Journal of the European Ceramic Society 37 (2017) 369–380

Contents lists available at www.sciencedirect.com

Journal of the European Ceramic Society

journal homepage: www.elsevier.com/locate/jeurceramsoc

Unbiased estimation of Weibull modulus using linear least squares analysis—A systematic approach

Ian J. Davies

Department of Mechanical Engineering, Curtin University, GPO Box U1987, Perth, WA 6845, Australia

ARTICLE INFO

Article history:
Received 2 May 2016
Accepted 8 July 2016
Available online 29 July 2016

Keywords:
Weibull
Extreme value distribution
Unbiased estimation
Weibull modulus
Linear least squares

ABSTRACT

The wide applicability of the Weibull distribution to fields such as hydrology and materials science has led to a large number of probability estimators being proposed, in particular for the widely used technique of obtaining the Weibull modulus, m, using unweighted linear least squares (LLS) analysis. In this work a systematic approach using the Monte Carlo method has been taken to determining the optimal probability estimators for unbiased estimation of m (mean, median and mode) using the general equation $F = (i - a)/(N + b)$ whilst simultaneously minimising the coefficient of variation for each of the average values. A wide range of a and b values were investigated within the region $0 \leq a \leq 1$ and $1 \leq b \leq 1000$ with the form of $F = (i - a)/(N + 1)$ being chosen as the recommend probability estimator equation due to its simplicity and relatively small coefficient of variation. Values of a as a function of N were presented for the mean, median and mode m values.

1. Introduction

Extreme value distributions such as Gumbel [1], Fréchet [2] and Weibull [3] have found use in a wide variety of applications including weather forecasting, hydrology, finance and engineering. Of particular concern to materials engineering researchers is the Weibull distribution which, in its two parameter form, has found great use for the analysis of failure data in brittle materials such as ceramics and glass [4,5] with:

$$F = 1 - e^{-\left(\frac{S}{S_o}\right)^m} \tag{1}$$

where F is the cumulative probability of failure, S is stress, S_o is the Weibull strength (scale) parameter and m is the Weibull shape parameter also known as the Weibull modulus. It should be stated that F denotes the probability that a specimen will have failed at or below a stress, S. The probability density function for the two parameter Weibull distribution is given by:

$$f = \frac{m}{S_o}\left(\frac{S}{S_o}\right)^{m-1} e^{-\left(\frac{S}{S_o}\right)^m} \tag{2}$$

A wide variety of methods [6,7] have been proposed for the estimation of S_o and m from experimental data based on Eq. (1) including maximum likelihood [8], inclusion of weighting factors [9,10], omission of specific data [11] and the empirical correction of biased m values [12,13]. However, despite limitations highlighted by several researchers, unweighted linear regression analysis (also known as linear least squares (LLS)) remains the most widely utilized technique for the estimation of S_o and m due to its simplicity and ease of use. In this method the form of Eq. (1) is linearized into the following:

$$\ln\left[\ln\left[\frac{1}{1-F}\right]\right] = -m \quad \ln S_o + m \quad \ln S \tag{3}$$

which is of the form:

$$Y = a + mx \tag{4}$$

where $Y = \ln\left[\ln\left[\frac{1}{1-F}\right]\right]$, $b = -m \quad \ln S_o$ and $x = \ln S$ with the values of m and S_o being easily obtained using the LLS method. It is important to note that the value of F for an individual data point, *i.e.*, individual specimen strength, may be obtained by ranking the specimens from weakest, $i = 1$, to strongest, $i = N$, where i is the rank of the specimen strength and N is the total number of specimens.

One of the simplest and most common ways to relate F to i and N is through the relationship [3]:

$$F = \frac{i}{N+1} \tag{5}$$

with examples of probability distributions for the resulting values of m and S_o being shown in Fig. 1 as normalised values, *i.e.*, values obtained from LLS divided by the actual value of m or S_o and denoted by m^* and S_o^*, respectively. In an ideal situation the curves in Fig. 1 should be centred about $m^* = 1$ or $S_o^* = 1$ and exhibit a nar-

E-mail address: i.davies@curtin.edu.au

http://dx.doi.org/10.1016/j.jeurceramsoc.2016.07.008
0955-2219/

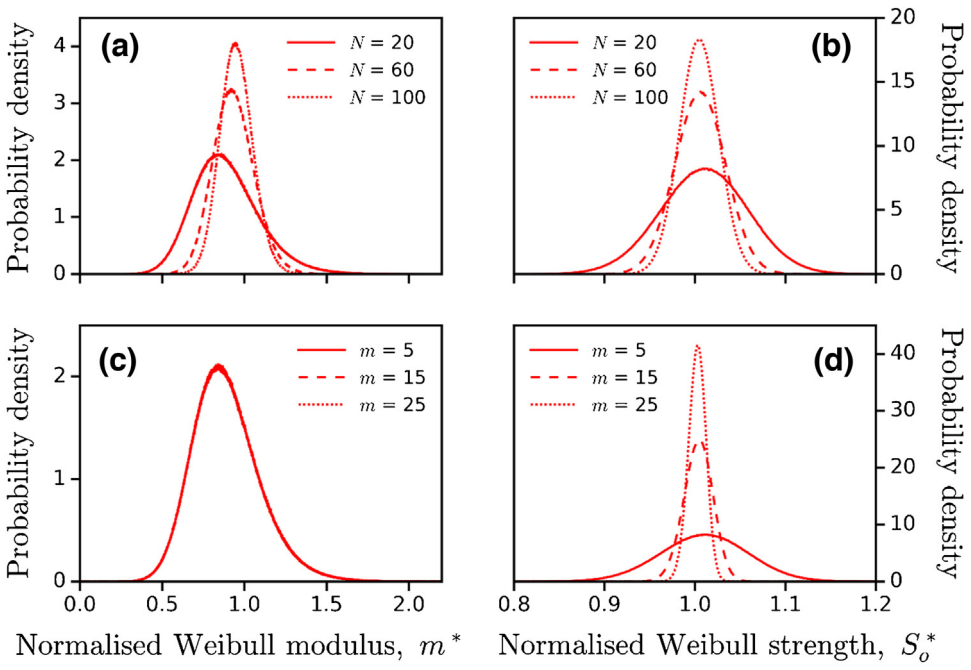

Fig. 1. Example probability distributions of normalised Weibull modulus, m^*, and normalised Weibull strength, S_o^*, obtained from 10^6 simulations of unweighted linear least squares analysis: (a), (b) $m = 5$, and (c), (d) $N = 20$.

row peak width. However, it can be seen from Fig. 1(a) and (c) that the m^* values in particular are unsymmetrical and biased towards smaller values of m^*, *i.e.*, the average value of m^* (whether it be mean, median or mode) is less than unity with the width of the peak decreasing as the number of specimens increased from 20 to 100 (Fig. 1(a)). Previous researchers have tended to describe the width of such probability distributions in terms of the coefficient of variation as follows:

$$CV = \frac{\sigma}{\mu} \tag{6}$$

where CV is the coefficient of variation, σ is the standard deviation and μ is the mean value of m^* or S_o^*, respectively. Another point of interest from Fig. 1 is that the probability distribution of m^* appears to depend on N but not on m whereas that of S_o^* depends on both N and m – this will be discussed later. Compared to m^*, the probability distributions for S_o^* tend to be significantly more symmetrical with mean values close to unity and with smaller CV values (note the difference in x axis scales for m^* and S_o^*) and this explains why most researchers have focussed on the bias in m^* results at the expense of S_o^*.

As mentioned previously, several methods have been utilized to reduce the bias in the m^* data but the most popular method by far has been to use an alternative equation to Eq. (5). Relationships between F, i and N for extreme value distributions have been available since at least the time of Laplace at the turn of the 19th century [14] and research into alternative forms of F has accelerated considerably since the formulation of the Weibull distribution in 1939 [15]. Table 1 shows expressions for F found in the literature [14–40] with early values tending to be focussed on hydrology applications whereas most recent research has tended to result from hydrology or materials engineering.

Whereas some of the probability estimators such as those from Hazen [16], Weibull [15] and Bernard and Bosi-Levenbach [19] remain popular, they all suffer from the issues of: (i) producing biased values of m^* and (ii) producing probability distributions of m^* that depend on N (as highlighted in Fig. 1(a)). It should thus be clear that no simple expression for F exists which gives unbiased values of m^* for all materials engineering relevant values of N (typically in the range of 10–100).

Due to the limitations of these simple expressions for F, researchers started to consider a more general form for F as follows:

$$F = \frac{i - a}{N + b} \tag{7}$$

with

$$0 \le a \le 1$$

$$0 \le b \le 1$$

where a and/or b are a function of N and have been denoted as a "Variable" probability estimator in Table 1. The first researcher to take this approach appears to have been Arnell in 1986 [28] with it now being a standard approach for the estimation of unbiased m^* using the standard unweighted LLS technique. Despite the emerging popularity of Eq. (7) amongst researchers attempting to reduce the bias in m^*, actual use of the "Variable" probability estimators appears scarce amongst researchers analysing experimental data despite its acknowledged superiority compared to standard expressions for F [15,16,19]. The present author attributes this to three main factors, namely: (i) the lack of expressions for F which are valid for all engineering relevant values of N (it often being the case that F is defined only for specific discrete values of N), (ii) inconsistency in the recommended values of "variable" F between different researchers and (iii) the lack of expressions from a single researcher for F that can be applied to all possible average values, *i.e.*, mean, median and mode.

Therefore, in the present work the author aims to use a systematic approach to investigate the range and appropriateness of Eq. (7) to provide simple expressions for F that are suitable to values of N in the range 10–150 and accurate for all average values, *i.e.*, mean, median and mode.

2. Experimental procedure

The general approach for this research was to use a Monte Carlo procedure to pick strength data for a given number of spec-