I. J. Davies, "Empirical correction factor for the best estimate of Weibull modulus obtained using linear least squares analysis", J. Maters. Sci. Letts., 20(11) pp. 997-999 (2001).

Empirical correction factor for the best estimate of Weibull modulus obtained using linear least squares analysis

I. J. Davies

Kyoto Institute of Technology, Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan
(Tel: +81 75 724 7850, Fax: +81 75 724 7800, e-mail: davies@ipc.kit.ac.jp)

It has long been known that the strength of brittle materials exhibits a degree of scatter that may be characterised using a cumulative distribution function originally proposed by Weibull [1] and given by:

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

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 normalising factor. In most cases, σ_u is assumed to be zero [2] and, for specimens of constant geometry, Equation 1 may be reduced to a simplified form:

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

Although initially proposed as an empirical relationship, the form of Equation 1 has since been closely linked to the properties of flaw distributions typically present in brittle materials [3-6].

The most widely used method to obtain best estimates of σ_o and m from a set of data has involved the ranking of σ data from smallest to largest and the assignment of respective P values according to the following:

$$P = \frac{i}{N+1} \tag{3}$$

where i is the rank and N is the total number of specimens. Equation 2 can be linearised into the form:

$$y = A + Bx \tag{4}$$

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

The best estimates of m and σ_o (referred to as m^* and σ^*_o, respectively) may be obtained using:

$$A = \frac{\Sigma x^2 \Sigma y - \Sigma x \Sigma xy}{N\Sigma x^2 - (\Sigma x)^2} \tag{5a}$$

$$B = \frac{N\Sigma xy - \Sigma x \Sigma y}{N\Sigma x^2 - (\Sigma x)^2} \tag{5b}$$

with Σ, x, and y in Equations 5a and 5b being abbreviations for $\sum_{i=1}^{N}$, x_i, and y_i, respectively. Although widely used, this method of estimating σ_o and m, known as the linear least squares (LLS) technique, suffers from the fact that the mean value of m^*, m^*_{mean}, is biased and generally smaller than the actual value of m; the bias decreases with increasing N such that for for the case of m=10, m^*_{mean}/m=0.869 for N=10 and 0.927 for N=50 [7]. The reason behind the bias in m^*_{mean} is the assumption utilised in the LLS technique that the error for each set of $\{x, y\}$ data points in Equation 4 is distributed symmetrically about a mean value according to a Gaussian function [8]. Whilst usually true for the $\{\sigma, P\}$ data set, application of a nonlinear transformation on Equation 2 to result in Equation 4 makes this assumption invalid. Several attempts have been made to reduce the bias in m^*_{mean} including: (i) use of weighting factors for each data point [9-11], (ii) use of different relationships to calculate P [7,12], (iii) use of parameter estimation methods other than LLS (such as "method of moments" and "maximum likelihood") [11,13], and (iv) omission of certain data points [14]. Although all of these methods have been shown to result in less biased estimates of m^*_{mean}, it is apparent that each method suffers from one or more of the following: (i) the improved m^*_{mean} value still exhibits a degree of bias, (ii) similarly invalid assumptions are utilised as for the case of LLS, (iii) lack of physical meaning, and (iv) difficulty of use. Indeed, there appears to be little evidence in the materials science and engineering literature that any of these "improved" methods have gained widespread acceptance.

In light of this, the author proposes that the best estimate of *m* should be obtained by applying an empirical correction factor to *m** obtained using the standard LLS technique. Compared to any of the previous methods, such a procedure would have the advantage of providing a best estimate of *m* with a significantly smaller bias in addition to being easily applied to both future and historic data sets.

One important feature of the LLS technique is that the bias in m^*_{mean} may be accurately determined by performing LLS analysis on a large number of simulated data sets for any given *m* and *N*. The present work will provide an example of how to obtain an empirical correction factor using a large number (5.1×10^5) of simulated data sets generated using a Monte Carlo method; the value of *m* was chosen to be 5 (as this is typical for glass and ceramic fibres [15-18]) whilst *N* was chosen to be {{10, 11, 12 .... 50}, {60, 70, 80 .... 150}} and σ_o was set to unity. For each data set, *N* values between 0 and 1 were randomly chosen $\{P_1, P_2, P_3 .... P_N\}$ and used to generate values of σ $\{\sigma_1, \sigma_2, \sigma_3 .... \sigma_N\}$ using Equation 2. The σ data was then ranked from smallest to largest and a revised *P* data set obtained using Equation 3. The {σ, *P*} data was then converted to linear form according to Equation 4 and values of *m** and S_o* obtained using Equation 5. The procedure was repeated 10^4 times for each *N* to result in a frequency distribution of *m** values; all data being normalised with respect to *m*.

Distributions of *m** for *N*=10 and *N*=150 have been presented in Figure 1 with the area under each curve being set equal. Whereas the *N*=10 curve is highly unsymmetrical (i.e., the mean, median, and mode are different) and shifted towards lower values of *m** (i.e., biased towards lower values), the *N*=150 curve appears almost symmetrical (i.e., the mean, median, and mode are similar) and centered close to *m* (i.e., almost no bias). One point generally not mentioned by previous researchers [7,9-14], and worthy of emphasis here, is that of the mean, median (m^*_{median}), and mode (m^*_{mode}) being significantly different for low *N* values. Past research has tended to compare the effectiveness of different fitting procedures using the bias in m^*_{mean}. However, workers in the fields of materials science and engineering typically utilise a limited number of data sets to obtain best estimates of *m* and σ_o. In this case, 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 the bias in *m**; the reason being that, for any given test, the region of highest probability for the resulting *m** value is centered around m^*_{mode}. However, for the sake of completeness, empirical correction factors have been calculated for both m^*_{mode} and m^*_{mean} in the present work.

For each distribution of *m** obtained for any particular *N*, the values of m^*_{mode}, m^*_{mean}, and the 95% confidence limits were calculated and these have been presented in Figure 2. The bias in *m** has been reflected in m^*_{mode} and m^*_{mean} being less than unity whilst the lack of symmetry in the *m** distributions is indicated by the lack of symmetry in the 95% confidence limits about m^*_{mode} and m^*_{mean}. Normalised values of m^*_{mode} and m^*_{mean} for N=10 were 0.795 and 0.863, respectively, with both values increasing and converging such that $m^*_{mode} \approx m^*_{mean} \approx 0.96$ for N=150.

Empirical correction factors for m^*_{mode} and m^*_{mean} were then obtained by fitting a large number of arbitrary equations to the m^*_{mode} vs. N and m^*_{mean} vs. N data sets using standard computer software. A large number of equations were found to fit the m^*_{mode} and m^*_{mean} data with similar accuracy (correlation coefficient > 0.997) and it was thus decided to choose equations that were relatively simple in form and which contained a small number of parameters. The empirical correction factors were thus chosen to be:

$$m = m^*_{mode}\left[a_0 + a_1 N^3 + \frac{a_2 \ln N}{N}\right] \tag{6a}$$

where a_0=1.027, a_1=-4.945 x 10^{-9}, a_2=1.056, and

$$m = m^*_{mean}\left[\frac{a_0 + a_1\sqrt{N}}{1 + a_2\sqrt{N}}\right] \tag{6b}$$

where a_0=1.506, a_1=0.612, a_2=0.598.

The effect of applying these correction factors on m^*_{mode} and m^*_{mean} in the range $10 \le N \le 50$ (where the bias is greatest) has been illustrated in Figure 3 with a significant decrease in bias being noted for both m^*_{mode} and m^*_{mean}. For the range $10 \le N \le 150$ it was calculated that, on average, application of the respective correction factors reduced the bias in m^*_{mode} by a factor of 28 whilst the bias in m^*_{mean} was reduced by a factor of 47. In addition, the average deviation of the corrected m^*_{mode} and m^*_{mean} values from m was 0.0034 and 0.0017, respectively, and significantly lower than that obtained by any of the fitting methods previously proposed [6-12] over such a wide range of N.

In summary, it has been shown that a significant decrease in bias for the best estimate of m may be achieved by applying a simple empirical correction factor to the value of m^* obtained using linear least squares analysis. The corrected m^* value was found to more closely approximate the real value of m compared to any of the fitting methods previously proposed.

References

1. W. Weibull, J. Appl. Mech. 18 (1951) 293.
2. K. Trustrum and A. De S. Jayatilaka, J. Mater. Sci. 14 (1979) 1080.
3. R. A. Hunt and L. N. McCartney, Int. J. Fract. 15 (1979) 365.
4. L. N. McCartney, Int. J. Fract. 15 (1979) 477.
5. A. De S. Jayatilaka and K. Trustrum, J. Mater. Sci. 12 (1977) 1426.
6. K. Trustrum and A. De S. Jayatilaka, J. Mater. Sci. 18 (1983) 2765.
7. B. Bergman, J. Mater. Sci. Lett. 3 (1984) 689.
8. J. R. Taylor, in "An introduction to error analysis, second edition", (University Science books, Sausalite, 1982) p. 183.
9. B. Berman, J. Mater. Sci. Lett. 5 (1986) 611.
10. B. Faucher and W. R. Tyson, J. Mater. Sci. Lett. 7 (1988) 1199.
11. R. Langlois, J. Mater. Sci. Lett. 10 (1991) 1049.
12. J. D. Sullivan and P. H. Lauzon, J. Mater. Sci. Lett. 5 (1986) 1245.
13. A. Khalili and K. Kromp, J. Mater. Sci. 26 (1991) 6741.
14. N. Kamiya and O. Kamigaito, J. Mater. Sci. 19 (1984) 4021.
15. R. D. Maurer, Appl. Phys. Letts. 27 (1975) 220.
16. F. W. Zok, X. Chen, and C. H. Weber, J. Am. Ceram. Soc. 78 (1995) 1965.
17. I. J. Davies, T. Ishikawa, M. Shibuya, T. Hirokawa, and J. Gotoh, Composites: part A 30 (1999) 587.
18. I. J. Davies, T. Ishikawa, M. Shibuya, and T. Hirokawa, Comp. Sci. Technol. 59 (1999) 801.

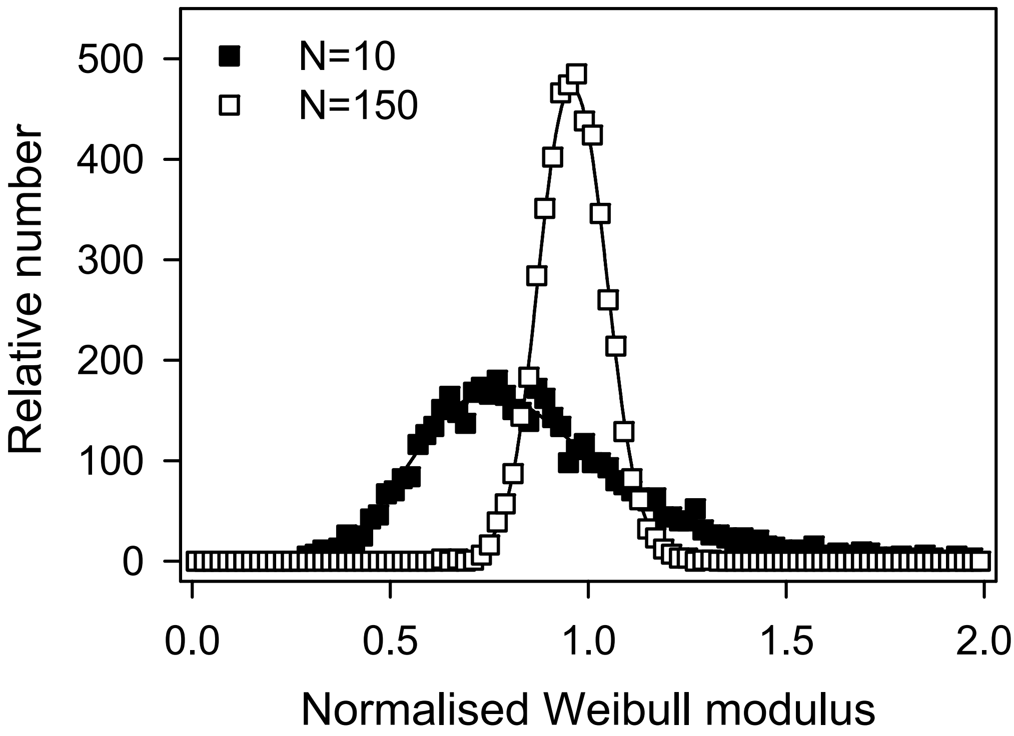

Figure 1: Frequency distribution of Weibull modulus values, m^*, obtained using the linear least squares method for data sets containing 10 and 150 specimens. For the sake of clarity, not all data points have been shown on each curve.

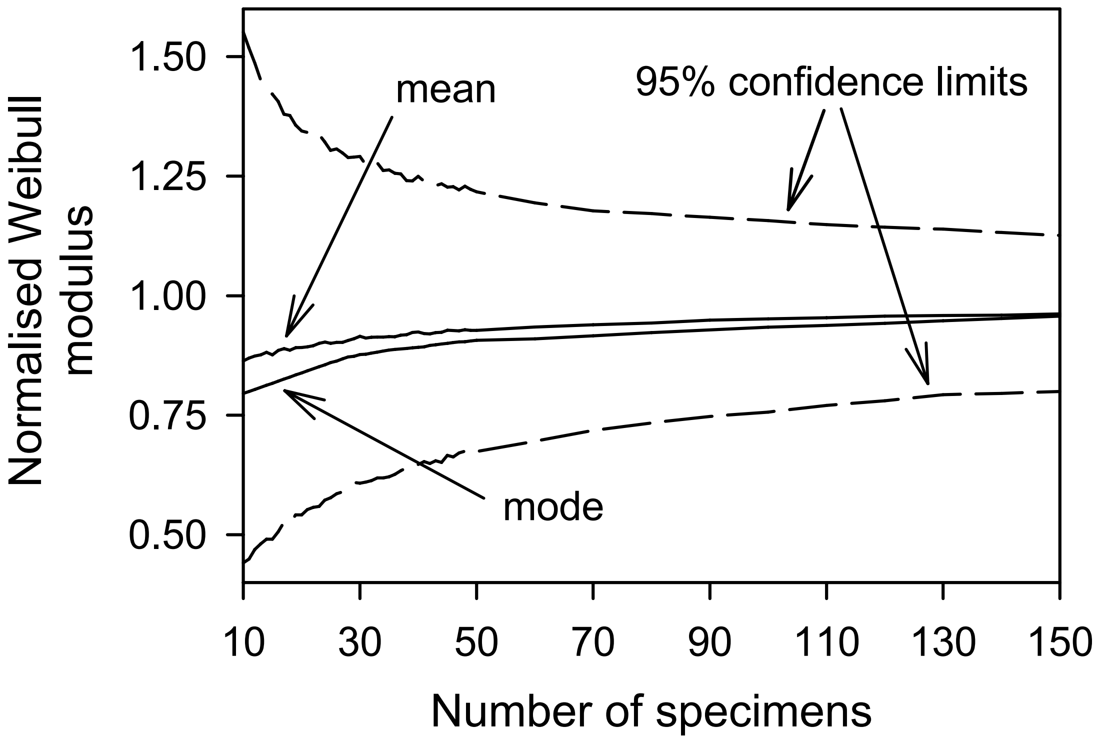

Figure 2: Effect of specimen number on properties of the normalised Weibull modulus obtained using linear least squares analysis.

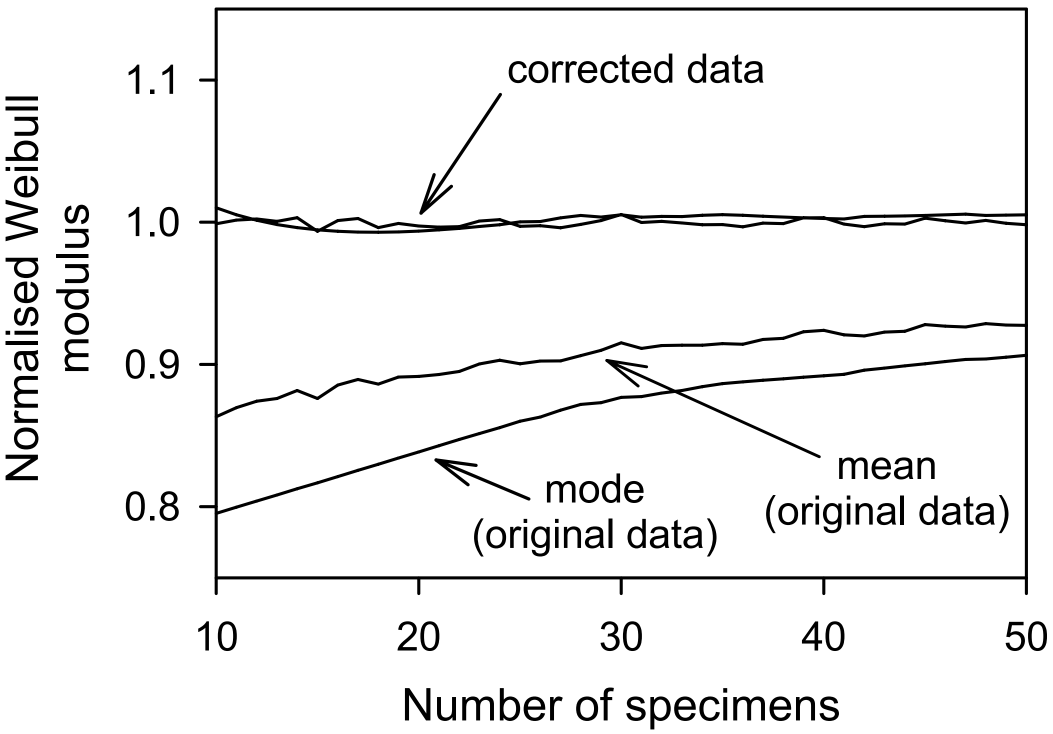

Figure 3: Comparison between mode, m^*_{mode}, and mean, m^*_{mean}, Weibull modulus data: (i) obtained from linear least squares analysis ("original data"), and (ii) following application of the respective empirical correction factors ("corrected data").