



L-moments-based uncertainty quantification for scarce samples including extremes

Deepan Jayaraman¹ · Palaniappan Ramu¹

Received: 28 December 2020 / Revised: 9 March 2021 / Accepted: 9 April 2021 / Published online: 30 June 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Sampling-based uncertainty quantification demands large data. Hence, when the available sample is scarce, it is customary to assume a distribution and estimate its moments from scarce data, to characterize the uncertainties. Nonetheless, inaccurate assumption about the distribution leads to flawed decisions. In addition, extremes, if present in the scarce data, are prone to be classified as outliers and neglected which leads to wrong estimation of the moments. Therefore, it is desirable to develop a method that is (i) distribution independent or allows distribution identification with scarce samples and (ii) accounts for the extremes in data and yet be insensitive or less sensitive to moments estimation. We propose using L-moments to develop a distribution-independent, robust moment estimation approach to characterize the uncertainty and propagate it through the system model. L-moment ratio diagram that uses higher order L-moments is adopted to choose the appropriate distribution, for uncertainty quantification. This allows for better characterization of the output distribution and the probabilistic estimates obtained using L-moments are found to be less sensitive to the extremes in the data, compared to the results obtained from the conventional moments approach. The efficacy of the proposed approach is demonstrated on conventional distributions covering all types of tails and several engineering examples. Engineering examples include a sheet metal manufacturing process, 7 variable speed reducer, and probabilistic fatigue life estimation.

Keywords Uncertainty quantification · Conventional moments · L-moments · Extremes · Scarce data

1 Introduction

In structural design, uncertainties in the system inputs are considered the root cause of poor product performance and this leads to variation in the system output responses. These uncertainties typically occur in material properties, manufacturing parameters, externally applied forces, etc. (Jin et al. 2003; Lee et al. 2009; Ramu and Arul 2016; Voinov et al. 2016; Lee et al. 2019). Irrespective of the variability in the input, a designer wishes to achieve a system response that satisfies the design objectives, preferably

with less variability (Ramu et al. 2017). It is a common statistical practice to quantify uncertainties by conventional moments (C-moments), mean (μ), and variance (σ^2) (Hosking 1990). Higher order C-moments such as skewness (γ_1) and kurtosis (α_4) are also used to describe the system outputs more precisely (Mekid and Vaja 2008; Anderson and Mattson 2012). In order to design under uncertainty, it is important to quantify uncertainties of system output which is the result of uncertain inputs propagated through the system. This is referred to as uncertainty propagation (Mekid and Vaja 2008; Lee and Chen 2009; Anderson and Mattson 2012; Jayaraman et al. 2018; Liu et al. 2019). Sometimes, input distributions are known and moments are computed to be propagated through the system model to obtain statistics of output response. When moments of distributions are used for propagation, it is imperative to compute higher order moments accurately (Lee and Chen 2009).

Most engineering applications have scarce data and description of the uncertainties associated with the input variables is usually not available readily (Ramu and Arul 2016). One way of quantifying this uncertainty is to use nonparametric methods such as histogram, kernel density

Responsible Editor: Makoto Ohsaki

✉ Deepan Jayaraman
deepanaero8@gmail.com

¹ Department of Engineering Design, Advanced Design, Optimization and Probabilistic Techniques (ADOPT) Laboratory, Indian Institute of Technology Madras, Chennai, India

estimation (KDE) techniques, and interval analysis (Lee et al. 2019). Though histogram is a nonparametric density estimation approach, it is less robust and the challenge lies in choosing the bin size. Changing the bin size leads to different inference from the data (Rudemo 1982; Silverman 1986; Lee et al. 2019). In KDE, the smoothing of the density curve is controlled by bandwidth and selecting the right bandwidth is a major challenge (Hall et al. 1991; Kang et al. 2018). Often, higher bandwidth leads to over-smoothing and lower bandwidth leads to under-smoothing of the density curve (Lee et al. 2019). Bandwidth selection for heavy tailed distributions is relatively difficult (Buch-Larsen et al. 2005). Integrated squared error, mean integrated squared error, and least squares cross validation error are used to choose the optimal bandwidth (Silverman 1986; Hall et al. 1991; Shirahata and Is 1992; Turlach 1993). These error metrics are more sensitive to the extremes present in the small data and suffer from sampling variability (Silverman 1986; Park and Marron 1990; Hall et al. 1991). Sometimes KDE can represent a very irregular shape of distribution for small data (Kang et al. 2018). Interval analysis is also used to represent the uncertainty as an interval or variable that has lower and upper bound. Fuzzy set theory and evidence theory are generally used to estimate the interval. Major drawback of this approach is that it requires interval arithmetic for statistical model comparison and validation (Rokne 2001; Gao et al. 2010; Lee et al. 2019). Though nonparametric methods are robust, they require sufficient data for accurate modelling (Lee et al. 2019). When the interest is in extreme probabilities, tail modeling techniques are also used to quantify the uncertainties. These techniques approximate the tail portion of the cumulative density function (CDF) using limited samples. However, the tail itself is highly volatile and very sensitive to number of data points (Ramu P 2013; Acar and Ramu 2014). These techniques are usually suitable for low probabilities and not to model the entire probability space, which is the general interest in uncertainty quantification.

A widely used parametric approach is the Pearson system which is based on the hypothesis that higher order moments provide a good representation of PDF. However, researchers (Hosking 1990) observe that there is lack of clarity on what information does the higher order moments impart, on the shape of the distribution. Also, moments computed from scarce data can be markedly different from those of the probability distribution from which the sample was drawn. The choice of underlying probabilistic distribution plays a very crucial role. When the choice is influenced by scarce samples, it is likely to have errors, which when propagated through the model result in amplified errors in the system output. In addition, the scarce samples might

also include extremes, which will only worsen the errors. In scarce data, extremes have large influence on the moments estimation (Moon et al. 2020). C-moments are sensitive to extremes present in the data leading to large variation in the computed moments. Extremes can sometimes be classified as outliers and excluded from the statistical study (Jayaraman et al. 2018; Jayaraman and Ramu 2019). This again will lead to large error between computed moments and actual moments. Since the available data itself is scarce, excluding one or a few data points leads to significant information loss. Therefore, it is desirable to develop an approach that can:

- (i) Identify appropriate probabilistic distribution with scarce data.
- (ii) Estimate moments accurately from scarce data with possible extremes. This translates to the estimation technique being less or insensitive to the extremes.

To overcome the hurdle of computing robust moments while accounting for extreme realizations, researchers recommend using linear moments (L-moments) which uses linear combinations of order statistics. L-moments are known to be robust and less sensitive to extremes or outliers in the scarce samples. In addition, L-moments are less subject to bias in estimation and their approximations on the asymptotic normal distribution are better, when the samples are finite (Sillitto 1969; David 1981; Hosking 1989, 1990, 1992; Gubareva and Gartsman 2010). L-moments are widely used in various applications in hydrology, water resource applications, and regional frequency analysis (Hosking and Wallis 1997; Sankarasubramanian and Srinivasan 1999; Adamowski 2000; Kumar and Chatterjee 2005; Atiem and Harmancioğlu 2006). L-moment approach is preferable for insufficient or lesser data in flood or rainfall analysis (Smithers and Schulze 2001; Haddad et al. 2011). L-moments are used to identify the probability distribution of the censored data in the fields of environmental quality and quantity monitoring (Zafirakou-Koulouris et al. 1998; Sankarasubramanian and Srinivasan 1999; Elamir and Seheult 2004). In reliability analysis, L-moments are used to analyze and characterize the life time data (Nair and Vinesh Kumar 2010). Based on the discussions presented above, the use of L-moments in the context of design with scarce samples that might include extremes is not available. Hence, the authors propose a framework using L-moments to quantify and propagate uncertainties in an optimization framework, when the available data is scarce and could include potential extremes.

To this end, the purpose of this study is to identify the PDF of response, with scarce samples and in the presence of extremes. We propose using L-moment ratio to identify the PDF. The identified PDF is compared to

the exact PDF using the Jensen-Shannon divergence. This comparison is only to validate the proposed approach, because in reality, one will not have the information about the exact PDF. Higher order L-moments are used to compute the L-moment ratios such as L-skewness (τ_3) and L-kurtosis (τ_4). We compare L-moment estimates against C-moment estimates and their effects on identifying the underlying distribution using L-moment ratio and the Pearson system, respectively. The novelty lies in aspects such as the following: (i) comprehensive understanding of performance of L-moments on different types of tails when the data is scarce and with potential extremes; (ii) combining scarce data, extremes, and L-moments in a design context; (iii) using L-moments to quantify the uncertainty and propagate it in a design framework, while implementing on real-life engineering examples. The proposed approach is demonstrated on a suite of distributions that cover all type of tails and several engineering examples.

The rest of the paper is organized as follows: In Section 2, it is explained how the PDFs are identified using the Pearson system and L-moment ratio diagram. Investigation strategy of the proposed approach and the comparison of obtained PDFs with the exact PDF using the Jensen-Shannon divergence and how extremes are generated and incorporated in the data is discussed in Section 3. Demonstration of the proposed approach on a suite of statistical distributions and engineering examples is carried out in Section 4 followed by conclusions in Section 5.

2 C-moments and L-moments

In this study, we focus on scarce data that might include extremes. The goal is to identify the underlying PDF and use its moments for characterising uncertainties. The two approaches that we compare here (C- and L- moment approaches) compute the moments and identify the respective PDFs using the Pearson system and L-moment ratio diagram, respectively. The specific objective is to compare the performance of the approaches in the presence of extremes. The process of investigation is presented in Fig. 1 and the elements of Fig. 1 are discussed below.

2.1 Identification of distribution using C-moments and Pearson system

The Pearson system (Pearson 1916) is a method of choosing appropriate distribution from a set of distributions called the Pearson distribution, based on first four C-moments of the system response. The Pearson distribution contains several distributions such as beta, normal, and gamma. Each PDF in

the Pearson distribution satisfies the generalized differential equation, in (1) (Kenney and Keeping 1947; Craig 1991; Weisstein and et al. 2004).

$$\frac{dy}{dx} = \frac{y(m-x)}{a_0 + a_1x + a_2x^2} \tag{1}$$

where y is the PDF, a_0, a_1, a_2 , and m are the parameters expressed in terms of C-moments of the system, in (2).

$$\begin{aligned} a_0 &= \frac{2+\delta}{2(1+2\delta)} \\ a_1 &= -m = \frac{\gamma_1}{2(1+2\delta)} \\ a_2 &= \frac{\delta}{2(1+2\delta)} \end{aligned} \tag{2}$$

where $\delta = \frac{2\alpha_4 - 3\gamma_1^2 - 6}{\gamma_1 + 3}$, γ_1 is the skewness, and α_4 is the kurtosis of the data.

The roots c_1, c_2 and the coefficients of $a_0 + a_1x + a_2x^2$ provide information that can be used to classify the distributions. Few possible types of distribution are (i) $a_1 = a_2 = 0, a_0 > 0$: normal distribution, (ii) $a_1^2/4a_0a_2 < 0, c_1 \leq x \leq c_2$: beta distribution (iii) $a_1^2/4a_0a_2 = 0, a_2 > 0, -\infty < x < \infty$: Student's t -distribution, etc. The solution of (1) needs to satisfy the following conditions:

- i. the distribution curve should vanish at the ends of the range, i.e., as $y \rightarrow 0, dy/dx \rightarrow 0$
- ii. when $x = m, dy/dx \rightarrow 0$

In this study, the Pearson system is adopted to identify the distribution using C-moments computed from the scarce samples.

2.2 L-moments

L-moments (Hosking 1989; 1990; 1992; Hosking and Wallis 1997; Hosking 2006) are expectations of certain linear combinations of order statistics. These combinations provide information about location, scale, and shape of the distribution from which the samples are drawn. Since L-moments are linear functions of the data, they suffer less from the effects of sampling variability and are reported to be robust in the presence of extremes. L-moments which are modifications of probability weighted moments (PWM) can be used to define the shape of the probability distribution (Greenwood et al. 1979). PWM (β_r) (Hosking and Wallis 1997) for a probability distribution with cumulative density function, F and quantile function, $x(F)$ is given by $\beta_r = \int_0^1 x(F) F^r dF$. The population L-moment is

$$\lambda_{r+1} = \sum_{k=0}^r p_{r,k}^* \beta_k, r = 0, 1, \dots, n-1; \tag{3}$$

where λ is the population L-moment, $p_{r,k}^*$ is $(-1)^{r-k} \binom{r}{k} \binom{r+k}{k}$, n is the sample size.

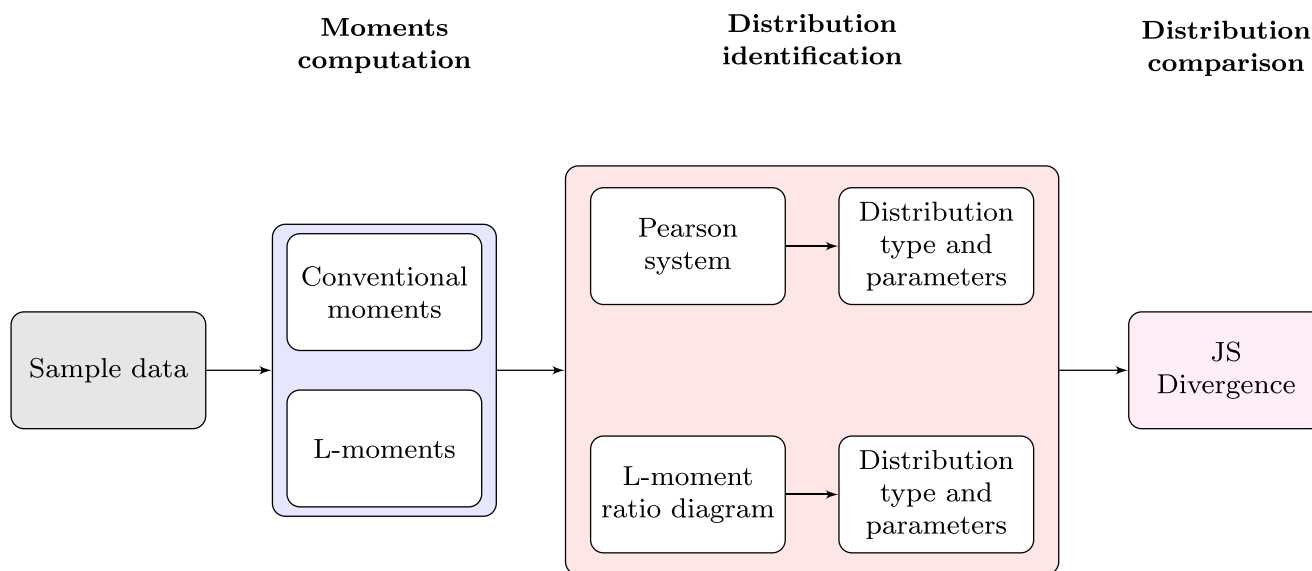


Fig. 1 Generalized distribution identification procedure

Equation (3) can be rewritten in terms of the expectations of order statistics of a random variable, X , as

$$\lambda_r = \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} E(X_{r-k:r}), \tag{4}$$

$r = 1, 2, \dots, n;$

where $E(\cdot)$ is the expectation value. The first four L-moments are given in terms of PWMs derived from (3), and presented in (5)

$$\begin{aligned} \lambda_{1,r=0} &= \beta_0, \\ \lambda_{2,r=1} &= 2\beta_1 - \beta_0, \\ \lambda_{3,r=2} &= 6\beta_2 - 6\beta_1 + \beta_0, \\ \lambda_{4,r=3} &= 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \end{aligned} \tag{5}$$

In practice, L-moments are estimated from the ordered sample, $x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}$. The sample PWM, b_r is presented in (6).

$$b_r = \frac{1}{n} \sum_{j=r+1}^n \frac{(j-1)(j-2)\dots(j-r)}{(n-1)(n-2)\dots(n-r)} x_{j:n}, \tag{6}$$

$r = 0, 1, 2, \dots, n-1;$

The general form of sample L-moment (l) in terms of sample PWM is

$$l_{r+1} = \sum_{k=0}^r p_{r,k}^* b_k, \quad r = 0, 1, 2, \dots, n-1; \tag{7}$$

The first sample L-moment (l_1) is the sample mean, a measure of location. The dispersion measure of the sample is provided by the second L-moment (l_2), a scalar multiple of Gini’s mean difference statistic (Hosking 1990; Ceriani and Verme 2012). Sample L-moment ratios (t_r) are obtained

by dividing the higher order sample L-moments by the dispersion measure (l_2), $t_r = l_r/l_2$. These are dimensionless quantities, independent of the units of measurement data. The L-moment analogue of the coefficient of variation (σ/μ), is L-coefficient of variance, $\tau_{L-CV} = \lambda_2/\lambda_1$ and it varies as $0 \leq \tau_{L-CV} < 1$ (Hosking 1990; Hosking and Wallis 1997).

2.3 Identification of distribution using L-moment ratio

Similar to conventional moments, L-moments are used to summarize the characteristics of the sample data. Many researchers (Hosking 1990; Hosking and Wallis 1997; Zafirakou-Koulouris et al. 1998; Adamowski 2000; Kumar and Chatterjee 2005) have used L-moment ratio diagram to describe the PDF. L-moment ratio diagram shown in Fig. 2, is a plot of L-skewness vs L-kurtosis of standard distributions such as uniform, normal, exponential, log-normal, logistic, gumbel, generalized extreme value (GEV), generalized pareto (GP) and gamma. It is useful to provide simple explicit expressions for τ_4 (L-kurtosis) in terms of τ_3 (L-skewness) for some widely used three-parameter distributions of probability such as lognormal, gamma, GEV and GP (Hosking and Wallis 1997). The L-moment ratio of two-parameter distributions is shown in Fig. 2 as single dots and mentioned in Table 1. Polynomial approximations of the form $\tau_4 = A_0 + A_1\tau_3 + A_2\tau_3^2 + \dots + A_8\tau_3^8$, have also been obtained and the coefficients are provided in Hosking and Wallis (1997). Overall bound is the lower bound on L-moment ratios of the all distributions obtained while satisfying the constraints, $\lambda_2 \geq 0, -1 < \tau_3 < 1$ and $\frac{1}{4}(5\tau_3^2 - 1) \leq \tau_4 < 1$.

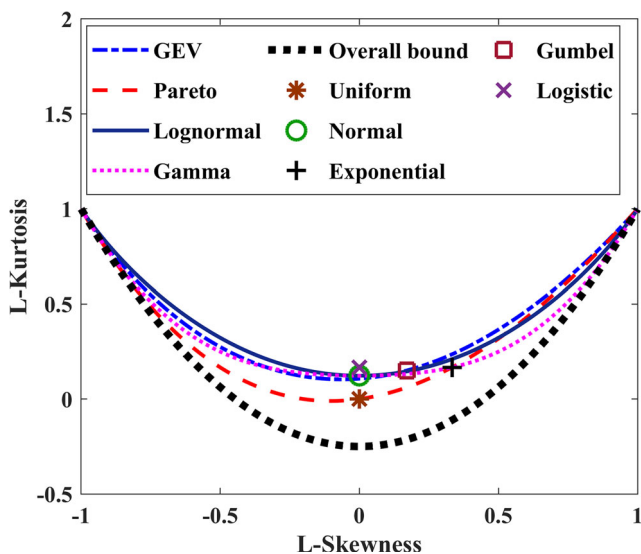


Fig. 2 L-moment ratio diagram (Hosking 1990)

The approximations yield values of τ_4 for a given τ_3 , are accurate within 0.0005 provided that the range of τ_3 is -0.9 to 0.9 . For GEV distribution, an accuracy of 0.0005 is achieved only if the range of τ_3 is -0.6 to 0.9 (Hosking and Wallis 1997).

L-moment ratio diagram permits a visual indication of which distribution maybe a good fit to the data. L-moment ratios t_3 and t_4 , computed from sample can be plotted in the L-moment ratio diagram to see which distribution is the closest. It is to be noted that our interest in the current work lies in computing the probabilities correctly rather than identifying the true parent distribution. Hence, we select distribution that is closest to the sample L-moment ratio plotted in Fig. 2. To understand the performance of the proposed approach, we compare the distribution identified through t_3 and t_4 , to the actual distribution. In order to select a suitable distribution from L-moment ratio diagram, euclidean norm in (8) is used for the distributions that are represented as points.

$$\|Ed_i\|_2 = \sqrt{(\tau_4^i - t_4)^2 + (\tau_3^i - t_3)^2} \tag{8}$$

Table 1 L-moment ratio of distributions

Distribution	L-moment ratio	
	L-skewness (τ_3)	L-kurtosis (τ_4)
Uniform	0	0
Normal	0	0.1226
Exponential	0.3333	0.1667
Gumbel	0.1699	0.1504
Logistic	0	0.1667

where $\|Ed_i\|_2$ is the euclidean norm, i is the i th distribution in the L-moment ratio diagram, $(.)_4$ is the L-kurtosis, $(.)_3$ is the L-skewness, τ_r is the population L-moment ratio, and t_r is the sample L-moment ratio. t_3 and t_4 are computed from available scarce data. For the distributions that are represented as curves in L-ratio diagram, the distance of the (t_3, t_4) from the respective curve is computed as the value of the curve evaluated at (t_3, t_4) . The distribution with the least distance is chosen as the best possible distribution. The algorithm for the best distribution selection is provided in Appendix 1. Once the PDF is chosen from the L-moment ratio diagram, the Jensen-Shannon divergence is used to compare the predicted PDF to the actual PDF.

3 Investigation strategy

The hypothesis in the current work is that L-moments are robust compared to C-moments, from an estimation perspective and L-moments provide a better characterization of the response PDF. We are interested in investigating this for scarce samples (small sample size), especially in the presence of extremes. We thus proceed as follows, for scarce samples, PDFs are obtained using C- and L-moment approaches and compared with the actual population PDF. Extremes are then introduced in the scarce samples and the PDFs are recomputed to be compared with the actual population PDF. Sample sizes considered in this study are 10, 25, 50, and 100. The Jensen-Shannon divergence is used as a comparison metric. The process is repeated 100 times to account for sampling variability.

3.1 Jensen-Shannon divergence

The Jensen-Shannon divergence (D_{JS}) is a widely used distance measure, to study the similarity between finite number of probability distributions (Melville et al. 2005). D_{JS} can be understood as symmetrized extension of Kullback-Leibler divergence (D_{KL}) with finite bounds. Sometimes, D_{JS} is also called as the smooth version of D_{KL} . D_{KL} is also a similarity measure between two probability distributions but unbounded. D_{JS} is presented in (9)

$$D_{JS}(P||Q) = \frac{1}{2}D_{KL}(P||M) + \frac{1}{2}D_{KL}(Q||M) \tag{9}$$

where $M = \frac{1}{2}(P + Q)$,

$D_{KL}(P||M) = \sum_{x \in X} P(x) \log_2 \left(\frac{P(x)}{M(x)} \right)$, P is probability of actual distribution and Q is probability of distribution to be compared with P . D_{JS} is bounded between 0 and 1,

given that D_{JS} uses base 2 logarithm. $D_{JS} = 0$ represents identical distributions.

3.2 Incorporation of extremes

Extreme event or extremes are rare events which have very less probability to occur but cannot be ignored as an outlier (Abarbanel et al. 1992; Moustapha and Sudret 2019). The impact of extremes is crucial in many areas of application, such as environmental disasters (flood, climate change, forest fire, etc.), finance (risk analysis, insurance losses), engineering (structural, traffic analysis), and biomedical (side effects of drugs) (Abarbanel et al. 1992; Alvarado et al. 1998; Davison and Huser 2015). In real-life scenarios or in any engineering application, often only small samples are available to perform the analysis. These samples need not necessarily come as a result of design of experiment (DoE). Sometimes, the small sample might include extremes and are prone to be classified as outliers (erroneous data). But extremes (belonging to the parent distribution) cannot be excluded from the data set. In small sample, omitting data will lead to significant information loss. Hence, data points, which are significantly different from other samples or has less probability, need to be accounted in data analysis. To understand the effect of extremes and the performance of proposed approach, we deliberately introduce extremes in the scarce data, for the examples discussed. 10^6 samples from the parent distribution are generated and the maximum value is considered as extreme. This extreme is then incorporated into the scarce samples.

4 Results and discussion

In this section, we demonstrate the proposed approach on (i) a suite of distributions that covers all type of tails, (ii) uncertainty quantification of sheet metal manufacturing process, (iii) design of speed reducer, and (iv) probabilistic fatigue life assessment. The study on the performance of proposed approach on different types of tail provides a comprehensive perspective on the superiority of the approach on data from any type of distribution. Section 4.2.1 discusses the proposed approach on a single variable problem where uncertainty quantification and propagation are also discussed. Section 4.2.2 discusses the proposed approach on large dimension example. Section 4.2.3 contains real-life data on which the proposed approach is demonstrated, to arrive at critical decisions.

4.1 Statistical distributions

The tails of any probabilistic distribution can be approximated by a generalized pareto distribution (GPD). ψ , the

scale and ξ , the shape parameters characterize the GPD (Ramu et al. 2010). ξ plays a significant role in quantifying the weight of the tail or tail heaviness. Probability distributions can be classified as light ($\xi < 0$), medium ($\xi = 0$), and heavy tail ($\xi > 0$) distributions based on the tail heaviness. Detailed description of ξ and the relation with tail heaviness of the probability distributions is discussed in the Appendix. Since it is impractical to test the approach on all possible distributions, we classify the distributions based on tail heaviness and demonstrate the superiority of the method on all possible types of tails. The distributions, their shape (ξ), and size parameters are mentioned in Table 2. The PDF and CDF of the distributions listed in Table 2 are presented in Fig. 3a and b respectively. Here, we study the effect of extremes in scarce data on identifying the PDFs using samples of sizes 10, 25, 50, and 100.

Samples for all the distributions listed in Table 2 are generated and the corresponding C- and L-moments are computed. With the help of the Pearson system and L-moment ratio diagram, PDFs are predicted. The predicted PDFs are compared to the parent PDF using D_{JS} . In the sample data, an extreme is inserted to study its effect and the PDFs are repredicted with the extreme and compared. This process is repeated 100 times. Figure 4 presents the performance of the proposed approach compared to C-moment approach for light tail. The figure is separated into four columns, representing different sample sizes. In each column, first two boxes are the results for sample with extremes (ex: sample size 11) and the rest of the two boxes are the results of sample with extremes removed (ex: sample size 10). From Fig. 4, it can be observed that the L-moment approach while considering the extreme has minimum D_{JS} compared to its C-moment counterpart. This implies that L-moment approach is robust to extremes. When the sample size increased, the variability in D_{JS} decreased as expected. However, with samples as scarce as 10 or 25, L-moments certainly works far better while including extremes. The values corresponding to median and quartiles are listed in the Appendix, for the interested reader. For a sample size of 10, the median values of D_{JS} based on L-moment approach is lesser than the C-moment approach. This clearly shows that PDF identification using L-moment approach works better. The interquartile range is lesser for L-moment approach, implying they perform better under the sampling variability. For larger sample size, the interquartile range is less for L-moment approach compared to C-moment approach. This indicates that L-moment approach works well when sample size are increased. The results for medium tail are presented in Fig. 5. The performance of L-moment approach is similar to that of in the light tail distribution. However, the effect of extreme in C-moment is large, even in a sample of size 100. The results of heavy tail I and heavy tail II are presented in Figs. 6

Table 2 Details of distributions

Distribution	Type of tail	Shape parameter (ξ)	Distribution parameters	
			Mean	Standard deviation
Normal	Light ($\xi < 0$)	-0.14	2	2
Exponential	Medium ($\xi = 0$)	0	2	-
Lognormal	Heavy I ($\xi > 0$)	0.11	2	$\sqrt{2}$
Lognormal	Heavy II ($\xi > 0$)	0.65	0.25	$\sqrt{2}$

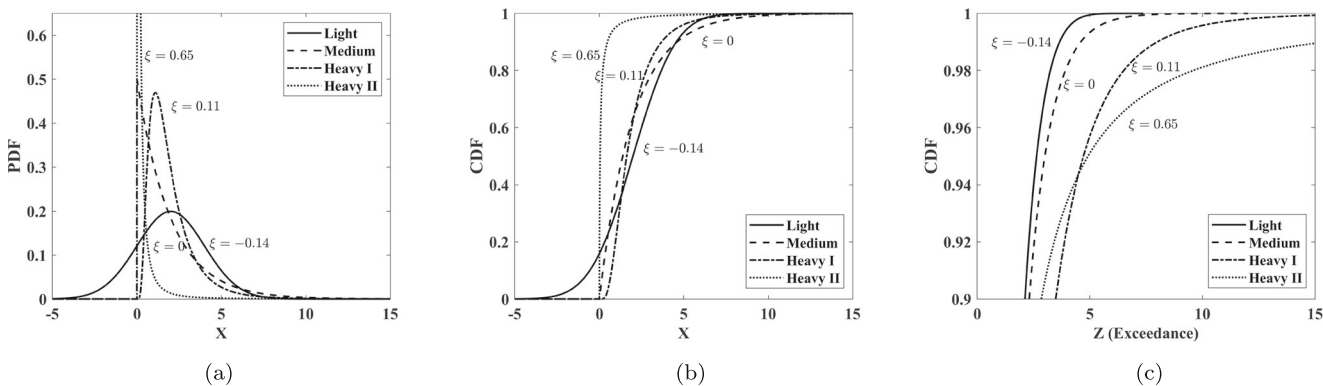


Fig. 3 PDF, CDF, and GPD of distributions. (a) PDF of distributions (b) CDF of distributions (c) GP t for selected distributions

Fig. 4 JS divergence for light tail

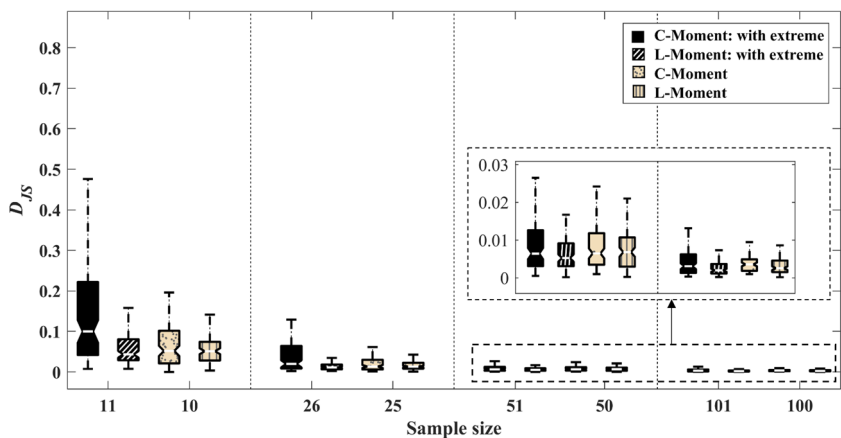


Fig. 5 JS divergence for medium tail

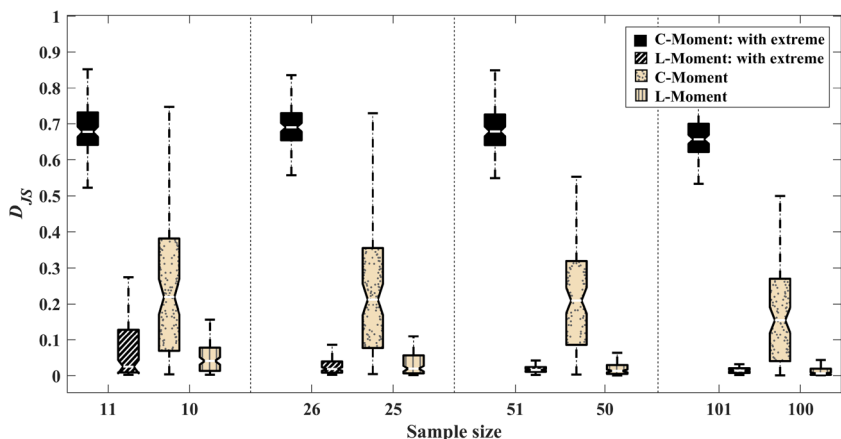
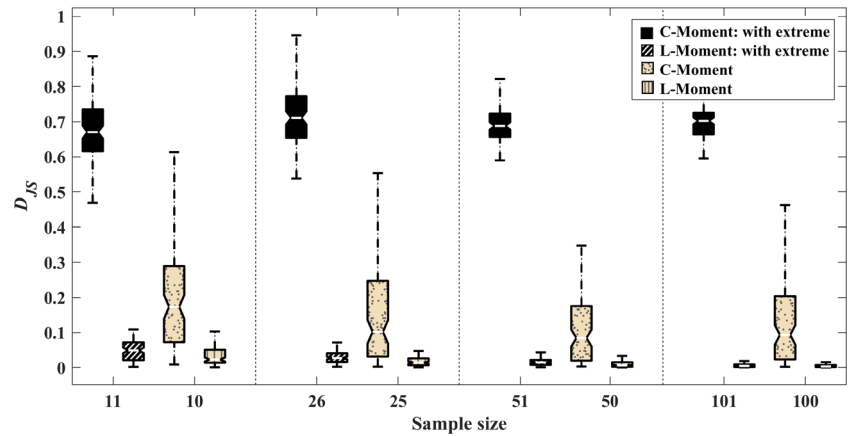


Fig. 6 JS divergence for heavy tail I



and 7 respectively. In heavy tail II case, there is large difference in the performance of C-moment and L-moment approach, even for samples as large as 100. The L-moment approach performance is similar to the other distribution types establishing the fact that L-moments perform better in terms of identifying the distribution in scarce data that includes extreme. Though L-moment approach works better compared to its C-counterpart, the difference is marginal between L-moment with and without the extremes. For heavy tail II, it can be seen that L-moment with extremes performs the best, in terms of variability. Hence using L-moment instead of C-moment approach is necessarily a better choice but within L-moments, whether to leave out the extreme or not might depend on tail heaviness. Nevertheless, it is important to note that one does not possess the information on tail heaviness with scarce data and one does not lose much by considering the extremes while using the L-moments. Information about quartile values for all distributions are presented in Appendix. Effects of two extremes (instead of 1 discussed above) in the scarce samples are also studied and the corresponding PDF comparison results are presented in the Appendix. The results indicate that PDF identified using L-moment ratio

diagram has less D_{JS} , strengthening our earlier conclusion on the performance of L-moment approach.

4.2 Engineering examples

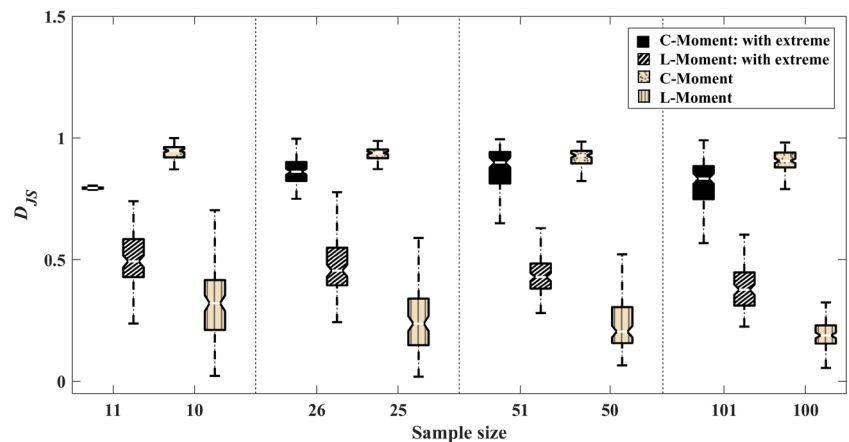
In this section, we discuss three engineering examples to demonstrate the proposed approach.

4.2.1 A flat rolling process: sheet metal manufacturing

Flat rolling process is used to manufacture metal sheets or plates. In this process, the metal is fed between the working rolls or rollers and the distance between the rolls is used to obtain the desired thickness of the metal sheet. The material is pulled by the rollers and eventually results in elongation with reduced thickness. The uncertainty in friction between rollers and the material affects the outcome of the manufacturing process. The process illustrated in Fig. 8 can be performed either below (cold rolling) or above (hot rolling) the material temperature.

The aim of this process is to minimize the number of passes required to achieve final plate thickness (Anderson and Mattson 2012). The total deformation possible in a

Fig. 7 JS divergence for heavy tail II



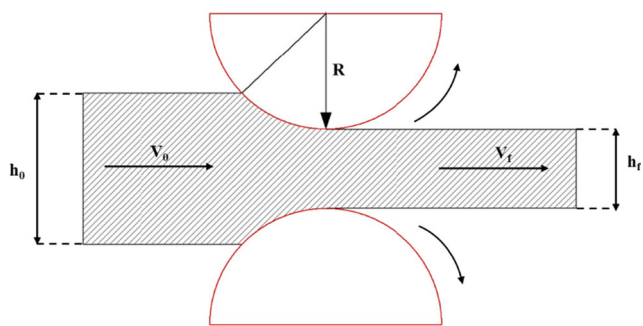


Fig. 8 Flat rolling metal sheet manufacturing process

single pass depends on the friction between the rollers and the material at the interface. Factors such as rolling speed (V), material properties, surface finish of the rollers cause friction, and lubricants are used to control friction during the process. Maximum change in thickness that can be achieved in a single pass, ΔH is presented in (10)

$$\Delta H = \mu_f^2 R \tag{10}$$

where μ_f is the friction coefficient between the working rollers and the sheet metal, and R is the radius of rollers. In this example, we treat μ_f as random variable that follows a generalized pareto distribution (GPD). The GPD of μ_f is depicted in Fig. 9 and the parameters are location, $\mu = 0.282$, scale, $\sigma = 0.103$, and shape, $\xi = -0.325$. We consider $R = 40.6$ cm.

We intend to study the effect of extremes in uncertainty quantification and propagation in this example. For uncertainty quantification, we pretend that the distribution of μ_f is not known and only scarce samples are available. Using scarce samples of μ_f , we compute ΔH and its distribution is predicted using the C- and L-moment approach. This process is repeated 100 times to account for sampling variability. One out of the 100 repetition is presented in Fig. 10 for illustration purpose. The PDFs obtained for various sample sizes in the presence of extremes based on L-moment ratio diagram and Pearson system approaches are presented in Fig. 10. The plots clearly indicate the superiority of the L-moment approach

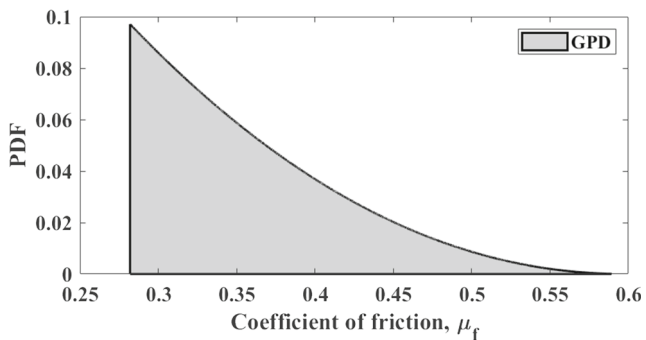


Fig. 9 Distribution of coefficient of friction

and in obtaining approximations closest to the exact PDF. D_{JS} results are presented in Fig. 11. As observed in earlier examples, L-moment approach is superior in identifying the distributions compared to C-moment approach. For uncertainty propagation, we predict the distribution of μ_f using limited samples and draw 10^6 samples from the predicted distribution, propagate it through (10), and compute the corresponding ΔH . In this example, we know the explicit expression of the system to propagate the uncertainties. When the expression is not known, implicit metamodells or response surfaces can be constructed and uncertainties can be propagated through them.

Finally, the PDF of ΔH is obtained from this 10^6 samples and compared against the original PDF. D_{JS} of L-moment approach shows a trend similar to previous examples and is presented in Fig. 12. Similarly, effects of two extremes present in the scarce data are studied. The results presented in the Appendix indicate the superiority of the L-moment approach.

4.2.2 Design of speed reducer with 7 variables

Speed reducers are used in industrial and domestic machinery to safely and efficiently reduce speed of the electric motors. It is also an integral part of gearbox of a mechanical system. Speed reducer adjusts the motor speed to ensure that a machine operates correctly (Lin et al. 2013; Shi et al. 2013; Chatterjee et al. 2017). Figure 13 presents the variables of the speed reducer which are face width (x_1), module of teeth (x_2), number of teeth on pinion (x_3), length of first shaft between bearings (x_4), length of second shaft between bearings (x_5), diameter of first shaft (x_6), and diameter of second shaft (x_7). The aim is to minimize the total weight of the speed reducer which is computed as presented in (11). In the speed reducer design, all the design variables are considered to follow a normal distribution. The description of all variables are mentioned in Table 3 and coefficient of variation (CoV) is 10%.

$$\begin{aligned} W = & 0.7854x_1x_2^2 \times (3.3333x_3^2 \\ & + 14.9334x_3 - 43.0934) \\ & - 1.508x_1(x_6^2 + x_7^2) + 7.4777(x_6^3 + x_7^3) \\ & + 0.7854(x_4x_6^2 + x_5x_7^2) \end{aligned} \tag{11}$$

In this example, different number of extremes are incorporated at a time. The variables in which the extreme is generated is randomly picked. That is, when only 1 extreme is added, it is randomly added to one of the 7 variables. Similarly, extremes are added to 3 variables at a time, 5 variables at a time, and all the random variables at a time. Figure 14 presents the results of distribution obtained using C- and L-moments for extremes added to one random variable. It indicates that results obtained using L-moments are

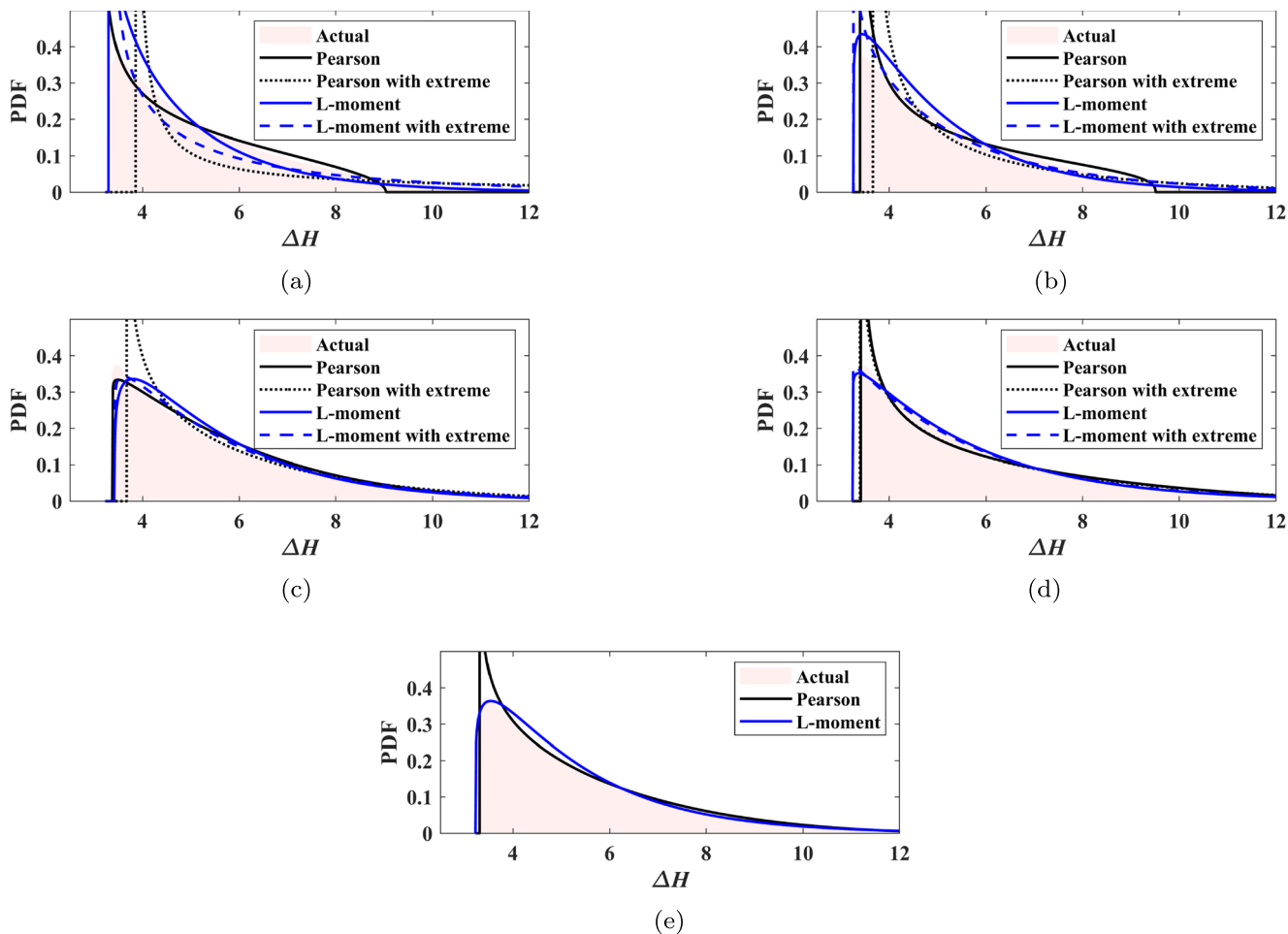


Fig. 10 PDFs of maximum change in thickness, ΔH . (a) 10 samples (b) 25 samples (c) 50 samples (d) 100 samples (e) 10,000 samples

better compared to C-moments for all sample sizes. Within L-moments, the median and variability of D_{JS} are less when the extremes are considered. For the other cases presented in Figs. 15, 16, and 17, similar observations are made. In all cases, L-moment approach is robust to the presence of outliers. The median and variability of D_{JS} in L-moment

approach is less than the median and variability of D_{JS} in C-moment counterpart, for all the cases. Within, the L-moment approach, sometimes the sample with extremes perform better than the sample without extremes. In particular, the difference between the L- and C-moment approaches are large when extremes are considered and included in more

Fig. 11 JS divergence - uncertainty quantification of ΔH of flat rolling sheet metal manufacturing process

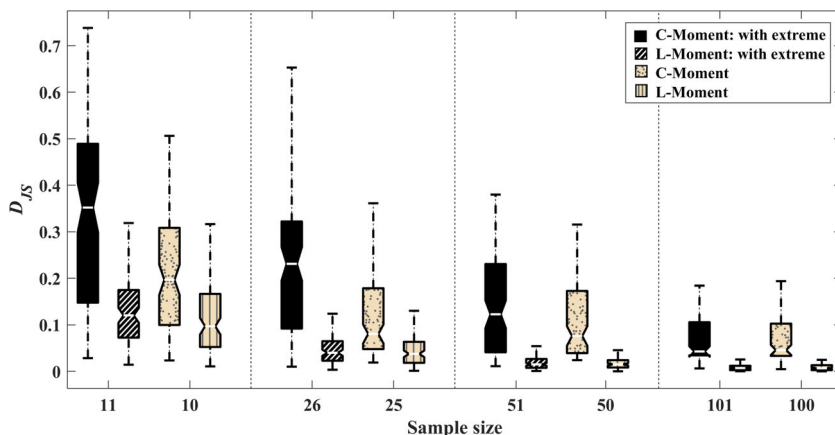
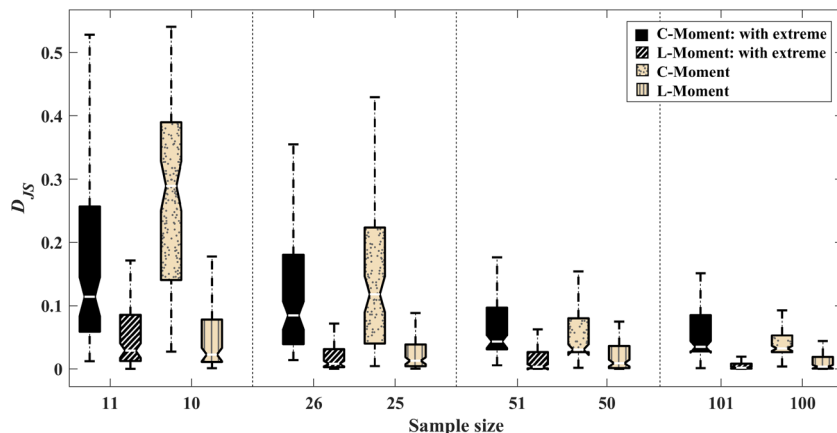


Fig. 12 JS divergence - uncertainty propagation of ΔH of flat rolling sheet metal manufacturing process



variables. The results of two extremes in the scarce data are presented in the Appendix and it is clear that L-moment approach works better than C-moment for all the cases.

In the examples discussed above, we have the prior knowledge of the exact PDF to evaluate the performance of the two approaches. However, in reality, one cannot estimate the D_{JS} . Nevertheless, it is established that the L-moment approach does better or as good as the C-moment approach because of the linear combination of order statistics as against the moment computation in classical approach. The variability in the estimate could be computed using approaches such as bootstrap.

4.2.3 Probabilistic fatigue life assessment

The stress-life (S-N) curve is used to describe fatigue life and it is experimentally measured but usually dispersed in the life axis. Hence, it is preferable to probabilistically measure the fatigue life. Often probabilistic stress-life (P-S-N) curve is used to characterize the uncertainty in the fatigue life. P-S-N curve can be interpreted as a family of S-N curves in which each curve reflect a different % of survival of specimens. Researchers usually model the dispersion by

a log-normal distribution (Hu et al. 2014; Ramu and Arul 2016).

In the current work, P-S-N for ultra-high cycle fatigue (UHCF) vibration tests are derived from scarce samples (fewer experiment data) of screws of strength class 8.8. The experimental data for different load cases 2.5 kN, 4 kN, and 7 kN are provided in Schäfer (2008) and the experimental data size of the three load cases are 101, 500, and 103 respectively. In Schäfer (2008), experimental data is assumed to follow normal distribution but as per the Bayesian inference criterion (BIC) and Akaike information criterion (AIC), the experimental data does follow a lognormal distribution. Hence, assuming a distribution a priori might lead to erroneous results. Here, we pretend that only scarce samples are available and are interested in finding the different percentiles.

Sample sizes used in this study are 10, 15, 20, and 25. Samples are randomly drawn from the experimental data and the detailed sample selection process is mentioned in the Appendix. The maximum value in the experimental data is considered as extreme for each load case. Probabilistic distributions are identified from C- and L-moment approaches and the survival percentiles of 90 and 95 of sample S-N are computed. Survival % obtained from scarce

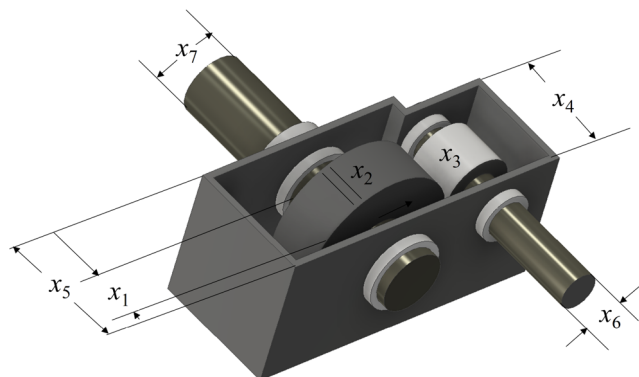


Fig. 13 Speed reducer design

Table 3 Description of random variables

Design variables	Distribution parameter Mean (μ)	Lower bound	Upper bound
x_1 (cm)	3.1	2.6	3.6
x_2 (cm)	0.75	0.7	0.8
x_3	23	17	28
x_4 (cm)	7.8	7.3	8.3
x_5 (cm)	7.8	7.3	8.3
x_6 (cm)	3.4	2.9	3.9
x_7 (cm)	5.25	5	5.5

Fig. 14 JS divergence - extreme at 1 variable at a time

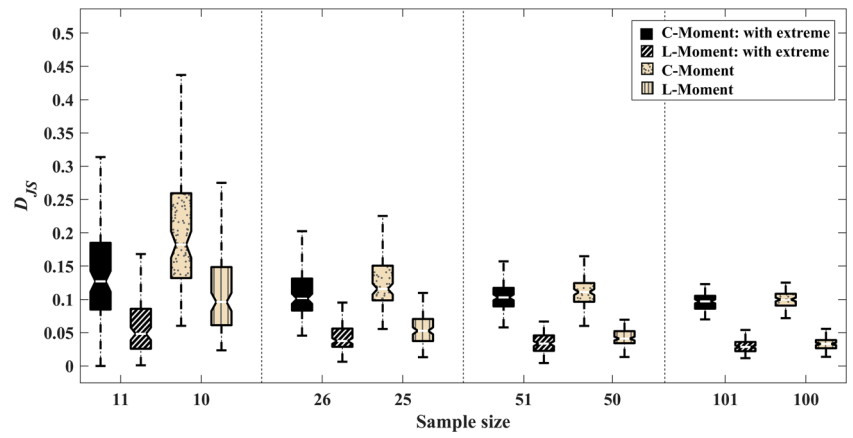


Fig. 15 JS divergence - extreme at 3 variables at a time

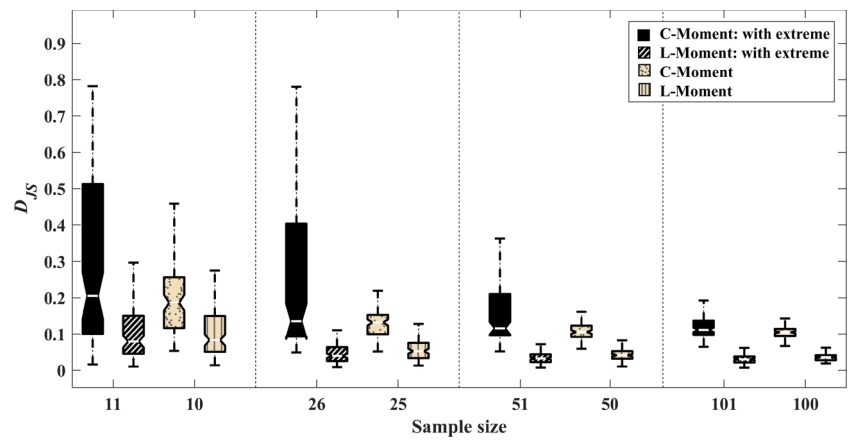


Fig. 16 JS divergence - extreme at 5 variables at a time

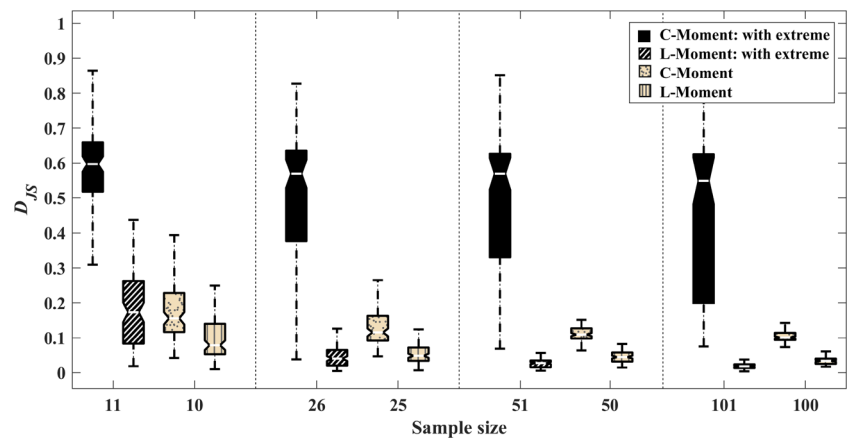
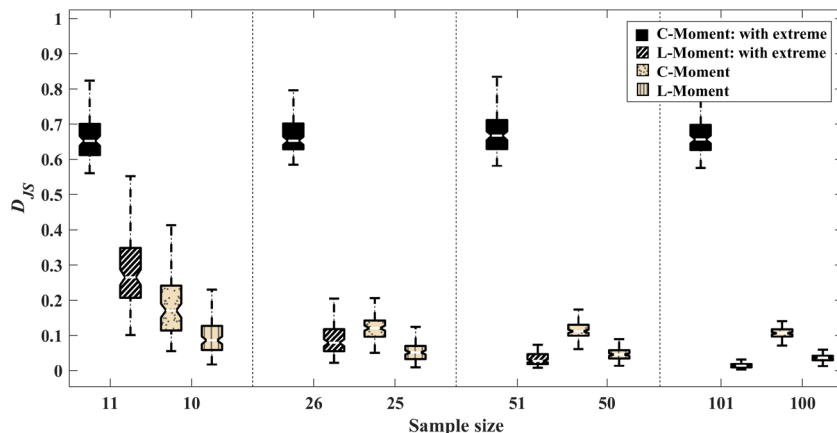


Fig. 17 JS divergence - extreme at 7 variables at a time



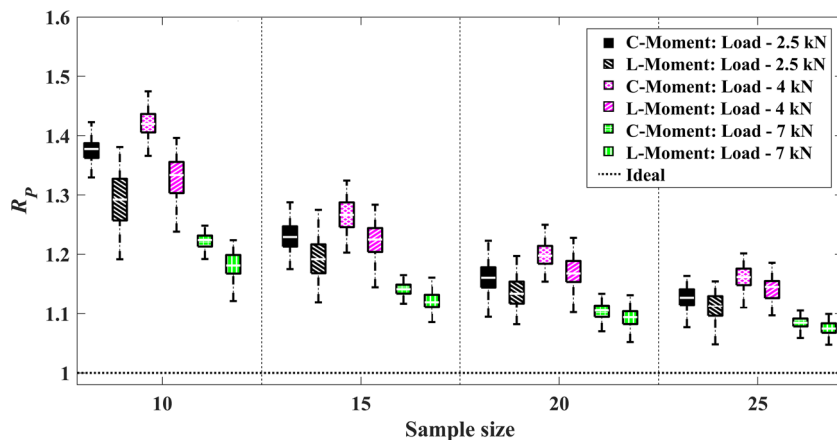
samples with extremes are compared with the ones obtained from the population. Error metric used in this study is ratio of sample percentile (P_S) to the population percentile (P_P), which is $R_P = P_S/P_P$. This process is repeated for 100 times to account for sampling variability.

Figure 18 presents the ratio of the 90th percentile obtained from the sample distribution to the parent distribution for different sample sizes. L-moment approach works well when the extremes are present in the scarce samples and gives the number of cycles value closer to the population for all the percentiles. While designing with fatigue life as a design criteria, one needs to account for the variability in the life estimates and usually require percentiles of life that depends on how critical is the application. Similar observations are obtained in 95th percentile case and the results are presented in the Appendix. Median and the confidence bounds of the all cases are provided in Appendix (25th and 75th percentile of % survival) for 90th and 95th percentiles. Predicted fatigue life from the L-moment approach is closer to the population estimate compared to the C-moment approach.

5 Conclusion

L-moments based approach is proposed to identify PDF and estimate probabilistic quantities, for uncertainty quantification, with scarce data and in the presence of extremes. It is found through numerical examples that L-moments are very robust to extremes, compared to C-moments, in estimating probabilistic quantities. L-moment ratio diagram is adopted to identify closest possible distribution. The identified distributions are compared to the original distribution using JS divergence. The proposed approach is demonstrated on a suite of statistical distribution covering all type of tails and three engineering examples. In all examples, L-moment approach performs better than the C-moment approach. Though the variability of the estimates cannot be obtained in real-life problems, approaches such as bootstrap can be used to obtain confidence bounds, which is scope of future work. The scope of application of the proposed approach was discussed on several engineering examples covering quantile estimation and large variable uncertainty propagation.

Fig. 18 Comparison of 90th survival % of sample distribution to parent distribution with different sample sizes



Appendix 1: Distribution selection based on L-moment

Based on generalized distribution identification procedure mentioned in Fig. 1, the step-by-step distribution selection process based on L-moment ratio diagram is provided in Algorithm 1.

Algorithm 1 Illustration of distribution selection process using L-moment ratio diagram.

1. Input Data: Generate ‘N’ samples in the input variable ‘x’ and compute the corresponding response ‘y’.
2. Compute the L-moment ratios L- skewness (t_3) and L-kurtosis (t_4) using Eqs. $t_3 = \frac{\frac{1}{3}E(X_{3:3}-2X_{2:3}+X_{1:3})}{\frac{1}{2}E(X_{2:2}-X_{1:2})}$ and $t_4 = \frac{\frac{1}{4}E(X_{4:4}-3X_{3:4}+3X_{2:4}-X_{1:4})}{\frac{1}{2}E(X_{2:2}-X_{1:2})}$.
3. τ_3 and τ_4 are the population L-skewness and L-kurtosis respectively. Polynomial approximations of the form $\tau_4 = A_0 + A_1\tau_3 + A_2\tau_3^2 + \dots + A_8\tau_3^8$ are used to represent the distributions that are curves. The coefficients provided in Hosking and Wallis (1997) are used to express the curves in the form $\tau_4 - f(\tau_3) = 0$.
4. The distance between the sample L-moment ratios (t_3, t_4) and the distributions that are represented as curves are computed as the absolute value of the curves evaluated at (t_3, t_4).
5. For the distributions which are points in the L-moment ratio diagram, the distance between them and the sample L-moment ratios (t_3, t_4) are computed using the euclidean norm given in (8).
6. The distribution with the least distance is chosen as the best possible distribution.

Appendix 2: Statistical distributions

2.1 Shape parameter and tail heaviness

Generalized pareto distribution (GPD) is used to approximate the tails of any parent distribution. ψ , the scale, and ξ , the shape, parameters characterize the GPD fit (Ramu et al. 2010). ξ plays a significant role in quantifying the weight of the tail or tail heaviness. Probability distributions can be classified as light ($\xi < 0$), medium ($\xi = 0$), and heavy tail ($\xi > 0$) distributions based on the tail heaviness. Let G be the performance measure which is random and u is threshold of G . The observations of G that surpass u are called, exceedance, Z in the following (12).

$$Z = G - u \tag{12}$$

The conditional CDF of exceedance is modeled by GPD and denoted by $\hat{F}_{\xi, \psi}(Z)$. The distribution function of ($G - u$) is mentioned in (13).

$$\hat{F}_{\xi, \psi}(Z) = \begin{cases} 1 - \left(1 + \frac{\xi}{\psi} Z\right)^{-\frac{1}{\xi}} & \text{if } \xi \neq 0 \\ 1 - \exp\left(-\frac{Z}{\psi}\right) & \text{if } \xi = 0 \end{cases} \tag{13}$$

ξ for exponential distribution is approximately 0. Similarly, ξ is lesser than 0 for light tail and greater than 0 for heavy tail. Figure 19 shows different CDF of tails approximated by GPD of the selected distributions.

2.2 Results—One extreme

In Tables 4, 5, 6, and 7, the quartile (Q) values of the D_{JS} are presented which are obtained for different sample sizes based on the Pearson system and L-moment ratio diagram. In all cases, it can be observed that L-moment approach works better.

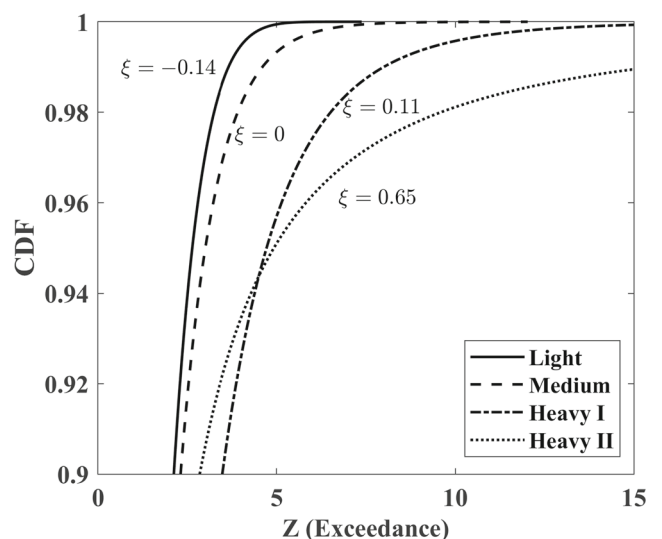


Fig. 19 CDF and shape parameter values of selected distribution

Table 4 Quartile values of JS divergence - light tail

Samples	Distribution identification	Sample size	Quartile values of JS divergence											
			Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
		11												
		26												
		51												
		101												
With extreme	Pearson system		0.042	0.100	0.222	0.009	0.020	0.064	0.003	0.006	0.013	0.001	0.003	0.006
	L-moment ratio diagram		0.029	0.043	0.081	0.007	0.011	0.018	0.003	0.005	0.009	0.001	0.002	0.004
Samples	Distribution identification	Sample size												
		10												
		25												
		50												
		100												
Without extreme	Pearson system		0.021	0.052	0.102	0.006	0.013	0.030	0.003	0.006	0.012	0.002	0.004	0.005
	L-moment ratio diagram		0.028	0.051	0.074	0.008	0.013	0.023	0.003	0.007	0.011	0.002	0.003	0.005

Table 5 Quartile values of JS divergence - medium tail

Samples	Distribution identification	Sample size	Quartile values of JS divergence											
			Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
		11												
		26												
		51												
		101												
With extreme	Pearson system		0.642	0.678	0.733	0.655	0.690	0.730	0.641	0.679	0.726	0.622	0.657	0.700
	L-moment ratio diagram		0.009	0.026	0.129	0.009	0.019	0.040	0.011	0.016	0.025	0.008	0.014	0.021
Samples	Distribution identification	Sample size												
		10												
		25												
		50												
		100												
Without extreme	Pearson system		0.070	0.226	0.380	0.077	0.212	0.355	0.086	0.209	0.319	0.041	0.155	0.270
	L-moment ratio diagram		0.013	0.042	0.078	0.007	0.020	0.057	0.005	0.013	0.030	0.003	0.006	0.019

Table 6 Quartile values of JS divergence - heavy tail I

Samples	Distribution identification	Sample size								
		11	26	51	101					
		Quartile values of JS divergence								
		Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
With extreme	Pearson system	0.616	0.670	0.735	0.654	0.711	0.773	0.656	0.688	0.725
	L-moment ratio diagram	0.021	0.049	0.072	0.016	0.027	0.041	0.008	0.015	0.009
Samples	Distribution identification	Sample size								
		10	25	50	100					
		Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
Without extreme	Pearson system	0.072	0.173	0.290	0.032	0.102	0.247	0.020	0.084	0.204
	L-moment ratio diagram	0.014	0.023	0.051	0.006	0.014	0.027	0.002	0.006	0.007

Table 7 Quartile values of JS divergence - heavy tail II

Samples	Distribution identification	Sample size								
		11	26	51	101					
		Quartile values of JS divergence								
		Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
With extreme	Pearson system	0.791	0.795	0.797	0.823	0.861	0.902	0.813	0.900	0.885
	L-moment ratio diagram	0.429	0.493	0.584	0.395	0.455	0.549	0.381	0.429	0.448
Samples	Distribution identification	Sample size								
		10	25	50	100					
		Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
Without extreme	Pearson system	0.920	0.947	0.962	0.917	0.939	0.953	0.896	0.928	0.940
	L-moment ratio diagram	0.211	0.321	0.417	0.148	0.237	0.340	0.157	0.205	0.230

Table 8 Quartile values of JS divergence - light tail

Samples	Distribution identification	Sample size								
		12	27	52	102					
		Quartile values of JS divergence								
		Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
With extreme	Pearson system	0.036	0.077	0.172	0.008	0.013	0.033	0.006	0.009	0.014
	L-moment ratio diagram	0.021	0.033	0.066	0.006	0.011	0.020	0.004	0.006	0.009
Samples	Distribution identification	Sample size								
		10	25	50	100					
		Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
Without extreme	Pearson system	0.016	0.058	0.124	0.008	0.016	0.028	0.006	0.008	0.013
	L-moment ratio diagram	0.023	0.040	0.090	0.007	0.014	0.024	0.005	0.008	0.012

Table 9 Quartile values of JS divergence - medium tail

Samples	Distribution identification	Sample size								
		12	27	52	102					
		Quartile values of JS divergence								
		Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
With extreme	Pearson system	0.650	0.711	0.768	0.670	0.713	0.754	0.648	0.690	0.733
	L-moment ratio diagram	0.013	0.036	0.120	0.007	0.014	0.037	0.007	0.012	0.020
Samples	Distribution identification	Sample size								
		10	25	50	100					
		Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
Without extreme	Pearson system	0.050	0.197	0.361	0.068	0.173	0.317	0.035	0.193	0.269
	L-moment ratio diagram	0.016	0.054	0.117	0.004	0.017	0.047	0.002	0.008	0.021

Table 10 Quartile values of JS divergence - heavy tail I

Samples	Distribution identification	Sample size											
		27			52			102					
		Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
With extreme	Pearson system	0.706	0.755	0.793	0.729	0.769	0.799	0.723	0.750	0.784	0.724	0.741	0.761
	L-moment ratio diagram	0.008	0.015	0.042	0.007	0.015	0.030	0.007	0.014	0.028	0.002	0.005	0.009
Samples	Distribution identification	Sample size											
		25			50			100					
Without extreme	Pearson system	0.047	0.116	0.246	0.036	0.109	0.278	0.019	0.068	0.163	0.037	0.079	0.168
	L-moment ratio diagram	0.013	0.021	0.045	0.007	0.014	0.024	0.004	0.009	0.017	0.001	0.003	0.006

Table 11 Quartile values of JS divergence - heavy tail II

Samples	Distribution identification	Sample size											
		27			52			102					
		Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
With extreme	Pearson system	0.839	0.849	0.859	0.935	0.960	0.979	0.950	0.974	0.987	0.974	0.982	0.989
	L-moment ratio diagram	0.549	0.608	0.670	0.513	0.556	0.612	0.480	0.521	0.570	0.462	0.503	0.556
Samples	Distribution identification	Sample size											
		25			50			100					
Without extreme	Pearson system	0.961	0.973	0.983	0.968	0.975	0.981	0.759	0.851	0.914	0.826	0.886	0.935
	L-moment ratio diagram	0.235	0.347	0.493	0.159	0.255	0.348	0.158	0.214	0.304	0.172	0.238	0.311

2.3 Results—Two extremes

Two extremes are incorporated in the samples in this study. 10^6 samples from the parent distribution are generated and the maximum value and the value at 9.9×10^5 is considered as extremes in the ordered data. In Tables 8, 9, 10, and 11, and Figs. 20, 21, 22, and 23, the results are presented for two extremes case. For all the sample sizes, L-moment approach work better than the C-moment approach.

Appendix 3: Flat rolling metal sheet manufacturing process

3.1 Distribution selection process based on L-moment

As mentioned in Algorithm 1, the step by step distribution selection process is demonstrated for this example. In this example, coefficient of friction (μ_f) is considered as random variable and follows the distribution mentioned in Fig. 9. Sample size considered to demonstrate the distribution selection process is 100.

1. Samples are generated and the response, maximum change in thickness (ΔH) is computed using (10).
2. L-moment ratios L-skewness (t_3) and L-kurtosis (t_4) are estimated and the values are 0.3, 0.19. Estimated sample L-moment ratios are plotted in L-moment ratio diagram and shown in Fig. 24.
3. Shortest distance is computed for all possible distributions in the L-moment ratio diagram and presented in Table 12.
4. Generalized pareto distribution is considered as the best possible distribution which has less distance compared to all other distributions.
5. Parameters of generalized pareto distribution obtained from the sample and corresponding PDF is plotted with the actual PDF (sample size for population is 10000), presented in Fig. 25.
6. Distributions obtained from sample is compared with the population distribution using D_{JS} . The D_{JS} value is 0.002.

In Tables 13 and 14, the quartile (Q) values of the D_{JS} are presented, which are obtained for different sample sizes. L-moment approach works well in all cases compared to the C-moment approach.

3.2 Results—Two extremes

In this section, results of the two extremes in the scarce data are presented. The D_{JS} obtained for various sample size are presented in Figs. 26 and 27 and the corresponding Q values

Table 12 Shortest distance of all possible distributions in L-moment ratio diagram

Distribution	Uniform	Normal	Exponential	Gumbel	Logistic	GEV*	GP**	Lognormal	Gamma
$\ E d_i\ _2$	0.28	0.26	0.10	0.09	0.26	0.08	0.01	0.07	0.04

*Generalized extreme value, **Generalized pareto

Fig. 20 JS divergence for light tail

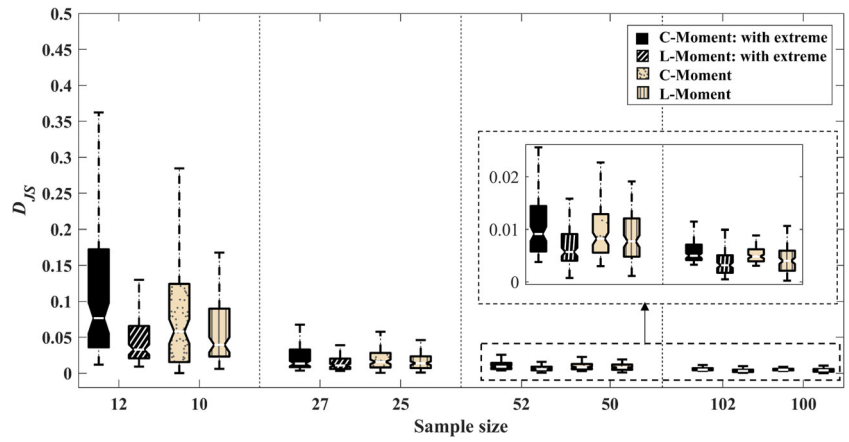


Fig. 21 JS divergence for medium tail

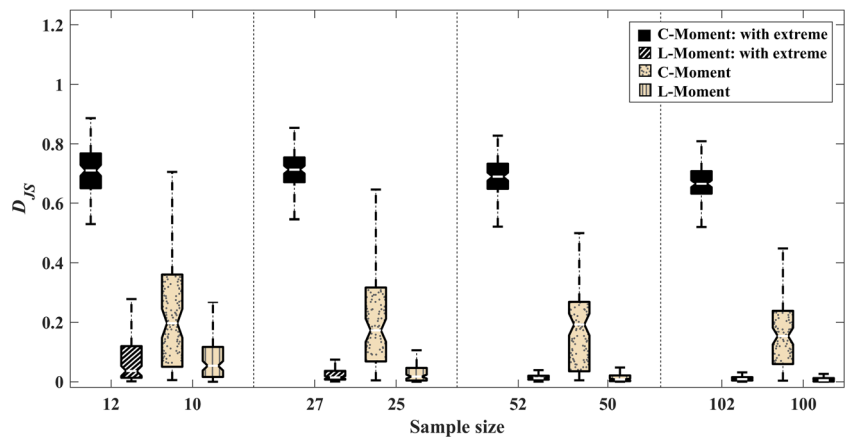


Fig. 22 JS divergence for heavy tail I

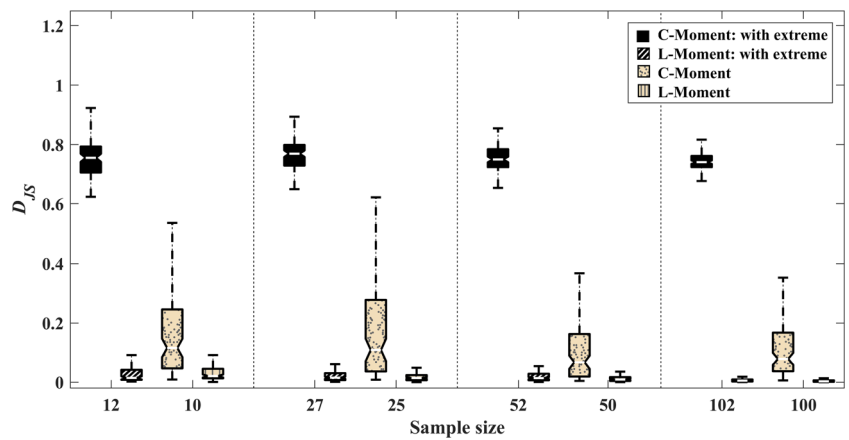
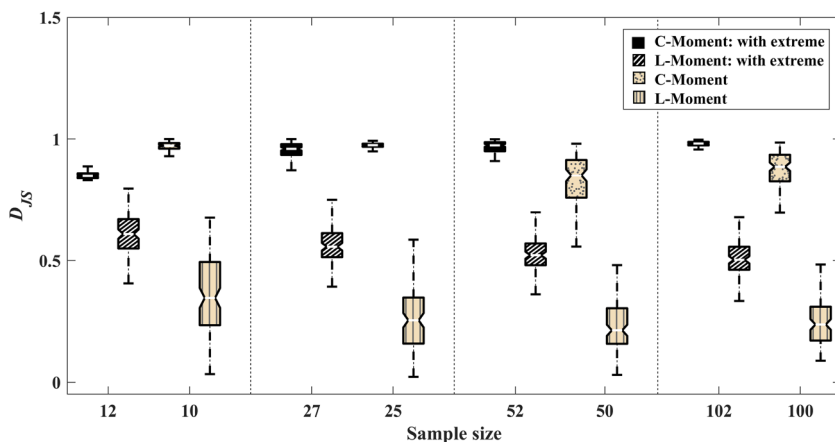


Fig. 23 JS divergence for heavy tail II



are provided in the Tables 15 and 16. Similar observations are obtained as mentioned in one extreme case.

Appendix 4: Design of speed reducer

4.1 Results—One extreme

In this section, the quartile (Q) values of the D_{JS} are presented for different sample sizes. The Q values clearly indicate that L-moment approach are less affected by the sampling variability and works better than C-moment approach.

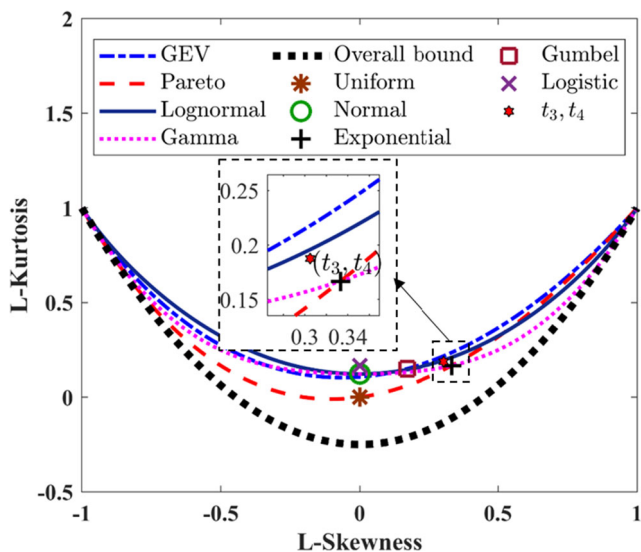


Fig. 24 L-moment ratio diagram with sample L-moment ratio

4.2 Results—Two extremes

D_{JS} results are presented in Figs. 28, 29, 30, and 31 and the corresponding Q values are presented in Tables 21, 22, 23, and 24. The results clearly show that the superiority of the L-moment approach over C-moment approach.

Appendix 5: Probabilistic fatigue life assessment

In this section, probabilistic fatigue life assessment for different sample sizes are presented. $\mu + 0.5\sigma$ of fatigue life is considered the upper bound and the minimum value of ordered experimental data is considered as the lower bound for scarce sample selection. Samples are randomly drawn between the upper and lower bound of the ordered experimental data for all load cases. The survival % comparison results are presented in Fig. 32. The confidence bounds of 90th and 95th survival % are provided in Tables 25, 26, 27, and 28. In all cases, L-moments approach predicts the probability of fatigue life closer to the actual for different loading conditions.

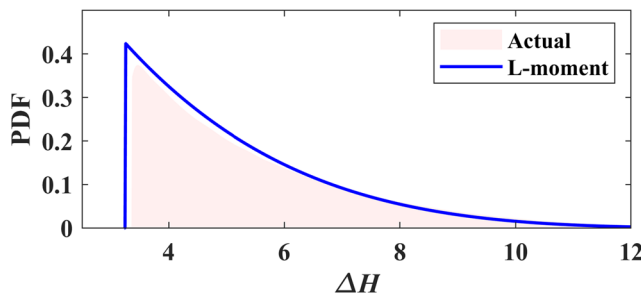


Fig. 25 PDF obtained with sample size 100

Fig. 26 JS divergence - uncertainty quantification

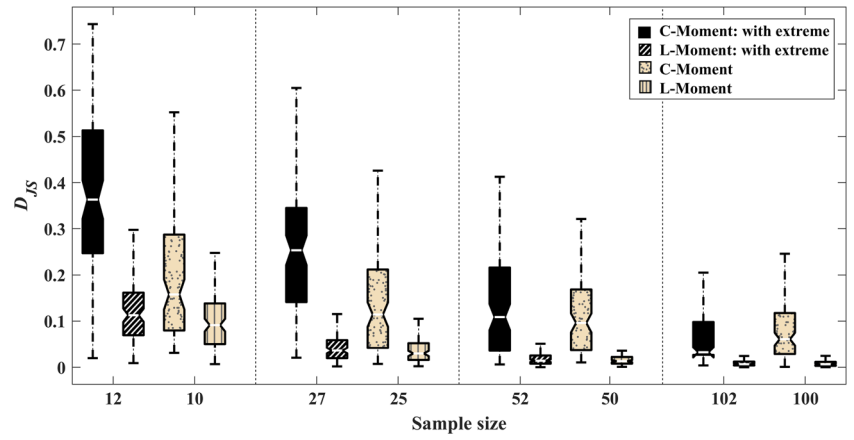


Fig. 27 JS divergence - uncertainty propagation

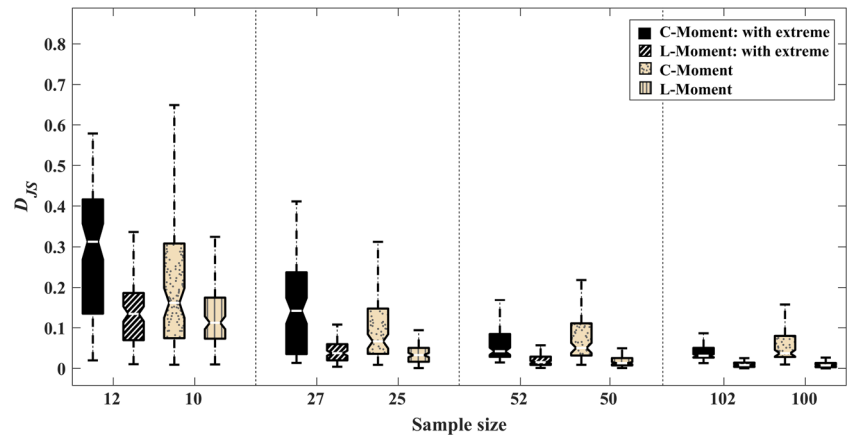


Table 15 Quartile values of JS divergence - uncertainty quantification of ΔH of flat rolling sheet metal manufacturing process

Samples	Distribution identification	Sample size	Quartile values of JS divergence								
			Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
		12	0.247	0.363	0.513	0.141	0.253	0.345	0.036	0.109	0.216
With extreme	Pearson system	27	0.070	0.113	0.162	0.020	0.036	0.059	0.008	0.014	0.026
Samples	L-moment ratio diagram	52									
	Distribution identification	102									
			Sample size								
		10	0.080	0.158	0.287	0.042	0.114	0.212	0.037	0.096	0.169
Without extreme	Pearson system	25	0.051	0.091	0.138	0.016	0.030	0.053	0.008	0.013	0.022
Samples	L-moment ratio diagram	50									
		100									

Table 16 Quartile values of JS divergence - uncertainty propagation of ΔH of flat rolling sheet metal manufacturing process

Samples	Distribution identification	Sample size	Quartile values of JS divergence								
			Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
		12	0.135	0.313	0.416	0.036	0.142	0.237	0.029	0.042	0.085
With extreme	Pearson system	27	0.070	0.135	0.186	0.020	0.038	0.060	0.009	0.016	0.029
Samples	L-moment ratio diagram	52									
	Distribution identification	102									
			Sample size								
		10	0.075	0.162	0.308	0.036	0.067	0.148	0.032	0.051	0.111
Without extreme	Pearson system	25	0.074	0.113	0.175	0.017	0.033	0.051	0.008	0.013	0.026
Samples	L-moment ratio diagram	50									
		100									

Fig. 28 JS divergence - extreme at 1 variable at a time

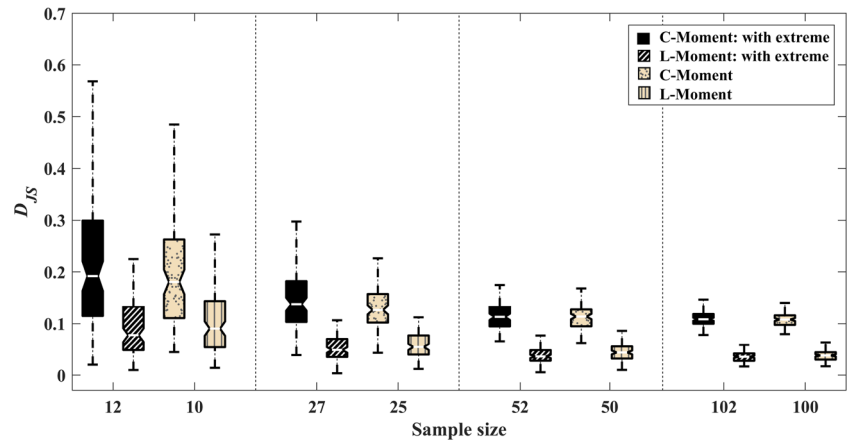


Fig. 29 JS divergence - extreme at 3 variables at a time

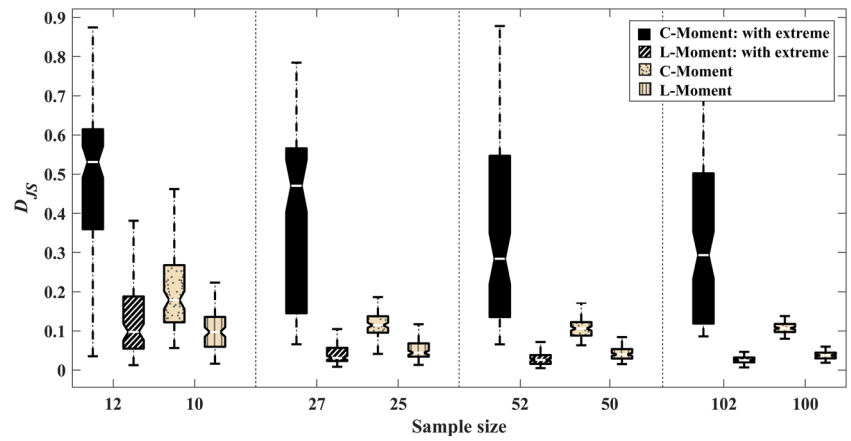


Fig. 30 JS divergence - extreme at 5 variables at a time

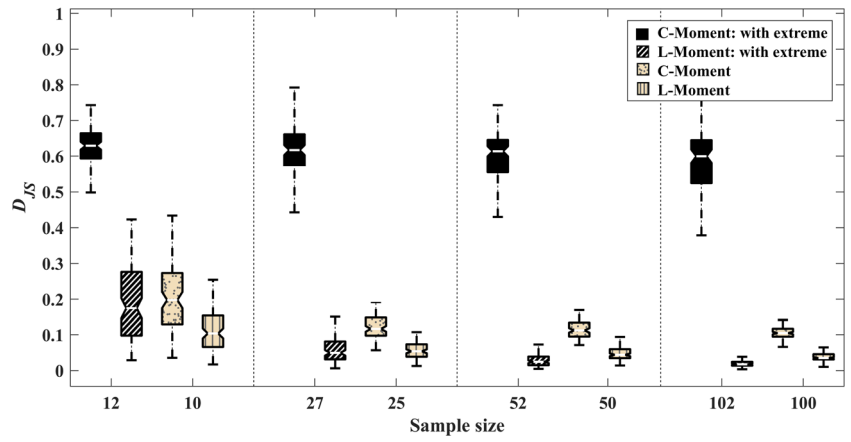


Fig. 31 JS divergence - extreme at 7 variables at a time

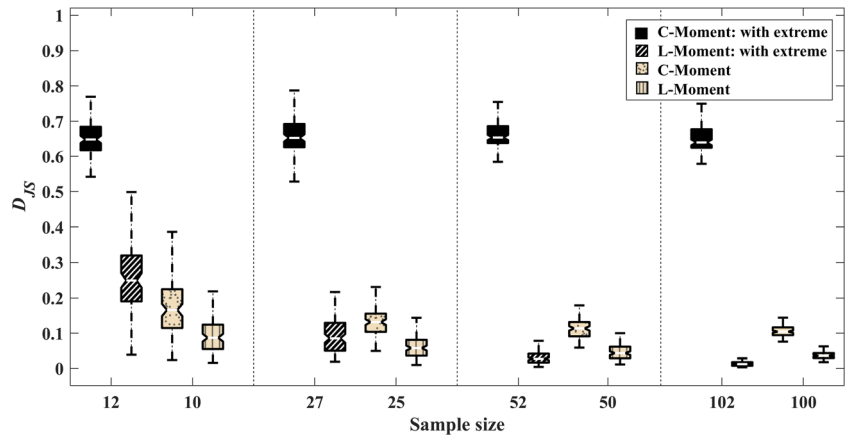


Table 17 Quartile values of JS divergence - extreme at 1 variable at a time

Samples	Distribution identification	Sample size											
		11	26	51	101								
		Quartile values of JS divergence											
		Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
With extreme	Pearson system	0.085	0.127	0.185	0.083	0.101	0.131	0.089	0.103	0.117	0.086	0.097	0.105
	L-moment ratio diagram	0.026	0.048	0.086	0.029	0.037	0.056	0.023	0.034	0.046	0.022	0.029	0.036
Samples	Distribution identification	Sample size											
		10	25	50	100								
		Quartile values of JS divergence											
		Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
Without extreme	Pearson system	0.132	0.182	0.259	0.099	0.116	0.151	0.096	0.111	0.125	0.091	0.100	0.108
	L-moment ratio diagram	0.061	0.096	0.149	0.037	0.053	0.071	0.034	0.041	0.052	0.027	0.033	0.039

Table 18 Quartile values of JS divergence - extreme at 3 variable at a time

Samples	Distribution identification	Sample size											
		11	26	51	101								
		Quartile values of JS divergence											
		Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
With extreme	Pearson system	0.101	0.205	0.513	0.093	0.135	0.403	0.097	0.116	0.210	0.098	0.112	0.137
	L-moment ratio diagram	0.046	0.079	0.150	0.026	0.040	0.064	0.023	0.033	0.044	0.022	0.031	0.039
Samples	Distribution identification	Sample size											
		10	25	50	100								
		Quartile values of JS divergence											
		Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
Without extreme	Pearson system	0.117	0.187	0.257	0.100	0.132	0.153	0.092	0.106	0.123	0.095	0.105	0.114
	L-moment ratio diagram	0.051	0.084	0.150	0.034	0.053	0.076	0.032	0.042	0.053	0.028	0.036	0.042

Table 19 Quartile values of JS divergence - extreme at 5 variables at a time

Samples	Distribution identification	Sample size											
		11	26	51	101								
Quartile values of JS divergence													
With extreme	Pearson system	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
	L-moment ratio diagram	0.518	0.597	0.659	0.377	0.569	0.635	0.331	0.569	0.626	0.198	0.549	0.625
Samples	Distribution identification	0.083	0.173	0.263	0.021	0.041	0.065	0.016	0.028	0.035	0.013	0.018	0.024
Sample size													
		10	25	50	100								
Without extreme	Pearson system	0.116	0.156	0.228	0.092	0.116	0.163	0.098	0.109	0.127	0.094	0.101	0.114
	L-moment ratio diagram	0.053	0.079	0.140	0.034	0.049	0.072	0.032	0.045	0.058	0.025	0.034	0.040

Table 20 Quartile values of JS divergence - extreme at 7 variables at a time

Samples	Distribution identification	Sample size											
		11	26	51	101								
Quartile values of JS divergence													
With extreme	Pearson system	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
	L-moment ratio diagram	0.612	0.652	0.701	0.628	0.652	0.702	0.629	0.667	0.711	0.626	0.656	0.698
Samples	Distribution identification	0.207	0.265	0.349	0.055	0.079	0.118	0.018	0.027	0.047	0.009	0.014	0.019
Sample size													
		10	25	50	100								
Without extreme	Pearson system	0.114	0.171	0.241	0.096	0.121	0.142	0.099	0.112	0.130	0.096	0.106	0.117
	L-moment ratio diagram	0.058	0.086	0.127	0.033	0.051	0.070	0.034	0.045	0.058	0.029	0.037	0.042

Table 21 Quartile values of JS divergence - extreme at 1 variable at a time

Samples	Distribution identification	Sample size											
		27			52			102					
		Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
With extreme	Pearson system	0.115	0.192	0.299	0.103	0.138	0.182	0.095	0.113	0.134	0.100	0.109	0.119
	L-moment ratio diagram	0.049	0.078	0.133	0.035	0.049	0.070	0.028	0.036	0.049	0.028	0.036	0.042
Samples	Distribution identification	Sample size											
		25			50			100					
Without extreme	Pearson system	0.111	0.181	0.263	0.102	0.127	0.157	0.095	0.114	0.128	0.098	0.108	0.116
	L-moment ratio diagram	0.054	0.090	0.143	0.040	0.055	0.077	0.032	0.044	0.056	0.030	0.038	0.044

Table 22 Quartile values of JS divergence - extreme at 3 variables at a time

Samples	Distribution identification	Sample size											
		27			52			102					
		Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
With extreme	Pearson system	0.359	0.531	0.615	0.145	0.470	0.566	0.135	0.284	0.547	0.119	0.293	0.502
	L-moment ratio diagram	0.055	0.097	0.188	0.023	0.031	0.057	0.016	0.026	0.038	0.020	0.025	0.032
Samples	Distribution identification	Sample size											
		25			50			100					
Without extreme	Pearson system	0.122	0.179	0.268	0.096	0.114	0.137	0.088	0.106	0.122	0.097	0.107	0.118
	L-moment ratio diagram	0.060	0.097	0.136	0.034	0.044	0.068	0.029	0.040	0.054	0.030	0.037	0.045

Table 23 Quartile values of JS divergence - extreme at 5 variables at a time

Samples	Distribution identification	Sample size											
		12	27	52	102								
		Quartile values of JS divergence											
		Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
With extreme	Pearson system	0.593	0.629	0.664	0.572	0.617	0.661	0.555	0.614	0.645	0.525	0.600	0.645
	L-moment ratio diagram	0.098	0.175	0.276	0.031	0.049	0.081	0.015	0.024	0.039	0.014	0.017	0.024
Samples	Distribution identification	Sample size											
		10	25	50	100								
Without extreme	Pearson system	0.129	0.198	0.273	0.098	0.117	0.148	0.095	0.112	0.134	0.095	0.105	0.117
	L-moment ratio diagram	0.066	0.104	0.154	0.039	0.054	0.073	0.035	0.044	0.060	0.031	0.035	0.046

Table 24 Quartile values of JS divergence - extreme at 7 variables at a time

Samples	Distribution identification	Sample size											
		12	27	52	102								
		Quartile values of JS divergence											
		Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
With extreme	Pearson system	0.618	0.648	0.684	0.626	0.653	0.692	0.638	0.654	0.686	0.625	0.640	0.677
	L-moment ratio diagram	0.190	0.249	0.320	0.051	0.085	0.129	0.017	0.027	0.042	0.008	0.012	0.018
Samples	Distribution identification	Sample size											
		10	25	50	100								
Without extreme	Pearson system	0.115	0.166	0.225	0.104	0.132	0.155	0.091	0.114	0.132	0.095	0.104	0.117
	L-moment ratio diagram	0.055	0.087	0.124	0.037	0.058	0.081	0.029	0.044	0.062	0.029	0.036	0.044

Fig. 32 Comparison of 95th survival % of sample distribution to parent distribution with different sample sizes

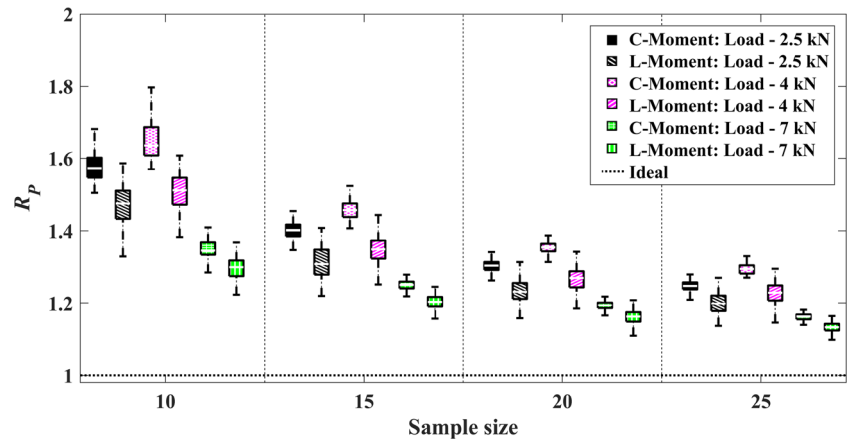


Table 25 Survival % of fatigue life at different load cases and confidence bounds for sample size 10

Load (kN)	Survival (%)								
	90			95					
	C-moment			L-moment					
	Median	Confidence bounds Lower	Upper	Median	Confidence bounds Lower	Upper			
2.5	1287604	1130209	1352811	1032600	1277545	1090925	1358098	977480	2749667
4	192221	189831	209893	165502	189397	166979	199919	151513	367789
7	39738	39717	41174	37316	40867	38357	42231	36802	58183
		61661	55516		60124				

Table 26 Survival % of fatigue life at different load cases and confidence bounds for sample size 15

Load (kN)	Survival (%)								
	90			95					
	C-moment			L-moment					
	Median	Confidence bounds Lower	Upper	Median	Confidence bounds Lower	Upper			
2.5	1375837	1346742	1494383	1186434	1430145	1193661	1486548	1108378	2436936
4	188874	156526	194838	147718	196553	196134	202303	180593	334723
7	39203	30690	39279	31130	38747	36148	39792	34839	53845
		55028	54477		54936				

Table 27 Survival % of fatigue life at different load cases and confidence bounds for sample size 20

Load (kN)	95							
	C-moment			L-moment				
Survival (%)	Median	Confidence bounds Lower	Confidence bounds Upper	Median	Confidence bounds Lower	Confidence bounds Upper		
2.5	1330312	1280966	1409248	1160228	1377295	1140762	1440411	1057059
4	196810	171852	208273	160451	185592	167188	196802	150127
7	40452	36695	41009	35696	40440	32308	40427	32916
		54373		52916		53722		52856

Table 28 Survival % of fatigue life at different load cases and confidence bounds for sample size 25

Load (kN)	95							
	C-moment			L-moment				
Survival (%)	Median	Confidence bounds Lower	Confidence bounds Upper	Median	Confidence bounds Lower	Confidence bounds Upper		
2.5	1393595	1067589	1412782	1038586	1315883	1160147	1378947	1054182
4	173832	154891	182842	141354	183664	157449	192096	143614
7	40589	33286	40607	33625	40386	37346	41032	36329
		51870		51445		52969		51483

Declarations

Conflict of interests The authors declare that they have no conflict of interest.

Replication of results MATLAB® codes to reproduce the L-moment approach results for the examples discussed are made available in this link <https://github.com/DeepanJayaraman/UQ.git>.

References

- Abarbanel H, Koonin S, Levine H, MacDonald G, Rothaus O (1992) Statistics of Extreme Events with Application to Climate; Technical Report; DTIC Document: McLean, VA, USA
- Acar E, Ramu P (2014) Reliability estimation using guided tail modeling with adaptive sampling. In: 16th AIAA non-deterministic approaches conference, pp 1–9. <https://doi.org/10.2514/6.2014-v>
- Adamowski K (2000) Regional analysis of annual maximum and partial duration flood data by nonparametric and L-moment methods. *J Hydrol* 229(3–4):219–231. [10.1016/S0022-1694\(00\)00156-6](https://doi.org/10.1016/S0022-1694(00)00156-6)
- Alvarado E, Sandberg D, Pickford S (1998) Modeling large forest fires as extreme events. *Northwest Science* 72:66–75
- Anderson TV, Mattson CA (2012) Propagating skewness and kurtosis through engineering models for low-cost, meaningful, nondeterministic design. *Journal of Mechanical Design* 134(10):100911. <https://doi.org/10.1115/1.4007389>
- Atiem IA, Harmancioğlu NB (2006) Assessment of regional floods using L-moments approach: the case of the River Nile. *Water Resources Management* 20(5):723–747. <https://doi.org/10.1007/s11269-005-9004-0>
- Buch-Larsen T, Nielsen JP, Guillén M, Bolancé C (2005) Kernel density estimation for heavy-tailed distributions using the champernowne transformation. *Statistics* 39(6):503–518. <https://doi.org/10.1080/02331880500439782>
- Craig CC (1991) A new exposition and chart for the pearson system of frequency curves. *Ann Stat* 7:16–28
- Ceriani L, Verme P (2012) The origins of the gini index: extracts from *variabilità e mutabilità* (1912) by Corrado Gini. *J Econ Inequal* 10(3):421–443
- Chatterjee T, Chakraborty S, Chowdhury R (2017) A critical review of surrogate assisted robust design optimization. *Archives of Computational Methods in Engineering*, pp 1–30. <https://doi.org/10.1007/s11831-017-9240-5>
- David HA (1981) Order statistics. Wiley, New York
- Davison A, Huser R (2015) Statistics of extremes. *Annual Review of Statistics and its Application* 2(1):203–235. <https://doi.org/10.1146/annurev-statistics-010814-020133>
- Elamir EA, Seheult AH (2004) Exact variance structure of sample L-moments. *Journal of Statistical Planning and Inference* 124(2):337–359. [https://doi.org/10.1016/S0378-3758\(03\)00213-1](https://doi.org/10.1016/S0378-3758(03)00213-1)
- Gao W, Song C, Tin-Loi F (2010) Probabilistic interval analysis for structures with uncertainty. *Struct Saf* 32(3):191–199
- Greenwood J, Landwehr J, Matalas N, Wallis J (1979) Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resour Res*, pp 1049–1054. <https://doi.org/10.1029/WR015i005p01049>
- Gubareva TS, Gartsman BI (2010) Estimating distribution parameters of extreme hydrometeorological characteristics by L-moments method. *Water Resources* 37(4):437–445. <https://doi.org/10.1134/S0097807810040020>
- Haddad K, Rahman A, Green J (2011) Design rainfall estimation in australia: a case study using l moments and generalized least squares regression. *Stoch Env Res Risk A* 25(6):815–825
- Hall P, Sheather SJ, Jones M, Marron JS (1991) On optimal data-based bandwidth selection in kernel density estimation. *Biometrika* 78(2):263–269. <https://doi.org/10.1093/biomet/78.2.263>
- Hosking JRM (1989) Some theoretical results concerning L-moments. IBM Thomas J. Watson Research Division
- Hosking JRM (1990) L-moments: analysis and estimation of distributions using linear combinations of order statistics. <https://doi.org/10.2307/2345653>
- Hosking JRM (1992) Moments or L-moments - an example comparing 2 measures of distributional shape. *Am Stat* 46(3):186–189. <https://doi.org/10.2307/2685210>
- Hosking JRM (2006) On the characterization of distributions by their L-moments. *Journal of Statistical Planning and Inference* 136(1):193–198. <https://doi.org/10.1016/j.jspi.2004.06.004>
- Hosking JRM, Wallis JR (1997) Regional frequency analysis: an approach based on L-moments. Cambridge University Press
- Hu Z, Du X, Conrad D, Twohy R, Walmsley M (2014) Fatigue reliability analysis for structures with known loading trend. *Struct Multidiscip Optim* 50(1):9–23
- Jayaraman D, Ramu P (2019) Uncertainty propagation using L-moments with scarce samples including extremes. In: Proc. 13th world congress of structural and multidisciplinary optimization, pp 15–21
- Jayaraman D, Ramu P, Suresh SK, Ramanath V (2018) Treating uncertainties to generate a robust design of gas turbine disk using l-moments and scarce samples including outliers. In: Turbo expo: power for land, sea, and air, vol 51135, p V07AT32A008. <https://doi.org/10.1115/GT2018-76431>
- Jin R, Du X, Chen W (2003) The use of metamodeling techniques for optimization under uncertainty. *Struct Multidiscip Optim* 25(2):99–116. <https://doi.org/10.1007/s00158-002-0277-0>
- Kang YJ, Noh Y, Lim OK (2018) Kernel density estimation with bounded data. *Struct Multidiscip Optim* 57(1):95–113. <https://doi.org/10.1007/s00158-017-1873-3>
- Kenney J, Keeping E (1947) Mathematics of Statistics. No. pt. 2 in *Mathematics of Statistics*, Van Nostrand
- Kumar R, Chatterjee C (2005) Regional flood frequency analysis using l-moments for north brahmaputra region of India. *J Hydrol Eng* 10:1–7. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2005\)10:1\(1\)](https://doi.org/10.1061/(ASCE)1084-0699(2005)10:1(1))
- Lee G, Kim W, Oh H, Youn BD, Kim NH (2019) Review of statistical model calibration and validation—from the perspective of uncertainty structures. *Structural and Multidisciplinary Optimization*. <https://doi.org/10.1007/s00158-019-02270-2>
- Lee SH, Chen W (2009) A comparative study of uncertainty propagation methods for black-box-type problems. *Struct Multidiscip Optim* 37(3):239–253. <https://doi.org/10.1007/s00158-008-0234-7>
- Lee SH, Chen W, Kwak BM (2009) Robust design with arbitrary distributions using Gauss-type quadrature formula. *Struct Multidiscip Optim* 39(3):227–243. <https://doi.org/10.1007/s00158-008-0328-2>
- Lin MH, Tsai JF, Hu NZ, Chang SC (2013) Design optimization of a speed reducer using deterministic techniques. *Math Probl Eng* 2013. <https://doi.org/10.1155/2013/419043>
- Liu H, Jiang C, Liu J, Mao J (2019) Uncertainty propagation analysis using sparse grid technique and saddlepoint approximation based on parameterized p-box representation. *Struct Multidiscip Optim* 59(1):61–74
- Mekid S, Vaja D (2008) Propagation of uncertainty: expressions of second and third order uncertainty with third and fourth moments.

- Measurement: Journal of the International Measurement Confederation 41(6):600–609. <https://doi.org/10.1016/j.measurement.2007.07.004>
- Melville P, Yang SM, Saar-Tsechansky M, Mooney R (2005) Active learning for probability estimation using Jensen-Shannon divergence. In: Gama J, Camacho R, Brazdil PB, Jorge AM, Torgo L (eds) Machine learning: ECML 2005. Springer, Berlin, pp 268–279
- Moon MY, Kim HS, Lee K, Park B, Choi KK (2020) Uncertainty quantification and statistical model validation for an offshore jacket structure panel given limited test data and simulation model. *Structural and Multidisciplinary Optimization*. <https://doi.org/10.1007/s00158-020-02520-8>
- Moustapha M, Sudret B (2019) Surrogate-assisted reliability-based design optimization: a survey and a unified modular framework. *Structural and Multidisciplinary Optimization* 60(5):2157–2176. <https://doi.org/10.1007/s00158-019-02290-y>, 1901.03311
- Nair NU, Vineshkumar B (2010) L-moments of residual life. *Journal of Statistical Planning and Inference* 140(9):2618–2631
- Park BU, Marron JS (1990) Comparison of data-driven bandwidth selectors. *J Am Stat Assoc* 85(409):66–72. <https://doi.org/10.1080/01621459.1990.10475307>
- Pearson K (1916) Mathematical contributions to the theory of evolution. XIX. Second Supplement to a Memoir on Skew Variation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 216(538–548):429–457. <https://doi.org/10.1098/rsta.1916.0009>
- Ramu P, Arul S (2016) Estimating probabilistic fatigue of nitinol with scarce samples. *Int J Fatigue* 85:31–39. <https://doi.org/10.1016/j.ijfatigue.2015.11.022>
- Ramu P, Kim NH, Haftka RT (2010) Multiple tail median approach for high reliability estimation. *Struct Saf* 32(2):124–137
- Ramu P, Kumar GS, Neelakantan P, Bathula KK (2017) Cost-reliability trade-off of path generating linkages using multi-objective genetic algorithm. *International Journal of Reliability and Safety* 11(3–4):200–219. <https://doi.org/10.1504/IJRS.2017.089706>
- P Ramu (2013) Modified third order polynomial approach for reliability analysis with scarce samples. In: Proceedings of the 10th world congress on structural and multidisciplinary optimization, Orlando, USA (2010), pp 1–10
- Rokne JG (2001) Interval arithmetic and interval analysis: an introduction. In: Granular computing. Springer, Berlin, pp 1–22
- Rudemo M (1982) Empirical choice of histograms and kernel density estimators. *Scand J Stat* 9(2):65–78
- Sankarasubramanian A, Srinivasan K (1999) Investigation and comparison of sampling properties of L-moments and conventional moments. *Journal of Hydrology* 218(1–2):13–34. [https://doi.org/10.1016/S0022-1694\(99\)00018-9](https://doi.org/10.1016/S0022-1694(99)00018-9)
- Schäfer HJ (2008) Auswertalgorithmus auf der Basis einer Modifikation des Goniometrischen Modells zur stetigen Beschreibung der Wöhlerkurve vom Low-Cycle-Fatigue-bis in den Ultra-High-Cycle-Fatigue-Bereich. Mainz
- Shi L, Fu Y, Yang RJ, Wang BP, Zhu P (2013) Selection of initial designs for multi-objective optimization using classification and regression tree. *Structural and Multidisciplinary Optimization* 48(6):1057–1073. <https://doi.org/10.1007/s00158-013-0947-0>
- Shirahata S, Is Chu (1992) Integrated squared error of kernel-type estimator of distribution function. *Ann Inst Statist Math* 44(3):579–591
- Sillitto GP (1969) Derivation of approximants to the inverse distribution function of a continuous univariate population from the order statistics of a sample. *Biometrika* 56(3):641–650
- Silverman BW (1986) Density estimation for statistics and data analysis 26
- Smithers J, Schulze R (2001) A methodology for the estimation of short duration design storms in South Africa using a regional approach based on l-moments. *J Hydrol* 241(1–2):42–52
- Turlach BA (1993) Bandwidth selection in kernel density estimation: a review. In: CORE and Institut de Statistique, Citeseer
- Voinov A, Kolagani N, McCall MK, Glynn PD, Kragt ME, Ostermann FO, Pierce SA, Ramu P (2016) Modelling with stakeholders - next generation. *Environmental Modelling and Software* 77:196–220. <https://doi.org/10.1016/j.envsoft.2015.11.016>
- Weisstein EW, et al. (2004) Pearson system, from mathworld—a wolfram web resource
- Zafirakou-Koulouris A, Vogel RM, Craig SM, Habermeier J (1998) L moment diagrams for censored observations. *Water Resour Res* 34(5):1241–1249. <https://doi.org/10.1029/97WR03712>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.