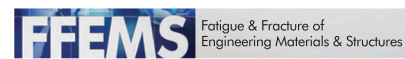


ORIGINAL ARTICLE



WILEY

Bayesian optimized collection strategies for fatigue strength testing

Christopher Massimo Magazzeni^{1,2} | Rory Rose¹ | Chris Gearhart³ |
Jicheng Gong¹ | Angus J. Wilkinson¹

¹Department of Materials, University of Oxford, Oxford, UK

²Materials and Processes Section, European Space Agency, ESA-ECSAT, Fermi Ave, Harwell, UK

³Center for Integrated Mobility Sciences, National Renewables Energy Laboratory, Golden, USA

Correspondence

Christopher Massimo Magazzeni and Rory Rose, Department of Materials, University of Oxford, Parks Road, OX1 3PH, Oxford, UK.
Email: christopher.magazzeni@queens.ox.ac.uk and rory.rose@alloyed.com

Funding information

Engineering and Physical Sciences Research Council; Royal Commission for the Exhibition of 1851; National Renewable Energy Laboratory; U.S. Department of Energy (Office of Science, Office of Basic Energy Sciences and Energy Efficiency and Renewable Energy, Solar Energy Technology Program)

Abstract

A statistical framework is presented enabling optimal sampling and analysis of constant life fatigue data. Protocols using Bayesian maximum entropy sampling are built based on conventional staircase and stress step methods, reducing the requirement of prior knowledge for data collection. The Bayesian Staircase method shows improved parameter estimation efficiency, and the Bayesian Stress Step method shows equal accuracy to the standard method at larger step size allowing experimentalists to lessen concerns of loading history. Statistical methods for determining model suitability are shown, highlighting the influence of protocol. Experimental validation is performed, showing the applicability of the methods in laboratory testing.

KEYWORDS

fatigue, high cycle fatigue, optimum design, statistical model, statistics of extremes

1 | INTRODUCTION

The fatigue of materials affects the longevity of engineered components, causing failures in service with often-costly outcomes.¹ The life of materials is known to decrease with increasing cyclic plastic strain and/or stress. Populating the characteristic stress versus life—or SN curve—can be

carried out by subjecting test specimens to cyclic loading. Fatigue data are often collected by considering the number of cycles for each specimen to fail “life testing”, though sometimes—such as in the high cycle fatigue (HCF) region—the fatigue strength (distribution in strength at a given life) is needed. Obtaining better estimates of fatigue strength parameters could inform more efficient engineering designs, and faster data collection enables shorter, lower cost, development cycles. The staircase and step methods are the most common protocols for obtaining these results with single tests or with multiple tests on an individual specimen, respectively. These protocols are

Abbreviations: HCF, High cycle fatigue; LCF, Low cycle fatigue; MAP, Maximum a posteriori probability; HPD, Highest posterior density; BIC, Bayes information criterion; CDF, Cumulative distribution function; PDF, Probability density function.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Fatigue & Fracture of Engineering Materials & Structures* published by John Wiley & Sons Ltd.

TABLE 1 Description of conventional fatigue strength protocols

Protocol	Description	Class
Staircase/ Up- down	The first specimen is tested at the <i>starting stress</i> . For the next specimen, the stress is decreased or increased by a <i>step size</i> if the preceding specimen failed or ran out, respectively. The final state (failed or runout) of each specimen is recorded.	Single test
Step	Specimens are initially tested at the <i>starting stress</i> . If the specimen runs out, it is tested again increasing the stress by the <i>step size</i> until failure. The runout and failure stress are recorded for each specimen.	Multiple test

described in brief in Table 1, with more extensive descriptions in papers and standards.^{2–8}

In single-test strategies such as the staircase method, protocol parameters and data analysis are ambiguous, conflicting, or require prior knowledge of the system. As reported by Lin, Lu and Lee,⁹ there is disagreement within the standards on key protocol parameters for step testing such as sample size, starting stress, and step size as well as the method for analyzing test data. The Dixon–Mood³ method, a common analysis technique used in industry for staircase data, discards half of the test data (failures or runouts), assumes a normal distribution in failure life, and recommends stress increments approximately equal to the sample set standard deviation which is not initially known. Pollak's investigation¹⁰ on simulated test data showed the effect of sample size and starting stress on estimates of the standard deviation and developed correction functions to increase the speed of convergence of the mean estimate. Engler–Pinto et al. developed a maximum likelihood-based framework for analyzing staircase data that includes the information from all the specimens, further incorporating life testing to generate parametric probabilistic SN curves.^{11,12} These studies all highlight the sensitivity of the protocol parameters in the analysis of single-test results.

In comparison to staircase protocols, stress step methods have the advantage that all specimens are tested to failure. However, there is significant debate over the reliability of step test data due to concerns over the effect of previous tests, through damage or coxing.^{2,13–17} Furthermore, as each specimen is tested multiple times, the experimental time taken per specimen can be higher than staircase testing. Nevertheless, these effects have not been fully characterized, and the testing strategy remains in use across the academic literature.^{18–37} For these reasons, step

sizes are generally advised to be above 5% of the fatigue limit.⁴ However, this requires prior knowledge of material properties and is dependent on the life being tested, as well as the continued debate over whether this is sufficient to avoid damage/coxing. While larger step sizes reduce the effect of loading history, there is a trade-off between accuracy and precision: the details of sample distribution are lost when carrying out tests at larger step size.

There is also disagreement in the analysis of these results: the American Society for Testing and Materials guide suggests using the mean of the runout and failure stress as the fatigue strength of that specimen,⁴ whereas Lerch et al. use the runout stresses only.¹⁷ While some analysis methods aiming to account for coxing or damage effects are available,^{30,33} extensive data are required to conclude their applicability and all agree that a larger step size reduces the effect.

We present two improved protocols for the collection of characteristic fatigue strength distribution parameters using a Bayesian Optimized Maximum Entropy Sampling method. Based on Lindley's approach,^{38–40} this broadly aims to design an experiment that maximizes some expected experimental aim, such as the information gained from each subsequent specimen. In this field, a function is created which expresses the utility of testing at any location given the information from a prior. Bayesian optimal experimental design is discussed in reviews from DasGupta,⁴¹ Chaloner et al.,⁴² and more recently by Ryan, Drovandi et al.^{43,44} We present this method applied to simulated results followed by its application in a laboratory experiment. The toolbox is available open source on the GitHub repository “BayesOptFatigue.”⁴⁵

2 | APPLYING BAYES THEORY TO FATIGUE TESTING

The experimental design is optimized such that, based on information from all previous specimens, the subsequent specimen provides a maximized increase in the predictive capability of the model. To describe this, the i^{th} specimen is tested at a stress level x_i with result $j_i \in \{1, 0\}$, indicating that the specimen has runout ($j_i = 0$) or failed ($j_i = 1$) by the target life. In tests with multiple stresses per specimen, we denote the stress level by a second index k : $x_{i,k}$. As we consider only step testing protocols with constant step size, the test levels denoted by k start from $x_{i,1}$ and incrementing by the step size ξ such that for some specimen x_i :

$$x_{i,k} = x_{i,k-1} + \xi \quad (1)$$

The data from all tested specimens is collected to form the matrices \mathbf{x} and \mathbf{j} , which contain the stresses and

results from every specimen and step increment such that every row corresponds to a specimen. For single test protocols, this matrix becomes a vector where each entry is a new specimen. The distribution in failure stress is modeled with parameter θ . For demonstration purposes, we start with the example of a normal distribution that is represented by the parameters mean (μ) and standard deviation (σ), that is, $\theta = [\mu, \sigma]$.

A Bayesian approach evaluates the probabilities of model parameters given observations, built on Bayes rule (Equation 2):

$$p(\theta | [\mathbf{j}, \mathbf{x}]) = \frac{p(\mathbf{j} | [\mathbf{x}, \theta]) p(\theta)}{p(\mathbf{j}, \mathbf{x})} \quad (2)$$

The posterior $p(\theta | [\mathbf{j}, \mathbf{x}])$ describes the probability supporting the model parameters. This posterior is inferred from the prior, $p(\theta)$, which describes the probability supporting the model without any data, the likelihood function $p(\mathbf{j} | [\mathbf{x}, \theta])$ that gives the likelihood of collecting the data given the model parameters, and finally the marginal likelihood $p(\mathbf{j}, \mathbf{x})$ of collecting the data (Equation 3):

$$p(\mathbf{j}, \mathbf{x}) = \int d\theta p(\mathbf{j} | [\mathbf{x}, \theta]) p(\theta) \quad (3)$$

The prior used is constant with a range in mean and standard deviation that should cover a parameter space to confidently include the model parameters. A reasonable assumption in a metallic case could, for example, cover a mean from 0 MPa to the Ultimate Tensile Stress (UTS) and standard deviation between 1 and half the UTS. A discussion of initial prior choice is given in Appendix A.1.

The likelihood L_i of a single test i is evaluated following the approach of Engler-Pinto et al.^{11,12} and Smalakys et al.,⁴⁶ represented graphically in Figure 1 and Equation (4), where $\Phi(x)$ is the cumulative distribution function. The staircase test can either runout (where $j_i = 0$, and said as “right-censored”) or fail (where $j_i = 1$, or it is “left-censored”). For step testing, the stress is incremented until failure, so there are no runouts. The specimen can either fail on the first step ($j_i = 1, k = 1$, left-censored) or fail on any other step ($j_i = 1, k \neq 1$, interval-censored). Therefore, Equation (4) can be rewritten as Equation (5):

$$L_i = p(j_i | [x_{i,k}, \theta]) = \begin{cases} 1 - \Phi(x_i), & \text{Runout} \rightarrow \text{right-censored} \\ \Phi(x_i), & \text{Failure} \rightarrow \text{left-censored} \\ \Phi(x_{i,k}) - \Phi(x_{i,k-1}), & \text{Step} \rightarrow \text{interval-censored} \end{cases} \quad (4)$$

$$L_i p(j_i | [x_{i,k}, \theta]) = \begin{cases} \Phi(x_i) \cdot \delta_{1,j_i} + [1 - \Phi(x_i)] \cdot \delta_{0,j_i} - \text{Staircase} \\ \Phi(x_{i,k}) \cdot \delta_{1,k} \cdot \delta_{1,j_i} + [\Phi(x_{i,k}) - \Phi(x_{i,k-1})] \cdot (1 - \delta_{1,k}) \cdot \delta_{1,j_i} - \text{Step} \end{cases} \quad (5)$$

where it is assumed that the step test stress is incremented until failure. The likelihood $p(\mathbf{j} | [\mathbf{x}, \theta])$ with N specimens is obtained by multiplying the likelihoods for all specimens (Equation 6):

$$L = p(\mathbf{j} | [\mathbf{x}, \theta]) = \prod_{i=1}^N L_i \quad (6)$$

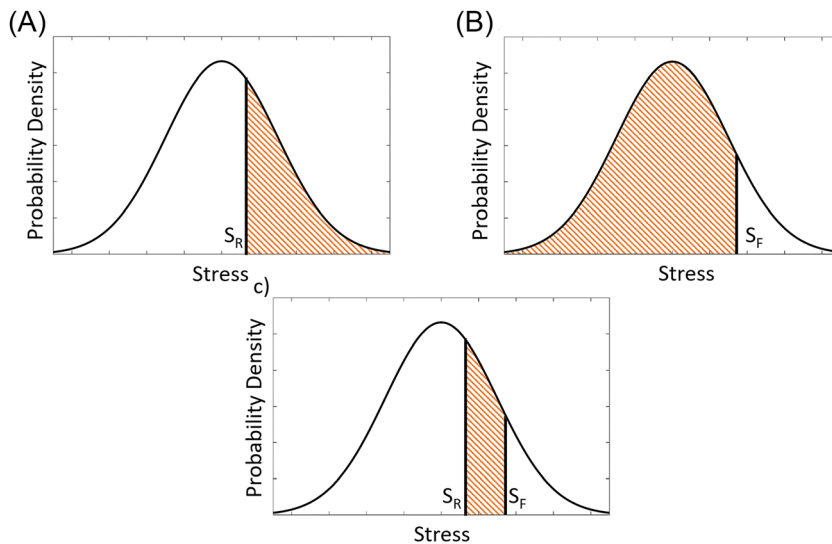


FIGURE 1 Representation of probabilities of collecting results for fatigue strength tests. (A) Probability of a runout, implying that the specimen fatigue strength is not below S_R —that is, that it is right-censored. (B) A failure at S_F only and (C) a runout and failure pair. [Colour figure can be viewed at wileyonlinelibrary.com]

2.1 | Applying Bayesian maximum entropy sampling to fatigue data

The goal of precisely estimating distribution parameters can be thought of as generating the most compact, or tightly constrained, joint posterior parameter distribution. To choose the next stress to test that will maximize this predictive power, Ryan, Drovandi et al.⁴³ recommend maximizing the expected gain in Shannon information I Equation (7) in the joint posterior. The next stress to test is therefore chosen to maximize the expected information as in Equation (8) from sampling at a subsequent stress x_{i+1} . Often called entropy, this expected information can be intuited as a quantitative measure of the compactness of the posterior distribution; a derivation can be found in Chaloner et al., Ryan, Drovandi et al., and Han and Chaloner^{42–44,47}:

$$I(p(\theta | [j_{i+1}, x_{i+1}, \mathbf{j}, \mathbf{x}])) = -\ln(p(\theta | [j_{i+1}, x_{i+1}, \mathbf{j}, \mathbf{x}])) \quad (7)$$

$$\text{ent}(p(\theta | [j_{i+1}, x_{i+1}, \mathbf{j}, \mathbf{x}])) = E[I(p(\theta | [j_{i+1}, x_{i+1}, \mathbf{j}, \mathbf{x}]))] = -\int d\theta p(\theta | [j_{i+1}, x_{i+1}, \mathbf{j}, \mathbf{x}]) \ln(p(\theta | [j_{i+1}, x_{i+1}, \mathbf{j}, \mathbf{x}])) \quad (8)$$

To maximize the compactness of the posterior distribution, we define a gain in entropy $\Upsilon(j_{i+1} | x_{i+1}, \mathbf{j}, \mathbf{x})$ from the addition of a possible data j_{i+1} (a failure or runout) at x_{i+1} :

$$\Upsilon(j_{i+1} | x_{i+1}, \mathbf{j}, \mathbf{x}) = \text{ent}(p(\theta)) - \text{ent}(p(\theta | [j_{i+1}, x_{i+1}, \mathbf{j}, \mathbf{x}])) \quad (9)$$

As $\text{ent}(p(\theta))$ does not vary with data, we can neglect this term. We therefore build a utility function $U(x_{i+1})$ to evaluate the expected value in the change of entropy $E[\Upsilon(j | x_{i+1}, \mathbf{j}, \mathbf{x})]$ of including the datapoint x_{i+1} over all possible outcomes J of the next specimen:

$$\begin{aligned} U(x_{i+1}) &= -E[\Upsilon(j | x_{i+1}, \mathbf{j}, \mathbf{x})] \\ &= \sum_J p(J | [x_{i+1}, \mathbf{j}, \mathbf{x}], \theta) \int d\theta p(J | [x_{i+1}, \mathbf{j}, \mathbf{x}], \theta) \ln(p(J | [x_{i+1}, \mathbf{j}, \mathbf{x}], \theta)) \end{aligned} \quad (10)$$

This can be intuited as evaluating the expected gain of information for testing the next specimen at some stress, for all outcomes. The maximization of this utility function therefore allows for the greatest expected entropy gain.

For staircase-type testing, the utility function $U^{\text{staircase}}(x_{i+1})$ is built by considering both possible outcomes—a runout or failure at stress x_{i+1} —and evaluates this expected entropy change as a function of the test stress:

$$U^{\text{staircase}}(x_{i+1}) = \Upsilon(0 | x_{i+1}, \mathbf{j}, \mathbf{x}) p(0 | [x_{i+1}, \mathbf{j}, \mathbf{x}], \theta) + \Upsilon(1 | x_{i+1}, \mathbf{j}, \mathbf{x}) p(1 | [x_{i+1}, \mathbf{j}, \mathbf{x}], \theta) \quad (11)$$

This can be understood as determining the expected value for each possible outcome through the entropy gain and its associated probability given the previous data. In this case, we aim to find the test stress x_{i+1} that maximizes $U^{\text{staircase}}$.

For step-type testing, the design space allows for varying two test parameters: the starting stress $x_{i,0}$, as well as the step size ξ . The utility function $U^{\text{step}}(x_{i+1,1}, \xi)$ is similarly built by considering possible outcomes as discussed above. Since the specimen could fail on any step—the expected Υ from each step contributes to the expected utility. Therefore, the utility function becomes Equation (12):

$$\begin{aligned} U^{\text{step}}(x_{i+1,1}, \xi) &= \Upsilon(1 | x_{i+1,1}, \mathbf{j}, \mathbf{x}) p(1 | [x_{i+1,1}, \mathbf{j}, \mathbf{x}], \theta) \\ &\quad + \sum_{k=2}^{N_k} \Upsilon(1 | x_{i+1,k}, \mathbf{j}, \mathbf{x}) p(1 | [x_{i+1,k}, \mathbf{j}, \mathbf{x}], \theta) \end{aligned} \quad (12)$$

where the first term denotes failing on the first step, and the sum across k evaluates the utility of failing at each subsequent step, recalling that $x_{i+1,k} = x_{i+1,1} + k\xi$. Here, N_k is the number of possible test stresses (k) from the starting stress to the maximum possible test stress. The sum needs only be defined up to the point where the probability of collecting a runout and fail pair, $p(1 | x_{i+1,k}, \theta)$, becomes infinitesimally small. In the computational implementation, as seen on GitHub under “BayesOptFatigue”, this was arbitrarily taken as the maximum value of the location or mean parameter plus 6 times the maximum spread value, which was shown to be sufficient.

3 | COMPUTATIONAL FRAMEWORK

To validate our framework, a simulated material’s fatigue response is generated. We arbitrarily generalize the fatigue response for any life as a modified Basquin model with a fatigue limit, to determine the fatigue strength at a certain life (Equation 13):

$$S_{\text{mean}} = \alpha \log(N)^{-\beta} + c \quad (13)$$

where the Basquin constants α, β , and c are taken in this simulated system to be 600 MPa/log-(cycle), 1, and 300 MPa, respectively. As such, the mean fatigue strength

at 10^6 cycles is 400 MPa. The modeling of the entire fatigue curve is at this stage arbitrary, but this is useful when extending the Bayesian approach to multiple lives.⁴⁸

We use a Monte Carlo based method for generating a simulated sample set with many fatigue responses by perturbing this mean response in stress by a distribution with some characteristic spread. These simulated results use a normal distribution as in Equation (3) which is chosen as a simple model, though readers are invited to the literature on the basis for fatigue strength models,^{9,49–61} and six other distributions are implemented in the code. Thus, a specimen with a determined fatigue response is a member of a sample set with a broad fatigue response. The simulation treats each sample curve as unique and to be tested as in the experimental protocols (e.g., they can only fail once). This can be written as Equation (14):

$$S_i = \alpha \log(N)^{-\beta} + c + \delta(0, \sigma) \quad (14)$$

where $\delta(0, \sigma)$ is a random number with distribution centered about 0, and a spread (e.g., standard deviation) defined by σ . This results in a mean fatigue response at 10^6 cycles is 400 MPa, and a spread of σ . In this framework, a single specimen's curve is selected to be subject to the test protocol, for either a single test or multiple step tests. For the selected specimen, if the fatigue strength at life $\log(N)$ is higher than the test stress, it has not failed, while if the fatigue strength is lower than the test stress, it has failed. This allows for the simulated testing of large sample sets and comparison between protocols (Figure 2).

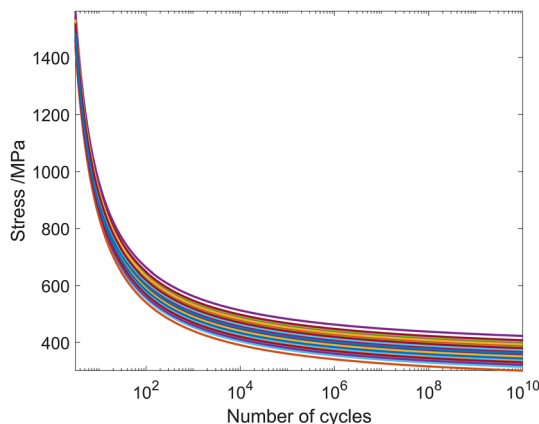


FIGURE 2 Example simulated sample set with each specimen modeled as a Basquin curve perturbed in stress by a normally distributed random variable [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

In this computational method, the prior is built as a discrete N -dimensional array with the characteristic distribution parameters along the axes such as location and spread, or mean and standard deviation in the case of a normal distribution. The range of values of parameters defines the extent of the flat prior, and the grid size affects the precision of model parameter estimates. The sequential evaluation and normalization of the posterior is therefore (Equation 15):

$$p(\theta|\mathbf{x}) = \frac{p(\theta|\mathbf{x}_{1:i-1})p(x_i|\theta)}{\sum p(\theta|\mathbf{x}_{1:i-1})p(x_i|\theta)} \quad (15)$$

As such, each element in this array with coordinate of $(\theta_1, \theta_2, \dots)$ gives the probability that these discrete parameter values represent the data. $ent(p)$ is therefore calculated by summing the probability contributions from every possible combination of model parameters of the desired distribution.

A flat prior is initialized by creating an array of equal values such that $\sum p(\theta|\mathbf{x}) = 1$. A more detailed view of this process is included in the README of the GitHub repository.⁴⁵

3.1 | Bayesian staircase protocol

The Bayesian staircase collection strategy (Figure 3) broadly resembles the conventional staircase method in three ways: the interpretation of the algorithm's behavior, an approximate similarity in collection in the initial points, and an analysis based on a sample set containing failed specimens and runout data.

The conventional staircase protocol implicitly infers that the fatigue strength is below the test stress if a specimen has failed, and vice versa. The Bayesian protocol can be seen to follow this similar process but considers all prior samples when determining the next test stress rather than following a rigid protocol. The first few specimens perform a binary search to estimate the mean fatigue strength. As the estimate becomes more precise, there is less information to be gained by testing at the center of the distribution, and the algorithm then tries to determine the standard deviation. This can be seen in Figure 3D as the transition in the utility function to a bimodal curve with peaks at approximately one standard deviation from the mean. This transition quantifies the intuitive approach of estimating the mean followed by the spread of a sample set, similar in behavior to Never's D-optimal test strategy.⁶²

The evolution of the estimate in the joint posterior can be seen in Figure 4. The Bayesian protocol can be seen to constrain the estimates for mean and standard

FIGURE 3 Staircase plot (A, C) and associated utility function plots (B, D) for 4 and 90 specimens. The simulated sample set had a mean of 400 MPa and distribution width of 40 MPa [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1111/ffe.13859)]

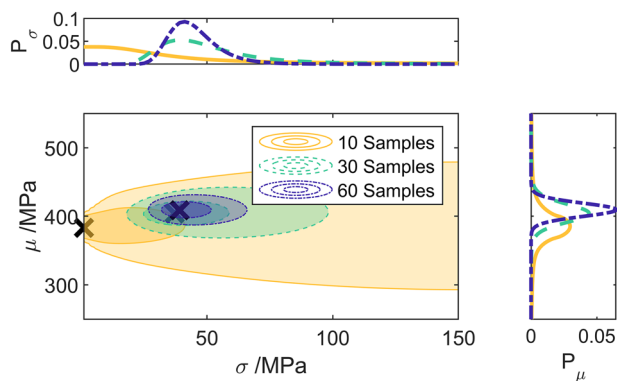
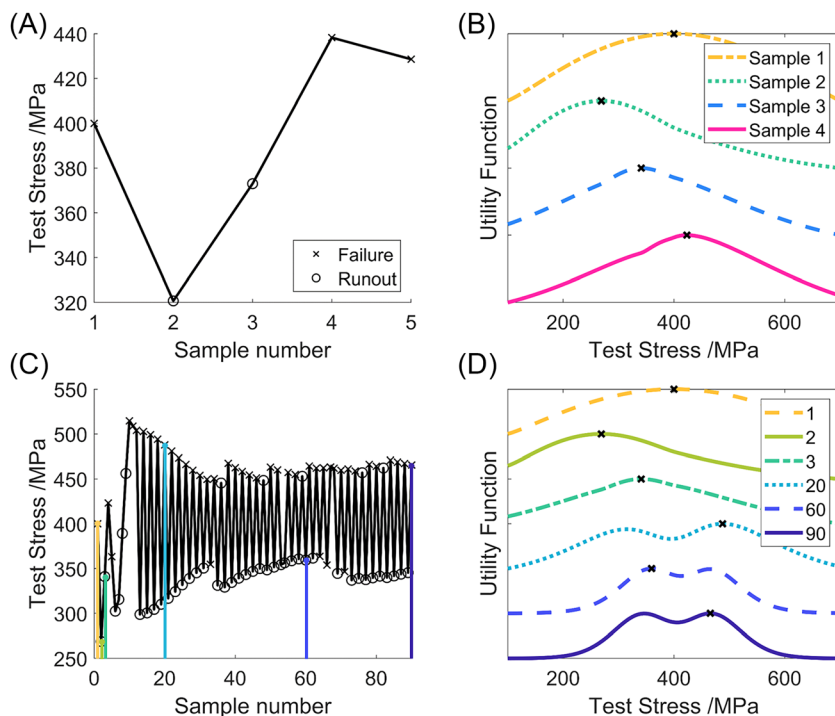


FIGURE 4 Evolution of joint posterior surface for Bayesian staircase protocol. The simulated sample set had a mean of 400 MPa and a distribution width of 40 MPa. Contours are at 50 and 90% HPD interval. The 2D array of model parameters can be seen, with projections showing the probability for each individual parameter [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1111/ffe.13859)]

deviation, with fewer model parameters falling within the 90% highest posterior density (HPD) surface as data are collected. It can likewise be seen that the estimate for mean is determined within 10 samples to a reasonable degree of confidence, while the standard deviation requires further testing.

The Bayesian protocol has several advantages to the regular staircase method: very little information is required about the sample set prior to testing, testing parameters are not reliant on prior knowledge, and a faster convergence is obtained. The first two points follow

from the protocol's process, and the latter can be seen in Figure 5. This figure shows the trend in error as a function of specimen number for various protocol parameters, as well as the error after 100 samples as a function of the ratio of stress increment to the true distribution standard deviation. The percentage error is calculated relative to the known true distribution parameters, for example, for the mean $\text{Error}_\mu = 100 * \frac{\mu_{\text{estimate}} - \mu_{\text{real}}}{\mu_{\text{real}}}$.

The behavior of the conventional method can be seen to be strongly sensitive to the protocol parameters. This implies that the choice of step size greatly influences the error in the result for conventional staircase testing. The Bayesian protocol outperforms or matches the conventional protocol for all possible step sizes.

3.2 | Bayesian step protocol

We create a protocol that finds an optimal step size and starting stress, setting out the stress levels for an individual specimen. For constant life testing, the optimum step size is always the smallest step size as it provides the smallest interval of fatigue strength for the specimen. This trivial result reflects the trade-off described above but without regard for damage effects. Therefore, the benefit of this Bayesian approach in comparison to the conventional protocol arises when using larger step sizes: where damage is less of a concern and the role of starting stress can greatly influence accuracy. Thus, the dimensionality of the utility surface can be reduced to only

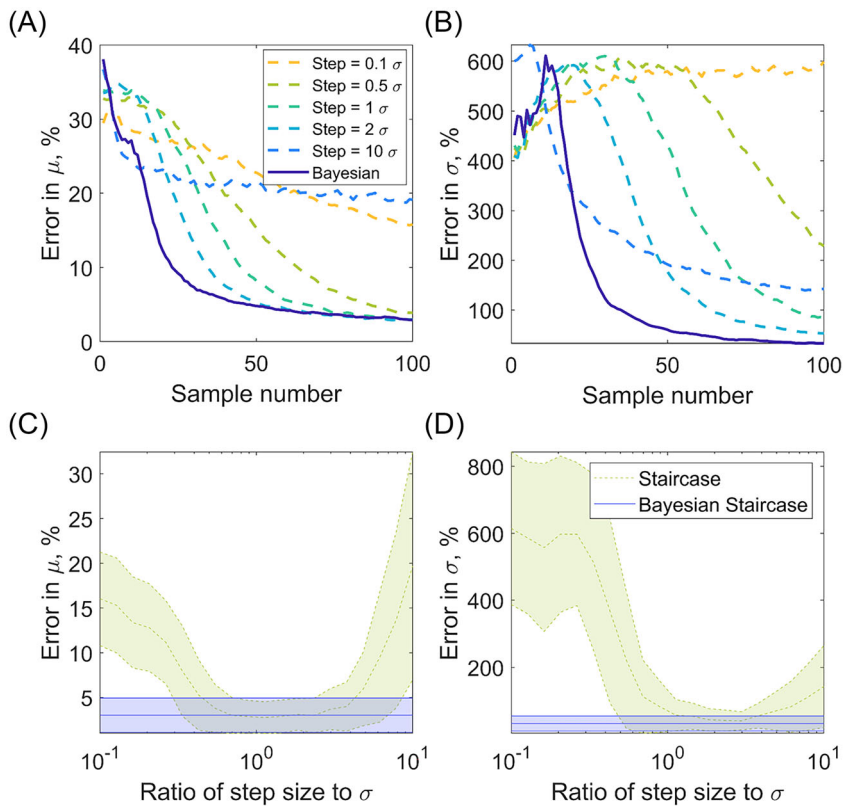


FIGURE 5 Convergence plot of (A) mean and (B) standard deviation for a variety of step sizes, and comparison of error after 100 samples in (C) mean and (D) standard deviation between staircase (green) and Bayesian staircase (blue) protocol. The simulated sample set had a mean of 400 MPa and a distribution with of 40 MPa, and the simulations in Figure 5C and D are repeated 200 times to generate a standard deviation in the values (shaded regions). The comparison plot in Figure 5C and D can be seen to be a cross section of Figure 5A and B after 100 samples for a larger number of step ratios. Figure 5A and B lines are smoothed in the x axis with a window of five samples [Colour figure can be viewed at wileyonlinelibrary.com]

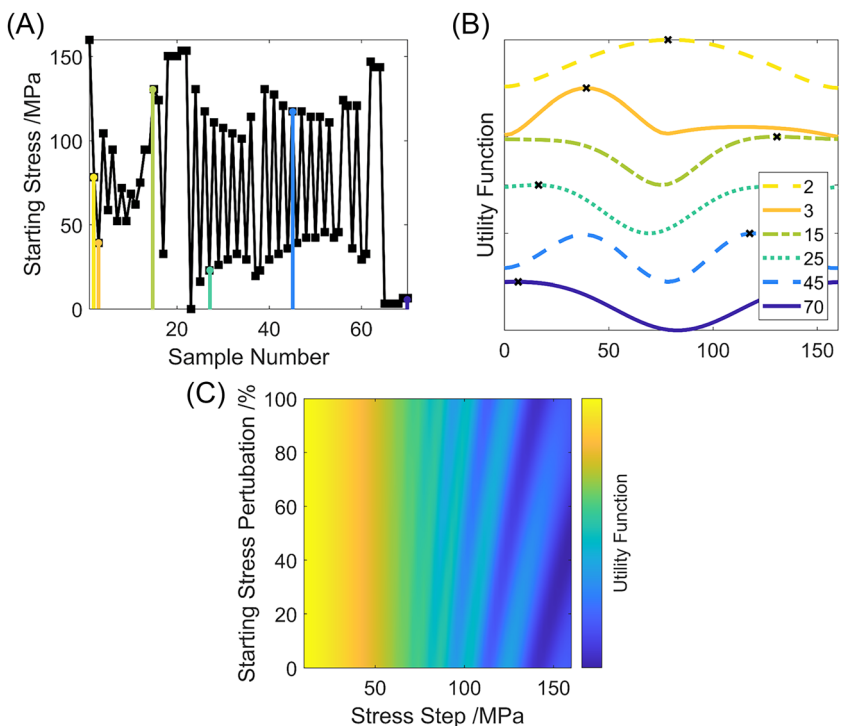
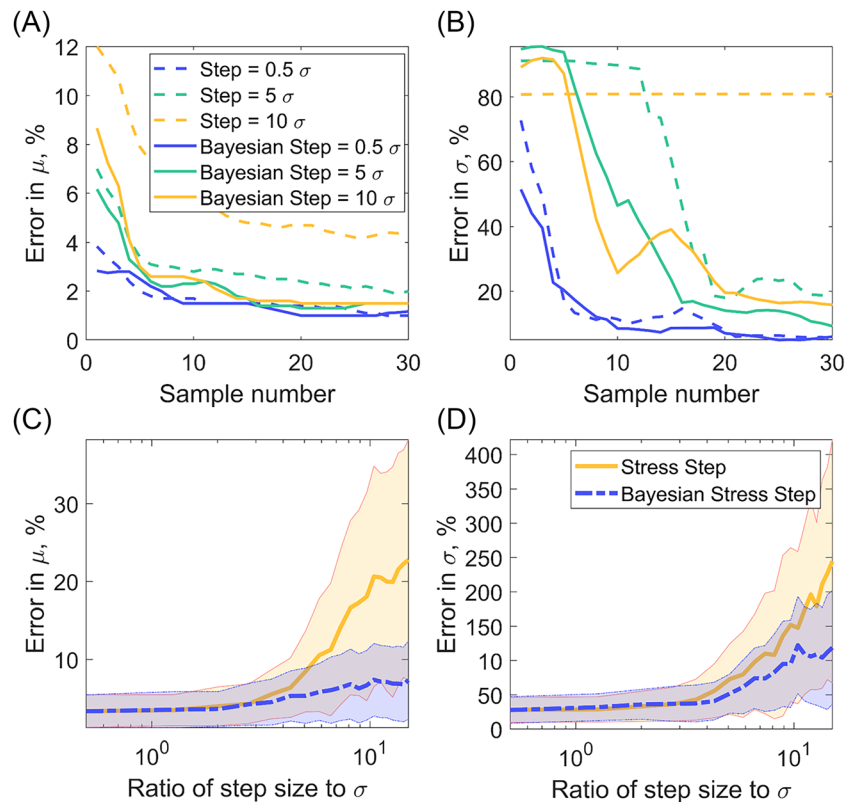


FIGURE 6 (A) Staircase plot of the Bayesian stress step protocol, showing the starting stress as a function of specimen number. The specimens simulated had a mean strength of 400 MPa and a distribution width of 40 MPa. The stress step was 160 MPa, with a maximum starting stress of 160 MPa. (B) Utility function showing the relative utility of various starting stresses, at a selection of specimen numbers, color-coded as in Figure 6A. (C) Utility at 70 samples is shown when allowing for a variation in step size, showing the preference for smaller step sizes. In this case, the vertical slice at a stress step of 160 MPa is equal to the curve at 70 specimens in Figure 6B [Colour figure can be viewed at wileyonlinelibrary.com]

considering varying the starting stress with a constant step size. Here, we choose 160 MPa, or 4 times the standard deviation, shown in Figure 6. This significantly larger step size is used only as a demonstrator of the effect of starting stress.

Generally, the protocol follows similar behavior to the Bayesian staircase protocol. The starting stress has large variations at low numbers of samples where varying the possible failure stresses gives a quicker estimation of the mean fatigue strength. When the step size is larger

FIGURE 7 Convergence plot of (A) mean and (B) standard deviation for a variety of step sizes, and comparison of error after 30 samples in (C) mean and (D) standard deviation between conventional (yellow) and Bayesian (blue) protocols for a variety of ratios between step size and true distribution width. The simulated sample set had a mean of 400 MPa and a distribution width of 40 MPa, and the simulations in Figure 7C and D are repeated 200 to generate a standard deviation in the values (shaded regions). The comparison plot in Figure 7C and D can be seen to be a cross section of Figure 7A and B after 30 samples for a larger number of step ratios. Figure 7A and B lines are smoothed in the x axis with a window of five samples [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1111/ffe.13859)]



than the width of the sample strength distribution, the starting stress tends towards where the stress levels intersect stresses approximately one standard deviation away from the mean, analogous to the behavior noted in section 3.1. Conversely, when the step size is much smaller than the width of the sample strength distribution, there is comparatively less to gain by varying starting stresses.

As in the case of the staircase protocols, a comparison can be made between the conventional stress step protocol and its Bayesian counterpart, as a function of the stress increment size shown in Figure 7. As discussed in Appendix A.2, mean values of error for step testing do not capture the bi-modal distribution in error, that is, an experimentalist is most likely to observe higher error for small numbers of samples. It should also be noted that the error level, even at this reduced number of specimens, is lower than for staircase testing at small step sizes. This is in line with common practice, highlighting the trade-off in choosing protocols between efficiency and concern about damage.

The benefit from the Bayesian step protocol comes when using larger step sizes and the relative difference of starting a fraction of a step away (Figure 7). For these larger step sizes, the Bayesian protocol has a significantly better performance in estimating both mean and standard deviation. For example, the Bayesian protocol with a 15 σ step size obtained a better estimate of mean than a

conventional step protocol with a 5 σ step after 30 samples.

Stress step tests can therefore be used with comparatively less concern over damage by using significantly larger step sizes and varying the starting stress. Compared to the conventional protocol, a stress step size approximately twice as large can be used to obtain the same error level for a set number of specimens.

3.3 | Statistical model selection

The choice of statistical distribution model to describe fatigue strength will strongly affect the probability of failure especially away from the median stress. This has motivated work in determining model suitability.^{9,49–60} We provide a tool for the comparison of different statistical models given experimental data collected, such that model suitability can be explored.

This work focuses on Bayes Information Criterion (BIC), shown in Equation (16). BIC is based on the maximum log-likelihood \hat{L} , number of observations n , and the number of parameters k in the model. The term with number of parameters penalizes against models with more parameters as they will be more liable to overfitting. Models with a lower BIC should be selected over other models.

$$BIC = k \ln(n) - 2 \ln(\hat{L}) \quad (16)$$

$$\hat{L} = \max_{\theta} p(\mathbf{j} | [\mathbf{x}, \theta]) \quad (17)$$

\hat{L} is the maximum value of the likelihood function (Equation 17) which describes the probability of observing data given a model M , data $\mathbf{j} | \mathbf{x}$, and parameters θ . Thus, \hat{L} gives a measure of the goodness-of-fit of the data to the best model of that functional form. Following Kass and Raftery,⁶³ the value of the BIC can give evidence for a model over another model with lower values indicating more evidence. To compare two models, ΔBIC can be defined as $\Delta BIC = BIC_1 - BIC_2$, and shows evidence for BIC_1 if it is negative. Values of ΔBIC between 0 and 2 indicate weak evidence, 2 and 6 indicate substantial evidence, and 6 to 10 indicate strong evidence for the lower BIC.

A simulated experiment was performed to observe the effectiveness of conventional and Bayesian protocols in distinguishing different distribution models, with the result shown in Figure 8. The ability of this criterion to select from these distributions can be seen to be dependent on the protocol used as well as the model of the input sample distribution: step-type protocols resolve the distribution type much more effectively than staircase-type protocols. Figure 8 repeat experiments comparing to three-parameter Weibull, Gumbell, and Frechet produced similar distinction between distribution models, with stress-step methods providing stronger distinction between distributions. More generally, step-type

protocols with larger step size relative to distribution were more effective at distinguishing distribution type, and it is harder to distinguish log-normal from normal.

However, it must be noted that the distinction between these distributions remains unclear after 80 specimens, and further testing is required to develop a confident distinction with these protocols. This does suggest the possibility of developing a utility function with an experimental aim to maximize the distinction between distributions.

3.4 | Comparing protocols and limitations

The comparison between conventional and Bayesian-type protocols as shown in Figures 5 and 7 are a result of our definition of convergence. However, this visualization significantly reduces the information present in the simulations, and our definition of convergence is just one of many ways to compare results. The current definition only considers the deviation of the MAP from the true value. Further considerations can be made, for instance, by observing the gradient at which the MAP reaches a value and does not change, the number of specimens required for this converged value not to change by some threshold, and analysis considering the distribution in parameter probabilities. However, regardless of the definition, it was observed that Bayesian counterparts to conventional methods lead to faster, narrower, and more accurate estimates of both parameters, independent of prior knowledge. Furthermore, in comparing the

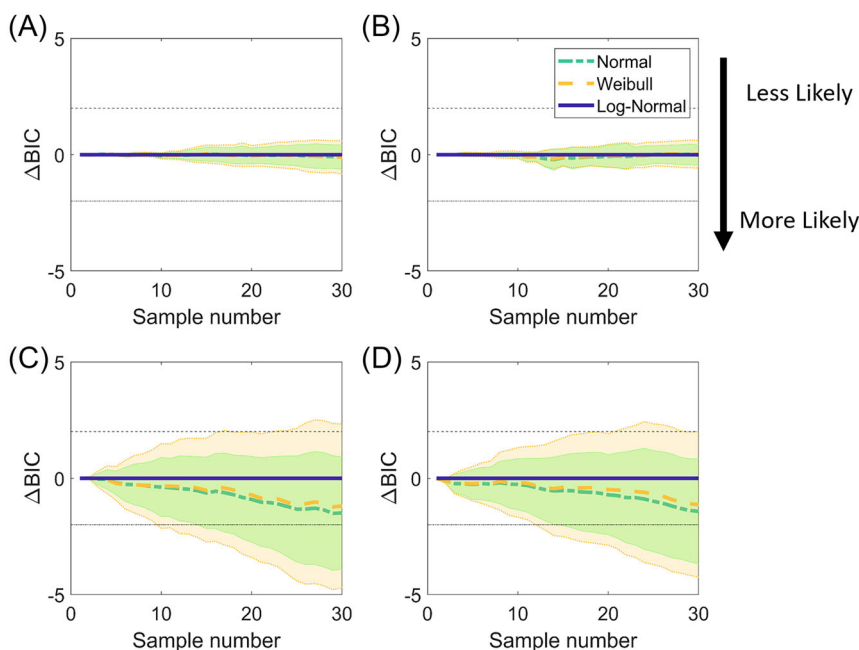
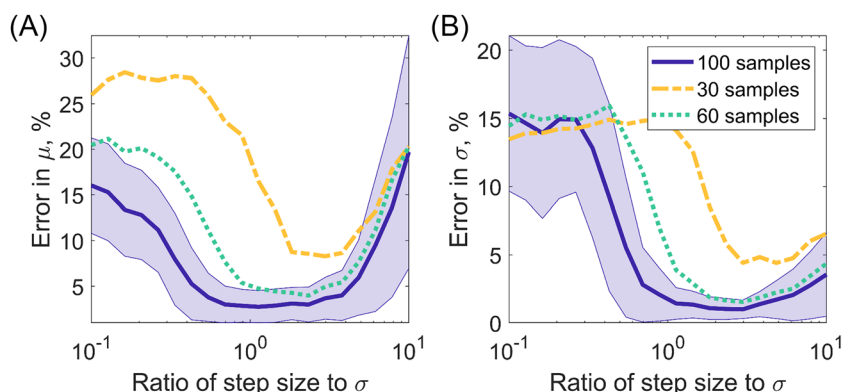


FIGURE 8 Evolution in ΔBIC (difference in Bayes Information Criterion) for data with normal distribution and mean of 400 MPa with standard deviation 40 MPa. ΔBIC is calculated relative to BIC for two-parameter Weibull distributions, with the lowest BIC indicating a distribution with stronger evidence. (A) Staircase, (B) Bayesian staircase, (C) stress step, and (D) Bayesian stress step protocols based on 100 repeats (shaded region corresponding to one standard deviation). Step size was 40 MPa, and starting stress was 380 MPa for staircase and 120 MPa for step. The dashed lines represent the level of evidence shown as in Kass and Raftery⁶³ [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 9 Comparison of error in (A) mean and (B) standard deviation as shown in Figure 5C and D examining only the staircase protocol for a variety of sample numbers [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]



protocols considered in this analysis, there is a clear difference between the rate of convergence between single-stress tests and multiple-stress tests. Stress step methods may be favorable when fewer specimens are available, whereas single-stress tests may be favorable when conducting tests in a shorter amount of time or without concern of loading history.

Examining further, these results also give an insight into the common practice of the conventional protocols. The staircase method can be seen to have the lowest error in determining the mean when the stress increment is near the standard deviation, reflecting the common practice discussed as early as in the original Dixon–Mood reports.³ However, it can also be seen that this common practice forgoes the nuance in this decision. Figure 9 shows the convergence plot as shown in Figure 5 with only the staircase results after 100 samples, overlaid with the errors as obtained after 30 and 60 samples.

While the optimal increment for measuring the mean to the lowest error is indeed equal to the standard deviation, this is not true for measuring the standard deviation. Furthermore, the optimal increment for measuring both mean and standard deviation changes with sample number. Initially, it is more effective to collect data with a larger increment, a result that is reflected in the Bayesian protocol during the binary search and ensuing collection. This highlights the significant advantage of the Bayesian protocol: an evolving sequential adaptation of the testing parameters to the data.

However, while the Bayesian methods avoid requiring estimates of the fatigue system to be measured in determining starting stress or step size, the construction of the prior itself is necessary. Both the range in parameters values as well as the coarseness of this discrete mesh require some broad estimation, though this requirement is substantially more flexible: the binary search is sufficiently quick at determining the mean if it is within the extent of the prior. Coarseness of the discrete mesh is only a concern when, for reasons of computational efficiency, the mesh has been coarsened to such a degree

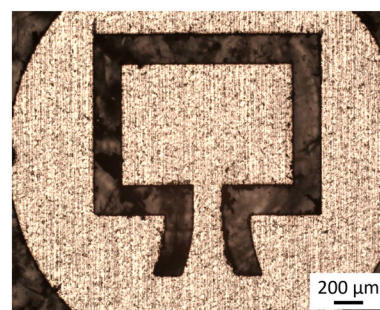


FIGURE 10 Optical micrograph of specimen geometry. The length and width of the cantilever are 600 and 200 μm , and the foil from which it is cut is 100- μm thick. The system acts as a tuned harmonic oscillator with the large rectangular mass acting as a counterweight. The geometry is designed in finite element analysis to determine the stress during vibration [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

that the model predicts likelihood for one set of parameters and zero probability for neighboring parameters. During collection of either simulated or experimental results, this can be easily visualized and amended.

This method lends itself to generalization to allow for a broader set of experimental goals. The method can be generalized to extend across the P-SN space.⁴⁸ From earlier discussions of distribution types, it may be possible to generate a utility function that maximizes the distinction between distribution types.⁴⁴ This represents a potential for future work, highlighting the need in engineering practice to ensure predictive capability at the extremes of the distribution.

4 | EXPERIMENTAL VALIDATION

An experimental validation of the above simulations was carried out to confirm the basis of the Monte Carlo sample set and demonstrates the protocols in a physical laboratory setting.

Fatigue data are generated using a novel small-scale ultrasonic fatigue setup developed by J Gong. The data are collected at 20 kHz to rapidly generate SN data on a laser cut foil of 304 L stainless steel. Figure 10 shows the unique geometry, with the exact fatigue response not being used to infer any property of the base material.

TABLE 2 Experimental protocol parameters

Protocol	Starting stress (MPa)	Stress step (MPa)
Staircase	325	30
Step	150	30

The specimens were randomly extracted for positions across the foil of material and assigned to protocols in a randomized order to reduce time-dependent effects. Conventional protocol parameters used are summarized in Table 2 and were based on best estimates from previous testing on similar specimens with the same method: a mean of approximately 300 MPa and a spread of ~30 MPa. It should be noted that this guidance from preliminary testing is a requirement to set up conventional testing but is not needed for the Bayesian protocols.

Figure 11 shows the raw data from each of the protocols. All protocols follow the expected behavior, including a similarity to the Bayesian behavior as shown in section 3.

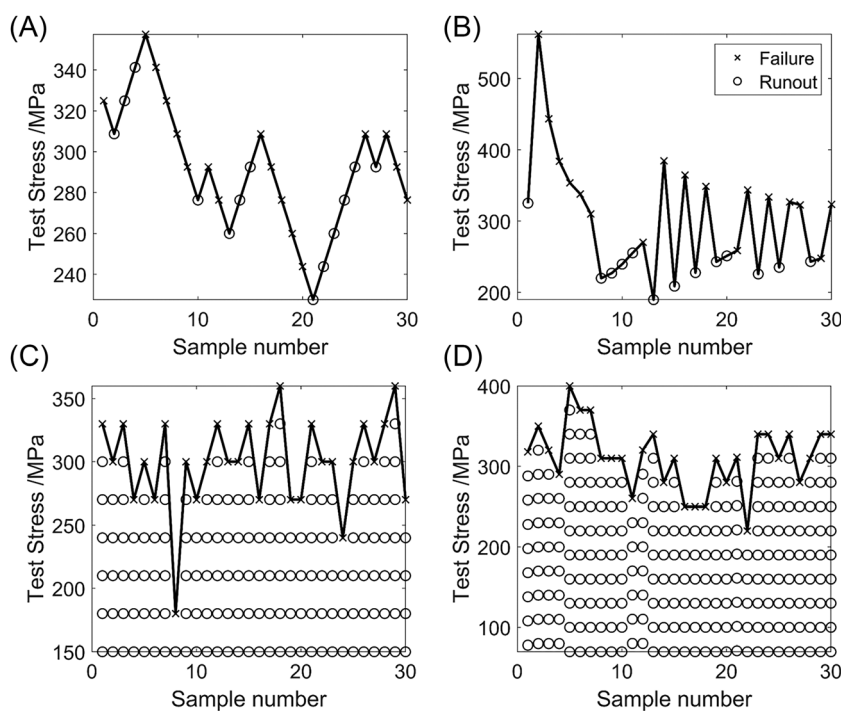


FIGURE 11 Staircase plots for the four experimental datasets. (A) Regular staircase test, (B) Bayesian staircase, (C) regular stress step, and (D) Bayesian stress step. The stress increment in all tests (except Bayesian staircase) is 30 MPa, and the starting stress for the stress step protocol is 150 MPa

TABLE 3 Results for the experimental campaign to compare the four sampling protocols

Protocol	Conventional analysis, MPa		Bayesian analysis, MPa: MAP, (95% HPD interval)	
	Mean	Standard deviation	Mean	Standard deviation
Staircase	Dixon–Mood analysis. N. B, the qualification factor is 0.18 (>0.3 required)		277 [113, 368]	83 [33,300]
	290	10		
Bayesian staircase			276 [233, 319]	35 [17, 114]
Step	Fitted normal distribution through regression: $R^2 = 0.39$. Value \pm 95% confidence interval:		285 [268, 302]	35 [26,51]
	285 \pm 15	37 \pm 8		
Bayesian step			296 [277,315]	39 [29, 56]
Whole dataset			288 [277,298]	39 [32, 49]

FIGURE 12 Evolution in parameter estimates as a function of sample number, for (A) mean and (B) standard deviation for the four protocols tested. Line represents the MAP, and shaded regions indicate the 95% HPD interval [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

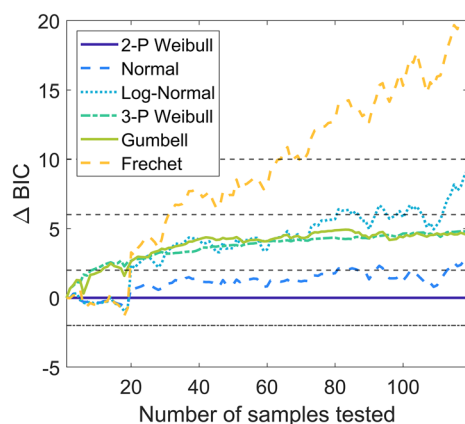
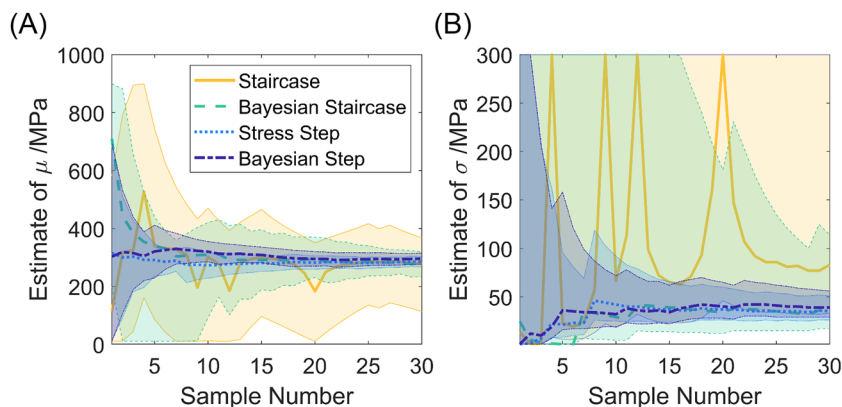


FIGURE 13 Experimental evolution of ΔBIC combining data collected using all protocols, showing positive evidence for the Weibull distribution over normal and strong over log-normal.⁶³ Specimen 5 of the Bayesian step method results (Figure 5D) was omitted due to experimental error. See Figure 8 for more details. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

When analyzing the results of all 120 specimens together—representing the most precise estimate of the fatigue behavior—the material system has a fatigue strength of 290 ± 10 MPa at 10^6 cycles, with a spread of approximately 40 ± 10 MPa. It can be seen in Table 3 and Figure 12 that the Bayesian protocol equivalents outperformed their conventional methods, with more accurate MAPs and narrower HPD intervals, despite good initial assumptions for starting stress and stress increment size in the conventional protocols. The analysis of these results via conventional methods, such as Dixon–Mood for staircase or fitted distributions for stress steps, is included alongside their respective Bayesian estimates in Table 3.

The conventional methods are unable to determine with reasonable confidence the parameter values that are best determined by considering the complete dataset—both the Dixon–Mood qualification factor and the fitted distribution R^2 being too low. In addition, the Bayesian analysis of



conventional data is better at generating reasonable estimates of parameters with a quantified error, with values within the range of the whole dataset estimate.

In general, the stress step methods unsurprisingly provide the tightest estimates of parameters. Ideally, a larger step size would have been used for the experiments to create more distinction between the Bayesian and conventional protocols and would have minimized the effects of damage or coaxing that may have affected the result. However, in this experiment, no appreciable difference was seen in model parameters between step and staircase results.

Finally, a comparison can be made between distribution models. Accounting for all the tested specimens, there is more evidence for the two parameter Weibull distribution over all other common distributions as shown in Figure 13.

5 | CONCLUSIONS

We show that Bayesian statistics can be applied to fatigue strength testing. We present two open source alternatives to the conventional staircase and stress step protocols. These can be shown to provide better estimates for model parameters such as location and spread, such as mean and standard deviation. Furthermore, these methods are independent of rigid definitions for step size or starting stress, as required in conventional protocols. Using a Monte Carlo simulation framework, we demonstrate the benefit that these protocols bring and validate these through experiment. Furthermore, simulation shows insight into the behavior underlying the conventional protocols, further demonstrating the limitations of fixed step sizes.

Finally, this framework permits comparisons of distribution models, allowing design engineers to make better trade-offs for weight and reliability. We also show with experimental data that a two-parameter Weibull distribution may better fit fatigue strength data and offer the

reader the tool to compare distribution models for other fatigue data.

These tools can be applied to any pass-fail testing application, fatigue or otherwise, and is material independent. The method, built for both the collection of data during experiments as well as the analysis of collected data, is made freely available under the GitHub repository “BayesOptFatigue”.

ACKNOWLEDGMENTS

CMM wishes to acknowledge Rolls Royce plc, the Royal Commission for the Exhibition of 1851, and the Engineering and Physical Sciences Research Council for financial support. This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

AUTHOR CONTRIBUTIONS

Christopher Massimo Magazzeni: Conceptualization, data curation, software, formal analysis, validation, investigation, visualization, methodology, writing—original draft, project administration. **Rory Rose:** Conceptualization, data curation, software, formal analysis, validation, investigation, visualization, methodology, writing—original draft, project administration. **Chris Gearhart:** Conceptualization, resources, software, formal analysis, supervision, methodology, writing—review and editing. **Jicheng Gong:** Investigation, methodology. **Angus J. Wilkinson:** Funding acquisition, writing—review and editing.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in BayesOptFatigue at <https://github.com/cmmagazz/BayesOptFatigue>.

NOMENCLATURE

S	Stress amplitude
θ	Vector of model parameters
μ	Mean
σ	Standard deviation
Φ	Cumulative distribution function
p	Probability
\mathbf{x}	Vector of previous test stresses

\mathbf{j}	Vector of previous test results
$x_{i,k}$	Test stress of specimen i at stress increment k
$j_{i,k}$	Test result of specimen i at stress increment k
f_S	Probability density function of fatigue strength
F_S	Cumulative distribution function of fatigue strength
I	Shannon information
ent	Expected information (Entropy)
γ	Gain in entropy
α, β, c	Basquin constants
U	Utility function
M	Probability model
L	Likelihood function
\hat{L}	Maximum value of the likelihood function
k	Number of parameters in the probability model
n	Number of observations

ORCID

Christopher Massimo Magazzeni  <https://orcid.org/0000-0003-3049-143X>

Rory Rose  <https://orcid.org/0000-0002-4255-279X>

REFERENCES

1. Reed R, Smith J, Christ B. The economic effects of fracture in the United States. *Natl Bur Stand Spec Publ.* 1983;1-7.
2. Nicholas T. Step loading for very high cycle fatigue. *Fatigue Fract Eng Mater Struct.* 2002;25(8-9):861-869.
3. Dixon WJ, Mood AM. A method for obtaining and analyzing sensitivity data. *J Am Stat Assoc.* 1948;43(241):109-126.
4. ASTM. *A tentative guide for fatigue testing and the statistical analysis of fatigue data.* 1958.
5. Pascual FG, Meeker WQ. Analysis of fatigue data with runouts based on a model with nonconstant standard deviation and a fatigue limit parameter. *J Test Eval.* 1997;25(3):292-301.
6. ISO. 12107. Metallic materials—fatigue testing—statistical planning and analysis of data—ISO 12107. *Test.* 2006;2006:13.
7. Nelson W. Accelerated life testing—step-stress models and data analyses. *IEEE Trans Reliab.* 1980;R-29(2):103-108.
8. Morrissey RJ, McDowell DL, Nicholas T. Frequency and stress ratio effects in high cycle fatigue of Ti-6Al-4V. *Int J Fatigue.* 1999;21(7):679-685.
9. Lin SK, Lee YL, Lu MW. Evaluation of the staircase and the accelerated test methods for fatigue limit distributions. *Int J Fatigue.* 2001;23(1):75-83.
10. Pollak RD. Analysis of methods for determining high cycle fatigue strength of a material with investigation of Ti-6Al-4V gigacycle fatigue behavior. *Fac Grad Sch Eng Manag Force Inst Technol.* 2005;278.
11. Engler-Pinto CC, Lasecki JV, Frisch RJ, Allison JE. Statistical approaches applied to very high cycle fatigue. In: Allison JE, Jones JW, Larsen JM, Ritchie RO, eds. *VHCF-4-4th International Conference on Very High Cycle Fatigue.* TMS; 2007: 369-376.
12. Engler-Pinto CC Jr, Lasecki JV, Frisch RJ Sr, et al. Statistical approaches applied to fatigue test data analysis. *SAE Trans.* 2005;114:422-431.

13. Xi L, Songlin Z. Changes in mechanical properties of vehicle components after strengthening under low-amplitude loads below the fatigue limit. *Fatigue Fract Eng Mater Struct*. 2009; 32(10):847-855.
14. Zioupos P, Casinos A. Cumulative damage and the response of human bone in two-step loading fatigue. *J Biomech*. 1998;31(9): 825-833.
15. Hectors K, De Waele W. Cumulative damage and life prediction models for high-cycle fatigue of metals: a review. *Metals (Basel)*. 2021;11(2):1-32.
16. Risitano A, Risitano G. Cumulative damage evaluation of steel using infrared thermography. *Theor Appl Fract Mech*. 2010; 54(2):82-90.
17. Lerch BA, Draper SL, Pereira JM. Conducting high-cycle fatigue strength step tests on gamma TiAl. *Metall Mater Trans a Phys Metall Mater Sci*. 2002;33(12):3871-3874.
18. Bergant M, Werner T, Madia M, Yawny A, Zerbst U. Short crack propagation analysis and fatigue strength assessment of additively manufactured materials: an application to AISI 316L. *Int J Fatigue*. 2021;151:106396.
19. Andreau O, Pessard E, Koutiri I, Peyre P, Saintier N. Influence of the position and size of various deterministic defects on the high cycle fatigue resistance of a 316L steel manufactured by laser powder bed fusion. *Int J Fatigue*. 2021;143:105930.
20. Andreau O, Pessard E, Koutiri I, et al. A competition between the contour and hatching zones on the high cycle fatigue behaviour of a 316L stainless steel: analyzed using X-ray computed tomography. *Mater Sci Eng A*. 2019;757:146-159.
21. Pan X, Guo S, Tian Z, et al. Fatigue performance improvement of laser shock peened hole on powder metallurgy Ni-based superalloy labyrinth disc. *Surf Coatings Technol*. 2021;409:126829.
22. Song Q, Sun C. Mechanism of crack initiation and early growth of high strength steels in very high cycle fatigue regime. *Mater Sci Eng A*. 2020;771:138648.
23. Dehghani M, Tremblay R, Leclerc M. Fatigue failure of 350WT steel under large-strain seismic loading at room and subfreezing temperatures. *Construct Build Mater*. 2017;145:602-618.
24. Mehdizadeh M, Khonsari MM. On the application of fracture fatigue entropy to variable frequency and loading amplitude. *Theor Appl Fract Mech*. 2018;98:30-37.
25. Zhao Z, Wang L, Liu C, Liu L, Chen W. Prediction of high-cycle fatigue performance of 1Cr11Ni2W2MoV stainless steel plate after foreign object damage. *Int J Aerosp Eng*. 2020;2020:1-13.
26. Mackay R, Byczynski G. An evaluation of the staircase and over-stress probe methods for fatigue life characterization in aluminum sand castings. *Int J Met*. April 2021;16(1):1-18.
27. Dallago M, Winiarski B, Zanini F, Carmignato S, Benedetti M. On the effect of geometrical imperfections and defects on the fatigue strength of cellular lattice structures additively manufactured via selective laser melting. *Int J Fatigue*. 2019;124: 348-360.
28. Nourian-Avval A, Khonsari MM. Rapid prediction of fatigue life based on thermodynamic entropy generation. *Int J Fatigue*. 2021;145:106105.
29. Gao JW, Pan XN, Han J, et al. Influence of artificial defects on fatigue strength of induction hardened S38C axles. *Int J Fatigue*. 2020;139:105746.
30. Guo Q, Guo X, Fan J, Syed R, Wu C. An energy method for rapid evaluation of high-cycle fatigue parameters based on intrinsic dissipation. *Int J Fatigue*. 2015;80:136-144.
31. Luo S, He W, Chen K, Nie X, Zhou L, Li Y. Regain the fatigue strength of laser additive manufactured Ti alloy via laser shock peening. *J Alloys Compd*. 2018;750:626-635.
32. Xu ZW, Wu SC, Wang XS. Fatigue evaluation for high-speed railway axles with surface scratch. *Int J Fatigue*. 2019;123: 79-86.
33. Murakami Y, Takagi T, Wada K, Matsunaga H. Essential structure of S-N curve: prediction of fatigue life and fatigue limit of defective materials and nature of scatter. *Int J Fatigue*. 2021; 146:106138.
34. Abroug F, Pessard E, Germain G, Morel F. HCF of AA7050 alloy containing surface defects: study of the statistical size effect. *Int J Fatigue*. 2018;110:81-94.
35. Baragetti S, Božić Ž, Arcieri EV. Stress and fracture surface analysis of uncoated and coated 7075-T6 specimens under the rotating bending fatigue loading. *Eng Fail Anal*. 2020;112: 104512.
36. Hu X, Jia X, Bao Z, Song Y. Effect of notch geometry on the fatigue strength and critical distance of TC4 titanium alloy †. *J Mech Sci Technol*. 2017;31(10):4727-4737.
37. Gao JW, Yu MH, Liao D, et al. Fatigue and damage tolerance assessment of induction hardened S38C axles under different foreign objects. *Int J Fatigue*. 2021;149:106276.
38. Lindley DV. Bayesian statistics. *Society for Industrial and Applied Mathematics*. 1972. <https://epubs.siam.org/doi/abs/10.1137/1.9781611970654>
39. Lindley DV. The choice of variables in multiple regression. *J R Stat Soc Ser B*. 1968;30(1):31-53.
40. Lindley DV. On a measure of the information provided by an experiment. *Ann Math Stat*. 1956;27(4):986-1005.
41. DasGupta A. Review of Optimal Bayes Designs. Technical Report #95-4. Purdue University; 1995. <https://www.stat.purdue.edu/~dasgupta/publications/tr95-04.pdf>
42. Chaloner K, Verdinelli I, Chaloner K IV. Bayesian experimental design: a review. *Stat Sci*. 1995;10(3):273-304.
43. Ryan EG, Drovandi CC, McGree JM, Pettitt AN. Fully Bayesian optimal experimental design: a review. *Queensl Univ Technol*. 2016;1-39.
44. Ryan EG, Drovandi CC, McGree JM, Pettitt AN. A review of modern computational algorithms for Bayesian optimal design. *Int Stat Rev*. 2016;84(1):128-154.
45. Magazzeni CM, Rose R. cmmagazz/BayesOptFatigue: Code for "Bayesian Optimised Collection Strategies for Fatigue Testing Constant Life Testing." Available at: <https://github.com/cmmagazz/BayesOptFatigue>. Accessed September 17, 2021.
46. Smalakys L, Melninkaitis A. Predicting lifetime of optical components with Bayesian inference. *Opt Express*. 2021;29(2): 903-915.
47. Han C, Chaloner K. Bayesian experimental design for non-linear mixed-effects models with application to HIV dynamics. *Biometrics*. 2004;60(1):25-33.
48. Magazzeni CM. Towards local property assessment of advanced manufactured microstructures—a case study in linear friction welded titanium. 2022.

49. Li H, Wen D, Lu Z, Wang Y, Deng F. Identifying the probability distribution of fatigue life using the maximum entropy principle. *Entropy*. 2016;18(4):111.
50. Zhao YX, Yang B. Probabilistic measurements of the fatigue limit data from a small sampling up-and-down test method. *Int J Fatigue*. 2008;30(12):2094-2103.
51. Schneider CRA, Maddox SJ. Commission XIII-WG1 best practice guide on statistical analysis of fatigue data. 2003.
52. Chi-liu S. The statistical analysis of fatigue data. 1994.
53. Grove D, Campean F. A comparison of two methods of analysing staircase fatigue test data. *Qual Reliab Eng Int*. 2008;24(4):485-497.
54. Yadav OP, Rathod V, Rathore A, Jain R. Probabilistic modeling of fatigue damage accumulation for reliability prediction. *Int J Qual Stat Reliab*. 2011;2011:1-10.
55. Zhang J, Kececioglu DB. *Technologies IS of S and A. New Approaches to Determine the Endurance Strength Distribution*. ISSAT; 1998.
56. Choucair H, Jawad B. Analyzing field failures of engine valve springs in presence of nonmetallic inclusions by applying statistical and fracture mechanics models. *SAE Int J Mater Manuf*. 2009;2(1):265-273.
57. Anderson KV, Daniewicz SR. Statistical analysis of the influence of defects on fatigue life using a Gumbel distribution. *Int J Fatigue*. 2018;112:78-83.
58. Hor A, Saintier N, Robert C, Palin-Luc T, Morel F. Statistical assessment of multiaxial HCF criteria at the grain scale. *Int J Fatigue*. 2014;67:151-158.
59. Schönning A, Neal J, Eason PD, Tiryakioğlu M, Logan ZS. On fatigue life variability in bone cement. *Mater Sci Technol*. 2020;36:1695-1703.
60. El Khoukhi D, Morel F, Saintier N, et al. Experimental investigation of the size effect in high cycle fatigue: role of the defect population in cast aluminium alloys. *Int J Fatigue*. 2019;129:105222.
61. Hoole J, Sartor P, Booker J, Cooper J, Gogouvitis XV, Schmidt RK. Systematic statistical characterisation of stress-life datasets using 3-parameter distributions. *Int J Fatigue*. 2019;129:105216.
62. Never BT. AD-Optimality-Based Sensitivity Test. *A D-optimality-based sensitivity test Technometrics*. 1994;36(1):61-70.
63. Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc*. 1995;90(430):773-795.
64. Box GEP, Tiao GC. *Bayesian inference in statistical analysis*. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 1992.
65. Fox EP, Sivia DS. Data analysis: a Bayesian tutorial. *Dent Tech*. 1998;40:155-156.

How to cite this article: Magazzeni CM, Rose R, Gearhart C, Gong J, Wilkinson AJ. Bayesian optimized collection strategies for fatigue strength testing. *Fatigue Fract Eng Mater Struct*. 2023;46(1):228-243. doi:[10.1111/ffe.13859](https://doi.org/10.1111/ffe.13859)

APPENDIX A.

A.1 | Choice of initial prior

When choosing an initial prior for inference based on the normal distribution, two non-informative priors were investigated. As suggested by Box and Tiao,⁶⁴ the Jeffreys prior Equation (A1) promotes models with small standard deviations.

$$P(\theta|\mathbf{x}) \propto \sigma^{-1} \quad (\text{A1})$$

Alternatively, a constant prior over a range is used, which has the advantage that it is mathematically equivalent to a maximum likelihood approach. The choice of using a flat prior, as opposed to the Jeffreys prior, is justified by both its simplicity and notable feature in the trend of Shannon information at low specimen number.^{64,65} When using a Jeffreys prior, the Shannon information drops significantly after approximately 10 samples (Figure A1). This is generally observed as an initial under-estimation of the standard deviation, in line with the prior construction, followed by an increase in the estimate and uncertainty in the standard deviation. This overconfidence in a low standard deviation was deemed, particularly in the context of fatigue testing, as sub-optimal, despite the subsequent faster convergence to the

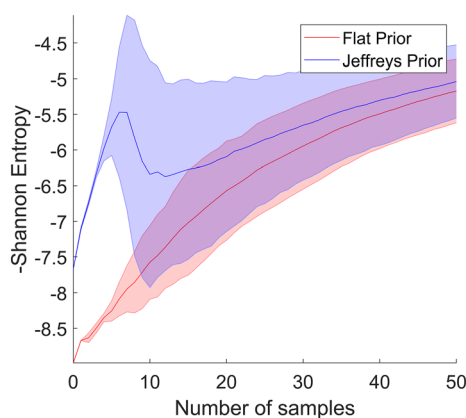


FIGURE A1 Comparison of negative Shannon entropy between tests with priors initialized as uniform probability (red) and non-informative (blue). The tests were performed on a simulated set with strength 400 MPa and distribution width 40 MPa at 10^6 cycles for 100 specimens using the Bayesian staircase protocol. The shaded error bars represent the standard deviation between 500 repeats [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

same parameter values. For these reasons, a fully flat prior was used in the testing and simulation reported in this work.

However, it must be clear that a flat prior is also not assumption free. Rather, the standard deviation is deemed equally likely to lie within the extent that is specified in the permitted 2D parameter space.

A.2 | Variability in step testing with large step sizes

As discussed in the main text, step sizes in step testing should be as large as possible to minimize the effects of coaxing and damage on the fatigue failure distribution observed from a test campaign. However, the larger the step size, the higher the chance that the failure stress is “jumped over” during testing: specimens always survive and fail at the same stress levels. 1515.

This can also be understood as placing a large importance on the initial test stress: if only few of the subsequent test levels intersect the distribution, it is easier to introduce bias. This bias produces a bimodal distribution in the MAP estimate with the modal value sometimes far away from the mean value (Figure A2), making the mean error for step testing not representative of what an experiment campaign is likely to observe.

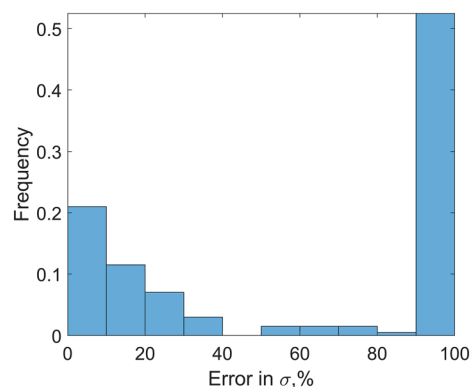


FIGURE A2 Histogram of errors in conventional step testing after five samples, with a step size 5 times the true standard deviation of the simulation dataset, at 200 repeats with a random starting stress between 0 MPa and the step size [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]