating that the right tail of the mixture resembled that of the Weibull distribution. For the remaining seven packages, the right tail of the mixture was located between the right tails of the lognormal and Weibull distributions ( $d_w < 1$ ). Figure 3 presents the graphs of the pdfs of the lognormal, Weibull, mixture, segmented and distributions for four packages.

The Anderson-Darling (AD) goodness-of-fit statistics (Anderson and Darling 1954) are presented in Table 2 for all four distributions. The hypothesis that the distribution fitted the fiber data was never rejected for the mixture and segmented distributions, but was rejected for the lognormal form in two cases (out of 12) and for the Weibull form in six cases. The AD statistics computed from the mixture and segmented distributions were reduced by 46–92% compared to the lognormal and by 54–97% compared to the Weibull distribution. Both the mixture and segmented distributions therefore provide a much better fit to the length data of fibers and particles than did either of the single distributions.

### Weight function used in the mixture distribution

Most mixtures of statistical distributions are based on constant weight values. For example, if a mixture distribution is used to model the diameter distribution of a mixed forest stand consisting of two species groups, the weights  $w$  and  $(1-w)$  are often indicative of the relative abundance of the two groups. The mixture distribution proposed in this study was used to model a population comprised of only one component. The need for the mixture of two distributions was justified because a single



**Figure 3** Graphs of the probability density functions (pdf) of the lognormal, Weibull, mixture, and segmented distributions for (a) package A2, (b) package B2, (c) package D2, and (d) package F1.

distribution (lognormal or Weibull) could not adequately fit the entire range of the fiber data. Another new procedure in this approach is the inclusion of a variable weight,  $w(x)$ , which was allowed to vary gradually from 0 to  $d_w$ . In the worst case scenario, having variable weights theoretically could possibly result in a decreasing  $F(x)$  and consequently a negative  $f(x)$  for some values of  $x$ . We did not experience this problem, however, probably because  $F_1(x)$ ,  $F_2(x)$ , and  $w(x)$  are all well-behaved functions.

### Quantile-quantile plots

Quantile-quantile (Q-Q) plots were used in this study to examine the goodness-of-fit of a distribution. In this type of plot, the observed quantile (fiber or particle length) is graphed against the predicted quantile, which is the fiber or particle length having the same cumulative probability

as predicted from the distribution. A distribution that fits the data perfectly is represented by a 1:1 straight line.

For all packages, the lognormal distribution fitted short fibers and particles better than the Weibull form, based on the Q-Q plot. The curve for the Weibull distribution tends to fall below the straight line, especially near zero (Figures 4–6). This fact was also noted by Lu et al. (2007).

The right tail of the lognormal distribution is generally heavier than that for the observed distribution; the lognormal form in the Q-Q plot for long fibers and particles is represented by a curve pointing upward. On the other hand, the curve for the Weibull distribution remains on the straight line in some cases for long fibers and particles (Figure 4), as do the curves for the mixture distribution ( $d_w=1$  in this case) and the segmented distribution. In other cases, the curve for the Weibull form points downward as the length of fibers or particles increases (Figure 5). The mixture distribution is then a weighted average between the lognormal and the Weibull distributions ( $d_w < 1$ ) for long fibers. The curve for the segmented distribution is similar to that for the mixture distribution.

Package D2 produced the highest AD values for the mixture and segmented distributions. Figure 6 shows that the curves for these two distributions center around the 1:1 line, indicating that both distributions are appropriate for characterizing the fiber length distribution for this package. The Q-Q plots demonstrate that they provide valuable information on whether the statistical function fits the data, and should be used in conjunction with goodness-of-fit tests.

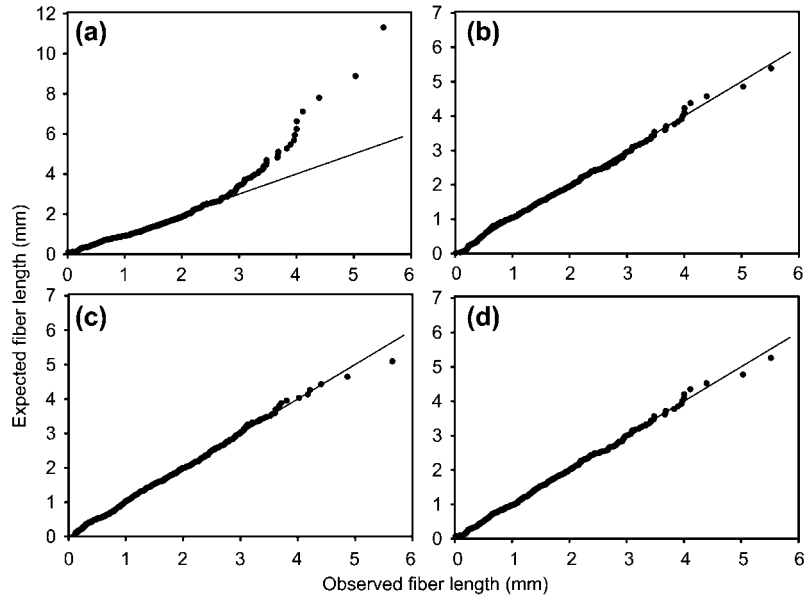
### Mixture versus segmented distribution

Mixture distributions are typically used to describe two or more types of observations. In this paper, however, a new approach is presented that uses a mixture distribution for modeling data from one type of observation (fiber or particle length). The mixture distribution

**Table 2** Anderson-Darling goodness-of-fit statistics by package for the lognormal, Weibull, mixture, and segmented distributions.

Package designation	Distribution			
	Lognormal	Weibull	Mixture	Segmented
A1	0.94	1.71	0.25	<b>0.16</b>
A2	0.54	3.47*	<b>0.20</b>	0.25
B1	5.71*	1.71	<b>0.65</b>	0.79
B2	2.86*	1.85	0.38	<b>0.38</b>
B3	1.81	2.57*	<b>0.50</b>	0.64
C1	1.13	5.43*	<b>0.48</b>	0.82
D1	2.18	6.53*	0.58	<b>0.48</b>
D2	2.06	5.12*	1.11	<b>0.91</b>
E1	1.99	1.98	<b>0.58</b>	0.60
F1	1.42	1.94	0.23	<b>0.19</b>
G1	1.92	1.58	<b>0.16</b>	0.25
H1	0.72	9.32*	0.26	<b>0.23</b>

Values in bold denote the smallest statistics among the four methods for each package. \*Significant at the 5% level.



**Figure 4** Quantile-quantile plots for package B2 for (a) lognormal, (b) Weibull, (c) mixture, and (d) segmented distributions. The x-axis is the observed fiber length, and the y-axis is the fiber length with the same cumulative probability as predicted from the distribution. The right tail of the mixture resembles that of the Weibull distribution ( $d_w = 1$ ).

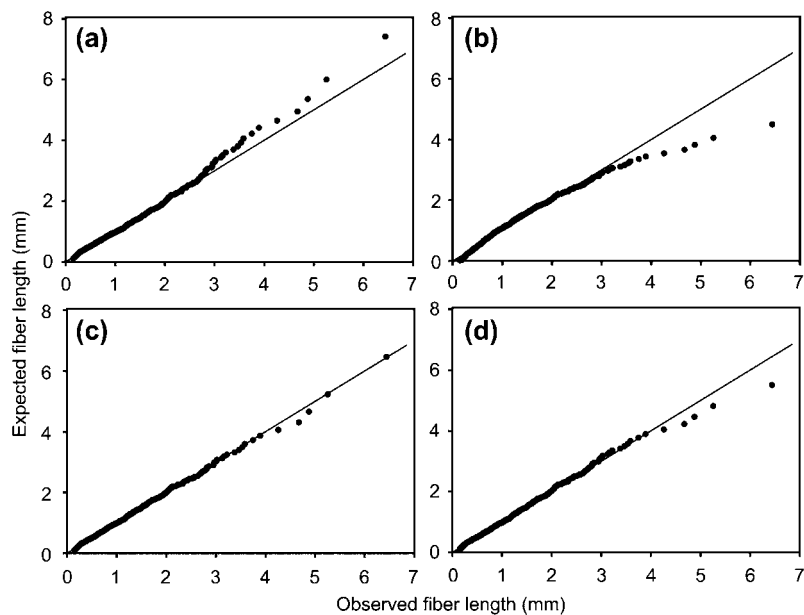
approach is applicable to problems involving characterization of a sample for which the underlying distribution might be too complicated to be approximated by a single statistical distribution such as the lognormal or Weibull form. Furthermore, use of a variable weight (or weight function) to replace the traditional constant weight should be worthy of investigation for use in solving other similar problems.

Both the mixture and segmented distributions delivered comparable performance in terms of AD statistics (Table 2). The mixture distribution was better for one half of the packages, whereas the segmented distribution was better for the other half. The average AD value was 0.45 for the mixture distribution and 0.48 for the seg-

mented distribution. The pdf shapes were similar (Figure 3), as were the Q-Q plots (Figures 4–6). The difference between the two distributions lies in the number of parameters to be estimated: five for the mixture distribution and four for the segmented distribution. In addition, the derivation for the segmented distribution seems more straightforward compared to that for the mixture distribution.

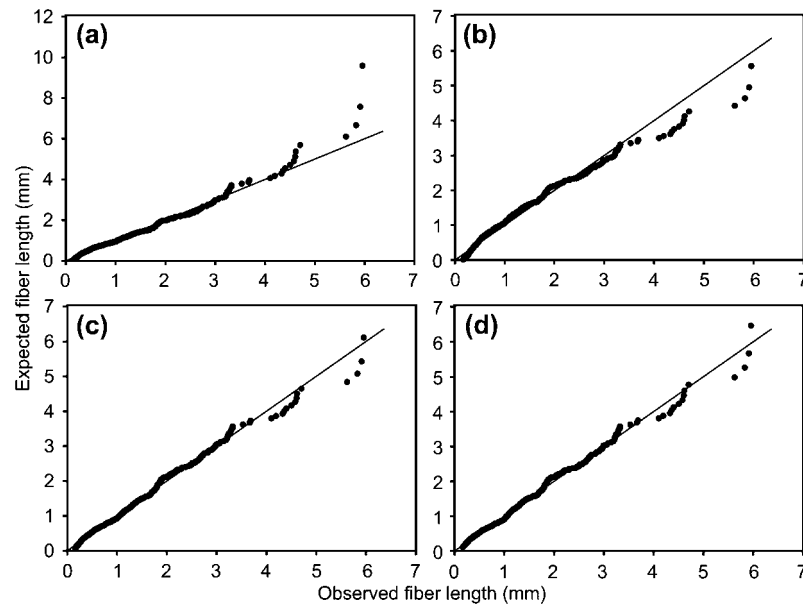
#### Adding an extra parameter

The sample median,  $x_{\text{med}}$ , was used in this study to fix the values for  $b_w$  of the mixture distribution and  $t$  of the segmented distribution. In addition, we also considered



**Figure 5** Quantile-quantile plots for package A2 for (a) lognormal, (b) Weibull, (c) mixture, and (d) segmented distributions. The right tail of the mixture is located between the right tails of the lognormal and Weibull distributions ( $d_w = 0.76977$ ).





**Figure 6** Quantile-quantile plots for package D2 for (a) lognormal, (b) Weibull, (c) mixture, and (d) segmented distributions. The distribution in (c) resulted in the highest Anderson-Darling goodness-of-fit statistic (1.11) among the mixture distributions. Similarly, the distribution in (d) resulted in the highest Anderson Darling statistics (0.91) among the segmented distributions.

$b_w$  as an extra parameter to be estimated at the same time as the other five parameters of the mixture distribution using the maximum likelihood method. Similarly, the number of parameters for the segmented distribution was increased from four to five by considering the join point  $t$  as a parameter to be estimated. For the segmented distribution, this action marginally reduced the average Anderson-Darling statistic from 0.48 to 0.43, while the resulting Q-Q plots showed no detectable improvements. For the mixture distribution, on the other hand, difficulties in obtaining the MLE for all six parameters were encountered for some packages, indicating that the new distribution might suffer from over-parameterization. Overall, the simpler versions of the mixture distribution (with five parameters) and the segmented distribution (with four parameters) seem appropriate for this data set.

## Conclusions

A mixture distribution and a segmented distribution were successfully used to describe the length distribution of wood fibers and particles. The main difference between this mixture distribution and others used in the past is the variable weight employed in this study. The weight function allowed the mixture distribution to take up the lognormal form for short fibers/particles and gradually change to the Weibull form for long fibers/particles. The segmented distribution, on the other hand, switched abruptly at the sample median length from the lognormal to the Weibull function. Q-Q plots showed that both the mixture and segmented distributions provided an excellent fit to the length data for fibers and particles, combining the best features of the lognormal and the Weibull distributions. These two distributions are therefore better alternatives to the single lognormal and Weibull distributions for this data set.

## Acknowledgements

The authors wish to thank Dr. George Woodson, Weyerhaeuser Company for helping in collection of all the MDF fibers used in this study. This work was partially supported by the Louisiana Governor's Biotechnology Initiative Grant Program (No. 101-40-0120).

## References

- Anderson, T.W., Darling, D.A. (1954) A test of goodness of fit. *J. Am. Stat. Assoc.* 49:765–769.
- Borders, B.E., Patterson, W.D. (1990) Projecting stand tables – a comparison of the Weibull diameter distribution method, a percentile-based projection method, and a basal area growth projection method. *For. Sci.* 36:413–424.
- Borders, B.E., Souter, R.A., Bailey, R.L. Ware, K.D. (1987) Percentile-based distributions characterize forest stand tables. *For. Sci.* 33:570–576.
- Cao, Q.V. (2004) Predicting parameters of a Weibull function for modeling diameter distribution. *For. Sci.* 50:682–685.
- Cao, Q.V., Burkhart, H.E. (1984) A segmented distribution approach for modeling diameter frequency data. *For. Sci.* 30:129–137.
- Dodson, C.T.J. (1992) The effect of fiber length distribution on formation. *J. Pulp Pap. Sci.* 18:J74–J76.
- Eckert, W.F., Masliyah, J.H., Afacan, A. (1997) Fractionation of softwood TMP by flotation. *Tappi J.* 80:210–216.
- Gavin, D.G., Hu, F.S. (2005) Bioclimatic modeling using Gaussian mixture distributions and multiscale segmentation. *Global Ecol. Biogeog.* 14:491–501.
- Huber, P., Roux, J.C., Mauret, E., Belgacem, N., Pierre, C. (2003) Suspension crowding for a general fiber-length distribution: Application to flocculation of mixtures of short and long paper making fibers. *J. Pulp Pap. Sci.* 29:77–85.
- Kropholler, H.W., Sampson, W.W. (2001) The effect of fiber length distribution on suspension crowding. *J. Pulp Pap. Sci.* 27:301–305.
- Lee, B.J., McDonald, A.G., James, B. (2001) Influence of fiber length on the mechanical properties of wood-fiber/polypropylene prepreg sheets. *Mater. Res. Innovat.* 4:97–103.

- Liu, C., Zhang, L., Davis, C.J., Solomon, D.S., Gove, J.H. (2002) A finite mixture model for characterizing the diameter distributions of mixed-species forest stands. *For. Sci.* 48:653–661.
- Lu, J.Z., Monlezun, C.J., Wu, Q., Cao, Q.V. (2007) Fitting Weibull and lognormal distributions to wood fiber length. *Wood Fiber Sci.* In press.
- Mao, C.X., Colwell, R.K. (2005) Estimation of species richness: mixture models, the role of rare species, and inferential challenges. *Ecology* 86:1143–1153.
- Mark, R.E., Gillis, P.P. (1983) Structure and structural anisotropy. In: *Handbook of Physical and Mechanical Testing of Paper and Paperboard*. Eds. Mark, R.E., Murakami, K. Marcel Dekker, Inc., New York, NY. pp. 283–371.
- Mörling, T., Sjöstedt-de Luna, S., Svensson, I., Fries, A., Ericsson, T. (2003) A method to estimate fibre length distribution in conifers based on wood samples from increment cores. *Holzforschung* 57:248–254.
- SAS Institute (2004) *SAS/STAT 9.1 User's guide*. SAS Institute, Inc., Cary, NC. 5121 pp.
- Svensson, I., Sjöstedt-de Luna, S., Bondesson, L. (2006) Estimation of wood fibre length distributions from censored data through an EM algorithm. *Scandinavian J. Stat.* 33:503–522.
- Takahashi, H., Suzuki, H., Endoh, K. (1979) The effect of fiber shape on the mechanical properties of paper and board. *Tappi J.* 62:85–88.
- Tasman, J.E. (1972) The fiber length of Bauer-McNett screen fractions. *Tappi J.* 55:136–138.
- Yan, J.F. (1975) A method for the interpretation of fiber length classification data. *Tappi J.* 58:191–192.
- Zhang, L., Liu, C., Davis, C.J. (2004) A mixture model-based approach to the classification of ecological habitats using Forest Inventory and Analysis data. *Can. J. For. Res.* 34: 1150–1156.

Received August 31, 2006. Accepted January 12, 2007.