Article

Reduced allowable strength of composite laminate for unknown distribution due to limited tests

Yiming Zhang[1], Nam H Kim[1] ⓘ**, Upul R Palliyaguru[2],
Jaco F Schutte[3] and Raphael T Haftka[1]**

Abstract

In design under uncertainty, random distributions are often determined by expensive sampling tests. A key question is whether to invest in more samples or live with a reduced performance by fewer samples due to large uncertainty. The question is particularly difficult to answer when the type of distribution is unknown. This paper investigates the tradeoff between performance and conservativeness in estimating B-basis allowables, using experiments on composite plates with holes. Two approaches that do not require a distribution type are examined: (1) bootstrap confidence intervals and (2) Hanson-Koopmans non-parametric method. Based on the study, it is found that the Hanson-Koopmans method was more conservative than the bootstrap method because the latter penalized allowables for small-size samples. For a small number of samples (less than 29), conservative estimations are preferred over accuracy to account for the large uncertainty. Based on this observation, the bootstrap-assisted Hanson-Koopmans method is proposed to enhance the conservativeness. For the tested cases, the performance penalty using the bootstrap-assisted Hanson-Koopmans method for a small number of samples is found to be substantial.

Introduction

In design under uncertainty, the distributions of random variables are often determined by expensive samples. The uncertainty related to the material properties of a composite laminate is significant and has to be modeled properly for reliable designs.[1] The manufacturing and testing process could be expensive and time-consuming which makes the design process more challenging. One remedy for limited tests is to introduce analytical models/simulations.[2,3] Vallmajó et al.[3] proposed an uncertainty quantification and management (UQ&M) framework based on analytical models to compute the B-basis design allowables of notched configurations. However, the analytical models are not always available, and this paper focuses on experimental tests only. With limited tests, a key question is whether to invest in more samples or live with the reduced performance by fewer samples due to large uncertainty. Identification of the distribution/uncertainty forms for statistical analysis would be also challenging with few samples.[4] Conservative designs are

preferred when uncertainty exists especially for safety-critical applications such as helicopter rotor blades.[5] However, the reduction in performance can be substantial because sampling uncertainty is epistemic uncertainty, which is usually treated more conservatively than aleatory uncertainty. Consequently, there has been substantial recent interest in the tradeoff between reducing sampling uncertainty and improving performance. Picheny et al.[6] studied the influence of sample sizes and target probability of failure on the conservative estimate. Bae et al.[7,8] showed a tradeoff between making design conservative and using more samples to

---

[1]Department of Mechanical and Aerospace Engineering, University of Florida, USA
[2]National Institute for Aviation Research, Wichita State University, USA
[3]Materials Sciences Corporation, USA

Corresponding author:
Nam H Kim, Department of Mechanical and Aerospace Engineering, University of Florida, PO Box 116250, Gainesville, FL 32611-6250, USA.
Email: nkim@ufl.edu

reduce sampling uncertainty. However, they assumed that the samples are normally distributed.

Establishing the tradeoff between the increased sample size and reduced performance is more challenging when the type of distribution is unknown. This paper presents a lesson learned from the regulations governing the design of composite materials used for aircraft structures. Composite materials are widely used for the design of various mechanical systems due to the outstanding capability to be tailored to specific load paths and conditions, resulting in weight efficient designs. For example, more than 50% of the Boeing 787 airframe and the Airbus A350XWB are made of carbon fiber composites.[9]

To determine the conservative strength of composite plates, multiple coupons at the same configuration (design) are tested. Design allowables (e.g. A- or B-basis allowables[10,11]), which are also termed as tolerance intervals/limits (TI), are usually obtained from the confidence interval of a low percentile of measured strengths. Various methods for computing the TI of composite material are well documented in the literature.[12,13] The calculation of TI depends on the type of distributions (e.g. normal, lognormal, and Weibull). Young[11] summarized the statistical approaches for calculation of TI for discrete and continuous distributions. When there is no clear indication of a specific type of distribution, MIL-HDBK-17-1F[14] recommends the non-parametric method such as the Hanson-Koopmans (HK) method.[15] The non-parametric method compensates for the lack of knowledge of the distribution by increasing conservativeness.

Conservative allowables imply more weight, but Bhachu[16] found that the non-parametric approach is efficient in achieving conservativeness without excessive increase in weight, compared to methods based on assumed distribution. Still, the HK method can fail when the sample size (number of the replicates) is smaller than a critical value, which varies with distributions.[15] Besides the HK method, bootstrap confidence bounds have been used to infer design allowables without specifying a statistical distribution. Cross et al.[17] estimated the confidence intervals of the crack growth model using bootstrap confidence bound. Bigerelle et al.[18,19] quantified the uncertainty in Paris law material constant using the bootstrap. Bhachu et al.[16,20] compared several common approaches for fatigue crack growth problems. Romero et al.[21,22] tested the performance of the TI method, kernel density method, Johnson method, and non-parametric method.

The TI approaches are based on rigorous mathematic assumptions and work well with high variability. However, with too small number of samples, it is difficul to identify the distribution type to apply TI. The non-parametric B-basis approach is invalid when the sample size is smaller than a critical value. Besides the effort to develop statistical tools for characterizing TI, extensive experimental studies have been reported to understand the variation of material properties. The world-wide failure exercises[23,24] provided experimental data and benchmarks for failure criteria of composites. The Laminate Variability Method[25] was proposed to incorporate the material properties at a lamina level to mitigate the adverse effects of limited numbers of test coupons while computing B-basis values. Carlsson et al.[26] and the Composite Materials Handbook-17 (CMH-17)[10] provided an in-depth guideline for systematic experimental analysis.

The B-basis allowables are often estimated from experimental test results on limited samples (e.g. less than 29 samples). This paper investigates the estimation of B-basis allowables, a typical TI, from limited samples using test results obtained from the composite laminates with a hole.[27] The tests were performed at eight configurations with 18 samples per configuration. The eight configurations are selected by changing two design variables: the size of the hole and the fraction of 45-degree plies in the laminates. Experimental data were collected on open-hole-tension (OHT) strength tests[28,29] following ASTM standard[27] for this study.

Two issues are investigated in this paper: (1) Are the B-basis allowables estimated from small-size samples as reliable as those from large-size samples? (2) What is the weight penalty when using the design allowables from a limited number of samples? In order to address these two issues, two approaches are evaluated in this paper: (1) bootstrap confidence intervals and (2) HK method. The former does not assume an underlying probability distribution, while the latter assumes log-concave CDF class, which is good for the distribution of composite failure strength.

One challenge for predicting B-basis allowables is balancing conservativeness (for a safe design) versus performance (weight penalty). Conservative prediction is a necessary requirement for certification by regulatory bodies, like the Federal Aviation Administration. B-basis allowables are sensitive to distribution form and sample size, which complicates the calculation. This paper utilizes a partial set of samples (out of 18 samples per configuration) to estimate B-basis allowables, from which the conservativeness and weight penalty are calculated. This paper also explores the usefulness of combining the non-parametric estimation with bootstrapping to account for the unidentified uncertainty and ensure better conservativeness.

The paper is structured as follows: The following section discusses the experiments of the benchmark OHT tests and statistics of the experimental results. Then, the estimation of B-basis at a given configuration with samples is detailed. 'Estimating B-basis allowables

of the OHT tests' section evaluates the B-basis estimations using the experimental results. In the penultimate section, the bootstrap-assisted HK is proposed to enhance the conservativeness from limited samples. In conclusion, we summarize the major outcomes and future work on the estimation of design allowables.

OHT tests

Experiments

OHT test[30,31] is a benchmark test to investigate the effect of an unfilled hole on the tensile failure strength. In this paper, it is used to investigate the approaches to estimate B-basis allowables. The test specimen geometry is shown in Figure 1. The composite laminates are made from MTM45-1 PWC2 3K PW G30-500 fabric prepregs, and the tests were performed according to ASTM D5766.[27] The width of the specimen and the diameter of the hole were denoted by w and D, respectively.

Two design parameters varied in the tests: the ratio w/D and layups measured by the fraction of $\pm 45°$ plies (R_{45}). Eight configurations are examined in the 2D variable space according to Table 1. Table 2 details the test matrix for different w/D. Table 3 lists the test matrix for different layups quantified by the fraction of $\pm 45°$ plies. Each configuration was composed of three prepreg batches, with each batch containing six samples.

Statistics of the experimental strengths

The test results are first examined using boxplot (see Figure 2) and statistics (see Table 4). The means of strengths vary between 37 and 100 ksi in different configurations. For a fixed R_{45} fraction, the strength gradually increases with w/D ratio. For a fixed w/D ratio, the strength gradually decreases with R_{45} fraction. The R_{45} fraction has a more significant impact on the strength than the w/D ratio. Configurations 2 and 7 have the smallest strength and variation ($R_{45} = 0.8$). The standard deviation (SD) varies substantially from 1.33 to 5.41 (5.41/ 1.33 = 4.06), whereas the coefficient of variation (CoV) varies between 0.033 and 0.059 (0.059/0.033 = 1.78). The variation of CoV reduces noticeably compared with SD, which indicates that SD is highly correlated with the mean. The strengths of some samples are far from the rest, such as the lowest sample at configuration 4 or the highest sample at configuration 5. The outliers of experiments have a significant impact on the statistics.[32] The maximum normed residual test[33] was recommended by CMH-17[10] to detect outliers. However, no outlier was identified for the OHT tests. Details of the maximum normed residual test are included in Appendix A.1.

Estimation of B-basis heavily depends on the type of distributions. Figure 3 shows the histograms of the

gage middle
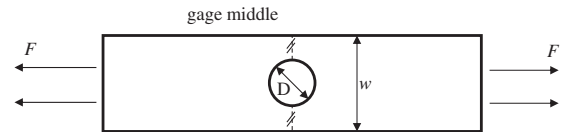
Figure 1. Open-hole tension test specimen configuration and observed failure mode.

Table 1. The test configurations in two design variables.

Points	1	2	3	4	5	6	7	8
(w/D,R_{45})	(3,0.2)	(3,0.8)	(4,0.2)	(4,0.5)	(6,0.2)	(6,0.5)	(6,0.8)	(8,0.2)

Table 2. w/D test matrix at a given R45.

| w (in.)	D (in.)	w/D	Batches/ replicates	No. of specimens
0.75	0.250	3	3 × 6	18
1.00	0.250	4	3 × 6	18
1.50	0.250	6	3 × 6	18
2.00	0.250	8	3 × 6	18

samples at different configurations. Configuration 2 showed a pattern of Weibull distribution with a heavy tail, configurations 1 and 7 follows a bimodal, and configuration 6 is close to a uniform distribution. The Kolmogorov–Smirnov (KS) test[12] was used to quantify which continuous distribution is the best fit. The KS test is a non-parametric test to quantify the goodness-of-fit between a given probability distribution and the empirical distribution of samples. The OHT samples are tested for a normal, uniform, and Weibull distributions, but there is no single distribution that fits all samples the best. Details of the KS tests are provided in Appendix A.2. The normal distribution fits best for four configurations, uniform distribution fits best at a configuration, and Weibull distribution fits best at three configurations. Based on the histograms and KS tests, the non-parametric approach was selected for calculating B-basis allowables.

Estimating B-basis allowables for unknown distributions

A design allowable is determined such that it is less than a large portion of the population with a high level of confidence. The B-basis allowable is a bound that is less than 90% of the population with 95% confidence, as shown in Figure 4. Two mainstream methods are examined for calculating B-basis allowables from unknown distributions.

Table 3. Layup test matrix at a given *w/D*.

R_{45}	Layup % 0°/45°/90°	Ply stacking sequence	Batches/replicates	No. of specimens
0.2	40/20/40	[0/90/0/90/45/-45/90/0/90/0]$_s$	3×6	18
0.5	25/50/25	[(45/0/-45/90)$_2$]$_s$	3×6	18
0.8	10/80/10	[45/-45/90/45/-45/45/-45/0/45/-45]$_s$	3×6	18
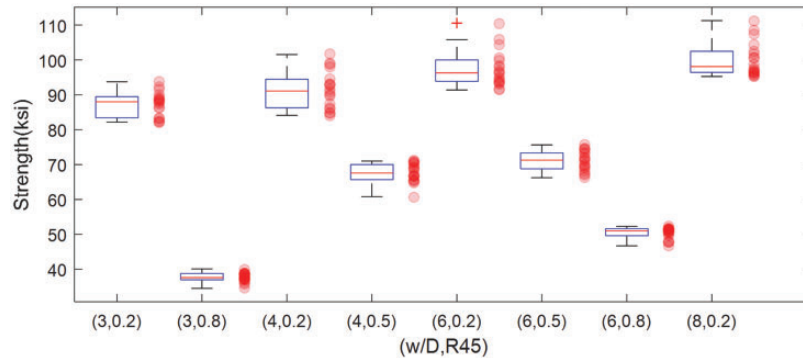
Figure 2. Boxplots of tensile strength using 18 samples at eight different configurations.

Table 4. Statistics of tensile strength using 18 samples at eight different configurations.

Configuration (*w/D*,R_{45})		1 (3,0.2)	2 (3,0.8)	3 (4,0.2)	4 (4,0.5)	5 (6,0.2)	6 (6,0.5)	7 (6,0.8)	8 (8,0.2)
Tensile strength (ksi)	Mean	87.35	37.68	91.45	67.62	97.43	71.06	50.29	100.11
	SD	3.54	1.33	5.41	2.70	5.19	2.84	1.67	4.90
	CoV	0.040	0.035	0.059	0.039	0.053	0.040	0.033	0.048

A non-parametric approach using order statistics

CMH-17[10] recommended a non-parametric approach for calculating B-basis allowables when samples do not demonstrate a clear distribution pattern (e.g. due to a limited number of samples). The non-parametric approach is based on the order statistics and varies with sample size. When more than 28 samples are available, the B-basis value is the *r*th lowest sample, where *r* varies with sample size. For example, *r* equals to one when 30 samples are available. The HK method is suggested for non-parametric estimation of B-basis allowables with less than 29 samples. We adopt the HK method for the non-parametric estimation (B_{HK}).

First, strength samples are ordered by magnitude; $x_{(1)}$ is the lowest strength, and $x_{(r)}$ is the *r*th lowest sample. Then, B-basis is determined by

$$B_{HK} = x_{(r)} \left[\frac{x_{(1)}}{x_{(r)}} \right]^k \qquad (1)$$

where *k* is a factor depending on sample size. The parameters *r* and *k* are found in Table 5.[10]

B-basis allowables using bootstrap confidence intervals

Bootstrapping[34,35] is a data-driven approach for statistical inference and commonly used for estimating bias and variance of a critical statistic. Various methods have been established for finding confidence bounds from bootstrap sampling distributions. The percentile method determines the confidence interval for the statistic of interest (e.g. 10th percentile) from bootstrapped distributions. The bias correction approach modifies the estimated statistic of interest by a bias to account for the small-size samples. The bias-corrected accelerated method corrects for both bias and skewness in the distribution of bootstrap estimates. Picheny et al.[6] incorporated bootstrapping for reliability analysis of a system response. Edwards et al.[36] proposed an approach to estimate the lower percentiles of material properties using bootstrapping. Lee et al.[37,38] evaluated
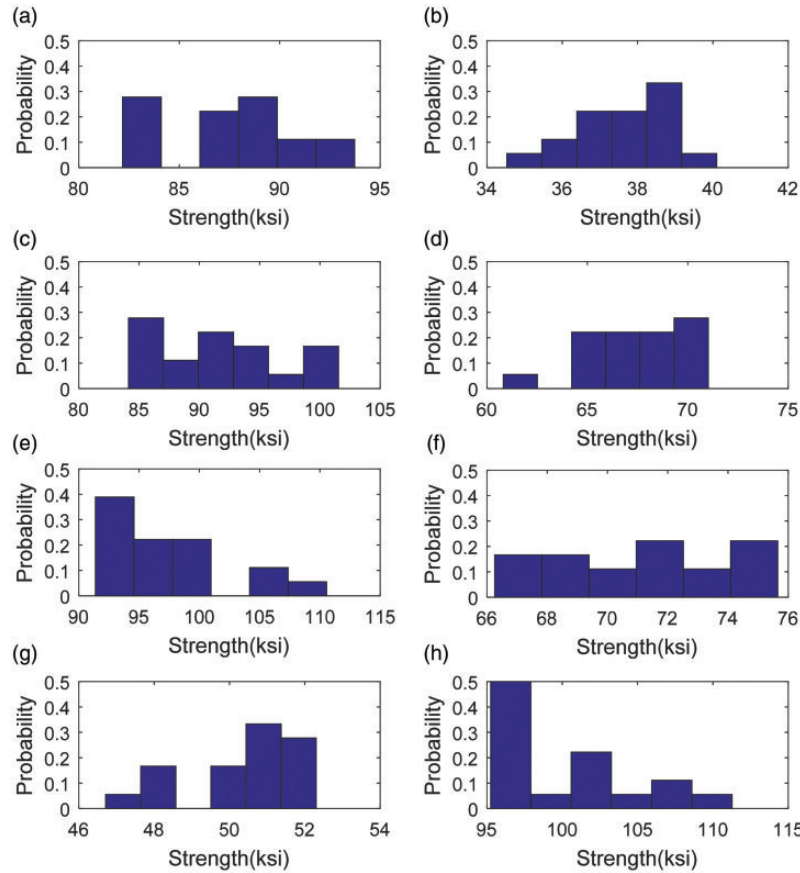
Figure 3. Histograms of samples at different configurations.(a) Configuration 1, (b) Configuration 2, (c) Configuration 3, (d) Configuration 4, (e) Configuration 5, (f) Configuration 6, (g) Configuration 7 and (h) Configuration 8.
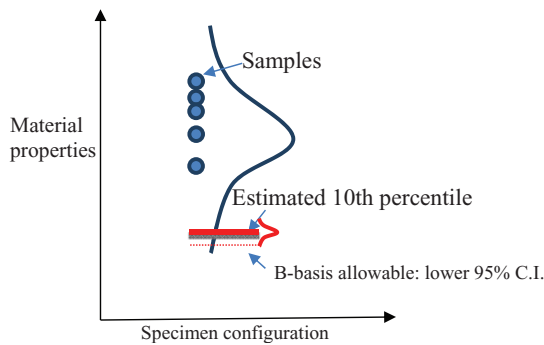
Figure 4. Illustration of B-basis allowables based on samples.

Table 5. Parameters for the non-parametric B-basis allowables with different number of samples.

Number of samples	6	8	10	18
r	5	6	6	9
K	3.064	2.382	2.137	1.354

and improved the accurate coverage of the bootstrap confidence interval.

A typical bootstrapping procedure resamples (with replacement) from the available sample set without assuming any probability model. The distribution is approximated by the population of existing samples. Samples with the same size are drawn (with replacement) from the population multiple times. The existing samples are resampled with an equal chance of being

selected. Then, a sample statistic (e.g. mean or variance) is calculated from each bootstrapped sample. An empirical distribution of the statistic is obtained to define the uncertainty. The percentile for small-size samples is computed according to Langford.[39]

We adopted the bias-corrected accelerated method which corrects the B-basis estimation for both bias and skewness (B_{boot}). The one-sided lower confidence bound is determined by

$$B_{\text{boot}} = \hat{G}^{-1}\left(\Phi\left(z_0 + \frac{z_0 + z_\alpha}{1 - A(z_0 + z_\alpha)}\right)\right) \qquad (2)$$

where $\hat{G}^{-1}(\cdot)$ is the inverse empirical CDF of the bootstrap sampling distribution, z_α is the z-score from the

standard normal distribution, z_0 is the bias correction, A is the acceleration parameter, and $\Phi(\cdot)$ is the standard normal CDF. A detailed procedure for determining the B_{boot} is described in DiCiccio and Efron.[40] The bias correction z_0 is used to modify the confidence bounds and to account for a limited number of samples. A is the adjustment to correct for the accelerating standard error. We used the "bootci" function in MATLAB to predict B_{boot}.

Estimating B-basis allowables of the OHT tests

Test plan to evaluate the estimations of B-basis allowables

We resort to the experimental results to estimate B-basis allowables with different sample sizes. The resampling scheme[31,41] is adopted to make the most use of experiments and investigate the B-basis estimation with varying sample size. The lower 10th percentile of $B^{(18)}$ is used as the baseline for comparison. Out of the full set of 18 samples at each configuration, partial samples are selected (without replacement) to examine the performance of estimated B-basis allowables. The B-basis allowables with k partial samples are denoted as $B^{(k)}$, where $k = 6, 8, 10$. Partial sampling is repeated $N_R = 1000$ times to calculate their statistics. The empirical distributions from N_R sets of $B_{HK}^{(k)}$ and $B_{boot}^{(k)}$ are used to evaluate the mean and variance of estimated B-basis allowables.

Two metrics are used to evaluate the margin (weight penalty) and conservativeness. The weight penalty is based on the assumption that the weight of the laminate is inversely proportional to the load it carries. Therefore, it is defined as the relative difference between the baseline and estimated B-basis. Since there are N_R weight penalties, the weight penalty factor (WPF) is defined as the mean of them, as

$$\text{WPF}^{(k)} = \frac{1}{N_R} \sum_{j=1}^{N_R} \left(\frac{B^{(18)} - B_j^{(k)}}{B^{(18)}} \right) \quad (3)$$

where j is the index of the resampled set. The positive WPF indicates how much the estimate is conservative on average.

The estimated B-basis allowables $B^{(k)}$ from partial samples is conservative if it is smaller than the baseline $B^{(18)}$. The conservativeness fraction (CF) is the fraction of conservative $B^{(k)}$ from N_R sets of partial resamples

$$\text{CF}^{(k)} = \frac{1}{N_R} \sum_{j=1}^{N_R} H \left(\frac{B^{(18)} - B_j^{(k)}}{B^{(18)}} \right) \quad (4)$$

where $H(x < 0) = 0$ and $H(x \geq 0) = 1$. CF $= 1$ means that all N_R sets are conservative. Note that a good method will have a high conservative fraction with a low weight penalty factor. CF and WPF are calculated using $B_{HK}^{(k)}$ and $B_{boot}^{(k)}$.

Estimation of B-basis allowables from experimental replicates

We first compared the B-basis allowables estimated from $B_{HK}^{(k)}$ and $B_{boot}^{(k)}$. The baseline B-basis allowables from all 18 samples are summarized in Table 6. $B_{HK}^{(18)}$ was more conservative than the $B_{boot}^{(18)}$ at all configurations. The differences between $B_{HK}^{(18)}$ and $B_{boot}^{(18)}$ varied within [0.78, 2.19] ksi.

The B-basis allowables estimated from $N_R = 1000$ sets of partial resamples are shown as boxplots in Figure 5. The mean values of $B_{HK}^{(k)}$ were conservative compared to that of $B_{boot}^{(k)}$ by 4%–20%. Another interesting observation was that the mean values of $B_{boot}^{(k)}$ remained almost the same, whereas that of $B_{HK}^{(k)}$ increased gradually by a few percents as k increases.

The performance of $B_{boot}^{(k)}$ and $B_{HK}^{(k)}$ with different sample sizes is further compared using the normalized mean value of the B-basis estimations from N_R sets of the resampled dataset, as

$$NM^{(k)} = \frac{\text{mean}(\mathbf{S}) - \text{mean}(B^{(k)})}{\text{std}(\mathbf{S})} \quad (5)$$

where mean($\mathbf{S}$) and std($\mathbf{S}$) stand for the mean and standard deviation of 18 full-set strength samples at a given configuration, and mean($B^{(k)}$) is the mean of either $B_{boot}^{(k)}$ or $B_{HK}^{(k)}$ from 1000 resample sets. $NM^{(k)}$ is essentially a measure of B-basis allowables scaled by standard deviations below the population mean values, which is summarized in Table 7. $NM_{HK}^{(k)}$ varied within [1.82, 5.36] and was more conservative than $NM_{boot}^{(k)}$, which was within [0.91, 1.84]. With an increasing number of samples, $NM_{boot}^{(k)}$ increased gradually and $B_{boot}^{(k)}$ became more conservative. In contrast, $NM_{HK}^{(k)}$ decreased noticeably with increasing samples as $B_{HK}^{(k)}$ increased.

The tradeoff between WPF and CF for the estimated B-basis allowables was examined next. The means of WPF and CF from the 1000 sets of resamples are visualized in Figure 6. It is noticeable that for $B_{boot}^{(k)}$, WPFs were non-positive at most configurations and CFs were less than 0.55. $B_{boot}^{(k)}$ became more conservative with an increasing number of samples. For $B_{HK}^{(k)}$, in contrast, WPFs were all positive between 0.07 and 0.18 for six samples and decreased with an increasing number of

Table 6. Estimated baseline B-basis allowables (unit: ksi).

Configuration	1	2	3	4	5	6	7	8
$(w/D, R_{45})$	(3,0.2)	(3,0.8)	(4,0.2)	(4,0.5)	(6,0.2)	(6,0.5)	(6,0.8)	(8,0.2)
$B_{\text{boot}}^{(18)}$	82.17	34.53	84.12	60.81	91.38	66.26	46.71	95.25
$B_{\text{HK}}^{(18)}$	80.30	33.54	81.93	58.8	89.79	64.58	45.28	94.47
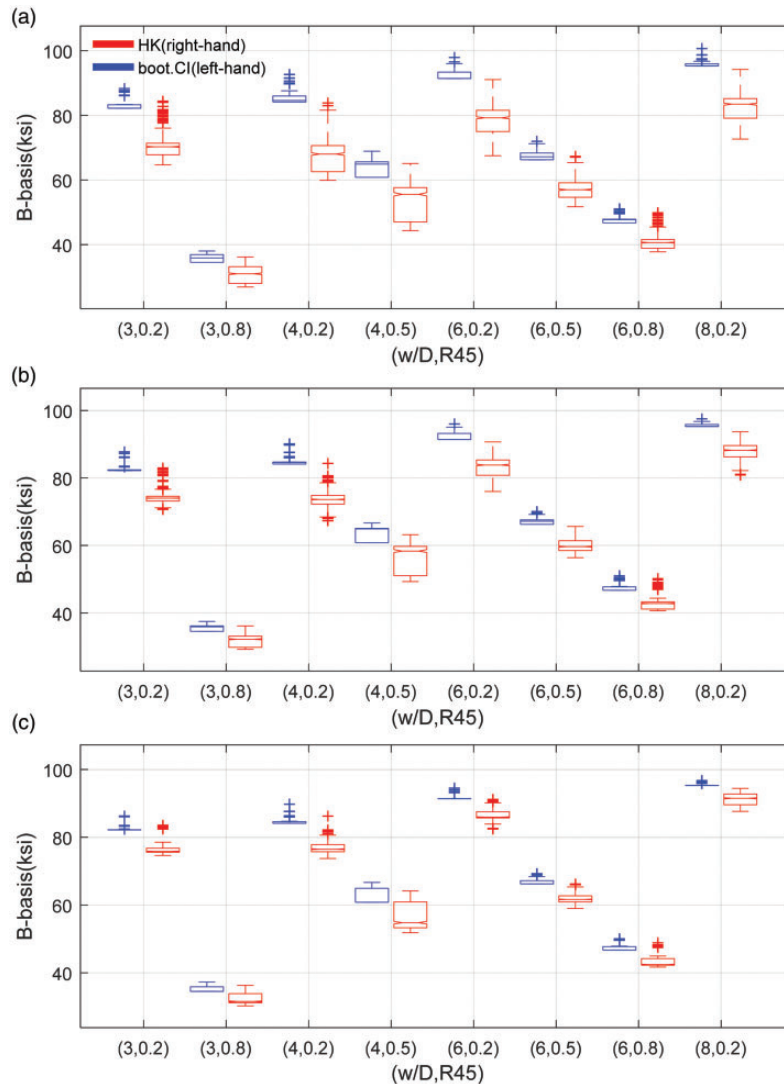
Figure 5. Estimated B-basis allowables from the 1000 sets of resamples. B-basis allowables are estimated from B_{HK} and B_{boot}. (a) $k = 6$ samples, (b) $k = 8$ samples and (c) $k = 10$ samples.

samples. It means that $B_{\text{HK}}^{(k)}$ penalized the allowables for small-size samples by increasing WPF. CFs were much better than that of $B_{\text{boot}}^{(k)}$. The overall trend of CF decreased slightly with an increasing number of samples.

$B_{\text{HK}}^{(k)}$ at configurations 2 and 4 were of interest, as they have the smallest CF. Empirical cumulative distribution function (ECDF) of $B_{\text{HK}}^{(k)}$ for the two configurations are summarized in Figure 7, which is basically

the potential price to pay because of using fewer samples to estimate a B-basis. In the case of configuration 2, for example, with $k = 6$, it was very likely to have a B-basis less than 30 ksi, but with $k = 10$, it was nearly impossible for a B-basis to be less than 30 ksi. For configuration 4, a jump is observed around ECDF $= 0.55$, which is more pronounced when using 10 samples. This jump is due to the close-to-outlier sample at

**Table 7.** Normalized mean values of the B-basis estimations from 1000 sets of the resampled dataset.

Configuration		1	2	3	4	5	6	7	8
$(w/D, R_{45})$		(3,0.2)	(3,0.8)	(4,0.2)	(4,0.5)	(6,0.2)	(6,0.5)	(6,0.8)	(8,0.2)
$B_{boot}^{(k)}$	$k=6$	1.28	1.41	1.17	1.39	0.98	1.28	1.46	0.91
	$k=8$	1.38	1.64	1.25	1.61	1.04	1.4	1.68	0.95
	$k=10$	1.42	1.84	1.29	1.83	1.09	1.5	1.83	0.97
$B_{HK}^{(k)}$	$k=6$	4.83	5.21	4.38	5.19	3.67	4.97	5.36	3.59
	$k=8$	3.74	4.41	3.33	4.33	2.68	3.92	4.56	2.52
	$k=10$	3.14	3.93	2.69	3.91	2.11	3.28	4.21	1.82

Note: Mean values are subtracted from the mean strength of the 18 samples and then divided by the standard deviation of the 18 samples.
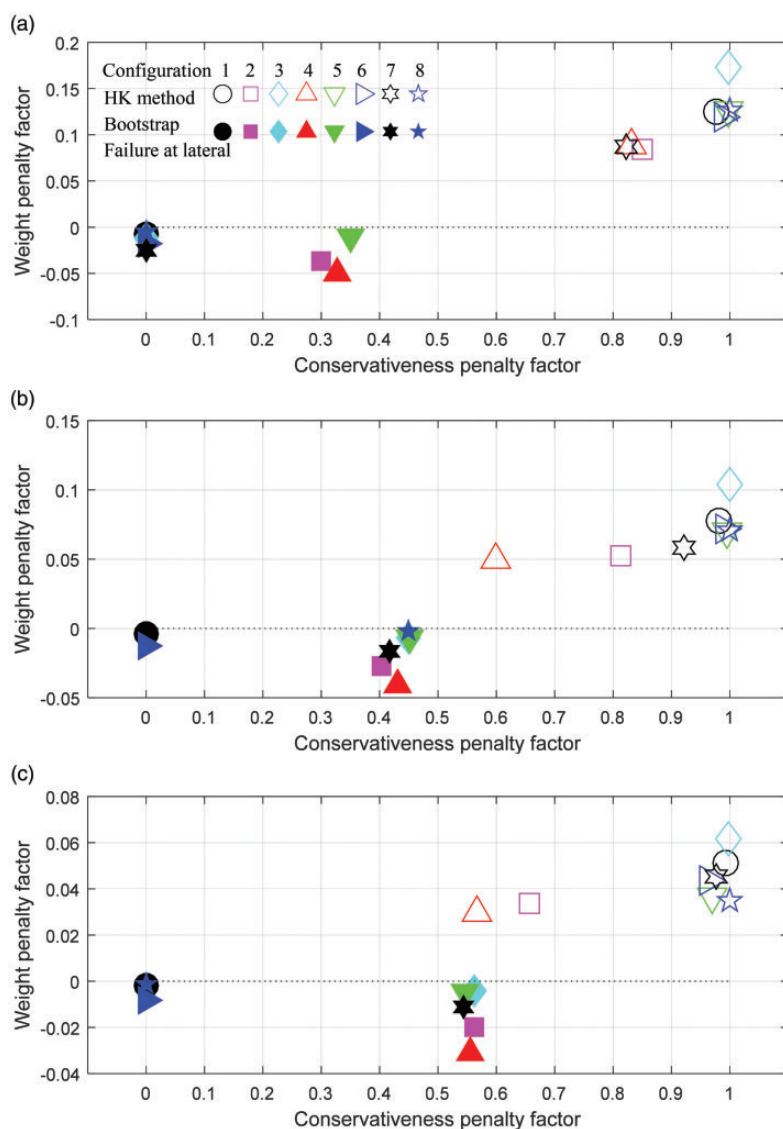
Figure 6. Weigh penalty factor versus conservativeness fraction from the 1000 sets of resamples. Each set of selected data has 6,8, or 10 samples. B_{HK} and B_{boot} are adopted for B-basis estimations. (a) Estimation from 6 samples, (b) Estimation from 8 samples and (c) Estimation from 10 samples.
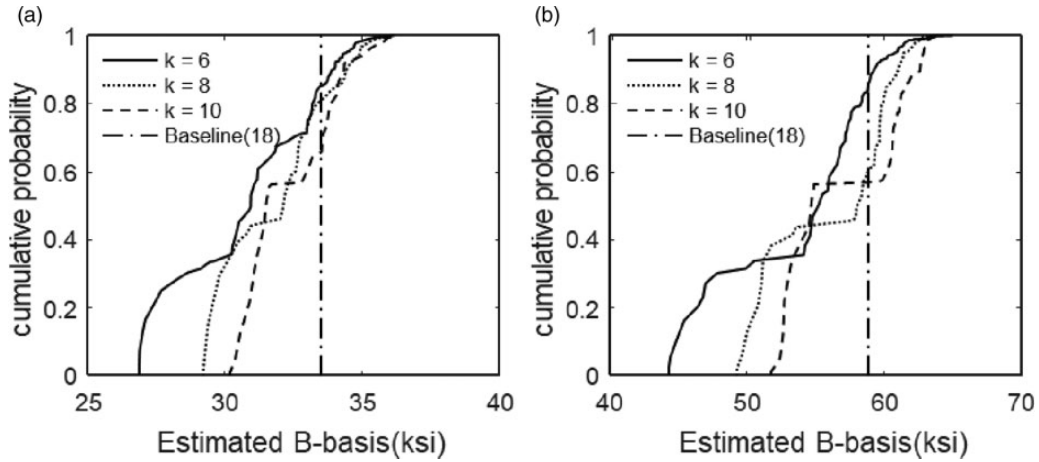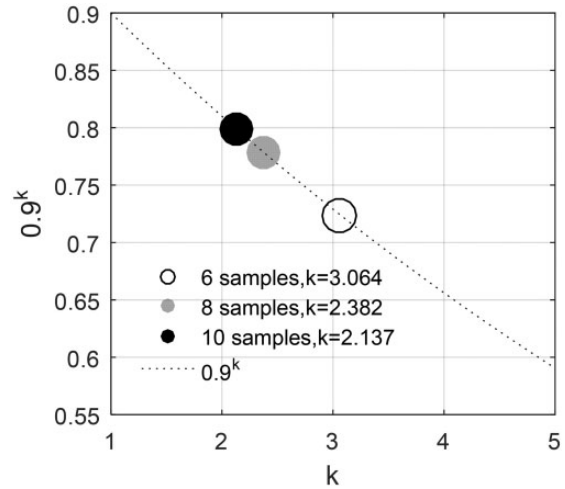
Figure 7. Empirical cumulative distribution of the estimated B-basis allowables using Hanson-Koopmans method: (a) at configuration 2 and (b) at configuration 4.

configuration 4, which dominates the estimation of B-basis allowables. In Figure 2, we can see a close-to-outlier sample far from the other samples, which results in the jump. When compared to the close-to-outlier sample at configurations 3 and 5, the sample at configuration 4 is at the lower end of the strength values and leads to undesirable conservativeness.

A further study is performed to investigate the trend of decreasing CF with an increasing number of samples for $B_{HK}^{(k)}$. $B_{HK}^{(k)}$ strongly depends on the parameters r and k as shown in equation (1). The parameters for estimating $B_{HK}^{(k)}$ are given in Table 5 with different numbers of samples. The magnitude of $x_{(r)}$ might increase or decrease and does not show a clear trend. If we assume the $x_{(r)}$ and $x_{(1)}$ remain the same, $B_{HK}^{(k)}$ increases with decreasing k. By assuming the ratio $x_{(1)}/x_{(r)} = 0.9$, an exponential function with a 0.9 base is plotted in Figure 8. The estimations using 8 and 10 samples are close to each other compared with that of six samples. As expected, $B_{HK}^{(k)}$ increased and became less conservative with an increasing number of samples. With increasing samples, $x_{(1)}$, $x_{(r)}$ and k would be stabilized and $B_{HK}^{(k)}$ is expected to have a reduced variance.

Bootstrap-assisted HK method for small-size samples

Bootstrap-assisted HK Method

The B_{HK} was originally proposed for B-basis estimations using a limited number of samples and proved to be reliable when the sample size is larger than a critical value.[10] However, the critical sample size varies with application and the B_{HK} might not be conservative for limited tests. As shown in 'Estimating B-basis allowables of The OHT tests' section, although B_{HK} provided conservativeness in WPF, CF was less

Figure 8. Plot of an exponential function with a 0.9 base.

desirable at configurations 2 and 4. The estimated B-basis allowables could be improved further to account for the unidentified uncertainties, such as too small sample size, to meet the threshold. We proposed the bootstrap-assisted HK method $B_{boot.HK}$ to make the estimated allowables with small-size samples comparable to those with large-size samples.

A flowchart for calculating $B_{boot.HK}$ is shown in Table 8. Assuming r_c test coupons are available from tests, N_r sets of samples are bootstrapped from the r_c coupons. r_c samples are included for each bootstrapped dataset allowing replacement. The B-basis allowable is calculated for each bootstrapped dataset and results in N_r estimations, from which ECDF is developed. The modified B-basis allowables from $B_{boot.HK}$ is defined using the lower 5th percentile of the ECDF. The number of bootstrapped estimations N_r is chosen

Table 8. Flowchart for the Bootstrap-assisted Hanson-Koopmans method from replicates.

Step	Procedure
1	Obtain failure strength samples from r_c test coupons (data acquisition)
2	Bootstrap N_r sets of samples from r_c coupons.
3	Calculate the B-basis allowables for each bootstrapped dataset using HK method and obtain N_r estimations of B-basis allowables
4	Develop the ECDF of the N_r estimations of B-basis allowables using interpolation
5	Define the modified B-basis allowables using the lower 5th percentile of the ECDF
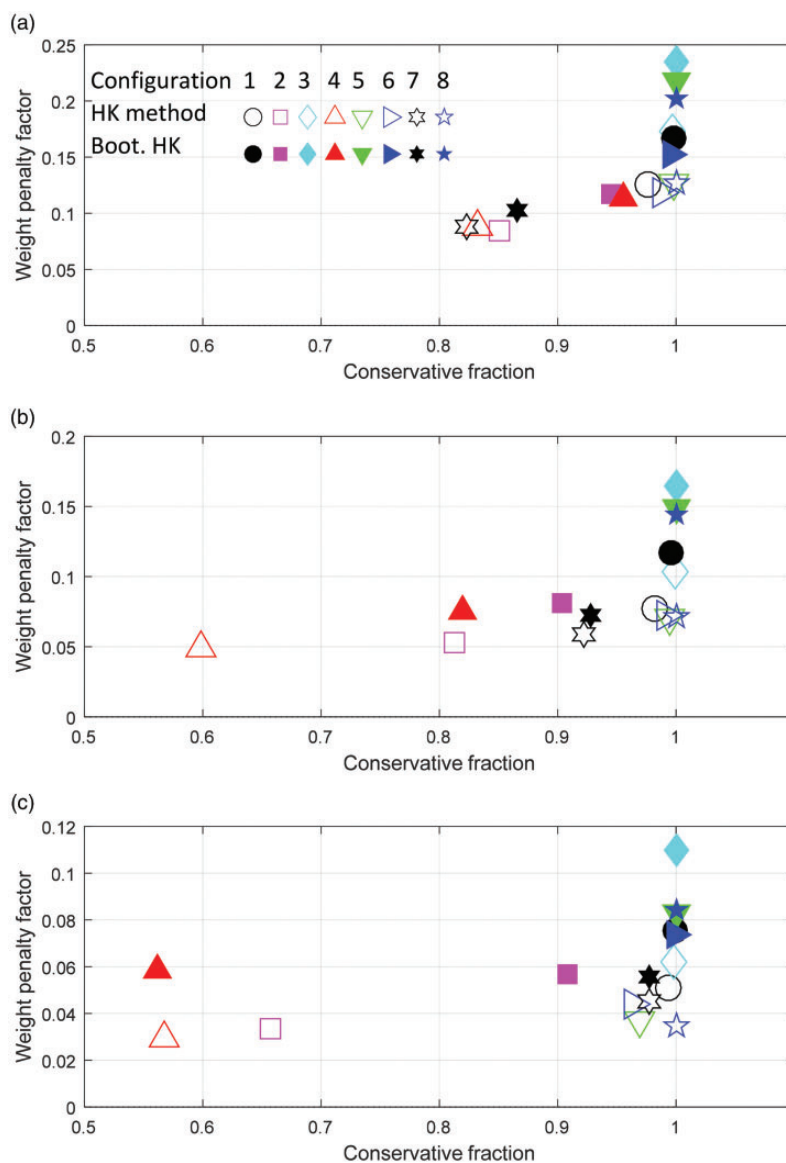
Figure 9. Weight penalty factor versus conservativeness penalty factor from the 1000 sets of replicates. Each set of selected data has 6,8, or 10 replicates. B_{HK} and $B_{boot.HK}$ are adopted for B-basis estimations. (a) Estimation from 6, (b) Estimation from 8 and (c) Estimation from 10.
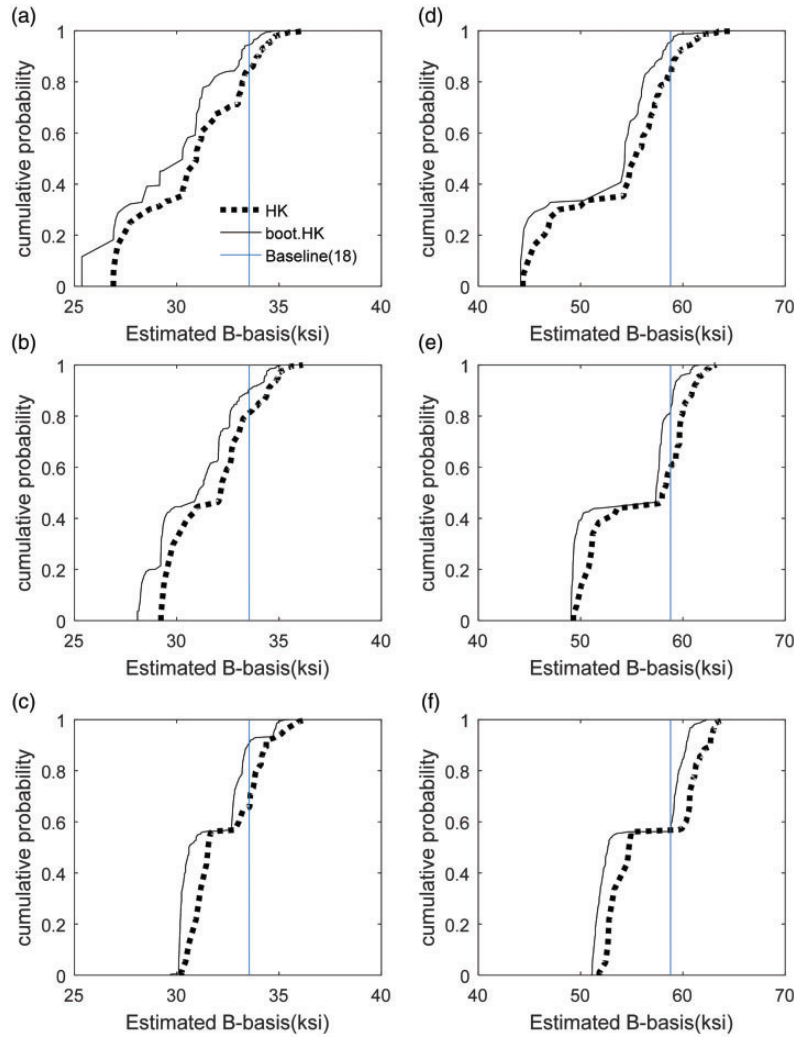
Figure 10. Empirical cumulative distribution of the estimated B-basis allowables at the configuration 2 (a) to (c) and configuration 4 (d) to (f). (a) Configuration 2 with $k = 6$, (b) Configuration 2 with $k = 8$, (c) Configuration 2 with $k = 10$, (d) Configuration 4 with $k = 6$, (e) Configuration 4 with $k = 8$ and (f) Configuration 4 with $k = 10$.

to be large enough, so that it does not affect the quality of the results (relatively stable estimation).

B-basis allowables of OHT tests using the bootstrap-assisted HK method

$B_{boot.HK}$ is compared with the B_{HK} following the test plan in 'Test plan to evaluate the estimations of B-basis allowables' section. The means of the WPF and the CF from the 1000 sets of resamples are visualized in Figure 9. For the estimation from six samples, CF of $B_{HK}^{(6)}$ varies between 0.82 and 1.00. CF of $B_{boot.HK}^{(6)}$ varies between 0.86 and 1.00, which is slightly more conservative. CF at configurations 2 and 4 increased noticeably from 0.85 to 0.95 with increased WPF around 0.05. For the estimation from eight samples, CF of $B_{HK}^{(8)}$ varies between 0.59 and 1.00. The CF of $B_{boot.HK}^{(8)}$ at configuration 4 increased significantly from 0.59 to

0.81. For the estimation of 10 samples, the low CF of $B_{HK}^{(10)}$ was also observed at configurations 2 and 4. The CF of $B_{boot.HK}^{(10)}$ increased significantly at configuration 2 but not at configuration 4. Based on the preliminary check, we found that the $B_{boot.HK}^{(10)}$ improved CF at the configurations with least CF. The exception for the $B^{(10)}$ at configuration 4 is further discussed.

The $B^{(10)}$ at configurations 2 and 4 are of critical interest because they have the least CF. ECDFs of $B^{(10)}$ for the two configurations are summarized in Figure 10. With six samples, ECDF of $B_{boot.HK}^{(6)}$ is conservative than that of $B_{HK}^{(6)}$. In contrast, the ECDFs showed a clear jump with 10 samples. For configuration 4, the jump is around ECDF=0.55. This jump is due to the close-to-outlier sample, which dominates the estimation of B-basis allowables. $B_{boot.HK}^{(k)}$ seemed invalid when a close-to-outlier sample dominates the B-basis estimation in the lower end.

Conclusions

Composite materials have been routinely used in load-bearing structures due to their outstanding capability to be tailored to specific load paths and conditions, resulting in weight efficient designs. Reliable stress limits, called design allowables, are of critical interest to designers to balance safety, performance, and economic value. The properties of composite laminates usually suffer from significant variation due to the complexity and inherent variability of the manufacturing process. Estimating the design allowables is challenging because of complicated failure mechanisms and the limited number of samples.

This paper examines the estimation of B-basis allowables from a limited number of samples demonstrated in OHT strength testing experiments. The conservativeness and weight penalty (margin of B-basis allowables) are evaluated for the bootstrap confidence interval (B_{boot}) and HK (B_{HK}) methods. The experiments are thoroughly investigated by examining the effects of outliers, the goodness-of-fit on different assumed statistical distributions, and data visualization. The B-basis allowable estimation using 18 samples (large-size samples) is used as the baseline. Partial subsets of samples (limited number of samples) are used for evaluating different ways of calculating the B-basis allowables and compared with the baseline.

Based on the study, it was observed that B_{HK} was more conservative than B_{boot}. The former penalized B-basis allowables for small-size samples and incorporated the effect of sample size better than the latter. It was also observed that B_{HK} was sensitive to outliers which dominated the estimations of B-basis allowables. In this paper, the bootstrap-assisted HK method ($B_{boot.HK}$) was proposed to enhance the reliability of B-basis allowables for small-size samples. The proposed method was especially beneficial when only a limited number of samples are available, yielding the least amount of conservativeness among the methods evaluated.

Acknowledgements

Declaration of Conflicting Interests

Funding

ORCID iD

Nam H Kim https://orcid.org/0000-0002-0221-9749

References

1. Toft HS, Branner K, Mishnaevsky L Jr, et al. Uncertainty modelling and code calibration for composite materials. *J Compos Mater* 2013; 47: 1729–1747.
2. Zhang Y, Schutte J, Meeker J, et al. Predicting B-basis allowable at untested points from experiments and simulations of plates with holes. In: *Proceedings of the 12th world congress on structural and multidisciplinary optimization*, Braunschweig, Germany, www.researchgate.net/publication/318909364. (accessed 20 January 2020).
3. Vallmajó O, Cózar I, Furtado C, et al. Virtual calculation of the B-value allowables of notched composite laminates. *Compos Struct* 2019; 212: 11–21.
4. Nilakantan G, Obaid AA, Keefe M, et al. Experimental evaluation and statistical characterization of the strength and strain energy density distribution of Kevlar KM2 yarns: exploring length-scale and weaving effects. *J Compos Mater* 2011; 45: 1749–1769.
5. Satheesh R, Narayana Naik G and Ganguli R. Conservative design optimization of laminated composite structures using genetic algorithms and multiple failure criteria. *J Compos Mater* 2010; 44: 369–387.
6. Picheny V, Kim NH and Haftka RT. Application of bootstrap method in conservative estimation of reliability with limited samples. *Struct Multidisc Optimiz* 2010; 41: 205–217.
7. Bae S, Kim NH and Jang S. Reliability-based design optimization under sampling uncertainty: shifting design versus shaping uncertainty. *Struct Multidisc Optimiz* 2018; 57: 1845–1855.
8. Bae S, Kim NH and Jang S-G. System reliability-based design optimization using trade-off between uncertainty reduction and design shift. *J Mech Des* 2019; 141: 041403.
9. Mrazova M. Advanced composite materials of the future in aerospace industry. *Incas Bull* 2013; 5: 139.
10. Defense US Do. Composite materials handbook, volume 1 – polymer matrix composites guidelines for characterization of structural materials, http://app.knovel.com/hotlink/toc/id:kpMHMILH5M/military-handbook-mil-3/military-handbook-mil-3 (2002, accessed 20 January 2020).
11. Young DS. Tolerance: an R package for estimating tolerance intervals. *J Stat Softw* 2010; 36: 1–39.
12. Chakravarti IM and Laha RG. *Handbook of methods of applied statistics*. New Jersey: John Wiley & Sons, 1967.
13. Rust SW, Todt FR, Harris B, et al. *Statistical methods for calculating material allowables for MIL-HDBK-17. Test methods and design allowables for fibrous composites* (2nd volume). West Conshohocken, PA: ASTM International, 1989.
14. Handbook M. *MIL-HDBK-17-1F: composite materials handbook, volume 1-polymer matrix composites guidelines for characterization of structural materials*. Virginia: US Department of Defense, 2002.

15. Hanson DL and Koopmans L. Tolerance limits for the class of distributions with increasing hazard rates. *Ann Math Stat* 1964; 35: 1561–1570.
16. Bhachu KS. *Effects of measured data uncertainty on damage tolerant design of aircraft structures*. Florida: University of Florida, 2015.
17. Cross RJ, Makeev A and Armanios E. A comparison of predictions from probabilistic crack growth models inferred from Virkler's data. *J ASTM Int* 2006; 3: 1–11.
18. Bigerelle M and Iost A. Bootstrap analysis of FCGR, application to the Paris relationship and to lifetime prediction. *Int J Fatigue* 1999; 21: 299–307.
19. Bigerelle M, Najjar D, Fournier B, et al. Application of lambda distributions and bootstrap analysis to the prediction of fatigue lifetime and confidence intervals. *Int J Fatigue* 2006; 28: 223–236.
20. Bhachu KS, Haftka RT and Kim NH. Comparison of methods for calculating B-basis crack growth life using limited tests. *AIAA J* 2016; 54: 1287–1298.
21. Romero V, Swiler L and Urbina A. An initial comparison of methods for representing and aggregating experimental uncertainties involving sparse data. In: *Proceedings of the 13th AIAA non-deterministic approaches conference*, Denver, CO, 4–7 April 2011.
22. Romero V, Swiler L, Urbina A, et al. A comparison of methods for representing sparsely sampled random quantities. Sandia National Laboratories report SAND2013-4561, September, 2013.
23. Hinton M and Kaddour A. Triaxial test results for fibre-reinforced composites: the second world-wide failure exercise benchmark data. *J Compos Mater* 2013; 47: 653–678.
24. Hinton MJ, Kaddour AS and Soden PD. *Failure criteria in fibre reinforced polymer composites: the world-wide failure exercise*. Amsterdam: Elsevier, 2004.
25. Tomblin J and Seneviratne W. *Laminate statistical allowable generation for fiber-reinforced composite materials: lamina variability method*. USA: Office of Aviation Research and Development, Federal Aviation Administration, 2009.
26. Carlsson LA, Adams DF and Pipes RB. *Experimental characterization of advanced composite materials*. Boca Raton: CRC Press, 2014.
27. ASTM International ASTM D5766/D5766M-11. *Standard test method for open-hole tensile strength of polymer matrix composite laminates*. West Conshohocken, PA: ASTM International, 2011.
28. Hallett SR, Green BG, Jiang W-G, et al. The open hole tensile test: a challenge for virtual testing of composites. *Int J Fract* 2009; 158: 169–181.
29. Tomblin J, Sherraden J, Seneviratne W, et al. *Advanced general aviation transport experiments. A–Basis and B–Basis design allowables for epoxy–based prepreg Toray T700GC-12K-31E/# 2510 unidirectional tape*. Wichita: National Institute for Aviation Research Wichita State University, 2002.
30. Zhang Y, Meeker J, Schutte J, et al. On approaches to combine experimental strength and simulation with application to open-hole-tension configuration. In: *Proceedings of the American Society for composites: thirty-first technical conference*, Williamsburg, VA, 19–22 September 2016.
31. Zhang Y, Haftka RT, Kim N-H, et al. Allocation of samples between exploration and replication for open-hole-tension test. In: *Proceedings of the American Society of Composites-30th Technical Conference*, East Lansing, MI, 28–30 September 2015.
32. Osborne JW and Overbay A. The power of outliers (and why researchers should always check for them). *Pract Assess Res Eval* 2004; 9: 1–12.
33. Stefansky W. Rejecting outliers by maximum normed residual. *Ann Math Stat* 1971; 42: 35–45.
34. Mooney CZ and Duval RD. *Bootstrapping: a nonparametric approach to statistical inference*. Thousand Oaks: Sage, 1993.
35. Efron B. *The jackknife, the bootstrap and other resampling plans*. New Delhi: SIAM, 1982.
36. Edwards DJ, Guess FM and Young TM. Improved estimation of the lower percentiles of material properties. *Wood Sci Technol* 2011; 45: 533–546.
37. Lee SM and Young GA. Asymptotic iterated bootstrap confidence intervals. *Ann Stat* 1995; 23: 1301–1330.
38. Lee SM and Young GA. The effect of Monte Carlo approximation on coverage error of double-bootstrap confidence intervals. *J R Stat Soc: Series B (Statistical Methodology)* 1999; 61: 353–366.
39. Langford E. Quartiles in elementary statistics. *J Stat Educ* 2006; 14: 1–27.
40. DiCiccio TJ and Efron B. Bootstrap confidence intervals. *Stat Sci* 1996; 11: 189–212.
41. Zhang Y, Schutte JF, Seneviratne WP, et al. Sampling by exploration and replication for estimating experimental strength of composite structures. *AIAA J* 2017; 55: 3594–3602.

Appendix 1. Outlier detection and identification of distributions in the open-hole-tension tests

A.1 Outlier detection

The outliers of samples have a significant impact on the statistics.[32] The maximum normed residual (MNR) test[33] was recommended by CMH-17[10] to detect outliers. A sample is identified as an outlier if the absolute deviation from the sample mean is too large. This procedure assumes that nominal samples follow a normal population. The MNR is defined as

$$\text{MNR} = \max_i \frac{|x_i - \text{mean}(\mathbf{x})|}{\text{std}(\mathbf{x})} \qquad (6)$$

where x_i denotes the experimental strength of r_c test coupons. mean($\mathbf{x}$) and std($\mathbf{x}$) are the mean and standard deviation of samples, respectively. The MNR is compared to a critical value depending on the sample size with a specific significance level. The critical MNR for 18 samples is 2.65 using a significance level of 0.05 (95% confidence). The MNR for the OHT tests is

Table 9. Value of the maximum normed residual (MNR) tests for the detection of an outlier.

Points	1	2	3	4	5	6	7	8
$(w/D, R_{45})$	(3,0.2)	(3,0.8)	(4,0.2)	(4,0.5)	(6,0.2)	(6,0.5)	(6,0.8)	(8,0.2)
MNR	1.79	2.35	1.86	2.52	2.52	1.69	2.13	2.27

Note: The critical value to detect the outlier is 2.65.

Table 10. The p-value of the KS tests against Normal, uniform and Weibull distributions.

Points	1	2	3	4	5	6	7	8
$(w/D, R_{45})$	(3,0.2)	(3,0.8)	(4,0.2)	(4,0.5)	(6,0.2)	(6,0.5)	(6,0.8)	(8,0.2)
Normal	0.79	0.94	0.96	0.91	0.60	0.97	0.21	0.39
Uniform	0.63	0.16	0.43	0.02	0.01	0.99	0.01	0.01
Weibull	0.91	0.97	0.83	0.71	0.48	0.87	0.41	0.36
Best fit	Weibull	Weibull	Normal	Normal	Normal	Uniform	Weibull	Normal

provided in Table 9. The largest two MNRs are 2.52 from configurations 2 and 4, which are smaller than the critical value 2.65. Therefore, no outliers are detected for the OHT tests.

A.2. Kolmogorov–Smirnov tests

Kolmogorov–Smirnov (KS) test was adopted to identify the distribution of OHT tests. The KS test is a non-parametric test to quantify the goodness-of-fit between a given probability distribution and the empirical distribution of samples. The p-value of KS test indicates the probability that samples do not reject the hypothetical distribution. A high p-value denotes the high probability that samples are from the hypothetical distribution. The OHT tests are examined against Normal, uniform and Weibull distributions. The p-values are summarized in Table 10. The Normal distribution fits best for 4 configurations, uniform distribution fits best for 1 configuration and Weibull distribution fits best for 3 configurations. No significant indication favors a single distribution while assuming the experiments follow the same type of distributions at different configurations.