JOURNAL OF MATERIALS SCIENCE **39** (2004) 1441–1444

Best estimate of Weibull modulus obtained using linear least squares analysis: An improved empirical correction factor

I. J. DAVIES
Department of Mechanical Engineering, Curtin University of Technology, GPO Box U1987, Perth, WA 6845, Australia
E-mail: daviesi@vesta.curtin.edu.au

The strength data of brittle materials, e.g., ceramics, have long been known to exhibit a wider degree of scatter when compared to that of ductile materials, resulting in implications for the reliability of structural ceramics [1]. This scatter of material strength is often found to follow the well-known empirical relationship proposed by Weibull [2]:

$$P = 1 - \exp\left\{-\int_X \left(\frac{\sigma - \sigma_u}{\sigma_o}\right)^m dX\right\} \quad (1)$$

where P is the probability of failure at a stress, σ, m is known as the Weibull modulus, σ_u is the stress at which $P = 0$, X is the strength-limiting dimension of the material (usually either volume or surface area), and σ_o is a normalizing factor. The value of σ_u, the stress threshold below which failure does not occur, is normally taken to be zero [3] and thus, for specimens of constant geometry, Equation 1 may be rewritten as:

$$P = 1 - \exp\left\{-\left(\frac{\sigma}{\sigma_o}\right)^m\right\} \quad (2)$$

Although for many years this relationship was applied to experimental data on an empirical basis, the work of Jayatilaka and Trustrum [4] has provided a theoretical background with the form of Equation 2 being related to the probability density, $f(a)$, of flaw sizes within the material with $f(a)$ being approximated by [5]:

$$f(a) = \frac{c^{n-1}}{(n-2)!} a^{-n} e^{-c/a} \quad (3)$$

where n is the rate at which $f(a)$ tends to zero for $a \gg c/n$ and c is a scaling parameter. Jayatilaka and Trustrum showed that, assuming there to be a large number of randomly oriented flaws, m and n are related through [4]:

$$m = 2n - 2 \quad (4)$$

Thus, the scatter in strength data is directly related to the shape of the flaw size distribution. The most common method of obtaining σ_o and m from a series of data has been to rank the σ data from smallest to largest and assign P values according to:

$$P = \frac{i}{N+1} \quad (5)$$

where i is the rank and N is the total number of specimens. Equation 2 may then be linearized using the form:

$$y = A + Bx \quad (6)$$

where

$$y = \ln\left[\ln\left(\frac{1}{1-P}\right)\right], \quad A = -m \ln \sigma_o,$$
$$B = m, \quad \text{and} \quad x = \ln \sigma.$$

From this, the best estimates of σ_o and m, respectively σ_o^* and m^*, can be obtained using the linear least squares (LLS) method, i.e.,

$$A = \frac{\Sigma x^2 \Sigma y - \Sigma x \Sigma xy}{N \Sigma x^2 - (\Sigma x)^2} \quad (7a)$$

$$B = \frac{N \Sigma xy - \Sigma x \Sigma y}{N \Sigma x^2 - (\Sigma x)^2} \quad (7b)$$

with Σ, x, and y in Equations 7a and b being abbreviations for $\sum_{i=1}^{N}$, x_i, and y_i, respectively. The application of LLS analysis to Weibull curve fitting remains popular amongst scientists and engineers due to its simplicity.

Whereas the majority of researchers wishing to obtain m from a set of data using the LLS technique assume $m = m^*$, this is not generally the case as the mean value of m^*, m^*_{mean}, is biased with $m^*_{mean} < m$. The bias increases with decreasing values of N such that, for a value of $m = 5$ typical for glass and ceramic fibers [6–9], m^*_{mean}/m is approximately 0.96 for $N = 150$ and 0.863 for $N = 10$ [10].

The reason for the bias in m^*_{mean} is directly related to an assumption utilized in the LLS technique that the error for each set of $\{x, y\}$ data points in Equation 6 is distributed symmetrically about a mean value according to a Gaussian function [11]. While usually true for the $\{\sigma, P\}$ data set, application of the non-linear transformation on Equation 2 renders this assumption invalid.

Attempts to reduce the bias in m^*_{mean} have followed several approaches including: (i) use of different relationships to calculate P [12, 13], (ii) use of weighting factors for each data point [14–16], (iii) use of

parameter estimation methods other than LLS (such as "method of moments" and "maximum likelihood") [16, 17], and (iv) omission of certain data points [18]. While these methods have resulted in less biased values of m^*_{mean}, they each suffer from one or more of the following deficiencies: (i) the improved m^*_{mean} value still exhibits a degree of bias, (ii) similarly invalid assumptions are utilized as for the case of LLS, (iii) lack of physical meaning, and (iv) difficulty of use. It is perhaps as a result of these issues that utilization of bias reduction methods in the literature appears to be limited at best.

In light of this, the author recently proposed [10] that the bias in values of m^* obtained by the LLS technique could be significantly reduced using a simple empirical correction factor. The rationale behind the work was that the bias in m^*_{mean} obtained using the LLS technique can be accurately determined for any value of N and m given a sufficient number of data sets. Compared to the techniques mentioned above, such a procedure would have the advantages of: (i) significantly reducing the bias, (ii) providing a best estimate of m with minimum effort, and (iii) being easily applied to both future and historic data sets.

In the initial investigation [10], an empirical correction factor was determined for the case of $m = 5$ and $10 < N < 150$ with only a single value of m being chosen due to a lack of computing resources. Despite this, the correction factor was shown to be successful in reducing the bias in m^*_{mean} to within less than 1% of m and would thus appear to be a good candidate for reducing the bias in m^* data. In contrast to the approach of the previous investigation, the aim of the present work will be to obtain an empirical correction factor that is applicable over a wide range of useful m values.

The empirical correction factor in the present investigation was calculated from a large number of data sets generated using a Monte Carlo method in the range $1 < m < 50$ and $10 < N < 50$. For each data set, N values between 0 and 1 were randomly chosen $\{P_1, P_2, P_3, \ldots P_N\}$ and used to generate values of σ, $\{\sigma_1, \sigma_2, \sigma_3, \ldots \sigma_N\}$, according to Equation 2 (with σ_o being nominally set to unity). The σ data was then ranked from smallest to largest with a revised set of P data being obtained using Equation 5. The $\{\sigma, P\}$ data was then linearized according to Equation 6 with values of m^* and σ^*_o being calculated using Equations 7a and b. This procedure was repeated 10^7 times for each value of N and m (as opposed to 10^4 times in the previous work [10]) to result in a frequency distribution of m^*. The value of m^*_{mean} was simply calculated from $m^*_{mean} = \Sigma m^*/10^7$ whereas the mode of m^*, m^*_{mode}, was calculated from the frequency of m^* values within incrementally increased ranges of m^*.

Frequency distributions of m^* for the case $m = 4$ and $N = 10, 20, \ldots 50$ have been presented in Fig. 1 with the area under each curve being set equal. Whereas the $N = 10$ curve was noted to be highly unsymmetrical (i.e., the mean, median, and mode were significantly different) and shifted (i.e., biased) towards lower values of m^*, larger values of N resulted in improved symmetricity of the curves and reduced bias in m^*. Another important point apparent from this figure was the increasing difference in mean and mode values as N decreased. Whereas past research [12–18] has tended to concentrate on the bias exhibited by m^*_{mean}, materials science and engineering researchers generally only have limited sets of data available and it is the author's believe that m^*_{mode} (i.e., the value of m^* with the highest probability of occurrence) is a more important indicator of bias in m^*; the reason being that, for any given test, the region of highest probability for the resulting m^* value will be centered around m^*_{mode}. However, for the sake of completeness, empirical correction factors have been calculated for both m^*_{mean} and m^*_{mode} in the present work.

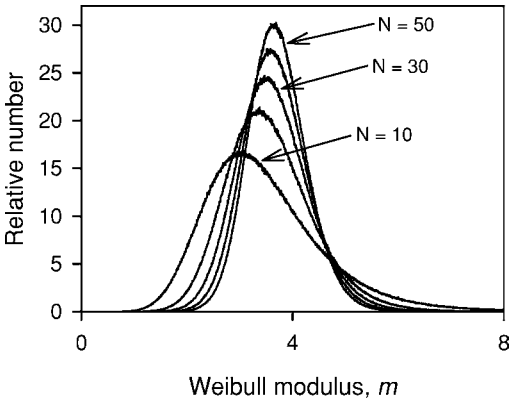

Figure 1 Frequency distributions of Weibull modulus, m^*, obtained using the linear least squares method for data sets containing N specimens ($N = 10, 20, \ldots 50$) and $m = 4$.

The effect of N and m on m^*_{mean} and m^*_{mode} have been presented in Fig. 2a and b, respectively; the smoother

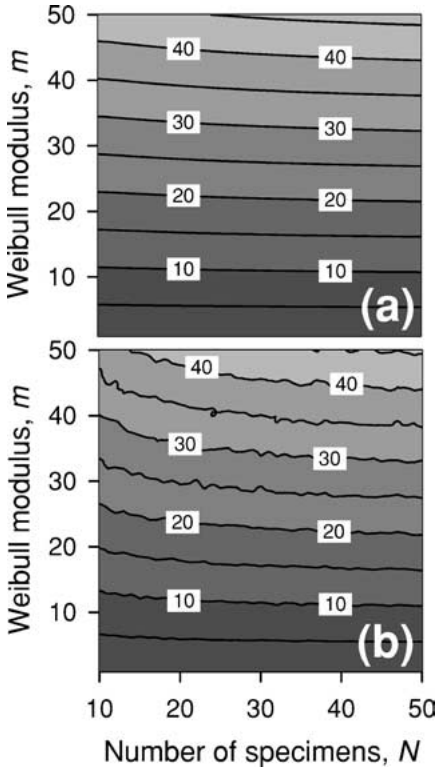

Figure 2 Values of Weibull modulus, m^*, calculated from least squares analysis as a function of specimen number and theoretical Weibull modulus: (a) mean (m^*_{mean}) and (b) mode (m^*_{mode}).

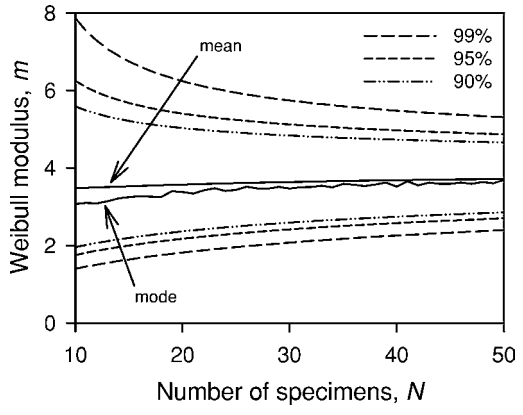

Figure 3 Comparison between mean (m^*_{mean}) and mode (m^*_{mode}) Weibull modulus data obtained from least linear squares analysis together with confidence limits of the m^* data for the case $m = 4$.

contour lines for the m^*_{mean} data (Fig. 2a) being a result of the more accurate estimation of m^*_{mean} when compared to that of m^*_{mode}. It can be seen that m^*_{mean} and m^*_{mode} were both biased towards lower values of m for the entire range of N investigated in this work with $m^*_{mode} < m^*_{mean} < m$. This trend has been clearly indicated in Fig. 3 which shows m^*_{mean} and m^*_{mode} as a function of N for the case $m = 4$. Overall, the fractional bias ranged from 0.0709 to 0.1306 for m^*_{mean} (average: 0.0945) and 0.0710 to 0.2660 for m^*_{mode} (average: 0.1368) and generally decreased with increasing N and m. The confidence limits in Fig. 3 also indicated a large degree of uncertainty associated with estimating m for small N values; the fractional error in m being on the order of ± 0.8 for the case of $N = 10$ and a 99% confidence limit. The confidence limits also highlight the highly unsymmetrical nature of the m^* distributions for small N values.

The data shown in Fig. 2 was fitted using a large number of arbitrary equations in order to obtain empirical correction factors for m^*_{mean} and m^*_{mode}. While many equations were found to fit the m^*_{mean} and m^*_{mode} data sets to a similarly high degree of accuracy (correlation coefficient > 0.999), it was decided to focus on equations that: (i) were relatively simple in form, and (ii) possessed a small number of parameters. The empirical correction factors were thus chosen to be:

$$m = Exp\left[A_o + \frac{A_1}{N^{1/2}} + \frac{A_2}{N^{3/2}} + A_3 \ln(m^*_{mean})\right] \quad (8a)$$

$$m = Exp\left[A_o + \frac{A_1 \ln(N)}{N} + \frac{A_2}{N^{3/2}} + A_3 \ln(m^*_{mode})\right] \quad (8b)$$

with the respective constants being given in Table I. In order to understand the goodness of fit for Equations 8a and b, residual magnitudes after fitting these equations to their respective data in Fig. 2 have been presented in Fig. 4 with the residuals being on the order of 0.01% for m^*_{mean} and 1% for m^*_{mode}. The significantly higher residual values for m^*_{mode} were attributed to the noisier data set in Fig. 2b as a result of the m^*_{mode} calculation method rather than to any issue with Equation 8b.

TABLE I Values of constants used in Equations 8a and b

Constant	Mean (Equation 8a)	Mode (Equation 8b)
A_0	−0.015110 (+/−0.000033)	0.015844 (+/−0.003102)
A_1	0.661820 (+/−0.000187)	0.951540 (+/−0.025727)
A_2	−1.717104 (+/−0.001249)	1.545974 (+/−0.150392)
A_3	0.999998 (+/− 0.000006)	0.999769 (+/−0.000675)

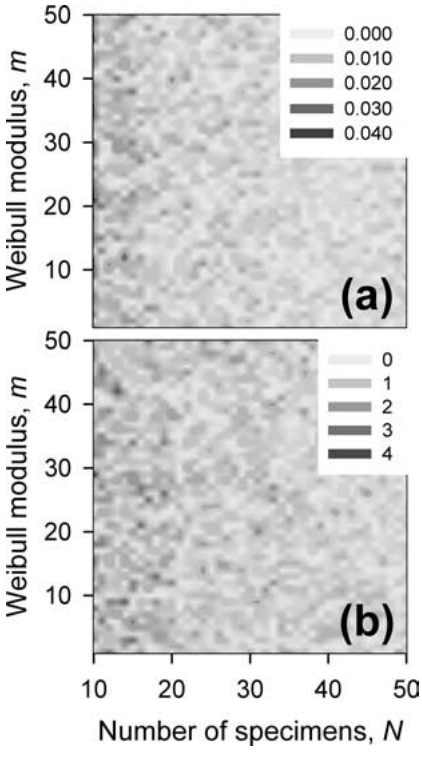

Figure 4 Magnitude of the residuals (%) after fitting the empirical equations to Weibull modulus data calculated from least squares analysis: (a) mean (Equation 8a), and (b) mode (Equation 8b).

It was also noted from Fig. 4 that the residual magnitudes indicated a general decreasing trend with increasing N. This can be more clearly seen in Fig. 5, which presents residual magnitudes averaged across the range of m values as a function of N. Over the range of N and m investigated, use of the empirical correction factors reduced the fractional bias in m^*_{mean} by a factor of 1464 to <0.0001 (range: <0.0001

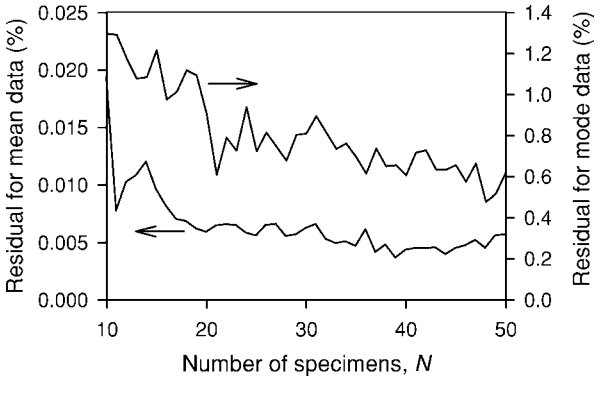

Figure 5 Magnitudes of the residuals as a function of specimen number after fitting the empirical equations to m^*_{mean} (Equation 8a) and m^*_{mode} (Equation 8b). Each curve was obtained by averaging over the range of m values (i.e., 1–50) investigated in this work.

to 0.0004) and by a factor of 17 in m^*_{mode} to 0.0081 (range: <0.0001 to 0.0459); significantly lower than that obtained by any of the fitting methods previously proposed [12–18] over such a wide range of N and m.

In summary, empirical correction factors have been calculated that significantly reduce the bias in the value of Weibull modulus obtained using linear least squares analysis. It is hoped that the relatively simple nature of the correction factors will promote widespread use in the materials science and engineering communities.

References

1. D. B. MARSHALL and J. E. RITTER, *Ceram. Bull.* **66** (1987) 309.
2. W. WEIBULL, *J. Appl. Mech.* **18** (1951) 293.
3. K. TRUSTRUM and A. DE S. JAYATILAKA, *J. Mater. Sci.* **14** (1979) 1080.
4. A. DE S. JAYATILAKA and K. TRUSTRUM, *ibid.* **12** (1977) 1426.
5. J. D. POLONIECKI and T. R. WILSHAW, *Nature* **229** (1971) 226.
6. R. D. MAURER, *Appl. Phys. Letts.* **27** (1975) 220.
7. I. J. DAVIES, T. ISHIKAWA, M. SHIBUYA, T. HIROKAWA and J. GOTOH, *Composites: Part A* **30** (1999) 587.
8. I. J. DAVIES, T. ISHIKAWA, M. SHIBUYA and T. HIROKAWA, *Comp. Sci. Technol.* **59** (1999) 801.
9. F. W. ZOK, X. CHEN and C. H. WEBER, *J. Amer. Ceram. Soc.* **78** (1995) 1965.
10. I. J. DAVIES, *J. Mater. Sci. Lett.* **20** (2001) 997.
11. J. R. TAYLOR, in "An Introduction to Error Analysis," 2nd ed. (University Science Books, Sausalite, 1982) p. 183.
12. B. BERGMAN, *J. Mater. Sci. Lett.* **3** (1984) 689.
13. J. D. SULLIVAN and P. H. LAUZON, *ibid.* **5** (1986) 1245.
14. B. BERMAN, *ibid.* **5** (1986) 611.
15. B. FAUCHER and W. R. TYSON, *ibid.* **7** (1988) 1199.
16. R. LANGLOIS, *ibid.* **10** (1991) 1049.
17. A. KHALILI and K. KROMP, *J. Mater. Sci.* **26** (1991) 6741.
18. N. KAMIYA and O. KAMIGAITO, *ibid.* **19** (1984) 4021.

Received 12 June
and accepted 16 September 2003