Using Bootstrap to Assess Sampling Uncertainty in Fatigue Crack Growth Life Due to Limited Test Data

K. S. Bhachu,[*] R. T. Haftka,[†] and N.-H. Kim[‡]

University of Florida, Gainesville, Florida, 32611, U.S.A

Substantial sampling uncertainty is introduced due to limited number of fatigue crack growth tests to characterize material's fatigue crack growth behavior. We explore the use of bootstrap resampling technique to estimate the sampling uncertainty, and construct lower confidence bounds on a given percentile of fatigue crack growth life (FCGL) distribution. With a small sample size, desired confidence level cannot be achieved without using a correction factor to adjust the confidence bounds. We calculate correction factors for number of test samples ranging from 8-40 to achieve 95% confidence level for the lower confidence bounds on 0.01 percentile of FCGL's lognormal distribution. Three different bootstrap based methods for setting confidence intervals are explored, i.e. basic percentile method, normal approximation, and bias corrected accelerated percentile method. Latter two methods were found to achieve better confidence levels for a given number of test samples without using correction factor. Although, basic percentile method gave least conservative bounds than others before and after correction. We use artificially generated fatigue crack growth rate data for simulation, and calculate FCGL for the geometry that assumes a single through crack at a hole.

Introduction

This paper explores the use of bootstrap resampling technique for constructing one-sided confidence interval (lower) on a given percentile '*p*' of fatigue crack growth life (FCGL) distribution. A confidence interval for a point estimate of certain percentile of distribution (e.g. p = 0.01 percentile) from limited data is supposed to contain the true/population (unknown) percentile value. The one-sided confidence interval on an estimated percentile value is analogous to a statistical tolerance interval that contains certain proportion of a population. For example, one-sided lower confidence bound that contains 5 percentile population value is the same lower tolerance bound that contains 95% of a population. So, this paper does not make distinction between confidence and tolerance intervals. A probability statement is also attached with a confidence interval to express the confidence level or assurance that confidence interval would contain the true percentile value. A confidence level of 95% means that 95% of the confidence intervals constructed from different set of random samples would contain the true percentile value.

A usual approach to construct confidence bounds (or tolerance bounds) on a point estimate from small sample ($n_s \leq 40$) is to assume a normal distribution for a random variable, and also for the sampling distribution of a given percentile. However, if distribution of a random variable is non-normal (e.g. lognormal for FCGL) and sample size is small ($n_s \leq 40$), then normal approximation for sampling distribution of a given percentile may give confidence bounds that does not achieve a target confidence level (e.g. 88% confidence level instead of targeted 95%). In such cases, bootstrap (a resampling technique) may provide a reasonable approximation to the shape of a sampling distribution without making normal distribution assumption for the random variable. If bootstrap indicates that sampling distribution could be reasonably approximated by the normal distribution then one can resort to the usual method of setting confidence bounds (for a given percentile) based on normal distribution assumption. On the other hand, if bootstrap indicates that sampling distribution is severely non-normal then confidence bounds should be set using one of the many bootstrap confidence interval methods, e.g. percentile method, and bias corrected accelerated method.

---

[*] Doctoral Candidate, Dept. of Mechanical & Aerospace Engineering, kdsbhachu@gmail.com, AIAA Student Member.
[†] Distinguished Professor, Dept. of Mechanical & Aerospace Engineering, haftka@ufl.edu, AIAA Fellow.
[‡] Associate Professor, Dept. of Mechanical & Aerospace Engineering, nkim@ufl.edu, AIAA Associate Fellow.

However, bootstrap may also underestimate a confidence level (e.g. 85% instead of 95%) when small sample size is used for building a sampling distribution of '*p*' percentile FCGL. The underestimation happens partly due to small sample sizes, and partly due to the fact that resampling techniques have trouble with extreme values i.e. sampling distribution may take significantly different shapes when estimated from one set of random sample to another. Also, bootstrap would not be any good if very small data is available (e.g. $n_s < 8$). We only apply bootstrap to an example problem that can afford minimum of 8 samples (i.e. 8 fatigue crack growth tests). The underestimation of a confidence level can be corrected by using a correction factor that adjusts (lengthen) the lower confidence intervals for a given percentile of FCGL to achieve the target confidence level. It would be appropriate to derive correction factors for very small percentile values (e.g. 0.01 percentile) if underlying probability model for a random variable is known or if reasonable choice could be made by fitting probability models to a small random sample. That is, a suitable probability model would allow to extrapolate into the tails and find very low percentile values (perhaps more accurately) instead of estimating it directly from the empirical cumulative distribution function (CDF) of a small sample. In case of fatigue crack growth life (FCGL), experimental tests indicated that [1-2] lognormal distribution may be a reasonable choice. Therefore, we will calculate correction factors for different sample sizes ranging from 8 – 40 by assuming a lognormal distribution for FCGL.

Probabilistic Fatigue Crack Growth Life Prediction with Limited Test Data

We consider the problem of characterization of material's fatigue crack growth (FCG) behavior from limited test data. Material's FCG behavior is captured by estimating constants (coefficients) of a FCG model, which is done by fitting FCG model to a FCG data obtained via laboratory testing. The most common FCG model used is the Paris law,

$$\frac{da}{dN} = C\left(\Delta K\right)^{-n} \tag{1}$$

Where, da/dN is the crack growth rate (in/cycle); ΔK is the crack tip stress intensity; C is Paris constant, and n is Paris exponent. The material constants (e.g. C and n) of a FCG model are random variables that capture randomness due to material variability from one test specimen to another. The variability in material constants further leads to variability in the predicted fatigue crack growth life (FCGL). The predicted FCGL can be found by integrating Eq.(1),

$$N = \frac{1}{C}\int_{a_{ini}}^{a_{cri}} \left(\Delta K\right)^{-n} da \tag{2}$$

Where, N is the crack growth life; a_{ini} is initial crack length, and a_{cri} is the critical crack length. Therefore, predicted FCGL is also becomes a random variable, and analyst is interested in designing a structure that meets some probability of failure constraint. That is, structure is sized such that '*p*' percentile of predicted FCGL distribution (i.e. $N_{p\%}$) is equal to the desired structural life (N^*) as shown in Fig. 1.

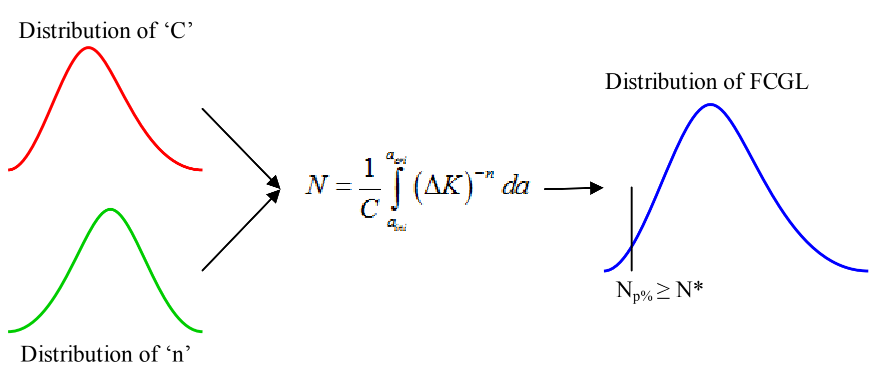

Fig. 1 Propagation of uncertainty in material constants to uncertainty in FCGL

However, substantial sampling uncertainty in introduced in the $N_{p\%}$ estimated from the limited test data available for estimating the probability distribution function PDFs of the material constants. That is, PDFs of material constants are generally obtained from a few laboratory tests (8 – 40) that are conducted under specific load conditions. These tests are time consuming and each costs about $ 1,000 - $ 2,000, which often proves prohibitive in conducting large number of tests. So, first objective for an analyst is to construct a lower (one-sided) confidence interval that is reliable (e.g. 95% confidence level) enough to contain the '*p*' percentile of the true PDF (unknown to analyst) as shown in Fig. 2. The second objective is to have a lower confidence bound that is not overly conservative and still achieves the desired confidence level.

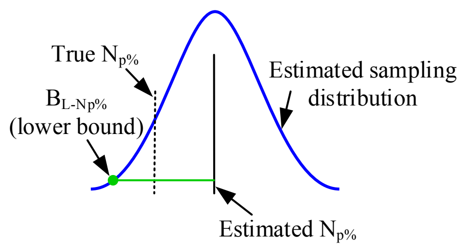

Fig. 2 Sampling distribution and lower confidence bound on estimated $N_{p\%}$

This paper explores the use of bootstrap resampling technique to construct one-sided lower confidence bound on the 0.01 percentile FCGL with 95% confidence level. Three methods of deriving confidence intervals based on bootstrap are explored, i.e. basic percentile method (BPM), normal approximation (NA), and bias corrected accelerated percentile method (BCAP). The confidence bounds from these three methods will be first checked to determine the confidence level achieved ($\alpha_{achieved}$) against the target confidence level of α_{target} = 95% for the number of FCG tests samples ranging from n_s= 8 – 40. Then correction factors F_{corr} will be estimated that when multiplied with the lower confidence bounds will achieve the α_{target} of 95%. Finally, the confidence bounds (after correction) derived from the three methods will be compared to find the one that gives least conservative bounds (i.e. higher crack growth life) for a given sample size of FCG tests. An artificially generated FCGL data is discussed next that is used for the simulation.

Synthetic Crack Growth Life Data and Simulation

The synthetic FCGL data needed for simulation is generated by using parameters of the material constant distributions given by [1], and are assumed here to define the true (population) PDFs of the material constants. The distributions of material constants were estimated by fitting Paris law to the data obtained from 68 fatigue crack growth (FCG) tests performed by Virkler et al. [2] on the test specimens made from 2024-T3 aluminum alloy. The tests were performed for the stress ratio of R = 0. The parameters of the marginal PDFs of the material constants are given in Table 1. Further, these material constants have strong negative correction with correction coefficient of r = -0.982. For the purpose of random sample generation, '$log(C)$' and 'n' are modeled jointly as multivariate normal distribution.

Table 1 Parameters of distributions of material constants for *R* = 0

Material constant	Distribution	Location	Scale
C	lognormal	-14.972	0.328
n	normal	2.872	0.165

Note that, the fatigue crack growth rate (da/dN vs. K) curves generated using the random samples from multivariate normal distribution seemed little off from that shown in [2]. It could be due to unit conversion, i.e. when random samples of 'C' are multiplied by 3.937×10^{-2} (mm/cycle to in/cycle conversion) it approximately gives similar FCGR graph given in [2].

The schematic shown in Fig. 3 draws similarity between real physical FCG testing and simulation. In reality, FCG testing is done for different test conditions (e.g. for stress ratios, R = 0.05, 0.30, 0.50, and 0.70) using the test specimens similar to the one shown in Fig. 4 (a). The real testing is repeated n_{rep} times for a particular test condition, and Walker equation,

$$\frac{da}{dN} = C\left[\Delta K\left(1-R\right)^{m-1}\right]^{n} = C\left(\Delta K_{eff}\right)^{n}$$
$$N = \frac{1}{C}\int_{a_{ini}}^{a_{cri}}\left[\Delta K\left(1-R\right)^{m-1}\right]^{-n} da = \frac{1}{C}\int_{a_{ini}}^{a_{cri}}\left(\Delta K_{eff}\right)^{-n} da \tag{3}$$

, can be used to fit and collapse all the FCG data to single test condition of $R = 0$. Then, total samples of C and n from $n_s = 4n_{rep}$ tests can be used to approximate their marginal PDFs. The marginal PDFs from limited testing can be used to estimate the distribution of FCGL by using simulation geometry shown in Fig. 4 (b), and load conditions listed in Table 2. The estimated lognormal PDF of FCGL is further used to approximate $N_{p\%}$ FCGL that is followed by constructing α% one-sided lower confidence bound $B_{L\text{-}Np\%}$ that should bound the true $N_{p\%}$.

Similarly in simulation environment, the physical testing is replaced with FCG data generation from the known true distributions of C and n. Here we assume the parameters given in Table 1 as true parameters and use those samples to generate FCG data and follow the same procedure of building confidence bounds as done for real physical testing. However, in our case an efficient way is to bypass the walker equation fitting by directly sampling n_s samples from the true PDF of FCGL and proceed from there. The simulation shown in red dotted box in Fig. 3 is repeated 10,000 times to estimate the proportion of confidence bounds that contain the true value of $N_{p\%}$. The lower confidence bounds are multiplied with a correction factor to achieve the targeted α% confidence level.

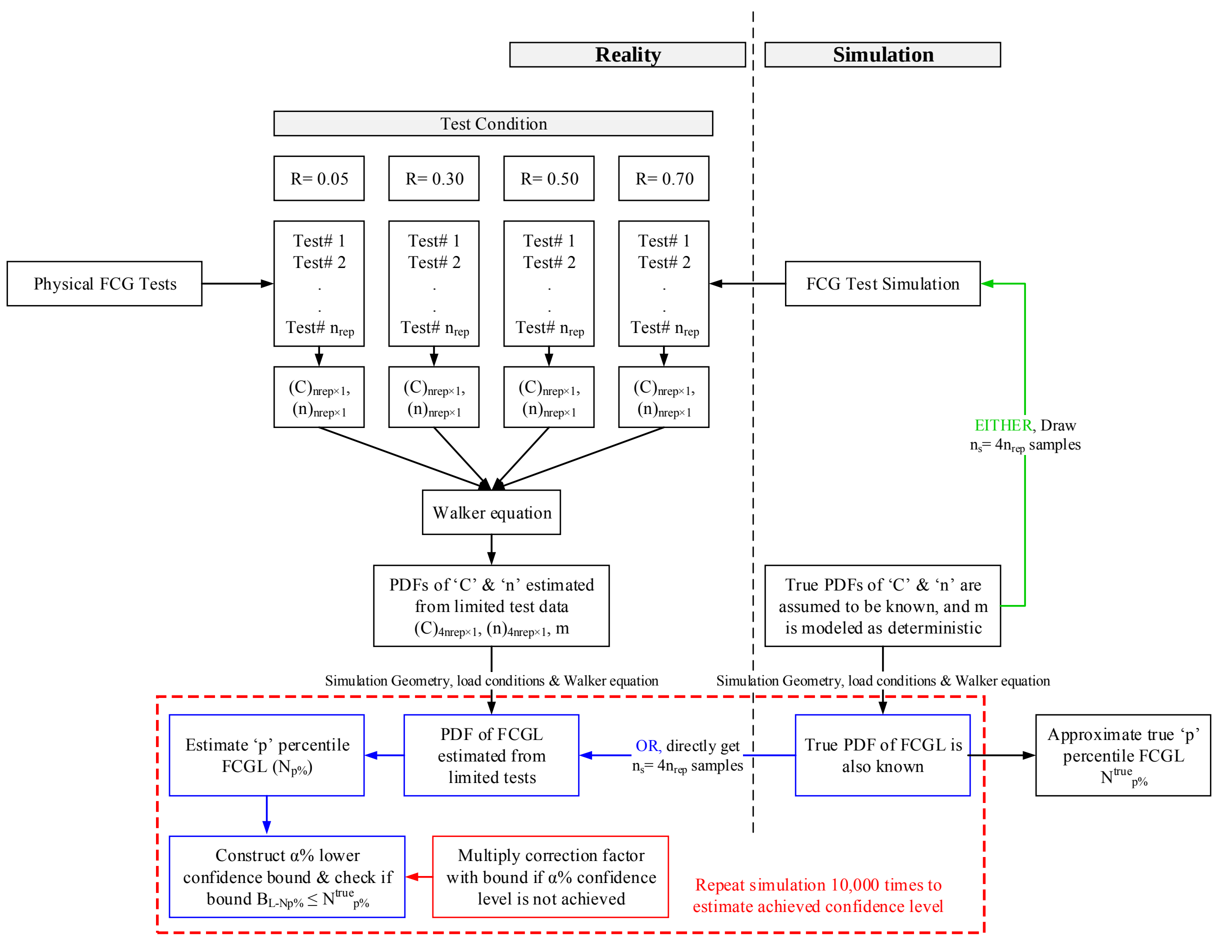

Fig. 3 Schematic representation of FCGL data generation and simulation

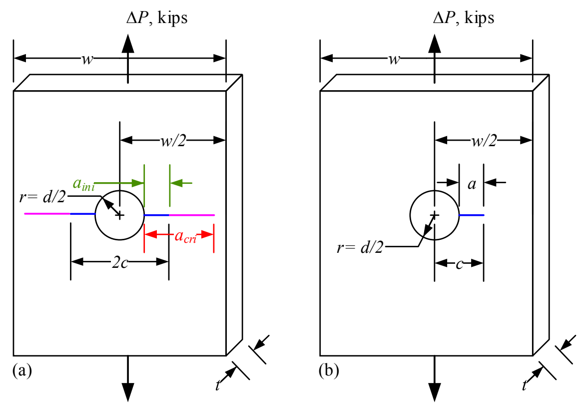

Fig. 4 (a) Double through crack (test geometry), (b) Single through crack (simulation geometry)

Table 2 Dimensions and load conditions for the simulation geometry

Dimension	Unit	Value
w	inch	4
a_{ini}	inch	0.05
a_{cri}	inch	2
d	inch	0.2
t	inch	0.16
P_{max}	kips	5
R	-	0.05

True PDF of FCGL

The true PDF of FCGL is estimated by generating 5×10^7 samples from the multivariate normal distribution of log(C) and n. The distribution is then used to approximate the true $N_{0.01\%}$ (i.e. true 0.01 percentile FCGL). The histogram of 5×10^7 samples of FCGL are shown in

Fig. 5. The true $N_{0.01\%}$ estimated from the samples is 200,192 cycles, and uncertainty in the calculation of probability of failure of 10^{-4} from sample of 5×10^7 is about 1.4%. The lognormal distribution was fitted to the histogram and had following parameters (μ = 12.4418, σ= 0.0633). The $N_{0.01\%}$ found from lognormal distribution was 200,063 cycles (error of only 129 cycles, which is very small). So lognormal distribution can be used to generate samples for simulation, and is a reasonable probability model to estimate the $N_{0.01\%}$ from the small sample set.

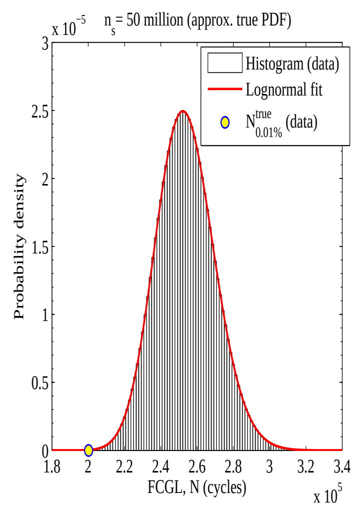

Fig. 5 True (approx.) PDF of FCGL

True Sampling Distribution of $N_{0.01\%}$ due to Limited Data

The true sampling distribution of $N_{0.01\%}$ can be estimated by generating n_s samples repeatedly from the FCGL's lognormal distribution shown above in

Fig. **5**. Each set of n_s samples can be fitted with lognormal distribution and $N_{0.01\%}$ can be estimated from the fit. That is, the Monte Carlo simulation with n_s samples drawn 100,000 times will give the true sampling distribution of $N_{0.01\%}$. Mean of the sampling distribution is 202,128 cycles and standard deviation of 13,230 cycles, i.e. mean has slight positive bias of 202,128 - 200,192 = 1,936 cycles (1%). The true sampling distribution of $N_{0.01\%}$ is approximated well by a normal distribution as indicated in the probability plots shown in Fig. 6.

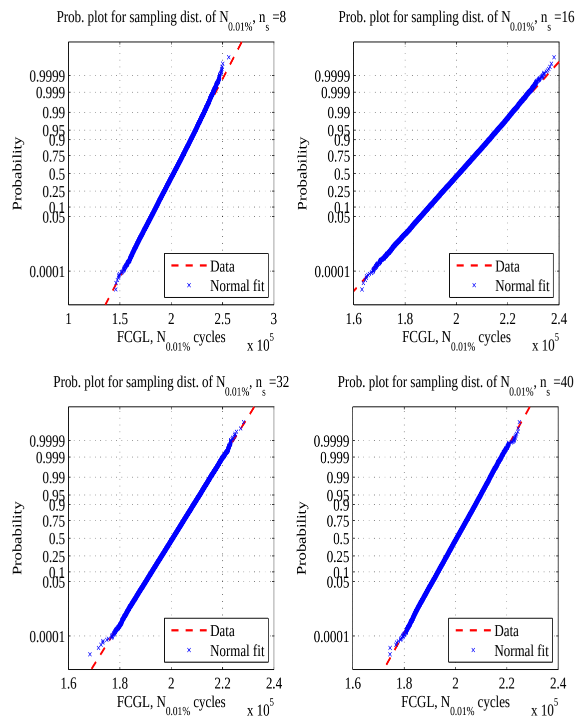

Fig. 6 True sampling distribution approximated well by normal distribution

Bootstrap Confidence Bounds From Limited Data

In reality, the nature of the sampling distribution has to be estimated from the limited test sample n_s available. Bootstrap is a resampling technique that can provide an approximation to the shape or nature of the sampling distribution. From the discussion above, we have all the reason to assume that sampling distribution is normal and set one-sided lower 95% confidence bound based on normal assumption. First, bootstrap based small sample estimate of the bias in $N_{0.01\%}$ and standard error can be used to construct the lower 95% confidence bounds ($B_{L\text{-}N0.01\%}$). Second, bootstrap based basic percentile method (BPM) that does not require assumption of normality can be used to set $B_{L\text{-}N0.01\%}$. BPM directly calculates the 5 percentile value from the sampling distribution of $N_{0.01\%}$ and uses that as a lower bound. Third, a very popular bootstrap based bias corrected accelerated percentile (BCAP) method is used to set the lower confidence bounds.

As discussed earlier that all the three bootstrap based methods are likely to underestimate the confidence level achieved e.g. only 88% (instead of 95%) of the confidence bounds contain that true value of $N_{0.01\%}$. Notice from **Table 3** that none of the three methods achieved the target confidence level of 95%, i.e. less than 95% of the lower confidence bounds contained the true 0.01 percentile of FCGL. Normal approximation for sampling distribution of 0.01 percentile FCGL gave most accurate confidence levels (closer to the target of 95%) for sample sizes ranging from 8-16 as compared to the BCAP (bias corrected accelerated percentile) method. For sample size ranging from 20-40, confidence levels given by BCAP method were marginally more accurate than the normal approximation. Also, both NA and BCAP methods always gave more accurate confidence levels than BP (basic percentile) method.

Table 3 Confidence level achieved by the three bootstrap based methods

n_s	$\alpha_{achieved}$ % (α_{target} = 95%),		
	BPA*	NA*	BCAP*
8	71.8	87.1	81.6
12	78.3	88.8	87.0
16	81.3	89.9	89.2
20	82.8	90.8	90.9
24	85.1	91.4	91.7
28	85.1	91.3	91.6
32	86.3	91.9	92.5
36	87.9	92.5	93.2
40	87.8	92.5	93.3

*Based on 10,000 bootstraps for each of the 10,000 simulations

Next, we estimate the correction factors (F_{corr}), which when multiplied with uncorrected lower confidence bounds gives corrected confidence bounds that helps in achieving the desired confidence level of 95%.

$$B_{L-N0.01\%}^{corr} = F_{corr} B_{L-N0.01\%}^{uncorr} \tag{4}$$

The correction factors are listed in Table 4 and are only valid for the setting 95% confidence bound on $N_{0.01\%}$ assuming that FCGL has a lognormal distribution.

Table 4 Correction factors for achieving 95% confidence bounds on $N_{0.01\%}$

n_s	Correction factors, F_{corr}		
	BPA	NA	BCAP
8	0.9280	0.9597	0.9441
12	0.9478	0.9704	0.9643
16	0.9594	0.9780	0.9761
20	0.9693	0.9845	0.9840
24	0.9745	0.9870	0.9882
28	0.9774	0.9886	0.9897
32	0.9805	0.9905	0.9928
36	0.9838	0.9928	0.9943
40	0.9851	0.9936	0.9954

Another, important requirement from the confidence bounds is that bounds should not be overly conservative. That is, lower bound should be tight enough to give the desired confidence level without giving too small design life. Now, even though BPM gave least accurate confidence bounds of the three methods but it gave tightest bounds both before and after correction, i.e. highest design life that would lead to weight savings. For example, notice from Fig. 7 that BPM gives highest design life (tightest lower bounds) by comparing the cumulative distribution functions (CDFs) of the lower bounds from the three methods. These empirical CDFs were constructed from 10,000 simulations. The statistics of empirical CDFs before and after correction are shown in Appendix for different sample sizes.

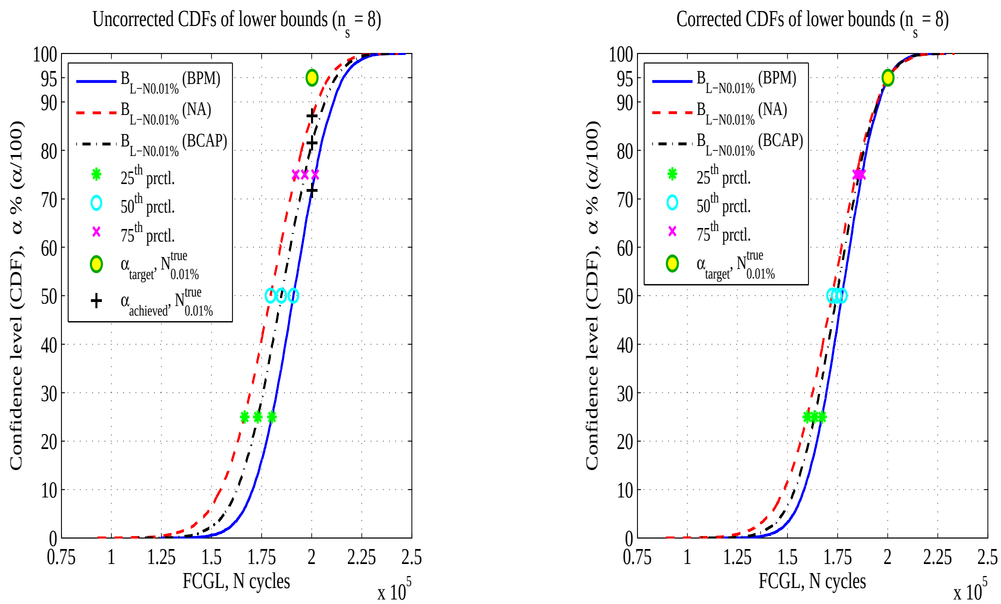

Fig. 7 CDFs of lower bounds for n_s = 8

Similarly, BPM was found to give tightest bounds (or higher FCGL) for n_s ranging from 8-40 as shown in Fig. 8. Note that, the difference between bounds given by the three methods tend to reduce with increase in number of samples or tests. So, perhaps it is fine to use bounds given by normal approximation if $n_s \geq 24$ as design life achieved by all the methods is pretty much same. For $n_s < 24$, it may be worthwhile to use BPM to set confidence bounds as weight savings might be substantial.

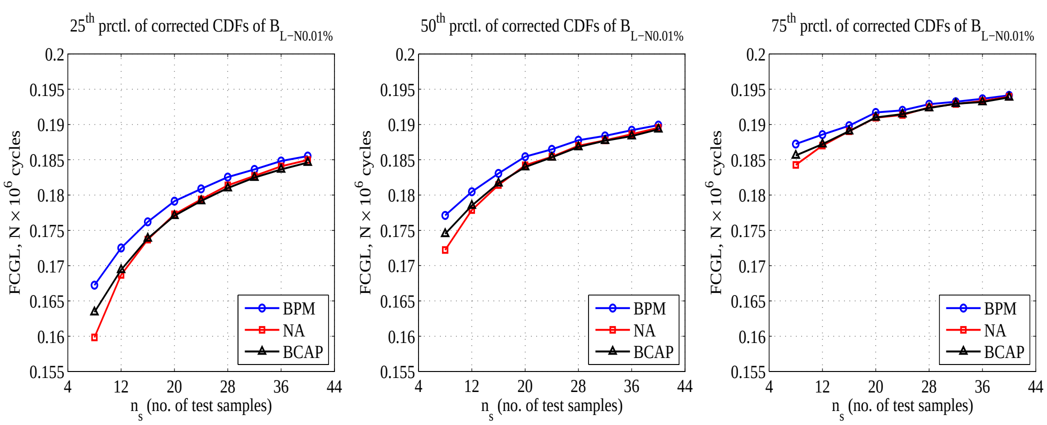

Fig. 8 25th, 50th, and 75th percentile points of the CDFs of lower bounds from different methods as a function of number of test samples

Conclusions

Key findings,

1. Normal approximation for sampling distribution of 0.01 percentile FCGL gave most accurate confidence levels (closer to the target of 95%) for sample sizes ranging from 8-16 as compared to the BCAP (bias corrected accelerated percentile) method. It always gave more accurate confidence levels than BP (basic percentile) method.

2. For sample size ranging from 20-40, confidence levels gives by BCAP method were marginally more accurate than the normal approximation.

3. Percentile method gave shortest lower bounds both before and after correction for any number of test samples ranging from 8-40. Shortest lower bound translates into higher design life, which would further lead to weight savings. Therefore, it may be beneficial to use percentile method (with correction) for setting lower confidence bounds when sample size is less than 24.

4. All three methods tend to give comparable bounds as sample size increases. So, one could use any of the three methods if the sample size is greater than 24.

5. As expected, sampling uncertainty reduces with the increase in sample size irrespective of the method used for setting lower confidence bounds. So, by performing more tests one can increase the chances of saving weight as lower confidence bound based on limited test data may give higher design life. Note that weight saving is not guaranteed as lower confidence bound would depend on the random sample one gets from testing.

Appendix

Table 5 Statistics of uncorrected CDFs of lower bounds using BPM

n_s	μ, mean (cycles)	σ, stand. dev. (cycles)	25th prctl. (cycles)	50th prctl. (cycles)	75th prctl. (cycles)
8	190,786	15,617	180,201	190,872	201,753
12	190,441	12,530	181,997	190,417	198,955
16	190,759	10,736	183,668	190,813	197,859
20	191,135	9,518	184,797	191,319	197,777
24	191,321	8,521	185,605	191,364	197,021
28	191,998	7,861	186,763	192,118	197,330
32	192,173	7,262	187,289	192,135	197,058
36	192,358	6,729	187,871	192,306	196,823
40	192,725	6,460	188,330	192,779	197,057

Table 6 Statistics of corrected CDFs of lower bounds using BPM

n_s	μ, mean (cycles)	σ, stand. dev. (cycles)	25th prctl. (cycles)	50th prctl. (cycles)	75th prctl. (cycles)
8	177,050	14,492	167,226	177,130	187,227
12	180,500	11,876	172,497	180,477	188,569
16	183,014	10,300	176,211	183,066	189,826
20	185,267	9,226	179,124	185,446	191,705
24	186,442	8,304	180,872	186,485	191,997
28	187,659	7,684	182,542	187,776	192,871
32	188,426	7,121	183,637	188,388	193,216
36	189,242	6,620	184,828	189,190	193,635
40	189,853	6,364	185,524	189,906	194,121

Table 7 Statistics of uncorrected CDFs of lower bounds using NA

n_s	μ, mean (cycles)	σ, stand. dev. (cycles)	25th prctl. (cycles)	50th prctl. (cycles)	75th prctl. (cycles)
8	178,935	18,818	166,556	179,450	192,009
12	183,202	14,095	173,843	183,306	192,711
16	185,408	11,742	177,597	185,508	193,291
20	186,849	10,228	180,057	187,118	193,964
24	187,807	9,040	181,751	187,870	193,876
28	189,029	8,268	183,474	189,137	194,634
32	189,596	7,585	184,473	189,547	194,754
36	190,067	6,995	185,409	189,994	194,721
40	190,682	6,689	186,159	190,753	195,168

Table 8 Statistics of corrected CDFs of lower bounds using NA

n_s	μ, mean (cycles)	σ, stand. dev. (cycles)	25th prctl. (cycles)	50th prctl. (cycles)	75th prctl. (cycles)
8	171,724	18,059	159,844	172,218	184,271
12	177,779	13,678	168,697	177,880	187,006
16	181,329	11,484	173,690	181,427	189,038
20	183,953	10,070	177,266	184,218	190,957
24	185,366	8,922	179,388	185,428	191,356
28	186,874	8,174	181,382	186,981	192,415
32	187,795	7,513	182,720	187,747	192,904
36	188,698	6,945	184,074	188,626	193,319
40	189,462	6,646	184,968	189,532	193,919

Table 9 Statistics of uncorrected CDFs of lower bounds using BCAP

n_s	μ, mean (cycles)	σ, stand. dev. (cycles)	25th prctl. (cycles)	50th prctl. (cycles)	75th prctl. (cycles)
8	184,668	17,035	173,098	184,846	196,599
12	184,922	13,727	175,673	185,115	194,154
16	185,846	11,760	178,112	186,117	193,683
20	186,750	10,386	179,951	186,965	194,096
24	187,443	9,247	181,302	187,560	193,749
28	188,536	8,516	182,879	188,769	194,333
32	189,031	7,804	183,809	189,061	194,336
36	189,437	7,210	184,689	189,442	194,304
40	190,065	6,913	185,460	190,197	194,742

Table 10 Statistics of corrected CDFs of lower bounds using BCAP

n_s	μ, mean (cycles)	σ, stand. dev. (cycles)	25th prctl. (cycles)	50th prctl. (cycles)	75th prctl. (cycles)
8	174,345	16,083	163,422	174,513	185,609
12	178,320	13,237	169,401	178,507	187,223
16	181,404	11,479	173,855	181,669	189,054
20	183,762	10,220	177,072	183,974	190,990
24	185,231	9,138	179,163	185,347	191,463
28	186,595	8,429	180,996	186,825	192,331
32	187,670	7,748	182,485	187,700	192,937
36	188,357	7,169	183,636	188,362	193,197
40	189,191	6,881	184,607	189,322	193,846

References

[1] C. Annis, "Probabilistic Life Prediction Isn't as Easy as It Looks," *ASTM STP-1450,* vol. 1, no. 2, 2004.

[2] D. A. Virkler, B. M. Hillberry and P. K. Goel, "The Statistical Nature of Fatigue Crack Propagation," AFFDL-TR-78-43, Air Force Flight Dynamics Laboratory, April, 1978.