FITTING WEIBULL AND LOGNORMAL DISTRIBUTIONS TO MEDIUM-DENSITY FIBERBOARD FIBER AND WOOD PARTICLE LENGTH

John Z. Lu†

Postdoctoral Research Associate
Department of Wood Science and Engineering
Oregon State University
Corvallis, OR 97330

Charles J. Monlezun

Associate Professor
Department of Experimental Statistics

Qinglin Wu†

Roy O. Martin Sr. Professor
School of Renewable Natural Resources

and

Quang V. Cao

Professor
School of Renewable Natural Resources
Louisiana State University Agricultural Center
Baton Rouge, LA 70803-6202

(Received July 2005)

ABSTRACT

Fiber lengths were analyzed for random samples of medium-density fiberboard (MDF) fibers and wood particles taken from eleven different populations. For six of the samples, the lognormal distribution fit the data, while the Weibull distribution did not. For three of the samples, the Weibull fit the data, while the lognormal did not. For two of the samples, both the lognormal and Weibull fit the data. Conclusions were based on hypothesis tests imposing a bound of 0.05 on the probability of making a Type I error for each test. Tests were based on large sample 95% nonparametric simultaneous confidence bands for the underlying cumulative distribution functions of the data.

Keywords: Goodness of fit tests, MDF fiber, maximum likelihood estimation, non-parametric confidence bands, probability plots, statistical analyses.

INTRODUCTION

Several studies have established the effect of wood fiber dimension and morphology on certain mechanical properties (e.g., bending strength and internal bonding strength) of wood fiber-based products such as paper, paper board, insulation board, medium-density fiberboard (MDF), hardboard, and wood-polymer composites (Takahashi et al. 1979; Mark and Gillis 1983; Eckert et al. 1997; Marklund et al. 1998; Lee et al. 2001; Myers 2002; Huber et al. 2003). However, suitable statistical functions that can be used to accurately describe wood fiber length distribution over various fiber length regimes are still missing.

For fiber dimension determination, the general measurement methods include microscopy,

† Member of SWST.

Wood and Fiber Science, 39(1), 2007, pp. 82–94

projection, screen classification (e.g., Bauer-McNett classifier), coulter counter and other particle size analyzers (e.g., Kajaani FS-200, Galai CIS-100), and image analyzers (Mark and Gillis 1983; Bentley et al. 1994; Carvalho et al. 1997). Particle size analyzers provide the most popular automated techniques for fiber length measurement in the pulp and paper industry. Compared with other automated techniques, image analyzers have high accuracy, fast speed, and high reliability for determining fiber length, width, surface area, and coarseness (Mark and Gillis 1983). Image analyzers are, perhaps, most useful when large numbers of fibers must be accurately detected and measured.

Rotation age, especially juvenile and adult woods (Haygreen and Bowyer 1994), and growth-accelerating treatments such as fertilization (Bannan 1967) and irrigation (Klem 1968; Murphey and Bowier 1975) have an effect upon average fiber dimensional sizes in stem. Individual wood fibers are different in shape and size for different tree species. Even in the same tree species and the same stem, individual fibers may have different lengths for different locations (e.g., height, distance from pith, early and late woods, and heart- and sapwoods). Therefore, fiber dimensional size (e.g., length, diameter or width, and surface area) in wood is inherently continuous in distribution.

As early as 1972, Tasman (1972) analyzed fiber lengths with a Bauer-McNett classifier. He discovered that the distribution of fiber fractions is approximately normal. It was also reported that fiber length distribution for papermaking furnish is approximately lognormal (Yan 1975; Dodson 1992; Kropholler and Sampson 2001). However, the data were simply fit with the lognormal distribution in these publications. The details on statistical analyses are not available.

Mark and Gillis (1983) used the Erlang family of distributions (a subset of the Gamma family) to describe the lengths of fibers from a bleached softwood kraft sheet. They applied chi-square goodness of fit tests to both the Erlang probability density function (pdf) and the Erlang cumulative distribution function (cdf). The resulting P-values were 0.870 and 0.947, respectively. They concluded that the Erlang model provided an excellent fit to this data set.

Failure to properly characterize an underlying distribution can lead to drastic errors in both analytical and simulation models (Law and Kelton 2000). The Weibull and lognormal families of distributions have been used with great success to model positively (or right) skewed random processes when the associated random variable is bounded below (right skewed refers to a process for which a large random sample would produce a histogram with the right or upper tail "stretched out" more than the left or lower tail). Most of the modeling has been empirical in nature, although mechanistically the Weibull arises naturally as the minimum observation for a certain class of distributions, while the product of a sufficient number of positive random variables, none of which dominates the others, is lognormal by the Central Limit Theorem (see Bury 1975; Johnson et al. 1994; Kalbfleisch and Prentice 2002). The Weibull distribution may also provide the added flexibility required for an accurate analysis when theoretical considerations indicate that an exponential distribution may be adequate (Johnson et al. 1994). The Weibull and lognormal distributions may be preferable to other distributions since both families assume a wide variety of shapes, and analyses are tractable for both the Weibull and lognormal. This is not surprising, since the lognormal is a function of the normal, which has been studied extensively, while the cumulative distribution for the Weibull is available in closed form; furthermore, the natural logarithm of each produces a location and scale parameter family of distributions (see SAS Institute Inc. 2002; Meeker and Escobar 1998). In addition, the Weibull and lognormal are "complementary" families of distributions in the following sense. When fitting both distributions to the same data set by the same method, the fitted lognormal pdf invariably has a heavier right hand tail than the fitted Weibull pdf, while the fitted Weibull pdf takes on larger values in the vicinity of zero (Law and Kelton 2000; Meeker and Escobar 1998). Two situations are then possible. Either one or both adequately fit the entire data set. If neither is satisfactory for the entire data set, then it is plausible that one of the two would fit large data values, the other would fit small data values, and both would fit the middle values. In either situation, by fitting both the Weibull and lognormal distributions, the process generating the data can be successfully modeled across the entire spectrum of the data.

The MDF fibers used in this study were created by a thermal-mechanical grinding process. The lengths of wood fibers obtained from such a process are bounded below by zero, since no fiber length can be negative. Such a process will produce extremely short fibers (having lengths close to zero). Most of the fibers would have lengths clustering around a positive value. The majority of naturally occurring long fibers would be crushed into shorter ones by the grinding process. Some long fibers would survive, but relatively few when compared to fibers of shorter length, resulting in a long right hand tail for the distribution of lengths. Thus, the distribution of fiber lengths should be continuous, bounded below by zero and skewed to the right. For this reason, Weibull and lognormal distributions were fit to each of the data sets of this study. The goodness of fit was determined for each. Details of all statistical analyses are presented.

EXPERIMENTAL

An imaging system was used for the determination of wood fiber length. This system consisted of a Leica MZFIII microscope (Leica Microsystems GmbH, Wetzlar, Germany), a CCD digital camera (Diagnostic Instruments, Sterling Heights, MI), a υ-Lux 1000 optical lighter (Volpi MFG. USH Co., Auburn, NY), a RT SP402-115 power supply (Diagnostic Instruments, Sterling Heights, MI), and a computer. The magnification number of the imaging system was 100 times.

The MDF fibers and wood particles were obtained in approximate 100 g packages from seven different sources. Several sources provided more than one package due to differences in fiber size, usages, or manufacturing lines for a total of eleven packages, designated by the letters A through K (Table 1). From each package, at least 500 fibers or particles were randomly sampled, oven-dried at 80°C for 24 h, and sealed in plastic bags. Exact sample sizes for all packages appear in Table 1. Fiber moisture content was kept between three and five percent. From each plastic bag, the fibers were removed, placed evenly on a glass dish, and clearly focused under the microscope. The optical lighter was adjusted for better picture quality. With the Spot Advance imaging software (Diagnostic Instruments, Sterling Heights, MI), the image was then recorded with the digital camera, and the length of each fiber was measured in millimeters using the Image-Pro Plus software (Media Cybernetics, Inc., Silver Spring, MD).

PROBABILITY DISTRIBUTION MODELS

The histograms of all wood fiber length data sets of this study were skewed to the right. The histograms for four of the data sets are representative of the shapes for all histograms and are given in Fig. 1. Because of the (right) skewness of the histograms, the Weibull and lognormal distributions were chosen to fit the data sets of this study.

A random variable X follows a Weibull distribution if and only if the cdf for X is

$$G(x; \gamma, \beta) = \begin{cases} 1 - \exp[-(x/\beta)^{\gamma}] & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{1}$$

where both γ and β are positive numbers. The

TABLE 1. *Summary of fiber information.*

Package designation	Fiber source	Fiber-based product[a]	Fiber type	Number of fibers per sample
A	1	MDF	Core	503
B	1	MDF	Face	505
C	2	Particleboard	Core	559
D	3	MDF	Core	512
E	3	MDF	Face	512
F	3	MDF	Face/core	527
G	4	Flakeboard	Core	595
H	5	MDF	Face/core	504
I	6	MDF	Face/core	505
J	6	MDF	Face/core	508
K	7	MDF	Face/core	503

[a] MDF = Medium density fiberboard.

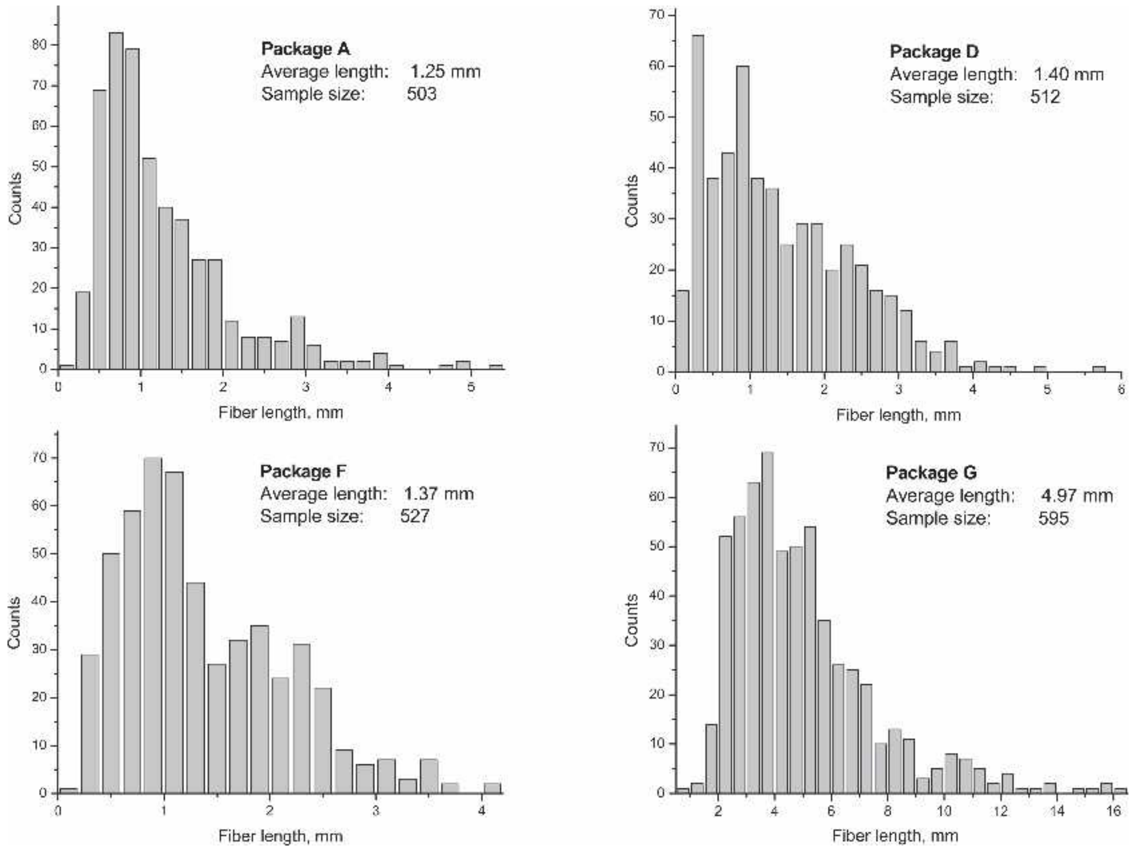

FIG. 1. Histograms of fiber length distribution for packages A, D, F, and G.

parameter γ is called the shape parameter, while β is called the scale parameter.

The pdf of the Weibull distribution

$$g(x;\gamma,\beta)=\begin{cases}\gamma\,(x^{\gamma-1}/\beta^{\gamma})\exp[-(x/\beta)^{\gamma}] & x\geq 0\\ 0 & x<0\end{cases} \tag{2}$$

is obtained by differentiating $G(x;\gamma,\beta)$ with respect to x.

The random variable X follows the lognormal distribution if and only if $Y = \ln X$ follows a standard normal distribution $N(\mu, \sigma^2)$. The pdf for X is

$$h(x;\mu,\sigma^2)=\begin{cases}\dfrac{1}{x\sqrt{2\pi\sigma 2}}\exp\left[-\dfrac{(\ln x-\mu)2}{2\sigma 2}\right] & x>0\\ 0 & x\leq 0;\end{cases} \tag{3}$$

μ is the location parameter ($-\infty < \mu < \infty$) and $\sigma > 0$ is the scale parameter. There is no closed form expression for the cdf $H(x; \mu, \sigma^2) = \int_{-\infty}^{x} h(t; \mu, \sigma^2)\, dt$ of X.

METHOD OF FITTING DISTRIBUTIONS

For each of the eleven data sets, location and scale parameters for the lognormal distribution, and shape and scale parameters for the Weibull distribution were estimated by the method of maximum likelihood using the RELIABILITY procedure in SAS®9.0 (SAS Institute Inc. 2002). Numeric values for the estimates appear in Table 2.

We shall now briefly describe the method of maximum likelihood. For detailed discussions of

TABLE 2. *Summary of distribution fitting**

Package designation/ **underlying distribution	Weibull				Lognormal			
	Shape γ	Scale β	Maximum loglikelihood	SSE	Location μ	Scale σ	Maximum loglikelihood	SSE
A/L	1.6875	1.4172	−493.6611	1.25454	0.0496	0.5878	−446.4296	0.15591
B/L	1.4946	1.2591	−564.6805	0.46562	−0.1175	0.7021	−537.9290	0.07897
C/W	2.5171	6.5311	−345.0989	0.32381	1.6619	0.4511	−348.2235	0.21108
D/W	1.4855	1.5541	−603.7361	0.23316	0.0539	0.8196	−624.6386	0.86972
E/W	1.4625	1.4760	−603.6720	0.28956	0.0070	0.8020	−613.5521	0.50384
F/L, W	1.8753	1.5486	−486.6175	0.43868	0.1430	0.6109	−488.0446	0.24566
G/L	2.0995	5.6332	−452.2452	1.34597	1.4897	0.4712	−396.5792	0.07008
H/L	1.5382	1.0966	−545.9236	0.74569	−0.2418	0.6719	−514.7229	0.17077
I/L	1.3363	1.1096	−620.7313	1.05338	−0.2844	0.7667	−584.4324	0.35276
J/L	1.3224	1.3496	−633.3904	0.65579	−0.0982	0.7987	−606.6133	0.27381
K/L, W	1.6548	1.4602	−522.1455	0.23869	0.0512	0.6709	−512.9337	0.33959

* Weibull and lognormal distributions were fit to each of eleven sets of fiber lengths.

** Results of testing H_0: the lognormal (or Weibull) distribution fits the data versus H_A: the lognormal (or Weibull) distribution does not fit the data, P(I) = 0.05 for each individual test. W = accept H_0 that the Weibull distribution fits the data. L = accept H_0 that the lognormal distribution fits the data.

maximum likelihood estimation, see Bickel and Doksum (1977), Casella and Berger (2002), Lawless (2003), Meeker and Escobar (1998), and Mood et al. (1974).

The likelihood function $L(\boldsymbol{\theta} \mid \boldsymbol{x})$ for a random sample $\boldsymbol{X} = \{X_1, X_2, \ldots, X_n\}$ of a random variable X is the joint pdf $f(\boldsymbol{x}; \boldsymbol{\theta})$ for $\boldsymbol{X}$ as a function of the parameter vector given $\boldsymbol{X} = \boldsymbol{x}$. The maximum likelihood estimate of $\boldsymbol{\theta}$ is that value of $\boldsymbol{\theta}$, say $\hat{\boldsymbol{\theta}}$, which maximizes $L(\boldsymbol{\theta} \mid \boldsymbol{x})$ for a fixed sample $\boldsymbol{x}$; i.e.,

$$L(\boldsymbol{\theta}|\boldsymbol{x}) \leq L(\hat{\boldsymbol{\theta}}|\boldsymbol{x}) \tag{4}$$

for all possible values of $\boldsymbol{\theta}$. As a function of $\boldsymbol{X}$, the random vector $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\boldsymbol{X})$ is called the maximum likelihood estimator of $\boldsymbol{\theta}$, while the numeric value of $\hat{\boldsymbol{\theta}}(\boldsymbol{X})$, namely $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$, is referred to as the maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$.

For many distributions, including the two parameter Weibull and lognormal distributions used in this study, the maximum likelihood estimators exist for any sample size. Under mild regularity conditions on the (joint) pdf for $\boldsymbol{X}$, $\hat{\boldsymbol{\theta}}$ converges in probability to $\boldsymbol{\theta}$. Also, under mild regularity conditions, the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ converges in distribution to a multivariate normal random vector having mean vector $\boldsymbol{\theta}$ and covariance matrix the inverse of the Fisher information matrix. This allows confidence intervals for $\boldsymbol{\theta}$ to be constructed and tests of hypotheses concerning $\boldsymbol{\theta}$ to be made using the chi-square and standard normal distributions, respectively, for large sample sizes. For small to moderate sample sizes, maximum likelihood estimators perform at least as well as other estimators.

Obtaining numeric values for maximum likelihood estimates invariably requires the use of a computational algorithm (usually imbedded in a software package such as SAS or S-Plus), since closed form expressions are not available except in a few elementary cases. In the majority of situations, MLE's can easily be obtained using the appropriate software packages. However, there are exceptions. It is possible that (a) the likelihood functions has several relative maxima (as opposed to a single global maximum), (b) the log of the likelihood function is unbounded, or (c) the likelihood function has a global maximum on a boundary of the parameter space. These situations present computational difficulties and should be analyzed with extreme care.

THE STATISTICAL TEST FOR GOODNESS OF FIT

Instead of one of the traditional goodness of fit tests, several of which are identified by name in the next section, we shall use a procedure that is both analytical and graphical, and is based on a large sample simultaneous $(1 - \alpha)\cdot 100\%$ confidence band for the underlying cdf $F(x)$ of a random variable X. The confidence band is of the form

$$\hat{F}(x) \pm e^*(x; 1 - \alpha/2)\sqrt{\hat{F}(x)[1 - \hat{F}(x)]/n}, \tag{5}$$

where $\hat{F}(x)$ is a nonparametric sample estimator of $F(x)$ based on a random sample of size n from X. The analytical part of the procedure is the determination of $e^*(x; 1 - \alpha/2)$ for each value of x. The graphical part consists in examining the probability plot (for a specific family of probability distributions) of $\hat{F}(x)$ together with the confidence band. Formally the null hypothesis that $F(x)$ is a member of the family of probability distributions is rejected in favor of the alternate hypothesis (with the bound of α on the probability of making a Type I error; i.e. $P(I) = \alpha$) if and only if it is not possible to construct a straight line extending the entire range of the data (i.e., from the smallest X_i to the largest X_i) that lies entirely within the confidence band. The alternate hypothesis is the negation of the null; i.e., that $F(x)$ is not a member of the family of probability distributions. The reader is referred to Meeker and Escobar (1998) for a detailed discussion of this procedure.

Weibull and lognormal probability plots of $\hat{F}(x)$ together with the 95% confidence band were constructed for each of the eleven data sets of this study using the SPLIDA package in S-Plus®7.0 (Insightful Corporation, 2005). SPLIDA is fully automated. Once the user specifies the family of distributions and the level of confidence, SPLIDA automatically calculates $\hat{F}(x)$, $e^*(x; 1 - \alpha/2)$, and $\sqrt{\hat{F}(x)[1 - \hat{F}(x)]/n}$, and then prints out the requested probability plot of $\hat{F}(x)$ together with the confidence band. It is, however, up to the user to determine whether or not it is possible to construct a straight line through the band. Each of our original probability plots of the 95% bands filled an 8-inch by 11.5-inch sheet of paper. There was no difficulty in deciding whether or not straight lines could be drawn through the bands for plots of this size. Should there be a problem, we recommend increasing the level of confidence until the bands expand sufficiently in size so that they clearly accommodate a straight line. Then reverse the process. Gradually decrease the level of confidence until it is clear that the bands contain no straight lines. In this manner, the P-value of the test can at least be approximated.

THREE PROCEDURES FOR COMPARING FITTED DISTRIBUTIONS

Likelihood ratio, chi-square, Kolmogorov-Smimov, Cramer-von Mises, Shapiro-Wilk, Anderson-Darling are but several of many traditional tests available for determining goodness of fit. Depending upon the distributional characteristics of interest and the objectives of the study, some goodness of fit tests may be more appropriate than others. Some have more power than others against specific alternatives. Some emphasize tail behavior, rather than behavior at the center of distribution. Thus, results may differ, depending on the test used. A common drawback of all such tests is that they all reject suitable models if the sample size is sufficiently large. For detailed discussion of these issues, the reader is referred to Law and Kelton (2000), Lawless (2003), and Vose (2000).

In this article, the goodness of fit test based on the simultaneous 95% confidence band of the previous section is adopted as the "gold standard," and supplemented by visual inspection of the probability plot of the fitted and nonparametric cdf's (discussed later in this section). Results from the other two procedures of this section are compared to those obtained from the simultaneous 95% confidence band.

Comparing maximum loglikelihood values

The maximum likelihood estimators of the parameters are obtained by maximizing the likelihood functions, or, equivalently, the (natural) logarithm of the likelihood functions. For a given data set, two statistics for comparing the Weibull fit to the lognormal fit are the maximum values of the loglikelihood functions $lnL(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}|X)$

for the Weibull and $lnL(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}| \boldsymbol{X})$ for the lognormal. If $lnL(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}|\boldsymbol{X}) >$ (or $<$) $lnL(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}|\boldsymbol{X})$, conclude that the lognormal (or Weibull) fit is preferable to the Weibull (or lognormal) fit. This selection procedure is based on the likelihood ratio test.

Comparing residual sums of squares

Let $X_1, X_2, \ldots, X_n$ be a random sample of X. Let $\hat{F}(x)$ be a nonparametric sample estimator of $F(x)$, the cdf of X. Let $\hat{G}(x) = \hat{G}(x; \hat{\beta}, \hat{\gamma})$ and $\hat{H}(x) = \hat{H}(x; \hat{\mu}, \hat{\sigma}^2)$ be the fitted Weibull and lognormal cdf's, respectively. Define $SSE_{Weibull}$ $\sum_{i=1}^{n}[\hat{F}(x_i) - \hat{G}(x_i)]^2$ and $SSE_{lognormal} = \sum_{i=1}^{n}[\hat{F}(x_i) - \hat{H}(x_i)]^2$. Since SSE consists of the sum of the squared deviations between the fitted cdf and the nonparametric sample estimator of the cdf, it measures how well the fitted distribution describes the data set over the entire range of the data. Thus, $SSE_{Weibull}$ and $SSE_{lognormal}$ may be used to compare the two fitted distributions over the entire range of the data. If $SSE_{lognormal} <$ (or $>$) $SSE_{Weibull}$, then the sum of the squared deviations between the fitted lognormal (or Weibull) cdf and the nonparametric sample estimator of $F(x)$ is smaller than the corresponding sum for the fitted Weibull (or lognormal) cdf, so that the fitted lognormal (or Weibull) cdf is "closer" to the nonparametric sample estimator of $F(x)$ than the fitted Weibull (or lognormal) cdf is. In this case, declare the lognormal (or Weibull) fit preferable to the Weibull (or lognormal) fit. This procedure is similar to that of Cramer-von Mises.

SSE values for each package were easily obtained using the basic calculation capabilities of SAS. All fitted Weibull cdf's were obtained from Eq. (1) using the SAS exponential function. The standard normal cdf function of SAS was required to get numeric values for each fitted lognormal cdf, since, unlike the Weibull, there is no easily implemented formula for the lognormal cdf.

Comparing probability plots

Visual assessment of goodness of fit may be made by inspection of the appropriate probability plots (Meeker and Escobar 1998). When the fitted lognormal (or Weibull) distribution function is plotted on lognormal (or Weibull) probability paper, the fitted distribution function appears as a straight line. When the nonparametric sample estimator $\hat{F}(x)$ of $F(x)$ is plotted together with the fitted distribution, the closer $\hat{F}(x)$ is to the fitted distribution, the better is the fit. For a given data set, then, whichever of the lognormal or Weibull probability plots has $\hat{F}(x)$ closer to the fitted distribution provides the better fit to the data. Weibull and lognormal probability plots of the fitted distributions together with the nonparametric sample estimators of the cdf's were obtained using the S-Plus/SPLIDA computing package.

RESULTS AND DISCUSSION

As previously mentioned, large sample 95% nonparametric simultaneous confidence bands were constructed for each of the eleven sets of fiber lengths and plotted on both lognormal and Weibull probability paper using the SPLIDA software package. Henceforth, these large sample 95% nonparametric simultaneous confidence bands will be referred to as the 95% lognormal and Weibull confidence bands, respectively.

For the data obtained from packages A, B, G, H, I, and J, it is possible to construct straight lines extending the entire range of the data that lie entirely within the lognormal confidence bands, while the curvature of the Weibull confidence bands prohibits the construction of any such lines that lie entirely within them. (For illustrative purposes, the 95% confidence bands for the data from packages A and G are shown in Figs. 2 and 3). Therefore, "H_0: the data come from a lognormal distribution" cannot be rejected with $P(I) = 0.05$, while "H_0: the data come from a Weibull distribution" must be rejected in favor of "H_A: the data do not come from a Weibull distribution" with $P(I) = 0.05$

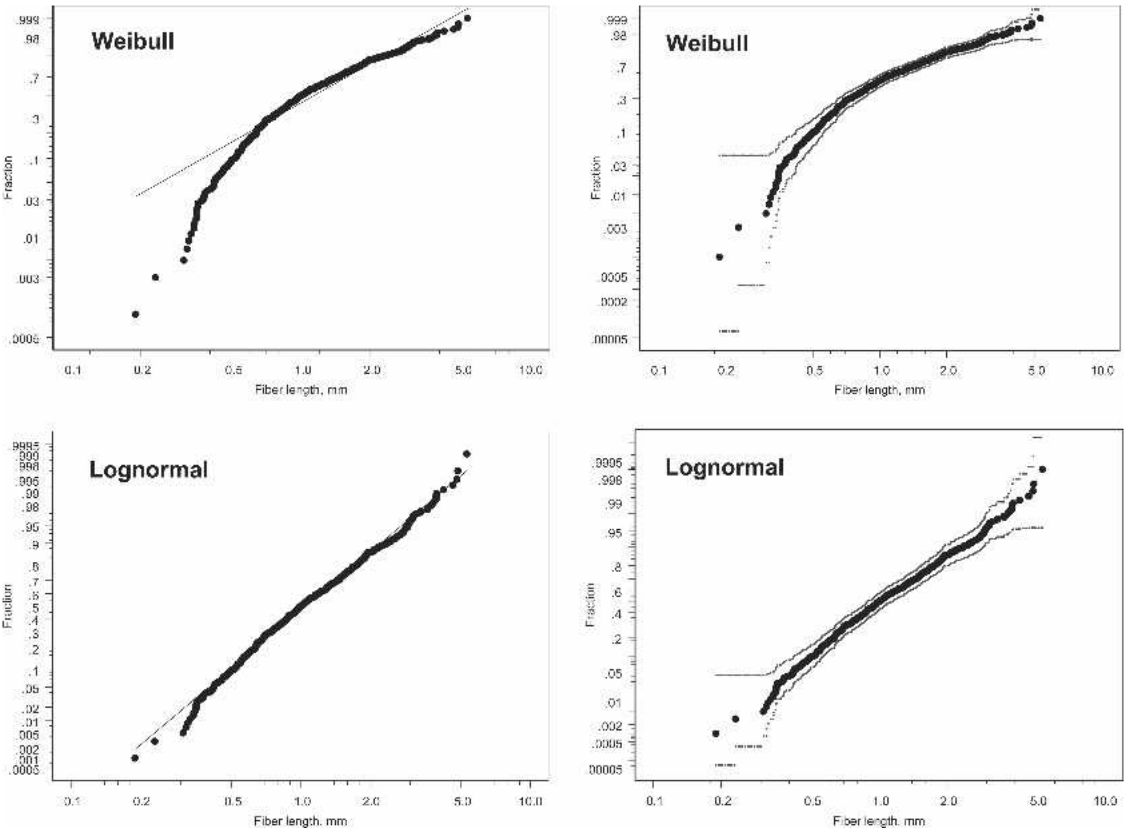

FIG. 2. Weibull and lognormal probability plots with fitted distributions (upper and lower left) and with simultaneous 95% confidence bands (upper and lower right) for package A.

for the data sets from packages A, B, G, H, I, and J. Thus, with P(I) = 0.05 for each test, the lognormal distribution provides an adequate fit to the fiber length data sets from these packages, while the Weibull distribution does not.

The opposite is true for the fiber lengths from packages C, D, and E. Straight lines that extend from the smallest length to the largest length may be placed through the 95% Weibull confidence bands for the data from packages C, D, and E, while the curvature of the corresponding 95% lognormal confidence bands prohibits the construction of any such straight lines that lie entirely within them. (The 95% confidence bands for the fiber lengths from package D are given in Fig. 4). Thus, with P(I) = 0.05 for each test, the Weibull distribution adequately fits the fiber lengths from these three packages, while the lognormal distribution does not.

For the fiber lengths from packages F and K, both the lognormal and Weibull distributions adequately describe the data with P(I) = 0.05. Straight lines extending the entire range of the data may be placed within the 95% Weibull confidence bands and the 95% lognormal confidence bands for these two data sets. (The 95% lognormal and Weibull confidence bands for the fiber lengths of package F are given in Fig. 5).

The lognormal distribution fit eight of the eleven data sets of this study with P(I) = 0.05, while the Weibull distribution fit five with P(I) = 0.05. For the smallest fiber lengths, each of the eleven fitted Weibull distributions was larger than the corresponding nonparametric sample estimator of the cdf, with the differences between the two increasing in magnitude as fiber length decreases to the minimum length. These differences are larger than the corresponding dif-

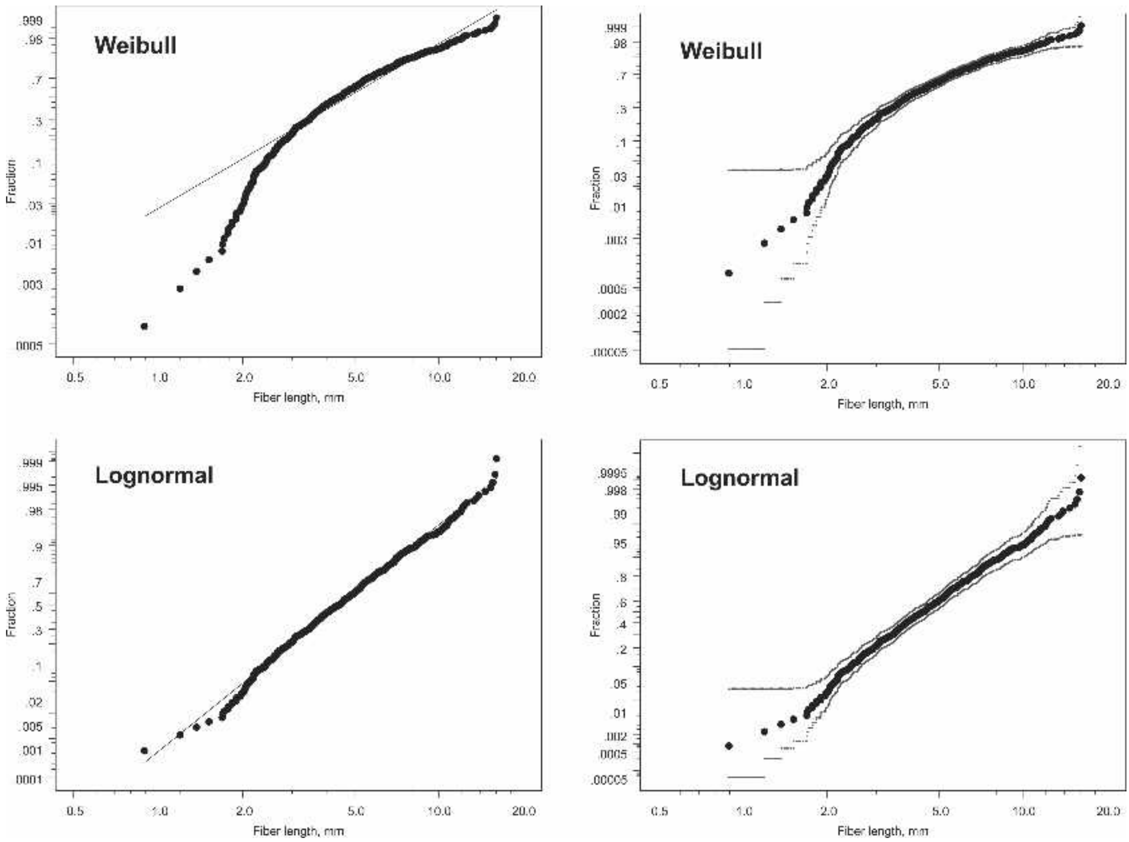

FIG. 3. Weibull and lognormal probability plots with fitted distributions (upper and lower left) and with simultaneous 95% confidence bands (upper and lower right) for package G.

ferences between the fitted lognormal cdf and the nonparametric sample estimator of the cdf. This phenomenon is clearly illustrated in Figs. 2, 3, 4, and 5. For each set of fibers from packages A, G, D, and F, the fitted Weibull cdf's are greater than the corresponding nonparametric sample estimators of the cdf's for those fibers less than 0.65 mm, 3.0 mm, 0.25 mm, and 0.55 mm in length, respectively. The corresponding fitted lognormal cdf's are closer to the nonparametric sample estimators of the cdf's for these fiber lengths than are the fitted Weibull cdf's. For estimating probabilities at the smaller fiber lengths, then, it would appear that estimates calculated using the fitted lognormal distribution would be more accurate than those calculated using the fitted Weibull distribution, even when the Weibull distribution and not the lognormal was judged to fit the data set [each at P(I) = 0.05].

An examination of the probability plots of the fitted Weibull and lognormal cdf's for the fibers from packages C, D, and E reveals that for all three packages, the fitted lognormal cdf's are smaller than the nonparametric sample estimators of the cdf's for the longest fibers. For example, the fitted lognormal cdf is smaller than the nonparametric sample estimator of the cdf for the fibers of package D having length greater than 2.7 mm (see Fig. 4). Packages C, D, and E were the very packages for which the fiber lengths were judged to follow the Weibull distribution but not the lognormal (all tests being conducted with P(I) = 0.05). It may be concluded, then, that the fit of the lognormal distribution to these data sets was rejected because of

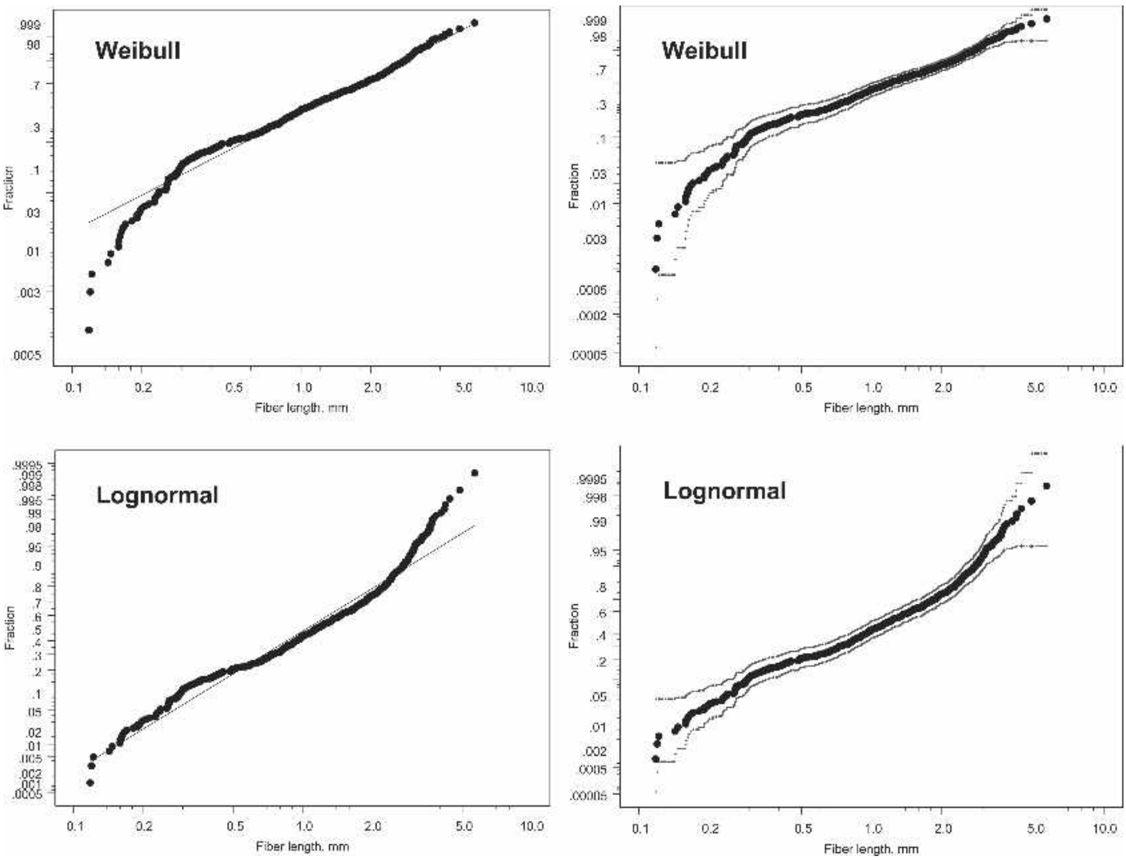

FIG. 4. Weibull and lognormal probability plots with fitted distributions (upper and lower left) and with simultaneous 95% confidence bands (upper and lower right) for package D.

the poor fit to the longest fibers. For the data sets from packages D and E, the Weibull cdf's were virtually identical to the nonparametric sample estimators of the cdf's for the longest fibers, and very close for package C, although not identical. The clear superiority of the fitted Weibull cdf's to the fitted lognormal cdf's at the largest fiber lengths was also observed for the fibers from packages F and K.

The relative magnitudes of $SSE_{lognormal}$ and $SSE_{Weibull}$ are generally consistent with the conclusions of the statistical hypothesis tests for determining goodness of fit. For the data sets adequately fit by the lognormal distribution but not the Weibull (i.e. the data sets obtained from the fibers from packages A, B, G, H, I, and J), $SSE_{lognormal}$ was considerably smaller than the corresponding $SSE_{Weibull}$, especially for package G (see Table 2). For packages D and E, for which the Weibull distribution fit the data sets but the lognormal did not, $SSE_{Weibull}$ was considerably less than $SSE_{lognormal}$. For the data sets from packages F and K, both distributions adequately fit the data according to the hypothesis tests. Therefore, it is not surprising that $SSE_{lognormal}$ is the smaller of the two for package F, while $SSE_{Weibull}$ is the smaller of the two for package K. For the data set from package C, $SSE_{lognormal}$ is considerably smaller than $SSE_{Weibull}$, so that SSE criterion would select the lognormal distribution over the Weibull. Yet the Statistical hypothesis tests chose the Weibull distribution as adequately fitting the data, and not the lognormal (with P(I) = 0.05 for each). The curvature in the probability plot of the nonparametric sample estimator of the underlying

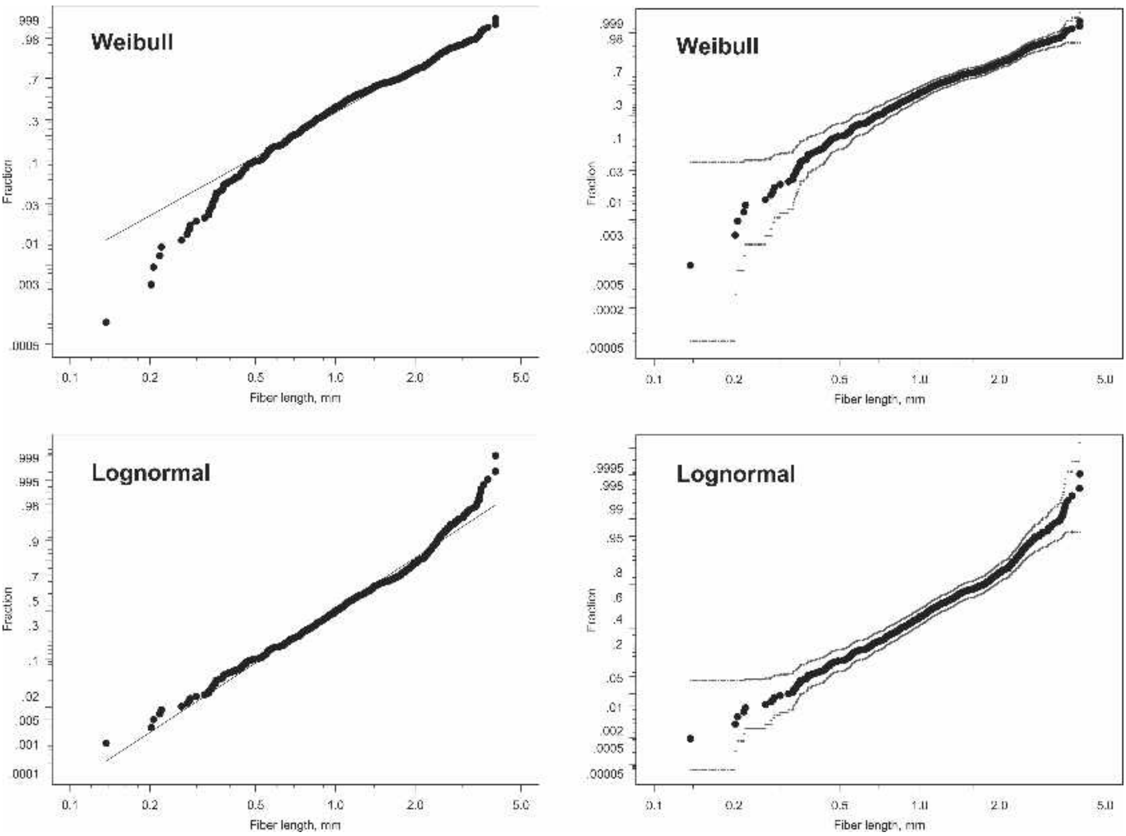

FIG. 5. Weibull and lognormal probability plots with fitted distributions (upper and lower left) and with simultaneous 95% confidence bands (upper and lower right) for package F.

cdf relative to the fitted lognormal cdf reveals that the lognormal distribution does not provide an acceptable fit to this data set. Thus, SSE by itself cannot always be relied upon to select the appropriate fit.

The maximum values for the loglikelihood functions were in perfect agreement with the hypothesis tests based on the 95% confidence bands for determining goodness of fit when one of the two distributions was judged to fit the data and the other was not. The lognormal maximum loglikelihood values were larger than their Weibull counterparts when the lognormal distribution was judged by the hypothesis tests to adequately fit the data but the Weibull was not (packages A, B, G, H, I, J; see Table 2). When the hypothesis tests declared that the Weibull distribution fit the data but the lognormal did not, the Weibull maximum loglikelihood values exceed their lognormal counterparts (packages C, D, and E). Both the lognormal and Weibull distribution were judged to adequately fit the data from packages F and K by the hypothesis tests. The Weibull maximum loglikelihood value exceeded that for the lognormal for the data from package F, while the lognormal maximum loglikelihood value was the larger of the two for the data from package K. Note that the SSE and maximum loglikelihood criteria are in perfect disagreement for the data from packages F and K.

CONCLUSIONS

Either the lognormal distribution or the Weibull distribution or both fit each data set of this study. For each data set, the fitted lognormal provided a closer fit to the data at the smallest fiber lengths. For five of the eleven data sets, the fitted Weibull provided the closer fit at the largest fiber lengths. With few exceptions, both fitted distributions closely fit the data for all but the shortest and longest fibers.

It may be concluded, then, that together, the lognormal and Weibull distributions form a robust set of distributions for fitting wood fiber length data sets. The distribution functions can be used to aid the development of larger mathematical models for predicting composite properties based on properties of wood fibers and manufacturing processes.

ACKNOWLEDGMENTS

The authors wish to thank Dr. George Woodson, Weyerhaeuser Company for helping collect all wood fibers used in this study. The authors would like to acknowledge Dr. Louis A. Escobar of the Department of Experimental Statistics at Louisiana State University for graciously providing access to the S-Plus/SPLIDA software. His insightful comments were invaluable in the preparation of this manuscript. The authors are indebted to Ms. Patricia Lebow of the USDA Forest Products Laboratory for serving as referee. Her careful review of the original manuscript led to a vastly improved revision. This work was partially supported by the Louisiana Governor's Biotechnology Initiative Grant Program (No. 101-40-0120).

REFERENCES

BANNAN, M. W. 1967. Anticlinal divisions and cell length in conifer cambium. Forest Prod. J. 17(6):63–69.

BENTLEY, R. G., P. SCUDAMORE, AND J. S. JACK. 1994. A comparison between fiber length measurement methods. Pulp Paper Canada 95(4):T155-T159.

BICKEL, P. J., AND K. A. DOKSUM. 1977. Mathematical statistics: Basic ideas and selected topics. Kolde-Day, Inc. Oakland, CA.

BURY, K. V. 1975. Statistical models in applied science. John Wiley & Sons, New York, NY.

CARVALHO, M. G., P. J. FERREIRA, A. A. MARTINS, AND M. M. FIGUEIREDO. 1997. A comparison study of two automated techniques for measuring fiber length. TAPPI 80(2):137–142.

CASELLA, G., AND R. L. BERGER. 2002. Statistical Inference. 2nd ed. Duxbury, Pacific Grove, CA.

DODSON, C. T. J. 1992. The effect of fiber length distribution on formation. J. Pulp Paper Sci. 18(2):J74-J76.

ECKERT, W. F., J. H. MASLIYAH, AND A. AFACAN. 1997. Fractionation of softwood TMP by flotation. TAPPI 80(5): 210–216.

HAYGREEN, J. G., AND J. L. BOWYER. 1994. Forest products and wood science: An introduction. 2nd ed. Iowa State University Press, Ames, IA.

HUBER, P. J.-C. ROUX, E. MAURET, N. BELGACEM, AND C. PIERRE. 2003. Suspension crowding for a general fiber-length distribution: Application to flocculation of mixtures of short and long paper making fibers. J. Pulp Paper Sci. 29(3):77–85.

Insightful Corporation 2005. S-Plus®7.0 for Windows, http://www.insightful.com/products/ splus.

JOHNSON, N. L., S. KOTZ, AND N. BALAKRISHNAN. 1994. Continuous univariate distributions, Vol. 1, 2nd ed. John Wiley & Sons, New York, NY.

KALBFLEISCH, J. D., AND R. S. PRENTICE. 2002. The statistical analysis of failure time data, 2nd ed. John Wiley & Sons, Hoboken, NJ.

KLEM, G. S. 1968. Some aspects of the qualities of wood from fertilized forests. TAPPI 51(11):99A-103A.

KROPHOLLER, H. W., AND W. W. SAMPSON. 2001. The effect of fiber length distribution on suspension crowding. J. Pulp Paper Sci. 27(9):301–305.

LAW, A. M., AND W. D. KELTON. 2000. Simulation modeling and analysis, 3rd ed., McGraw-Hill, Inc. Boston, MA.

LAWLESS, J. F. 2003. Statistical models and methods for lifetime data, 3rd ed. John Wiley & Sons. Hoboken, NJ.

LEE, B. J., A. G. MCDONALD, AND B. JAMES. 2001. Influence of fiber length on the mechanical properties of wood-fiber/polypropylene prepreg sheets. Mat. Res. Innovat. 4:97–103.

MARK, R. E., AND P. P. GILLIS. 1983. Structure and structural anisotropy. Pages 283 and 371 *in* R. E. Mark, K. Murakami, eds. Handbook of physical and mechanical testing of paper and paperboard. Marcel Dekker, Inc., New York, NY.

MARKLUND, A., M. PAPER, J. B. HAUKSSON, U. EDLUND, AND M. SJOSTROM. 1998. Prediction of strength parameters for softwood kraft pulps. Multivariate data analysis based on physical and morphological parameters. Nordic Pulp Paper Res. J. 13 (3):211–219.

MEEKER, W. Q, AND L. A. ESCOBAR. 1998. Statistical methods for reliability data. John Wiley & Sons, Inc., New York, NY.

MOOD, A. M., F. A. GRAYBILL, AND D. C. BOES. 1974. Introduction to the theory of statistics. 3rd ed., McGraw Hill, New York, N.Y.

Murphey, W. K., and J. J. Bowier. 1975. The response of aspen to irrigation by municipal waste water. TAPPI 58(5):128–129.

Myers, G. 2002. Thermomechanical pulping of loblolly pine juvenile wood. Wood Fiber Sci. 34(1):108–115.

SAS Institute Inc. 2002. SAS Online DOC®9.0, http://V9doc.sas.com/sasdoc.

Takahashi, H., H. Suzuki, and K. Endoh. 1979. The effect of fiber shape on the mechanical properties of paper and board. TAPPI 62(7):85–88.

Tasman, J. E. 1972. The fiber length of Bauer-McNett Screen Fractions. TAPPI 55(1):136–138.

Vose, D. 2000. Risk analysis: A quantitative guide. John Wiley & Sons, Chichester, West Sussex.

Yan, J. F. 1975. A method for the interpretation of fiber length classification data. TAPPI 58(8):191–192.