On the estimation of percentiles of the Weibull distribution

E. BARBERO, J. FERNÁNDEZ-SÁEZ, C. NAVARRO
Mechanical Engineering Department, Carlos III University of Madrid, Calle Butarque, 15. 28911 Leganés, Madrid, Spain
E-mail: navarro@elrond.uc3m.es

Advanced ceramics exhibit brittle behavior. The lack of ductility and the presence of flaws and defects of different sizes and orientations lead to scatter in failure strength. This variability depends also on the specimen size, stress distribution and stress state. The Weibull theory explains correctly this dependence [1], so the fracture strength of ceramic materials has been studied using the Weibull statistic [2–5] as recommended by ASTM Standards for reporting uniaxial strength data of these materials [6]. The Weibull statistic is also applicable to describe the scatter of the fracture toughness of steels in the ductile-brittle transition region, where failure occurs by cleavage [7–10].

The two-parameter Weibull distribution function is given by:

$$F(\sigma) = 1 - \exp[-(\sigma/\sigma_0)^m] \tag{1}$$

where F is the probability of rupture under an applied uniaxial tensile stress σ, m is the shape parameter or Weibull modulus and σ_0 is the scale parameter.

A subject of great interest in this field is the determination of the σ value, σ_p, corresponding to a predefined probability of failure p, i.e. the σ_p values such that

$$\Pr(\sigma \le \sigma_p) = p \tag{2}$$

These values coincide with the percentile of the distribution and can be obtained as:

$$\sigma_p = \sigma_0\,[\ln(1/(1-p))]^{1/m} \tag{3}$$

From the estimate values ($\hat{\sigma}_0$ and $\hat{m}$) of the true values of the Weibull parameters, σ_0 and m, computed from a set of tests, an estimation $\hat{\sigma}_p$ of σ_p can be calculated by:

$$\hat{\sigma}_p = \hat{\sigma}_0\,[\ln(1/(1-p))]^{1/\hat{m}} \tag{4}$$

For percentile points, combining Equations 3 and 4, the following relation is obtained [11]:

$$\hat{m}\ln\left(\frac{\hat{\sigma}_p}{\sigma_p}\right) = \hat{m}\ln\left(\frac{\hat{\sigma}_0}{\sigma_0}\right) + \left(1 - \frac{\hat{m}}{m}\right)\ln\left(\ln\frac{1}{1-p}\right) \tag{5}$$

In a previous paper, Fernández-Sáez *et al.* [11] pointed out that the $(1-\alpha)$ confidence intervals for σ_p have the limits $\hat{\sigma}_p\exp(-c_2/\hat{m})$ and $\hat{\sigma}_p\exp(-c_1/\hat{m})$, i.e.

$$\Pr[\hat{\sigma}_p\exp(-c_2/\hat{m}) \le \sigma_p \le \hat{\sigma}_p\exp(-c_1/\hat{m})] = 1-\alpha \tag{6}$$

where c_1 and c_2 fulfils the following relationships:

$$\Pr[\hat{m}\ln(\hat{\sigma}_p/\sigma_p) \le c_1] = \alpha/2 \tag{7a}$$

and

$$\Pr[\hat{m}\ln(\hat{\sigma}_p/\sigma_p) \le c_2] = 1-\alpha/2 \tag{7b}$$

These equations show the importance of the variable $\hat{m}\ln(\hat{\sigma}_p/\sigma_p)$ in determining the percentiles of the Weibull distribution and it may be estimated from estimations of the distribution parameters m and σ_0.

The aim of this work is to select the best procedure to estimate the percentage points of this variable.

Several procedures for the determination of Weibull parameters have been proposed in the literature. The most commonly used are:

(1) *The maximum-likelihood method*, according to which the estimation of parameters should satisfy the following equation:

$$\sum_{i=1}^{i=n}\ln\sigma_i - n\left(\frac{\sum_{i=1}^{i=n}\ln\sigma_i(\sigma_i)^{\hat{m}}}{\sum_{i=1}^{i=n}(\sigma_i)^{\hat{m}}}\right) + \frac{n}{\hat{m}} = 0 \tag{8}$$

and

$$\hat{\sigma}_0 = \left(\frac{1}{n}\cdot\sum_{i=1}^{i=n}(\sigma_i)^{\hat{m}}\right)^{\frac{1}{m}} \tag{9}$$

McCool [12] showed that the maximum likelihood equation (Equation 8) has a unique positive solution and can be solved by the iterative Newton-Raphson procedure.

(2) *The general linear regression method*. Equation 1 becomes a straight line when a double logarithmic transformation is made. That is:

$$\ln\left[\ln\left(\frac{1}{1-F(\sigma)}\right)\right] = m\ln\sigma - m\ln\sigma_0 \tag{10}$$

F-values are assigned on the basis of the position ith of an observation among the n ordered σ-values forming the sample. The most commonly used estimations of F are:

$$F_i = \frac{(i-0.3)}{(n+0.4)} \quad (11a)$$

$$F_i = \frac{i}{(n+1)} \quad (11b)$$

Several authors [13, 14] have indicated the convenience of using weight functions in performing the linear regression. Bergman [13] proposed the weight factor given by

$$W_i = [(1-F_i)\ln(1-F_i)]^2 \quad (12)$$

The weight factor proposed by Faucher and Tyson [14] can be approximated by:

$$W_i = 3.3F_i - 27.5[1-(1-F_i)^{0.025}] \quad (13)$$

The procedure to estimate the Weibull parameters by the weighted linear regression method can be seen in Ref. [11].

(3) *The moments method*, in which the sample moments are identified with those of the distributions.

Thoman *et al*. [15] showed that when the maximum likelihood method is used, the variables $\hat{m}/m$ and $\hat{m}\ln(\hat{\sigma}_0/\sigma_0)$ are distributed independently of m and σ_0, and have the same distribution as $\hat{m}_{11}$ and $\hat{m}_{11}\ln(\hat{\sigma}_0)_{11}$, respectively, which correspond to $m=1$ and $\sigma_0=1$. These properties are also valid when weighted linear regression is used [11]. However, these properties do not hold for the estimation obtained by the moments method.

TABLE I Estimation methods investigated

Method	Type	Equation for F_i	Equation for W_i
1	Linear regression	11a	
2	Weighted linear regression	11a	12
3	Weighted linear regression	11a	13
4	Maximum likelihood		
5	Linear regression	11b	
6	Weighted linear regression	11b	12
7	Weighted linear regression	11b	13

Because the variables $\hat{m}/m$ and $\hat{m}\ln(\hat{\sigma}_0/\sigma_0)$ are distributed independently of the real values of m and σ_0, when maximum likelihood or weighted linear regression methods are used, the variable sample size, percentage point, p, and confidence level, α, but they are not on the real values of Weibull parameters. This $\hat{m}\ln(\hat{\sigma}_p/\sigma_p)$ is distributed independently of the true values of the Weibull distribution, therefore, the coefficients c_1 and c_2 of Equation 6 only depend on the sample size, percentage point, p, and confidence level, α, but they are not on the real values of Weibull parameters. This property permits the use of simulation procedures based on the Monte Carlo method to analyze the cited variable. Thus, in this work, a set of n values from a Weibull distribution with parameters σ_0 and m (in this paper $\sigma_0=1$ and $m=1$) were generated as:

$$\sigma_i = \sigma_0[\ln(1/r)]^{1/m} \quad (14)$$

where r is a random number uniformly distributed in the range $0 \le r \le 1$. From this sample, estimations of Weibull parameters can be obtained, using the different methods summarized in Table I, and from Equation 5, the value of variable $\hat{m}\ln(\hat{\sigma}_p/\sigma_p)$ may be calculated.

Repeated application of this procedure provides a set of values of this variable and can be characterized statistically. In this work we repeated the above procedure 20000 times for each sample size for each method. The sample size was increased progressively from 5 to 50.

Tables II and III give average and standard deviation of variable $\hat{m}\ln(\hat{\sigma}_p/\sigma_p)$ for $p=0.01$ and 0.05, respectively, from the 20000 values calculated in each case. The Tables show that the bias of the variable studied depends on the method used, and the standard deviation decreases as the sample size n increases.

Langlois [16] selected the best method to obtain normalized estimation of m, based on the smallest coefficients of variation. From this point of view the procedure to be used for estimating the Weibull modulus is the maximum likelihood method.

In order to establish the best estimation method of the variable $\hat{m}\ln(\hat{\sigma}_p/\sigma_p)$, the criterion considered here is that of smallest standard deviation. In fact, the standard deviation is directly related to the precision of the estimation obtained. On the other hand, the bias of the

TABLE II Average ± standard deviation of variable $\hat{m}\ln(\hat{\sigma}_p/\sigma_p)$ for the seven methods and different sample sizes (20000 samples for each size; $p=0.01$)

	Method						
n	1	2	3	4	5	6	7
5	0.256 ± 2.958	−0.047 ± 2.767	0.047 ± 2.802	1.970 ± 3.745	−0.408 ± 2.533	−0.585 ± 2.402	−0.527 ± 2.432
6	0.064 ± 2.300	−0.193 ± 2.141	−0.126 ± 2.153	1.498 ± 2.988	−0.519 ± 2.000	−0.664 ± 1.882	−0.625 ± 1.899
7	−0.004 ± 2.011	−0.238 ± 1.880	−0.186 ± 1.871	1.205 ± 2.457	−0.537 ± 1.769	−0.661 ± 1.666	−0.634 ± 1.670
8	−0.061 ± 1.806	−0.253 ± 1.697	−0.221 ± 1.676	1.010 ± 2.148	−0.553 ± 1.605	−0.642 ± 1.515	−0.633 ± 1.510
9	−0.089 ± 1.685	−0.256 ± 1.580	−0.236 ± 1.544	0.858 ± 1.911	−0.549 ± 1.510	−0.615 ± 1.418	−0.617 ± 1.402
10	−0.112 ± 1.597	−0.256 ± 1.502	−0.246 ± 1.465	0.758 ± 1.741	−0.544 ± 1.441	−0.589 ± 1.357	−0.602 ± 1.340
20	−0.168 ± 1.063	−0.183 ± 1.014	−0.205 ± 0.944	0.336 ± 1.042	−0.459 ± 0.995	−0.372 ± 0.947	−0.427 ± 0.891
30	−0.150 ± 0.877	−0.128 ± 0.838	−0.154 ± 0.770	0.222 ± 0.813	−0.381 ± 0.834	−0.258 ± 0.798	−0.317 ± 0.736
40	−0.142 ± 0.763	−0.103 ± 0.724	−0.125 ± 0.662	0.163 ± 0.684	−0.337 ± 0.733	−0.201 ± 0.697	−0.253 ± 0.638
50	−0.132 ± 0.678	−0.082 ± 0.647	−0.103 ± 0.592	0.127 ± 0.601	−0.302 ± 0.656	−0.161 ± 0.628	−0.209 ± 0.573

TABLE III Average ± standard deviation of variable $\hat{m}\ \ln(\hat{\sigma}_p/\sigma_p)$ for the seven methods and different sample sizes (20000 samples for each size; $p = 0.05$)

	Method						
n	1	2	3	4	5	6	7
5	0.173 ± 1.966	−0.029 ± 1.840	0.027 ± 1.865	1.249 ± 2.494	−0.243 ± 1.683	−0.363 ± 1.597	−0.329 ± 1.618
6	0.051 ± 1.542	−0.121 ± 1.440	−0.083 ± 1.449	0.950 ± 2.013	−0.315 ± 1.341	−0.414 ± 1.266	−0.392 ± 1.278
7	0.008 ± 1.356	−0.149 ± 1.272	−0.120 ± 1.269	0.768 ± 1.668	−0.325 ± 1.193	−0.412 ± 1.128	−0.398 ± 1.132
8	−0.029 ± 1.220	−0.158 ± 1.151	−0.142 ± 1.140	0.643 ± 1.463	−0.337 ± 1.083	−0.401 ± 1.028	−0.398 ± 1.026
9	−0.047 ± 1.140	−0.161 ± 1.074	−0.152 ± 1.053	0.545 ± 1.303	−0.335 ± 1.021	−0.384 ± 0.965	−0.389 ± 0.956
10	−0.061 ± 1.079	−0.161 ± 1.020	−0.158 ± 0.999	0.483 ± 1.193	−0.333 ± 0.973	−0.368 ± 0.923	−0.380 ± 0.913
20	−0.100 ± 0.719	−0.115 ± 0.692	−0.131 ± 0.650	0.215 ± 0.722	−0.283 ± 0.673	−0.233 ± 0.648	−0.270 ± 0.614
30	−0.090 ± 0.593	−0.081 ± 0.572	−0.099 ± 0.532	0.142 ± 0.564	−0.234 ± 0.563	−0.161 ± 0.546	−0.201 ± 0.509.
40	−0.085 ± 0.515	−0.065 ± 0.494	−0.080 ± 0.458	0.104 ± 0.475	−0.207 ± 0.494	−0.126 ± 0.477	−0.160 ± 0.442
50	−0.080 ± 0.458	−0.052 ± 0.442	−0.066 ± 0.410	0.081 ± 0.419	−0.186 ± 0.442	−0.101 ± 0.430	−0.132 ± 0.397

pivotal variable analyzed can be removed by adding or subtracting the corresponding value from Tables II and III, without modification of the standard deviation calculated.

In accordance with this criterion, it may be seen, from results of Tables II and III, corresponding to the two values of p considered in this work, that for sample sizes greater than 7, the best method of estimation is the number 7 (see Table I) corresponding to a weighted linear regression procedure.

For sample sizes less than 7 the best method is number 6, corresponding also to a weighted linear regression procedure. In all the cases studied, results from the cited weighted linear regression methods are better than those of the maximum likelihood method for the estimation of parameters c_1 and c_2 in Equation 6.

Acknowledgment

The authors are indebted to the Fundación Ramón Areces (Área de Materiales, IX Concurso Nacional) for its financial support of this research.

References

1. J. J. PETROVIC, *Met. Trans A* **18** (1987) 1829.
2. S. C. NAMJANGUD, R. BREZNY and D. J. GREEN, *J. Amer. Ceram. Soc.* **78** (1995) 266.
3. G. D. QUINN and R. MORRELL, *ibid.* **74** (1991) 2037.
4. E. M. ASLOUN, J. B. DONNET, G. GUILPAN, M. NARDIN and J. SCHULTZ, *J. Mater Sci.* **24** (1989) 3504.
5. H. GODA and H. FUKUNAGA, *ibid.* **21** (1986) 4475.
6. "Standard Practice for Reporting Uniaxial Strength Data and Estimating Weibull Parameters for Advanced Ceramics," ASTM C1239 94a (American Society for Testing and Materials, Philadelphia, PA, 1994).
7. A. G. EVANS, *Met. Trans. A* **14** (1983) 1349.
8. F. M. BEREMIN, *ibid.* **14** (1983) 2277.
9. K. WALLIN, *Eng. Fracture Mech.* **19** (1984) 1085.
10. T. L. ANDERSON, ASTM STP 995 (American Society for Testing and Materials, Philadelphia, Pennsylvania, 1989) p. 563.
11. J. FERNÁNDEZ-SÁEZ, J. CHAO, J. DURÁN and J. M. AMO, *J. Mater. Sci. Lett.* **12** (1993) 1493.
12. J. T. McCOOL, *IEEE Trans. Reliability* **19** (1970) 2.
13. B. BERGMAN, *J. Mater. Sci. Lett.* **5** (1986) 611.
14. F. FAUCHER and W. R. TYSON, *ibid.* **7** (1988) 1199.
15. D. R. THOMAN, L. J. BAIN and C. E. ANTLE, *Technometrics* **11** (1969) 445.
16. R. LANGLOIS, *J. Mater. Sci. Lett.* **10** (1991) 1049.