



Simulation-based inference for reliability model selection and parameter calibration: An application to fatigue

Mohamed Anis Ben Abdessalem

► To cite this version:

Mohamed Anis Ben Abdessalem. Simulation-based inference for reliability model selection and parameter calibration: An application to fatigue. Quality and Reliability Engineering International, 2024, 40 (4), pp.1855 - 1874. 10.1002/qre.3497 . hal-04606922

HAL Id: hal-04606922

<https://univ-angers.hal.science/hal-04606922>

Submitted on 10 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Simulation-based inference for reliability model selection and parameter calibration: An application to fatigue

Anis Ben Abdessalem

LARIS, SFR MATHSTIC, Univ Angers,
Angers, France

Correspondence

Anis Ben Abdessalem, LARIS, SFR
MATHSTIC, Univ Angers, F-49000
Angers, France.

Email:

anis.ben-abdessalem@univ-angers.fr

Abstract

Selecting a “useful” model among a set of competing candidate models is a common problem in many disciplines and particularly in engineering applications. For a modeller, the best model fulfills a certain purpose best (e.g., the ability to accurately predict new data, good interpolation, precise estimates of lower/upper percentiles, etc.), which is typically assessed by comparing model simulations to the observed data. Approximate Bayesian computation (ABC) which is a statistical learning technique is used in this work for reliability model selection and parameter calibration using small/moderate fatigue life data sets. This is always the case in material fatigue due to the high cost and low efficiency of fatigue tests which are bottleneck problem when designing components/structures with regard to fatigue. The ABC is a likelihood-free based method which means that it needs only a generative model to simulate the data. The proposed method provides a formal rank of the competing reliability models by eliminating gradually the least likely models in a parsimonious manner. It is flexible since it offers the possibility to use different metrics to measure the discrepancy between simulated and observed data. The choice of the appropriate distance function depends essentially on the purpose of model selection. Through various examples, it has been demonstrated that the ABC method has a number of very attractive properties, which makes it especially useful in probabilistic risk assessment and reliability analysis.

KEY WORDS

approximate Bayesian computation, fatigue life distributions, low percentiles, probabilistic fatigue design, reliability model selection, risk assessment, simulation-based inference

1 | INTRODUCTION

Model selection aims at identifying the most appropriate model(s) among a set of competing models with respect to a given calibration data set. Despite the progress being made, model selection is still a subject of debate and a major challenge for statisticians and practitioners. The problem is statistically and numerically more challenging when the data is limited.

This is an open access article under the terms of the [Creative Commons Attribution License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Quality and Reliability Engineering International* published by John Wiley & Sons Ltd.

As it is well known, it is the requirement of fatigue reliability analyses to accurately determine the probability model of a group of fatigue lives. Since the data are generally seldom sufficient to define a statistical distribution, we normally must rely on assumed distributions which arise from convenience or, preferably, from relevant physical arguments. Among the assumed distributions, which have been widely used in fatigue studies, are lognormal (LN), Birnbaum-Saunders and Weibull distributions,^{1,2} to mention a few. These distributions are very similar but the predictions at lower/upper tails which are usually the areas of interest for inference could be significantly different, for this reason a robust discrimination technique according to a desired goal could be a useful alternative for practitioners.

Many different approaches have been proposed in the statistical literature to help select the “best” model among a group of competing hypotheses. The Bayesian solution originates from Jeffreys^{3,4} and builds on the principles of Occam’s razor is arguably the most popular and widely adopted in different domains particularly in fatigue.^{5–10} Bayesian inference is aimed at computing a posterior probability distribution over a set of hypotheses or models, in terms of their relative support from the data. In this discussion I shall present some of them. The Bayes factor (BF) proposed by Jeffreys et al.¹¹ is referred to as the standard Bayesian solution to the hypothesis testing and model selection problems¹² and the primary tool used in Bayesian inference for hypothesis testing and model selection.¹³ A comprehensive review of BFs is presented by Kass and Raftery.¹⁴ The idea is to select the model that presents the best point estimate of the posterior density ratios. Nevertheless, the main drawback of BFs is their sensitivity to the choice of priors as well documented in the literature. The method is also problematic if there are many possible models. To overcome the issues with the BF, other variants including partial BF,¹⁵ the intrinsic BF,¹⁶ and the fractional BF¹⁷ have been proposed in the literature. These variants basically split the data into a training sample and a testing sample. The training sample is used to update an uninformative prior to obtain an informative prior. Unfortunately, they suffer from more or less arbitrary choices of training samples, weights for averaging training samples, and fractions, respectively. Model selection criterions based on loss of information, the well-known are the Akaike information criterion (AIC),¹⁸ Akaike information criterion with correction (AICc),¹⁹ Bayesian information criterion (BIC)²⁰ and Hannan-Quinn information criterion (HQIC)²¹ have been widely used as well. They are based on a penalization of the likelihood function as the model becomes more complex, that is, models with more parameters. Following Spiegelhalter et al.,²² the model whose information criterion has a smaller value is the better. This is coherent with the principle of Occam’s razor which states that unnecessarily complex models should not be preferred to simpler ones. The principle states that among competing theories that lead to the same prediction, the one that relies on the fewest assumptions is the best. In particular, when choosing among a set of models, the simplest valid model is the best choice. Exactly, how to quantify simplicity has been a subject of debate for centuries. In this study, the simplicity of a model is assumed to be determined by the number of parameters in the model. An inherent problem with these criteria is that they do not allow prior input for model choice. In a somewhat similar spirit, Spiegelhalter et al.²² devised a selection criterion, called the deviance information criterion (DIC) based on Bayesian measures of the complexity level and of how well the model fits the data. A lower value of DIC indicates a better model fit. The major limitation of these approaches is that they undertake separately the competing models.

The reversible jump Markov chain Monte Carlo (RJMCMC) method, originally given by Green,²³ is another strategy that samples over the model and parameter spaces. The RJMCMC method offers an across-model simulation approach for Bayesian estimation and model comparison, by exploring the sampling space that consists of several models of possibly varying dimensions. A naive implementation of RJMCMC to models like Gibbs random fields suffers from computational difficulties. To overcome the inefficiency of a standard RJMCMC algorithm, a noisy RJMCMC sampler²⁴ has been developed. The Transitional Markov Chain Monte Carlo (TMCMC) algorithm proposed by Ching et al.²⁵ has been proven to outperform other approaches for its capability of estimating model evidence and smooth convergence merit. However, it has potential problems in tackling higher dimension parameters,²⁶ since its intermediate stage number will augment for the need of more samples while the accuracy of estimators may decrease with increasing parameter dimension.²⁵ It should be noted that most of the proposed selection procedures converge to a single model. However in some circumstances more than one model could explain the data. Thus, there is no need to choose one model and it is possible to average the predictions from several models. Bayesian Model Averaging (BMA) is a technique designed to incorporate the uncertainty inherent in the model selection process, which is often ignored by traditional statistical analysis. As a statistical procedure to infer consensus predictions, BMA method weighs individual predictions based on their posterior model probabilities, with the better performing predictions receiving higher weights than the worse performing ones. BMA method can thus generate an averaged model, especially in cases where more than one model has a non-negligible posterior probability.²⁷ BMA is ideally suited to guide this search, because it implicitly honors the principle of parsimony.^{4,28} The BMA ranking reflects an optimal tradeoff between goodness-of-fit and model complexity, with model complexity being encoded in the prior probability distributions of the model parameters. Another selection

procedure that is popular is the Nested sampling (NS) which is a highly efficient and easily implemented sampling algorithm that has been successfully incorporated into Bayesian inference for model updating and model selection. The key step of this algorithm lies in proposing a new sample in each step that has a higher likelihood to replace the sample that has the lowest likelihood evaluated in the previous iteration. This process is a constrained sampling step which has significant impact on the algorithm efficiency. It has been successfully applied to cosmological field,²⁹ finite element model updating and model selection,³⁰ and surface flow problems.^{31,32} The main problem of the standard NS algorithm lies in the constrained step in which a new sample with higher likelihood is proposed to replace the sample with the lowest likelihood. This issue becomes even more challengeable after several iterations as the likelihood has reached a higher value and the parameter space shrinks to a very sharp region.

In this paper, a simulation-based framework that simultaneously addresses parameter calibration and model selection using the Approximate Bayesian computation (ABC) method is proposed. It offers the possibility to compare between several reliability models and to estimate the marginal distributions of the model parameters in a straightforward way. The ABC method belongs to the class of sequential particle filter methods which does not require the specification of a likelihood function. Its basic idea is to compare simulated data to the observed data using an appropriate distance/metric; if their difference is small, then the parameter values that generate the simulated data are selected as a sample from the posterior distribution. Successful applications of the ABC algorithms can be found in various areas including structural dynamics,^{33,34} biology,³⁵ epidemiology³⁶ and cosmology,³⁷ to mention a few. The uniqueness of this study comes from the fact that thus far, no attempt has been made to compare several competing models simultaneously by using a variety of metrics/distances. It should be noted that the ABC algorithms take into account the complexity of the model (i.e., parameter number) automatically without the need to define a penalty term. In other words, the ABC algorithms promote systematically the selection of the simpler model and switches to complex models only when a high accuracy is required. It is similar to the classical penalised information criterions based on the log-likelihood whose penalise complex models, especially to discourage overfitting. In addition, we will demonstrate through this study that the proposed strategy is robust against model mis-specification (or at least could mitigate the effects of model mis-specification) compared with the classical likelihood-based methods.

The outline of the paper is as follows. Section 2 introduces the Bayesian paradigm for parameter estimation and model class selection and sets out the necessary background on ABC algorithms. The distance functions/metrics that can be used for Bayesian inference are given in the same section. In Section 3, we present a motivating problem with a focus on the merits of the ABC method. In Section 4, we present three applications in which we propose several competing statistical distributions to model lifetime data and we illustrate the procedure for selecting the most likely reliability model. Section 5 summaries the conclusions of this paper and discusses future extensions of this research.

2 | SIMULATION-BASED INFERENCE USING ABC

2.1 | Background and details

ABC is employed in this work for simultaneous model selection and parameter estimation.^{35,38,39} Let us consider first the problem of parameter estimation,^{35,40,41} where θ is the vector of model parameters. Thus, given the prior distribution for the parameters, $\pi(\theta)$, the objective is to approximate the posterior distribution given by:

$$\pi(\theta|D) \propto \mathcal{L}(D|\theta)\pi(\theta)$$

where $\mathcal{L}(D|\theta)$ is the likelihood function (conditional probability of observed data D given the parameters θ).

A simpler ABC algorithm for parameter estimation, which avoids the requirement of evaluating likelihood functions directly $\mathcal{L}(D|\theta)$ in order to obtain the posterior distributions of unknown parameters, can be summarized in general terms as follows. By starting with $k = 0$, repeat the following steps until a sufficiently large number of samples θ^k are accepted:

- **Step 1:** sample a parameter vector θ^* from the prior distribution $\pi(\theta)$;
- **Step 2:** generate simulated data D_* from a generative model (in this study, a parametric probability model);
- **Step 3:** compare the simulated data, D_* , with the observations, D , utilising a user-selected distance function $\rho(D, D_*)$ and a user-selected tolerance ε (the tolerance $\varepsilon > 0$ is the level of desired agreement between D and D_* , measured by $\rho(D, D_*)$);
- **Step 4:** if $\rho(D, D_*) \leq \varepsilon$, accept θ^* , that is, make $k = k + 1$ and then set $\theta_k = \theta^*$. Otherwise, return to **Step 1**.

Each output θ^k of the above algorithm is a sample of the distribution $P(\theta|\rho(D, D_*) \leq \varepsilon)$. If ε is sufficiently small and $\rho(D, D_*)$ is appropriately selected, it is expected that this distribution would be a good approximation to the true posterior distribution $\pi(\theta|D)$. Therefore, difficulties in the successful implementation of the above algorithm obviously include the specification of the distance function $\rho(D, D_*)$, as well as of the tolerance ε .

The concepts outlined in the above algorithm can be extended to incorporate model selection in addition to parameter estimation.^{35,42,43} The marginal posterior distribution of a model M_i from the given observations is given by:

$$\pi(M_i|D) = \frac{\pi(D|M_i)\pi(M_i)}{\pi(D)} \quad (1)$$

where $\pi(D|M_i)$ is the marginal likelihood distribution and $\pi(M_i)$ is the prior probability function for model $M_i, i = 1, 2, \dots, n$. The posterior distribution in the combined space of models and parameters, which is sought in this work, is $\pi(\theta_i, M_i|D)$, where θ_i is the vector of parameters for model M_i . Such posterior distribution is approximately represented in this work by a set of samples obtained from $P(\theta_i, M_i|\rho(D, D_s) \leq \varepsilon)$, using a more efficient sampling technique. It should be noted that the acceptance rate of ABC in its basic form is computationally prohibitive for small ε . Several variants have been proposed in the literature to overcome this issue and to gain in efficiency. In this work, the ABC algorithm based on an ellipsoidal sampling technique and a reweighing scheme is used.³⁴ It can be summarised in Algorithm 1.

The algorithm starts by selecting a candidate model from the competing models supposed to be equally probable. A particle is then sampled from the prior over the model parameters, simulating the data using the model and accepted the particle if $\rho(D, D_*) \leq \varepsilon_1$). The process is repeated until N particles distributed over the competing models are obtained. The particles are then weighted using a kernel and normalised according to each model. For the next population, a tolerance threshold ε_2 is defined based on the discrepancy values ranked in descending order. For each model, the particles with $\rho(D, D_*) > \varepsilon_2$ are dropped. From the retained particles, a proportion are randomly selected and propagated to the next population. The remaining particles are then enclosed in an ellipsoid in which the mass centre and the covariance matrix are based on the propagated particles. The ellipsoid are then enlarged by a factor f_0 to ensure that the particles on the borders will be inside. Finally, the population is replenished by re-sampling particles inside the ellipsoid associated to each model. In the subsequent iterations, the threshold is updated adaptively in the same way and samples selection are subjected to more stringent threshold. The priors on the models are also updated when one of the competing model is eliminated. Through the populations and as $\varepsilon \rightarrow 0$, a larger number of particles are selected for the most likely model(s) and the samples for the parameters better reflect the real posterior distributions as the most interesting region in the parameter space is well identified. Several stopping criterions could be used to stop the algorithm. In this work, the algorithm is stopped when the difference between two successive tolerance threshold values falls below a pre-specified value. Let denote by ξ this value. A detailed discussion concerning the effects of these settings can be found in Ref. [34].

At the last population, the algorithm produces a Markov chain on (\mathcal{M}, θ) for which the marginal distribution is $\pi(\mathcal{M}|D)$. The posterior model probability can then be estimated by:

$$\hat{\mathbb{P}}(M_i|D) = \frac{1}{N} \sum_{k=1}^N \mathbb{I}(M^{(k)} = M_i), \quad k = 1, \dots, n \quad (2)$$

As shown in Algorithm 1, the ABC-NS algorithm requires to choose a number of hyper-parameters. In this study, the algorithm is implemented with the following hyper-parameters:

- the initial threshold ε_1 is set to 1000.
- the number of particles per population is set to 1000.
- the proportion of dropped particles and surviving particles are set respectively, to 30% and 60%.
- The ellipsoid enlargement factor f_0 , is set to 1.1.
- The stopping criterion threshold value is equal to 10^{-6} .

For further detail about ABC-NS in general as well as the concrete implementation, the interested reader can refer to the following examples released in GitHub (the same notation has been used here): ABC-NS for model calibration and ABC-NS for model selection.

ALGORITHM 1 ABC-NS for model selection and parameter estimation

Require: data: D , initial tolerance: ε_1 , distance function: $\rho(\cdot)$, prior distributions: $p(\theta)$, tuning parameters: (α_0, β_0, f_0) , N : number of particles, T : number of populations

```

1: At iteration  $t = 1$ 
2: for  $i = 1$  to  $N$  do
3:   repeat
4:     Draw  $\mathcal{M}^*$  from the prior  $p(\mathcal{M})$ 
5:     Sample  $\theta^*$  from the prior  $p(\theta|\mathcal{M}^*)$ 
6:     Simulate a candidate data set  $D^* \sim F(\cdot|\theta^*, \mathcal{M}^*)$ 
7:   until  $\rho(D, D^*) \leq \varepsilon_1$ 
8:   Store  $\mathcal{M}_t^{(i)} = \mathcal{M}^*$ ,  $\theta_t^{(i)} = \theta^*$ ,  $e_t^{(i)} = \rho(D, D^*)$ 
9: end for
10: Evaluate weights for all particles:  $\omega_i = \frac{1}{\varepsilon_1}(1 - (\frac{e_i}{\varepsilon_1})^2)$ ,  $i = 1, \dots, N$ 
11: Sort  $e_t^{(i)}$  in descending order and store them.
12: Compute the next tolerance threshold  $\varepsilon_2 = e_t^{(\alpha_0 N)}$ 
13: Drop particles with  $\rho(D, D^*) \geq \varepsilon_2$ ,  $\{\omega_k\}_{k=1}^{\alpha_0 N} = 0$   $\triangleright$  A weight of zero is assigned to the dropped particles
14: Normalise the weights such that  $\sum_{\ell=1}^{(1-\alpha_0)N} \omega_\ell = 1$ 
15: Select  $\mathcal{A}_t = \beta_0 N$  particles from the remaining based on the weights
16: For each model, define the ellipsoid by its centre of the mass and covariance matrix  $\mathcal{E}_t^{(k)} = \{\mu_t^{(k)}, C_t^{(k)}\}$ ,  $k = 1, \dots, \kappa$ 
17: Enlarge the ellipsoids by  $f_0$   $\triangleright$  The same notation for the updated ellipsoids is kept
18: for  $t = 2$  to  $T$  do
19:   for  $j = 1$  to  $(1 - \beta_0)N$  do
20:     repeat
21:       Draw  $\mathcal{M}^*$  from the prior  $p(\mathcal{M})$   $\triangleright$  updated through the iterations
22:       Sample  $\theta^*$  inside the ellipsoid  $\mathcal{E}_{t-1}^*$ 
23:       Simulate a candidate data set  $D^* \sim F(\cdot|\theta^*, \mathcal{M}^*)$ ,
24:     until  $\rho(D, D^*) \leq \varepsilon_t$ 
25:     Store  $\mathcal{M}_t^{(j)} = \mathcal{M}^*$  and add  $\theta_t^{(j)} = \theta^*$  to the population of particles of  $\{\theta(\mathcal{M}^*)\}$ ,  $e_t^{(j)} = \rho(D, D^*)$ 
26:   end for
   Store the new particles in  $\mathcal{S}_t$ 
   Obtain the new particle set for each model,  $\mathcal{N}_{new} = [\mathcal{A}_{t-1}; \mathcal{S}_t]$  with their correspondent distance value  $e^t$ 
   Sort  $e^t$  and define  $\varepsilon_{t+1} = e_t^{(\alpha_0 N)}$ 
   Compute new weights for all particles as in step (10) and normalise them
   Define the new set of selected particles  $\mathcal{A}_t$  as in step (15)
   Update the hyper-parameters of the ellipsoid using  $\mathcal{A}_t$ ,  $\mathcal{E}_t = \{\mu_t^{(k)}, C_t^{(k)}\}$ ,  $k = 1, \dots, \kappa$   $\triangleright$  The same factor is used for all the simulations
27: end for
Ensure: Model posterior probabilities, Marginal distributions

```

2.2 | Minimum distance estimators

A key ingredient in ABC procedure is the choice of a discrepancy function that describes how different the simulated and observed data are. In this section we give the computational forms of some of the main distance functions which can be used to deal with reliability model selection and parameter calibration.

- **The Cramer-von-Mises distance**

The Cramer-von-Mises (CM) distance measures the distance between the cumulative density function of the derived distribution against the dataset's cumulative histogram. It is defined by:

$$W_n^2 = n \int_{-\infty}^{+\infty} [F(t) - S_n(t)]^2 \Psi(t) dF(t) \quad (3)$$

where $\Psi(t)$ is a weight function such that $\Psi(t) = 1$. For given ordered sample, t_1, t_2, \dots, t_n , from a statistical distribution, the distance can be written as follows:

$$\rho_{CM} = \frac{1}{12n} + \sum_{i=1}^n \left[z_i - \frac{2i-1}{2n} \right]^2 \quad (4)$$

where $z_i = F(t_i)$ is the predicted value for the i th observation. In the following, the same alphabets denote the same meaning.

- **The Anderson-Darling distance**

The Anderson-Darling (AD) distance first introduced by Anderson & Darling⁴⁴ to place more weight at the tails of the distribution.⁴⁵ It is defined by the following equation:

$$A_n^2 = n \int_{-\infty}^{+\infty} [F(t) - S_n(t)]^2 \Psi(t) dF(t) \quad (5)$$

where $\Psi(t) = [F(t)(1 - F(t))]^{-1}$ is a non-negative weight function. Equation (5) can be written for a finite data sample as:

$$\rho_{AD} = -n - \frac{1}{n} \sum_{i=1}^n (2i-1)[\ln(z_i) + \ln(1-z_{n+1-i})] \quad (6)$$

Two modified distances of the AD distance called ADL and AD2L which give more weight to the left-tail have been tested. They are computed by the following formulas:

$$\rho_{ADL} = -\frac{3n}{2} + 2 \sum_{i=1}^n z_i - \frac{1}{n} \sum_{i=1}^n (2i-1) \ln z_i \quad (7)$$

$$\rho_{AD2L} = 2 \sum_{i=1}^n \log z_i + \frac{1}{n} \sum_{i=1}^n \frac{2i-1}{z_i} \quad (8)$$

Compared to the ρ_{AD} distance, the ρ_{ADL} distance gives more weight to the left-tail region. Similarly, the ρ_{AD2L} assigns larger weight to the left-tail region of the distribution compared to the ρ_{ADL} .

3 | MOTIVATING PROBLEM

In this section, I shall illustrate the potentiality offered by the ABC-NS algorithm for model selection and parameter calibration by means of real data. The advantages of the simulation-based inference method over the most popular maximum likelihood (ML) based-methods will be highlighted.

3.1 | Example 1. “endosulfan” data set

In this example a data named “endosulfan”⁴⁶ is used to illustrate some of the merits of the ABC method very important in risk assessment and reliability analysis particularly for designing against fatigue. Through this example, we will show that the ABC method is flexible, robust against model mis-specification or at least, it can mitigate the effects of model mis-specification through an appropriate choice of a distance function/metric and can be tailored according to the user’s interest.

To demonstrate the usefulness of the ABC, we intend to estimate lower percentiles in a misspecified setting and we make a comparison with a likelihood-based method. It is worthwhile to note that percentiles of the fatigue lifetime distribution are often used as an indicators of reliability and structure safety. Consequently a precise estimates of some important percentiles is of paramount importance for designer. This data set used in this example is taken from Refs. [46] which contains acute toxicity values for the organochlorine pesticide endosulfan, tested on Australian and non-Australian

TABLE 1 AIC and BIC scores for the different competing models.

Information criterion	Lognormal	Loglogistic	Pareto	Burr
AIC	1069	1069	1048	1046
BIC	1074	1075	1053	1054

Abbreviations: AIC, Akaike information criterion; BIC, Bayesian information criterion.

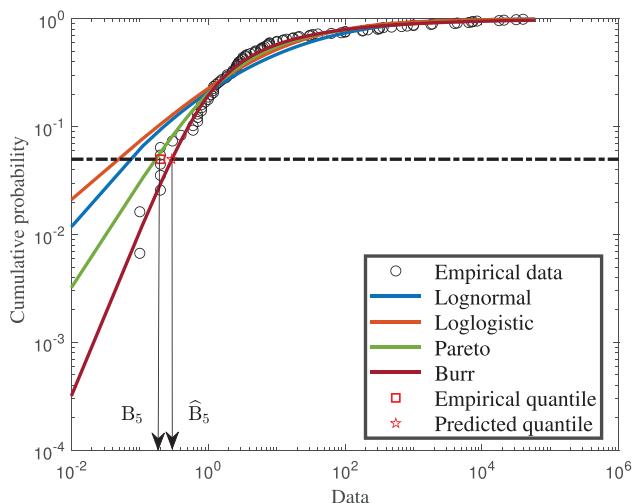


FIGURE 1 Empirical data and theoretical CDFs plot to compare the fit of four distributions with ML estimates (B_5 and \hat{B}_5 indicate the empirical and estimated 5th quantile, respectively.). CDF, cumulative distribution function; ML, maximum likelihood.

laboratory-species. A LN or a log-logistic distribution is often fitted in order to characterise the species sensitivity distribution for a pollutant. A low percentile of the fitted distribution, generally the 5% percentile, is then calculated and called the hazardous concentration 5% (HC5). It is interpreted as the value of the pollutant concentration protecting 95% of the species.⁴⁷ In addition, the two-parameter Pareto distribution and the three-parameter Burr distribution have been fitted. The specifications of the competing reliability models are given in Appendix A. The four competing models are the same considered in Ref. [48]. Model selection is first conducted by using a likelihood-based approach and through computing two information criterions for comparison purpose. Our analysis will incorporate both Akaike's and BICs. AIC balances between the complexity of the model and the statistical goodness-of-fit of the model by imposing a penalty for increasing the number of parameters in the reliability model. It is defined as:

$$\text{AIC} = -2\mathcal{L}(\hat{\theta}) + 2p \quad (9)$$

where $\mathcal{L}(\hat{\theta})$ is the maximised log-likelihood function and p is the number of parameters in the model.

The preferred model is the one corresponding to the lowest index. BIC is an improvement of the AIC in the sense that BIC factors in the size of the sample data in determining the amount of penalty to impose on a model due to increased number of parameters. It has been argued to penalise overfitting more effectively than the AIC measure. It is defined as:

$$\text{BIC} = -2\mathcal{L}(\hat{\theta}) + p \ln(n) \quad (10)$$

where $\mathcal{L}(\hat{\theta})$ and p are as defined above and n is the sample size of the data.

As in AIC, the preferred model is the one corresponding to the lowest index. As one can see from Table 1, the AIC and BIC criterions give the preference to the Burr distribution or the Pareto distribution. The choice between the two distributions seems thus less obvious. Looking at classical penalised criterions based on the log-likelihood seems thus also interesting, especially to discourage over-fitting. Figure 1 shows the data and the fitted cumulative distribution function (CDF), it can be seen that the left-tail seems to be better fitted by the Burr distribution. Its use could then be considered to estimate the HC5 value as the 5% quantile of the distribution. The LN and loglogistic reliability models seem the least likely models for this data set.

The ABC-NS algorithm is now employed to deal first with parameter calibration. We compare the performance of the Burr distribution (using the ML estimates) with the LN distribution calibrated using the distances given in Section 2.2 in

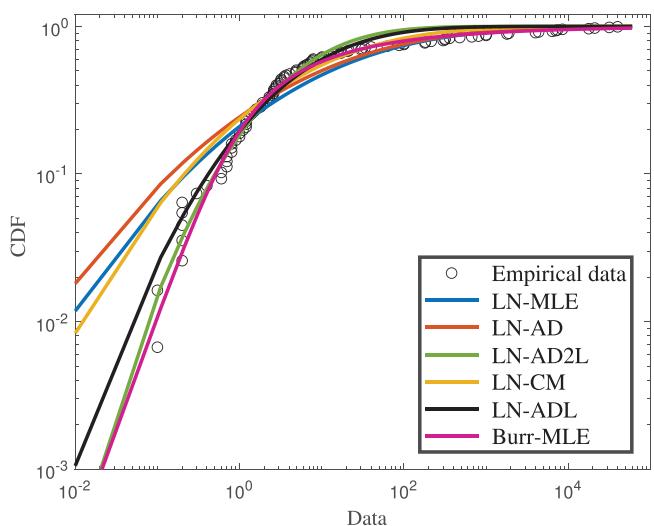


FIGURE 2 Empirical data and theoretical CDFs plot using different models and different distance functions. CDF, cumulative distribution function.

TABLE 2 Empirical and predicted percentiles using a mis-specified reliability model coupled with different distances.

Empirical percentile	Burr model	Mis-specified model with different distances			
		ρ_{CM}	ρ_{AD}	ρ_{ADL}	ρ_{AD2L}
0.2	0.2939	0.0783	0.0440	0.1959	0.2588

a simple misspecified setting. Based on the fact that many practitioners prefer to use simpler models, the LN distribution is selected to fit the data and to predict the HC5 percentiles. We investigate now the efficiency of the ABC method for mitigating model mis-specification. We will show that by coupling a “wrong” model with an appropriate distance function measuring the similarity between the real and simulated data, one may get good predictions as well. The choice of the distance depends on the desired stated goal.

Figure 2 displays the empirical data and the fitted CDFs using the best model following the IC and the LN models coupled with the different distances. One can clearly see from Figure 2 that the ML estimates cannot provide a good fitting quality to the empirical data. In contrast the use of the mis-specified model coupled with an appropriate distance function could provide a better fitting quality to the data. Through the following example, it has been demonstrated that the likelihood-free approach using a suitable distance is robust against model mis-specification.

Comparing the 5% percentiles (HC5) calculated using these fits to the one calculated from the ML estimates fit to the Burr distribution, we can observe on this example that fitting the LN distribution by maximising left-tail AD distances of first or second order enables to approach the value obtained by fitting the Burr distribution using ML estimates. Table 2 displays the empirical and computed quantiles given by the best model based on the information criterions and the mis-specified model with different distances. It has shown that the mis-specified model coupled with an appropriate distance (i.e., ρ_{ADL} , ρ_{AD2L}) makes good predictions much better than the model that has been initially identified as the most likely model (the Burr distribution). For this data set, the distances ρ_{CM} and ρ_{AD} show poor predictions. In conclusion, the proposed likelihood-free method seems interesting when precise predictions of some percentiles is required which is a main interest in fatigue.

4 | APPLICATION TO FATIGUE DESIGN

4.1 | Example 1. 16Mn steel welded specimens

In the first example, the test data of two kinds of 16Mn steel welded specimens whose tests were performed at Zhengzhou Mechanical Institute^{49–51} are taken. The dimensions of the two kinds of specimens, machined under CO₂ gas protective welding procedures, are given in Figure 3. The tests was carried out under the four-point bending sine wave loading

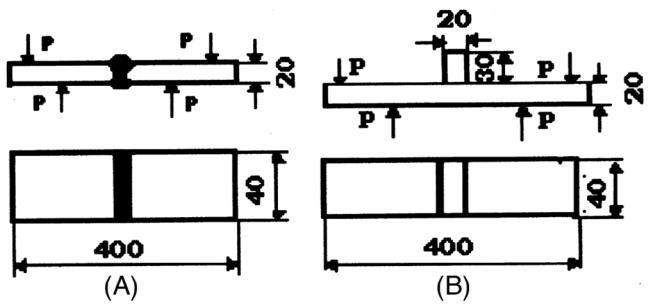


FIGURE 3 Schematic diagrams of the two kinds of 16Mn steel welded specimens.

TABLE 3 Details of the test arrangements.

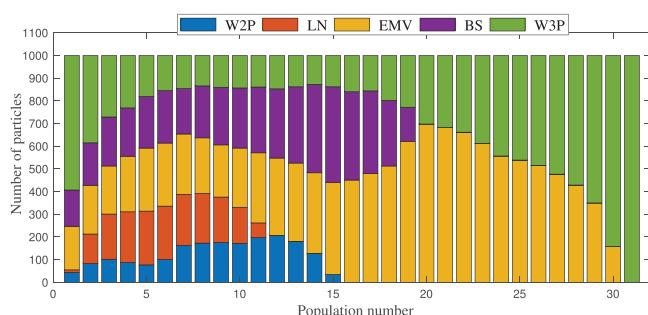
Ordinal of group, k	1	2	3	4	5
Stress range, $\Delta\sigma$ (MPa)	320	290	260	230	202
Number of specimen, n	6	7	9	10	9

TABLE 4 Fatigue life data of the 16Mn steel welded plate specimens.

N_i , Cycles										
k	1	2	3	4	5	6	7	8	9	10
1	290290	297000	327860	339560	406800	564400				
2	359900	418680	440400	459500	499040	528480	635000			
3	274000	501640	522270	685300	690390	730000	750000	894670	980580	
4	436000	555000	786000	818000	838000	940000	1029000	1073000	1282000	1724000
5	912000	940000	1032000	1280000	1530000	1630000	2315000	2754000	3633000	

mode and stress ratio of 0.1. The test arrangements, including the specimen joint pattern, ordinal of group k , number of specimens n and the corresponding stress amplitude level $\Delta\sigma$, are given in Table 3.

Cycles to failure is defined to be the cycles for which the maximum of crack grows to the half the specimen thickness. At this time, the deformation of specimens is too large to continue the normal testing and for the safety of the structure, this definition is necessary for avoiding disaster accident. The test results of the considered specimen are given in Table 4 (for illustrative purpose, only the tested welded specimens shown in Figure 3A are considered). For the data collected at each stress range, one aims to identify a suitable reliability model from a number of competing models. Five competing models widely used in fatigue have been considered: the LN, the two-parameter Weibull (W2P), the three-parameter Weibull (W3P), the extreme maximum value (EMV) and the Birnbaum-Saunders (BS). The probability density function (PDF) and the CDF of the candidate models are given in Appendix B. The ABC-NS algorithm using the different distance functions is now employed to discriminate between the candidate reliability models. Equal prior probabilities are assigned to the competing models and the same stopping criterion is used as in the previous example. Figure 4 shows the distributions

FIGURE 4 Particles repartition over the populations using ρ_{ADL} distance.

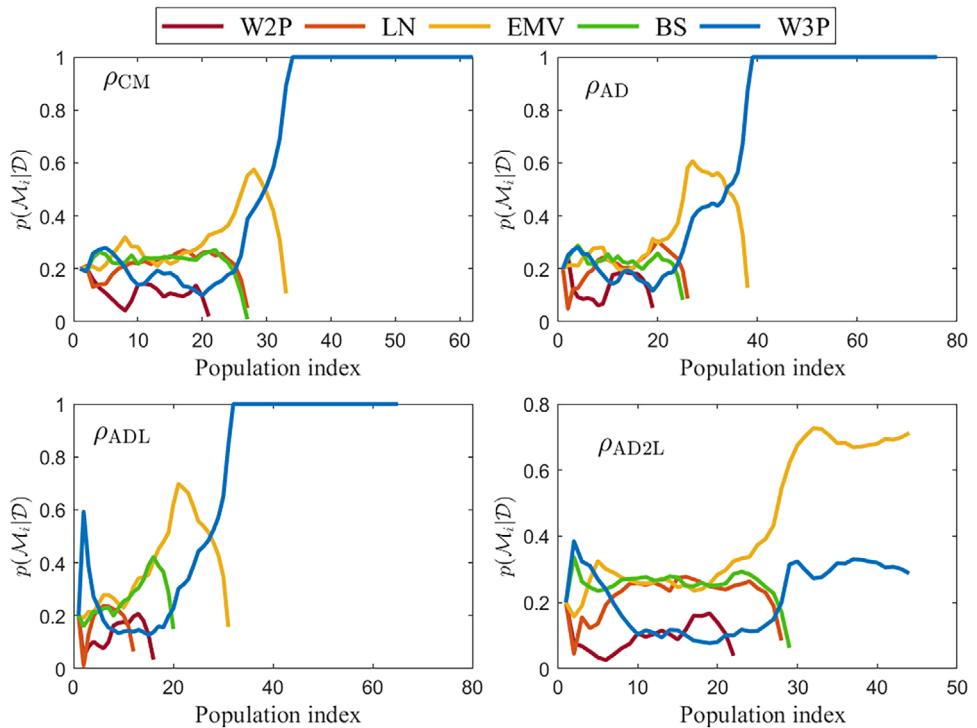


FIGURE 5 Evolution of the model posterior probabilities over the populations for the different distances.

TABLE 5 Optimal estimates for $\Delta\sigma = 320$ MPa using different distances for the considered competing models.

$\Delta\sigma$	Distance	W2P	LN	EMV	BS	W3P
320 MPa	ρ_{CM}	$\hat{\alpha} = 369996.5796$ $\hat{\beta} = 5.7065$	$\hat{\mu} = 12.7435$ $\hat{\sigma} = 0.1904$	$\hat{\mu} = 320180.5229$ $\hat{\sigma} = 54102.9529$	$\hat{\alpha} = 342314.0803$ $\hat{\beta} = 0.1904$	$\hat{\gamma} = 287717.7348$ $\hat{\alpha} = 74559.0656$ $\hat{\beta} = 0.6802$
	ρ_{AD}	$\hat{\alpha} = 403295.3929$ $\hat{\beta} = 3.4548$	$\hat{\mu} = 12.7758$ $\hat{\sigma} = 0.2635$	$\hat{\mu} = 323689.8978$ $\hat{\sigma} = 71346.3097$	$\hat{\alpha} = 353816.5250$ $\hat{\beta} = 0.2652$	$\hat{\gamma} = 288236.0905$ $\hat{\alpha} = 72715.0256$ $\hat{\beta} = 0.6654$
	ρ_{ADL}	$\hat{\alpha} = 345363.1880$ $\hat{\beta} = 10.7200$	$\hat{\mu} = 12.7301$ $\hat{\sigma} = 0.1470$	$\hat{\mu} = 319998.2136$ $\hat{\sigma} = 44284.5634$	$\hat{\alpha} = 337795.5389$ $\hat{\beta} = 0.1472$	$\hat{\gamma} = 288798.9968$ $\hat{\alpha} = 73669.6859$ $\hat{\beta} = 0.6180$
	ρ_{AD2L}	$\hat{\alpha} = 342246.0272$ $\hat{\beta} = 11.8135$	$\hat{\mu} = 12.7128$ $\hat{\sigma} = 0.1180$	$\hat{\mu} = 317297.4637$ $\hat{\sigma} = 36260.4266$	$\hat{\alpha} = 331968.6335$ $\hat{\beta} = 0.1181$	$\hat{\gamma} = 289214.3754$ $\hat{\alpha} = 75155.0137$ $\hat{\beta} = 0.5740$

Abbreviations: BS, Birnbaum-Saunders; EMV, extreme maximum value; LN, lognormal; W2P, two-parameterWeibull; W3P, three-parameterWeibull.

particle over the populations considering the data obtained under $\Delta\sigma = 320$ MPa. One can clearly see that the algorithm selects the W3P model as the most likely model to fit the available fatigue lives. The results are shown by using the ρ_{ADL} distance for illustrative purposes. The algorithm rules out the least likely models over the populations and selects the model(s) which performs better. Figure 5 shows for the same stress level the evolution of the model posterior probabilities over the populations using the different distance functions. The W3P model is selected as the most plausible model by minimising the $(\rho_{CM}, \rho_{AD}, \rho_{ADL})$ distances while with the ρ_{AD2L} distance, the algorithm converges to the EMV and W3P models with a posterior probabilities equal to 0.645 and 0.355, respectively.

Next the ABC-NS algorithm is used to infer the competing models separately in order to compare the fitting quality. Table 5 shows the optimal estimates for each model using the different distance functions. Figure 6 displays the failure times and the fitted CDF, one can see that the W3P outperforms the other models for this data set. The W3P model provides the best fitting quality and has the ability to accurately describe the lower left-tail. Finally, it should be noted that he

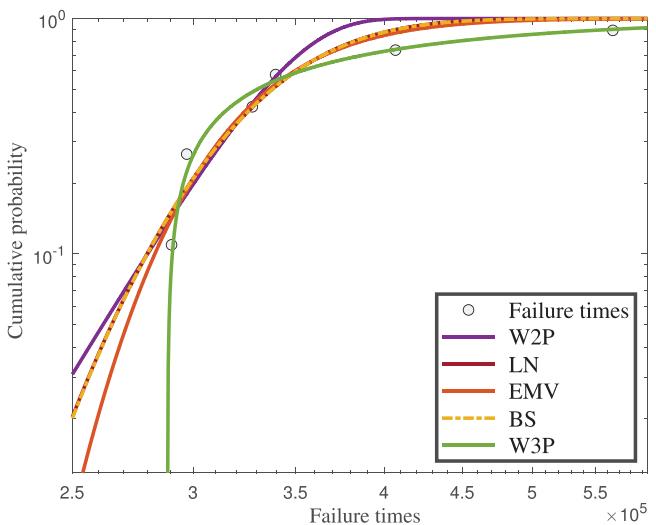


FIGURE 6 A comparison between the fitted CDFs using the ρ_{ADL} distance. CDF, cumulative distribution function.

ABC-NS algorithm converges for all the models despite the small sample size. For the selected model, one can see that the shape parameter is less than 1 for all the distances which means that the hazard function is decreasing as the number of cycles is increased which is not coherent with the physics of fatigue. It has been observed that for all the considered sample sizes and for all the distance functions, the ABC-NS algorithm works very well and never displays any problem with convergence.

To make a comparison, we employ the ABC-NS using the other distances and we estimate the posterior model probabilities for all the tested specimens. Table 6 shows the model posterior probabilities at the last population using the different distances. One can clearly see that the W3P and the EMV models perform equally well. From the same table, one can clearly see that the W2P performs poorly. It is ranked the last for all the metrics. Table 6 shows the model ranking for the other stress ranges, one can see that none of the models is always supported by the data. The selected model vary with respect of the distance function and the collected data. For instance for specimen 3, the best model is the W2P while for specimen 4, the best model is the EMV.

4.1.1 | Fit effects in the left-tail region

After assessing how well each distribution fit to the overall data sets, the focus now is on the left-tail region of the distribution since this is the region of importance to engineering design and planning applications. In order to describe the fit effects in the tail regions, the error parameters d_{F1} and d_{F2} between the real value and theoretical value of two life data (so-called fit-differences in the left-tail region)^{52–54} are, respectively, defined as:

$$d_{F1} = \hat{F}(t_1) - F(t_1) = \frac{1 - 0.3}{n + 0.4} - F(t_1) \quad (11)$$

$$d_{F2} = \hat{F}(t_2) - F(t_2) = \frac{2 - 0.3}{n + 0.4} - F(t_2) \quad (12)$$

where t_1 and t_2 are, respectively, the minimum life and sub-minimum life of a group of n data.

The smaller the $|d_F|$ value, the better the fit effects in the left-tail region. When $d_F < 0$, this implies a conservative evaluation. In contrast, when $d_F > 0$, a non-conservative evaluation results. In addition, by comparing d_{F1} with d_{F2} , the trend of the evaluation can be determined. In fact, when $d_{F1} < d_{F2}$, the evaluation, as the fatigue life is less than t_1 , may be conservative. But, when $d_{F1} > d_{F2}$, the evaluation may be non-conservative. The calculated values of d_{F1} and d_{F2} for the crack lengths given by the competing models are given in Table 7.

From Table 7, the following observations can be drawn:

TABLE 6 Posterior probabilities and the corresponding model ranking (in parentheses) for specimens subjected to different stress levels.

$\Delta\sigma$ [MPa]	Distance function	Competing models				
		W2P	LN	EMV	BS	W3P
320	ρ_{CM}	0 (5)	0 (3)	0 (2)	0 (4)	1 (1)
	ρ_{AD}	0 (5)	0 (3)	0.517 (1)	0 (4)	0.483 (2)
	ρ_{ADL}	0 (4)	0 (5)	0 (2)	0 (3)	1 (1)
	ρ_{AD2L}	0 (5)	0 (4)	0.767 (1)	0 (3)	0.233 (2)
290	ρ_{CM}	0 (5)	0 (3)	0 (2)	0 (4)	1 (1)
	ρ_{AD}	0 (5)	0 (3)	0.517 (1)	0 (4)	0.483 (2)
	ρ_{ADL}	0 (4)	0 (5)	0 (2)	0 (3)	1 (1)
	ρ_{AD2L}	0 (5)	0 (4)	0.767 (1)	0 (3)	0.233 (2)
260	ρ_{CM}	1 (1)	0 (4)	0 (5)	0 (3)	0 (2)
	ρ_{AD}	1 (1)	0 (4)	0 (3)	0 (5)	0 (2)
	ρ_{ADL}	1 (1)	0 (4)	0 (3)	0 (5)	0 (2)
	ρ_{AD2L}	0.715 (1)	0 (4)	0 (3)	0 (5)	0.285 (2)
230	ρ_{CM}	0.392 (2)	0 (4)	0 (3)	0 (5)	0.608 (1)
	ρ_{AD}	0 (5)	0 (2)	1 (1)	0 (4)	0 (3)
	ρ_{ADL}	0.702 (1)	0 (3)	0 (4)	0 (5)	0.298 (2)
	ρ_{AD2L}	0 (2)	0 (4)	0 (3)	0 (5)	1 (1)
202	ρ_{CM}	0 (5)	0 (3)	0 (4)	0.360 (2)	0.640 (1)
	ρ_{AD}	0 (5)	0 (3)	0 (4)	0.441 (2)	0.559 (1)
	ρ_{ADL}	0 (5)	0 (3)	0 (4)	0 (2)	1 (1)
	ρ_{AD2L}	0 (5)	0 (3)	0 (4)	0.730 (1)	0.270 (2)

Abbreviations: AD, Anderson-Darling; BS, Birnbaum-Saunders; CM, Cramer-von-Mises; EMV, extreme maximum value; LN, lognormal; W2P, two-parameterWeibull; W3P, three-parameterWeibull.

- The W3P model provides the best fit in the left-tail region followed by the EMV model. The BS and the LN models are quite similar while the W2P model provides a relatively poor fit.
- The predictions in the left-tail region of the W3P model shows the same trends: $d_{F1} < d_{F2}$ except for $\Delta\sigma = 260$ MPa.
- The predictions of the LN and EMV models shows the same trend $d_{F1} < d_{F2}$ for all the stress ranges. This means that the evaluation, as the fatigue life is less than t_1 , may be conservative for all the crack lengths.
- The results show that the EMV, LN and BS models give a conservative evaluations in the left-tail region.
- Overall, a better fitting quality of the left-tail region is observed by using the (ρ_{AD} , ρ_{ADL} , ρ_{AD2L}) distances compared with the ρ_{AD2L} distance which is expected.

Analysing all the data, one can see from the total fit effects that the W3P model is the best. The shape parameter is less than 1 of all the crack lengths, which results in failure rates that decrease with fatigue cycling. This is not coherent with the fatigue physics. Moreover, the W3P model may give a non-conservative evaluations in the left-tail region for some stress ranges as one can see in Table 7. The EMV model may be an appropriate model to describe fatigue life both from safety and from the fatigue physics points of view. In addition, we have seen that it gives a conservative evaluations in the left-tail region. The BS and the LN distributions may be a good distributions to assume as well due to the good total fit effects, although the hazard functions in the large cycles number decrease with fatigue cycling.

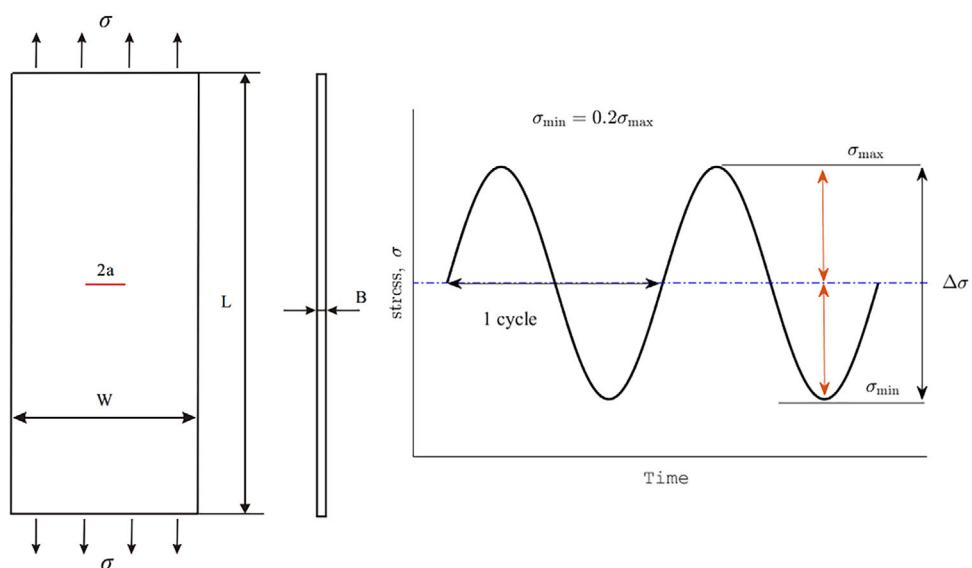
4.2 | Example 2. Fatigue tests and reliability analysis

In the second illustration, we consider the fatigue crack growth data available in the literature.⁵⁵ The experimental data set consists of 68 crack propagation trajectories obtained from a series of fatigue tests performed on identical center-cracked 2024-T3 aluminum plates (specimen geometry is illustrated in Figure 7). The tested specimens have the following dimensions: length L = 558.8 mm, width w = 152.4 mm, thickness B = 2.54 mm and a crack of initial size $2a = 18$ mm

TABLE 7 The fitting parameters d_{F1} and d_{F2} for the considered cracks using the ABC-NS with the different distance functions.

$\Delta\sigma$	Model	ρ_{CM}		ρ_{AD}		ρ_{ADL}		ρ_{AD2L}	
		d_{F1}	d_{F2}	d_{F1}	d_{F2}	d_{F1}	d_{F2}	d_{F1}	d_{F2}
320	W2P	-0.1122	0.0174	-0.1653	-0.0279	0.0746	0.2254	-0.0239	0.0948
	LN	-0.0839	0.0377	-0.1178	0.0115	-0.0421	0.0748	-0.0185	0.0927
	EMV	-0.0666	0.0501	-0.0931	0.0319	-0.0321	0.0794	-0.0123	-0.0919
	BS	-0.1730	-0.0334	-0.1181	0.0113	-0.0420	0.0748	-0.0185	0.0927
	W3P	-0.0108	0.0498	0.0174	0.0487	0.0197	0.0413	0.0253	0.0283
290	W2P	0.0001	-0.0274	-0.0377	-0.0538	0.0667	0.0863	0.0226	-0.0049
	LN	0.0380	-0.0233	0.0239	-0.0347	0.0336	-0.0271	0.0266	-0.0300
	EMV	0.0614	-0.0165	0.0426	-0.0387	0.0383	-0.0386	0.0270	-0.0402
	BS	-0.0836	0.0379	0.0240	-0.0350	0.0338	-0.0274	0.0266	-0.0304
	W3P	0.0946	0.2297	0.0346	-0.0472	0.0940	0.2231	0.0240	-0.0098
260	W2P	0.0536	-0.0274	0.0479	-0.0294	0.0414	-0.0413	0.0255	-0.0643
	LN	0.0731	0.0174	0.0651	-0.0556	0.0526	-0.0785	0.0261	-0.1059
	EMV	0.0727	0.0125	0.0581	-0.0591	0.0448	-0.0679	0.0244	-0.0790
	BS	0.0383	-0.0234	0.0657	-0.0607	0.0534	-0.0842	0.0261	-0.1130
	W3P	0.0597	-0.0052	0.0483	-0.0295	0.0422	-0.0417	0.0256	-0.0645
230	W2P	0.0166	0.0495	-0.0112	0.0132	0.0129	0.0443	0.0114	0.0422
	LN	0.0535	0.0931	0.0346	0.0526	0.0271	0.0421	0.0174	0.0293
	EMV	0.0499	0.0879	0.0277	0.0493	0.0212	0.0422	0.0150	0.0355
	BS	0.0549	0.0944	0.0355	0.0513	0.0281	0.0407	0.0179	0.0271
	W3P	0.0271	0.0580	0.0220	0.0347	0.0148	0.0442	0.0133	0.0392
202	W2P	-0.1239	-0.0287	-0.1205	-0.0247	-0.0888	0.0060	-0.0536	0.0397
	LN	-0.090	0.0011	-0.0820	0.0104	-0.0669	0.0248	-0.0441	0.0462
	EMV	-0.1001	-0.0060	-0.0953	-0.005	-0.0673	0.0255	-0.0417	0.0484
	BS	-0.0905	0.0013	-0.0816	0.0105	-0.0663	0.0251	-0.0437	0.0464
	W3P	-0.0392	0.0362	-0.0141	0.0530	0.0070	0.0465	0.0179	0.0297

Abbreviations: ABC, Approximate Bayesian computation; AD, Anderson-Darling; BS, Birnbaum-Saunders; CM, Cramer-von-Mises; EMV, extreme maximum value; LN, lognormal; NS, Nested sampling; W2P, two-parameterWeibull; W3P, three-parameterWeibull.

**FIGURE 7** (Left) Test piece geometry with a central crack (taken from Ref. [56]), (right) harmonic loading.

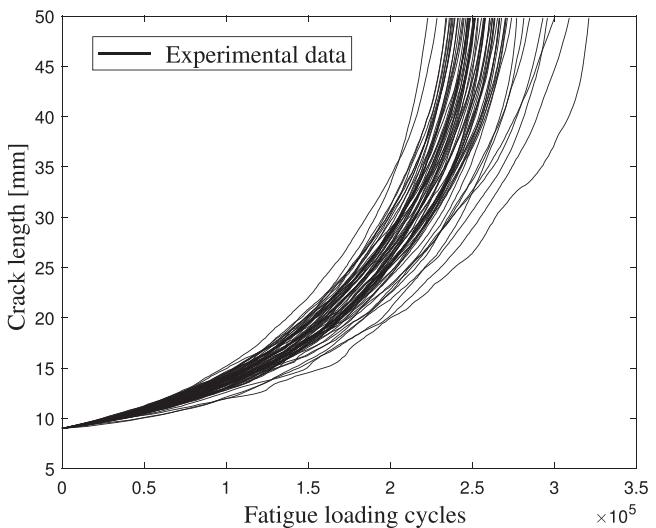


FIGURE 8 Experimental fatigue crack trajectories.

located at the centre of specimen. The tests were performed under load control using a sinusoidal input at 20 Hz producing a constant stress range of $\Delta\sigma = 48.28$ MPa with a stress ratio equal to 0.2 (see, Figure 7). The accumulated number of cycles was recorded at specified crack lengths between 9 and 49.8 mm. The 68 experimental trajectories are shown in Figure 8.

In this study, one aims to select the most suitable reliability model to describe the number of cycles for a pre-defined critical crack length in order to estimate the probability of failure at a target number of cycles t_s . Let us consider a structure subject to a main degradation process ($Z_t, t \in \mathbb{R}_+$) which increases randomly on $\mathbb{R}_+^* = (0, \infty)$ until it reaches a critical value $a_{cr} > 0$, meaning the failure of the structure.⁵⁷ It is assumed that Z_t is an observable process whose sample paths are obtained from experimental feedback indicating the service time for crack size from a_0 to a_{cr} . The reliability of such a system may be expressed by:

$$R(t) = \mathbb{P}(Z_t < a_{cr}) \quad (13)$$

that is the probability that the process Z_t does not reach the failure boundary on the whole observation interval $[0, t_s]$.

Thus, the reliability analysis consists in estimating the probability of failure given by Equation (13). It has shown in Annis⁵⁸ that first-and second-order reliability methods FORM/SORM were inappropriate for such a problem for estimating the probability of failure. Bourinet⁵⁹ fixed this issue by introducing an extra random variable which is representative of a model error. In this study, we present a simple and a straightforward method to compute the probability of failure. It is based on an appropriate selection of the form of the probability distribution describing the cycles-to-failure of the tested specimens. To better illustrate the purpose of this study, Figure 9 shows the probability of failure to be estimated. The target time service t_s is selected as the average value between the 7th and 8th lowest numbers of cycles of the Virkler tests:

$$t_s = \frac{237292 + 237794}{2} = 237543 \text{ cycles} \quad (14)$$

Based on the empirical cumulative density function, the probability of failure should be between 0.0980 and 0.1126.

The ABC-NS algorithm is now employed using the ρ_{AD} distance for illustrative purpose to discriminate between the candidate models with the same control parameters as in Example 1. Our goal here is to select simultaneously the reliability model but also the suitable distance to precisely estimate the probability of failure after a specified service time t_s . The posterior model probabilities over some intermediate populations are shown in Figure 10. The algorithm converges to the EMV model. To get more insights, we infer the competing models to see the impact of the selected distance on model selection. Figure 11 shows the posterior probabilities for the four distances using the same stopping criterion. One clearly see a substantial advantage to the EMV model except for the ρ_{AD2L} distance where the LN model is supported by the data. Table 8 shows the estimated probabilities for all the distances. One can clearly see that the EMV and LN models performs well in terms of probability of failure prediction. It is worthwhile to mention that except the ρ_{CM} distance, all the the selected models with (ρ_{AD}, ρ_{ADL} and ρ_{AD2L}) distances provide an estimation of the probability of failure inside the

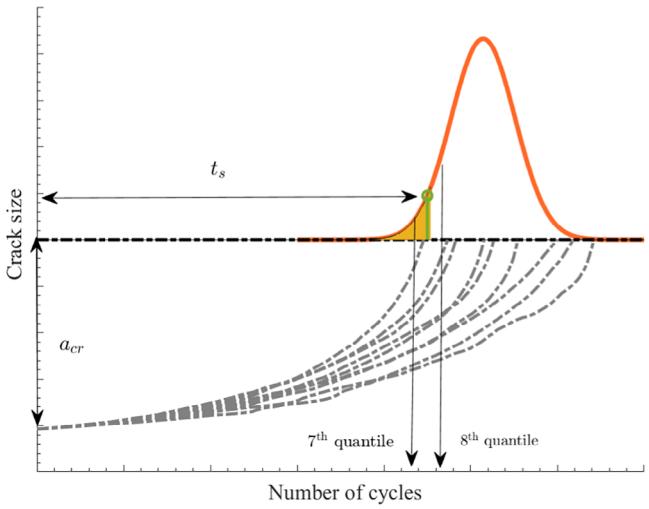


FIGURE 9 Schematic diagram of the probability of failure to be estimated and the random time distribution with a target service time t_s .

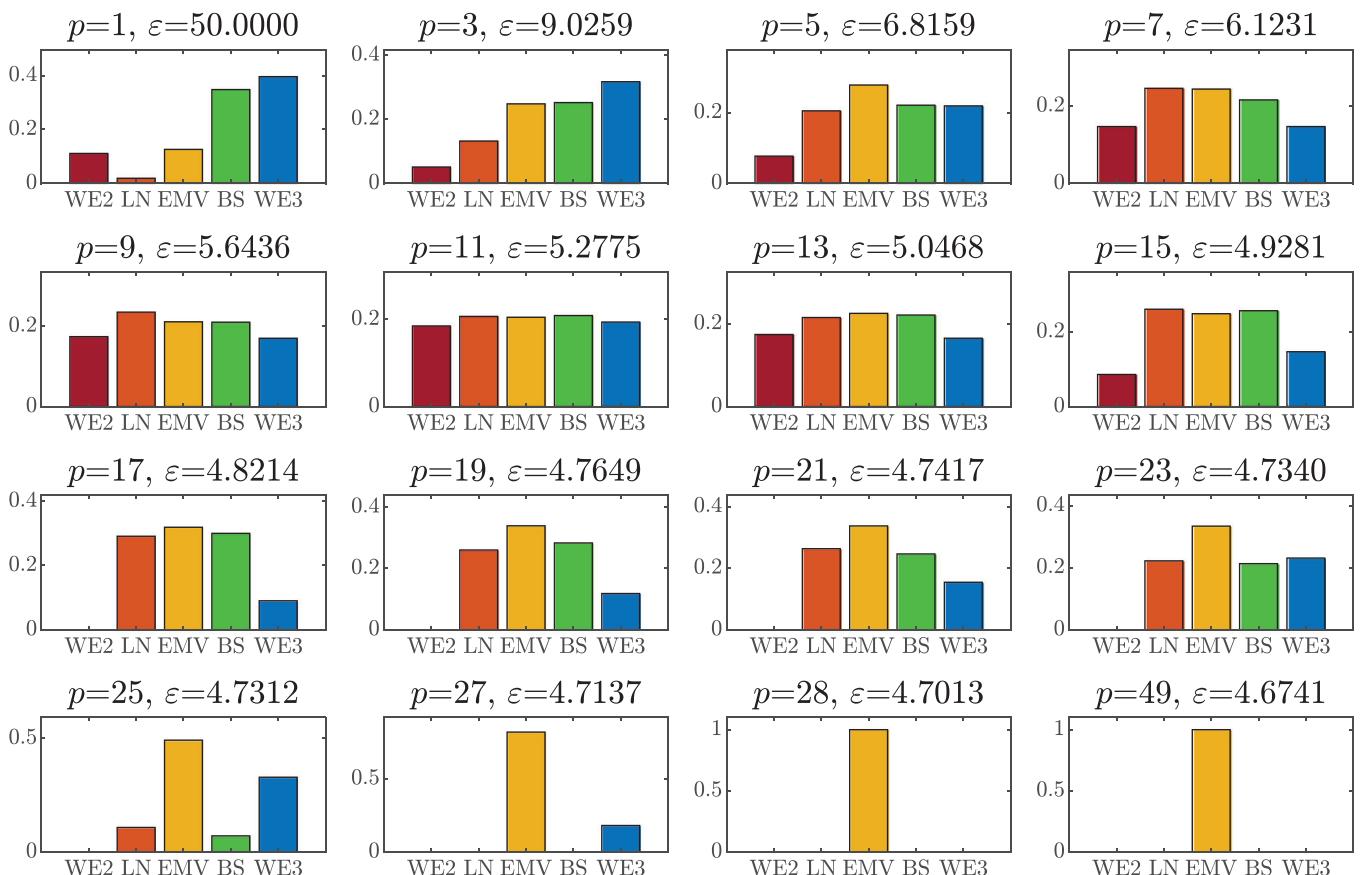


FIGURE 10 Model posterior probabilities over some intermediate populations using ρ_{AD} distance.

TABLE 8 Estimation of the probability of failure for the different distances for $a_{cr} = 49.8$ mm and $t_s = 237543$ cycles.

Distance function	ρ_{CM}	ρ_{AD}	ρ_{ADL}	ρ_{AD2L}
Selected model	EMV	EMV	EMV	LN
Probability of failure P_f	0.0918	0.0994	0.0998	0.1004

Abbreviations: AD, Anderson-Darling; CM, Cramer-von-Mises; EMV, extreme maximum value; LN, lognormal.

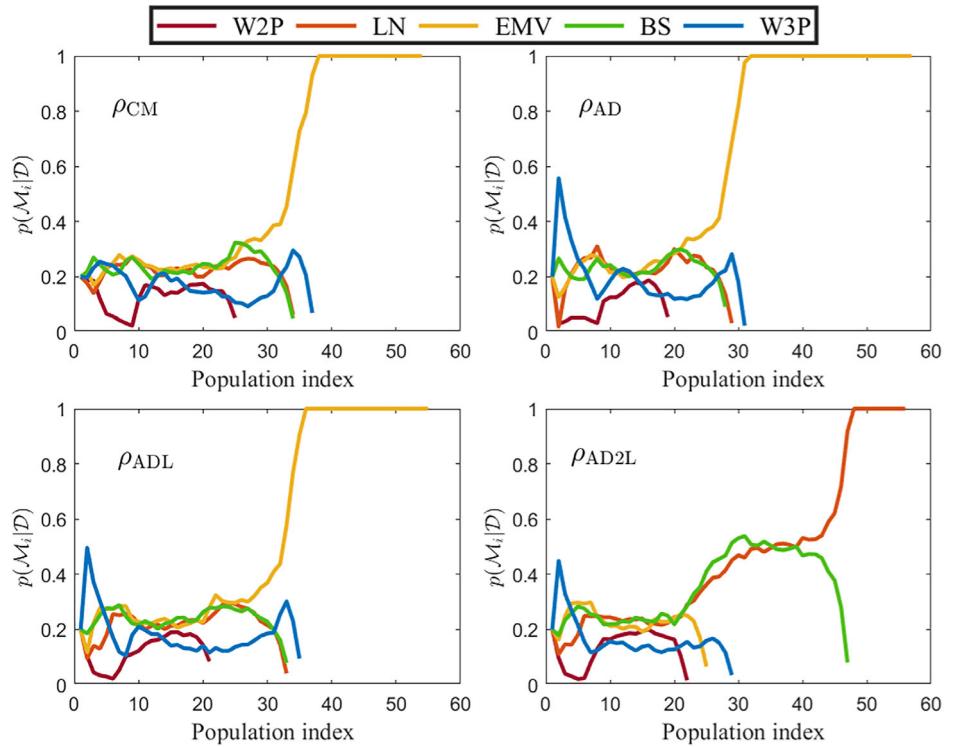


FIGURE 11 Evolution of the model posterior probabilities over the populations for the different distances.

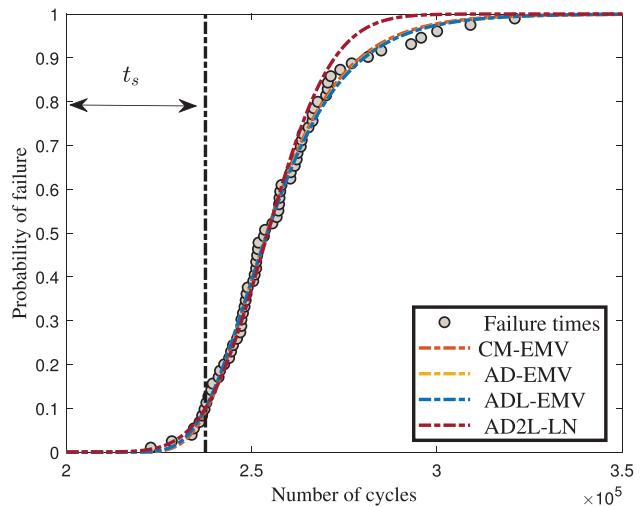


FIGURE 12 Failure times to reach the critical crack size and the fitted CDFs using the different distance functions. CDF, cumulative distribution function.

empirical bounds. This means that by selecting a plausible model and appropriate distance function, it is possible to better capture the pattern of small extreme values more “closely” and more importantly, the data on the lower tail.

Figure 12 shows the empirical data and the fitted CDF using the selected model for each distance function. Overall, one can see a good fitting quality.

Now let us investigate the influence of the available times to failure on the probability of failure accuracy for a target time service. The main goal is to get an idea about the robustness of the different distances in estimating the probability of failure. To do so, we drop the available data from 5 to 30 using a step of 5 and we estimate the probability of failure. To get a representative results, we replicate the simulations 500 times. From Figure 13, one can see that overall the ρ_{AD} , ρ_{AD} distances can give relatively accurate predictions. Based on the obtained results, it is obvious that the ρ_{AD2L} distance is more conservative as it overestimates the probability of failure while the ρ_{CM} distance provides a non conservative

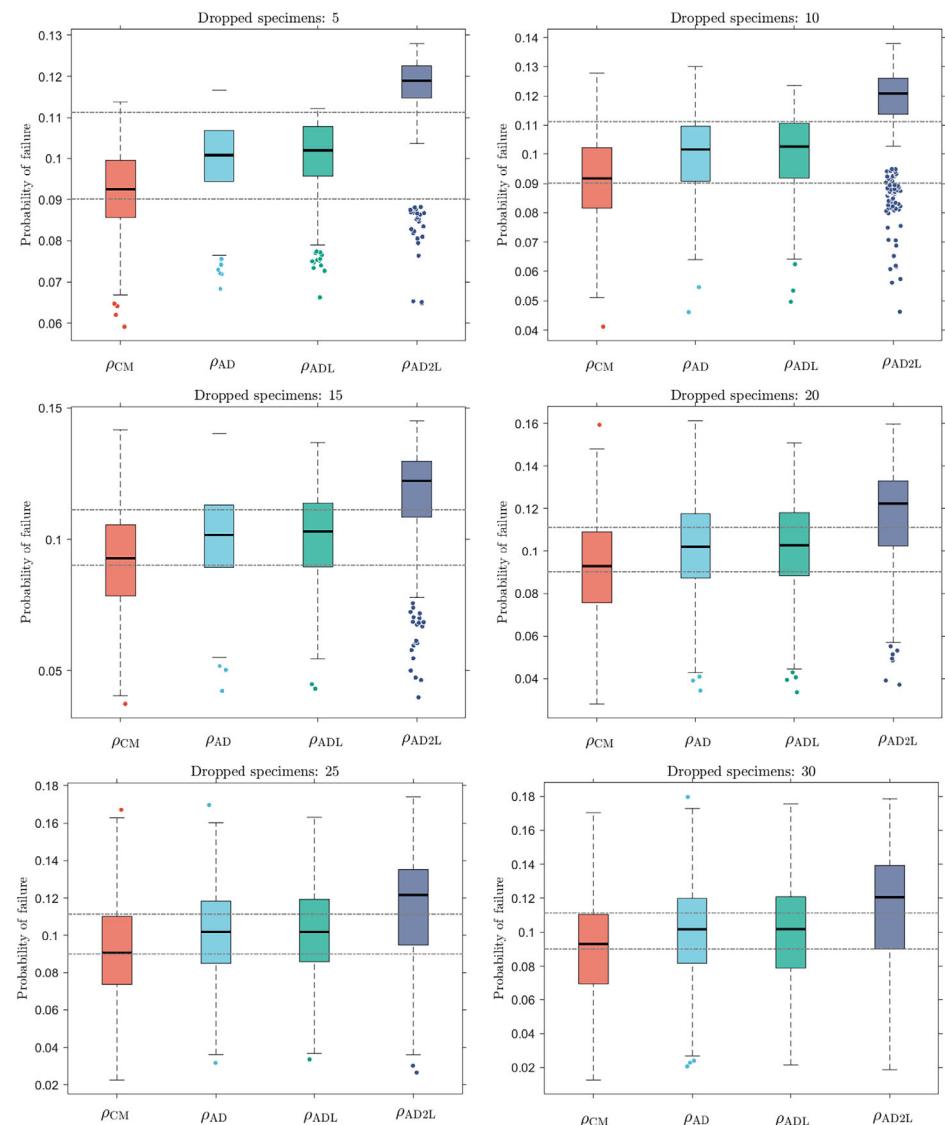


FIGURE 13 Boxplots of the probabilities of failure estimated using the selected models with different distances and 500 replications.

estimates of the probability of failure as it tends towards the lower bound. Although the estimated probability is out the empirical bounds, the ρ_{AD2L} distance may be of interest if a conservative estimates are preferred in the practical context of reliability evaluation.

5 | CONCLUDING REMARKS

In this paper, a simulation-based inference algorithm called ABC-NS was adopted to discriminate between a number of competing models widely used in fatigue. The ABC-NS ranks the competing models according to their plausibility in light of a calibration data set. We have illustrated the proposed methodology with various data sets considering small/moderate sample sizes. Overall the method works well without any numerical problems despite the limited data. Several distances have been considered each with a specified properties. By identifying the purpose of model selection, modellers and experimenters can decide which distance should be preferred. Through different simulation studies we have shown the good performance of the method, which requires minimal tuning. The obtained results show that no model is always supported by the data. Thus, a straightforward method to discriminate simultaneously between a number of candidate models is very useful. Additionally, the optimal model may vary regarding the distance function used to measure the discrepancy between the observed and the simulated data. In summary, the simulation-based method presents some important advantages compared with the classical methods and could be useful for practitioners. Some of the merits of the simulation-based inference method are give below:

- The competing models are compared simultaneously and the least likely models are eliminated through the simulation.
- The proposed method is flexible in the sense that the modellers can use different distances.
- The principle of parsimony is well embedded in the ABC-NS algorithm (it tends to select simpler models first).
- The proposed methodology is easy to implement, intuitive and can address simultaneously parameter calibration and model selection.
- The algorithm can be easily tailored following the user's interest which is embedded in the distance function.
- Very useful when the main concern is about the prediction of low percentiles fatigue life.
- It has shown that the ABC algorithm coupled with a suitable distance is less affected by model misspecification.

DATA AVAILABILITY STATEMENT

The data and the Matlab codes that support the findings of this study are available from the corresponding author.

REFERENCES

1. Schijve J. Statistical distribution functions and fatigue of structures. *Int J Fatigue*. 2005;27:1031-1039.
2. Klemenc J, Fajdiga M. Estimating S-N curves and their scatter using differential ant-stigmergy algorithm. *Int J Fatigue*. 2012;43:90-97.
3. Jeffreys H. Some tests of significance, treated by the theory of probability. *Math Proc Cambridge Philos Soc*. 1935;31(2):203-222.
4. Jeffreys H. *Theory of Probability*. 3rd ed. Oxford University Press; 1939.
5. Zhu SP, Huang HZ, Smith R, Ontiveros V, He LP, Modarres M. Bayesian framework for probabilistic low cycle fatigue life prediction and uncertainty modeling of aircraft turbine disk alloys. *Probab Eng Mech*. 2013;34:114-122.
6. Berg T, von Ende S, Lammerin R. Bayesian consideration of unknown sensor characteristics in fatigue-related structural health monitoring. *Probab Eng Mech*. 2019;56:71-81.
7. Shoji H, Shinozuka M, Sampath S. A Bayesian evaluation of simulation models of multiple-site fatigue cracks. *Probab Eng Mech*. 2001;16:355-361.
8. Guan X, Jha R, Liu Y. Model selection, updating, and averaging for probabilistic fatigue damage prognosis. *Struct Saf*. 2011;33:242-249.
9. Chiachio M, Chiachio J, Rus G, Beck JL. Predicting fatigue damage in composites: a Bayesian framework. *Struct Saf*. 2014;51:57-68.
10. Guida M, Penta F. A Bayesian analysis of fatigue data. *Struct Saf*. 2009;32(1):64-76.
11. Jeffreys H. *Theory of Probability*. Oxford University Press; 1939.
12. Lewis SM, Raftery AE. Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator. *J Am Statist Assoc*. 1997;92:648-655.
13. Berger JO. Bayes factors. In: Kotz S, Balakrishnan N, Read C, Vidakovic B, Johnson NL, eds. *Encyclopeda of Statistical Sciences*. Vol 1. 2nd ed. Wiley; 2006:378-386.
14. Kass RE, Raftery AE. Bayes factors. *J Amer Statist Assoc*. 1995;90:773-795.
15. O'Hagan A. Discussion of posterior Bayes factors by M. Aitkin. *J R Stat Soc Ser B*. 1991;53:136.
16. Berger JO, Perrichi LR. The intrinsic Bayes factor for model selection and prediction. *J Am Stat Assoc*. 1996;91:109-122.
17. O'Hagan A. Fractional Bayes factors for model comparison, (with discussion). *J R Stat Soc Ser B*. 1995;57:99-138.
18. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Control*. 1974;19:716-723.
19. Hurvich CM, Tsai CL. Regression and time series model selection in small samples. *Biometrika*. 1989;76:297-307.
20. Schwarz G. Estimating the dimension of a model. *Ann Stat*. 1978;6:461-464.
21. Hannan EJ, Quinn BG. The Determination of the order of an autoregression. *J R Stat Soc, Series B*. 1979;41:190-195.
22. Spiegelhalter DJ, Best NG, Carlin BP, Vander Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc Stat Methodol*. 2002;64(4):583-639.
23. Green PJ. Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biomnetrika*. 1995;82:711-732.
24. Bouranis L, Friel N, Maire F. Model comparison for Gibbs random fields using noisy reversible jump Markov chain Monte Carlo. *Comput Stat Data Anal*. 2018;128:221-241.
25. Ching J, Chen YC. Transitional Markov Chain Monte Carlo method for Bayesian updating, model class selection, and model averaging. *J Eng Mech*. 2007;133:816-832.
26. Cheung S, Beck JL. Bayesian model updating using hybrid monte carlo simulation with application to structural dynamic models with many uncertain parameters. *J Eng Mech*. 2009;135(4):243-255.
27. Wasserman L. Bayesian model selection and model averaging. *J Math Psych*. 2000;44(1):92e107.
28. Gull SF. Bayesian inductive inference and maximum entropy. In: *Maximum-entropy and Bayesian methods in Science and Engineering*. Vol. 31-32. Springer; 1988:53-74.
29. Mukherjee P, Parkinson D, Liddle AR. A nested sampling algorithm for cosmological model selection. *Astrophys J Lett*. 2006;638(2):L51.
30. Mthembu L, Marwala T, Friswell MI, Adhikari S. Model selection in finite element model updating using the Bayesian evidence statistic. *Mech Syst Signal Process*. 2011;25(7):2399-2412.
31. Elsheikh AH, Wheeler MF, Hoteit I. Hybrid nested sampling algorithm for Bayesian model selection applied to inverse subsurface flow problems. *J Comput Phys*. 2014;258:319-337.
32. Elsheikh AH, Ibrahim H, Wheeler MF. Efficient Bayesian inference of subsurface flow models using nested sampling and sparse polynomial chaos surrogates. *Comput Methods Appl Mech Eng*. 2014;269:515-537.

33. Abdessalem AB, Dervilis N, Wagg D, Worden K. Model selection and parameter estimation in structural dynamics using approximate Bayesian computation. *Mech Syst Signal Process.* 2018;99:306-325.
34. Abdessalem AB, Dervilis N, Wagg D, Worden K. Model selection and parameter estimation of dynamical systems using a novel variant of approximate Bayesian computation. *Mech Syst Signal Process.* 2019;122:364-386.
35. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J R Soc Interface.* 2009;6:187-202.
36. Tanaka MM, Francis AR, Luciani F, Sisson SA. Using approximate Bayesian computation to estimate tuberculosis transmission parameters from genotype data. *Genetics.* 2006;173:1511-1520.
37. Jennings E, Madigan M. astroABC: An Approximate Bayesian Computation Sequential Monte Carlo sampler for cosmological parameter estimation. *Astron Comput.* 2017;19:16-22.
38. Drovandi CC, Pettitt AN. Estimation of parameters for macroparasite population evolution using approximate bayesian computation. *Biometrics.* 2011;67(1):225-233.
39. da Costa JMJ, Orlande HRB, da Silva WB. Model selection and parameter estimation in tumor growth models using approximate Bayesian computation-ABC. *J Comput Appl Math.* 2018;37:2795-2815.
40. Beaumont MA, Zhang W, Balding DJ. Approximate bayesian computation in population genetics. *Genetics.* 2002;162(4):2025-2035.
41. Wegmann D, Leuenberger C, Excoffier L. Efficient approximate bayesian computation coupled with markov chain monte carlo without likelihood. *Genetics.* 2009;182(4):1207-1218.
42. Toni T, Stumpf MPH. Parameter inference and model selection in signaling pathway models. In: Fenyö D, ed. *Computational Biology, Methods in Molecular Biology (Methods and Protocols).* Vol 673. Humana Press; 2010:283-295.
43. Toni T, Stumpf MP. Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics.* 2010;26(1):104-110.
44. Anderson TW, Darling DA. A Test of Goodness-of-fit. *J Am Stat Assoc.* 1954;49:765-769.
45. Farrel PJ, Stewart KR. Comprehensive study of tests for normality and symmetry: extending the Spiegelhalter test. *J Stat Comput Simul.* 2006;76:803-816.
46. Hose GC, Van den Brink PJ. Confirming the species-sensitivity distribution concept for Endosulfan using laboratory, Mesocosms, and field data. *Arch Environ Contam Toxicol.* 2004;47:511-520.
47. Posthuma L, Suter G, Traas T. *Species sensitivity distributions in ecotoxicology.* Environmental and Ecological Risk Assessment Series. Taylor & Francis; 2010.
48. Delignette-Muller ML, Dutang C. fitdistrplus: an R Package for fitting distributions. *J Stat Softw.* 2015;64(4):1-34.
49. Chen H-N, Zhang Y-H. Summing-up report of the welded joint basic curve test. Report No. 02-00, National Mechanical Industry Science and Technology Development Bounds, Project 86JB5017, Zhengzhou Mechanical Institute, Zhengzhou, China, 1991, in Chinese.
50. Fang Q-Z, Li H-L. The test report of the welded joint basic curves. Report No. 02-01, National Mechanical Industry Science and Technology Development Bounds, Project 86JB5017, Zhengzhou Mechanical Institute, Zhengzhou, China, 1991, in Chinese.
51. Liu T, Zhang S, Yue J-P. The test report of the welded joint basic curves. Report No. 02-02, National Mechanical Industry Science and Technology Development Bounds, Project 86JB5017, Zhengzhou Mechanical Institute, Zhengzhou, China, 1991, in Chinese.
52. Zhao YX, Gao Q, Sun XF. A statistical investigation of the fatigue lives of Q235 steel-welded joints. *Fatigue Fract Eng Mater Struct.* 1998;21:781-790.
53. Zhao YX, Gao Q, Wang JN. An approach for determining an appropriate assumed distribution of fatigue life under limited data. *Reliab Eng Syst Saf.* 2000;67:1-7.
54. Li H, Wen D, Lu Z, Wang Y, Deng F. Identifying the probability distribution of fatigue life using the maximum entropy principle. *Entropy.* 2016;18:111.
55. Virkler DA, Hillberry BM, Goel PK. The statistic nature of fatigue crack propagation. *ASME J Eng Mater Technol.* 1979;101:148-153.
56. Qiu Z, Zheng Y. Predicting fatigue crack growth evolution via perturbation series expansion method based on the generalized multinomial theorem. *Theor Appl Fract Mech.* 2016;86:361-369.
57. Chiquet J, Limnios N, Eid M. Piecewise deterministic Markov processes applied to fatigue crack growth modelling. *J Stat Plan Inference.* 2009;139(5):1657-1667.
58. Annis C. How reliable are first order and second order reliability methods? Statistical Engineering website. 2017. <https://statistical-engineering.com/form-sorm/>
59. Bourinet J-M. Reliability analysis and optimal design under uncertainty-Focus on adaptive surrogate-based approaches. Computation [stat.CO]. Université Clermont Auvergne; 2018. tel-01737299v2.

How to cite this article: Ben Abdessalem A. Simulation-based inference for reliability model selection and parameter calibration: An application to fatigue. *Qual Reliab Eng Int.* 2024;40:1855-1874.

<https://doi.org/10.1002/qre.3497>

APPENDIX A: THE COMPETING RELIABILITY MODELS FOR DATA SET 1

Reliability model		PDF/CDF	Model size, p
Lognormal	PDF:	$f(t \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma t} \exp\left\{-\frac{1}{2}\left(\frac{\ln(t)-\mu}{\sigma}\right)^2\right\}$	2
	CDF:	$F(t \mu, \sigma) = \Phi\left\{\frac{\ln(t)-\mu}{\sigma}\right\}$	
Loglogistic	PDF:	$f(t \theta, \nu) = \frac{\nu}{\theta^\nu} t^{\nu-1} \left\{1 + \left(\frac{t}{\theta}\right)^\nu\right\}^{-2}$	2
	CDF:	$F(t \theta, \nu) = 1 - \frac{1}{1 + \left(\frac{t}{\theta}\right)^\nu}$	
Pareto	PDF:	$f(t \alpha, \beta) = \frac{e^z}{\sigma(1+e^z)^2}$	2
	CDF:	$F(t \alpha, \beta) = \frac{1}{1+(x/\alpha)^\beta} = \frac{\left(\frac{x}{\alpha}\right)^\beta}{1+\left(\frac{x}{\alpha}\right)^\beta}$	
Burr	PDF:	$f(t \alpha, c, k) = \frac{\frac{k\alpha}{\alpha} t^{c-1}}{\{1+(\frac{t}{\alpha})^c\}^k}$	3
	CDF:	$F(t \alpha, c, k) = 1 - \frac{1}{\{1+(\frac{t}{\alpha})^c\}^k}$	

APPENDIX B: THE COMPETING RELIABILITY MODELS FOR DATA SET 2

Reliability model		PDF/CDF	Model size, p
W2P	PDF:	$f(t \alpha, \beta) = \beta \alpha^{-\beta} t^{\beta-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\beta\right\}$	2
	CDF:	$F(t \alpha, \beta) = 1 - \exp\left\{-\left(\frac{t}{\alpha}\right)^\beta\right\}$	
EMV	PDF:	$f(t \mu, \sigma) = \frac{1}{\sigma} \exp\left(-\frac{t-\mu}{\sigma}\right) \exp\left(-\exp\left(-\frac{t-\mu}{\sigma}\right)\right)$	2
	CDF:	$F(t \mu, \sigma) = \exp\left\{-\exp\left(-\frac{t-\mu}{\sigma}\right)\right\}$	
BS	PDF:	$f(t \alpha, \beta) = \frac{1}{2\sqrt{2\pi}} \frac{t^{-3/2}(t+\beta)}{\alpha\sqrt{\beta}} \exp\left\{-\frac{1}{2\alpha^2}\left(\frac{t}{\beta} + \frac{\beta}{t} - 2\right)\right\}$	2
	CDF:	$F(t \alpha, \beta) = \Phi\left\{\frac{1}{\alpha}\left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right)\right\}$	
W3P	PDF:	$f(t \alpha, \beta, \tau) = \beta \alpha^{-\beta} (t-\tau)^{\beta-1} \exp\left\{-\left(\frac{t-\tau}{\alpha}\right)^\beta\right\}$	3
	CDF:	$F(t \alpha, \beta, \tau) = 1 - \exp\left\{-\left(\frac{t-\tau}{\alpha}\right)^\beta\right\}$	

AUTHOR BIOGRAPHY



Anis Ben Abdessalem (A.B.A.) is a lecturer in the Department of Mechanical Engineering at the University of Angers and a member of the LARIS laboratory. His background is structural dynamics, nonlinear system identification, reliability, stochastic modelling and Bayesian inference. ABA obtained a PhD in Physical, Mathematics and Information Sciences for Engineers (2007–2011) from INSA Rouen. From 2015 to 2017 he joined the Dynamics Research Group at the University of Sheffield to work on the EPSRC Programme Grant ‘Engineering Nonlinearity’ (EK/K003836). His current research sits between reliability analysis of complex systems, model selection, verification and validation in situations where data is sparse in multiple engineering disciplines.