Structural Safety 73 (2018) 29–41

Contents lists available at ScienceDirect

Structural Safety

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

Weibull parameter estimation and goodness-of-fit for glass strength data

Kyriaki Corinna Datsiou *, Mauro Overend

Glass and Façade Technology Research Group, Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK

A R T I C L E   I N F O

Article history:
Received 11 December 2016
Received in revised form 4 February 2018
Accepted 6 February 2018

Keywords:
Glass strength
Statistical analysis
2-Parameter Weibull distribution
Goodness of fit
Small samples

A B S T R A C T

Strength data from macroscopically identical glass specimens is commonly described by a two-parameter Weibull distribution, but there is lack of research on the methods used for fitting strength data to the Weibull distribution. This study investigates 4 different methods for fitting data and estimating the parameters of the Weibull distribution namely, good linear unbiased estimators, least squares regression, weighted least squares regression and maximum likelihood estimation. These methods are implemented on fracture surface strength data from 418 annealed soda-lime-silica glass specimens, grouped in 30 nominally identical series, including as-received, naturally aged and artificially aged specimens. The strength data are evaluated based on their goodness of fit. Comparison of conservativeness of strength estimates is also provided. It is found that a weighted least squares regression is the most effective fitting method for the analysis of small samples of glass strength data.

1. Introduction

Glass strength is governed by the condition of its surface, including microscopic flaws on the surface of glass that may be indiscernible to the naked eye. There is a large variation in fracture strength obtained from seemingly identical specimens which are produced, stored and tested destructively under the same conditions. Therefore, destructive testing of several nominally identical glass specimens and the subsequent statistical analysis of their strength data is essential for establishing an accurate design strength, corresponding to a sufficiently low probability of failure. Glass is susceptible to sub-critical crack growth, therefore in order to normalise the effects of glass specimens failing after different load durations, the fracture strength data from the destructive tests is often expressed as a time-equivalent strength. This is achieved by converting the stress history exerted during the destructive test over the time to failure, t_f to an equivalent constant stress, $\sigma_{f,ref}$, for a reference time period, t_{ref}, (60 s is a typical value) as shown in Eq. (1) [1]:

$$\int_0^{t_f} \sigma^n(t)dt = \int_0^{t_{ref}} \sigma^n_{f,ref} dt \tag{1}$$

There are three statistical distributions that have historically been used to describe strength data: Weibull, normal and lognormal [2–5]. The 2-parameter Weibull distribution is often preferred because: (a) it is more accurate in describing glass strength data than a normal distribution [2] and; (b) is always more conservative in the tail of the distribution than a lognormal distribution [3] (this is also verified for the strength data used in the present study as shown in Appendix B). Conservative estimates are more desirable for engineering design applications. As a result the Weibull distribution is the established way of describing glass strength data in both research [5–11] and engineering applications [12–14].

The general equation for the cumulative distribution function (*CDF*) of the Weibull distribution [15] is:

$$P_f(\sigma_{f,60}) = 1 - \exp\left[-\left(\frac{\sigma_{f,60} - \sigma_u}{\theta}\right)^{\beta}\right] \tag{2}$$

where β is the shape parameter, θ is the scale parameter, $\sigma_{f,60}$ is the equivalent fracture stress for a reference time period of 60 s and σ_u is the location parameter.

The location parameter, σ_u, represents the stress level below which the material never fails (i.e. $P_f = 0$). Safety reasons dictate that σ_u is set to 0 as recommended in Trustrum and Jayatilaka [16] for brittle materials. Therefore, Eq. (2) is reduced to a two-parameter Weibull function and the *CDF* can be linearized (Eq. (3)) in the form of $y = bx + c$ by taking the logarithm of each side twice:

$$\ln\left(\ln\left(\frac{1}{1-P_f}\right)\right) = \beta \cdot \ln\sigma - \beta \cdot \ln\theta \tag{3}$$

Hence, the CDF becomes a linear plot of $\ln\left(\ln\left(\frac{1}{1-P_f}\right)\right)$ vs. $\ln\sigma$ as illustrated in Fig. 1, and where the gradient of the distribution is

* Corresponding author.
E-mail address: kd365@cam.ac.uk (K.C. Datsiou).

https://doi.org/10.1016/j.strusafe.2018.02.002
0167-4730/© 2018 The Authors. Published by Elsevier Ltd.

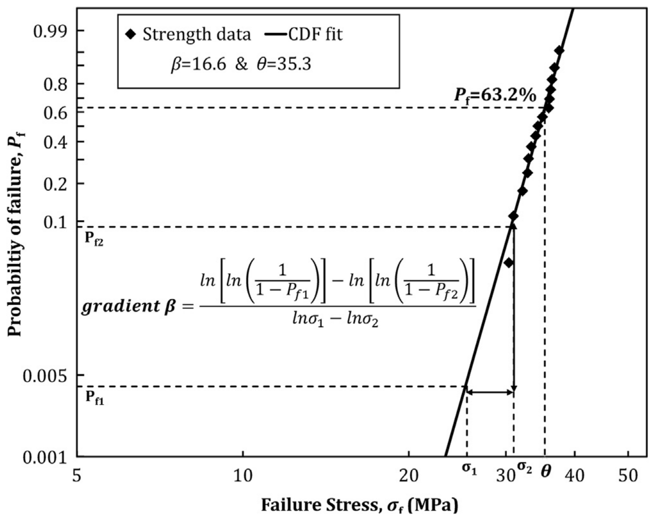

Fig. 1. Cumulative distribution function (CDF) of glass strength data.

equal to the shape parameter, β and the intercept is $-\beta{\cdot}\ln\theta$. The shape parameter, β, indicates the variability of the data and thus, higher values of β lead to a steeper *CDF* and represent a smaller scatter of strength in the data. The scale parameter, θ, represents the stress level, below which 63.2% of the specimens fail and together with the shape parameter dictates the position of the *CDF* along the horizontal axis.

There are various approaches for estimating the Weibull parameters from a given set of strength data. They can be classified either as manual or computational methods. Manual calculations can be performed by: (a) least square regression (*LR*); (b) weighted least squares regression (*WLR*) and; (c) a linear approach based on good linear unbiased estimators (*GLUEs*); while computational (computer-based) methods are: (a) the maximum likelihood estimation (*MLE*) and; (b) the method of moments estimation (*MME*).

The aim of this study is to review these different estimation methods for Weibull parameters and to propose the most effective method for the statistical analysis of small sized samples of glass strength. To the best of the authors' knowledge, this study is the first to use real glass strength data in order to assess methods for their statistical analysis. Therefore, the observations and conclusions from this study are valuable for researchers and practitioners who have performed destructive tests on a relatively small number of nominally identical glass components and wish to perform a statistical analysis of the strength data. An overview of the existing methods for estimating the Weibull parameters and goodness-of-fit for glass strength data are first reviewed in Section 2. The existing methods (*LR*, *WLR*, *GLUEs* and *MLE*) are then implemented on 30 real data sets, obtained from destructive tests on naturally aged, as-received and artificially aged glass in Section 3. The goodness-of-fit and strength estimate results of each method are presented and discussed in Section 4 and the conclusions are provided in Section 5.

2. Review of Weibull statistics methods

The two principal steps when performing a statistical analysis are: estimating the statistical parameters and evaluating the goodness-of-fit. These are reviewed in this section in the context of a Weibull distribution for glass strength data.

2.1. Parameter estimation

The most commonly used approaches, within the Weibull statistics community, for the estimation of the shape and scale parameters of the Weibull distribution are described below.

2.1.1. Manual calculations methods

Equivalent strength data are ranked in ascending order ($i = 1$ to n) for the manual calculation methods. Equal probabilities of failure, P_f, are assigned to each data point in cumulative form with functions called probability estimators, E_i. The simplest forms of probability estimators are $E = i/n$ or $E = (i-1)/n$ but these estimators eliminate the highest or lowest data point of the sample in the *CDF* graph for $P_f = 1$ or $P_f = 0$ respectively; the highest/lowest strength point are therefore, also eliminated during the estimation of the Weibull parameters so that instead of n specimens, only $(n - 1)$ would be considered. Therefore, these estimators are avoided and probability estimators of the following form are preferred instead:

$$E_j = \frac{i - C_j}{n + 1 - 2C_j} \tag{4}$$

where C_j is a constant $0 \leq C_j < 1$, i is the index of the ascending order and n is the sample size. The following four probability estimators (E_j, $j = 1$– 4, [17–19]) are most commonly used in Weibull statistics:

$$E_1 \text{ (mean rank)}: \; C_1 = 0 \rightarrow E_1 = \frac{i}{n+1} \tag{4a}$$

$$E_2 \text{(Hazen's)}: \; C_2 = 0.5 \rightarrow E_2 = \frac{i - 0.5}{n} \tag{4b}$$

$$E_3 \text{ (median rank)}: \; C_3 = 0.3 \rightarrow E_3 = \frac{i - 0.3}{n + 0.4} \tag{4c}$$

$$E_4 \text{ (small sample)}: \; C_4 = 0.375 \rightarrow E_3 = \frac{i - 0.375}{n + 0.25} \tag{4d}$$

2.1.1.1. Least Squares Regression (LR). The Weibull parameters are determined in the Least Squares Regression method (*LR*), by minimizing the sum of squared residuals of the x values about Eq. (3):

$$\beta = \frac{n \cdot \sum_{i=1}^{n}[\ln(\sigma_i) \cdot y_i] - \sum_{i=1}^{n}(\ln \sigma_i) \cdot \sum_{i=1}^{n}(y_i)}{n \cdot \sum_{i=1}^{n}[(\ln \sigma_i)^2] - [\sum_{i=1}^{n}(\ln \sigma_i)]^2} \tag{5a}$$

$$-\beta \cdot \ln \theta = \frac{\sum_{i=1}^{n}(y_i) - \beta \times \sum_{i=1}^{n}(\ln \sigma_i)}{n} \tag{5b}$$

However, *LR* implicitly applies the same unit weight to each data point without accounting for the uncertainty of $y = \ln\left(\ln\left(\frac{1}{1-E}\right)\right)$ or E_i. and thus provide biased estimates.

2.1.1.2. Weighted Least Squares Regression (WLR). Weibull parameters with smaller bias than those deriving from *LR*, can be obtained (Eq. (6a) and (6b)) by introducing weight functions based on the uncertainty of y and E within the *LR* method leading to a Weighted Least Squares Regression, *WLR* [20].

$$\beta = \frac{\sum_{i=1}^{n} W_i \cdot \sum_{i=1}^{n}[\ln(\sigma_i) \cdot y_i \cdot W_i] - \sum_{i=1}^{n}[\ln(\sigma_i) \cdot W_i] \cdot \sum_{i=1}^{n}(y_i \cdot W_i)}{\sum_{i=1}^{n} W_i \cdot \sum_{i=1}^{n}[(\ln \sigma_i)^2 \cdot W_i] - [\sum_{i=1}^{n}(\ln \sigma_i) \cdot W_i]^2} \tag{6a}$$

$$-\beta \cdot \ln \theta = \frac{\sum_{i=1}^{n}(y_i \cdot W_i) - \beta \cdot \sum_{i=1}^{n}[\ln(\sigma_i) \cdot W_i]}{\sum_{i=1}^{n} W_i} \tag{6b}$$

where W_i is the weight applied to each data point.

Various weight functions have been proposed over the years [20–22] with Bergman's (Eq. (7a), [20]) and Faucher & Tyson's weight function (Eq. (7b), [21]) being mostly used. Faucher and Tyson's (*F&T*) was found to produce the most accurate estimates for data sets produced with Monte Carlo simulation [22–24]. However, these studies disagree on the choice of estimator used in con-