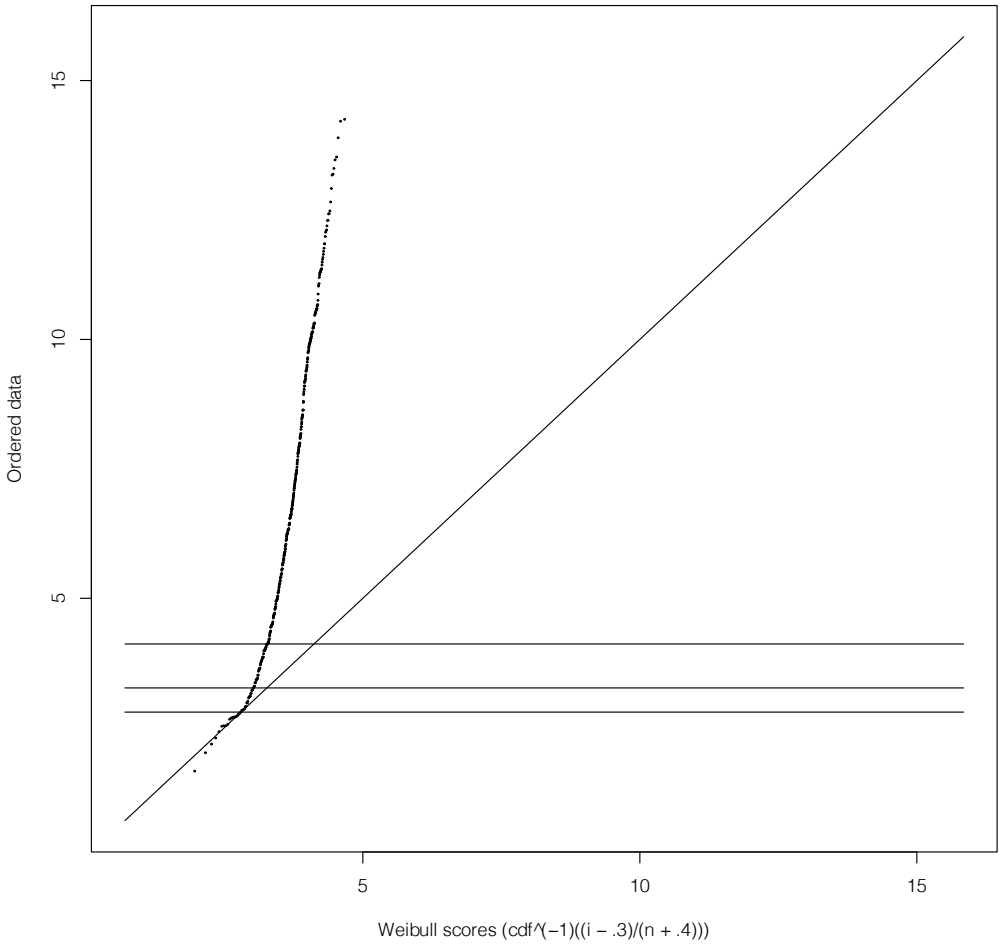
United States Department of Agriculture

Visual and MSR Grades of Lumber Are Not Two-Parameter Weibulls and Why It Matters (with a Discussion of Censored Data Fitting)

Steve P. Verrill
Frank C. Owens
David E. Kretschmann
Rubin Shmulsky
Linda S. Brown



Forest Service
Forest Products Laboratory
Research Paper FPL–RP–703
December 2019

Abstract

It has been common practice to assume that a two-parameter Weibull probability distribution is suitable for modeling lumber strength properties. Verrill and others (papers published in 2012, 2013, 2014, and 2015) demonstrated theoretically and empirically that the modulus of rupture (MOR) distribution of a visual grade of lumber or of lumber that has been "binned" by modulus of elasticity (MOE) is not a two-parameter Weibull. Instead the tails of the MOR distribution are thinned via "pseudo-truncation." Verrill and others (papers published in 2013 and 2014) performed simulations that established that fitting two-parameter Weibulls to pseudo-truncated data via either full or censored data methods can yield poor estimates of probabilities of failure. In this paper we support the 2013 and 2014 simulation results by analyzing large "In-Grade type" data sets and establishing that two-parameter Weibull fits yield inflated estimates of the probability of lumber failure when specimens are subjected to loads near allowable properties. We also discuss the censored data or "tail fitting" methods permitted under ASTM D5457, and we extend, via simulations, the empirical "In-Grade type" data results for individual pieces of lumber to a simple seven-member assembly.

December 2019

Verrill, Steve P.; Owens, Frank C.; Kretschmann, David E.; Shmulsky, Rubin; Brown, Linda S. 2019. Visual and MSR grades of lumber are not two-parameter Weibulls and why it matters (with a discussion of censored data fitting). Research Paper FPL-RP-703. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 40 p.

A limited number of free copies of this publication are available to the public from the Forest Products Laboratory, One Gifford Pinchot Drive, Madison, WI 53726-2398. This publication is also available online at www.fpl.fs.fed.us. Laboratory publications are sent to hundreds of libraries in the United States and elsewhere.

The Forest Products Laboratory is maintained in cooperation with the University of Wisconsin.

Keywords: two-parameter Weibull distribution, pseudo-truncated Weibull distribution, machine stress rated lumber, MSR lumber, MOE binned lumber, visually graded lumber, thinned tail, lumber property distribution, lumber reliability, lumber assembly reliability, censored data two-parameter Weibull fits

Contents

Visual and MSR Grades of Lumber Are Not Two-Parameter Weibulls and Why It Matters (with a Discussion of Censored Data Fitting)

Steve P. Verrill, Mathematical Statistician
USDA Forest Service, Forest Products Laboratory, Madison, Wisconsin, USA

Frank C. Owens, Assistant Research Professor
Department of Sustainable Bioproducts, Mississippi State University, Mississippi, USA

David E. Kretschmann, Research General Engineer (retired[1])
USDA Forest Service, Forest Products Laboratory, Madison, Wisconsin, USA

Rubin Shmulsky, Professor and Department Head
Department of Sustainable Bioproducts, Mississippi State University, Mississippi, USA

Linda S. Brown, Engineer
Southern Pine Inspection Bureau, Pensacola, Florida, USA

1 Introduction

Background and Objective

Verrill et al. (2012, 2013, 2014, 2015) established theoretically and empirically that the strength properties of visual and machine stress rated (MSR) grades of lumber are not distributed as two-parameter Weibulls. Instead, the strength properties of grades of lumber must have (at least to a first approximation) "pseudo-truncated" distributions.

"Pseudo-truncation" has a technical meaning. The concept, at least, of pseudo-truncation was recognized in an American Society of Civil Engineers (ASCE) pre-standard report (ASCE 1988). Section B3 of that standard notes that "an improved strength distribution can be obtained by ... thinning the lower tail by sorting on a correlated variable." For example, if the full ("mill run") bivariate modulus of elasticity–modulus of rupture (MOE–MOR) distribution were a bivariate Gaussian (normal)–Weibull, then truncating or "binning" on the basis of MOE values (as in machine stress rated lumber) would lead to a pseudo-truncated MOR distribution. That is, because MOE and MOR are not perfectly correlated, truncating on the basis of lower and upper MOE limits does not lead to perfect truncation of the MOR distribution, but it does, of course, lead to a MOR distribution whose tails are thinned. For the case in which the mill run joint MOE–MOR distribution is a bivariate Gaussian–Weibull, Verrill et al. (2012, 2015) derived the exact form of this "pseudo-truncated Weibull" distribution. (They obtained its probability density function.) They also showed that it *cannot* have tail behavior that matches that of a Weibull distribution.

Verrill et al. (2013, 2014) performed simulations that strongly suggested that both uncensored and censored (see section X2 of ASTM D5457 (ASTM 2019)) fits of two-parameter Weibulls to pseudo-truncated data can lead to significant over- or underestimates of probabilities of failure when loads are near allowable properties. Censored data techniques are also known among wood scientists as "tail fitting" and involve situations in which we have (or use) full information for only

[1] now President, American Lumber Standard Committee, Inc., Frederick, Maryland, USA

a subset of the data. Censored data techniques are discussed in many statistical textbooks that deal with reliability or lifetime estimation methods. See, for example, Lawless (2003).

In this paper we support the 2013 and 2014 simulation results by analyzing large "In-Grade type" data sets and establishing that modeling the MOR distributions of visual grades of lumber by two-parameter Weibull distributions can lead to poor reliability estimates. In particular, when loads are close to the allowable properties calculated for those data sets, estimates of probabilities of failure will tend to be inflated. Our preceding simulation work — see section 5.4 of Verrill et al. (2013) — demonstrated that this positive bias in the mean tends to decrease with censoring, but also, as we would expect, censored data estimates are *more variable* than full data estimates. Consequently, censored data techniques can, with appreciable probability, yield failure probability estimates that are considerably too high and, again with appreciable probability, yield estimates that are considerably too low. The problems associated with estimating wood grade strength distributions via censored data techniques are further discussed in Section 5.

A Note about Censoring, Truncation, and Pseudo-truncation

Some readers might be confused about the differences among censoring, truncation, and pseudo-truncation. In fact, in the statistical (and more general) literature, the terms "censored" and "truncated" are sometimes conflated. In this paper, we draw a sharp distinction between the two terms. The third term, "pseudo-truncation," is related to "truncation" and was introduced in Verrill et al. (2012). We include the next few paragraphs as an introduction to these topics.

"Censoring" applies to *a type of data* and the *corresponding statistical analysis techniques*. For example, one might be interested in estimating the distribution of the lifetimes of a certain electronic component. One might obtain a random sample of 1000 of the components and put them "under load" for 60 days. At the end of 60 days, some of the components — N of them — will have failed at known times (their "failure times"), and the remaining $1000 - N$ specimens will not have failed. A censored data analysis will take into account the N known failure times, and the fact that $1000 - N$ of the specimens did not fail in the course of the 60 day experiment, and will yield a censored data estimate of the lifetime distribution of the components. (See, for example, Lawless (2003) for general censored data techniques and appendix X2 of ASTM (2019) for some right-censored two-parameter Weibull data techniques.) In a wood science setting, we might be working with strength rather than lifetime, and a maximum experimental load rather than a time-limited experiment. The same censored data statistical techniques are used in both cases.

We emphasize that in a censored data situation, the underlying population distribution from which a sample is obtained *is not altered*. Instead, in the presence of censoring (e.g., a limit on time or loads in an experiment), we have only partial information about some of our sample observations (e.g., that they are above some value) and special "censored data techniques" must be employed to draw correct conclusions about the characteristics of the full underlying population distribution.

On the other hand (at least in this paper), "truncation" and "pseudo-truncation" apply to *population distributions* (rather than a type of data and the associated types of statistical analyses). For example, the mill run MOE distribution of all lumber (of a specific size) produced by a mill might have a normal distribution. In this case, the distribution of all MOEs of lumber of the specific size from that mill that lie between a specific lower bound, c_{l}, and a specific upper bound, c_{u}, would be a truncated normal. In Figure 1, the solid line is the probability density function (pdf) of a standard normal (N(0,1)) distribution, and the dotted line is the pdf of a N(0,1) distribution truncated at its 40th and 80th percentiles. In contrast to a censored data situation, in a truncated situation (or a pseudo-truncated situation, see below) we are interested in drawing inferences about a distribution (the truncated or pseudo-truncated distribution) that *does* differ from the original

full distribution.

A pseudo-truncated distribution (Verrill et al. 2012, 2013) is related to a truncated distribution in the following manner. Suppose, for example, that a mill run population of lumber (of a specific size) has a joint MOE–MOR distribution that is a bivariate Gaussian–Weibull (that is, the mill run MOE population for lumber of the particular size has a normal distribution, the mill run MOR population for lumber of the particular size has a two-parameter Weibull distribution, and the MOE and MOR values are positively correlated). Now further suppose that we consider only those MOR values for which the corresponding MOE values lie between limits c_{l} and c_{u} (as in machine stress-rated (MSR) lumber). Then the MOR values are those associated with truncated MOE values. However, because the correlation between MOE and MOR is not perfect (the correlation is not 1), the population of MOR values associated with the truncated MOE values will *not* in turn be truncated (there will not be sharp left and right MOR boundaries). Instead the MOR distribution will have "thinned tails." (Actually, changes in the MOR distribution due to hard limits on MOE values occur throughout the distribution, not just in its tails.) The MOR distribution will be "pseudo-truncated." For the Gaussian–Weibull case and the mixture of two bivariate normals case, the exact mathematical forms of the pseudo-truncated MOR distributions have been determined. They are not two-parameter Weibulls (Verrill et al. 2012, 2018). The exact mathematical forms of pseudo-truncated distributions can also be calculated in other cases.

In Figure 2, the solid line is the pdf of a full two-parameter Weibull distribution with shape parameter 3.1 and scale parameter 1.0. The corresponding pseudo-truncated Weibull pdf for the case in which

1. we begin with a bivariate Gaussian–Weibull MOE–MOR distribution,

2. the bivariate Gaussian–Weibull correlation is 0.7, and

3. the MOE is truncated at its 40th and 80th percentiles

is plotted as a short-dashed line in Figure 2. The corresponding pseudo-truncated Weibull pdfs for the cases in which the bivariate Gaussian–Weibull correlations are 0.9, 0.99, and 0.9999 are plotted respectively as a dotted line, a dot-dashed line, and a long-dashed line in Figure 2. We can see that as the MOE–MOR correlation approaches one, a pseudo-truncated distribution approaches a fully truncated distribution. However, for "realistic" correlations, the distributions continue to look vaguely "bell-shaped" with tails that are "shorter" than the original distribution.

2 The Data

The data come from 19 of the original In-Grade data cells (species-size-grade-property combinations), 6 data cells from a 2011 Southern Pine Inspection Bureau (SPIB) repeat of the In-Grade testing program, and 1 data cell from a 2014 SPIB resource monitoring program study. The data cells are identified in columns 1–4 of Table 1. The In-Grade program and some of its results are discussed in Green et al. (1988, 1989) and Evans and Green (1988). Testing procedures for the In-Grade testing program are described in ASTM D4761 (ASTM 2013), and the data were adjusted in accordance with ASTM D1990 (ASTM 2016). The original SPIB resource monitoring program (1994–2010) is discussed in Kretschmann, Evans, and Brown (1999). In recent years, the program has been modified to ensure conformity with the requirements in the most recent version of ASTM D1990 and to add action points that depend upon both strength and stiffness measurements.

3 An Extension of an Earlier Analysis

In their table 1, Verrill et al. (2014) provided Cramér-von Mises and Anderson-Darling goodness-of-fit test p-values for tests of the null hypotheses that 19 In-Grade data cell MOR distributions were two-parameter Weibulls. In Table 1 of the current paper, we extend this 2014 table to include the 2011 and 2014 SPIB data. To perform the goodness-of-fit tests, we used an R (R Core Team 2013) goodness-of-fit function, WEDF.test (Krit 2014), that provides more precise estimates of the p-values than the estimates provided in Verrill et al. (2014). This updated table continues to strongly suggest that visual grades of lumber are poorly fit by two-parameter Weibulls.

The 2014 paper also discussed Weibull probability plots of the data. In Figure 3, we provide an example of such a plot. The 2014 paper noted that 16 of the 19 data sets available at the time led to probability plots that had the "short or thinned left tails" that one would expect from pseudo-truncated data. (That is, the points in the left tails of the probability plots tended to lie above $y = x$ lines.) Twenty of the 26 data sets currently available to us display such a short left tail. To give readers an idea of how unlikely this would be if the data truly were two-parameter Weibull, we performed 26 simulations. To do this, for each of the 26 data sets, we first obtained the maximum likelihood fit of a two-parameter Weibull to the data. If the original data set contained n points, we then generated a sample of size n from the fitted two-parameter Weibull distribution and plotted the corresponding Weibull probability plot. An example of such a plot is provided in Figure 4. (Actual and generated probability plots for all 26 data sets can be viewed at `http://www1.fpl.fs.fed.us/weib2.pp.html`.) None of the 26 generated probability plots displayed shortened left tails. *If* the real MOR distributions are two-parameter Weibulls, an *approximate* estimate of the probability of seeing *all* 20 of the observed short left tails among the original 26 probability plots and none among the probability plots associated with generated data or vice versa is

$$2 \times \left(\begin{array}{c} 26 \\ 20 \end{array} \right) \times \left(\begin{array}{c} 26 \\ 0 \end{array} \right) / \left(\begin{array}{c} 52 \\ 20 \end{array} \right) = 4 \times 10^{-9}$$

Note that this estimate ignores differences that may be due to species, lumber size, grade, and sample size differences, so it is merely suggestive rather than definitive. In fact, the data and intuition suggest that Select Structural (SS) and No. 2 fits may behave differently — 13 of 14 No. 2 probability plots display overly heavy *predicted* (thin *observed*) left tails, while only 7 of 12 SS plots do. This might be associated with the fact that the SS data sets are all left-skewed, while the No. 2 data sets are all right-skewed. (See the full, actual data, skewness estimates in Table 2.)

Regardless, it is unlikely that the 26 MOR distributions are two-parameter Weibulls. Of course, we essentially already knew this from the Cramér-von Mises and Anderson-Darling goodness-of-fit tests. However, this second analysis focuses our attention on the thinned observed left tails as one source of the lack-of-fit.

So far, we have been seeking to conclude that two-parameter Weibull fits may be *statistically* rejected. In the next section, we identify a *practical* reason for rejecting two-parameter Weibull lumber strength models when making reliability predictions.

4 The New Analysis

We used full and censored data maximum likelihood methods to fit two-parameter Weibull distributions to each of the 26 data sets. A listing of the program that did the fitting (fit26.w2.cens.4.web.f) can be found at `http://www1.fpl.fs.fed.us/weib2.prog.html`. The program also obtained non-parametric estimates of the 5th percentiles of the 26 strength populations.

For In-Grade type data cell j, the two-parameter Weibull estimate of the probability that the strength of a member of the cell would fall below x was calculated as

$$\text{Prob}_{\text{W},j} = 1 - \exp\left(-\left(x/\hat{\lambda}_j\right)^{\hat{\beta}_j}\right)$$

where $\hat{\lambda}_j$, $\hat{\beta}_j$ were the maximum likelihood estimates of the scale and shape parameters for cell j. These parameter estimates are provided in Table 2 for full, censored 20, censored 10, and censored 5 fits. A "censored 20" fit, for example, is one in which in our maximum likelihood fit, we make use of the bottom 20% of the data, and the number, N_j, of specimens sampled for cell j. (N_j corresponds to the n in equation X2.1 of ASTM D5457.)

The empirical estimate of this probability was

$$\text{Prob}_{\text{emp},j} = n_j/N_j$$

where n_j was the number of specimens in cell j with MORs that fell below x, and N_j was the total number of specimens sampled for cell j. In this paper, we refer to these probabilities as "probabilities of failure." We considered cases in which x equaled the nonparametric estimate of the 5th percentile of the distribution divided by 1.9, 2.1, and 2.3 (that is, for cases in which x was in the neighborhood of the allowable property).

In Tables 3–5 (corresponding to the three divisors 1.9, 2.1, and 2.3), $\text{Prob}_{\text{W},j} \times N_j$ values (the expected number of "failures" under full, censored 20, censored 10, and censored 5 two-parameter Weibull fits) are presented in columns 6–9, and n_j (the observed number of "failures") is presented in column 10. A listing of the program that was used to produce Tables 3–5 can be found at `http://www1.fpl.fs.fed.us/weib2.prog.html`.

When $\text{Prob}_{\text{W},j} \times N_j \gg n_j$, the two-parameter Weibull fit is likely to be overestimating the true probability of a failure. In some cases, the estimates based on a two-parameter Weibull fit could be said to be highly inflated (consider the 20.8, 5.4, 11.0, and 5.9 predictions in column 6 of Table 4).

For the 2.1 divisor (Table 4), the total number of cases in which specimen strengths actually lay below nonparametric 5th/2.1 values was *9*. That is, the observed overall failure probability was $9/13275 = 0.00068$. For the full, censored 20, censored 10, and censored 5 two-parameter Weibull fits, the expected total numbers of failures were *83.8*, *29.8*, *15.9*, and *12.4*. The 83.8 value is more than nine times the observed number of failures and yields an overall probability of failure estimate of $83.8/13275 = 0.0063$.

This factor of nine suggests to us that a two-parameter Weibull model for the MOR distribution of grades of lumber is a poor one. A two-parameter Weibull model leads to inflated estimates of probabilities of failure when loads are near allowable properties. This result can be expected from the "tail-thinning" due to pseudo-truncation that was explored in Verrill et al. (2012, 2013, 2014, 2015). (Of course, the inflation factor might actually vary with the cell. For example, if we restrict ourselves to the SS cells and uncensored fits, the ratio is $10.8/4 = 2.7$, whereas if we restrict ourselves to the No. 2 cells and uncensored fits, the ratio is $73.0/5 = 14.6$.)

On the other hand, one could argue that despite the fact that Verrill et al. (2012, 2015) established theoretically that pseudo-truncated strength distributions are unlikely to be two-parameter Weibulls, and despite the fact that formal goodness-of-fit tests reject two-parameter Weibull models for the strength distributions of grades of lumber, and despite the fact that two-parameter Weibull fits to In-Grade data sets lead to overestimates of probabilities of failure for loads near allowable properties, one could — on a purely ad hoc basis — use *censored data* fits to two-parameter Weibulls to predict probabilities of failure for loads near allowable properties. Such an argument might be

made by someone who noted from Tables 3–5 that although censored data fits also lead to overestimates of the numbers of failures for loads near allowable properties, the overestimates appear to decline as the censoring increases. That is, the upward bias in the estimate of the number of failures appears to decrease as the censoring increases.

Our short response is that the simulations of Verrill et al. (2013, 2014) established that although the bias in the estimate of the probability of failure declines as censoring increases, the *variance* of the estimate significantly increases (as one would expect from censored data estimates) with the result that, for any given data set, censored data estimates of failure probabilities will have good chances of being seriously inflated or seriously deflated (while, *given* correct joint mill run distributions of visual grade, MOE, and MOR, and current visual and MSR grading practices, maximum likelihood theory guarantees that "pseudo-truncated" estimators will be asymptotically unbiased, with minimum variance among all unbiased estimators).

In the next section we provide a longer response to the suggestion that the strength distributions of visual or MSR grades of lumber can be well-approximated by censored data fits of two-parameter Weibull models.

5 Censored Data

ASTM D5457 permits two-parameter Weibull distributions to be fit to data via censored data methods. Some scientists refer to this as "low tail fitting." Two permitted censored data fitting techniques (maximum likelihood and "method of least squares") are described in section X2 of ASTM D5457. To produce a "low tail fit," one might explicitly use only the bottom (say) 20% of the data (the bottom "n_c" data values in the notation of section X2 of ASTM D5457) and the fact that 80% of the data ("n_s" of the data values in the notation of section X2 of ASTM D5457) exceed the maximum of the bottom 20%.

In section X1.1.2 of ASTM D5457 the authors of the standard write

> Refinements of these procedures suggest that estimates of the distribution and its parameters give the most accurate reliability estimates when they represent a tail portion of the distribution rather than the full distribution. This reflects the fact that, for common building applications, only the lower tail of the resistance and upper tail of the load distribution contribute to failure probabilities.

In section X1.1.3 of ASTM D5457 the authors of the standard write

> In addition, by permitting tail fitting of the data, it provides a way of fitting data in this important region that is superior to full-distribution types.

We would argue that this notion of the superiority of "tail fitting" is mistaken. This issue is relevant, important, and somewhat opaque. Thus, we feel that addressing it in some detail here is appropriate.

Statisticians refer to "tail fitting" methods as "censored data" methods and know that they are derived for the case in which we know the probability density function associated with the bottom (in the notation of section X2 of the standard) n_c strength values in a sample, know those strength values, and further know that the remaining $n - n_c$ values in the sample are larger than the largest of the bottom n_c values. We might encounter such data if we applied a fixed maximum load to all the specimens in a data set (rather than loading all the specimens to failure). We might also encounter such a data set if we chose to simply record strengths larger than the bottom n_c strengths as "larger than the bottom n_c strengths" (as is contemplated in section X2 of ASTM D5457).

Among statisticians, it is a well-accepted "fact" that *if our probability models are correct*, we obtain better (lower mean squared error) parameter estimates and thus, in reliability situations, better estimates of the probability of failure (anywhere, including the left tail) by performing full data fits rather than censored data fits. (Of course, this will not be possible if we do not load all the pieces to failure, or do not record failure loads above a certain level.)

That is, if we have, for example, 400 random draws from a distribution, we obtain better estimates of the parameters of this distribution (and thus estimates of the percentiles of this distribution) by using explicit knowledge of all 400 data values than by using just the explicit bottom 80 values and knowledge that the remaining 320 values exceed the maximum of the 80 values (the censored data approach covered in section X2 of D5457).

We performed 500 simulation trials of full, censored 20, and censored 10 fits to samples of size 400 that confirmed that when the full underlying distribution is a two-parameter Weibull, we do better with full data analyses than with censored data analyses. Results from these simulations are reported in Table 6. A listing of the Fortran computer program that was used to perform the simulations can be found at `http://www1.fpl.fs.fed.us/weib2.prog.html`.

Thus, there is no theoretical justification for performing a censored data fit if we believe that our model (e.g., a two-parameter Weibull) holds for the *whole* population. So, we must assume that an advocate of censored data fits believes that a two-parameter Weibull does not fit the whole population. What do they believe? We assume that they are thinking one of three things:

1. The population is a mixture — for *example*, 45% of MOR values are drawn from one normal (or Weibull or lognormal or ...) population and 55% of the MOR values are drawn from a different normal (or Weibull or lognormal or ...) population. (We considered a pseudo-truncated version of such a model in Verrill et al. (2018).)

2. The population is a "chimera" — for *example*, for data below x_c, the population's probability density function at x is the two-parameter Weibull density $\gamma^\beta \beta x^{\beta-1} \exp(-(\gamma x)^\beta)$, and for $x > x_c$, it is something else. Such a model might arise, for example, if we have competing failure modes that are associated with different portions of a strength distribution.

3. They have no proposed theoretical model for the distribution of the MOR population. They simply believe that, for practical purposes, they can perform a censored data two-parameter Weibull fit on some portion of the left-tail of the data and get fairly decent predictions of, say, the bottom 10 to 20 percent of the data.

The censored data methods described in section X2 of ASTM D5457 are not designed to handle case 1. In the notation of section X2 of ASTM D5457, the censored data techniques make use of the lowest n_c values in the data set, and the number n in the full random draw from the population. In the mixture case, we don't know the n associated with the "leftmost" sub-population. We know N, the total number of observations drawn from the various populations in the mixture. If we do a mixture analysis, we can *estimate* n as $\hat{p} \times N$ where $\hat{p}$ is our estimate of the proportion of the leftmost subpopulation in the mixture (even then, we would have to assume that the bottom n_c observations come solely from this leftmost population in the mixture). (Of course if we do a complete pseudo-truncated mixed analysis as we did in Verrill et al. (2018), we could calculate the complete resulting probability density function and predict probabilities of failure at various loads.)

In the chimera case, if we knew that the underlying probability density function were a two-parameter Weibull for x less than some x_c, then we could indeed, perform a censored data fit based on the x_i's that lie below x_c and the total number of observations in the sample. However, we have not yet seen a mechanistic model that would predict the cut-offs between various modes of failure

or the forms of the associated local strength distributions. We have certainly not seen anything that would link such differing models with the left tails of visual or MSR grades of lumber, and thus we do not believe that they motivate arguments for censored data fits.

Instead, we assume that proponents of modeling MOR distributions with censored data estimates of two-parameter Weibulls know that poor fits and probability plots are obtained when non-censored two-parameter Weibull fits are made to visual or MSR grade strength data, and they see that censored data fits yield left tail probability plots in which observed and predicted order statistics more closely align. They then argue that for reliability purposes, we are concerned about the "left tail" (leaving this loosely defined), not the whole distribution, so we need only get a "good fit" in some portion of the left tail. That is, they hope to use a censored data two-parameter Weibull fit as a good *interpolator* for some portion (bottom 10%?, bottom 20%?) of the left tail data.

The problem with this approach is that if we do not have a good mechanistic model for the generation of the data (for example, pseudo-truncation of a bivariate mill run distribution if we focus on MOE–MOR data or visual grade–MOR data, or pseudo-truncation of a tri-variate mill run distribution if we focus on visual grade–MOE–MOR data), but instead simply fit an interpolator to the left tail of a visual grade data set, we run into the weakness associated with all empirical models — *they tend to perform poorly when we attempt to apply them beyond the data used to fit them.* Thus, if we are interested in estimates of probabilities of failure on the order of 0.001 or even 0.0001, we might not do well if we base our predictions on interpolative fits to the bottom 10 or 20 percent of a sample of size 400, *even if* the bottom 10 or 20 percent of the observed and predicted data "align well" along the $y = x$ line in a probability plot. (Of course, we also will not do well if the load can fall above the 5 or 10 or 20 percent of the data fully included in the MOR censored-data fit, and the form of the appropriate probability density function in this area differs significantly from the two-parameter Weibull estimated from the censored-data fit.)

In support of this claim, we first note that, as stated in Section 4, Verrill et al. (2013, 2014) performed simulations that established that using censored data two-parameter Weibull methods on pseudo-truncated Weibull data can often lead to probability of failure estimates that are either well above or well below true values, and that this problem is much reduced when estimates are based on the correct pseudo-truncated model. A censored data two-parameter Weibull approach might yield probability plots that look better than those produced from a full data two-parameter Weibull approach, but *the censored data two-parameter Weibull approach still performs much more poorly than a correct pseudo-truncated Weibull approach.* Censored data two-parameter Weibull fits to pseudo-truncated Weibull data appear to lead to decreases in the biases of probability of failure estimates (compared to non-censored data two-parameter Weibull fits) but to *increases* in their variances. In short, increases in censoring lead to interpolation overfitting caused not by an increase in the number of parameters in the model, but by decreases in the range and effective number of data points. Again, for simulation details, see section 5.4 of Verrill et al. (2013).

Second, we note that if there really were an unvarying two-parameter Weibull that fit "low-tail data," then (depending upon where the "low-tail" is supposed to begin) censored-data maximum likelihood techniques would obtain similar (because maximum likelihood methods are asymptotically unbiased regardless of the degree of censoring) estimates whether we fit the bottom 20, 10, or 5 percent of the data (unless, of course, the "low-tail" doesn't begin until we are at or below some or all of these percentiles). However, as displayed in columns 4 and 5 of Table 2, for 12 of the 14 No. 2 "In-Grade type" data sets (and 4 of the 12 SS data sets), as censoring goes from none, to 20th percentile, to 10th percentile, to 5th percentile, the shape parameter obtained from a two-parameter Weibull censored data fit monotonically increases and the scale parameter monotonically decreases. (A listing of the Fortran program that produced the fits can be obtained at

http://www1.fpl.fs.fed.us/weib2.prog.html.) This is *not* what should occur theoretically or what does occur empirically when we use censored data two-parameter Weibull methods to fit *generated* two-parameter Weibulls — see columns 8 and 9 of Table 2 and see the discussion of table 3 in section 5.2 of Verrill et al. (2013). However, it *is* what we expect when we incorrectly use censored data two-parameter Weibull methods to fit *pseudo-truncated* Weibull data. See the discussion of table 5 in section 5.3 of Verrill et al. (2013). All of this suggests that "low-tail" In-Grade-type strength data are not well-modeled by the left tail of a two-parameter Weibull, despite, for example, two-parameter Weibull probability plots that might appear more nearly "linear" when they are based on censored data fits and only include low-tail data.

The Fortran program that produced Table 2 also produced plots that make our point visually. The full set of plots is available at http://www1.fpl.fs.fed.us/weib2.156plots.pdf. We display six of these plots as Figures 5 to 10.

In Figure 5 we plot a histogram of the 413 southern pine, 2x6, No. 2 In-Grade MOR values. We overlay this histogram with the estimated probability density function from a two-parameter Weibull maximum likelihood uncensored fit to the In-Grade MOR values, and with the estimated probability density functions from censored 20, censored 10, and censored 5 maximum likelihood fits to the In-Grade MOR values. This plot makes clear the systematic change in the fits as the censoring increases. *In contrast*, in Figure 6 we plot a histogram of 413 two-parameter Weibull values generated from the full data two-parameter Weibull fit of the 413 southern pine, 2x6, No. 2 In-Grade MOR values. We overlay this histogram with the estimated probability density function from a two-parameter Weibull maximum likelihood uncensored fit to the generated data, and with the estimated probability density functions from censored 20, censored 10, and censored 5 maximum likelihood fits to the generated data. These fits (to generated data that we *know* to be two-parameter Weibull) do not display the systematic changes with increased censoring displayed by the corresponding fits to In-Grade MOR values.

The dependency of left-tail predictions on the degree of censoring in our fitting procedures is further illustrated by Figures 7 to 10. In these figures we plot ordered data versus predicted ordered data. That is, these figures display two-parameter Weibull probability plots. They correspond, respectively, to full, censored 20, censored 10, and censored 5 two-parameter Weibull fits of the southern pine, 2x6, No. 2 In-Grade data. The horizontal lines in these plots mark the 20th, 10th, and 5th percentiles of the ordered data. In Figures 8–10 (corresponding to censored 20, censored 10, and censored 5 fits), the fits sharply deviate from the $y = x$ lines at or shortly above the data that is explicitly used in the fit. That is, 5% fits don't do a good job of predicting 10% data, and 10% fits don't do a good job of predicting 20% data. Locally good "interpolants" don't continue to perform well beyond the data used to produce the interpolants.

We admit that we are presenting plots from a case that does an especially good job of making our point visually. However, the plots in the 25 other cases make similar points. (All 156 plots can be viewed at http://www1.fpl.fs.fed.us/weib2.156plots.pdf.) In fact, for 7 of the 14 No. 2 censored 5 probability plots and 4 of the 12 SS censored 5 probability plots, there is a sharp bend at the 5th percentile of the data. For an additional 4 of the No. 2 censored 5 probability plots and an additional 4 of the SS censored 5 probability plots, this sharp bend is slightly higher (between the 5th and 10th percentiles of the data).

Given the observed poor predictions above the interpolated data, we see no reason to trust predictions for the important region below the interpolated data. As stated above, this empirical result is strongly bolstered by the simulation results reported in section 5.4 of Verrill et al. (2013), which establish the superiority of theoretically correct pseudo-truncated fits over theoretically incorrect censored data two-parameter Weibull fits. (Of course, if we do fits to sufficiently large data sets, relevant percentiles will lie within the interpolation region, and this will improve the performance

of interpolator-based percentile predictions.)

6 Load Distributions Rather than Fixed Loads

In Section 4 we evaluated estimated and empirical probabilities of failure at fixed loads (at approximate allowable properties). In this section we consider variable loads. We discuss calculations in which peak load distributions are modeled as lognormals. Such a model was suggested to us by a reviewer of Verrill et al. (2013). The reviewer of the 2013 paper suggested that we model the load distribution as a lognormal with coefficient of variation 0.3 that exceeds the allowable property with probability 0.02.

In our calculations the load was modeled as a lognormal with coefficient of variation 0.3, and we considered the cases in which the load exceeded the approximate allowable property (nonparametric estimate of the 5th percentile divided by 2.1) with probabilities 0.01, 0.02, 0.05, 0.10, and 0.2. The mathematical details of the calculations are provided in Appendix A. The results appear in Table 7. (A listing of the program that was used to produce Table 7 can be found at `http://www1.fpl.fs.fed.us/weib2.prog.html`.)

The results presented in Table 7 suggest that two-parameter Weibull fits to strength distributions can also yield inadequate estimates of failure probabilities when we incorporate load distributions into our calculations. For example, for load exceedance probabilities of 0.01 (the probability that the lognormal load exceeds the [approximate] allowable property is 0.01), and censored 20, censored 10, and censored 5 fits, the ratios of two-parameter Weibull-based estimates of the number of expected failures to data-based estimates of the number of expected failures are, respectively, $4.1/0.54 = 7.6$, $1.7/0.54 = 3.1$, and $1.2/0.54 = 2.2$. (Admittedly, these values do decline as the exceedance probability goes up. For example, for an exceedance probability of 0.02 rather than 0.01, the corresponding values are 5.8, 2.6, and 1.9.)

We note that the results suggest that failure rate biases decline as censoring increases, but we continue to emphasize that this is an improvement in interpolation rather than modeling and that simulations (again see section 5.4 of Verrill et al. (2013)) suggest that the apparent *reduction in the bias* of the estimate of failure rates given censored data fits is accompanied by an *increase in the variance* of the estimate of failure rates, and thus an *increased* chance of serious underestimates of failure rates.

7 Simulation of a Simple Lumber Assembly

In the course of studying repetitive member allowable property adjustments, Verrill and Kretschmann (2008) simulated simple seven-member load sharing assemblies. In their simulations, the seven members were connected to a "stiff diaphragm" and it was assumed that all seven assembly members were subjected to the same deflection. Given this simple model, the load on member i is given by

$$L_{\text{tot}} \times \text{MOE}_i/(\text{MOE}_1 + \ldots + \text{MOE}_7)$$

where L_{tot} denotes the total load on the assembly.

For the current paper, we used this simple *simulation* model to perform an initial investigation into the effect of using incorrect MOE and MOR models for visual grade data. In our simulations, the generating model for the mill run trivariate implicit visual grade/MOE/MOR data was a trivariate standard normal/general normal/two-parameter Weibull distribution. (For details on the manner in which we generated the trivariate distribution, see Appendix B.) In our simulation, visual grades were based on an implicit N(0,1) distribution. In particular, "No. 2" pieces of lumber

corresponded to pieces whose implicit visual grade values lay between the 40th and 80th percentiles of the N(0,1) distribution.

We also assumed that the (mill run) correlations between the implicit visual grade variable, the MOE, and the MOR were all 0.6, or all 0.7, or all 0.8; the (mill run) coefficients of variation of MOE were 0.25 or 0.30; and the (mill run) coefficients of variation of MOR were 0.30, 0.35, or 0.40.

We understand that the simulation model and the associated specific assumptions represent (at best) a very rough approximation to reality. We do note, however, that recent mill run MOE–MOR data (Verrill et al. 2017; Owens et al. 2018, 2019ab) suggest that the assumed correlations and coefficients of variation are within realistic ranges. (See Tables 8 and 9 for the observed sample correlations and coefficients of variation.)

In the simulations we compared the performances of correct and incorrect models for the simulated data.

Correct approach

We performed 100 trials in which we estimated the probabilities of failure of seven-member assemblies under the correct probability distribution. For our simplified model and No. 2 visual grade data, this is a trivariate truncated N(0,1)/pseudo-truncated N(μ, σ^2)/pseudo-truncated two-parameter Weibull distribution. (Choosing only No. 2 visual grade pieces amounts to truncating on the implicit visual grade variable, which leads to pseudo-truncated distributions for the correlated MOE and MOR variables.)

We used the methods described in section 5 of Verrill et al. (2012) to calculate the 5th percentile of the pseudo-truncated Weibull. (For simulation details, see Appendix B of the current paper and the FORTRAN program trisim.f available at `http://www1.fpl.fs.fed.us/weib2.prog.html`.) The "allowable property" was taken to be the 5th percentile of the pseudo-truncated Weibull divided by 2.1.

In each of the 100 trials, we generated 10,000 subtrials in which we drew triples from the trivariate N(0,1)/N(μ, σ^2)/two-parameter Weibull distribution (implicit visual grade variable/MOE/MOR) until we had seven pieces of lumber that had implicit visual grade values that lay between the 40th and 80th percentiles of a N(0,1) distribution. Thus the corresponding MOE,MOR pairs had a bivariate pseudo-truncated Gaussian, pseudo-truncated Weibull distribution (the tails of the resulting MOE and MOR distributions were not truncated, but they were "thinned" due to the truncation of the correlated implicit visual grade variable).

We counted the number of times (out of 10,000) that a seven-member assembly failed at the allowable property, that is the number of times that

$$\mathrm{MOR}_i < 7 \times (\text{allowable property}) \times \mathrm{MOE}_i/(\mathrm{MOE}_1 + \ldots + \mathrm{MOE}_7)$$

for at least one of the seven i's.

We report the mean of the resulting 100 failure rate estimates together with its standard error in the section of Table 10 that is headed by "Correct."

Incorrect approach

We also performed 100 trials under an incorrect model. The incorrect model is one that is probably still widely accepted. It assumes that the MOE values for a visual grade are normally distributed (not pseudo-truncated) and the MOR values are distributed as a two-parameter Weibull (not pseudo-truncated).

In each of the 100 trials, a sample of 400 pseudo-truncated MOE, pseudo-truncated MOR pairs was drawn. The pairs were generated as in the correct approach — implicit visual grade/MOE/MOR triples were drawn from a trivariate N(0,1)/N(μ, σ^2)/two-parameter Weibull distribution and a triple's MOE,MOR pair was kept if the implicit visual grade value lay between the 40th and 80th percentiles of a N(0,1) distribution. Triples were drawn until 400 MOE,MOR pairs were accepted. The resultant 400 pairs were then treated, incorrectly, as samples from a non-pseudo-truncated Gaussian distribution and a correlated non-pseudo-truncated two-parameter Weibull distribution. For each of the 400-pair data sets, standard methods were used to calculate parameter estimates $\hat{\mu}, \hat{\sigma}, \hat{\rho}, \hat{\gamma}, \hat{\beta}$ (mean, standard deviation, correlation, inverse scale, and shape parameters) under the incorrect assumption that the pairs were drawn from correlated (and non-pseudo-truncated) Gaussian and two-parameter Weibull distributions. Then 10,000 seven-member assemblies were generated from this incorrect bivariate Gaussian–Weibull fit to "No. 2 data," and counts of the number of failures (determined as in the subsection above) were obtained. We report the mean of the resulting 100 failure rate estimates together with its standard error in the section of Table 10 that is headed by "Incorrect."

From Table 10 we can see that the incorrect approach that ignores the pseudo-truncated nature of the MOE and MOR distributions when we truncate based on an implicit underlying visual grade variable yields inflated estimates of assembly failure rates. The inflation increases as MOR mill run coefficients of variation decrease and as mill run MOE/MOR correlations increase. Again, as in the case of single-member "In-Grade type" data sets discussed in Sections 2–4, incorrect models that ignore pseudo-truncation can lead to 10-fold overestimates of failure rates. In this case, the overestimates are overestimates of the failure rates of assemblies rather than of individual members.

8 Summary

Past theoretical work (Verrill et al. 2012, 2015) established that if a mill run implicit visual grade–MOR population or an MOE–MOR population has a bivariate normal–Weibull distribution, then the MOR distributions of visual grade or MSR sub-populations will be pseudo-truncated Weibulls (with thinned tails). Past empirical work (Verrill et al. 2013, 2014) and work reported in Section 3 confirm that MOR distributions of visual grades of lumber are not two-parameter Weibulls and do have thinned tails. Past simulation work (Verrill et al. 2013, 2014) suggested that two-parameter Weibull fits to pseudo-truncated Weibull data lead to inflated estimates of failure when loads are near allowable properties. Empirical work reported in Section 4 suggests that modeling visual grade MOR distributions with two-parameter Weibulls can lead to estimates of failure probabilities when loads are near allowable properties that are inflated by a factor of 9 (at least for No. 2 lumber and uncensored fits).

Past simulation work (Verrill et al. 2013, 2014) and work reported in Sections 5 and 6 suggest that, as one would expect, censored data two-parameter Weibull fits can also perform poorly when applied to data from pseudo-truncated distributions. In particular, they can lead to highly variable estimates of probabilities of failure and thus (for different data sets) to both serious overestimates and serious underestimates of probabilities of failure.

The preliminary work discussed in Section 7 suggests that the failure rate overestimates that occur for individual pieces of lumber when we ignore the pseudo-truncated nature of strength distributions also occur when we are working with simple assemblies of members.

Given these theoretical, empirical, and simulation results, we believe that additional full mill run data sets (see, for example, Verrill et al. (2017) and Owens et al. (2018, 2019ab)) need to be obtained, and additional pseudo-truncated distributions (see, for example, Verrill et al. (2018))

need to be developed in an attempt to identify alternatives to the two-parameter Weibull as a model for visual and MSR grade lumber strength distributions. We are engaged in such work.

However, given the fact that actual distributions may be complicated mixtures of base distributions that vary from mill to mill, region to region, time to time, size to size, and species to species, it may be that no satisfactory theoretical form(s) can be identified to form the basis of sophisticated reliability models that could yield improved design values.

We suspect that ultimately, if reliability engineers want to obtain fully accurate reliability estimates, they will need to develop detailed computer models that yield real-time, in-line estimates of lumber strength based on measurements of stiffness, specific gravity, knot size and location, slope of grain, and other strength predictors.

9 References

ASCE (1988). Load and Resistance Factor Design for Engineered Wood Construction: A Pre-Standard Report, prepared by the Task Committee on Load and Resistance Factor Design for Engineered Wood Construction, American Society of Civil Engineers, New York, NY, 177 pages.

ASTM (2013). Standard test methods for mechanical properties of lumber and wood-base structural material, D4761-13, *Annual Book of ASTM Standards*, ASTM International, West Conshohocken, PA.

ASTM (2016). Standard practice for establishing allowable properties for visually-graded dimension lumber from In-Grade tests of full-size specimens, D1990-16, *Annual Book of ASTM Standards*, ASTM International, West Conshohocken, PA.

ASTM (2019). Standard specification for computing reference resistence of wood-based materials and structural connections for load and resistance factor design, D5457-19b, *Annual Book of ASTM Standards*, ASTM International, West Conshohocken, PA.

Evans, J.W. and Green, D.W. (1988). Mechanical Properties of Visually Graded Dimension Lumber: Vol. 2. Douglas Fir-Larch. Publication PB-88-159-397, National Technical Information Service, Springfield, VA. p. 476.

Genz, A. and Bretz, F. (2009). *Computation of Multivariate Normal and t probabilities.* Lecture Notes in Statistics, Vol. 195, Springer-Verlag, Heidelberg.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2019). "mvtnorm: Multivariate Normal and t Distributions." R package version 1.0-11.
URL `http://CRAN.R-project.org/package=mvtnorm`

Green, D.W. and Evans, J.W. (1988). Mechanical Properties of Visually Graded Dimension Lumber: Vol. 4. Southern Pine. Publication PB-88-159-413, National Technical Information Service, Springfield, VA. p. 412.

Green, D.W., Shelley, B.E., and Vokey, H.P., eds. (1989). In-Grade Testing of Structural Lumber. Conference proceedings 47363. Forest Products Research Society, Madison, WI.

Kretschmann, D.E., Evans, J.W., and Brown, L. (1999). "Monitoring of Visually Graded Structural Lumber." Research Paper FPL-RP-576. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 18 pages.

Krit, M. (2014). EWGoF: Goodness-of-fit tests for the Exponential and two-parameter Weibull distributions. R package version 2.0. URL `http://CRAN.R-project.org/package=EWGoF`

Lawless, J.F. (2003). Statistical Models and Methods for Lifetime Data. John Wiley and Sons, Hoboken, New Jersey.

Owens, F.C, Verrill, S.P., Shmulsky, R., and Kretschmann, D.E. (2018), "Distributions of MOE and MOR in a Full Lumber Population," *Wood and Fiber Science*, **50**(3), pp. 265–279.

Owens, F.C, Verrill, S.P., Shmulsky, R., and Ross, R.J. (2019a), "Distributions of Modulus of Elasticity and Modulus of Rupture in Four Mill-Run Lumber Populations," *Wood and Fiber Science*, **51**(2), pp. 183–192.

Owens, F.C, Verrill, S.P., Shmulsky, R., and Ross, R.J. (2019b), "Distributions of Moduli of Elasticity and Rupture in Eight Mill-Run Lumber Populations (Four Mills at Two Times), under review

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org/`.

Verrill, S.P. and Kretschmann, D.E. (2008), "Material Variability and Repetitive Member Allowable Property Adjustments in Forest Products Engineering." Research Paper FPL-RP-649. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 36 pages.

Verrill, S.P. and Kretschmann, D.E. (2010), "Repetitive Member Factors for the Allowable Properties of Wood Products." Research Paper FPL-RP-657. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 38 pages.

Verrill, S.P., Evans, J.W., Kretschmann, D.E., and Hatfield, C.A. (2012). "Asymptotically Efficient Estimation of a Bivariate Gaussian–Weibull Distribution and an Introduction to the Pseudo-truncated Weibull." Research Paper FPL-RP-666. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 76 pages.

Verrill, S.P., Evans, J.W., Kretschmann, D.E., and Hatfield, C.A. (2013). "An Evaluation of a Proposed Revision of the ASTM D 1990 Grouping Procedure." Research Paper FPL-RP-671. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 34 pages.

Verrill, S.P., Evans, J.W., Kretschmann, D.E., and Hatfield, C.A. (2014), "Reliability Implications in Wood Systems of a Bivariate Gaussian–Weibull Distribution and the Associated Univariate Pseudo-truncated Weibull," *ASTM Journal of Testing and Evaluation*, **42**, Number 2, pages 412–419.

Verrill, S.P., Evans, J.W., Kretschmann, D.E., and Hatfield, C.A. (2015), "Asymptotically Efficient Estimation of a Bivariate Gaussian–Weibull Distribution and an Introduction to the Pseudo-truncated Weibull," *Communications in Statistics – Theory and Methods*, **44**, pages 2957–2975.

Verrill, S.P., Owens, F.C., Kretschmann, D.E., and Shmulsky, R. (2017). "Statistical Models for the Distribution of Modulus of Elasticity and Modulus of Rupture in Lumber with Implications for Reliability Calculations." Research Paper FPL-RP-692. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 51 pages.

Verrill, S.P., Owens, F.C., Kretschmann, D.E., and Shmulsky, R. (2018). "A Fit of a Mixture of Bivariate Normals to Lumber Stiffness–Strength Data." Research Paper FPL-RP-696. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 44 pages.

10 Appendix A — Calculations for Table 7

The calculations that yielded Table 7 were performed via a Fortran program that can be found at `http://www1.fpl.fs.fed.us/weib2.prog.html`. In this appendix we describe these calculations. None of the mathematics is novel. We describe it in some detail simply because this permits easy checking of our work. Some of the description replicates material found in the appendix to Verrill et al. (2013).

10.1 Obtaining the parameters of the lognormal load distribution

By definition, a random variable X is distributed as a lognormal(μ, σ^2) if $\ln(X)$ is distributed as a normal(μ, σ^2). Thus, characterizing the lognormal is equivalent to determining the two parameters μ and σ. For a lognormal distribution, it can be shown that the coefficient of variation, CV, is given by

$$\mathrm{CV} = \sqrt{\exp(\sigma^2) - 1}$$

or

$$\ln\left(1 + \mathrm{CV}^2\right) = \sigma^2 \tag{1}$$

Thus, for lognormals, the parameter σ can be determined from a knowledge of the (alternative) parameter CV. In the calculations that produced Table 7, CV = 0.3 so

$$\sigma^2 = \ln(1.09) \approx 0.0862 \tag{2}$$

Next, we show that we can obtain μ_i for the ith of the 26 In-Grade type data sets from σ and the probability, q, that the lognormal lies above the approximate allowable property, a_i (the nonparametric estimate of the 5th percentile divided by 2.1 of the ith In-Grade type data set). In Table 7, the q values appear in column 2.

We have

$$\mathrm{Prob}\left(\mathrm{LN}(\mu_i, \sigma^2) \le a_i\right) = 1 - q$$

or

$$\mathrm{Prob}\left(\mathrm{N}(\mu_i, \sigma^2) \le \ln(a_i)\right) = 1 - q$$

or

$$\mathrm{Prob}\left(\mathrm{N}(0,1) \le \frac{\ln(a_i) - \mu_i}{\sigma}\right) = 1 - q$$

or

$$\frac{\ln(a_i) - \mu_i}{\sigma} = \Phi^{-1}(1 - q)$$

where Φ denotes the N(0,1) cumulative distribution function. Thus,

$$\mu_i = \ln(a_i) - \sigma \times \Phi^{-1}(1 - q) \tag{3}$$

So, given the coefficient of variation of the lognormal load distribution (we assume it to be 0.3) and the probability, q (0.01, 0.02, 0.05, 0.1, or 0.2 in our calculations), that the lognormal load distribution exceeds the approximate allowable property, a_i, we can use (1) and (3) to calculate the mean and variance, μ_i and σ^2, needed to characterize the lognormal load distribution appropriate for data set i, and exceedance probability q.

10.2 Obtaining columns three and four of Table 7

Let

$$f_{\mathrm{LN},i}(y) = \frac{1}{\sqrt{2\pi}} \times \exp(-((\ln(y) - \mu_i)/\sigma)^2/2) \times \frac{1}{\sigma} \times \frac{1}{y} \tag{4}$$

denote the lognormal probability density function appropriate for data set i, a CV equal to 0.3, and exceedance probability q ($q \in \{0.01, 0.02, 0.05, 0.10, 0.20\}$). Then column 3 in Table 7 contains values of the form

$$\begin{aligned} \sum_{i=1}^{26} N_i \times \mathrm{Prob}(\mathrm{Load}_i > \mathrm{Strength}_i) &\approx \sum_{i=1}^{26} N_i \times \left(\sum_{j=1}^{N_i} \int_{w_{ij}}^{\infty} f_{\mathrm{LN},i}(y)dy \right) /N_i \\ &= \sum_{i=1}^{26} \sum_{j=1}^{N_i} \left(1 - \Phi\left((\log(w_{ij}) - \mu_i)/\sigma\right)\right) \end{aligned} \tag{5}$$

where Φ denotes the N(0,1) cumulative distribution function, w_{ij} is the jth observed strength value for In-Grade type data set i ($i \in \{1, \ldots, 26\}$), and N_i is the number of lumber specimens in data set i.

Here

$$\left(\sum_{j=1}^{N_i} \int_{w_{ij}}^{\infty} f_{\mathrm{LN},i}(y)dy \right) /N_i$$

is a data-based estimate of the probability that the lognormal load would be greater than the strength in the ith of the 26 cases.

Column 4 in Table 7 contains

$$\begin{aligned} &\sum_{i=1}^{26} N_i \int_0^{\infty} f_{\mathrm{LN},i}(x) \int_0^{x} \hat{\gamma}_i^{\hat{\beta}_i} \hat{\beta}_i w^{\hat{\beta}_i - 1} \exp\left(-(\hat{\gamma}_i w)^{\hat{\beta}_i}\right) dw\, dx \\ &= \sum_{i=1}^{26} N_i \int_0^{\infty} f_{\mathrm{LN},i}(x) \times \left(1 - \exp\left(-(\hat{\gamma}_i x)^{\hat{\beta}_i}\right)\right) dx \end{aligned} \tag{6}$$

where $\hat{\beta}_i$ and $\hat{\gamma}_i$ are the maximum likelihood estimates of the shape parameter and the inverse of the scale parameter for the ith of the 26 In-Grade type data sets and the corresponding censoring level. (The censoring level is indicated in column 1 of the table).

Here

$$\int_0^{\infty} f_{\mathrm{LN},i}(x) \int_0^{x} \hat{\gamma}_i^{\hat{\beta}_i} \hat{\beta}_i w^{\hat{\beta}_i - 1} \exp\left(-(\hat{\gamma}_i w)^{\hat{\beta}_i}\right) dw\, dx$$

is a theoretical estimate of the probability that the lognormal load would be greater than the strength in the ith of the 26 cases under the assumption (essentially) of a two-parameter Weibull left tail of the strength distribution (or at least of a two-parameter Weibull strength distribution in the region of practical overlap of the load and strength distributions).

11 Appendix B — Generating correlated triples of a N(0,1), a N(μ, σ^2), and a Weibull(γ, β) for the simulation of the strength properties of seven-member assemblies discussed in Section 7

The mathematical details described in this appendix might be of interest to only a few wood scientists. We provide it primarily for those interested in replicating or extending our simulations.

We use "Gaussian copula" methods to generate the correlated triples. We begin with a trivariate normal distribution that has covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$$

where, in our simulations, the correlation ρ is 0.6, 0.7, or 0.8. (An obvious extension is one in which the three correlations corresponding to the three possible pairs of variables are permitted to differ.)

Statisticians know that if

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \mathbf{\Sigma} \right) \tag{7}$$

then

1. $\Phi(X)$, $\Phi(Y)$, and $\Phi(Z)$ (where Φ is the N(0,1) cumulative distribution function) are uniformly distributed, and
2. if $V_1 = F_1^{-1}(\Phi(X))$, $V_2 = F_2^{-1}(\Phi(Y))$, $V_3 = F_3^{-1}(\Phi(Z))$, where F_1, F_2, F_3 are three cumulative distribution functions, then, for $i = 1, 2, 3$, V_i has cumulative distribution function F_i and the correlations among V_1, V_2, and V_3 are approximately ρ. (The correlation result is an empirical one, not a theoretical one, and is only approximately true. See, for example, appendix B of Verrill and Kretschmann (2010) and the associated table 10.)

Thus, in our simulation F_1 is a N(0,1) cumulative distribution function, F_2 is a N(μ, σ^2) cumulative distribution function, and F_3 is a Weibull(γ, β) cumulative distribution function.

To generate the original X, Y, Z, we begin by using a N(0,1) random number generator to generate three independent N(0,1)'s, T_1, T_2, T_3. Thus,

$$\begin{pmatrix} T_1 \\ T_2 \\ T_3 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right) \tag{8}$$

Statisticians know that if result (8) holds then

$$\mathbf{C} \begin{pmatrix} T_1 \\ T_2 \\ T_3 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \mathbf{C}\mathbf{C}^T \right) \tag{9}$$

Thus to generate X, Y, Z we need a C that satisfies

$$\mathbf{C}\mathbf{C}^T = \mathbf{\Sigma} \tag{10}$$

Such a $\mathbf{C}$ can be calculated as the product of the eigenvector matrix for $\mathbf{\Sigma}$ multiplied by the diagonal matrix whose diagonal elements are the square roots of the eigenvalues that correspond to the columns of the eigenvector matrix. Alternatively, one could obtain the Cholesky decomposition of $\mathbf{\Sigma}$, which would yield the necessary $\mathbf{C}$ directly. The R (R Core Team 2013) function eigen can be used to obtain the eigenvectors and corresponding eigenvalues of a matrix. The R (R Core Team 2013) function chol can be used to obtain a Cholesky decomposition of a symmetric matrix.

Our simulations were performed in a Fortran environment. If a scientist were willing to perform their simulations within the R (R Core Team 2013) environment, they could generate their multivariate normals via the mvtnorm package (Genz and Bretz 2009, Genz et al. 2019).

Data Set	Species	Lumber size	Grade	Sample size	Goodness-of-fit p-value	
					CVM	AD
In-Grade	DF	2×4	SS	414	.086	.033
	DF	2×8	SS	493	.472	.418
	DF	2×10	SS	414	.955	.771
	DF	2×4	2	386	.116	.066
	DF	2×8	2	1964	.001	.001
	DF	2×10	2	388	.001	.001
	HF	2×4	SS	428	.026	.002
	HF	2×8	SS	375	.034	.033
	HF	2×10	SS	368	.062	.049
	HF	2×4	2	406	.004	.002
	HF	2×8	2	372	.009	.004
	HF	2×10	2	361	.010	.002
	SP	2×4	SS	413	.029	.010
	SP	2×8	SS	626	.028	.023
	SP	2×10	SS	413	.002	.001
	SP	2×4	2	413	.001	.001
	SP	2×6	2	413	.001	.001
	SP	2×8	2	1367	.001	.001
	SP	2×10	2	412	.086	.113
2011	SP	2×4	SS	420	.254	.158
	SP	2×8	SS	409	.316	.226
	SP	2×10	SS	410	.043	.023
	SP	2×4	2	408	.001	.001
	SP	2×8	2	420	.196	.146
	SP	2×10	2	420	.091	.060
2014	SP	2×4	2	362	.807	.590

Table 1: p-values for Cramér–Von Mises and Anderson–Darling goodness-of-fit tests of two-parameter Weibull fits to In-Grade, 2011 SPIB "In-Grade", and 2014 SPIB resource monitoring program data. p-values listed as .001 might actually be lower.

Data Set	Data Cell	Fraction of data	Actual data				Generated data			
			shape	scale	skew	exc. kurt	shape	scale	skew	exc. kurt
In-Grade	DF2x4SS	full	4.62	11.07	-0.197	-0.176	4.65	11.15	-0.202	-0.172
		20%	5.56	10.56	-0.326	-0.033	4.96	10.74	-0.248	-0.126
		10%	6.07	10.10	-0.380	0.046	6.08	9.64	-0.382	0.048
		5%	8.58	8.46	-0.568	0.403	6.33	9.42	-0.406	0.087
	DF2x8SS	full	3.75	8.45	-0.034	-0.274	3.88	8.53	-0.062	-0.264
		20%	3.71	8.57	-0.024	-0.277	3.46	9.03	0.034	-0.288
		10%	4.24	7.81	-0.133	-0.226	3.62	8.65	-0.003	-0.282
		5%	4.41	7.50	-0.164	-0.204	3.98	7.83	-0.083	-0.254
	DF2x10SS	full	3.99	7.87	-0.084	-0.254	4.00	7.87	-0.088	-0.252
		20%	4.76	7.28	-0.220	-0.155	4.22	7.64	-0.129	-0.228
		10%	5.72	6.52	-0.343	-0.009	4.03	7.90	-0.094	-0.249
		5%	8.01	5.48	-0.534	0.328	5.40	6.27	-0.306	-0.058
	DF2x4_2	full	3.02	8.73	0.163	-0.272	3.21	8.84	0.104	-0.287
		20%	4.18	7.47	-0.121	-0.233	2.94	9.46	0.188	-0.263
		10%	4.50	7.16	-0.178	-0.192	3.35	8.37	0.065	-0.289
		5%	5.46	6.21	-0.314	-0.048	2.87	10.10	0.212	-0.252
	DF2x8_2	full	2.60	6.38	0.317	-0.181	2.62	6.28	0.309	-0.188
		20%	3.52	5.44	0.021	-0.287	2.78	6.01	0.245	-0.233
		10%	4.46	4.53	-0.172	-0.197	2.84	5.92	0.224	-0.245
		5%	4.67	4.40	-0.205	-0.169	3.02	5.44	0.162	-0.273
	DF2x10_2	full	2.44	5.68	0.388	-0.113	2.37	5.57	0.421	-0.076
		20%	4.96	3.80	-0.248	-0.126	2.07	6.06	0.583	0.160
		10%	5.59	3.56	-0.329	-0.028	2.10	5.92	0.566	0.129
		5%	5.68	3.50	-0.338	-0.015	2.24	5.34	0.489	0.012

Table 2a: In-Grade Douglas Fir data. 1.) Estimated shape and scale and corresponding skewness and excess kurtosis values for two-parameter Weibull fits (full, censored 20, censored 10, and censored 5) to six In-Grade Douglas Fir data sets, and 2.) Corresponding estimates for six data sets *generated* from the full sample two-parameter Weibull fits to the six In-Grade Douglas Fir data sets.

Additional details: Consider the In-Grade, DF2x4SS data cell. As noted in Table 1, there were 414 observations in this data cell. A full (uncensored) two-parameter Weibull fit to the 414 MOR values yields a shape estimate of 4.62 and a scale estimate of 11.07. A two-parameter Weibull with these shape and scale parameters has skewness -0.197 and excess kurtosis -0.176. The corresponding Generated data values were obtained by generating 414 observations from a two-parameter Weibull with shape 4.62 and scale 11.07. Fitting a two-parameter Weibull to these 414 *generated* MOR values yielded a shape estimate of 4.65 and a scale estimate of 11.15. A two-parameter Weibull with these shape and scale parameters has skewness -0.202 and excess kurtosis -0.172.

The 20% line in the DF2x4SS section was obtained by performing censored data two-parameter Weibull fits to the bottom 20% of the 414 DF2x4SS observations and to the bottom 20% of the 414 generated DF2x4SS values. And so on.

Data Set	Data Cell	Fraction of data	Actual data				Generated data			
			shape	scale	skew	exc. kurt	shape	scale	skew	exc. kurt
In-Grade	HF2x4SS	full	4.57	9.42	-0.190	-0.182	4.72	9.61	-0.213	-0.161
		20%	5.95	8.87	-0.368	0.027	4.66	9.45	-0.204	-0.170
		10%	5.53	9.24	-0.322	-0.038	4.56	9.56	-0.188	-0.184
		5%	5.71	8.95	-0.342	-0.010	3.50	12.18	0.026	-0.287
	HF2x8SS	full	4.24	7.51	-0.134	-0.225	4.38	7.55	-0.159	-0.207
		20%	4.78	7.24	-0.223	-0.152	4.24	7.66	-0.134	-0.225
		10%	5.80	6.49	-0.352	0.004	4.47	7.43	-0.174	-0.196
		5%	7.85	5.42	-0.524	0.308	8.03	5.15	-0.535	0.331
	HF2x10SS	full	4.25	6.42	-0.135	-0.224	4.31	6.48	-0.146	-0.217
		20%	4.73	6.39	-0.215	-0.160	3.44	7.19	0.041	-0.289
		10%	6.79	5.24	-0.446	0.156	3.17	7.71	0.114	-0.285
		5%	9.17	4.52	-0.599	0.475	8.46	3.61	-0.562	0.389
	HF2x4_2	full	2.90	7.34	0.203	-0.256	2.76	7.47	0.251	-0.230
		20%	4.36	6.09	-0.155	-0.210	2.44	7.86	0.384	-0.117
		10%	6.49	4.88	-0.420	0.111	2.03	10.11	0.613	0.212
		5%	5.48	5.48	-0.316	-0.046	2.30	8.14	0.453	-0.037
	HF2x8_2	full	2.72	5.86	0.268	-0.218	2.63	5.75	0.301	-0.194
		20%	3.70	5.07	-0.024	-0.277	2.44	5.94	0.386	-0.116
		10%	5.31	4.00	-0.296	-0.071	2.61	5.42	0.310	-0.187
		5%	6.18	3.60	-0.391	0.063	2.27	6.71	0.471	-0.013
	HF2x10_2	full	2.80	4.91	0.239	-0.237	2.90	4.98	0.203	-0.256
		20%	5.09	3.66	-0.267	-0.105	2.31	5.90	0.450	-0.039
		10%	5.76	3.40	-0.348	-0.002	2.62	5.02	0.308	-0.188
		5%	5.99	3.33	-0.372	0.034	2.83	4.44	0.227	-0.244

Table 2b: In-Grade HemFir data. 1.) Estimated shape and scale and corresponding skewness and excess kurtosis values for two-parameter Weibull fits (full, censored 20, censored 10, and censored 5) to six In-Grade HemFir datasets, and 2.) Corresponding estimates for six data sets *generated* from the full sample two-parameter Weibull fits to the six In-Grade HemFir data sets.

Additional details: Consider the In-Grade, HF2x4SS data cell. As noted in Table 1, there were 428 observations in this data cell. A full (uncensored) two-parameter Weibull fit to the 428 MOR values yields a shape estimate of 4.57 and a scale estimate of 9.42. A two-parameter Weibull with these shape and scale parameters has skewness -0.190 and excess kurtosis -0.182. The corresponding Generated data values were obtained by generating 428 observations from a two-parameter Weibull with shape 4.57 and scale 9.42. Fitting a two-parameter Weibull to these 428 *generated* MOR values yielded a shape estimate of 4.72 and a scale estimate of 9.61. A two-parameter Weibull with these shape and scale parameters has skewness -0.213 and excess kurtosis -0.161.

The 20% line in the HF2x4SS section was obtained by performing censored data two-parameter Weibull fits to the bottom 20% of the 428 HF2x4SS observations and to the bottom 20% of the 428 generated HF2x4SS values. And so on.

Data Set	Data Cell	Fraction of data	Actual data				Generated data			
			shape	scale	skew	exc. kurt	shape	scale	skew	exc. kurt
In-Grade	SP2x4SS	full	4.61	11.99	-0.196	-0.177	4.64	11.89	-0.201	-0.172
		20%	6.38	10.85	-0.410	0.094	5.12	11.50	-0.270	-0.102
		10%	5.65	11.56	-0.336	-0.019	6.17	10.46	-0.390	0.062
		5%	5.88	11.28	-0.361	0.017	6.86	9.87	-0.452	0.167
	SP2x8SS	full	4.63	9.26	-0.199	-0.175	4.51	9.23	-0.181	-0.190
		20%	4.95	9.06	-0.247	-0.127	4.49	9.13	-0.176	-0.193
		10%	4.34	9.87	-0.151	-0.213	4.57	9.05	-0.189	-0.183
		5%	4.50	9.59	-0.178	-0.192	4.01	10.07	-0.090	-0.251
	SP2x10SS	full	6.32	7.86	-0.405	0.085	6.13	7.88	-0.387	0.056
		20%	6.95	7.88	-0.459	0.180	6.82	7.73	-0.449	0.161
		10%	6.23	8.25	-0.396	0.071	5.64	8.47	-0.335	-0.020
		5%	8.02	7.20	-0.535	0.331	7.20	7.39	-0.479	0.217
	SP2x4_2	full	2.78	8.89	0.244	-0.234	2.96	8.71	0.182	-0.265
		20%	4.56	7.28	-0.188	-0.184	3.36	8.01	0.062	-0.289
		10%	5.69	6.42	-0.340	-0.013	3.24	8.30	0.095	-0.288
		5%	8.07	5.30	-0.538	0.338	4.15	6.40	-0.117	-0.236
	SP2x6_2	full	2.65	7.87	0.296	-0.198	2.56	7.72	0.334	-0.167
		20%	4.34	5.82	-0.152	-0.213	2.62	7.47	0.307	-0.190
		10%	6.27	4.68	-0.400	0.078	2.55	7.73	0.338	-0.163
		5%	9.51	3.84	-0.616	0.516	2.51	7.93	0.354	-0.148
	SP2x8_2	full	2.76	6.97	0.251	-0.230	2.72	6.85	0.268	-0.218
		20%	3.58	5.86	0.004	-0.284	3.01	6.44	0.164	-0.272
		10%	4.55	4.96	-0.187	-0.185	2.93	6.63	0.191	-0.261
		5%	4.96	4.66	-0.249	-0.126	3.46	5.50	0.034	-0.288
	SP2x10_2	full	3.47	6.44	0.033	-0.288	3.30	6.46	0.078	-0.289
		20%	4.24	5.65	-0.133	-0.225	3.77	5.92	-0.040	-0.272
		10%	5.17	4.93	-0.277	-0.093	3.34	6.62	0.066	-0.289
		5%	7.01	4.04	-0.464	0.189	3.49	6.36	0.028	-0.288

Table 2c: In-Grade Southern Pine data. 1.) Estimated shape and scale and corresponding skewness and excess kurtosis values for two-parameter Weibull fits (full, censored 20, censored 10, and censored 5) to seven In-Grade Southern Pine data sets, and 2.) Corresponding estimates for seven data sets *generated* from the full sample two-parameter Weibull fits to the seven In-Grade Southern Pine data sets.

Additional details: Consider the In-Grade, SP2x4SS data cell. As noted in Table 1, there were 413 observations in this data cell. A full (uncensored) two-parameter Weibull fit to the 413 MOR values yields a shape estimate of 4.61 and a scale estimate of 11.99. A two-parameter Weibull with these shape and scale parameters has skewness -0.196 and excess kurtosis -0.177. The corresponding Generated data values were obtained by generating 413 observations from a two-parameter Weibull with shape 4.61 and scale 11.99. Fitting a two-parameter Weibull to these 413 *generated* MOR values yielded a shape estimate of 4.64 and a scale estimate of 11.89. A two-parameter Weibull with these shape and scale parameters has skewness -0.201 and excess kurtosis -0.172.

The 20% line in the SP2x4SS section was obtained by performing censored data two-parameter Weibull fits to the bottom 20% of the 413 SP2x4SS observations and to the bottom 20% of the 413 generated SP2x4SS values. And so on.

Data Set	Data Cell	Fraction of data	Actual data				Generated data			
			shape	scale	skew	exc. kurt	shape	scale	skew	exc. kurt
2011	SP2x4SS	full	4.20	7.43	-0.126	-0.230	4.22	7.44	-0.130	-0.227
		20%	5.16	6.93	-0.276	-0.095	5.09	6.89	-0.267	-0.106
		10%	6.09	6.28	-0.382	0.049	5.17	6.82	-0.278	-0.093
		5%	5.16	6.99	-0.275	-0.096	5.14	6.82	-0.273	-0.099
	SP2x8SS	full	5.10	7.96	-0.268	-0.104	4.99	7.95	-0.253	-0.121
		20%	4.31	8.43	-0.145	-0.217	4.93	7.87	-0.244	-0.131
		10%	5.22	7.49	-0.283	-0.087	4.54	8.32	-0.185	-0.186
		5%	5.26	7.39	-0.289	-0.079	5.06	7.61	-0.262	-0.111
	SP2x10SS	full	5.93	7.34	-0.366	0.025	5.71	7.34	-0.343	-0.009
		20%	7.06	7.10	-0.468	0.196	7.44	6.84	-0.496	0.250
		10%	6.48	7.38	-0.420	0.110	6.95	6.98	-0.459	0.180
		5%	5.35	8.29	-0.300	-0.066	7.87	6.51	-0.525	0.310
	SP2x4_2	full	2.51	4.78	0.356	-0.146	2.68	4.94	0.285	-0.207
		20%	4.19	3.79	-0.124	-0.231	3.00	4.62	0.169	-0.270
		10%	6.45	2.98	-0.417	0.105	3.60	3.94	-0.001	-0.283
		5%	7.20	2.79	-0.478	0.216	2.98	4.91	0.175	-0.268
	SP2x8_2	full	2.95	5.88	0.184	-0.264	3.02	6.00	0.162	-0.273
		20%	3.52	5.31	0.021	-0.287	2.95	6.08	0.186	-0.264
		10%	4.22	4.62	-0.130	-0.228	2.82	6.26	0.231	-0.241
		5%	4.98	4.02	-0.251	-0.123	2.76	6.39	0.254	-0.228
	SP2x10_2	full	3.35	5.78	0.065	-0.289	3.31	5.82	0.074	-0.289
		20%	3.04	5.88	0.156	-0.275	3.68	5.56	-0.019	-0.279
		10%	3.61	4.99	-0.001	-0.283	3.17	6.33	0.115	-0.285
		5%	4.40	4.12	-0.162	-0.205	3.37	5.93	0.060	-0.289
2014	SP2x4_2	full	3.11	8.50	0.133	-0.281	3.29	8.36	0.079	-0.289
		20%	3.56	7.89	0.010	-0.285	3.80	7.71	-0.044	-0.271
		10%	5.60	5.79	-0.330	-0.027	4.10	7.24	-0.107	-0.241
		5%	7.95	4.78	-0.531	0.321	4.92	6.12	-0.243	-0.132

Table 2d: 2011 SPIB "In-Grade" and 2014 SPIB Southern Pine monitoring data. 1.) Estimated shape and scale and corresponding skewness and excess kurtosis values for two-parameter Weibull fits (full, censored 20, censored 10, and censored 5) to seven SPIB Southern Pine data sets, and 2.) Corresponding estimates for seven data sets *generated* from the full sample two-parameter Weibull fits to the seven SPIB Southern Pine data sets.

Additional details: Consider the 2011, SP2x4SS data cell. As noted in Table 1, there were 420 observations in this data cell. A full (uncensored) two-parameter Weibull fit to the 420 MOR values yields a shape estimate of 4.20 and a scale estimate of 7.43. A two-parameter Weibull with these shape and scale parameters has skewness -0.126 and excess kurtosis -0.230. The corresponding Generated data values were obtained by generating 420 observations from a two-parameter Weibull with shape 4.20 and scale 7.43. Fitting a two-parameter Weibull to these 420 *generated* MOR values yielded a shape estimate of 4.22 and a scale estimate of 7.44. A two-parameter Weibull with these shape and scale parameters has skewness -0.130 and excess kurtosis -0.227.

The 20% line in the SP2x4SS section was obtained by performing censored data two-parameter Weibull fits to the bottom 20% of the 420 2011, SP2x4SS observations and to the bottom 20% of the 420 generated SP2x4SS values. And so on.

Data Set	Species	Lumber size	Grade	Sample size	Predicted # failures				Observed # failures
					No cens	censored 20	censored 10	censored 5	
In-Grade	DF	2×4	SS	414	1.23	0.49	0.34	0.08	1
	DF	2×8	SS	493	2.32	2.33	1.60	1.52	1
	DF	2×10	SS	414	1.74	0.87	0.47	0.13	0
	DF	2×4	2	386	3.83	1.26	0.98	0.59	1
	DF	2×8	2	1964	26.99	10.43	5.79	5.08	1
	DF	2×10	2	388	6.88	0.79	0.52	0.52	0
	HF	2×4	SS	428	1.73	0.47	0.61	0.59	1
	HF	2×8	SS	375	1.26	0.73	0.36	0.13	0
	HF	2×10	SS	368	1.40	0.76	0.20	0.06	0
	HF	2×4	2	406	5.57	1.44	0.39	0.61	0
	HF	2×8	2	372	4.62	1.61	0.54	0.36	0
	HF	2×10	2	361	5.02	0.68	0.46	0.40	0
	SP	2×4	SS	413	1.56	0.34	0.54	0.48	0
	SP	2×8	SS	626	1.76	1.31	1.92	1.77	1
	SP	2×10	SS	413	0.41	0.20	0.34	0.13	0
	SP	2×4	2	413	5.85	0.97	0.45	0.12	1
	SP	2×6	2	413	5.04	1.12	0.32	0.05	0
	SP	2×8	2	1367	14.49	7.04	3.63	2.90	2
	SP	2×10	2	412	2.02	1.08	0.59	0.23	0
2011	SP	2×4	SS	420	1.98	0.83	0.49	0.80	0
	SP	2×8	SS	409	0.64	1.36	0.75	0.76	0
	SP	2×10	SS	410	0.72	0.27	0.39	0.70	1
	SP	2×4	2	408	7.54	1.37	0.30	0.21	0
	SP	2×8	2	420	3.71	2.15	1.34	0.97	0
	SP	2×10	2	420	1.67	2.63	1.84	1.30	1
2014	SP	2×4	2	362	2.56	1.64	0.42	0.11	0
Total					112.53	44.18	25.59	20.58	11
Inflation factor					10.2	4.0	2.3	1.9	

Table 3: Evidence from In-Grade, 2011 SPIB "In-Grade", and 2014 SPIB resource monitoring program data that two-parameter Weibull fits (both uncensored and censored) lead, on average, to inflated estimates of failure probabilities when loads are at the nonparametric 5th/1.9.

Data Set	Species	Lumber size	Grade	Sample size	Predicted # failures				Observed # failures
					No cens	Censored 20	Censored 10	Censored 5	
In-Grade	DF	2×4	SS	414	0.77	0.28	0.19	0.04	0
	DF	2×8	SS	493	1.60	1.61	1.05	0.98	1
	DF	2×10	SS	414	1.17	0.54	0.27	0.06	0
	DF	2×4	2	386	2.84	0.83	0.63	0.34	1
	DF	2×8	2	1964	20.84	7.34	3.71	3.18	0
	DF	2×10	2	388	5.41	0.48	0.30	0.29	0
	HF	2×4	SS	428	1.10	0.26	0.35	0.33	1
	HF	2×8	SS	375	0.82	0.45	0.20	0.06	0
	HF	2×10	SS	368	0.91	0.48	0.10	0.02	0
	HF	2×4	2	406	4.17	0.93	0.20	0.35	0
	HF	2×8	2	372	3.52	1.11	0.32	0.19	0
	HF	2×10	2	361	3.80	0.41	0.26	0.22	0
	SP	2×4	SS	413	0.98	0.18	0.31	0.26	0
	SP	2×8	SS	626	1.11	0.80	1.25	1.13	1
	SP	2×10	SS	413	0.22	0.10	0.18	0.06	0
	SP	2×4	2	413	4.44	0.62	0.25	0.05	1
	SP	2×6	2	413	3.87	0.72	0.17	0.02	0
	SP	2×8	2	1367	11.00	4.92	2.30	1.76	2
	SP	2×10	2	412	1.43	0.71	0.35	0.12	0
2011	SP	2×4	SS	420	1.30	0.50	0.27	0.48	0
	SP	2×8	SS	409	0.38	0.89	0.45	0.45	0
	SP	2×10	SS	410	0.40	0.13	0.20	0.41	1
	SP	2×4	2	408	5.88	0.90	0.16	0.10	0
	SP	2×8	2	420	2.76	1.52	0.88	0.59	0
	SP	2×10	2	420	1.19	1.94	1.28	0.83	1
2014	SP	2×4	2	362	1.88	1.15	0.24	0.05	0
Total					83.80	29.79	15.85	12.39	9
Inflation factor					9.3	3.3	1.8	1.4	

Table 4: Evidence from In-Grade, 2011 SPIB "In-Grade", and 2014 SPIB resource monitoring program data that two-parameter Weibull fits (both uncensored and censored) lead, on average, to inflated estimates of failure probabilities when loads are at the nonparametric 5th/2.1.

Data Set	Species	Lumber size	Grade	Sample size	Predicted # failures				Observed # failures
					No cens	Censored 20	Censored 10	Censored 5	
In-Grade	DF	2×4	SS	414	0.51	0.17	0.11	0.02	0
	DF	2×8	SS	493	1.14	1.15	0.71	0.66	0
	DF	2×10	SS	414	0.81	0.35	0.16	0.03	0
	DF	2×4	2	386	2.16	0.57	0.42	0.21	0
	DF	2×8	2	1964	16.47	5.33	2.47	2.08	0
	DF	2×10	2	388	4.34	0.31	0.18	0.18	0
	HF	2×4	SS	428	0.73	0.15	0.21	0.20	1
	HF	2×8	SS	375	0.56	0.29	0.12	0.03	0
	HF	2×10	SS	368	0.62	0.31	0.05	0.01	0
	HF	2×4	2	406	3.21	0.63	0.11	0.21	0
	HF	2×8	2	372	2.75	0.79	0.19	0.11	0
	HF	2×10	2	361	2.95	0.26	0.15	0.13	0
	SP	2×4	SS	413	0.65	0.10	0.18	0.16	0
	SP	2×8	SS	626	0.73	0.51	0.84	0.75	1
	SP	2×10	SS	413	0.12	0.05	0.10	0.03	0
	SP	2×4	2	413	3.45	0.41	0.15	0.03	0
	SP	2×6	2	413	3.05	0.49	0.10	0.01	0
	SP	2×8	2	1367	8.56	3.55	1.52	1.12	2
	SP	2×10	2	412	1.04	0.48	0.22	0.06	0
2011	SP	2×4	SS	420	0.89	0.31	0.15	0.30	0
	SP	2×8	SS	409	0.24	0.60	0.28	0.28	0
	SP	2×10	SS	410	0.23	0.07	0.11	0.25	1
	SP	2×4	2	408	4.69	0.62	0.09	0.05	0
	SP	2×8	2	420	2.11	1.10	0.60	0.38	0
	SP	2×10	2	420	0.88	1.48	0.92	0.56	1
2014	SP	2×4	2	362	1.42	0.83	0.14	0.03	0
Total					64.31	20.90	10.31	7.85	6
Inflation factor					10.7	3.5	1.7	1.3	

Table 5: Evidence from In-Grade, 2011 SPIB "In-Grade", and 2014 SPIB resource monitoring program data that two-parameter Weibull fits (both uncensored and censored) lead, on average, to inflated estimates of failure probabilities when loads are at the nonparametric 5th/2.3.

Coefficient of variation	Fitting method	"Parameter"	Root mean squared error		
			Uncensored	Censored 20	Censored 10
.1	regression	1/scale	0.438E−03	0.252E−02	0.462E−02
		shape	0.636E+00	0.181E+01	0.250E+01
		q05	0.116E+00	0.137E+00	0.134E+00
		q01	0.151E+00	0.250E+00	0.271E+00
		q001	0.183E+00	0.369E+00	0.446E+00
		pfallow	0.554E−05	0.267E−04	0.536E−04
	MLE	1/scale	0.431E−03	0.163E−02	0.301E−02
		shape	0.466E+00	0.145E+01	0.215E+01
		q05	0.949E−01	0.130E+00	0.131E+00
		q01	0.118E+00	0.210E+00	0.237E+00
		q001	0.139E+00	0.296E+00	0.366E+00
		pfallow	0.336E−05	0.120E−04	0.178E−04
.2	regression	1/scale	0.180E−02	0.103E−01	0.187E−01
		shape	0.306E+00	0.883E+00	0.131E+01
		q05	0.966E−01	0.112E+00	0.109E+00
		q01	0.108E+00	0.173E+00	0.189E+00
		q001	0.106E+00	0.208E+00	0.255E+00
		pfallow	0.321E−03	0.723E−03	0.905E−03
	MLE	1/scale	0.175E−02	0.666E−02	0.131E−01
		shape	0.239E+00	0.669E+00	0.107E+01
		q05	0.822E−01	0.107E+00	0.107E+00
		q01	0.890E−01	0.148E+00	0.170E+00
		q001	0.860E−01	0.169E+00	0.218E+00
		pfallow	0.235E−03	0.503E−03	0.633E−03
.3	regression	1/scale	0.384E−02	0.245E−01	0.429E−01
		shape	0.201E+00	0.571E+00	0.799E+00
		q05	0.780E−01	0.922E−01	0.920E−01
		q01	0.747E−01	0.123E+00	0.132E+00
		q001	0.588E−01	0.118E+00	0.139E+00
		pfallow	0.118E−02	0.233E−02	0.259E−02
	MLE	1/scale	0.374E−02	0.166E−01	0.304E−01
		shape	0.149E+00	0.467E+00	0.680E+00
		q05	0.645E−01	0.891E−01	0.904E−01
		q01	0.597E−01	0.109E+00	0.121E+00
		q001	0.462E−01	0.101E+00	0.124E+00
		pfallow	0.829E−03	0.180E−02	0.200E−02

Table 6: Root mean squared errors of estimates of quantities associated with two-parameter Weibulls. Based on samples of size 400 drawn from known two-parameter Weibulls (coefficents of variation in column 1, scales chosen so that the variances of the generating Weibulls will be 1). Estimated quantities: 1/scale — inverse of the Weibull scale parameter; shape — Weibull shape parameter; q05 — 0.05 quantile of the Weibull; q01 — 0.01 quantile of the Weibull; q001 — 0.001 quantile of the Weibull; pfallow — probability that a draw from the Weibull lies below its 5th percentile divided by 2.1. Note that the root mean squared errors increase as censoring increases (except in some of the q05 cases). Also note that root mean squared errors for regression estimators

are larger than the root mean squared errors for the corresponding maximum likelihood estimators.

Coefficient of variation	Fitting method	"Parameter"	Root mean squared error		
			Uncensored	Censored 20	Censored 10
.4	regression	1/scale	0.728E−02	0.394E−01	0.722E−01
		shape	0.134E+00	0.381E+00	0.552E+00
		q05	0.564E−01	0.655E−01	0.649E−01
		q01	0.455E−01	0.732E−01	0.793E−01
		q001	0.283E−01	0.560E−01	0.681E−01
		pfallow	0.187E−02	0.317E−02	0.344E−02
	MLE	1/scale	0.716E−02	0.264E−01	0.489E−01
		shape	0.106E+00	0.304E+00	0.452E+00
		q05	0.492E−01	0.640E−01	0.636E−01
		q01	0.385E−01	0.653E−01	0.751E−01
		q001	0.237E−01	0.487E−01	0.622E−01
		pfallow	0.147E−02	0.251E−02	0.293E−02
.5	regression	1/scale	0.118E−01	0.674E−01	0.121E+00
		shape	0.112E+00	0.313E+00	0.452E+00
		q05	0.451E−01	0.526E−01	0.510E−01
		q01	0.308E−01	0.498E−01	0.540E−01
		q001	0.151E−01	0.299E−01	0.365E−01
		pfallow	0.282E−02	0.453E−02	0.472E−02
	MLE	1/scale	0.115E−01	0.445E−01	0.858E−01
		shape	0.845E−01	0.243E+00	0.371E+00
		q05	0.379E−01	0.506E−01	0.509E−01
		q01	0.251E−01	0.433E−01	0.500E−01
		q001	0.122E−01	0.256E−01	0.335E−01
		pfallow	0.216E−02	0.363E−02	0.407E−02

Table 6 continued: Root mean squared errors of estimates of quantities associated with two-parameter Weibulls. Based on samples of size 400 drawn from known two-parameter Weibulls (coefficents of variation in column 1, scales chosen so that the variances of the generating Weibulls will be 1). Estimated quantities: 1/scale — inverse of the Weibull scale parameter; shape — Weibull shape parameter; q05 — 0.05 quantile of the Weibull; q01 — 0.01 quantile of the Weibull; q001 — 0.001 quantile of the Weibull; pfallow — probability that a draw from the Weibull lies below its 5th percentile divided by 2.1. Note that the root mean squared errors increase as censoring increases (except in some of the q05 cases). Also note that root mean squared errors for regression estimators are larger than the root mean squared errors for the corresponding maximum likelihood estimators.

Data used in Weibull fit	Probability that the load is above the allowable property	Data based expected failures	Weibull fit based expected failures	Column 4 divided by column 3
All	0.01	0.54	17.1	31.7
	0.02	0.96	21.4	22.3
	0.05	2.16	30.1	13.9
	0.10	4.27	41.0	9.6
	0.20	9.39	59.8	6.4
bottom 20%	0.01	0.54	4.1	7.6
	0.02	0.96	5.6	5.8
	0.05	2.16	8.9	4.1
	0.10	4.27	13.5	3.2
	0.20	9.39	22.4	2.4
bottom 10%	0.01	0.54	1.7	3.1
	0.02	0.96	2.5	2.6
	0.05	2.16	4.3	2.0
	0.10	4.27	7.2	1.7
	0.20	9.39	13.4	1.4
bottom 5%	0.01	0.54	1.2	2.2
	0.02	0.96	1.8	1.9
	0.05	2.16	3.3	1.5
	0.10	4.27	5.7	1.3
	0.20	9.39	11.2	1.2

Table 7: Evidence that Weibull fits yield inadequate estimates of failure probabilties *even when* we incorporate load distributions and censoring into the calculations.

Season	Mill	Correlations		
		MOE/MOR	ecomp/MOR	dire/MOR
Summer	1	0.71	0.59	0.61
	2	0.78	0.71	0.72
	3	0.78	0.71	0.72
	4	0.76	0.65	0.68
Winter	1	0.79	0.70	0.70
	2	0.78	0.68	0.71
	3	0.80	0.72	0.74
	4	0.76	0.65	0.67

Table 8: Correlations between stiffness measures and MOR for eight mill run samples of 200 pieces of lumber. MOE, ecomp and dire are stiffness measures. For details, see Verrill et al. (2017), Owens et al. (2018, 2019 a,b).

Season	Mill	Coefficients of Variation			
		MOE	ecomp	dire	MOR
Summer	1	0.25	0.23	0.23	.31
	2	0.30	0.31	0.31	.41
	3	0.26	0.25	0.25	.39
	4	0.26	0.24	0.24	.41
Winter	1	0.25	0.23	0.23	.38
	2	0.26	0.24	0.26	.39
	3	0.31	0.31	0.31	.44
	4	0.25	0.23	0.25	.37

Table 9: For eight mill run samples of 200 pieces of lumber, coefficients of variation for three stiffness measures and MOR. For details, see Verrill et al. (2017), Owens et al. (2018, 2019 a,b).

MOR CofV	MOE CofV	Stiffness–Strength Correlation											
		0.6				0.7				0.8			
		Correct		Incorr		Correct		Incorr		Correct		Incorr	
		mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
0.30	0.25	.0024	.0005	.0064	.0019	.0007	.0003	.0035	.0011	.00008	.00009	.0017	.0009
	0.30	.0023	.0005	.0057	.0020	.0007	.0002	.0030	.0012	.00006	.00007	.0013	.0006
0.35	0.25	.0052	.0007	.0117	.0035	.0020	.0004	.0075	.0025	.00032	.0002	.0039	.0015
	0.30	.0050	.0007	.0107	.0030	.0018	.0004	.0072	.0024	.00026	.0002	.0037	.0014
0.40	0.25	.0094	.0009	.0192	.0049	.0042	.0006	.0131	.0037	.0010	.0003	.0077	.0020
	0.30	.0087	.0011	.0164	.0046	.0038	.0006	.0122	.0035	.0008	.0003	.0070	.0022

Table 10: Estimated assembly failure rates under correct and incorrect models. "CofV" — coefficient of variation.

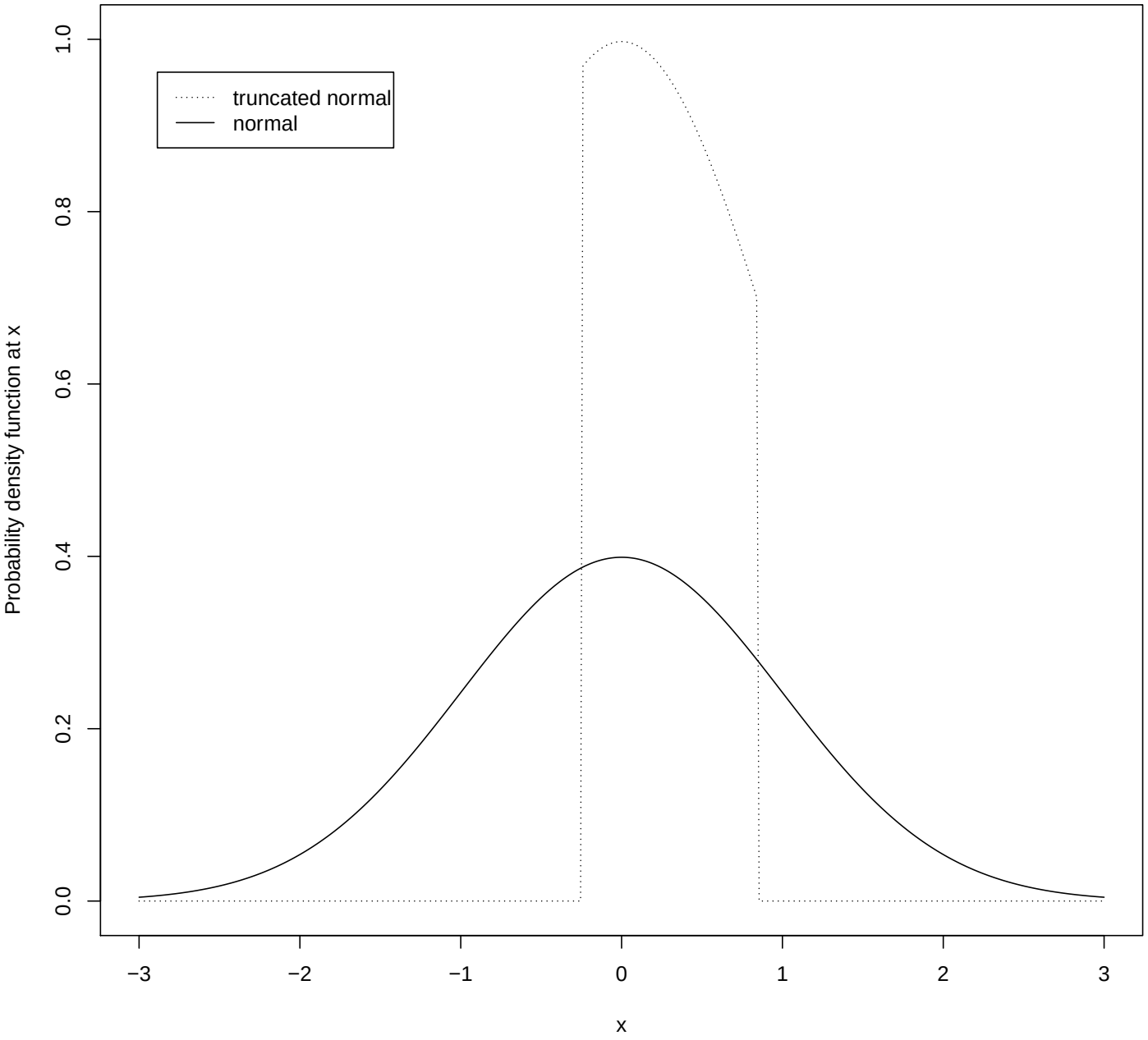


Figure 1: A plot of a N(0,1) (standard normal) probability density function (pdf) overlaid by the probability density function of a N(0,1) that has been truncated at its 40th and 80th percentiles.

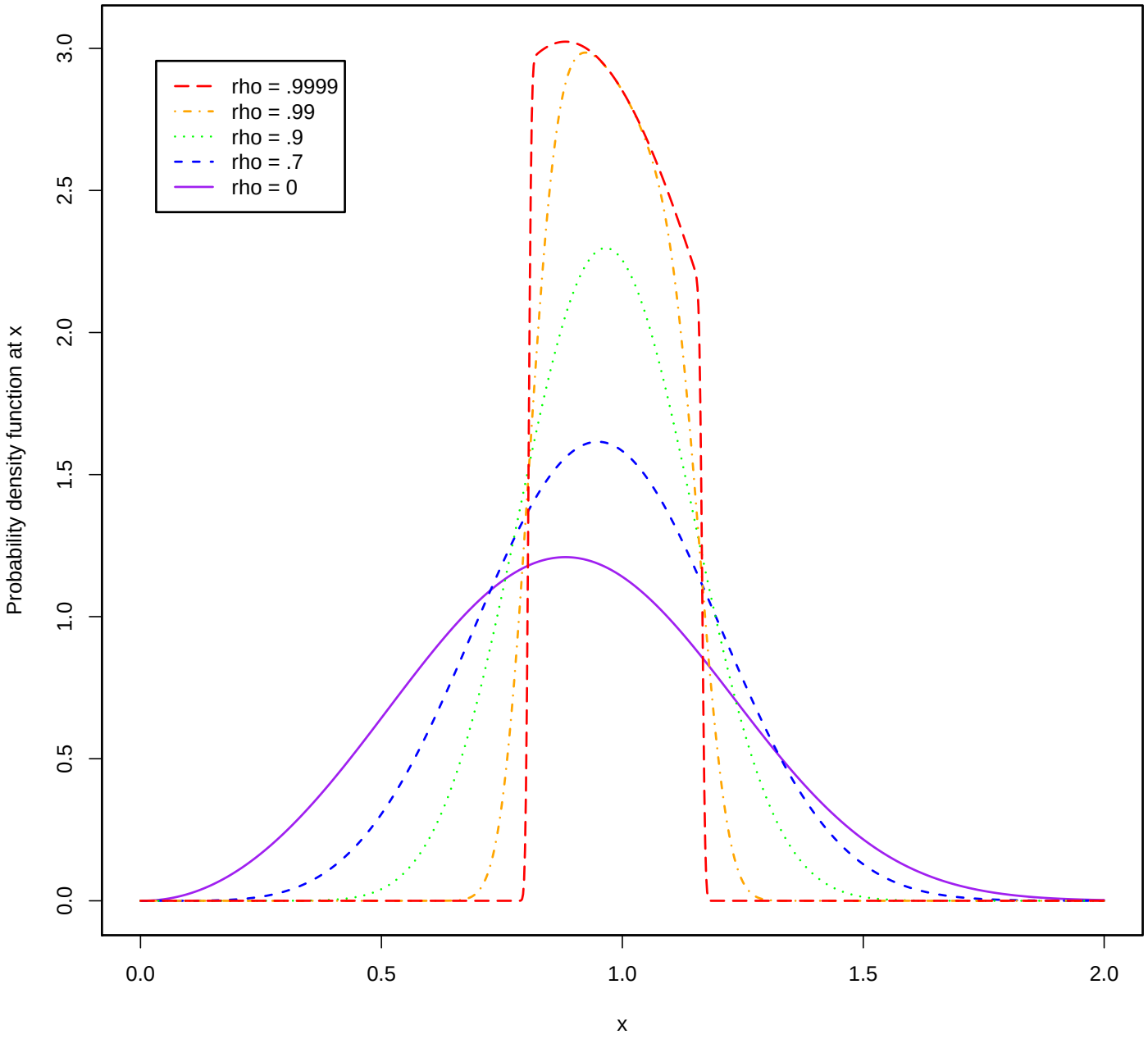


Figure 2: A plot of a two-parameter Weibull (shape = 3.1, scale = 1) pdf overlaid by the pdfs of the corresponding pseudo-truncated Weibulls for the cases in which the correlation parameters for the generating bivariate Gaussian–Weibulls are 0.7,0.9,0.99, and 0.9999

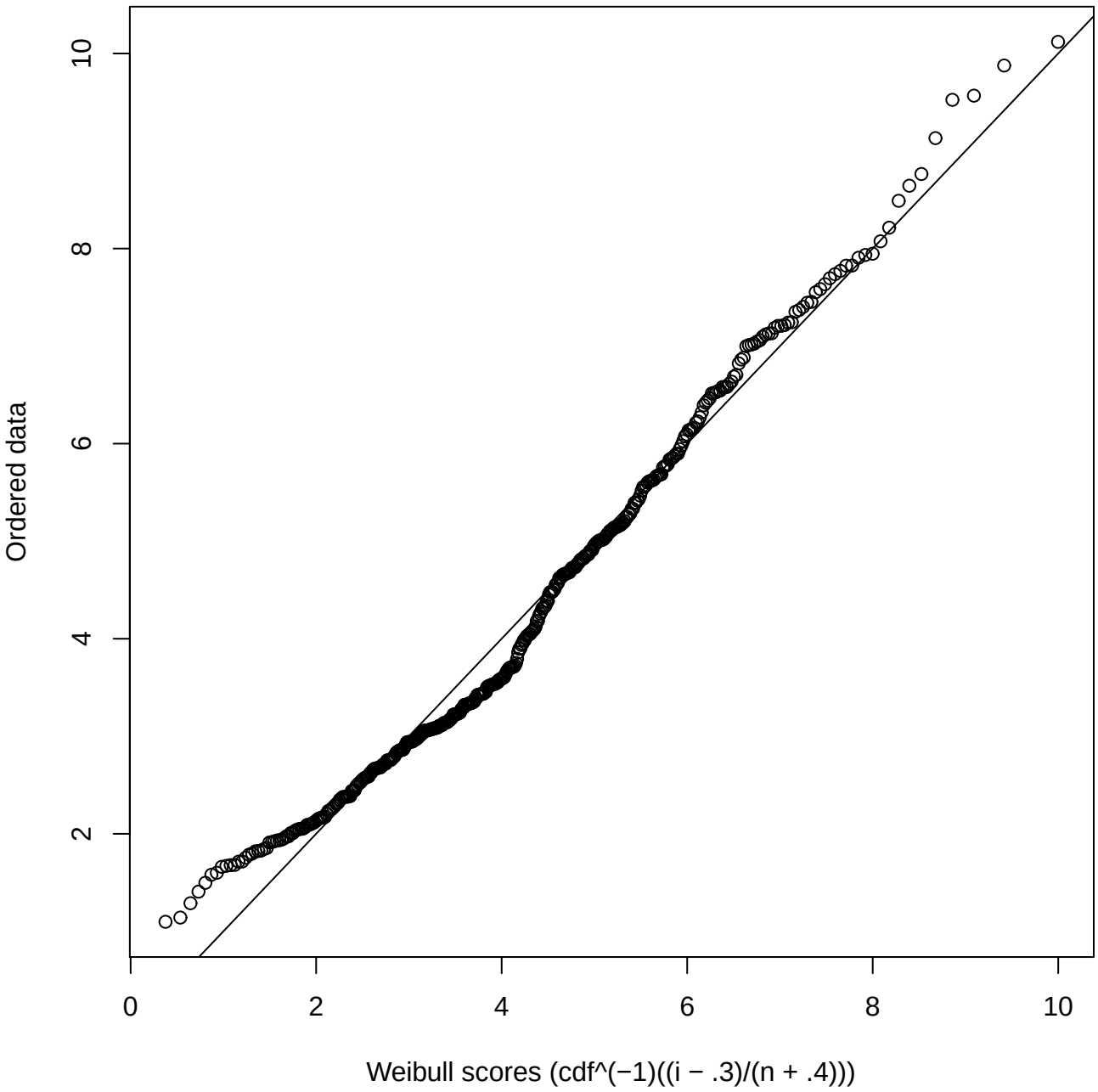


Figure 3: Weibull probability plot for the 2011 SPIB "In-Grade" Southern Pine, 2x4, No. 2 data. Ordered empirical data versus expected ordered data under a two-parameter Weibull model. The straight line is the $y = x$ line.

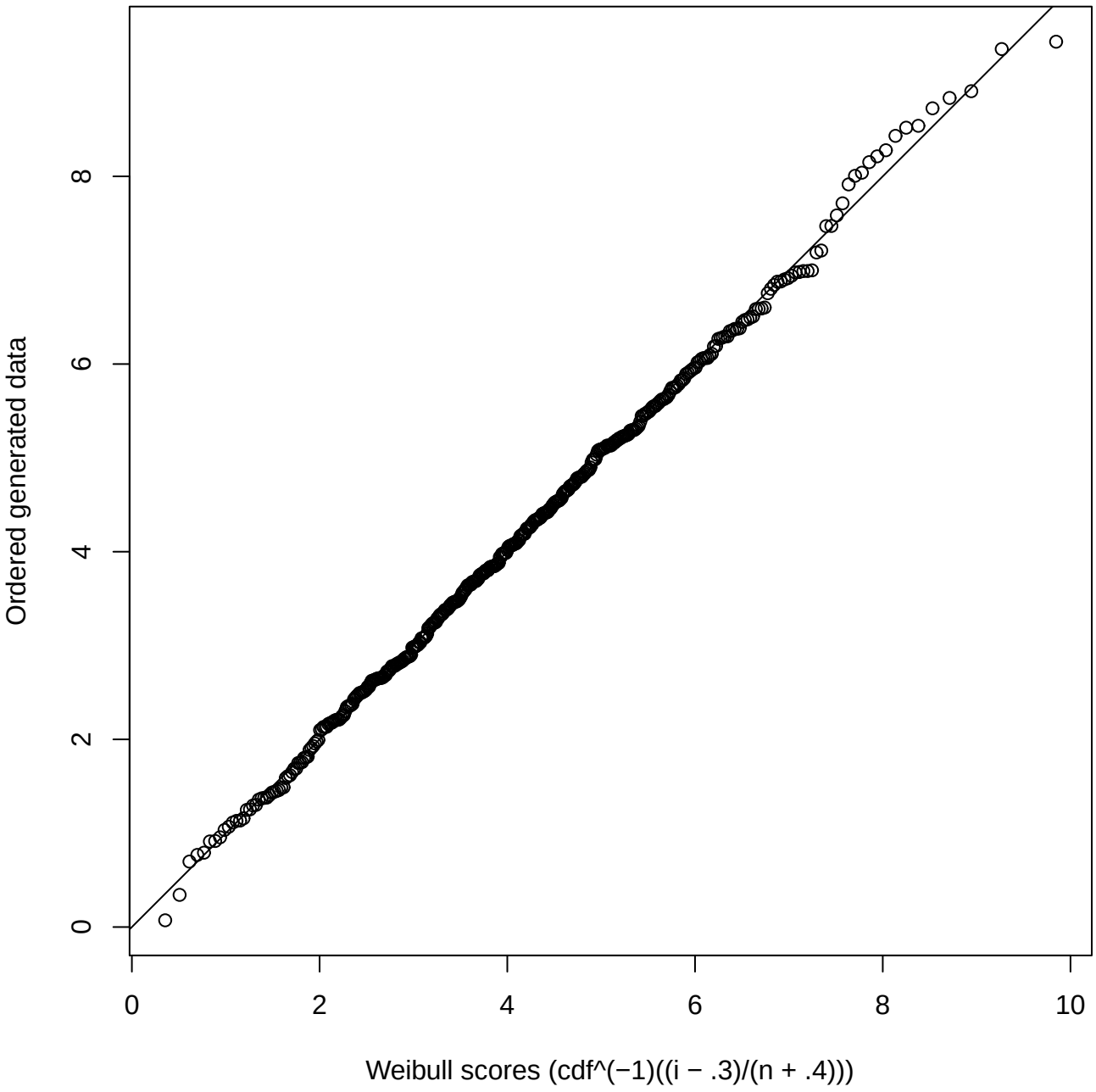


Figure 4: Weibull probability plot of generated data. Ordered generated data versus expected ordered data under a two-parameter Weibull model. The generated data was generated from a two-parameter Weibull maximum likelihood fit of the 2011 SPIB "In-Grade" Southern Pine, 2x4, No. 2 data. The straight line is the $y = x$ line.

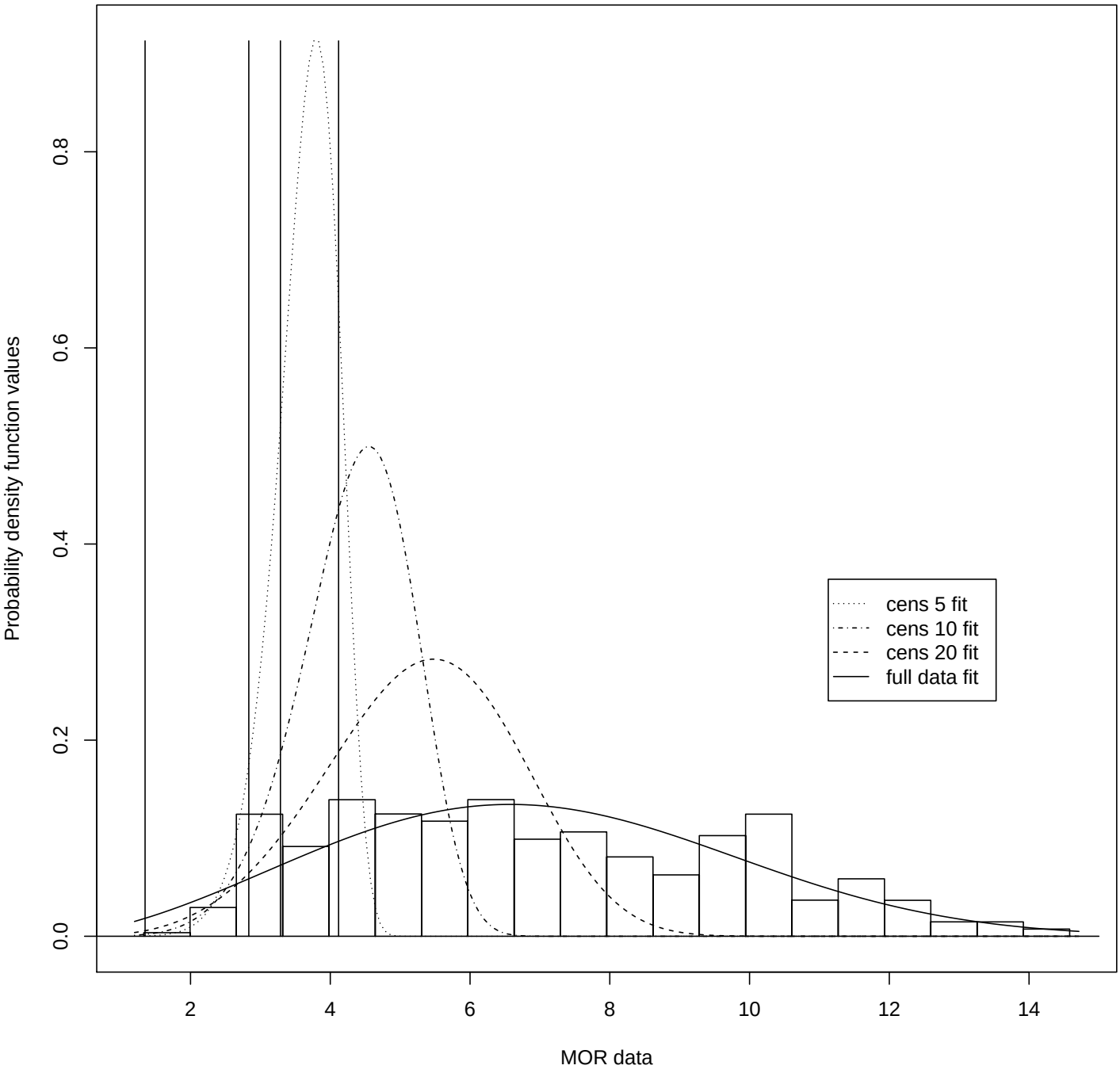


Figure 5: Histogram of In-Grade Southern Pine, 2x6, No. 2 data. It is overlaid with a two-parameter Weibull full data fit (solid line), a two-parameter Weibull censored 20 fit (dashed line), a two-parameter Weibull censored 10 fit (dot-dashed line), and a two-parameter Weibull censored 5 fit (dotted line). Solid vertical lines are plotted at the nonparametric 5th/2.1, and at the nonparametric 5th, 10th, and 20th percentiles of the data.

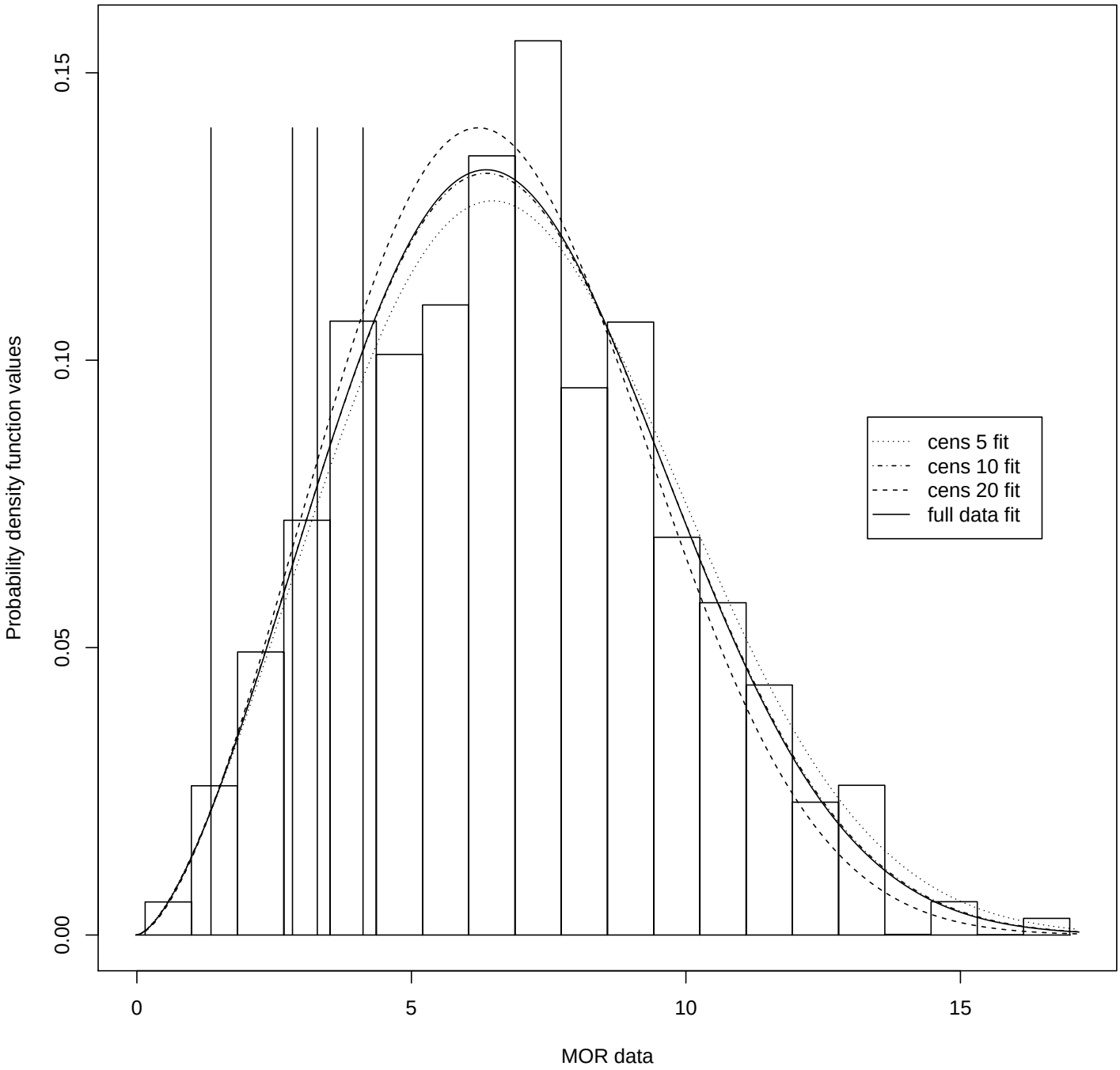


Figure 6: Histogram of generated In-Grade Southern Pine, 2x6, No. 2 data. It is overlaid with a two-parameter Weibull full data fit (solid line), a two-parameter Weibull censored 20 fit (dashed line), a two-parameter Weibull censored 10 fit (dot-dashed line), and a two-parameter Weibull censored 5 fit (dotted line). Solid vertical lines are plotted at the nonparametric 5th/2.1, and at the nonparametric 5th, 10th, and 20th percentiles of the data.

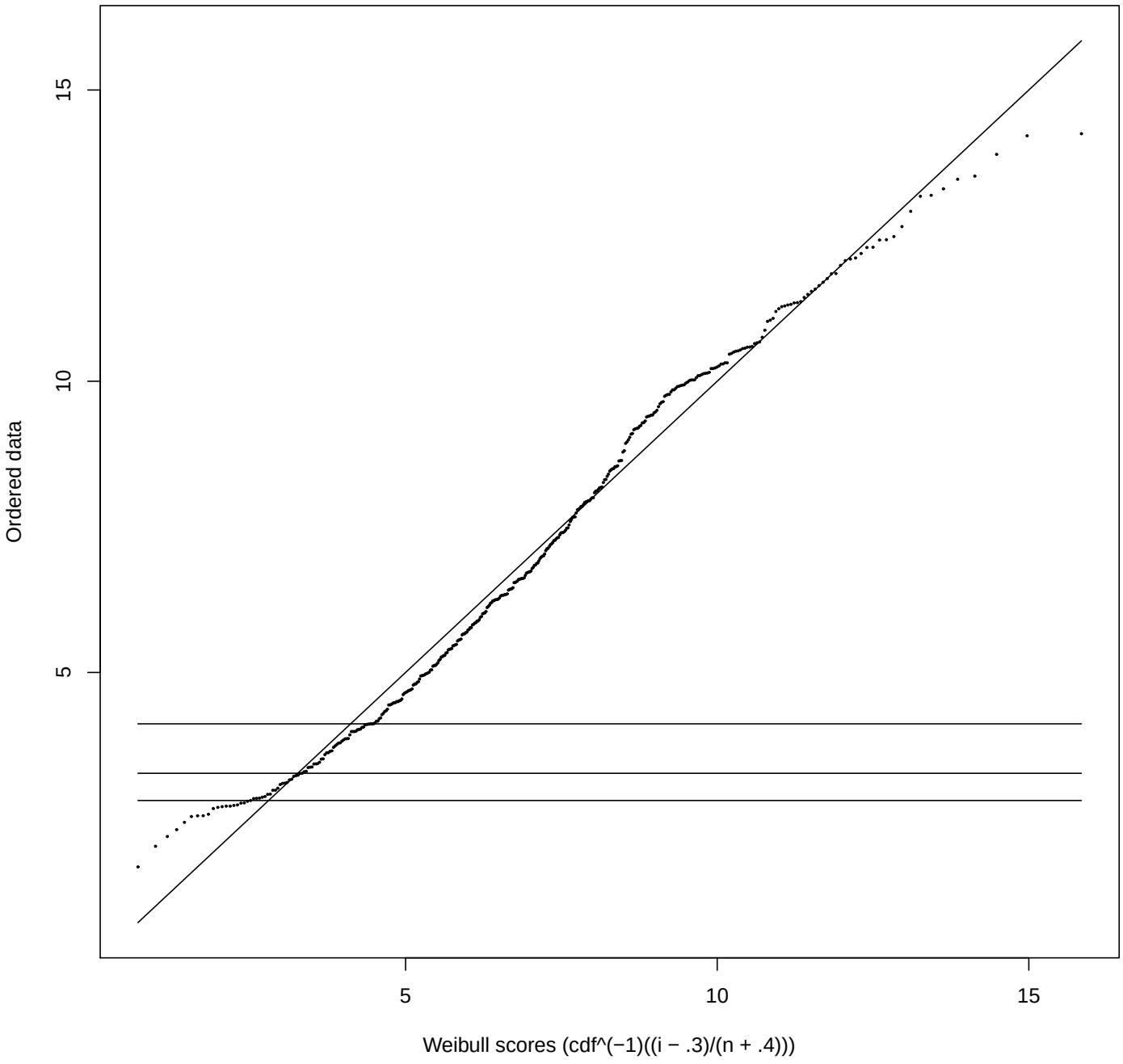


Figure 7: Weibull probability plot of In-Grade Southern Pine, 2x6, No. 2 data. Ordered observed data versus expected ordered data under a full-data two-parameter Weibull model fit. The straight non-horizontal line is the $y = x$ line. The horizontal lines are at the 0.20, 0.10, and 0.05 quantiles of the data.

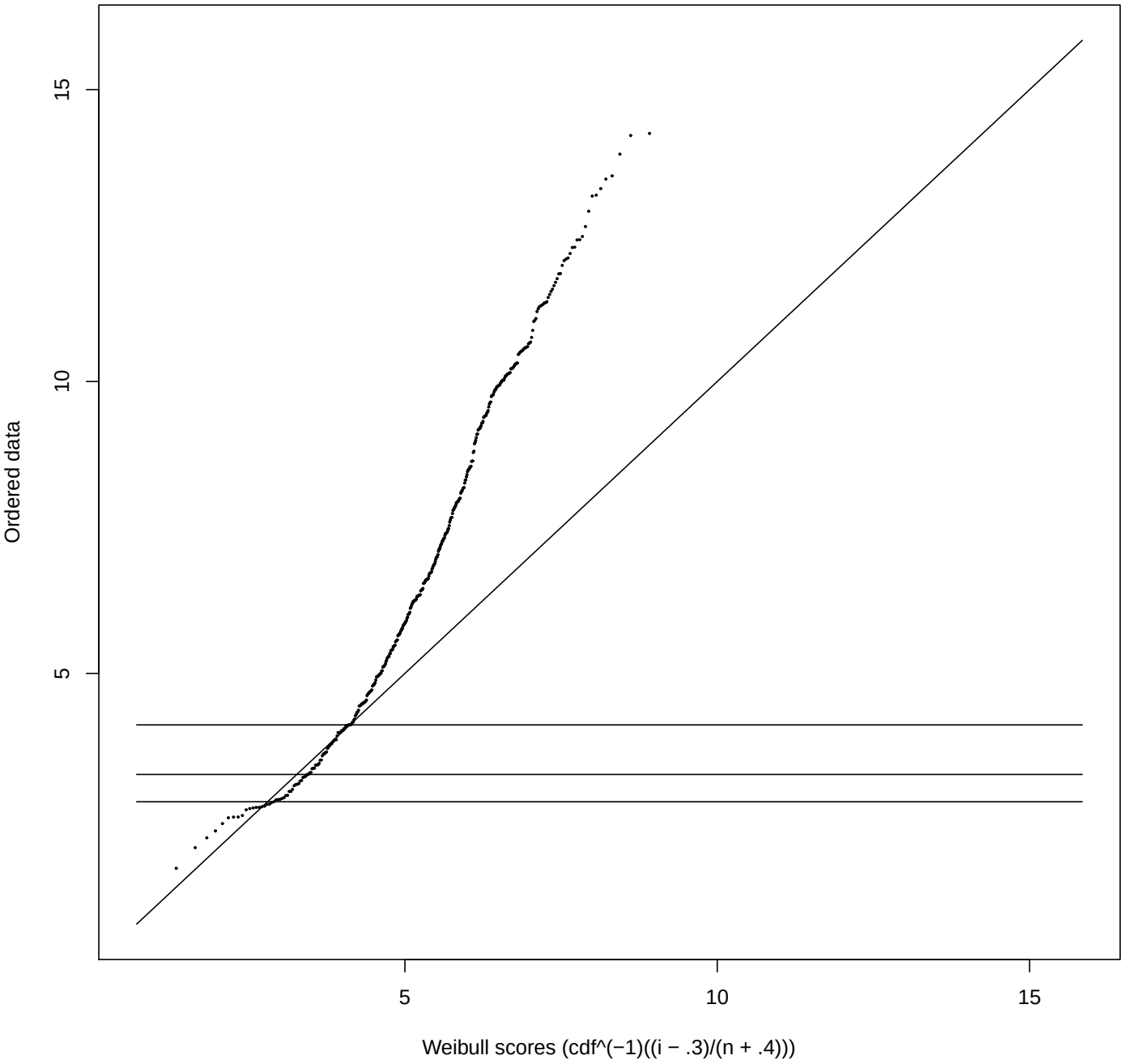


Figure 8: Weibull probability plot of In-Grade Southern Pine, 2x6, No. 2 data. Ordered observed data versus expected ordered data under a censored 20 fit. The straight non-horizontal line is the $y = x$ line. The horizontal lines are at the 0.20, 0.10, and 0.05 quantiles of the data.

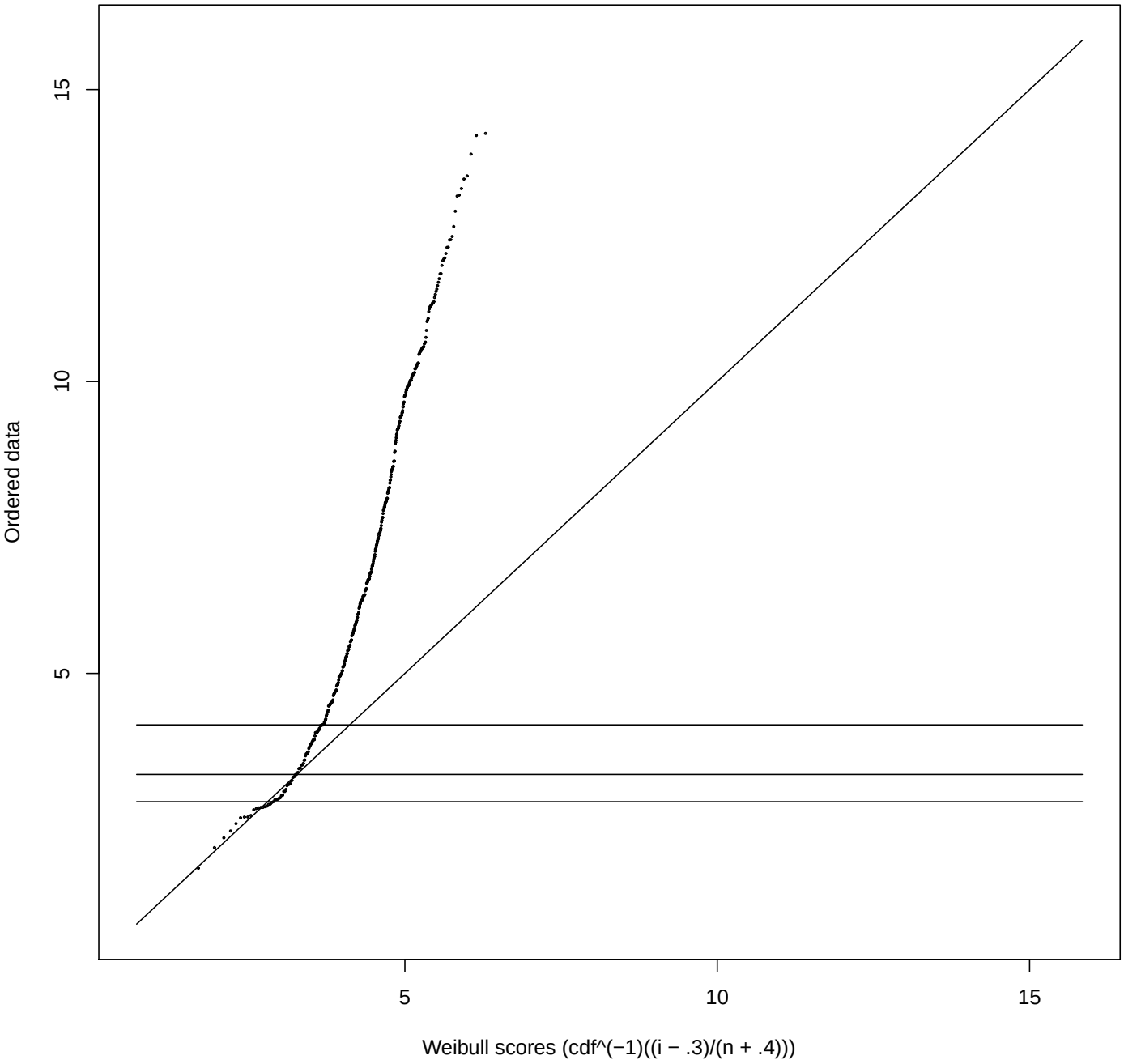


Figure 9: Weibull probability plot of In-Grade Southern Pine, 2x6, No. 2 data. Ordered observed data versus expected ordered data under a censored 10 fit. The straight non-horizontal line is the $y = x$ line. The horizontal lines are at the 0.20, 0.10, and 0.05 quantiles of the data.

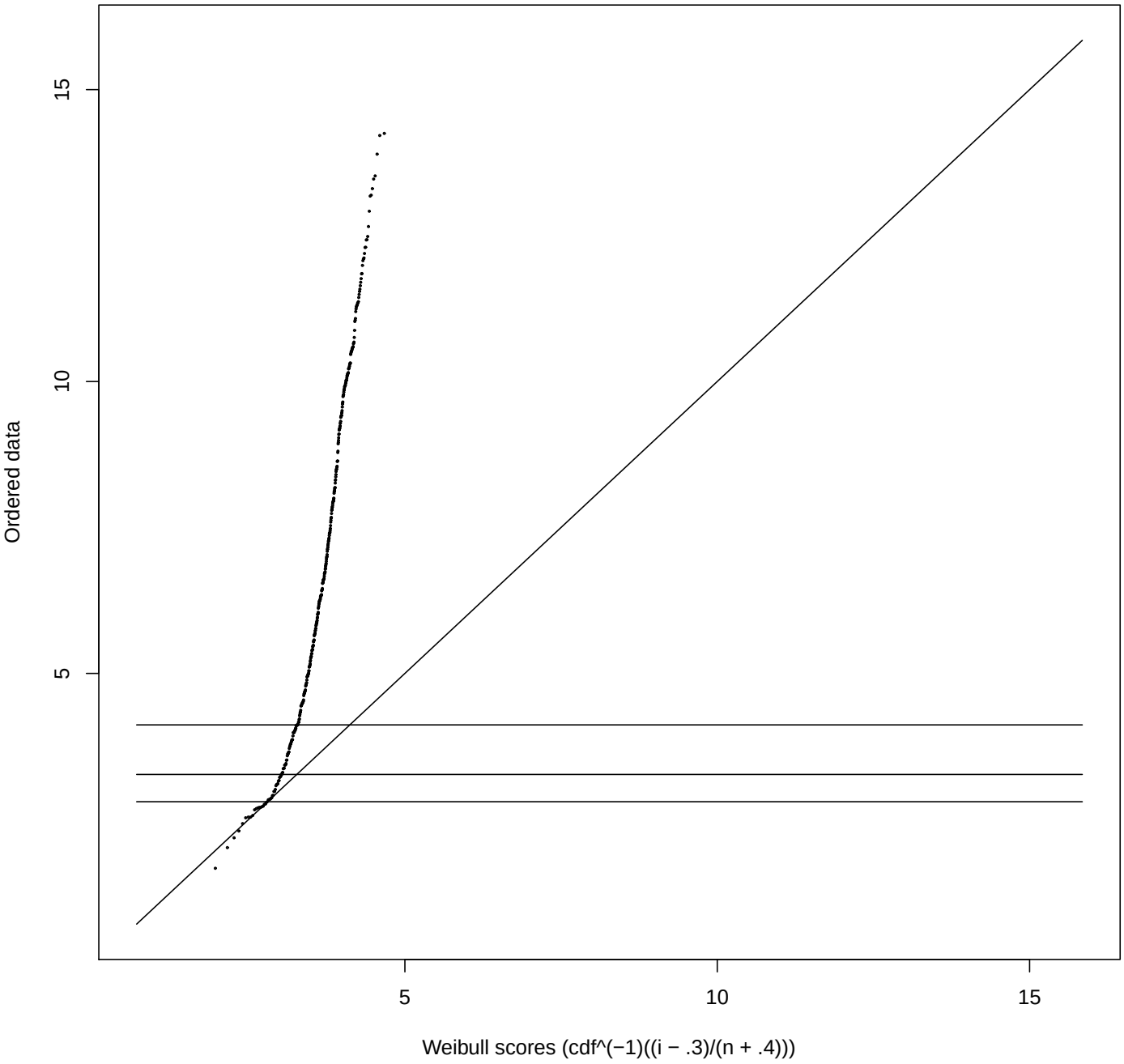


Figure 10: Weibull probability plot of In-Grade Southern Pine, 2x6, No. 2 data. Ordered observed data versus expected ordered data under a censored 5 fit. The straight non-horizontal line is the $y = x$ line. The horizontal lines are at the 0.20, 0.10, and 0.05 quantiles of the data.