Quality Engineering

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lqen20

Utilizing jackknife and bootstrap to understand tensile stress to failure of an epoxy resin

R. Caro-Carretero, A. Carnicero, J. R. Jiménez-Octavio & D. Cousineau

To cite this article: R. Caro-Carretero, A. Carnicero, J. R. Jiménez-Octavio & D. Cousineau (11 Dec 2023): Utilizing jackknife and bootstrap to understand tensile stress to failure of an epoxy resin, Quality Engineering, DOI: 10.1080/08982112.2023.2286500

To link to this article: https://doi.org/10.1080/08982112.2023.2286500

Published online: 11 Dec 2023.

QUALITY ENGINEERING
https://doi.org/10.1080/08982112.2023.2286500

Utilizing jackknife and bootstrap to understand tensile stress to failure of an epoxy resin

R. Caro-Carretero[a], A. Carnicero[b], J. R. Jiménez-Octavio[c], and D. Cousineau[d]

[a]Organizational Industrial Department, ICAI-Higher Technical School of Engineering, Universidad Pontificia Comillas, Madrid, Spain; [b]Department of Mechanical Engineering, Universidad Pontificia Comillas, Madrid, Spain; [c]Institute for Research in Technology, Madrid, Spain; [d]Université d'Ottawa, Ottawa, Canada

ABSTRACT

A study was conducted on the tensile stress of an epoxy resin (Resoltech® 1050/1056). This was done by gathering a sample of 39 tensile strength data under consistent levels of stress. The tensile stress resistance is often characterized using a three-parameter Weibull distribution and the reliability of this characterization, given by confidence intervals (CIs). This approach commonly utilizes data-resampling techniques to estimate the CI of its parameters. CIs are constructed from six existing point-estimation methods. Herein, the jackknife was carried out to calculate the CIs using 39 subsamples and bootstrap methods using 100 or 200 subsamples. To date, there have been no studies exploring the effectiveness of subsampling methods for constructing CIs related to tensile strength. In this study, jackknifed and bootstrapped samples are used to implement the percentile method and three variations of the bias correction methods. We then performed simulations to evaluate the reliability of these methods using a Weibull random number generator. Our results showed that while the bias-corrected approach generated the most stable CIs from replicate samples, its accuracy was contingent on the point-estimation method employed. We also found that the different methods for calculating CIs resulted in significantly varying widths of the CIs.

KEYWORDS
jackknife; bootstrap confidence intervals; three-parameter Weibull distribution; point-estimation methods

1. Introduction

Epoxy-based laminating systems are becoming popular because of their high strength, durability and low shrinkage during curing procedure, providing lighter and more resistant materials. Epoxy polymers are also being employed as adhesives or matrices in extensive industrial applications regarding the manufacturing processes of automobile, aeronautical and naval engineering, etc. As epoxy resins are being used more frequently in load-bearing structures where the safety of personnel or property is of great importance, relying solely on point-wise estimates is inadequate. Depending only on individual, isolated estimates or data points is not sufficient or reliable for making informed decisions or drawing conclusions in certain fields like statistics, risk assessment, decision-making and scientific research. In this regard, tensile strength, a fundamental material property, plays a pivotal role in industries. Accurate estimates of tensile strength are critical for ensuring the safety and reliability of products and structures. Confidence intervals (CIs hereafter) are essential statistical tools that provide a measure of the precision and uncertainty associated with these estimates. However, constructing reliable CIs for tensile strength is a complex endeavor due to the inherent variability in material properties and the often-limited availability of test specimens. Traditionally, constructing CIs for tensile strength has relied on classical statistical methods assuming that data follows specific parametric distributions. These methods, while useful in many scenarios, often struggle to capture the full complexity of material behavior, leading to potentially inaccurate or overly optimistic CIs. In light of these challenges, sub-sampling methods have emerged as promising alternatives for estimating CIs related to tensile strength. However, despite their potential advantages, there remains a significant gap in the literature (Mäki-Lohiluoma et al. 2021 and references herein). Previous studies have not adequately explored or evaluated the effectiveness of sub-sampling methods specific to tensile strength data.

To ensure a higher level of reliability, supplementary methods must be employed. Although prior tests

CONTACT R. Caro-Carretero rcaro@comillas.edu Organizational Industrial Department, ICAI- Higher Technical School of Engineering, Universidad Pontificia Comillas, Alberto Aguilera 23, 28015 Madrid, Spain.

have shown the mechanical performance of an adhesive joint, they do not allow quantitatively assessing its reliability. To achieve this objective, an analysis of experimental data using resampling techniques was conducted to assess the accuracy of CI estimations in terms of width and proportion of coverage of the true parameter. Application of this methodology to an epoxy resin provides insights into how different point-wise estimation methods may influence CIs width. Epoxy resin (Resoltech® 1050/1056) was chosen for this purpose because it has enhanced mechanical and thermal properties, making it a viable option for developing high-performance lightweight structures reinforced with glass, carbon, aramid and basalt. Stress rupture measured by tensile strength methods is a time-dependent failure mode. The variation in the strength of epoxy resin has been analyzed assuming that different tensile strengths follow a three-parameter Weibull distribution. Weibull statistics were successful in materials science for modeling this type of data. The Weibull distribution's adaptability to various types of data is the primary reason for its widespread use in a broad range of applications. Also, the probability distributions governing failures in materials exhibit a non-Gaussian behavior with a large tail on the high-strength side.

Engineering literature features a large number of studies that characterize strength data concerning flaw types and sizes determined through fractography using Weibull parameters (Quinn and Quinn 2010). The Weibull distribution is also widely- used for example to represent the brittle fracture failure of ceramic components (Garrido et al. 2019 and references therein). Wong et al. (2006) developed a methodology for estimating Weibull parameters to examine the microstructural factors contributing to a failure in a brittle rock model. In addition, the fracture of concrete structures was analyzed by assigning distinct tensile strengths to beams in a regular triangular lattice, using number generators based on the Weibull distribution instead of the Gaussian distribution (van Mier, van Vliet, and Wang 2002). Parambil and Gururaja (2017 and references herein) conducted an in-depth examination of the gradual development of damage at the micro-scale in polymer composites when subjected to longitudinal loading (loading along the direction of the fibers). Testing the tensile strength of individual fibers and bundles of fibers revealed that Weibull fits were representative of the stochastic fiber fragmentation characteristics (Parambil and Gururaja 2015). created a 3D model of a repeating unit cell consisting of numerous fibers randomly distributed throughout a polymer matrix and subjected it to loading in multiple directions (multi-axial loading). A unique strength value was assigned to each fiber, based on the Weibull distribution (Naito et al. 2012). Alía et al. (2013) conducted an experiment that utilized the Weibull statistical model to suggest the optimal adhesive and surface treatment that provides the highest level of technical performance and reliability for adhesive joints between aluminum and composite materials (probability that an adhesive joint works appropriately under specific load). Towse et al. (1999) utilized the Weibull distribution to estimate the strength at specific locations within the adhesive joint and evaluated the impact of minor fluctuations in the local geometry on failure prediction within an idealized section of the joint. Arenas, Narbón, and Alía (2010) assessed the optimal thickness of an acrylic adhesive in simple single-lap joints carried out using the Weibull distribution. Hadj-Ahmed, Foret, and Ehrlacher (2001) proposed a Weibull model to estimate the shear strength of an adhesive joint, using a similar approach.

1.1. Three-parameter Weibull approach

The two-parameter Weibull distribution has been extensively studied and refined in formal standards, norms, and scientific literature. Its popularity is mainly attributed to its simplicity and ease of use (Nourbakhsh et al. 2014). However, with the three-parameter Weibull distribution, data can be fitted with less dispersion. International standards do not provide a procedure for estimating the parameters of this distribution although (Akram and Hayat 2014; Cousineau 2009a; Goda, Kudaka, and Kawai 2011; Qin, Zhang, and Yan 2012), among other studies, present different methods to estimate them. Nevertheless, there is no consensus among the scientific community regarding the most effective method for estimating the parameters of this distribution. The challenge of obtaining accurate parameter estimates has led to the development of numerous techniques to address this issue. One such approach is presented in Garrido et al. (2019), which evaluates and analyzes six existing methods for fitting the three-parameter Weibull distribution to glass ceramic data. The study also introduces the use of neural networks for this purpose.

The definition of the cumulative distribution function for the three-parameter Weibull distribution is

$$F(t) = 1 - e^{-\left(\frac{t-t_0}{\eta}\right)^{\beta}} \quad (1)$$

in which the parameters are the shift (or location) parameter t_0, the scale parameter η and the shape parameter β, which is also called the Weibull modulus. When t_0 is set to 0, the resulting distribution is the two-parameter Weibull distribution.

Two common techniques for parameter estimation are the L-Moment method (LM, hereafter) and the maximum likelihood estimation (MLE, hereafter). Moment-based methods have a well-established history in statistics because they are typically less computationally complex and require fewer iterations compared to other techniques. However, they are not always satisfactory (Akram and Hayat 2014). When the data contains outliers or follows a heavy-tailed distribution, moment-based estimators of the Weibull parameters may produce biased or unreliable results. Outliers can excessively impact the moments and lead to poor parameter estimates. When dealing with truncated or censored data, alternative methods such as MLE or Bayesian estimation may be more suitable for estimating Weibull parameters. In situations where the available data is limited, other methods like MLE, which are known for their efficiency in small samples, may be preferred. For data with complex dependencies, specialized techniques like MLE with suitable models for dependence structures might be more appropriate. Moment-based methods do not naturally incorporate prior information or expert knowledge about the parameters, which could be valuable in some applications. While Modified Moment Estimators (MME, hereafter) are generally more efficient than moment-based estimators, the latter are easier to compute. Nonetheless, as with the LM approach the MLE technique may not always yield satisfactory results as the *regularity conditions* are not always met (Bartkutė and Sakalauskas 2008; Cheng and Amin 1983). Furthermore, the solutions obtained through MLE may display biases that vary based on the sample size and cannot be determined precisely. Finally, closed-form solutions for two out of three parameters are not available using this method (Cousineau 2009a). As a result, more elaborate and unorthodox techniques are required. The modified maximum likelihood estimators (MMLE, hereafter), as defined in Cohen and Whitten (1982), are designed to generate reliable estimates of Weibull's parameters in situations where MLE is inadequate. These estimators are formulated by incorporating additional restrictions into the log-likelihood function (Cheng and Amin 1983). In standard MLE, the shape parameter of the Weibull distribution can be unconstrained, leading to estimation difficulties, especially when the estimated shape parameter falls outside a meaningful range (e.g., negative values). MMLE may incorporate restrictions on the shape parameter to ensure it falls within a specified range, such as $\beta > 0$ to guarantee a proper Weibull shape. In cases where the data is not well-described by a Weibull distribution, MMLE may involve transforming the data or applying a different parametric form (e.g., log-transformed data to fit a linear model) to better conform to the distributional assumptions. In addition, MMLE may involve adjustments to the likelihood function to account for the presence of censored or truncated observations. This might include incorporating indicator functions to handle the truncated or censored data appropriately. In some cases, MMLE may seek to maximize the posterior distribution, incorporating prior distributions for the parameters.

Following Garrido et al. (2019), for the simulations that follow, we will employ the MMLE approach, which involves maximizing the log-likelihood, considering that

$$1 - e^{-\left(\frac{t_r - t_0}{\eta}\right)^{\beta}} = r/(N+1) \quad (2)$$

where, in a random sample of size N, t_r refers to the r^{th} order statistic. Setting r to 1, we get:

$$-\ln\left(\frac{n}{n+1}\right) = \left(\frac{t_1 - t_0}{\eta}\right)^{\beta} \quad (3)$$

Cousineau (2009a) examines alternative techniques to maximize the likelihood function, including the Maximum Product of Spacing (MPS hereafter) approach (Cheng and Amin 1983) and the Weighted Maximum Likelihood Estimation technique (w-MLE hereafter) (Cousineau 2009b). The MPS estimation technique provides consistent estimators under a broader range of conditions compared to MLE estimators. It can be regarded as a method that overcomes the aforementioned difficulties associated with MLE while preserving its fundamental characteristics. This approach involves maximizing the likelihood of the spacing between two consecutive data points (Cheng and Amin 1983; Qin, Zhang, and Yan 2012). Regarding w-MLE, Cousineau (2009b) demonstrates the process of computing weights that counteract the biases inherent to the MLE function.

In this manuscript, we present a bootstrap statistical analysis of the six existing point-estimation methods described above, namely LM, MLE, MME, w-MLE, MMLE, and MPS when they are used to build CIs. As of now, there has been no investigation conducted on the effectiveness of bootstrap confidence interval techniques in evaluating tensile strength. Although it is crucial to conduct a reliability analysis of brittle materials, most studies were focused on point estimates of the parameters (Zapata-Ordúz, Portela, and Suárez 2014). Estimated Weibull parameters in an attempt to investigate the concrete tensile strength using MLE, MME and w-MLE. Thus, in this study, we explore various bootstrap CI methods to assess their precision in estimating the tensile stress to failure of an epoxy resin. The code necessary to reproduce the analyses discussed in this manuscript is accessible in Caro-Carretero (2023). In the development of the resampling techniques, various alternatives have been considered such as Leave-One-Out Cross-Validation (Webb et al. 2011), Oversampling (Oliveira et al. 2021), Undersampling (Hoyos-Osorio et al. 2021; Oliveira et al. 2021) or Combined Sampling (Oliveira et al. 2021). The choice of the jackknife and bootstrap resampling methods was motivated by the need for robust and computationally feasible techniques for estimating the sampling distribution of statistics. This is particularly relevant here since the assumptions required for classical statistical methods (such as normality) are not met. Furthermore, these methods have become popular in statistics due to their flexibility and ability to handle a wide range of situations.

2. Experimental data

The sample of epoxy resin (Resoltech® 1050/1056) tested consists of 39 specimens. Empirical data on tensile strength was gathered at constant stress levels ranging from 54 MPa to 105 MPa. The maximum strength of this adhesive, according to the technical datasheet, is 81.5 MPa after 14 days of curing at 23 °C while a value of 97.2 MPa is reached with 16 h at 60 °C.

The international standards devoted to the determination of tensile properties in plastics, ISO 527-1 (2019) and ISO 527-2 (2012), whose ASTM equivalent standard is D638-14 (2022), have been followed in this research. Such standards encompass the principle, methods and apparatus for testing, the preparation, conditioning and measuring of test specimens, etc. Therefore, in spite of not being directly applicable to epoxy resins, they provide a reliable framework to define the procedure for our purpose. They also describe the shape and dimensions of the test specimens, shown in Figure 1.

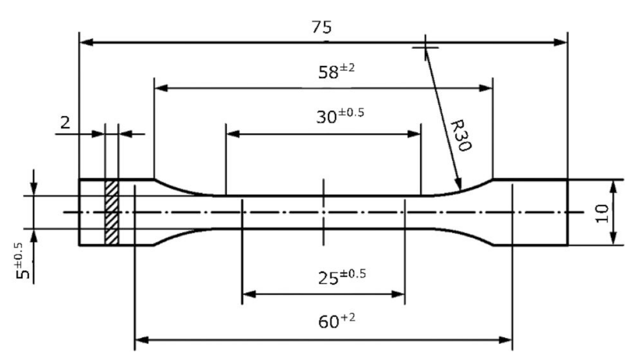

Figure 1. Geometry and size of the specimens according to ISO 527-2 (2012). Distances are given in millimeters (mm).

This standard D638-14 refers to the molding procedures and conditions set out in the relevant international standards for the corresponding material. It also refers to the strict control of all the conditions of preparation of the specimens of each batch and to the examination of their surfaces, which must be free from visible defects, scratches and other imperfections. The positioning of the samples on the grips is also specified, taking care to align the longitudinal axis of the grips with the axis of the testing machine, the pre-stress requirements, the set-up, adjustment and calibration of the extensometers as well as the test speed. Adhering to such a detailed procedure, the stress to failure of 39 pieces was recorded, which are shown in Table 1.

An additional factor to consider when employing the Weibull as a lifetime distribution in strength reliability is the number of test specimens that are needed to adequately estimate the parameters. The appropriate number of test specimens to use is influenced by several factors, such as material and testing expenses, the distribution parameter values and the desired precision for a particular application. Standards provide useful tables and calculations to aid in assessing this number. In the absence of specific requirements, a general rule (no strict rule specifying a minimum sample size) is that about 30 test specimens are enough to estimate Weibull distribution parameters. Nevertheless, utilizing more specimens will enhance the accuracy of the estimates (Quinn and Quinn 2010 and references herein). In practice, the sample size you can obtain may be limited by resource constraints or data availability. In such cases, you may need to work with the data you have and be aware of the

Table 1. Epoxy resin date.

54.403	56.537	64.442	64.759	66.082	66.686	67.442	69.526	70.316
73.323	73.395	73.517	74.144	74.302	75.028	75.269	75.275	76.214
78.150	80.212	80.378	80.950	81.875	82.674	85.489	86.301	87.466
87.773	92.125	93.317	94.576	94.966	97.715	98.706	99.408	101.852
104.069	104.774	104.829						

Note: The tensile stress to failure ranging from 54.4 to 104.8; the mean of this sample is 81.2 MPa, its standard deviation is 13.5 MPa and its skewness is 0.14.

limitations in the parameter estimates. Additionally, resampling methods, such as bootstrapping can be useful even with relatively small sample sizes.

3. Materials and methods

3.1. Resampling technique

Jackknife and bootstrap are based on the concept of resampling, where inferences about a population can be made from sample data by resampling or subsampling the data and performing inference about the population from the resampled data. This approach is appealing because it is versatile and can be utilized in various scenarios involving intricate data structures, both in parametric and non-parametric contexts. In addition, these techniques can be applied to statistics that originate from non-symmetric distributions, which is particularly useful for performance measures related to adhesive failure tension, as these cannot be easily transformed into a Gaussian distribution. This article explores two resampling techniques that can be employed to determine CI around point estimates for the adhesive failure tension model.

The accuracy of any particular point-estimation approximation is not known precisely. In statistics, a CI is a form of estimation that is derived from the observed data. It puts forward a range of credible values for an unknown parameter that lies between a lower and an upper limit with a given level of confidence (usually 95% confidence level is utilized).

Because the material parameters are nonlinearly interdependent, analytical methods to obtain the confidence intervals of the reliability or failure probability of the components are more complicated. Herein, resampling approaches are used instead, known as the jackknife and the bootstrap methods, which are based on the resampling of experimentally observed data.

3.1.1. The jackknife technique

The jackknife estimator for a parameter is obtained by systematically excluding one observation at a time from a dataset and computing the estimate based on these subsets. Afterwards, the jackknife estimate is the mean of these computations. This method for estimating parameters of data sets was originally introduced by Quenouille (1949) and later improved by Tukey (1958).

The jackknife can be viewed as a linear approximation of the bootstrap, as well as other subsampling techniques (Efron and Tibshirani 1993; Tukey 1958). These methods are built on the concept of estimating the distribution of a large population using only a small sample; they are especially useful when traditional methods are impractical, infeasible or inapplicable.

A significant advantage of this technique is that data analysis is performed without any prior assumptions regarding the population distribution. This is a desirable advantage particularly when the data's underlying distribution is uncertain or there is a suspicion that the data may not conform to a specific distribution. The jackknifed datasets provide an estimate of the theoretical distribution of the observed data and can be viewed as treating the sample as if it represents the entire population. Hence, this approach relies on the assumption that the jackknifed distribution of errors is a reliable approximation of the actual distribution of sampling errors, as outlined in Efron and Tibshirani (1993).

The basic jackknife algorithm can be schematized as follows:

We define a random variable T whose unknown distribution function, $P_f = P(T \leq t)$, depends on a population parameter θ that we want to estimate with an estimator $\hat{\theta}^*$.

1. Conduct the experiment and collect data $T = \{t_1, t_2, t_3 \ \ldots, t_N\}$ as the original sample of size N. 2. Generate a new data set, $T^* = \{t_1{}^*, t_2{}^*, t_3{}^*, \ldots, t_{N\text{-}1}{}^*\}$, from the original (i.e., copy each element except one). For each iteration, exclude a distinct element (or a distinct set of elements) from the original dataset T to form slightly varied datasets, T^*. 3. Compute the estimate of θ^* as t^* using the jackknifed dataset. 4. Perform steps 2 and 3 n times, excluding a different datum on each iteration. In this study, the observed dataset is subject to multiple repetitions. Quinn and Quinn (2010) and references herein found that the average sample size in test specimens is approximately 20. 5. Estimate the intervals limits using the distribution of t^*.

In an attempt to compare the different estimation methods and resampling techniques, the observed data is used to obtain the failure probability $F(t) = P_f$, generating jackknifed replications $P_f\ (t^*)$. The empirical probability distribution is constructed putting a mass of $1/(N - 1)$ to each point of the subsamples.

3.1.2. The bootstrap technique

The jackknife resampling is generally considered suitable for small data samples as in our case. We also examined an alternative technique called bootstrap. To implement bootstrap, it is necessary to construct samples that are of equal size to the original sample. To achieve this, subsampling is done with replacement. Hence, some of the data might be present more than once in the subsample while others may be missing entirely from the subsample. One advantage of bootstrap is that all the subsamples have identical sizes to the original sample. Consequently, if the parameter to be estimated has a biased estimator, and if the bias depends on sample size, then the influence of this bias is absent from the bootstrap technique.

Here, to perform bootstrap, we elected to generate 100 and 200 bootstrapped samples all with the same size as the original sample, taken using random sampling with replacement to approximate repeated sampling from the population (given that there are $39^{39} > 10^{62}$ possible subsamples of size 39 with replacements, these bootstrapped samples are far from exhausting the possible subsamples).

Resampling techniques are frequently employed by researchers as a broad approach to constructing confidence intervals for a variety of commonly used statistics. Bootstrapping is primarily based on drawing sub-samples from either the original sample (non-parametric bootstrapping) or from a model fitted to the original sample (parametric bootstrapping). This article will focus on non-parametric bootstrapping, which is more commonly used. It does not rely on any assumptions about the characteristics of the underlying population, and solely utilizes information from the original sample (Puth, Neuhäuser, and Ruxton 2015).

3.2. Deriving confidence intervals from jackknife and bootstrap distribution

When working with a sample from the population of interest, it is not possible to determine the population parameter with absolute certainty. Therefore, a confidence interval is constructed to estimate a range of values that a parameter of interest in a population may take, based on a random sample of data.

With relatively small sample sizes (Chernick and Labudde 2011), we are at risk that CIs have coverage lower than the specified nominal level (e.g., 95%). In fact, Carpenter and Bithell (2000) emphasize that this issue might be more severe for 99% confidence intervals generated *via* bootstrapping, particularly if the sample size is not sufficiently large ($n = 50$ as a minimum). Hence, in this study, we would caution against using 99% CIs because of the sample size of the observed data set.

The objective of this work is to evaluate the bootstrap CIs of the parameters related to the three-parameter Weibull distribution. To test this approach, we will utilize the data on the breaking stress of an epoxy resin described earlier. There are various bootstrapping techniques to determine a confidence interval for a population parameter. In this study, we concentrate on the two most widely used methods: percentile bootstrap (PB) and bias-corrected bootstrap (BC) confidence intervals (Chernick and Labudde 2011; Davison and Hinkley 1997; Efron and Tibshirani 1993). Bootstrap CIs are constructed from six proposed point-estimation methods where θ denotes a parameter of interest while α indicates the significance level (e.g., $\alpha = 0.05$ for 95% CIs; $1 - \alpha$ is commonly called the *confidence level*). Further details of these methods can be found in Efron and Tibshirani (1993).

Manly (2007) provides references to a number of other methods; the techniques outlined below are both the most commonly employed and the simplest to put into practice. In essence, one way to enhance the percentile method is to address any bias that may arise due to the true parameter value not being the median of the distribution of bootstrapped estimates. As pointed out by Caro-Carretero (2023) and Ruxton and Neuhäuser (2013), there is no explanation for why researchers choose to use a particular bootstrap method to generate CIs. Nevertheless, the choice of which bootstrap CI method to use depends on the specific statistical problem, data characteristics and the underlying assumptions (Davison and Hinkley 1997; Karavarsamis et al. 2013). Each bootstrap CI method has its strengths and weaknesses. Namely, Percentile Bootstrap CI is simple to implement, widely applicable and suitable for a wide range of statistics and parameter estimations. However, it may not perform well for skewed distributions or in cases where there are outliers. Studentized Bootstrap CI adjusts the bootstrap samples using the estimated standard error of the statistic, taking into account the variability within each resampled dataset. In contrast, it can be computationally intensive, especially for complex statistics and may not be ideal for small sample sizes. Bootstrap-t CI is similar to the studentized bootstrap approach but uses the t-distribution to construct the CI, which can be more robust against outliers. Despite that, it can be computationally intensive and may still be affected by skewness in the data. Bias-Corrected Bootstrap CI addresses the bias in the bootstrap estimates by adjusting them with the estimated bias of the statistic. While it provides more accurate CIs, especially for small sample sizes, it can be sensitive to the choice of bias-correction method and may not perform well with extreme outliers. Bias-corrected and accelerated CI adjusts for bias and skewness by estimating acceleration and skewness factors from the bootstrap samples. Conversely, it can introduce additional uncertainty. Bootstrapped Quantile CI constructs CIs based on quantiles from the bootstrap distribution, which can be robust against outliers. Alternatively, it may not capture the full shape of the bootstrap distribution if the quantiles chosen are not appropriate. Bootstrapped CIs for Ratios designed specifically by resampling from the data while maintaining the structure of the ratio, although it has limited applicability to situations where ratio estimations are not relevant. Bootstrapped Bayesian CI combines the principles of Bayesian modeling with bootstrap resampling to construct credible intervals, allowing the incorporation of prior information when available, which can be subjective.

3.3. Methods for jackknife and bootstrap confidence intervals

Here, we briefly review the basic concepts underlying the two distinct methods of using bootstrapping to estimate CIs.

1. Percentile bootstrap method (Efron 1982; Efron and Tibshirani 1993)
   In the percentile confidence interval, the bootstrapped (or jackknifed) interval is determined by calculating the range between the percentiles of $100 \times \alpha/2$ and $100 \times (1 - \alpha/2)$ of the distribution of θ estimates obtained from resampling. To obtain a percentile CI for θ using bootstrapping (or jackknife), follow these steps: (1) create R random bootstrapped (or jackknife) samples where R is either 100 or 200 in the bootstrap simulations that follow and R = 39 with the jackknife simulations; (2) calculate a parameter estimate for each bootstrapped (or jackknifed) sample; (3) arrange all R bootstrap (or jackknife) parameter estimates from lowest to highest; and (4) construct the 95% CI as outlined below,

$$CI_{\theta} \in \left[\hat{\theta}^*_{(j)}, \hat{\theta}^*_{(k)}\right]_{95\%} \tag{4}$$

where $\hat{\theta}^*_{(j)}$ denotes the j^{th} estimate when sorted in ascending order (lower limit), and $\hat{\theta}^*_{(k)}$ denotes the k^{th} estimate (upper limit); $j = \left\lceil \frac{\alpha}{2} \times R \right\rceil$ and $k = \left\lceil \left(1 - \frac{\alpha}{2}\right) \times R \right\rceil$. This method assumes that as the sample size increases, the empirical distribution function derived from bootstrapping progressively approaches the genuine distribution function, and that the empirical quantiles conform to the law of large numbers. According to Efron (1982), in most cases the jackknife method yields less accurate results than the bootstrap and tends to be more cautious, resulting in marginally higher estimated standard errors. However, bootstrapping introduces an additional source of variation, caused by the finite number of possible subsamples of size R that can be generated. This limitation can be severe for large R and small samples. In order to determine the method with the best estimates, coverage probability (the proportion of intervals containing the true parameter) has been calculated using the international standards devoted to the determination of tensile properties (true parameters according to standards). It is common to establish a nominal coverage probability of 95%. For instance, a 95% percentile bootstrap CI generated using 100 bootstrap samples would encompass the range between the 2.5% quantile value and the 97.5% quantile value of the 100 bootstrapped parameter estimates.

2. Bias-corrected method (DiCiccio, Romano, and Wolf 2019)
   The bias-corrected method corrects for bias in the bootstrap (or jackknife) parameter estimates by adding a bias-correction factor z_0 defined as the fraction of the bootstrap (or jackknife) estimates that are smaller than the original parameter estimate

$$z_0 = \phi^{-1}\left(\frac{Prob* \ \{\hat{\theta}^* < \hat{\theta}\}}{R}\right) \tag{5}$$

where ϕ^{-1} denotes the inverse function of the standard normal cumulative distribution function, (e.g., ϕ^{-1} (0.975) = 1.96), and $\hat{\theta}$ is the point estimation of the parameter (labelled BC-point in Tables 2–4), obtained from the original sample. We also have used the bootstrap (or jackknife) estimate for mean (labelled BC-mean in Tables 2–4) and the median (labelled BC-median in Tables 2–4) from the jackknifed and bootstrapped replications to obtain z_0 when we calculate bias-corrected bootstrap (or jackknife) CIs for the three parameters of the Weibull distribution.

In sum, we explored 72 different methods for building confidence intervals, that is, six-point estimate methods (LM, MLE, MME, w-MLE, MMLE, and MPS) × 3 subsampling techniques (jackknife or bootstrap with 100 and 200 subsamples) × 4 methods (percentile, BC-point, BC-mean, and BC-median). All these techniques are used to get CIs for the three parameters shown in three distinct tables.

To assess their coverage performance, we conducted a Monte Carlo simulation. CIs offer an estimation of the plausible range in which the true value of the statistic is expected to lie. A narrow CI suggests low variability of the statistic, which validates stronger

Table 2. Estimates and confidence intervals for the time-to-failure in a Weibull distribution shape parameter.

	Method					
Jackknifed sample = 39	LM	MLE	MME	w-MLE	MMLE	MPS
Point estimate	2.612	2.792	4.168	2.722	3.374	3.161
Jackknife estimate for mean	2.648	2.780	4.173	2.712	3.408	3.209
Jackknife estimate for median	2.719	2.771	4.076	2.702	3.426	3.202
95% confidence interval						
Percentile						
Lower limit	2.450	2.684	3.963	2.6	3.205	2.998
Upper limit	2.789	2.981	4.662	3	3.737	3.408
Width	0.338	0.297	0.698	0.4	0.531	0.410
BC-point						
Lower limit	2.309	2.684	3.999	2.621	3.205	2.998
Upper limit	2.786	2.981	4.830	3	3.676	3.394
Width	0.476	0.297	0.831	0.379	0.471	0.395
BC-mean						
Lower limit	2.309	2.684	3.999	2.621	3.205	2.998
Upper limit	2.786	2.981	4.830	3	3.676	3.408
Width	0.476	0.297	0.831	0.379	0.471	0.410
BC-median						
Lower limit	2.450	2.684	3.963	2.6	3.056	2.998
Upper limit	2.789	2.981	4.662	3	3.737	3.408
Width	0.338	0.297	0.698	0.4	0.531	0.410
	Method					
R bootstrap sample = 100	LM	MLE	MME	w-MLE	MMLE	MPS
Point estimate	2.612	2.792	4.168	2.722	3.374	3.161
Bootstrap estimate for mean	2.646	2.800	4.318	2.694	3.368	3.154
Bootstrap estimate for median	2.690	2.824	4.190	2.674	3.328	3.126
95% confidence interval						
Percentile						
Lower limit	2.438	2.677	3.964	2.5	3.205	2.997
Upper limit	2.786	2.969	5.446	2.862	3.632	3.408
Width	0.347	0.292	1.482	0.362	0.427	0.410
BC-point						
Lower limit	2.309	2.673	3.963	2.5	3.205	2.998
Upper limit	2.785	2.963	4.830	3	3.671	3.408
Width	0.475	0.290	0.866	0.5	0.465	0.409
BC-mean						
Lower limit	2.438	2.673	4.025	2.5	3.205	2.998
Upper limit	2.786	2.963	6.032	2.862	3.632	3.408
Width	0.347	0.290	2.007	0.362	0.427	0.409
BC-median						
Lower limit	2.438	2.677	3.966	2.5	3.205	2.997
Upper limit	2.786	2.969	6.032	2.862	3.632	3.408
Width	0.347	0.292	2.065	0.362	0.427	0.410
	Method					
R bootstrap sample = 200	LM	MLE	MME	w-MLE	MMLE-1	MPS
Point estimate	2.612	2.792	4.168	2.7221	3.374	3.161
Bootstrap estimate for mean	2.628	2.791	4.289	2.6933	3.405	3.169
Bootstrap estimate for median	2.632	2.787	4.154	2.6735	3.414	3.165
95% confidence interval						
Percentile						
Lower limit	2.438	2.673	3.966	2.5	3.205	2.997
Upper limit	2.786	2.974	4.860	3	3.676	3.400
Width	0.347	0.300	0.893	0.5	0.471	0.403
BC-point						
Lower limit	2.396	2.673	3.966	2.6	3.204	2.997
Upper limit	2.786	2.979	4.860	3	3.671	3.394
Width	0.390	0.305	0.893	0.4	0.466	0.397
BC-mean						
Lower limit	2.438	2.673	4.026	2.5	3.204	2.997
Upper limit	2.786	2.979	6.032	3	3.671	3.408
Width	0.347	0.3056	2.005	0.5	0.466	0.411
BC-median						
Lower limit	2.438	2.739	3.966	2.5	3.210	2.997
Upper limit	2.786	2.97	4.860	3	3.676	3.400
Width	0.347	0.305	0.893	0.5	0.466	0.403

Note: The point estimates are the same in all the three segments of the table.

Table 3. Estimates and confidence intervals for the time-to-failure in a Weibull distribution scale parameter.

	Method					
Jackknifed sample = 39	LM	MLE	MME	w-MLE	MMLE-1	MPS
Point estimate	37.604	38.287	54.263	38.238	46.816	45.642
Jackknife estimate for mean	38.013	38.065	54.114	38.242	47.218	46.295
Jackknife estimate for median	39.017	37.643	53.098	38.100	46.669	45.910
95% confidence interval						
Percentile						
Lower limit	35.293	36.998	50.741	36.516	45.026	43.726
Upper limit	40.149	40.154	59.504	41.603	50.700	48.482
Width	4.855	3.155	8.762	5.087	5.674	4.756
BC-point						
Lower limit	32.485	37.025	52.657	36.645	45.053	43.726
Upper limit	40.142	40.154	61.043	41.603	50.700	48.482
Width	7.656	3.128	8.386	4.958	5.646	4.756
BC-mean						
Lower limit	32.485	36.998	52.550	36.645	45.053	43.726
Upper limit	40.142	40.154	61.043	41.603	50.700	48.482
Width	7.656	3.155	8.493	4.958	5.646	4.756
BC-median						
Lower limit	35.293	36.998	50.741	36.516	45.026	43.726
Upper limit	40.149	40.154	59.504	41.603	50.700	48.482
Width	4.855	3.155	8.762	5.087	5.674	4.756
	Method					
R bootstrap sample = 100	LM	MLE	MME	w-MLE	MMLE-1	MPS
Point estimate	37.604	38.287	54.263	38.238	46.816	45.642
Bootstrap estimate for mean	38.134	38.315	55.772	37.897	46.804	45.673
Bootstrap estimate for median	38.441	38.026	54.966	37.911	4.641	45.557
95% confidence interval						
Percentile						
Lower limit	35.304	36.998	52.516	34.985	45.024	43.711
Upper limit	40.149	40.154	66.636	39.635	49.808	48.456
Width	4.844	3.155	14.120	4.650	4.784	4.744
BC-point						
Lower limit	32.485	37.006	52.507	36.432	45.053	43.723
Upper limit	39.916	40.549	61.043	41.609	50.169	48.438
Width	7.4303	3.148	8.536	5.177	5.116	4.715
BC-mean						
Lower limit	32.485	37.006	54.059	34.985	45.053	43.723
Upper limit	40.076	40.154	71.942	39.635	50.169	48.438
Width	7.591	3.148	17.882	4.650	5.116	4.715
BC-median						
Lower limit	35.293	36.998	52.507	34.985	45.024	43.723
Upper limit	40.149	40.154	61.331	39.635	49.808	48.438
Width	4.855	3.155	8.823	4.650	4.784	4.715
	Method					
R bootstrap sample = 200	LM	MLE	MME	w-MLE	MMLE-1	MPS
Point estimate	37.604	38.287	54.263	38.238	46.816	45.642
Bootstrap estimate for mean	37.854	38.265	55.440	37.925	47.194	45.800
Bootstrap estimate for median	38.333	38.080	54.390	37.680	46.744	45.557
95% confidence interval						
Percentile						
Lower limit	35.293	36.998	50.741	35.899	45.026	43.723
Upper limit	40.149	40.154	61.331	41.603	50.223	48.482
Width	4.855	3.155	10.589	5.704	5.196	4.759
BC-point						
Lower limit	32.485	37.006	50.741	36.569	45.101	43.726
Upper limit	40.053	40.154	61.331	41.609	50.700	48.482
Width	7.567	3.148	10.589	5.040	5.598	4.756
BC-mean						
Lower limit	33.759	37.006	53.064	36.301	45.101	43.726
Upper limit	40.142	40.154	71.942	41.606	50.700	48.482
Width	6.383	3.148	18.878	5.304	5.598	4.756
BC-median						
Lower limit	35.293	36.998	50.741	36.301	45.101	43.726
Upper limit	40.149	40.154	61.331	41.606	50.700	48.482
Width	4.855	3.155	10.589	5.304	5.598	4.756

Note: The point estimates are the same in all the three segments of the table.

Table 4. Estimates and confidence intervals for the time-to-failure in a Weibull distribution location parameter.

	Method					
Jackknifed sample = 39	LM	MLE	MME	w-MLE	MMLE-1	MPS
Point estimate	47.832	47.179	31.936	47.212	38.655	40.426
Jackknife estimate for mean	47.415	47.336	31.967	47.197	38.209	39.885
Jackknife estimate for median	46.636	47.268	32.722	47.359	38.351	40.235
95% Confidence interval						
Percentile						
Lower limit	45.699	45.867	27.206	44.347	35.312	38.097
Upper limit	49.503	48.386	35.916	48.928	39.976	41.887
Width	3.803	2.519	8.709	4.815	4.663	3.789
BC-point						
Lower limit	45.703	45.863	25.734	44.345	35.312	38.126
Upper limit	53.162	47.921	33.388	48.219	39.976	41.887
Width	7.459	2.058	7.654	3.874	4.663	3.761
BC-mean						
Lower limit	45.699	45.86	25.734	44.345	35.312	38.092
Upper limit	49.503	48.38	33.388	48.219	39.950	41.475
Width	3.803	2.519	7.654	3.874	4.637	3.382
BC-median						
Lower limit	45.699	45.867	27.206	44.347	35.312	38.097
Upper limit	49.503	48.386	35.916	48.928	39.976	41.887
Width	3.803	2.519	8.709	4.581	4.663	3.789
	Method					
R bootstrap sample = 100	LM	MLE	MME	w-MLE	MMLE-1	MPS
Point estimate	47.832	47.179	31.936	47.212	38.655	40.426
Bootstrap estimate for mean	47.372	47.139	30.443	47.586	38.598	40.361
Bootstrap estimate for median	47.116	47.160	31.439	47.422	38.803	40.529
95% confidence interval						
Percentile						
Lower limit	45.699	45.839	20.291	46.289	36.168	38.128
Upper limit	49.503	48.386	33.492	50.934	39.976	41.886
Width	3.803	2.546	13.200	4.644	3.808	3.757
BC-point						
Lower limit	45.876	45.839	26.769	44.338	36.168	38.126
Upper limit	53.162	48.386	33.496	48.928	39.976	41.885
Width	7.286	2.546	6.727	4.590	3.808	3.759
BC-mean						
Lower limit	45.699	45.839	15.125	46.296	36.168	38.126
Upper limit	49.516	48.386	32.838	50.934	39.976	41.885
Width	3.816	2.546	17.712	4.638	3.808	3.759
BC-median						
Lower limit	45.699	45.839	15.125	46.289	36.168	38.131
Upper limit	49.503	48.386	33.488	50.934	39.976	41.885
Width	3.803	2.546	18.363	4.644	3.808	3.754
	Method					
R bootstrap sample = 200	LM	MLE	MME	w-MLE	MMLE-1	MPS
Point estimate	47.832	47.179	31.936	47.2121	38.655	40.426
Bootstrap estimate for mean	47.589	47.172	30.821	47.4681	38.277	40.261
Bootstrap estimate for median	47.493	47.237	31.732	47.4374	38.351	40.529
95% confidence interval						
Percentile						
Lower limit	45.700	45.852	25.458	44.3473	35.794	38.097
Upper limit	49.503	48.010	35.916	48.9506	39.966	41.885
Width	3.802	2.158	10.457	4.6033	4.172	3.787
BC-point						
Lower limit	45.749	45.839	25.458	44.3386	35.843	38.097
Upper limit	51.961	48.008	35.916	48.2798	39.976	41.851
Width	6.212	2.168	10.457	3.9412	4.133	3.754
BC-mean						
Lower limit	45.703	45.839	15.125	46.2899	35.312	38.097
Upper limit	49.516	48.008	33.388	48.9901	39.950	41.851
Width	3.813	2.168	18.262	2.7003	4.637	3.754
BC-median						
Lower limit	45.701	45.852	25.458	44.3473	35.794	38.126
Upper limit	49.503	48.010	35.916	48.9506	39.966	41.885
Width	3.801	2.158	10.457	4.6033	4.172	3.759

Note: The point estimates are the same in all the three segments of the table.

conclusions made from the analysis. The proposed methodology has been explored by using the real data of breaking stress of an Epoxy resin presented in section 2.

4. Results and discussion

4.1. Confidence interval estimates

The results of the CIs for the six three-parameter Weibull point-estimation methods with the two resampling methods, one using jackknife and two using bootstrap techniques, are presented in Tables 2–4. Table 2 depicts the CIs for the shape parameter from 39, 100 and 200 replicate samples. The same CIs for scale and location parameters are given in Tables 3 and 4, respectively.

The results show that for all resampling approaches the CI width is roughly similar for MLE and MPS based on point-estimation methods from 39, 100 and 200 replicated samples (highlighted in orange in Tables 2–4). More specifically, in the case of the shape parameter, the utilization of the median estimate indicated that the bias-corrected bootstrap method was the preferred approach since it gave the most consistent confidence intervals across 39, 100 and 200 replicated samples, but its reliability depended on the point-estimation method examined. The width (upper limit minus lower limit) of the CIs is much bigger (cells highlighted in blue) for MME method, resulting in more conservative CIs. Wider CIs suggest less precision (overcoverage) whereas narrower CIs may indicate undercoverage. Nevertheless, the estimated CIs have more similar width (they are more consistent) when the number of subsamples is higher, and that, for all the methods (both percentile and the three bias-corrected bootstrap techniques) and all the point-estimation methods. This result was expected as increasing the number of subsamples reduces the variations from sampling error so that all the techniques, owing to the law of large numbers, converge onto a unique estimate.

Regarding the scale parameter it was found that bias-corrected bootstrap approaches using the median estimate and the percentile bootstrap technique gave the most consistent confidence intervals from 39, 100 and 200 replicated samples, but its accuracy, as was the case for the shape parameter, depended on the point-estimation method examined. It appears that for MME method, the CI widths are larger (highlighted in blue). However, unlike bootstrap CIs for the shape parameter, there have more similar values for all the point-estimation methods except the LM ones, regardless of the number of subsamples (either 39, 100 or 200).

Finally, we see for the location parameter, as was the case for the scale parameter, that bias-corrected bootstrap approaches using the median estimate and the percentile bootstrap technique gave the most consistent confidence intervals from 39, 100 and 200 replicate samples. However, its reliability, as was found for the other two parameters, depended on the point-estimation method examined. For MME method the CI widths are wider (highlighted in blue). For the location parameter, bootstrap CIs have more similar values along all the point-estimation methods, when the replicate sample is 39, 100 and 200. It seems to indicate that the true parameter value is roughly the median estimate for the three pseudo samples.

Unlike previous research (Ruxton and Neuhäuser 2013), the percentile method does not tend to produce more conservative CIs, because it depends on the point-estimation method used. The MLE and MPS point-estimation techniques returned roughly similar CI widths. In addition, the use of bias-corrected method using the median estimate gives more consistent CIs than the remaining techniques.

4.2. Reliability curves

Figures 2–4 show the reliability curves of tensile strength for the epoxy resin. These figures provide illustrations of stress-to-failure distributions based on CIs of three parameters of Weibull distribution for the original sample consisting of thirty-nine values employing the six different methods and from 39, 100 and 200 replicate samples. We first focus on sample size of 39. The results reveal that all the bootstrap methods utilized for the six estimation techniques are conservative. This effect is relatively minor for sample sizes of 100 and 200 (Figures 3 and 4) in all the bootstrap methods and excepted with MME point estimation (highlighted in blue in Tables 2–4) whose confidence interval is notably conservative (as it is excessively large, above all bias-corrected method using the mean and median estimates).

Our example further demonstrates that when dealing with a single sample, the six point-estimation methods can generate CIs that vary significantly in size from one another. It is not unexpected that the usage of smaller sample sizes resulted in broader confidence intervals overall.

It is worth highlighting that the reliability curves overlap frequently in many figures with percentile and bias-corrected methods. It means the reliability performance

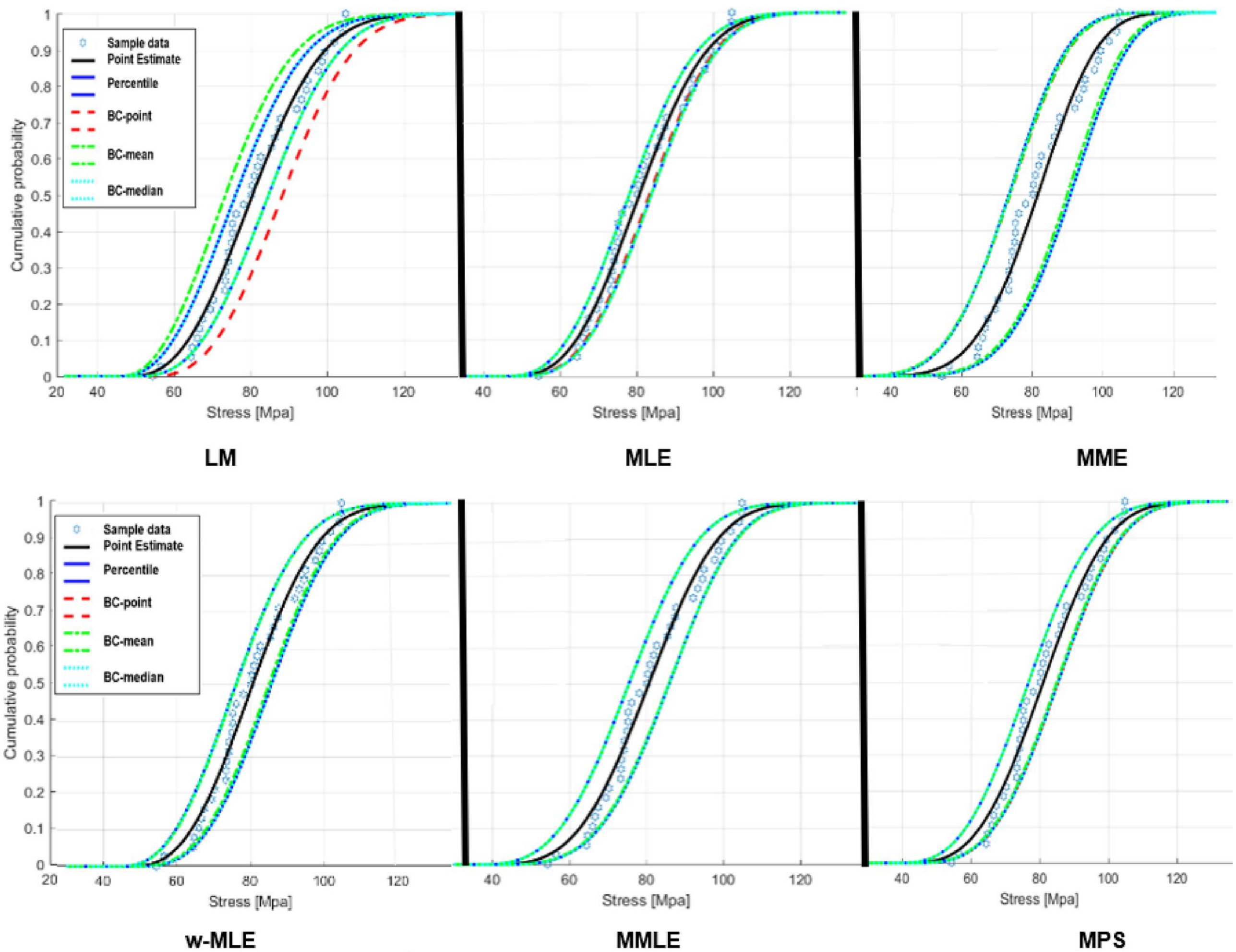

Figure 2. Reliability curves of the Weibull distribution according to bootstrap CIs derived from the percentile and bias-corrected methods. Number of subsamples = 39.

of these two methods is quite similar. In other words, there isn't a clear distinction in terms of which method consistently outperforms the other. This overlapping of reliability curves can present several implications: 1) the choice between them might introduce some uncertainty into the analysis 2) the overlap could also indicate that the quality of the data or the assumptions made in the analysis play a significant role in the performance of these methods. In such cases, improving data quality on the tensile stress of an epoxy resin may be more critical than choosing between the methods themselves. In summary, the overlapping of reliability curves between percentile-based and bias-corrected methods highlights the complexity of statistical analysis and the need for careful consideration when choosing and interpreting these methods in different scenarios. Nevertheless, a larger number of resamples can provide smoother and more clearly defined curves.

In contrast, the observation of similar reliability curves across different sub-sample sizes when using MMLE and MPS methods suggests robustness, consistency, and efficiency in these estimation techniques. This suggests that these methods are not highly sensitive to the number of sub-samples used for estimation. If so, it may have practical implications for decision-making and resource allocation. Depending on the context, it might be more cost-effective to use a smaller sample size (e.g., 39) if it yields similar reliability estimates as larger samples (e.g., 100 or 200). However, it's essential to consider the specific context, underlying assumptions, and potential tradeoffs when deciding on the appropriate sample size and estimation method for reliability analysis.

4.3. Monte Carlo with simulated data

In order to compare the various bootstrap techniques formally, we replaced the data from Table 1 with 39 simulated data generated with a Weibull distribution random number generator since the sample of epoxy resin tested consists of 39 specimens. The appropriate number of empirical data on tensile strength to use is

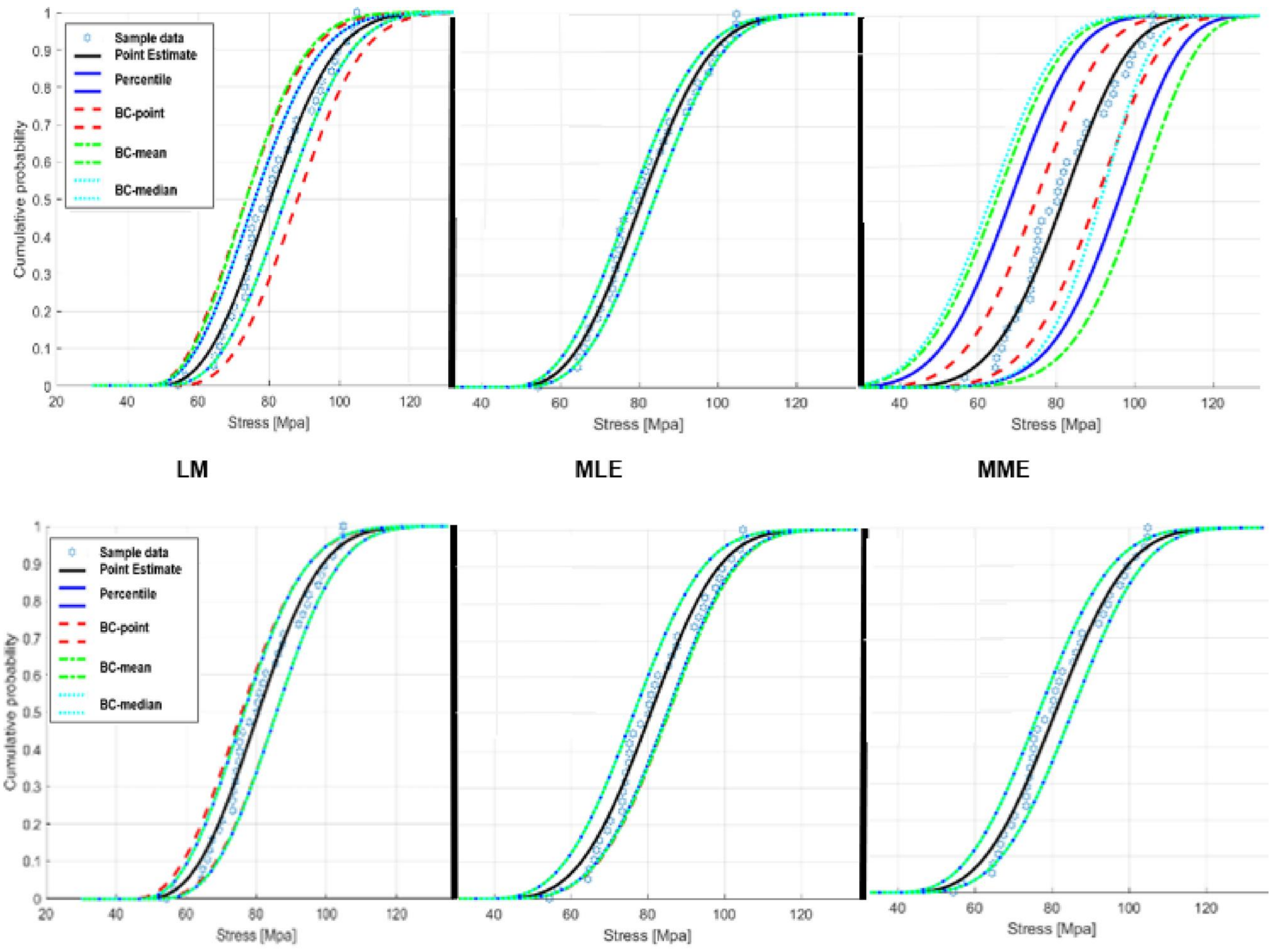

Figure 3. Reliability curves of the Weibull distribution according to bootstrap CIs from the percentile and bias-corrected approaches. Number of subsamples = 100.

influenced by testing expenses. The advantage is that now we know the true population parameters; it is therefore possible to compute the proportion of confidence intervals that indeed contain the true population parameters. We show in Figure 5 (in the order they appeared in Tables 2–4, that is, Percentile, BC-point, BC-mean, BC-median) the frequency with which 95% CIs do not include the true population value (D638-14 2022). Each simulation was repeated 5,000 times, each with a different simulated sample based on a Weibull random number generator (Manly 2007). The MATLAB codes utilized for data generation and functions for computing CIs are accessible in Caro-Carretero (2023).

For each method and value of Weibull point-estimation parameters, we assessed the proportion of CIs in which the true population value is outside the CI limits. We assessed separately the number of times the population value exceeds the upper limit of the interval (indicated in orange in Figure 5) and falls below the lower limit of the interval (indicated in blue in Figure 5). In an optimal scenario, a CI should be symmetric, with the population value exceeding the upper limit or falling below the lower limit in the same proportion, while also ensuring that the target level of coverage is met (Aguirre-Urreta, Rönkkö, and McIntosh 2019; Kyselý 2009).

In Figure 5, it can be observed that with a small sample size of 39, for MME point-estimation method, irrespective of which bootstrap techniques is used, the parameter of interest is included in the CI approximately 93% of the time (Chernick and Labudde 2009, 2011). The Percentile and BC-median bootstrap techniques also include the parameters of interest roughly 93% of the time over replications. For the remaining cases, we observed important deviations to 5% (12% in some case, up to 30% in others), indicating a lower than desired coverage level. Overall, the percentile method yielded CIs that were closer to the target coverage level of 95% than the bias-corrected bootstrap method. As mentioned earlier, it is not exactly 5% mainly because the sample is small. It is frequent

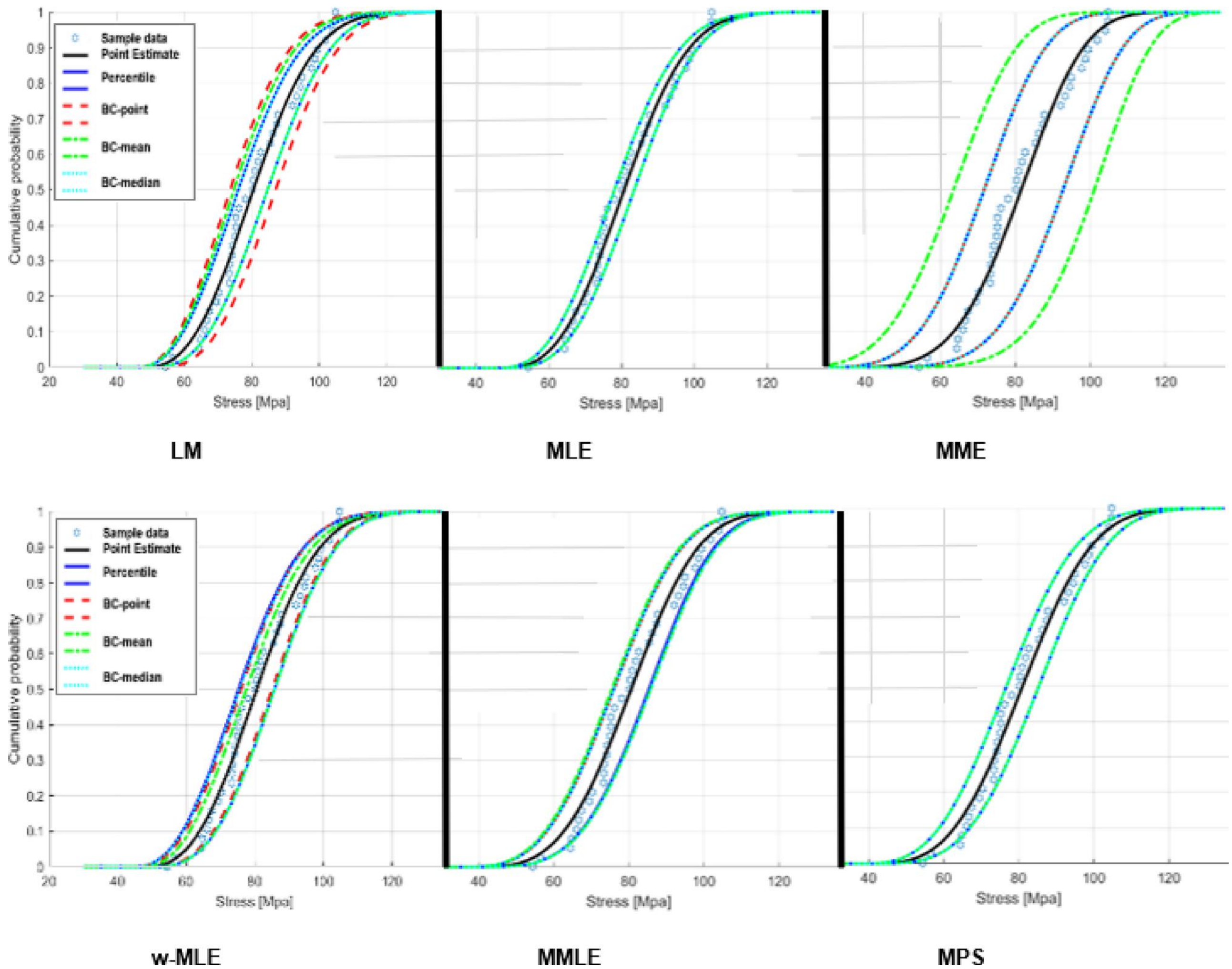

Figure 4. Reliability curves of the Weibull distribution plotted based on bootstrap CIs derived from the percentile and bias-corrected methods. Number of subsamples = 200. *Note:* Method 1 refers to percentile; 2 refers to BC-point; 3 refers to BC-mean; 4 refers to BC-median. The percentages below the bars are the sum of the red and blue bars.

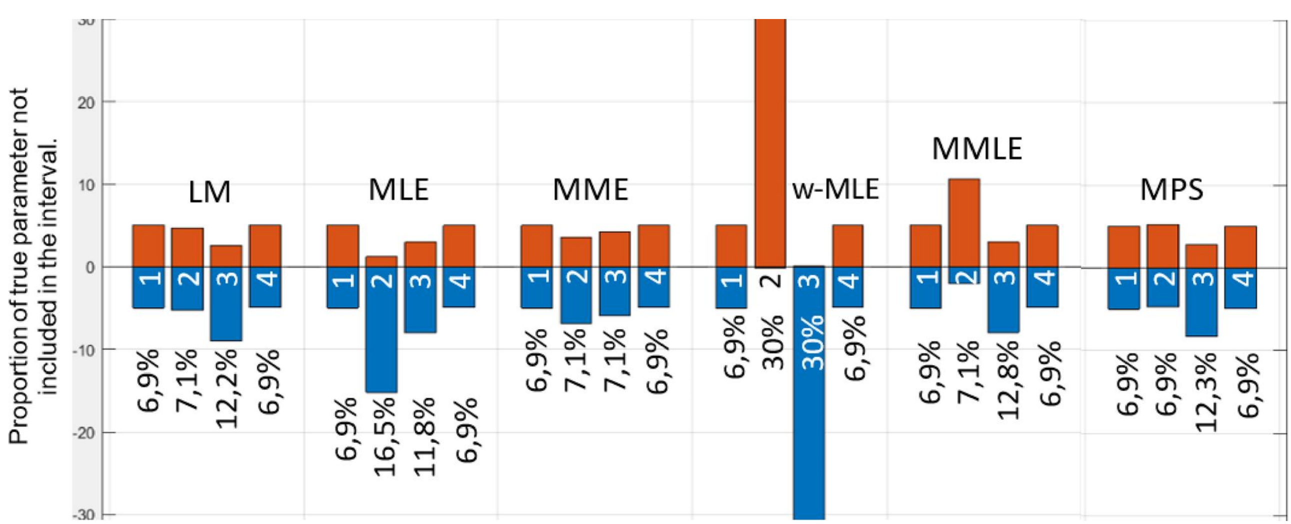

Figure 5. Proportion of samples for which calculated 95% CIs do not include the true population value of parameters within the CI limits based on 5,000 random samples.

to see CI estimates that are conservatives for small samples.

5. Conclusions

Although the experimental results allow knowing the adhesive joint's mechanical performance, a safe industrial application requires suitable reliability measures when the goods and safety of the staff is at risk. To achieve this purpose, subsampling techniques were carried out assuming a Weibull distribution consisting of three parameters to estimate CIs of its parameters providing example time-to-failure distribution. This procedure is in agreement with research carried out on samples of concrete with high strength (Zapata-Ordúz, Portela, and Suárez 2014) and is consistent with other works on brittle materials (Alqam, Bennett, and Zureick 2002) where it was reported that the three parameters Weibull distribution prevailed over the two parameters Weibull distribution. In studies that involve failure and damage development, the Weibull distribution is assumed to be a suitable model for the statistics of mechanical and failure properties (Cousineau 2011).

The findings of our study indicate that when calculating CIs, using six different point-estimation methods can result in significantly varying CIs for a given sample. However, while the choice of point-estimation methods can indeed influence the accuracy and properties of CIs, the effectiveness of a specific point-estimation method can vary depending on the data distribution, sample size and underlying assumptions. Even the presence of outliers can greatly affect point estimates. Therefore, it's essential to consider the context and characteristics of the data when selecting a point-estimation method. For data that follows a normal or approximately normal distribution, the sample mean is often the preferred point estimate and it leads to efficient CIs. When dealing with non-normal data, especially when it has heavy tails or outliers, consider using robust point estimates like the sample median or trimmed mean. In cases with small sample sizes, it's important to choose estimators that are less sensitive to extreme values. The sample median may be more suitable. For data with unknown or complex distributions, non-parametric point estimators like the sample median or quantiles can be valuable. Resampling techniques can help assess the robustness of point estimates and their impact on the resulting CIs under various scenarios.

In this regard, our simulations further demonstrate that the choice of point-estimation approach has a significant impact on the methodology's outcomes. The BC-point and BC-median returned the most consistent estimates across subsample sizes. Therefore, we suggest that researchers should explicitly disclose the bootstrapping method they employed when presenting such confidence intervals. An informal examination of the literature indicates that this is not a widespread practice at present. Additionally, researchers should state the number of bootstrap samples they used in their calculations. It is rare for researchers to specify the number of bootstrap samples used when reporting CIs obtained *via* bootstrapping or insufficient information was given to the reader to determine the number used, and amongst those that did, practice can be quite varied.

About the authors

Raquel Caro-Carretero, Professor in the Department of Industrial Organization at ICAI-Higher Technical School of Engineering (ICAI) of the Pontifical Comillas University in Madrid, has extensive teaching and research experience in various areas related to the application of statistics in engineering, economics and finance, tourism, migrations, disaster management, resilence and Space Sciences. She has taught classes at the Faculty of Economics and Business Administration of Comillas (ICADE) for 10 years. She has been recognized with two six-year research periods and is accredited by ANECA and ACAP.

Alberto Carnicero López studied at the Pontifical Comillas University (ICAI), where he graduated in 1995 and obtained his doctorate in 1999 under the supervision of Professor Enrique Alarcón and Ricardo Perera from the ETSII-UPM. He began his teaching career at ETSI-ICAI, where he has taught various subjects related to Continuum Mechanics. He has been recognized with two six-year research periods and is accredited by ANECA and ACAP.

Jesús R. Jiménez-Octavio obtained a degree in Industrial Engineering in 2004 from the Pontifical Comillas University and a Ph.D. in Industrial Engineering in 2009 from the same university, with a doctoral thesis titled "Dynamic Analysis and Optimization of Overhead Lines for High-Speed Rail". He currently works as a Professor in the Department of Mechanical Engineering at ICAI, specializing in the area of Continuum Mechanics and conducts research in the fields of Railway Systems and Bioengineering at the Technological Research Institute. He received the Talgo Prize for Technological Innovation in 2010 for his work on "Railway Simulation, Calculation and Optimization".

Denis Cousineau is a Professor at the Université d'Ottawa since 2011. He previously was a Professor at the Université de Montréal and before that, post-doctoral fellow at Indiana University. He holds B.Sc. degree in computer science, in education and in psychology. His areas of research include response time models, attention and visual search. From 2002 to 2007, he organized yearly the Summer School in Advanced Methodological Methods and he founded in 2005 the Tutorials in Quantitative Methods for

Psychology journal. He just published his second textbook (in French), Panorama des statistiques.

Author contributions

Raquel Caro-Carretero: writing—original draft, review & editing, conceptualization, investigation, methodology, software, formal analysis, validation visualization, supervision. Alberto Carnicero: conceptualization, investigation, software, visualization. Jesús Ramón Jiménez-Octavio: conceptualization, investigation. Denis Cousineau: writing—review & editing, conceptualization, methodology, software, visualization.

Acknowledgments

This publication is the result of a collaborative work conducted within the framework of the projects DEWEIN (102021599) and EPOXIGRAPHENIT (133201614N). The authors would like to express their gratitude to the reviewers for their insightful and constructive comments on the paper. Their feedback has been invaluable in improving the quality and rigor of this research.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was partially supported by the project DEWEIN and the project EPOXIGRAPHENIT.

ORCID

R. Caro-Carretero http://orcid.org/0000-0003-2233-7635

References

Aguirre-Urreta, M. I., M. Rönkkö, and C. N. McIntosh. 2019. A cautionary note on the finite sample behavior of maximal reliability. *Psychological Methods* 24 (2):236–52. doi: 10.1037/met0000176.

Akram, M., and A. Hayat. 2014. Comparison of estimators of the Weibull distribution. *Journal of Statistical Theory and Practice* 8 (2):238–59. doi: 10.1080/15598608.2014.847771.

Alía, C., J. M. Arenas, J. C. Suárez, J. J. Narbón, and R. Ocaña. 2013. ENF test in the adhesive bonding of aluminium–composite joints and evaluation of its reliability with Weibull distribution. *Journal of Adhesion Science and Technology* 27 (11):1236–46. doi: 10.1080/01694243.2012.736853.

Alqam, M., R. M. Bennett, and A. H. Zureick. 2002. Three-parameter vs. two-parameter Weibull distribution for pultruded composite material properties. *Composite Structures* 58 (4):497–503. doi: 10.1016/s0263-8223(02)00158-7.

Arenas, J. M., J. J. Narbón, and C. Alía. 2010. Optimum adhesive thickness in structural adhesives joints using statistical techniques based on Weibull distribution. *International Journal of Adhesion and Adhesives* 30 (3):160–5. doi: 10.1016/j.ijadhadh.2009.12.003.

Bartkutė, V., and L. Sakalauskas. 2008. The method of three-parameter Weibull distribution estimation. *Acta et Commentationes Universitatis Tartuensis de Mathematica* 12:65–78. doi: 10.12697/ACUTM.2008.12.06.

Caro-Carretero, R. 2023. https://github.com/raquelcaro1-caro/data-an-epoxy-resin-Resoltech-1050-1056-and-replicate-results (accessed November 22, 2023).

Carpenter, J., and J. Bithell. 2000. Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. *Statistics in Medicine* 19 (9):1141–64. doi: 10.1002/(SICI)1097-0258(20000515)19:9 < 1141::AID-SIM479 > 3.0.CO;2-F.

Cheng, R. C. H., and N. A. K. Amin. 1983. Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society: Series B (Methodological)* 45 (3):394–403. doi: 10.1111/j.2517-6161.1983.tb01268.x.

Chernick, M. R., and R. A. Labudde. 2009. Revisiting qualms about bootstrap confidence intervals. *American Journal of Mathematical and Management Sciences* 29 (3-4):437–56. doi: 10.1080/01966324.2009.10737767.

Chernick, M. R., and R. A. Labudde. 2011. *Introduction to bootstrap with R*. Hoboken, NJ: John Wiley & Sons, Inc.

Cohen, C. A., and B. Whitten. 1982. Modified maximum likelihood and modified moment estimators for the three-parameter Weibull distribution. *Communications in Statistics – Theory and Methods* 11 (23):2631–56. doi: 10.1080/03610928208828412.

Cousineau, D. 2009a. Fitting the three-parameter Weibull distribution: Review and evaluation of existing and new methods. *IEEE Transactions on Dielectrics and Electrical Insulation* 16 (1):281–8. doi: 10.1109/TDEI.2009.4784578.

Cousineau, D. 2009b. Nearly unbiased estimators for the three-parameter Weibull distribution with greater efficiency than the iterative likelihood method. *The British Journal of Mathematical and Statistical Psychology* 62 (Pt 1):167–91. doi: 10.1348/000711007x270843.

Cousineau, D. 2011. The fallacy of large shape parameters when using the two-parameter Weibull distribution. *IEEE Transactions on Dielectrics and Electrical Insulation* 18 (6):2095–102. doi: 10.1109/TDEI.2011.6118649.

D638-14. 2022. Standard test method for tensile properties of plastics (s. f.). https://www.astm.org/d0638-14.html (accessed November 20, 2023).

Davison, A. C., and D. V. Hinkley. 1997. *Bootstrap methods and their application. Cambridge series in statistical and probabilistic mathematics (1). Computational statistics, machine learning and information science, general statistics and probability, statistics and probability*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511802843.

DiCiccio, C. J., J. P. Romano, and M. Wolf. 2019. Improving weighted least squares inference. *Econometrics and Statistics* 10:96–119. doi: 10.1016/j.ecosta.2018.06.005.

Efron, B. 1982. *The jackknife, the bootstrap and other resampling plans. CBMS-NSF regional conference series in* *applied mathematics, monograph 38*. Philadelphia, PA: SIAM. doi: 10.1137/1.9781611970319.

Efron, B., and R. J. Tibshirani. 1993. *An introduction to the bootstrap*. New York, NY: Chapman & Hall. doi: 10.1007/978-1-4899-4541-9.

Garrido, A., R. Caro-Carretero, J. R. Jimenez-Octavio, A. Carnicero, and M. Such. 2019. A new approach to fitting the three-parameter Weibull distribution: An application to glass ceramics. *Communications in Statistics – Theory and Methods* 50 (14):3403–20. doi: 10.1080/03610926.2019.1702698.

Goda, Y., M. Kudaka, and H. Kawai. 2011. Incorporation of Weibull distribution in L-moments method for waves. *Coastal Engineering Proceedings* 1 (32):62. doi: 10.9753/icce.v32.waves.62.

Hadj-Ahmed, R., G. Foret, and A. Ehrlacher. 2001. Probabilistic analysis of failure in adhesive bonded joints. *Mechanics of Materials* 33 (2):77–84. doi: 10.1016/s0167-6636(00)00051-x.

Hoyos-Osorio, K., A. M. Álvarez-Meza, G. Daza-Santacoloma, Á. Á. O. Gutiérrez, and G. Castellanos-Domínguez. 2021. Relevant information undersampling to support imbalanced data classification. *Neurocomputing* 436:136–46. doi: 10.1016/j.neucom.2021.01.033.

ISO 527-1. 2019. (s. f.). ISO. https://www.iso.org/standard/75824.html (accessed November 20, 2023).

ISO 527-2. 2012. (s. f.). ISO. https://www.iso.org/standard/56046.html (accessed November 20, 2023).

Karavarsamis, N., A. P. Robinson, G. Hepworth, A. J. Hamilton, and G. W. Heard. 2013. Comparison of four Bootstrap-Based interval estimators of species occupancy and detection probabilities. *Australian & New Zealand Journal of Statistics* 55 (3):235–52. doi: 10.1111/anzs.12044.

Kyselý, J. 2009. Coverage probability of bootstrap confidence intervals in heavy-tailed frequency models, with application to precipitation data. *Theoretical and Applied Climatology* 101 (3-4):345–61. doi: 10.1007/s00704-009-0190-1.

Mäki-Lohiluoma, E., N. Säkkinen, M. Palomäki, O. Winberg, H. X. Ta, T. Heikkinen, E. Kiljunen, and A. Kauppinen. 2021. Use of machine learning in prediction of granule particle size distribution and tablet tensile strength in commercial pharmaceutical manufacturing. *International Journal of Pharmaceutics* 609:121146. doi: 10.1016/j.ijpharm.2021.121146.

Manly, B. 2007. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Third Edition, Chapman and Hall/CRC, London.

Naito, K., J. M. Yang, Y. Inoue, and H. Fukuda. 2012. The effect of surface modification with carbon nanotubes upon the tensile strength and Weibull modulus of carbon fibers. *Journal of Materials Science* 47 (23):8044–51. doi: 10.1007/s10853-012-6694-6.

Nourbakhsh, M., Y. Mehrali, A. Jamalizadeh, and G. Yari. 2014. On a Selection Weibull Distribution. *Communications in Statistics - Theory and Methods* 44 (8):1640–52. doi: 10.1080/03610926.2013.777738.

Oliveira, M., N. Moniz, L. Torgo, and V. S. Costa. 2021. Biased resampling strategies for imbalanced spatio-temporal forecasting. *International Journal of Data Science and Analytics* 12 (3):205–28. doi: 10.1007/s41060-021-00256-2.

Parambil, N. K., and S. Gururaja. 2015. Damage Development 3D-RUC of Polymer Matrix with Randomly Distributed Fibers. 56th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference. doi: 10.2514/6.2015-0971.

Parambil, N. K., and S. Gururaja. 2017. Micro-scale progressive damage development in polymer composites under longitudinal loading. *Mechanics of Materials* 111:21–34. doi: 10.1016/j.mechmat.2017.05.002.

Puth, M. T., M. Neuhäuser, and G. D. Ruxton. 2015. On the variety of methods for calculating confidence intervals by bootstrapping. *The Journal of Animal Ecology* 84 (4):892–7. doi: 10.1111/1365-2656.12382.

Qin, X., J. S. Zhang, and X. D. Yan. 2012. A Finite Mixture Three-Parameter Weibull Model for the Analysis of Wind Speed Data. *Communications in Statistics - Theory and Methods* 41 (12):2160–71. doi: 10.1080/03610926.2011.558657.

Quenouille, M. 1949. Problems in plane sampling. *The Annals of Mathematical Statistics* 20 (3):355–75. doi: 10.1214/aoms/1177729989.

Quinn, J. B., and G. D. Quinn. 2010. A practical and systematic review of Weibull statistics for reporting strengths of dental materials. *Dental Materials: Official Publication of the Academy of Dental Materials* 26 (2):135–47. doi: 10.1016/j.dental.2009.09.006.

Ruxton, G. D., and M. Neuhäuser. 2013. Improving the reporting of*P*-values generated by randomization methods. *Methods in Ecology and Evolution* 4 (11):1033–6. doi: 10.1111/2041-210X.12102.

Towse, A., K. Potter, M. Wisnom, and R. Adams. 1999. The sensitivity of a Weibull failure criterion to singularity strength and local geometry variations. *International Journal of Adhesion and Adhesives* 19 (1):71–82. doi: 10.1016/s0143-7496(98)00058-x.

Tukey, J. 1958. Bias and confidence in not-quite large samples. *Annals of Mathematical Statistics* 29 (2):614–

van Mier, J. G., M. R. van Vliet, and T. K. Wang. 2002. Fracture mechanisms in particle composites: Statistical aspects in lattice type analysis. *Mechanics of Materials* 34 (11):705–24. doi: 10.1016/s0167-6636(02)00170-9.

Webb, G. I., C. Sammut, C. Perlich, T. L. Horvath, S. Wrobel, K. B. Korb, W. S. Noble, C. Leslie, M. G. Lagoudakis, N. Quadrianto, et al. 2011. Leave-One-Out Cross-Validation. In *Encyclopedia of Machine Learning*, eds. C. Sammut and G. I. Webb, 600–1. Boston, MA: Springer. doi: 10.1007/978-0-387-30164-8_469.

Wong, T. F., R. H. Wong, K. Chau, and C. Tang. 2006. Microcrack statistics, Weibull distribution and micromechanical modeling of compressive failure in rock. *Mechanics of Materials* 38 (7):664–81. doi: 10.1016/j.mechmat.2005.12.002.

Zapata-Ordúz, L., G. Portela, and O. Suárez. 2014. Weibull statistical analysis of splitting tensile strength of concretes containing class F fly ash, micro/nano-SiO2. *Ceramics International* 40 (5):7373–88. doi: 10.1016/j.ceramint.2013.12.083.