J Mater Sci (2009) 44:1959–1964
DOI 10.1007/s10853-009-3306-1

On estimating percentiles of the Weibull distribution by the linear regression method

David Hudak · Murat Tiryakioğlu

Received: 22 September 2008 / Accepted: 29 January 2009 / Published online: 26 February 2009

Abstract Probability estimators developed previously by the authors have been used to obtain unbiased estimates of the Weibull parameters by the linear regression method. Using these unbiased estimators, percentiles of the Weibull distribution have been estimated. Since these percentiles are determined from the estimated parameters, they also have distributions and subsequently are determined for five sample sizes. Analysis has shown that the distributions of these estimated percentiles are neither normal, lognormal, three-parameter Weibull nor three-parameter log-Weibull. A new methodology to estimate the percentile with a specified level of confidence has been introduced. The step-by-step use of the methodology is demonstrated by examples in this paper.

Introduction

Weibull statistics is widely used to model the variability in the fracture properties of ceramics and metals, where the concept of weakest link applies. For the two-parameter Weibull distribution, the cumulative probability, P, that a part will fracture at a given stress, σ, or below can be predicted as [1];

D. Hudak
Department of Mathematics, Robert Morris University, Moon Township, PA 15108, USA
e-mail: hudak@rmu.edu

M. Tiryakioğlu (✉)
Department of Engineering, Robert Morris University, Moon Township, PA 15108, USA
e-mail: tiryakioglu@rmu.edu

$$P = 1 - \exp\left[-\left(\frac{\sigma}{\sigma_0}\right)^m\right] \quad (1)$$

where σ_0 is the scale parameter and m is the Weibull modulus, alternatively referred to as the shape parameter. There are several methods available in the literature to estimate the Weibull parameters such as linear regression, maximum likelihood method, and method of moments. Among these methods, the most popular method is linear regression method, mainly because of its simplicity; taking the logarithm of Eq. 1 twice, it yields a linear equation:

$$\ln[-\ln(1-P)] = m\ln(\sigma) - m\ln(\sigma_0) \quad (2)$$

Regardless of which method is used, the estimates of Weibull parameters, by definition, have their own distribution. The bias, i.e., the difference between the true parameter and the average of the average of the distribution of the estimates, as well as the standard error in the estimates depends on the method used [2–4].

Once the Weibull parameters are estimated, it is of interest to obtain conservative estimates such as in design values, by estimating a given percentile of the Weibull distribution. Estimated percentiles also have a distribution and therefore confidence levels are usually indicated for such conservative estimates. For instance, Mil-Handbook-5 [5] defines "A"-design allowables for alloys to be used in aerospace applications as the values for which at least 99% of the population are expected to equal or exceed (1st percentile) with a confidence of 95%. "B" allowables represent 90% exceedance (10th percentile) at the same confidence level of 95%. Because the method for estimating the Weibull parameters determines the average and standard deviation of estimates, it will also affect the estimated percentiles, as demonstrated by Barbero et al. [6, 7].

Although there have been many studies in which the distribution of the Weibull parameters were investigated, it is the authors' opinion that the estimated Weibull percentiles have not received the same level of attention. The present study has been motivated by this gap in the literature and builds on a recent study [8] by the authors in which probability estimators for unbiased estimates of Weibull parameters were introduced for 30 sample sizes (n) ranging from 5 to 100. A new approach for estimating percentiles is first introduced and then applied to the linear regression method, using probability estimators that yield no bias for the estimates of the Weibull modulus and the scale factor.

Background

To obtain a given percentile of the Weibull distribution, Eq. 1 can be rearranged as;

$$\sigma_P = \sigma_0[-\ln(1-P)]^{1/m} \quad (3)$$

where σ_P is the Pth percentile. The ratio of five percentiles ranging from 1 to 50 to the scale parameter is plotted in Fig. 1. Note that the value of m has a profound effect on the ratio. Because σ_0 and m are unknown, their estimates, $\hat{\sigma}_0$ and $\hat{m}$ have to be used to estimate the percentile. Because $\hat{\sigma}_0$ and $\hat{m}$ have their own distributions, the estimated percentiles will have a distribution, which can be expected to be dominated by the distribution of $\hat{m}$, because of the profound effect of m on the percentile, as shown in Fig. 1.

The distribution of the Weibull parameters have been investigated in several studies. In one of the earliest studies, Ritter et al. [9] ran Monte Carlo simulations and concluded that the distribution of the estimated Weibull modulus is approximately normal. These researchers ran Monte Carlo simulations only 100 times. It has since been shown [1, 7, 10] that the distribution of m is positively skewed. Gong and Wang [10] stated that m follows a lognormal distribution for linear regression and maximum likelihood methods. Barbero et al. [6] claimed that the distribution of m estimated by the maximum likelihood method is better expressed by a three-parameter Weibull distribution. In a later publication [7], the same authors found that three-parameter log-Weibull distribution provides a better fit to m estimated by the maximum likelihood method than lognormal and three-parameter Weibull distributions. Recently Tiryakioğlu [4] analyzed the distribution of m estimated by the maximum likelihood and moments methods using the Anderson–Darling goodness-of-fit test [11–13]. The author found that the distribution of $\hat{m}$ for $5 \leq n \leq 50$ is neither normal, lognormal, nor three-parameter Weibull for the maximum likelihood method. For the moments method, the distribution of $\hat{m}$ was found to be lognormal for $n \geq 40$. For any sample size, the normal three-parameter Weibull distributions did not provide a good fit.

More recently, the authors [8] developed probability estimators that yield unbiased estimates of the Weibull parameters for samples sizes (n) between 5 and 100. Moreover, they analyzed the distribution of normalized $\hat{\sigma}_0$ and $\hat{m}$ and found that (i) the distribution of normalized $\hat{\sigma}_0$ is normal [8], and (ii) the distribution of normalized $\hat{m}$ is neither normal, lognormal, three-parameter Weibull, nor three-parameter log-Weibull [3]. Although the distribution of $\hat{\sigma}_0$ is normal, it is reasonable to expect that the distribution of percentiles estimated by probability equations recently introduced by the authors [8] will not follow any of the four distributions indicated above.

This study is a continuation of the research on unbiased estimates of the Weibull parameters using the linear regression technique.

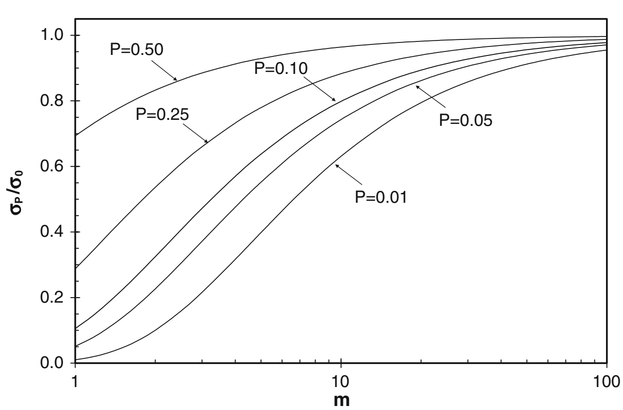

Fig. 1 The plot of σ_P/σ_0 as a function of m for various percentiles

Research methodology

Monte Carlo simulations were used to generate n data points from a Weibull distribution with $\sigma_0 = 1$ and $m = 1$. The probability estimators that yield unbiased estimates of the Weibull parameters reported by the authors [8] were used. These probability estimators have the form,

$$P = \frac{i-a}{n+b} \quad (4)$$

where i is the rank of the data point in the sample in ascending order, n represents the sample size, and a and b are numbers, such that $0 \leq a \leq 0.5$ and $0 \leq b \leq 1.0$. The authors provided a table (Table 1 in Ref. [8]), in which a and b for each sample size are listed. Thirty sample sizes ranging from 5 to 100 were investigated. At each iteration, n random numbers between 0 and 1 were generated to obtain a set of σ values. The estimates of Weibull modulus and scale factor ($\hat{m}$ and $\hat{\sigma}_0$, respectively) were obtained by Eq. 2. For each sample size, the experiment was repeated 20,000 times. Estimated percentile points were calculated as;

$$\hat{\sigma}_P = \hat{\sigma}_0\left[(-\ln(1-P))^{\frac{1}{\hat{m}}}\right] \quad (5)$$

Results and discussion

A histogram of estimated 5th percentiles for $n = 30$ is presented in Fig. 2 where f is probability density. The distribution is positively skewed, similar to the distribution of $\hat{m}$, showing the dominance of the estimated Weibull modulus on the distribution of $\hat{\sigma}_P$.

Hypothesis tests were conducted for goodness of fit for the normal, lognormal, three-parameter lognormal, three-parameter Weibull, and three-parameter log-Weibull distributions, using the Anderson–Darling statistic. In all cases, the hypothesis that the distribution of estimated percentile followed the inferred distribution was rejected. Consequently, percentage points of the distributions for each sample were generated. Tables for five sample sizes (n = 10, 20, 30, 40, 50) are provided in Tables 1, 2, 3, 4, and 5. Tables for other sample sizes can be obtained by contacting the authors.

The percentage points in Tables 1, 2, 3, 4, and 5 were generated for the case where $\sigma_0 = 1$ and $m = 1$. When the scale and shape factors are different from unity, then a transformation is required. For the case $\hat{\sigma}_0' = k$, the transformation is straight forward; because σ is a linear function of $\hat{\sigma}_0$, the desired percentile can be obtained from the appropriate table and multiplied by k. For instance, assume that 10th percentile for $n = 20$ is desired with a confidence level of 95% (as in B allowable) when $\hat{\sigma}_0' = 12$ and $\hat{m} = 1$. From Table 2, $\hat{\sigma}_P$ is found to be 0.101. Hence the 10th percentile with a confidence level of 95% is found to be 1.212 (0.101 × 12).

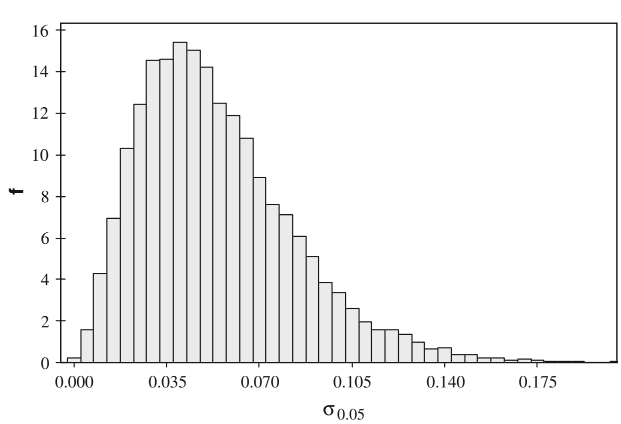

Fig. 2 Probability density of estimated 5th percentile results for $n = 30$

The transformation of the percentiles for Weibull modulus different from 1 is more complicated. Let us assume that $\sigma_0 = 1$ and $m' = k$. If we wish to determine the distribution of σ_P for a given percentile, P', we would again solve for σ as in Eq. 5:

$$\sigma_P' = \hat{\sigma}_0\left[(-\ln(1-P'))^{\frac{1}{\hat{m}'}}\right] \quad (5a)$$

where $\hat{m}' = k\hat{m}$. Inserting Eq. 5a into Eq. 1, we obtained:

$$P = 1 - \exp\left\{-\left(\frac{\hat{\sigma}_0\left[(-\ln(1-P'))^{\frac{1}{k\hat{m}}}\right]}{\hat{\sigma}_0}\right)^{\hat{m}}\right\}$$
$$= 1 - \exp\left\{-\left[(-\ln(1-P'))^{\frac{1}{k}}\right]\right\} \quad (6)$$

Therefore to find a percentile P' when $\sigma_0 = 1$ and $m' = k$, then it is necessary to compute the percentile P from Eq. 6, where P represents the percentiles when using $\sigma_0 = 1$ and $m = 1$, as presented in Tables 1, 2, 3, 4, and 5.

As an example, consider the case where $\sigma_0 = 1$ and $m = 3$ and it is necessary to find the distribution for the 10th percentile for $n = 20$. Substituting $k = 3$, and $P' = 0.10$ in Eq. 6 yields $P = 0.376$. In other words, the distribution for the 37.6th percentile when $\sigma_0 = 1$ and $m = 1$ will be the same as the distribution for the 10th percentile when $\sigma_0 = 1$ and $m = 3$. In a short paper, it is not feasible to give distributions for all percentiles and all multiple sample sizes to the accuracy above (such as the 37.6th percentile). However, it is possible to interpolate between the percentage points given in Tables 1, 2, 3, 4, and 5 to obtain approximate distributions.

After evaluating various examples, it became clear that a geometric rather than a linear interpolation gives significantly closer results. In the above case for $n = 20$, a geometric interpolation was used between the distributions for the 30th and 40th percentiles. For each value X in the 30th percentile distribution and each value Y in the 40th percentile, the value Z for the 37.6th percentile was computed as:

$$Z = X^{1-a}Y^a \quad (7)$$

Because 37.6 is 76% of the distance between 30 and 40 is $a = 0.76$. In this example, the geometric interpolation gave results within 2% and in many cases less than 1% of the true values. The results of this example are summarized in Table 6.

Calculation of a confidence interval for a percentile

To illustrate the calculation for a confidence interval for a specified percentile, let us assume that, from a sample size of 40, we wish to compute a value such that 90% of the population is greater than this value with a confidence level of 97%. This will be done in three steps:

Table 1 Percentage points for percentile estimators ($m = 1$, $\sigma_0 = 1$), $n = 10$

Distribution	Percentiles													
	1	2.5	10	20	30	40	50	60	70	80	90	95	97.5	99
0.005	0.0000	0.0001	0.004	0.024	0.076	0.18	0.37	0.63	0.87	1.10	1.35	1.56	1.72	1.90
0.010	0.0000	0.0003	0.006	0.034	0.095	0.21	0.40	0.66	0.90	1.15	1.43	1.65	1.84	2.07
0.025	0.0001	0.0007	0.011	0.049	0.123	0.25	0.44	0.69	0.96	1.21	1.54	1.80	2.05	2.33
0.050	0.0003	0.0014	0.018	0.066	0.151	0.28	0.48	0.73	0.99	1.27	1.64	1.95	2.24	2.58
0.100	0.0007	0.0030	0.028	0.091	0.188	0.33	0.52	0.77	1.04	1.35	1.78	2.16	2.51	2.94
0.900	0.0400	0.0766	0.211	0.359	0.504	0.66	0.83	1.05	1.43	2.14	3.69	5.53	7.67	10.82
0.950	0.0593	0.1052	0.257	0.413	0.557	0.71	0.87	1.09	1.50	2.37	4.39	6.95	10.01	14.81
0.975	0.0795	0.1341	0.299	0.459	0.603	0.75	0.91	1.13	1.57	2.63	5.25	8.82	13.36	20.59
0.990	0.1115	0.1759	0.352	0.512	0.654	0.80	0.95	1.16	1.68	3.03	6.71	11.91	19.03	31.36
0.995	0.1370	0.2062	0.394	0.556	0.695	0.83	0.99	1.19	1.77	3.46	8.23	15.48	26.01	45.30

Table 2 Percentage points for percentile estimators ($m = 1$, $\sigma_0 = 1$), $n = 20$

Distribution	Percentiles													
	1	2.5	10	20	30	40	50	60	70	80	90	95	97.5	99
0.005	0.0002	0.0009	0.014	0.057	0.138	0.27	0.47	0.71	0.96	1.21	1.54	1.82	2.07	2.35
0.010	0.0003	0.0013	0.017	0.067	0.154	0.29	0.49	0.74	0.99	1.25	1.60	1.90	2.17	2.49
0.025	0.0005	0.0024	0.025	0.085	0.182	0.32	0.52	0.76	1.02	1.31	1.70	2.04	2.35	2.74
0.050	0.0010	0.0039	0.033	0.101	0.205	0.35	0.54	0.79	1.05	1.35	1.78	2.16	2.51	2.95
0.100	0.0017	0.0062	0.044	0.123	0.234	0.38	0.57	0.81	1.08	1.40	1.88	2.32	2.73	3.24
0.900	0.0296	0.0604	0.181	0.323	0.466	0.62	0.79	1.01	1.35	1.95	3.14	4.50	5.96	8.07
0.950	0.0395	0.0761	0.209	0.358	0.502	0.65	0.83	1.04	1.40	2.09	3.54	5.24	7.14	9.95
0.975	0.0491	0.0904	0.233	0.386	0.532	0.68	0.85	1.07	1.45	2.24	3.95	6.06	8.52	12.25
0.990	0.0635	0.1114	0.265	0.422	0.568	0.72	0.88	1.09	1.52	2.46	4.63	7.41	10.71	15.93
0.995	0.0756	0.1283	0.292	0.451	0.596	0.74	0.90	1.12	1.57	2.64	5.18	8.64	12.84	19.59

Table 3 Percentage points for percentile estimators ($m = 1$, $\sigma_0 = 1$), $n = 30$

Distribution	Percentiles													
	1	2.5	10	20	30	40	50	60	70	80	90	95	97.5	99
0.005	0.0004	0.0018	0.020	0.076	0.168	0.31	0.51	0.75	1.00	1.28	1.66	1.99	2.28	2.65
0.010	0.0006	0.0025	0.026	0.087	0.185	0.33	0.53	0.77	1.02	1.31	1.71	2.05	2.37	2.76
0.025	0.0010	0.0040	0.034	0.104	0.211	0.36	0.55	0.79	1.05	1.35	1.79	2.17	2.53	2.97
0.050	0.0016	0.0058	0.043	0.121	0.232	0.38	0.57	0.81	1.08	1.39	1.85	2.28	2.67	3.17
0.100	0.0026	0.0084	0.054	0.141	0.258	0.41	0.60	0.83	1.11	1.44	1.94	2.41	2.86	3.42
0.900	0.0250	0.0527	0.167	0.306	0.448	0.60	0.78	1.00	1.33	1.89	2.96	4.15	5.42	7.22
0.950	0.0319	0.0640	0.188	0.333	0.476	0.63	0.80	1.02	1.37	1.99	3.25	4.68	6.25	8.52
0.975	0.0388	0.0751	0.208	0.356	0.500	0.65	0.82	1.04	1.41	2.11	3.57	5.28	7.20	9.96
0.990	0.0474	0.0880	0.230	0.382	0.527	0.68	0.85	1.06	1.46	2.27	4.05	6.21	8.73	12.55
0.995	0.0545	0.0976	0.245	0.400	0.546	0.69	0.87	1.08	1.51	2.45	4.48	7.04	10.09	14.77

1. First, assume that the sample shape and scale parameters are calculated as $\hat{m} = 7.0$ and $\hat{\sigma}_0 = 3.5$, respectively, by the linear regression method and using the probability estimators that yield unbiased estimates of the Weibull parameters. Table 2 in Ref. [8] gives 0.99 point of the distribution for $\hat{m}/m$ with $n = 40$ as 1.424. Therefore, the one-sided 99% confidence interval for m is calculated as $m > 4.92$ (=7/1.424). Similarly, a one-sided 99% confidence interval for an unbiased estimator for σ_0 is $\sigma_0 > 3.09$.

Table 4 Percentage points for percentile estimators ($m = 1$, $\sigma_0 = 1$), $n = 40$

Distribution	Percentiles													
	1	2.5	10	20	30	40	50	60	70	80	90	95	97.5	99
0.005	0.0007	0.0031	0.029	0.094	0.196	0.34	0.54	0.78	1.04	1.33	1.73	2.08	2.41	2.80
0.010	0.0010	0.0039	0.033	0.103	0.208	0.36	0.55	0.79	1.05	1.35	1.78	2.16	2.51	2.94
0.025	0.0015	0.0055	0.041	0.119	0.229	0.38	0.57	0.81	1.07	1.38	1.84	2.25	2.64	3.12
0.050	0.0021	0.0073	0.049	0.133	0.248	0.40	0.59	0.83	1.10	1.42	1.90	2.35	2.77	3.31
0.100	0.0031	0.0099	0.059	0.151	0.270	0.42	0.61	0.85	1.12	1.46	1.98	2.48	2.95	3.55
0.900	0.0224	0.0483	0.157	0.294	0.435	0.59	0.77	0.99	1.31	1.84	2.86	3.97	5.15	6.79
0.950	0.0278	0.0573	0.176	0.317	0.460	0.61	0.79	1.00	1.34	1.92	3.07	4.35	5.73	7.72
0.975	0.0332	0.0660	0.192	0.337	0.481	0.63	0.81	1.02	1.37	2.01	3.30	4.77	6.38	8.71
0.990	0.0406	0.0777	0.212	0.360	0.505	0.66	0.83	1.04	1.41	2.13	3.62	5.36	7.31	10.18
0.995	0.0461	0.0861	0.227	0.377	0.521	0.67	0.84	1.06	1.45	2.23	3.86	5.80	8.07	11.43

Table 5 Percentage points for percentile estimators ($m = 1$, $\sigma_0 = 1$), $n = 50$

Distribution	Percentiles													
	1	2.5	10	20	30	40	50	60	70	80	90	95	97.5	99
0.005	0.0009	0.0036	0.032	0.101	0.206	0.35	0.55	0.79	1.05	1.34	1.77	2.15	2.50	2.93
0.010	0.0012	0.0047	0.037	0.111	0.221	0.37	0.56	0.80	1.06	1.37	1.81	2.22	2.59	3.05
0.025	0.0019	0.0067	0.046	0.129	0.242	0.39	0.59	0.82	1.09	1.40	1.88	2.31	2.71	3.23
0.050	0.0026	0.0086	0.054	0.142	0.260	0.41	0.60	0.84	1.10	1.43	1.93	2.40	2.84	3.40
0.100	0.0037	0.0112	0.064	0.159	0.280	0.43	0.62	0.85	1.13	1.47	2.01	2.51	3.00	3.61
0.900	0.0212	0.0461	0.153	0.288	0.429	0.58	0.76	0.98	1.30	1.82	2.78	3.83	4.93	6.44
0.950	0.0258	0.0538	0.169	0.309	0.451	0.60	0.78	1.00	1.33	1.89	2.98	4.17	5.44	7.25
0.975	0.0300	0.0609	0.183	0.325	0.469	0.62	0.80	1.01	1.36	1.97	3.17	4.53	6.00	8.11
0.990	0.0359	0.0705	0.199	0.347	0.492	0.64	0.82	1.03	1.40	2.08	3.48	5.07	6.89	9.52
0.995	0.0403	0.0773	0.212	0.362	0.506	0.66	0.83	1.04	1.43	2.19	3.75	5.56	7.62	10.64

Table 6 Demonstration of the interpolation technique

Probability	30th percentile	40th percentile	37.6th percentile		
			Interpolated	Actual	Error
0.005	0.138	0.27	0.2305	0.2334	−1.2%
0.010	0.154	0.29	0.2482	0.2515	−1.3%
0.025	0.182	0.32	0.2813	0.2846	−1.2%
0.050	0.205	0.35	0.3070	0.3101	−1.0%
0.100	0.234	0.38	0.3387	0.3417	−0.9%
0.900	0.466	0.62	0.5792	0.5818	−0.5%
0.950	0.502	0.65	0.6135	0.6163	−0.5%
0.975	0.532	0.68	0.6423	0.6446	−0.4%
0.990	0.568	0.72	0.6774	0.6794	−0.3%
0.995	0.596	0.74	0.7025	0.7033	−0.1%

2. We now consider Eq. 6 using $m = 4.92$, with $\sigma_0 = 1.0$. The one-sided 99% confidence interval for the population 10th percentile when $m = 4.92$ and $\sigma_0 = 1.0$ corresponds to the confidence interval for the 46.9th percentile when $m = 1.0$ and $\sigma_0 = 1.0$. Interpolating using results for the 10th percentile in Table 4 gives a 99% confidence that the 10th percentile is greater than 0.48. That is, we are 99% confident that 90% of the population is greater than 0.48.
3. Finally, we combine the various confidence intervals. Thus, if $\sigma_0 = 1.0$, we can be 99% confident that $m > 4.92$ and given that m is in this range, we can be 99% confident that the 10th percentile is greater than 0.48. That is, we are approximately 98% (0.99^2) confident that the population 10th percentile is greater than 0.48 when $\sigma_0 = 1.0$. Further, we are 99% confident that σ_0 is greater than 3.09 so we can be approximately 97% (0.98 × 0.99) confident that the 10th percentile is greater than 1.48 (3.09 × 0.48).

Conclusions

The distributions of the estimated percentiles in this paper are neither normal, lognormal, three-parameter Weibull, nor three-parameter log-Weibull. As a result, it is necessary to generate tables for the distribution of percentiles. This was done for five sample sizes when the shape and scale parameters are both equal to one.

A geometric interpolation method is used to develop distributions of percentiles when the shape and scale parameters differ from one. Further, since in practice, the shape and scale parameters that would be used in this interpolation are themselves estimates from data, a step-by-step procedure for determining the distribution for the true percentiles is demonstrated through the use of examples.

References

1. Weibull W (1951) J Appl Mech 18:293
2. Khalili A, Kromp K (1991) J Mater Sci 26:6741. doi:10.1007/BF00553701
3. Tiryakioğlu M, Hudak D (2007) J Mater Sci 42:10173. doi:10.1007/s10853-007-2060-5
4. Tiryakioğlu M (2008) J Mater Sci 43:793. doi:10.1007/s10853-007-2095-7
5. Military handbook 5F: Metallic materials and elements for aerospace vehicle structures, Department of Defense (Nov 1990)
6. Barbero E, Fernandez-Saez J, Navarro C (2000) Compos Part B-Eng 31:375
7. Barbero E, Fernandez-Saez J, Navarro C (2001) J Mater Sci Lett 20:847
8. Tiryakioğlu M, Hudak D (2008) J Mater Sci 43:1914. doi:10.1007/s10853-008-2457-9
9. Ritter J, Bandyopadhyay N, Jakus K (1981) Am Ceram Soc Bull 60:788
10. Gong J, Wang J (2002) Key Eng Mater 224–226:779
11. Anderson TW, Darling DA (1954) J Am Stat Assoc 49:765
12. Stephens MA (1974) J Am Stat Assoc 69:730
13. Stephens MA (1986) In: D'Agostino RB, Stephens MA (eds) Goodness of fit techniques. Marcel Dekker, p 97