Bayesian Maximum Entropy Method for Stochastic Model Updating using Measurement data and Statistical Information

Chenxing Wang[a], **Lechang Yang**[a,b,1] **Min Xie**[b], **Marcos Valdebenito**[c], **Michael Beer**[d,e,f],

[a] *School of Mechanical Engineering, University of Science and Technology Beijing, Beijing 100083, China*

[b] *Department of Advanced Design and Systems Engineering, City University of Hong Kong, Hong Kong, China*

[c] *Chair for Reliability Engineering, TU Dortmund University, Dortmund 44227, Germany*

[d] *Institute for Risk and Reliability, Leibniz Universität Hannover, Hannover 30167, Germany*

[f] *Institute for Risk and Uncertainty, University of Liverpool, Liverpool L69 7ZF, United Kingdom*

[g] *International Joint Research Center for Resilient Infrastructure & International Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University, Shanghai 200092, China*

Abstract

The presence of summarized statistical information, such as some statistics of the system response, is not rare in practical engineering as the acquisition of precisely measured point data is expensive and may not be always accessible. In this paper, we integrate the Bayesian framework with the maximum entropy theory and develop a Bayesian Maximum Entropy (BME) approach for model updating in a scenario where measurement data and statistical information are simultaneously available. Within the scope of this contribution, it is assumed that measurement data denote direct observations, e.g. point data, representing system response measurements while statistical information involves summarized information, e.g. moment and/or reliability information, of the system response. The basic principle of our approach is to convert point data and various statistical information into constraints under the BME framework and use the method of Lagrange multipliers to find the optimal posterior distributions. We then extend this approach to imprecise probabilistic models which have not been addressed so far. The approximate Bayesian computation is employed to facilitate the estimation of cumbersome likelihood functions which results from the involvement of entropy terms accounting for statistical information. Furthermore, a Wasserstein distance-based metric is proposed and embedded into the framework to capture the divergence information in an effective and efficient way. The effectiveness of the proposed approach is verified by a numerical case of simply supported beam and an engineering problem of fatigue crack growth. It shows some promising aspects of this research as better calibration results are produced with less uncertainty, and hence potential of our approach for engineering applications.

Keywords

Bayesian Maximum Entropy; Stochastic model updating; Wasserstein distance; Approximate Bayesian computation; Measurement data; Statistical information

Abbreviation

Abbreviation	Term

[1] Corresponding author.
Email address: yanglechang@ustb.edu.cn

BME	Bayesian Maximum Entropy
UQ	Uncertainty Quantification
ME	Maximum Entropy
MrE	Maximum relative Entropy
ABC	Approximate Bayesian Computation
WD	Wasserstein Distance
ED	Euclidean Distance
BD	Bhattacharyya Distance
TMCMC	Transitional Markov Chain Monte Carlo
CDF	Cumulative Distribution Function
PDF	Probability Distribution Functions

1. Introduction

The characterization of the behavior of an engineering system by means of a numerical model usually comprises several deterministic and uncertain quantities. The uncertainty affecting a model can be broadly classified into epistemic and aleatory uncertainty [1]. To deal with such a model, uncertainty quantification (UQ) generally involves forward UQ and inverse UQ [2-6]. The former one allows obtaining the characteristics of the output response of the system given the model and its uncertain input variables while the later one focuses on the assessment of model characterizing the system's behavior and the estimation of unknown parameters given the observed data. Inverse problems are ubiquitous and calibration approaches that consider uncertainty explicitly are often necessary to enhance the accuracy and fidelity of numerical models. To this end, Bayes' theorem allows solving probabilistic calibration problems by combining prior distribution with measurement data via a likelihood function and hence attracts much attention in related communities [7-12].

Information for performing parameter calibration may be available in many different forms[13]. In practical applications, useful information in a report is usually presented in the form of observation frequency. In other cases, the performance of a population of components or systems may be synthesized as reliability data, or the measurements may be described by means of a statistical moment. These heterogeneous sources of information can lead to significant challenges for parameter calibration, as information may not be exclusively provided in traditional form, such as direct (point) observations [14], but also in frequency or summary form, such as reliability data, mean and variance, etc. [15]. The classical Bayesian method is mainly based on point measurements, and no formal rule is available for processing statistical information in other forms which hinders the comprehensive utilization of various sources of information. To fill this gap, Jayne proposed the method of constructing a prior by using the principle of Maximum Entropy (ME) [16, 17], whose idea is to construct a probability distribution function to maximize its information entropy under the constraints of moment information. Caticha and Giffin extended Jayne's idea and proposed to simultaneously process point observations and moment data for probability updating based on the principle of Maximum relative Entropy (MrE) [18-20]. The MrE could be considered as a generalization of Bayesian theory since it reduces to the Bayes' rule when only point data are available. Following this research line, Daoqing Zhou presents a general framework for probabilistic information fusion with point, moment, and interval data based on the principle of maximum relative entropy [21], VanDerHorn and Sankaran proposed to update model parameters by using statistical data and reliability data through Bayesian network [13]. However, till now, the research for coping with measurement data and statistical information (including reliability data and moment data) simultaneously is still limited, and a comprehensive framework addressing multi-source information fusion in the model updating has not been fully constructed yet.

Dealing with data available in different forms is certainly not the only challenge that must be faced when identifying input parameters of a model. Essentially, input parameters may be subject to both aleatoric and epistemic uncertainties and hence, must be characterized by means of specialized models, such as imprecise probabilities. In this context, parameters calibration for imprecise probabilistic models is one of the most challenging tasks in UQ taking into account the possibly high dimension uncertain parameters and the large epistemic uncertainty associated with the input. Specifically, when the dimension of the parameters to be calibrated is high, the derivation of the corresponding likelihood function becomes extremely challenging as in most cases, it does not have an explicit form and is analytical intractable. To facilitate the derivation of complex likelihood function, the approximate Bayesian computation (ABC) proposed by Pritchard et al. has attracted wide attention in recent years [22]. This method involves statistical distance to construct an approximate likelihood function instead of evaluating the full likelihood function, allowing to reduce the associated numerical cost. Although some useful statistical distances, e.g. Euclidean distance and Bhattacharyya distance, have been used to construct the required UQ metric, some inherent drawbacks hinder their further applications in more general engineering cases. For instance, Euclidean distance-based metrics are not able to capture higher-order information of imprecise probability distributions, while the Bhattacharyya distance tends to infinity when the two data sets under analysis are far apart. In this contribution, we explore the application of the Wasserstein distance (WD) [23] as a novel UQ metric in the imprecise probability model updating using multiple sources of information.

Considering the issues described above, this contribution formulates an approach for parameter identification considering heterogeneous sources of information (that allows coping with both point measurements and statistical information) under the Bayesian Maximum Entropy (BME) framework, which is implemented via the ABC where the WD-based UQ metric is fully embedded. The objective is applying this approach to problems where input parameters are characterized with imprecise probabilistic models. The novelty of the works lies in two aspects. First, this paper provides a feasible method for statistical information, i.e. the reliability data and moment data, to be integrated into the parameter calibration. The proposed framework resolves the problem of probability information fusion with both measurement data and statistical information that has not been fully addressed in the existing literature. Second, the Bayesian Maximum Entropy method is applied to problems comprising imprecise probability models, where the WD plays a key role to efficiently fuse heterogeneous data. This allows to accurately calibrate the hyperparameters associated with an imprecise probabilistic model.

The structure of this paper is as follows: In section 2, some preliminary works, including Bayesian model updating with mixed uncertainty, the basic concepts of maximum entropy and approximate Bayesian calculation are introduced. In section 3, the BME framework for measurement data and statistical information is proposed. In section 4, this framework is extended to incorporate imprecise probabilistic models where the ABC with novel WD-based metric is formulated for (hyper)parameters identification. In sections 5 and 6, a numerical example and an engineering application case are studied to verify the effectiveness of the proposed method. Section 7 provides concluding remarks.

2. Preliminary works

2.1 Uncertainty characterization

In the context of uncertainty quantification, input parameters of a model may involve epistemic uncertainty and aleatory uncertainty [24, 25]. Epistemic uncertainty is the uncertainty caused by the lack of knowledge, which can be reduced or even eliminated with the accumulation of additional knowledge. Aleatory uncertainty, on the other hand, is due to the inherent randomness, and is regarded as irreducible. For convenience, the associated uncertainty is classified into three categories and notated as follows.

- Category I is related with aleatory uncertainty. Input parameters in this category are denoted as ζ characterizing random variables with prescribed properties, such as distribution type, mean, variance.
- Category II corresponds to epistemic uncertainty. The input parameters belonging to this category are denoted as c, which possess an unknown but constant value.
- Category III denotes mixed uncertainty. The input parameters x with both aleatory and epistemic uncertainties are modeled as random variables with only vaguely determined uncertainty characteristics. An example of such model is a parametric probability box [24], where the parameters of a probability distribution (represented as θ_x in Figure 1) are characterized as interval variables.

Figure 1 gives an intuitive description of the three categories uncertainty.

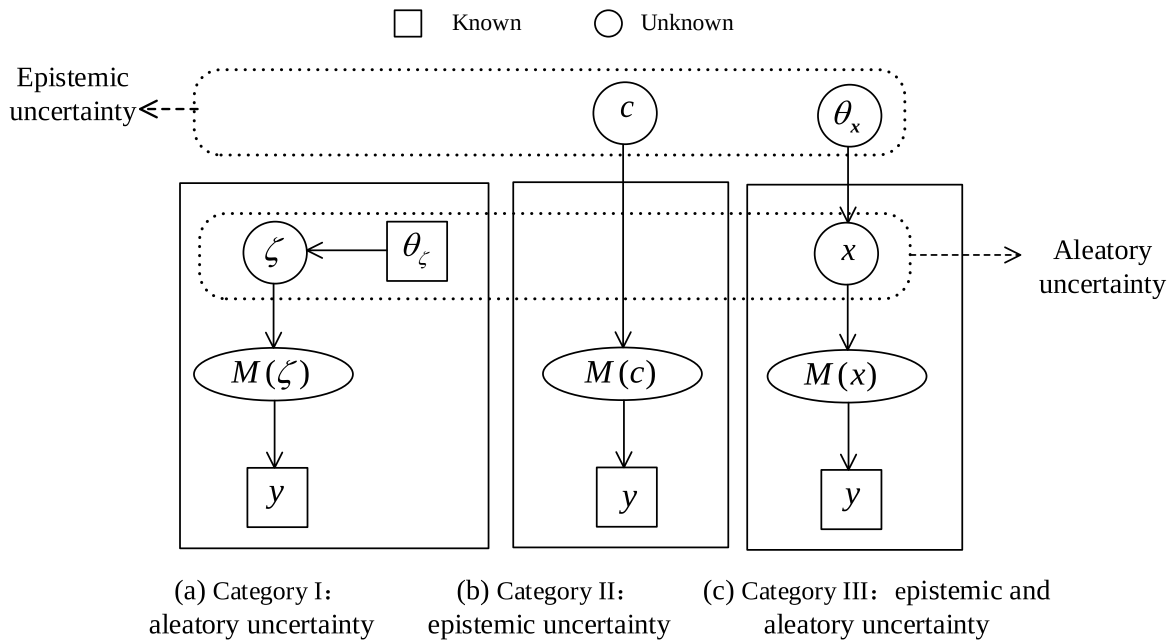


Figure 1 Intuitive description of the three categories uncertainty

2.2 Maximum Entropy method

The principle of maximum entropy was first proposed by Shannon [26]. Its main idea is that when only partial knowledge of an unknown distribution is available, the probability distribution that maximizes the entropy should be selected in accordance with that available knowledge. That is, when we need to establish the probability distribution of a random event, our prediction should meet all known conditions, and do not make any subjective assumptions about the unknown aspects. To illustrate, suppose our aim is to estimate the optimal (posterior) probability distribution $p(\xi)$ associated with parameter ξ in domain Ω based on the following information.

- A prior distribution $p_0(\xi)$ that summarizes a priori knowledge on the distribution associated with ξ.
- The *n*-order moment of ξ is given as follows [20]:

$$\int_\Omega p(\xi) f(\xi) d\xi = < f(\xi) > \tag{1}$$

  where $f(\xi)$ is ξ^n (*n* is an integer number) and $< f(\xi) >$ denotes the expected value of $f(\xi)$.
- The integral of (posterior) probability distribution over domain Ω should equal to 1, that is:

$$\int_\Omega p(\xi) d\xi = 1 \tag{2}$$

According to the Maximum Entropy principle, the posterior distribution is derived by maximizing the negative of the relative entropy (also known as Kullback-Leibler divergence) $S(p,p_0)$ between the posterior distribution $p(\xi)$ and prior distribution $p_0(\xi)$ [14]:

$$S(p, p_0) = -\int_\Omega p(\xi) \ln \frac{p(\xi)}{p_0(\xi)} d\xi \tag{3}$$

which is subject to the constraints imposed by Eq. (1) and (2). Such problem has a general solution as [20]:

$$p(\xi) = p_0(\xi) \mathrm{e}^{\beta f(\xi)} / Z \tag{4}$$

where $Z = \int_\Omega p_0(\xi) \mathrm{e}^{\beta f(\xi)} d\xi$ is the normalization constant, β is the so-called Lagrange multiplier, and the exponential term (entropy term) $\mathrm{e}^{\beta f(\xi)}$ represents constraints imposed by $f(\xi)$.

2.3 Approximate Bayesian Computation

Consider a deterministic forward model $\boldsymbol{y} = M(\boldsymbol{x})$ with *n*-dimensional input variables $\boldsymbol{x} = [x_1, x_2, ..., x_n]$ and *m*-dimensional output variables $\boldsymbol{y} = [y_1, y_2, ..., y_m]$. Given a collection of observations $\mathbf{Y}_{obs} = [\boldsymbol{y_1}, \boldsymbol{y_2}, ..., \boldsymbol{y}_{\boldsymbol{N_{obs}}}]^T$ with observation error, the Quantity of interest (QoI) can be estimated through the Bayesian inference [10]:

$$p(\boldsymbol{x} \mid \mathbf{Y}_{obs}) = \frac{L(\mathbf{Y}_{obs} \mid \boldsymbol{x}) p(\boldsymbol{x})}{\int_X L(\mathbf{Y}_{obs} \mid \boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x}} \tag{5}$$

where $p(\boldsymbol{x})$ is the prior distribution, $p(\boldsymbol{x} \mid \mathbf{Y}_{obs})$ is the posterior distribution, and $L(\mathbf{Y}_{obs} \mid \boldsymbol{x})$ is the likelihood function. $\int_X L(\mathbf{Y}_{obs} \mid \boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x}$ is the normalization constant. Given independent observations, the likelihood function is decomposed as:

$$L(\mathbf{Y}_{obs} \mid \boldsymbol{x}) = \prod_{i=1}^{N_{obs}} p(\boldsymbol{y}_i \mid \boldsymbol{x}) \tag{6}$$

In practical engineering problems, high dimensional input and output may lead to the time-consuming or even intractable calculation of full likelihood function. To address this obstacle, the ABC method has received much attention as it essentially reduces the computational cost by evaluating an approximate likelihood instead of the full likelihood. Typically, an approximate likelihood function with Gaussian kernel is considered [24]:

$$L(\mathbf{Y}_{obs} \mid \boldsymbol{x}) \propto \frac{1}{\sqrt{2\pi\varepsilon}} \exp\left\{-\frac{d(\mathbf{Y}_{obs}, \mathbf{Y}_{sim})^2}{2\varepsilon^2}\right\} \tag{7}$$

where ε is a predefined width factor controlling the concentration of the posterior distribution. A smaller ε leads to a more peaked posterior distribution but requires more calculation for convergence. The distance metric $d(\mathbf{Y}_{obs}, \mathbf{Y}_{sim})$ quantifies the discrepancy between the observed data $\mathbf{Y}_{obs}$ and simulated data $\mathbf{Y}_{sim}$. The selected metric plays a central role in the Bayesian model updating.

3. Bayesian Maximum Entropy framework for heterogeneous information fusion

This section proposes a Bayesian Maximum Entropy framework for model updating using heterogeneous information. For the sake of simplicity, the current section assumes that the input and output of the model are scalars, that is $y=M(x)$. However, the material reported here can be extended towards the more general case of vector-valued input and output $\boldsymbol{y}=M(\boldsymbol{x})$.

3.1. Measurement data

The measurement data here denotes the direct point observations of a system response e.g. sensor readings or experimental outcomes. The Bayesian updating model for the parameter x based on N_{obs} observations of the output response variable y, that is $\mathrm{Y}_{obs}= [y_1, y_2, \dots , y_{Nobs}]^{\mathrm{T}}$, is shown below:

$$p(x \mid \mathrm{Y}_{obs}) = \frac{L(\mathrm{Y}_{obs} \mid x)\, p_0(x)}{\int_X L(\mathrm{Y}_{obs} \mid x)\, p_0(x) dx} \propto L(\mathrm{Y}_{obs} \mid x)\, p_0(x) \tag{8}$$

As an alternative to the Bayes' rule, ME can be used to estimate the posterior distribution by maximizing the negative relative entropy between the prior joint distribution $p_0(y,x)$ and the posterior joint distribution $p(y,x)$ under the constraints of point data [14]. In this case, the Dirac delta (denoted as δ) is used to convert point data into constraints:

$$\int_X p(y,x)dx = \delta(y - y_i), i = 1,2,\dots,N_{obs} \tag{9}$$

and the normalization constraints is obtained as:

$$\int_{Y\times X} p(y,x)dydx = 1 \tag{10}$$

The optimal posterior distribution is obtained as Eq. (11) by maximizing the negative relative entropy under point data constraints and normalization constraints

$$p_1(x) = p_0(x) p_0(\mathrm{Y}_{obs} \mid x) / Z_1 = p_0(x) \prod_{i=1}^{N_{obs}} p_0(y_i \mid x) / Z_1 \tag{11}$$

where $Z_1 = \int_X p_0(x) p_0(\mathrm{Y}_{obs} \mid x) dx$ is the normalization constant. It is observed that the $p_1(x)$ derived in Eq. (11) is exactly the same as that derived in Eq. (8). Basically, the proposed BME framework is equivalent to the Bayes' rule when only point data is considered [18].

3.2. Statistical information

In some cases, direct measurements are not available and such information is given in a synthetized form as statistical information. Common forms of statistical information include summary statistics, such as the frequency of observations for discrete variables (e.g. reliability information) and the mean and variance of continuous variables (e.g. moment information). In the context of reliability analysis, a common form of available information is summarized reliability data for various mechanical components (e.g. failure rates or failure probabilities) instead of detailed actual test data.

3.2.1. Reliability information

One way to determine the reliability of a component or system is to perform *m* independent tests and record whether the performance $y_i, i=1,...,m$ of the system exceeds a certain specified threshold y_c. The system performance is then synthetized into the ratio of the number of tests not exceeding the threshold value to the total number of tests. Thus, the reliability of the sample is obtained as: $R^{'}=\sum_{i=1}^{m} H(y_i-y_c)/m$, where *H* is the Heaviside step function. This type of reliability data can be included to perform model updating in the proposed BME framework. Specifically, the negative relative entropy between prior and posterior distribution is given as:

$$S(p_2,p_0)=-\int_X p_2(x)\ln\frac{p_2(x)}{p_0(x)}dx \tag{12}$$

where $p_0(x)$ and $p_2(x)$ are the prior and posterior distributions respectively. The available reliability information is transformed into constraints as:

$$\int_X p_2(x)R(x)dx = R^{'} \tag{13}$$

where $R^{'}$ denotes the reliability value and $R(x)$ is reliability function given in Eq. (14) associated to parameter *x* considering output variable followed a Gaussian distribution [27]:

$$R(x)=P(y\le y_c)=\Phi\left[\frac{y_c-M(x)}{\sigma}\right] \tag{14}$$

where $P(\bullet)$ is probability distribution function, $\Phi(\bullet)$ is the cumulative function of the standard normal distribution, σ is the standard deviation of y and $M(x)$. The optimal posterior distribution is solved by maximizing the negative relative entropy under the reliability constrain. The method of Lagrange multipliers is used and its function is given as

$$F(p_2,p_0,\beta_1)=S(p_2,p_0)+\beta_1\left[\int_X p_2(x)R(x)dx-R^{'}\right] \tag{15}$$

where β_1 is the Lagrange multiplier corresponding to reliability data constraint and the optimal posterior distribution is derived by setting $\frac{\partial F}{\partial p_2}=0$ as

$$p_2(x)=p_0(x)\exp(\beta_1 R(x))/Z_2 \tag{16}$$

where $Z_2=\int_X p_0(x)\exp(\beta_1 R(x))dx$ is the integration constant, and the Lagrange multiplier β_1 is derived by solving the following equation:

$$\frac{\partial \ln Z_2}{\partial(\beta_1)}=\frac{\int_X p_0(x)\exp(\beta_1 R(x))R(x)dx}{\int_X p_0(x)\exp(\beta_1 R(x))dx}=R^{'} \tag{17}$$

Regarding numerical implementation, this equation can be solved , for example, with the *fsolve* function in MATLAB software, and the integral part is evaluated using the integral function integral/integral2.

3.2.2. Moment information

The moment information includes the mean, variance and percentage quantities of direct observations. In the proposed BME framework, such moment information is converted into various constraints and used for model updating. Typically, to moment information in its mean-value form, we have:

$$\int_X p_3(x)M(x)dx = \bar{y} \tag{18}$$

where $p_3(x)$ is the posterior probability distribution under the constraint of moment information, X is the domain of x, and $\bar{y}$ is the expected value of y. The second order moment information (variance) can be expressed as $\int p(x)(M(x)-\bar{y})^2 dx$. Similarly, the Lagrange function in the case of moment information is given as

$$F(p_3, p_0, \beta_2) = S(p_3, p_0) + \beta_2\left[\int_X p_2(x)R(x)dx - \bar{y}\right] \tag{19}$$

where $S(p_3,p_0)$ is the negative relative entropy between prior distribution $p_0(x)$ and posterior distribution $p_3(x)$, β_2 is the Lagrange multiplier corresponding to the moment constraint. The optimal posterior distribution is derived by setting $\frac{\partial F}{\partial p_3} = 0$ as

$$p_3(x) = p_0(x)\exp(\beta_2 M(x)) / Z_3 \tag{20}$$

where $Z_3 = \int_X p_0(x)\exp(\beta_2 M(x))dx$ is the integration constant, and β_2 is a constant obtained by solving the following equation:

$$\frac{\partial \ln Z_3}{\partial(\beta_2)} = \frac{\int_X p_0(x)\exp(\beta_2 M(x))M(x)dx}{\int_X p_0(x)\exp(\beta_2 M(x))dx} = \bar{y} \tag{21}$$

3.3. Bayesian Maximum Entropy method for heterogeneous information fusion

For the information aggregation of measurement data, reliability and moment information, we propose a comprehensive BME model. It is carried out by maximizing the negative relative entropy between the prior joint distribution $q(y,x)$ and posterior joint distribution $p(y,x)$

$$S(p,q) = -\int_{Y\times X} p(y,x)\ln\frac{p(y,x)}{q(y,x)}dydx \tag{22}$$

where $q(y,x) = p_0(x)p_0(y\,|\,x)$. The expression in Eq. (22) should be maximized considering the all available information shown in Fig. (2). Essentially, this information is transformed into multiple constrains in Eq. (23) that need to be satisfied.

$$s.t.\begin{cases} \int_X p(y,x)dx = \delta(y-y_i), i = 1,2,\ldots,N_{obs} \\ \int_{Y\times X} p(y,x)dydx = 1 \\ \int_{Y\times X} p(y,x)R(x)dydx = R' \\ \int_{Y\times X} p(y,x)M(x)dydx = \bar{y} \end{cases} \tag{23}$$

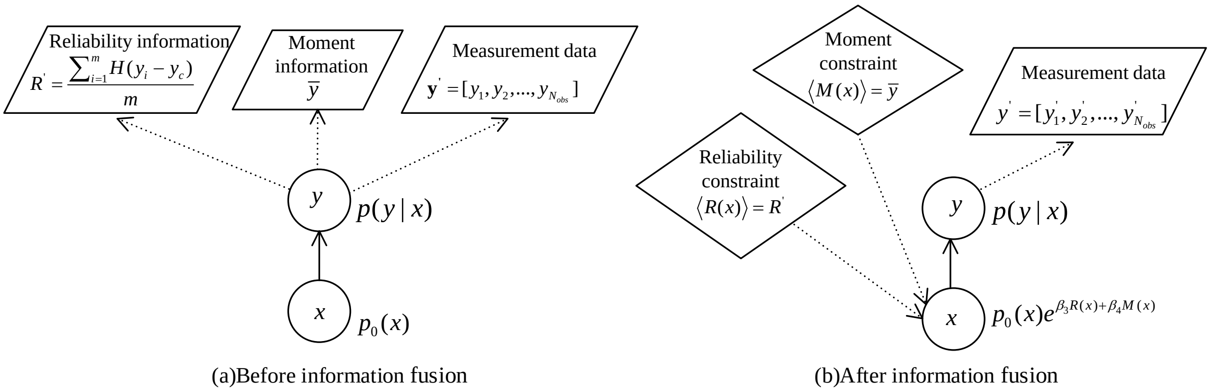


Figure 2 The structure for information fusion.

In the BME framework, the measurement data is processed in the traditional Bayesian updating way while the moment information is included in the associated entropy terms, which allows a broader information fusion. By solving the corresponding optimization problem (for detailed procedure, please refer the Appendix), the optimal posterior joint distribution is derived as Eq. (24) includes three different sources of information

$$p(x) = p_0(x)\exp(\beta_3 R(x) + \beta_4 M(x))\prod_{i=1}^{N_{obs}} p(y_i \mid x) / Z_4 \tag{24}$$

where $Z_4 = \int_X p_0(x)\exp(\beta_3 R(x) + \beta_4 M(x))\prod_{i=1}^{N_{obs}} p(y_i \mid x)dx$ is the integration constant, β_3 and β_4 are the Lagrange multipliers corresponding to reliability data constraint and moment data constraint.

4. Bayesian Maximum Entropy method for imprecise probabilistic model updating

The proposed BME approach is capable of incorporating multi-type information (i.e. precise measurements, reliability information and moment information) under a comprehensive framework for model updating. However, the extension towards imprecise probabilistic model can be challenging, as the derivation of an analytical likelihood function is cumbersome or even intractable as the mixed uncertainty (Category III) is involved. To fill this gap, we employ ABC to avoid the cumbersome evaluation of full likelihood function. In addition, to address the hyperparameters of imprecise probabilistic model, a novel WD-based UQ metric is developed to capture higher order information in presence of mixed uncertainty.

4.1. Bayesian Maximum Entropy method for Imprecise probability model updating

The imprecise probability model includes epistemic uncertainty and aleatory uncertainty. Consider the model $y=M(x \mid \theta_x)$ that the probability distribution function P of the intermediate variable x is governed by the hyper-parameter θ_x : X~$P(x \mid \theta_x)$. Note that the hyper-parameter is affected by epistemic uncertainty (due to lack of data) and make the distribution of intermediate variable x imprecise (that is, a parametric probability box [24]). When output response data Y_{obs} is available, Bayes' theorem can be used to update the parameters as follows:

$$P(x,\theta_x \mid \mathrm{Y}_{obs}) = \frac{L(\mathrm{Y}_{obs} \mid x)P(x \mid \theta_x)P(\theta_x)}{\int_{X\times\Theta} L(\mathrm{Y}_{obs} \mid x)P(x \mid \theta_x)P(\theta_x)dxd\theta_x} \tag{25}$$

where $P(\theta_x)$ is the hyper-prior distribution, $\int_{X\times\Theta} L(\mathrm{Y}_{obs} \mid x)P(x\mid\theta_x)P(\theta_x)dxd\theta_x$ is the normalization constant, $P(x,\theta_x \mid \mathrm{Y}_{obs})$ is the posterior distribution, $L(\mathrm{Y}_{obs} \mid x)$ is the likelihood function, and $P(x|\theta_x)$ is probability of intermediate variable x conditioned on the hyper-parameter θ_x. The (prior) joint probability distribution of output variable y, intermediate variable x and hyper-parameter θ_x is decomposed as:

$$p_0(y,x,\theta_x) = p_0(y\mid x)p_0(x\mid\theta_x)p_0(\theta_x) \tag{26}$$

To estimate the optimal posterior distribution, we calculate the negative relative entropy of prior joint probability distribution $p_0(y,x,\theta_x)$ and posterior joint probability distribution $p(y,x,\theta_x)$:

$$S(p,p_0) = -\int_{\mathrm{Y}\times\mathrm{X}\times\Theta} p(y,x,\theta_x)\ln\frac{p(y,x,\theta_x)}{p_0(y,x,\theta_x)}dydxd\theta_x \tag{27}$$

The available reliability information is rewritten as the following constraint:

$$\int_{\mathrm{Y}\times\mathrm{X}\times\Theta} p(y,x,\theta_x)R(\theta_x)dydxd\theta_x = R^{'} \tag{28}$$

where R' denotes the reliability value and $R(\bullet)$ corresponds to the reliability function.

The first order moment information of the output response y is transformed as:

$$\int_{\mathrm{Y}\times\mathrm{X}\times\Theta} p(y,x,\theta_x)M(x\mid\theta_x)dydxd\theta_x = \bar{y} \tag{29}$$

where $y=M(x|\theta_x)$, $\bar{y}$ is the expected value of y. The measurement data $\mathrm{Y}_{obs}=[y_1, y_2, \ldots, y_{Nobs}]^{\mathrm{T}}$ is transformed as:

$$\int_{\mathrm{X}\times\Theta} p(y,x,\theta_x)dxd\theta_x = \delta(y-y_i), i=1,2,\ldots,N_{obs} \tag{30}$$

in addition, there is the normalization constraint:

$$\int_{\mathrm{Y}\times\mathrm{X}\times\Theta} p(y,x,\theta_x)dydxd\theta_x = 1 \tag{31}$$

In this case, the Lagrange function is given as

$$\begin{aligned} F(p,p_0,\alpha,\beta,\lambda,\eta) = {} & S(p,p_0) + \alpha\left[\int_{\mathrm{Y}\times\mathrm{X}\times\Theta} p(y,x,\theta_x)dydxd\theta_x - 1\right] + \beta\left[\int_{\mathrm{Y}\times\mathrm{X}\times\Theta} p(y,x,\theta_x)R(\theta_x)dydxd\theta_x - R^{'}\right] \\ & \lambda\left[\int_{\mathrm{Y}\times\mathrm{X}\times\Theta} p(y,x,\theta_x)M(x\mid\theta_x)dydxd\theta_x - \bar{y}\right] + \int_{\mathrm{Y}} \eta(y)\left[\int_{\mathrm{X}\times\Theta} p(y,x,\theta_x)dxd\theta_x - \delta(y-y_i)\right]dy \end{aligned} \tag{32}$$

where the Lagrange multiplier $\alpha,\beta,\lambda,\eta$ corresponds to the constraint of Eq. (28-31). Then, the optimal posterior joint distribution is obtained by setting $\frac{\partial F}{\partial p} = 0$ as

$$p(y,x,\theta_x) = p_0(y,x,\theta_x)e^{1+\alpha+\beta R(\theta_x)+\lambda M(x|\theta_x)+\eta(y)} = p_0(y,x,\theta_x)e^{\beta R(\theta_x)+\lambda M(x|\theta_x)+\eta(y)}\frac{1}{Z} \tag{33}$$

where $Z = \int_{Y\times X\times\Theta} p_0(y,x,\theta_x)e^{\beta R(\theta_x)+\lambda M(x|\theta_x)+\eta(y)}dydxd\theta_x$ is the integration constant. Substitution of Eq. (33) into Eq. (30) yields

$$\frac{e^{\eta(y)}}{Z} = \frac{\delta(y-y_i)}{p_0(y)\int_{\mathrm{X}} p_0(x,\theta_x\mid y_i)e^{\beta R(\theta_x)+\lambda M(x|\theta_x)}dxd\theta_x} \tag{34}$$

Eq. (34) can be alternatively expressed as:

$$p(y,x,\theta_x) = p_0\left(x,\theta_x \mid y_i\right)\mathrm{e}^{\beta R(\theta_x)+\lambda M(x|\theta_x)}\frac{\delta(y-y_i)}{\int_{\mathrm{X}\times\Theta} p_0(x,\theta_x \mid y_i)e^{\beta R(\theta_x)+\lambda M(x|\theta_x)}dxd\theta_x} \tag{35}$$

The posterior distribution of unknown parameter x and hyperparameter θ_x is derived by integrating out y:

$$p(x,\theta_x) = \frac{p_0\left(x,\theta_x \mid \mathrm{Y}_{obs}\right)\mathrm{e}^{\beta R(\theta_x)+\lambda M(x|\theta_x)}}{\int_{\mathrm{X}\times\Theta} p_0\left(x,\theta_x \mid \mathrm{Y}_{obs}\right)e^{\beta R(\theta_x)+\lambda M(x|\theta_x)}dxd\theta_x} = \prod_{i=1}^{N_{obs}} p_0(y_i \mid x)p_0(x \mid \theta_x)p_0(\theta_x)\mathrm{e}^{\beta R(\theta_x)+\lambda M(x|\theta_x)} / Z' \tag{36}$$

where $Z' = \int_{X\times\Theta}\prod_{i=1}^{N_{obs}} p_0(y_i \mid x)p_0(x \mid \theta_x)p_0(\theta_x)\mathrm{e}^{\beta R(\theta_x)+\lambda M(x|\theta_x)}dxd\theta_x$ is the integration constant, and (β,λ) is a constant obtained by solving $\partial \ln Z' / \partial(\beta) = R'$ and $\partial \ln Z' / \partial(\lambda) = \bar{y}$.

4.2. Approximate Bayesian computation with Wasserstein distance-based UQ metric

The proposed BME approach has been extended to imprecise probability models in the preceding section. However, in presence of mixed uncertainty, the evaluation of full likelihood function is much more challenging as additional hyperparameters are introduced. To address this issue, ABC is employed as it facilitates the process by evaluating an approximate likelihood function instead a complete one. The original likelihood function in Eq. (36) is replaced by an approximate likelihood function, and the posterior distribution is obtained as follows:

$$p(\theta_x \mid D) \propto p_0(\theta_x)e^{-\frac{d^2}{2\varepsilon^2}}e^{\beta R(\theta_x)+\lambda M(x|\theta_x)} \tag{37}$$

To perform the ABC, a suitable UQ metric is required to measure the divergence between the observed dataset and the simulated samples.

Given the observed dataset $\mathrm{Y}_{\mathrm{obs}}$ and the simulated sample set $\mathrm{Y}_{\mathrm{sim}}$ obtained by running the forward model $y=M(x)$ multiple times, the popular Euclidean distance (ED)-based metric is presented as [28]:

$$d_E(\mathrm{Y}_{obs},\mathrm{Y}_{sim}) = \sqrt{(\bar{\mathrm{Y}}_{obs}-\bar{\mathrm{Y}}_{sim})(\bar{\mathrm{Y}}_{obs}-\bar{\mathrm{Y}}_{sim})^T} \tag{38}$$

where $\bar{\mathrm{Y}}_{obs}$ and $\bar{\mathrm{Y}}_{sim}$ represent the mean values of observed data and simulated samples respectively. It is apparent that ED measures the absolute distance between the mean values of the two sample sets, which is applicable for parameter calibration when few parameters with epistemic uncertainty (i.e. constants with unknown value) are identified. However, in case a large number of parameters with mixed uncertainty (i.e. imprecise random variables) is being identified, ED-based metric becomes problematic as it is not able to capture the divergence of higher-order information, such as variance and covariance.

In this context, some other alternatives such as the Bhattacharyya distance (BD) [24] is useful as it provides a means for measuring the degree between probability densities associated with observed and simulated data. Nonetheless, when these two data sets are far apart (i.e., there is no overlap), BD tends to infinity, which may lead to implementation issues. By contrast, the WD is capable of capturing the divergence information between two different sample sets while avoiding the aforementioned implementation issues, so as to serve a promising UQ metric for BME model updating [23]. The p-WD [29] is defined as follows:

$$d_{W_p}\left(\mathrm{Y}_{obs},\mathrm{Y}_{sim}\right)=\left(\int_{-\infty}^{+\infty}|F_{\mathrm{Y}_{obs}}(y)-F_{\mathrm{Y}_{sim}}(y)|^p\,dy\right)^{1/p}$$
$$=\left(\int_0^1|F_{\mathrm{Y}_{obs}}^{-1}(\alpha)-F_{\mathrm{Y}_{sim}}^{-1}(\alpha)|^p\,d\alpha\right)^{1/p},\ p\geq 1 \tag{39}$$

where $F_{\mathrm{Y}_{obs}}(y)$ and $F_{\mathrm{Y}_{sim}}(y)$ are cumulative probability density functions of $\mathrm{Y}_{\mathrm{obs}}$ and $\mathrm{Y}_{\mathrm{sim}}$ respectively, $F_{\mathrm{Y}_{obs}}^{-1}(\alpha)$ and $F_{\mathrm{Y}_{sim}}^{-1}(\alpha)$ are the α quantile functions, and p represents the dimension. For one dimensional case, i.e. $p = 1$, the WD is the area between the two marginal CDFs. The following Fig. 3 shows an intuitive description of the one-dimensional WD, which is used in our study.

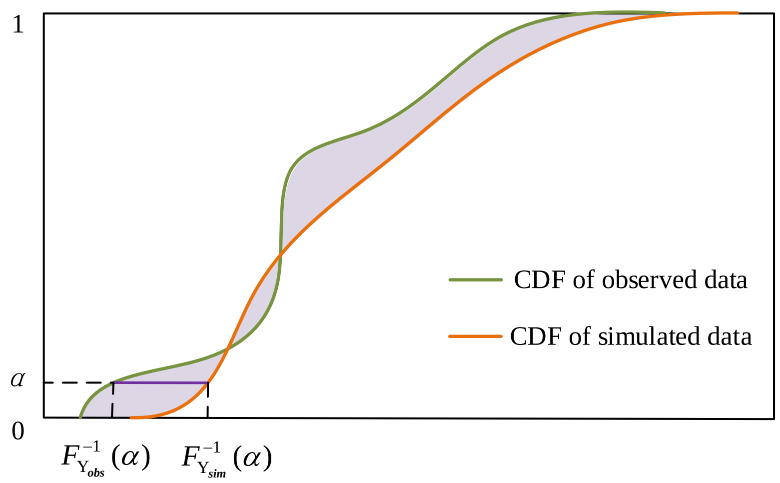


Figure 3 Intuitive description of one-dimensional Wasserstein distance

To illustrate the benefits of using WD as UQ metric, consider a simple model $y = x^2$ where x follows a Gaussian distribution with imprecise parameters (μ, σ). 50 observed samples $\mathrm{Y}_{\mathrm{obs}}= [y_1, \ldots, y_{50}]^{\mathrm{T}}$, moment information $\bar{y}$ =64 and reliability information R'=0.65 are generated based on the true values μ_0=8, σ_0=1. Given a prior distribution $\mu \sim Unif\,(0,20)$, $\sigma \sim Unif\,(0,5)$. The posterior distributions of (μ, σ) by employing the WD-based metric with different *p-value* are shown in Figure 4. Table 1 shows the statistical results of μ and σ for different *p-value*. The average deviation first decreased and then increased with the increase of *p-value*, and the result of p=3 is the best. A larger p requires more calculation time. The posterior distributions of (μ, σ) by employing the ED-based, BD-based and the WD-based metric with p=3 are shown in Figure 5. Table 2 shows statistical results of posterior distributions for different metrics. The average deviation with the WD-based metric is the smallest.

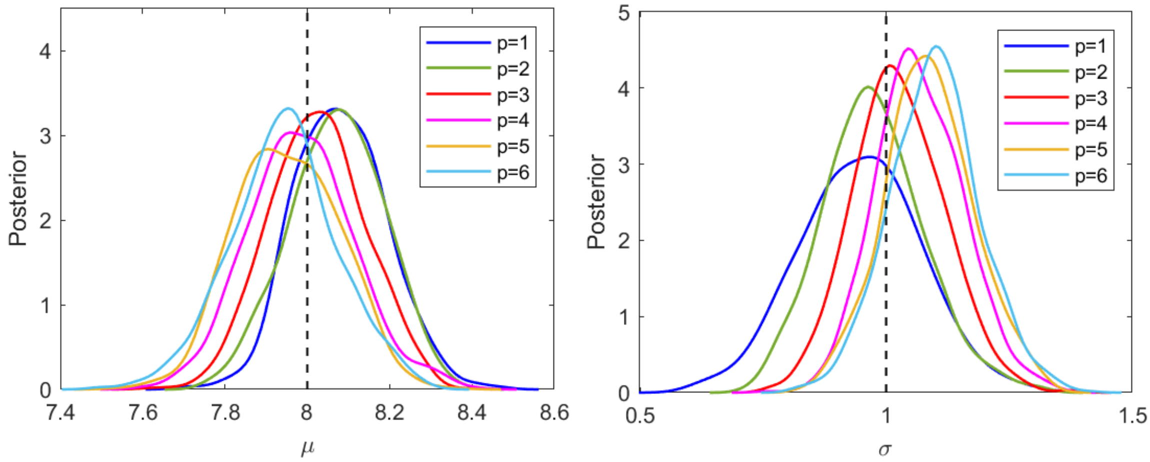


Figure 4 The results based on WD with different p-value.

Table 1 Statistical results of μ and σ for different p-value.

p-value	Posterior mean (deviation)		Average deviation	Time(s)
	μ	σ		
p=1	8.0811(1.01%)	0.9478(5.22%)	3.12%	52
p=2	8.0713(0.89%)	0.9745(2.55%)	1.72%	60
p=3	8.0240(0.30%)	1.0260(2.60%)	1.45%	80
p=4	7.9847(0.19%)	1.0685(6.85%)	3.52%	96
p=5	7.9520(0.60%)	1.0870(8.70%)	4.65%	119
p=6	7.9421(0.72%)	1.0986(9.86%)	5.29%	136

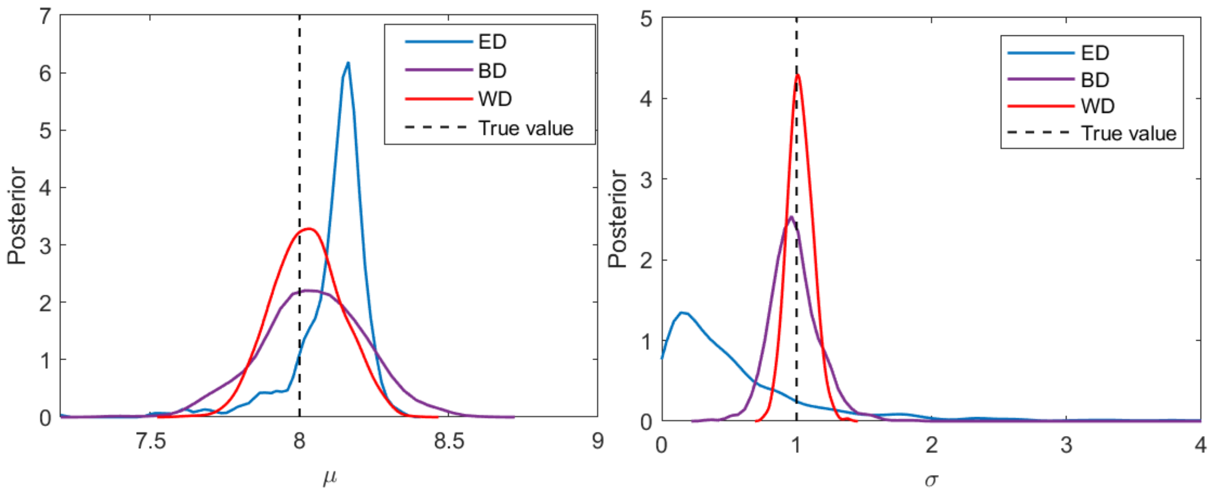


Figure 5 The posterior of parameters with different metrics.

Table 2 Statistical results of μ and σ for different metrics.

Metrics	Parameters	25% quantile	75% quantile	Mean	Average deviation
ED	μ	8.0736	8.1791	8.0903	1.43%
	σ	0.1609	0.7227	0.5733	51.44%
BD	μ	7.9310	8.1620	8.0015	0.97%
	σ	0.8731	1.0913	1.0382	8.55%
WD	μ	7.9434	8.1013	8.0240	0.76%
	σ	0.9623	1.0879	1.0260	5.05%

4.3. Algorithm for Implementation

From an implementation viewpoint, the main steps of our approach are summarized as follows:

***Step* 1** Determine the object function using Eq. (27) and construct the required constraints using Eq. (28) and (29). Then solve the corresponding optimization problem by using the Lagrange multiplier method (see in Appendix).

***Step* 2** Generate n samples $\boldsymbol{\theta}_x=[\theta_1, \theta_2, \ldots, \theta_n]$ from the prior distribution $p(\theta_x)$. For each sample θ_i, generate enough N samples for intermediate variable x given $X \sim p(x|\theta_i)$. These samples are then propagated through the forward model $y = M(x)$ to obtain the corresponding samples for the output Y_{sim}.

***Step* 3** Calculate the WD between the simulated sample set Y_{sim} and the observed data set Y_{obs} by using Eq. (39), and then build the approximate likelihood function based on the derived WD-based metric by using Eq. (7).

***Step* 4** According to Eq. (37), perform transitional Markov chain Monte Carlo (TMCMC) algorithm under the proposed BME framework. The TMCMC algorithm allows sampling from intermediate probability distribution functions (PDFs). Intermediate distributions $p_j \propto p_0(\theta_x)\left(e^{-\frac{d^2}{2\varepsilon^2}} e^{\beta R(\theta_x)+\lambda M(x|\theta_x)}\right)^{\alpha_j}$ are computed as the product of the prior distribution and likelihood and exponential function scaled by a transitional coefficient $0 < \alpha_j \leq 1$. The method starts from $\alpha_j = 0$ and gradually transitions to $\alpha_j = 1$. The first distribution p_0 is the prior distribution, and the last is the posterior distribution.

For the sake of clarity, a flowchart of the proposed BME approach based on WD is shown in Figure 6.

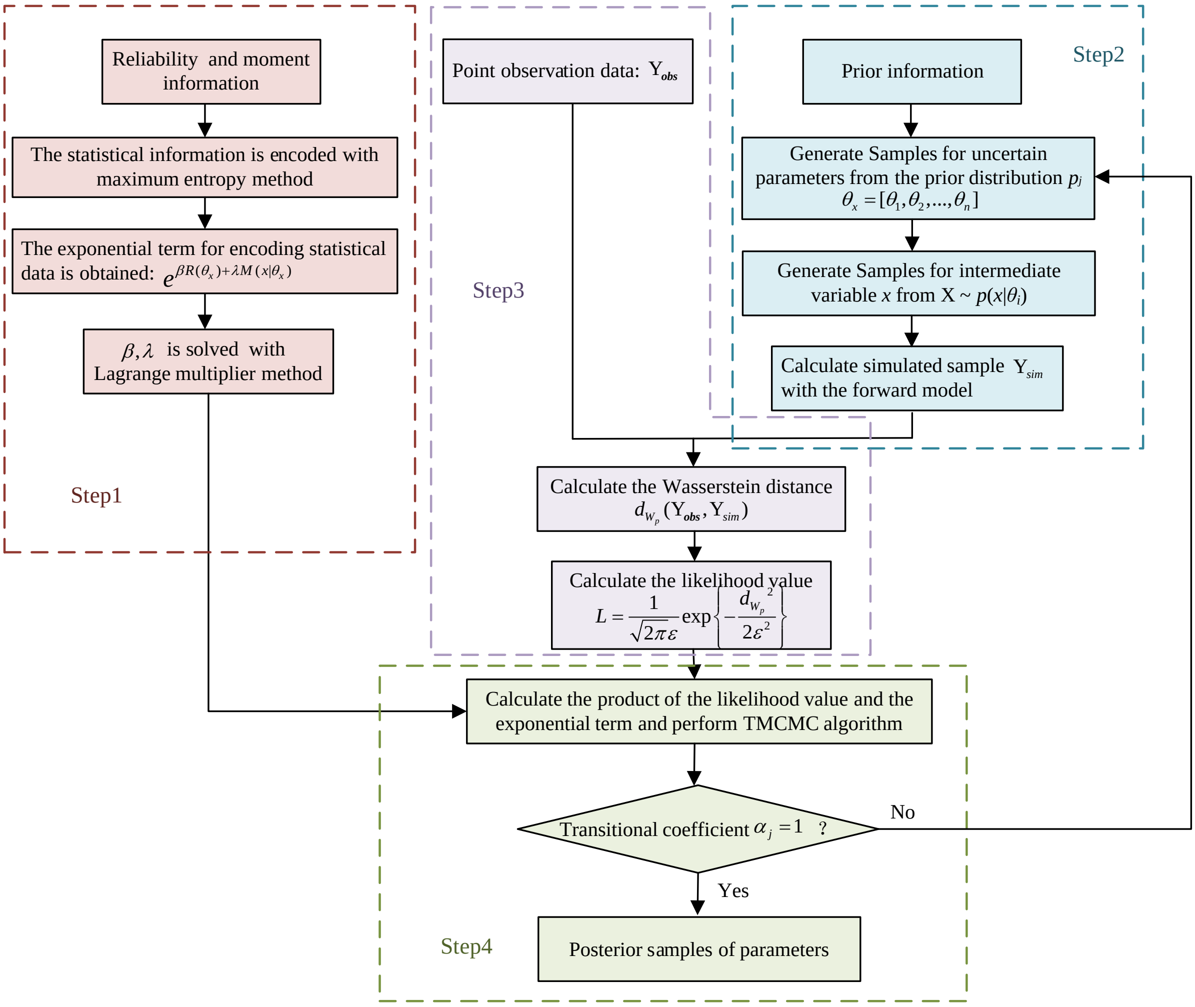


Figure 6 Flowchart of the proposed BME-WD approach.

5. Test Example: Simply supported beam

In this section, we use a numerical case to validate the proposed method. Fig 7. shows a simply supported beam subjected to uniformly distributed load p. Due to uncertainties in the fabrication process, all geometric parameters shown in Fig. 7 are not precisely known and are described with (truncated) normal distributions with coefficient of variation 0.1. According to the Euler-Bernoulli beam theory, the mid-span deflection V_{mid} of the beam is given as

$$V_{mid} = M(E, p, L, b, h) = \frac{5}{32} \frac{pL^4}{Ebh^3} \tag{40}$$

where E represents the Young's modulus while L, b and h are length, width and height of the beam respectively. For validation purpose, two subcases are demonstrated with different parameters settings: (1) parameters with category I (aleatory) and category II (epistemic) uncertainty are considered; (2) parameters with category I (aleatory) and category III (mixed) uncertainty are considered.

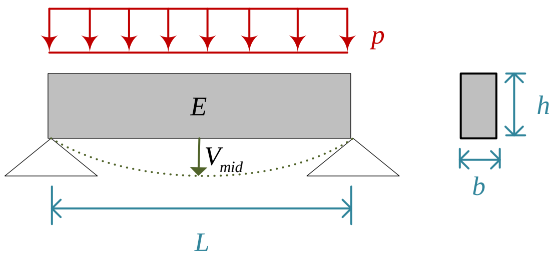


Figure 7 Simple supported beam.

5.1. Subcase 1: Probability model with disjoint aleatory and epistemic uncertainty

In this subcase, parameters (*b, h, L, p*) are set as aleatory uncertainty variables while the Young's modulus *E* is an epistemic uncertain parameter to be calibrated. The specific parameter settings are shown in Table 3. To model the Young's Modulus *E*, in the Bayesian framework, its prior distribution is assigned as a lognormal distribution where the distribution parameters are μ_E = 3.4GPa and σ_E = 0.15GPa [30]. An error term $e \sim Gaussian(0,0.008)$ is included to represent the noise associated with mid-span deflection.

180 raw data of the output V_{obs} are derived based on the "true" value of Young's modulus E=20GPa and independently generated samples of (*b, h, L, p*) by using Eq. (39). Among them, 50 data points are used to yield the first moment information $\overline{V} = 0.0145m$ and another 100 data points are used to yield the reliability information R=0.67 by setting the deflection threshold V_{crit} = 0.018m. The 30 left data points are kept as measurement data.

Table 3 Scenario 1: parameter settings for simple supported beam model.

Category	Parameter	Value	Uncertainty characteristics
I	b	$Guassian, \mu_1 = 0.15m, \sigma_1 = 0.015m$	Aleatory uncertainty
I	h	$Guassian, \mu_2 = 0.3m, \sigma_2 = 0.03m$	Aleatory uncertainty
I	L	$Guassian, \mu_3 = 5m, \sigma_3 = 0.5m$	Aleatory uncertainty
I	p	$Guassian, \mu_4 = 12000N/m, \sigma_4 = 100N/m$	Aleatory uncertainty
II	E	Unknown constant	Epistemic uncertainty

The TMCMC algorithm [31] is applied to perform the proposed BME approach as it allows to sample from intermediate probability distribution functions (PDFs) and can be used to effectively perform identification in the parameter space. Intermediate distributions are computed as the product of the prior distribution and likelihood and exponential function scaled by an exponent parameter $0<\alpha_j\leq 1$. The method starts from α_j=0 and gradually transitions to α_j=1. Thus, the first distribution is the prior PDF, and the last is the posterior distribution. In this subcase, a posterior distribution model using three types of data is formulated as

$$p(E|D) \propto p_0(E)\prod_{i=1}^{30} p(V_i \,|\, E)e^{\beta_1 R(E)+\beta_2 M(E)} \tag{41}$$

We first carry out the simulation by using the point data, reliability information, and moment information separately, corresponding to β_1=β_2=0, β_2=0 and β_1=0 in Eq. (41), then the posterior distributions of Young's modulus *E* are derived and shown in Fig. 8. It can be seen that the three posterior distributions in different scenarios all converge to the true value, however, the curve produced using point data is more concentrated with less uncertainty, followed by the curve of using moment information and that of using reliability information. This is not surprising as the reliability information and moment information are indeed "manufactured" from the raw point data and lose some of the information contained during the process. By contrast, Fig. 9 shows the estimation results of using both statistical information and point data. It can be seen that the proposed BME approach can effectively fuse multi-type information and produce better estimation results with less uncertainty. It is noted that, in this specific case, only using the point data already achieved satisfying estimation results. However, it does not mean that solely using point data is sufficient for all cases. Simply abandoning those imprecise (but valuable) information seems lack of scientific foundation. In a more general sense, it is reasonable and desirable to aggregate all available information to produce a comprehensive estimation result.

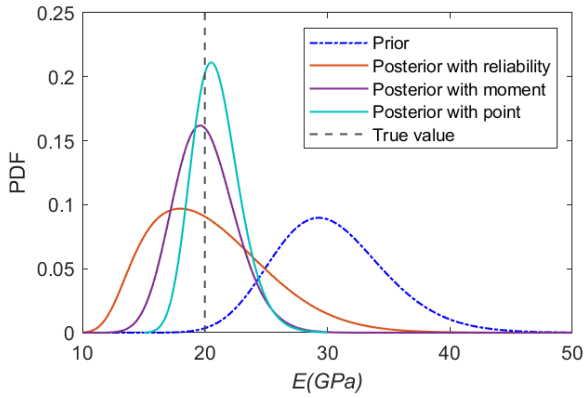


Figure 8 Posterior distribution with three types of data separately.

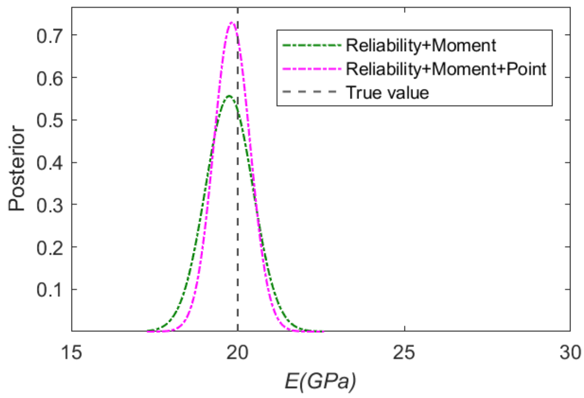


Figure 9 Posterior distribution of data fusion.

5.2. Subcase 2: imprecise probability model with mixed uncertainty

In this subcase, parameter settings of (*b, h, L, p*) remain unchanged while the Young's modulus *E* is assumed to follow a LogNormal distribution $LogNormal(\mu_E,\ \sigma_E)$ with unknown hyperparameters to incorporate the mixed uncertainty. We assign diffuse prior distributions $\mu_E \sim Uniform(0{,}20GPa)$ and $\sigma_E \sim Uniform(0{,}2GPa)$ to the unknown hyperparameters since we only roughly know their lower and upper bounds. Based on the predetermined true values $\mu_E = 3GPa$ and $\sigma_E = 0.15$GPa, 180 samples of *E* are generated and are combined with independently generated samples of (*b, h, L, p*) to produce corresponding output samples. Similarly as in the previous Subcase, 100 raw data of *V* are used to produce the reliability information *R*=0.67 by adopting the threshold V_{crit} = 0.018*m*. This process is equivalent to evaluate the reliability for a batch of 100 nominal identical beams. The moment information $\bar{V} = 0.0145m$ is obtained by calculating the mean value of another 50 data points. The left 30 data points are used as direct observations.

Table 4 Scenario 2: parameter settings for simple supported beam model.

Category	Parameter	Value	Uncertainty characteristics
I	*b*	$Guassian, \mu_1 = 0.15m, \sigma_1 = 0.015m$	Aleatory uncertainty
I	*h*	$Guassian, \mu_2 = 0.3m, \sigma_2 = 0.03m$	Aleatory uncertainty
I	*L*	$Guassian, \mu_3 = 5m, \sigma_3 = 0.5m$	Aleatory uncertainty
I	*p*	$Guassian, \mu_4 = 12000N/m, \sigma_4 = 100N/m$	Aleatory uncertainty
III	*E*	$LogNormal(\mu_E, \sigma_E)$	Mixed uncertainty

To address the imprecise probabilistic model updating, the developed WD-based UQ metric is embedded into the stochastic model updating framework. The posterior distributions model of the two hyperparameters of interest are formulated as

$$p(\mu_E, \sigma_E \mid D) \propto p_0(\mu_E, \sigma_E) e^{-\frac{d(V_{obs} - V_{sim})^2}{2\varepsilon^2}} e^{\lambda_1 R(\mu_E, \sigma_E) + \lambda_2 M(\mu_E, \sigma_E)} \tag{42}$$

As an imprecise parameter, the QoIs here are actually the hyperparameters μ_E and σ_E. The posterior samples are obtained by using the TMCMC algorithm, after 7 iterations, the transition coefficient α_j=1 indicates the updated samples have converged to the target distribution as shown in Fig. 10. It shows that the posterior distributions converge to the true value indicating a successful model updating. Figure 11 shows a comparison between the cases using point data, reliability information, moment information and fused information respectively. Table 5 gives a summary of the statistical results. It suggests that the uncertainty is gradually reduced as more information added in.

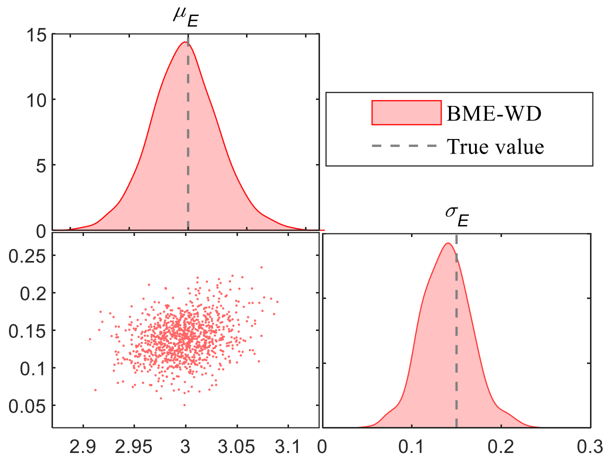


Figure 10 Posterior of hyperparameters with the proposed BME-WD method.

Table 5 Updated results of hyperparameters (μ_E, σ_E) .

Cases	Posterior mean (True value)		Relative error	
	μ_E	σ_E	μ_E	σ_E
Reliability	3.0716 (3.00)	0.8789 (0.15)	2.38%	485.93%
Moment	4.3291(3.00)	0.8147 (0.15)	44.30%	442.67%
Point	2.9936 (3.00)	0.1328 (0.15)	0.21%	11.46%
Reliability+Moment+Point	2.9963 (3.00)	0.1378 (0.15)	0.12%	8.13%

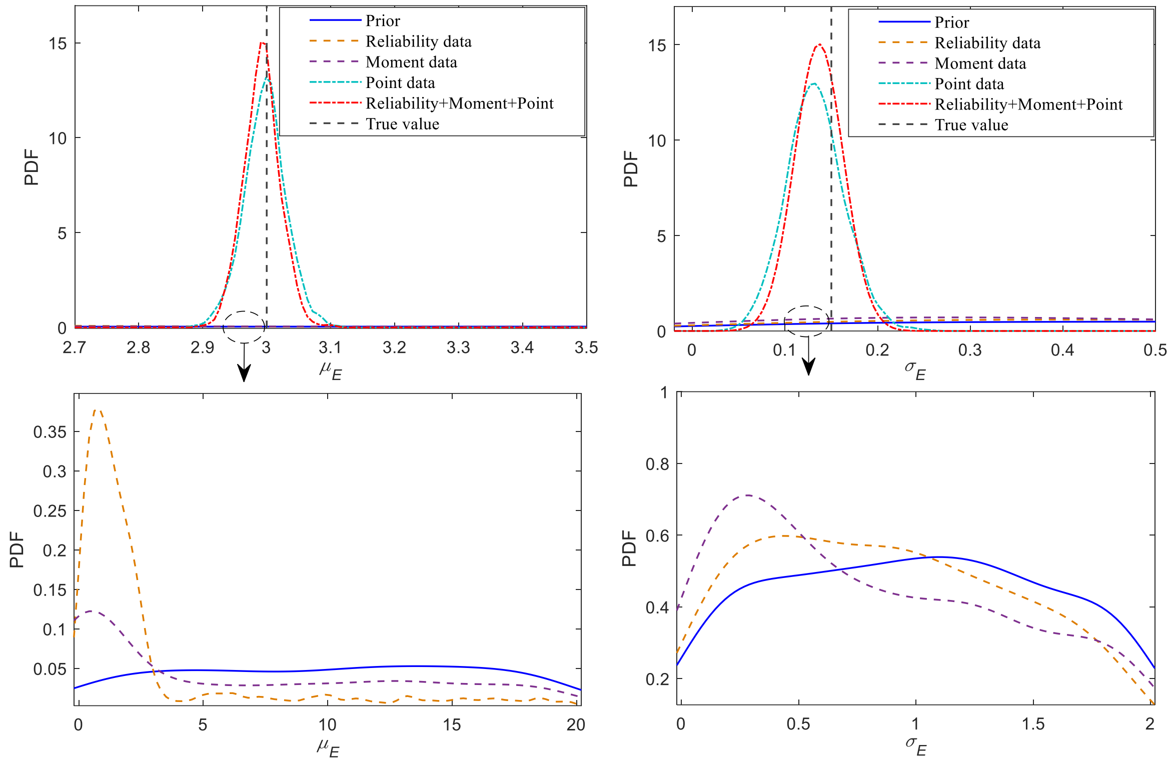
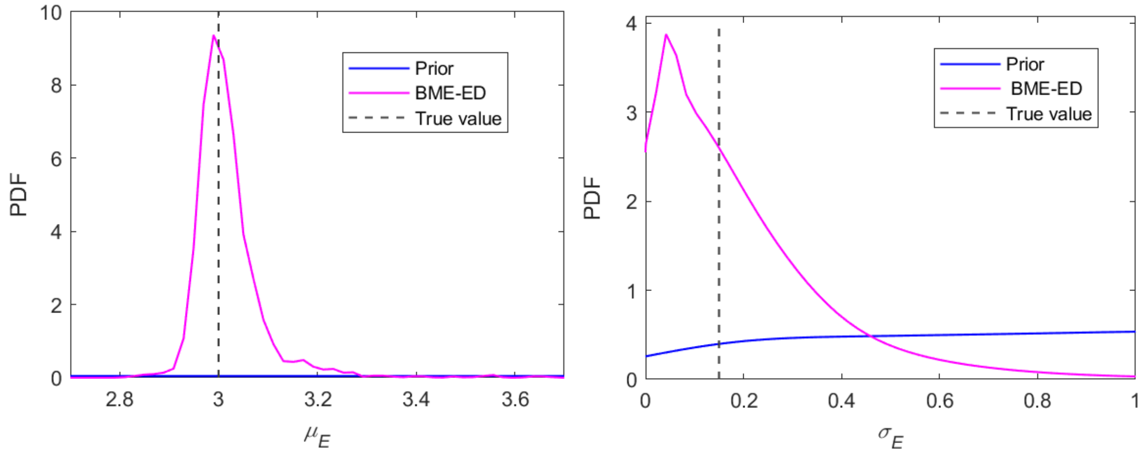


Figure 11 Posterior distributions with different data.

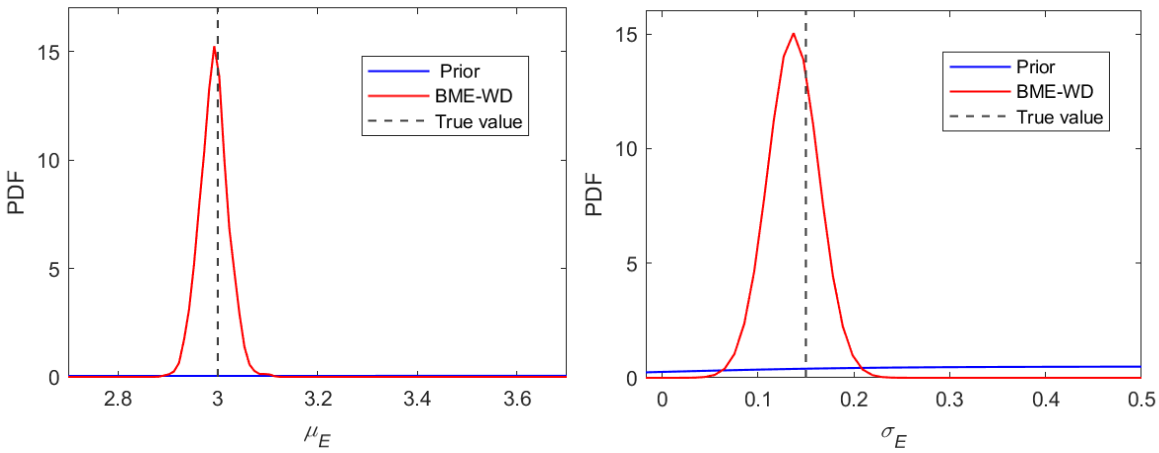
To highlight our contribution, a comparison is carried out towards with the existing BME methods with ED-based metric. The estimation results are shown in Fig. 12. It shows that both methods perform well in the calibration of μ_E (mean value), however, the traditional method (BME-ED) failed in calibrating σ_E (the standard deviation) in comparison with the proposed BME-WD method. This is because the ED-based metric only returns a distance measure between the mean of observed data set and that of simulated samples while neglecting the difference of higher order information. On the contrary, the proposed WD-based metric allows good estimation results for both μ_E and σ_E by sufficiently capturing the divergence of higher order information. This point is clearly demonstrated by the statistics shown in Table 6. To see the improvement in uncertainty reduction, the p-box of posterior distributions of deflection V are shown in Fig. 13. It is obtained by first generating enough samples for the imprecise parameters and then drawing the envelope based on these samples (each sample denotes a CDF curve).

Table 6 Updated results of hyperparameters (μ_E, σ_E) .

Cases	Posterior mean (True value)		Relative error	
	μ_E	σ_E	μ_E	μ_E
BME-ED	3.0199 (3.00)	0.2006 (0.15)	0.66%	33.73%
BME-WD	2.9963 (3.00)	0.1378 (0.15)	0.12%	8.13%



(a) Posterior distribution with BME-ED.



(b) Posterior distribution with BME-WD.

Figure 12 Comparison of the posterior distribution with two methods.

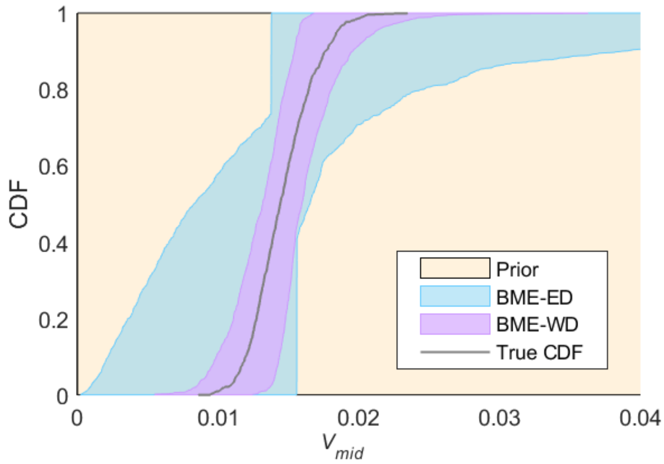


Figure 13 The p-box of posterior distribution of the deflection V_{mid} with different method.

6. Application Example: Fatigue crack growth

In this section, an engineering application example is demonstrated to show the potential of the proposed BME approach.

6.1. Problem description

The Paris law [32] is one of the most widely used models for the fatigue-induced crack growth analysis

[33-35]. It is adopted to describe the crack growth rate under constant amplitude cyclic loading:

$$\frac{da}{dN} = C(\Delta K)^m \tag{43}$$

where a is the half-crack size, N is the number of cycles, da/dN is the crack growth rate, $\Delta K = \Delta\sigma\sqrt{\pi a}$ is the stress intensity factor range which (for simplicity) does not include any correction factor due to the assumed finite size of the plate [36], $\Delta\sigma$ represents the applied stress range during one load cycle, and m and C are the unknown material parameters of interest. Given the initial half-crack size a_0 the half-crack size a_N is solved by integrating Eq. (43) as a function of N:

$$a_N = (NC(1-\frac{m}{2})(\Delta\sigma\sqrt{\pi})^m + a_0^{1-\frac{m}{2}})^{\frac{2}{2-m}} \tag{44}$$

Eq. (44) will serve the forward model in this case study. According to the accumulated knowledge, the logarithm of C and m could be modeled via a two-dimensional Gaussian distribution taking their correlation into consideration [37]. To incorporate the epistemic uncertainty, it is assumed that we only vaguely know the lower and upper bounds of the mean and covariance: $\mu_1 \in [-40,-20]$, $\mu_2 \in [0,10]$ $\sigma_1^2 \in [0,10]$, $\sigma_2^2 \in [0,1]$ and $\sigma_{12} \in [-1,1]$. For comparison purpose, the target value for parameter calibration are set based on the posterior values in [37], which are estimated (but without mixed uncertainty) based on Virkler's experiment data of 68 species for 2024-T3 aluminum alloy [38]. Detailed parameter settings can be found in Table 7.

Table 7 Parameter settings of the forward model.

Parameters	Value	Hyperparameter	Target value
$(C,\ m)$	$(\ln C, m) \sim MultivariateNormal(\mu, \Sigma)$ $\mu = [\mu_1, \mu_2], \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{21} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$	$\theta_1 = \mu_1, \theta_2 = \mu_2,$ $\theta_3 = \sigma_1^2, \theta_4 = \sigma_2^2,$ $\theta_5 = \sigma_{12} = \sigma_{21}$	$\theta_1 = -28.6301, \theta_2 = 3.2706$ $\theta_3 = 1.6049, \theta_4 = 0.048,$ $\theta_5 = -0.2775$
a_0	Deterministic, 9 mm	-	-
$\Delta\sigma$	Deterministic, 48.28 MPa	-	-

To carry out our study, 200 pairs of (C m) are generated based on the true (target) values of their mean and covariance, which lead to 200 samples of a_N for a fixed N=230000. 50 samples are used as direct measurements (i.e. point data), 100 data are used to produce the associated reliability information $R' = 0.95$ by setting a threshold a_c = 24.9mm, and the left 50 samples are used to obtain the corresponding first order moment $\bar{a} = 23.314$ mm. All this information is used to address the imprecise probabilistic model updating under the proposed BME framework.

6.2. Results and discussion

Fig. 14 shows the estimated results of 5 hyperparameters by using the proposed BME (with WD-based metric) and that of using the traditional BME method (with ED-based metric). It is seen that the proposed BME-WD method significantly outperforms the traditional BME-ED method both in accuracy and uncertainty reduction. Such benefits should be attributed to the capability of the proposed WD-based metric in capturing the required information for higher-order parameter calibration. This advantage is also demonstrated by the quantitative summarized results shown in Table 8.

For comparison, we show the prior and posterior distributions of the 5 hyperparameters using only point data and all available information (i.e. including reliability and moment information) in Fig. 15. Essentially, the estimation results of using all information slightly outperform those derived using only direct measurements. It indicates that involving all available information in the model updating improves the estimation results, suggesting the importance of statistical information as an complement to traditional direct point observations (especially when the number such measurements is limited). The statistics are summarized in Table 9.

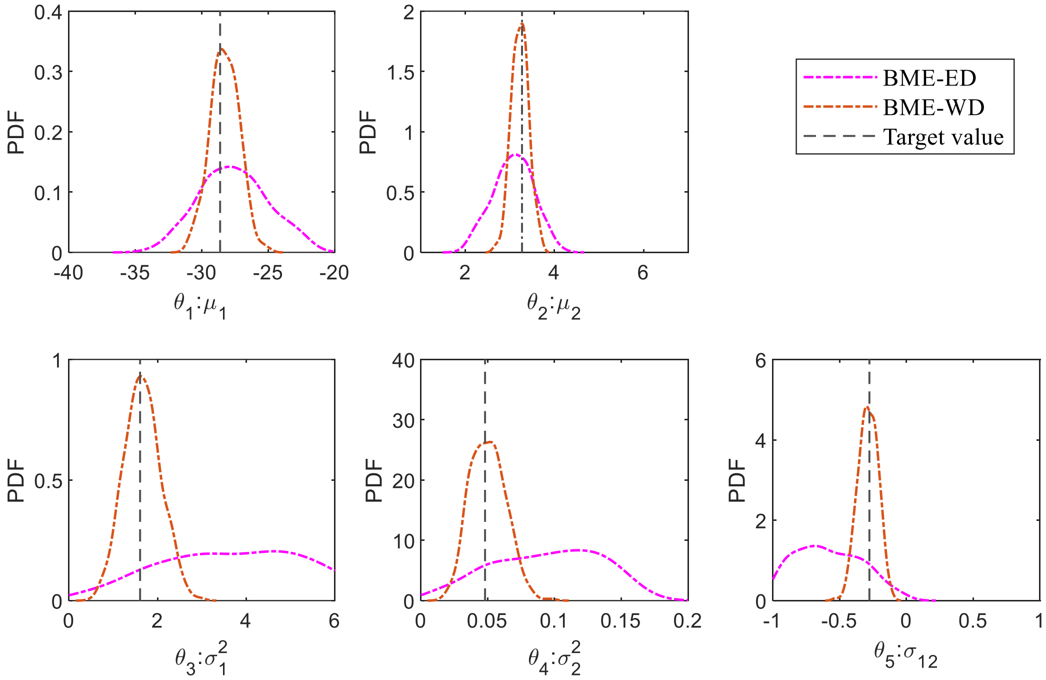


Figure 14 Posterior distribution of hyperparameters with different method.

Table 8 Updated results of hyperparameters.

Method	Updated posterior mean of hyperparameter (Relative error)				
	μ_1	μ_2	σ_1^2	σ_2^2	σ_{12} (σ_{21})
BME-ED	-27.6064 (3.58%)	3.0719 (6.07%)	3.7077 (131%)	0.0941 (96%)	-0.5806 (109%)
BME-WD	-28.3509 (0.98%)	3.2255 (1.38%)	1.6575 (3.28%)	0.0502 (4.58%)	-0.2878 (3.71%)

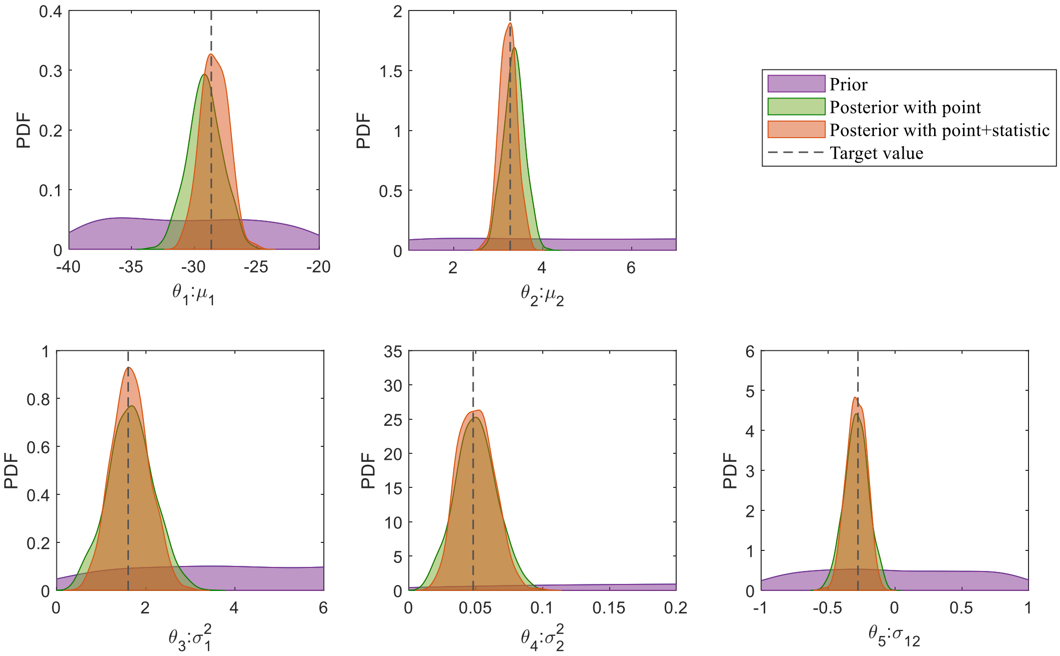


Figure 15 Probability distribution of hyperparameters under different cases.

Table 9 Updated results of hyperparameters.

Case	Updated posterior mean of hyperparameter (Relative error)				
	μ_1	μ_2	σ_1^2	σ_2^2	σ_{12} (σ_{21})
Point	-29.1067 (1.66%)	3.3594 (2.72%)	1.6673 (3.89%)	0.0500 (4.17%)	-0.2887 (4.04%)
Point +Statistic	-28.3509 (0.98%)	3.2255 (1.38%)	1.6575 (3.28%)	0.0502 (4.58%)	-0.2878 (3.71%)

The posterior distributions of (ln$C$, m) are derived based on the estimated hyperparameters and are shown in Figure 16. These sample pairs of (ln$C$, m) are then used to produce the corresponding p-box of a_N at a fixed N=230000. From Fig. 17, it is seen that the (posterior) p-box with BME-WD fully envelops the true CDF curve and shows a significant uncertainty reduction compared with the (posterior) p-box with BME-ED. To highlight our contribution, Fig. 18 shows the estimated a_N (with calibrated parameters) evolution with respect to the cycle number N for the proposed BME-WD method and that of the traditional BME-ED method. The curve derived using BME-WD matches the target one well while the curve produced by BME-ED shows a significant bias, indicating a successful crack growth prediction in the former case.

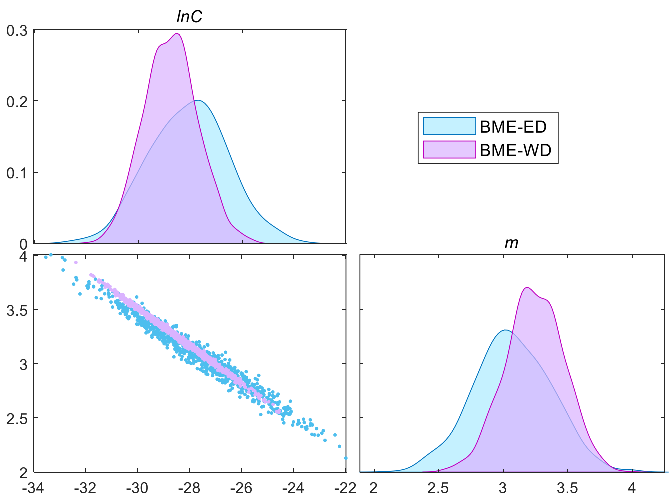


Figure 16 Posterior distribution and samples of parameter with different method.

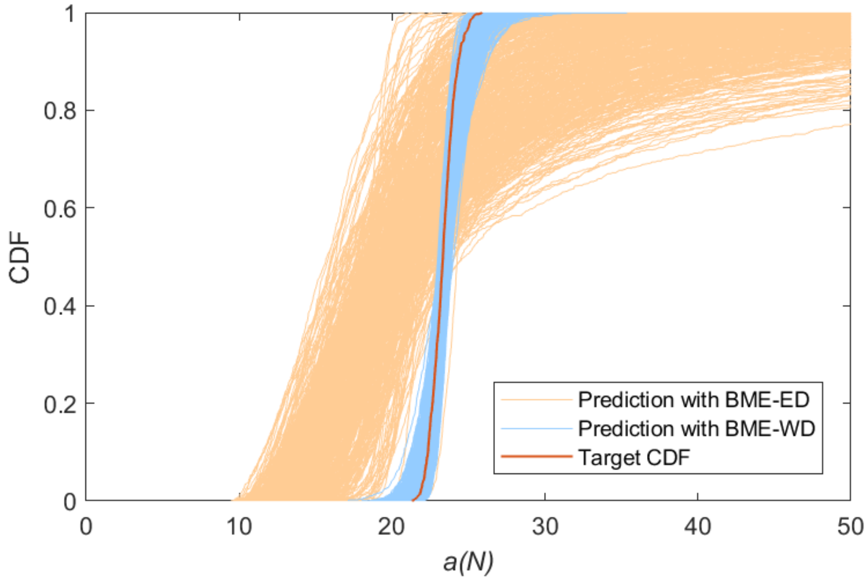


Figure 17 Comparison of predicted *p*-box and target CDF at *N*=230000cycles.

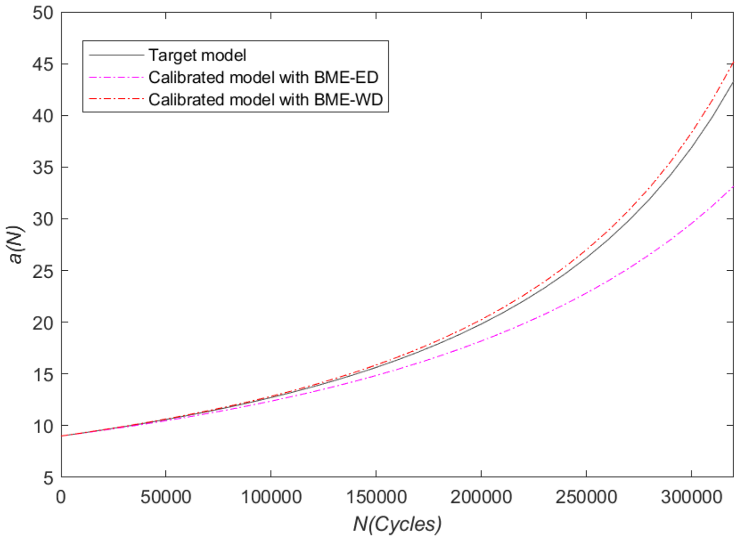


Figure 18 Comparison of calibrated model and target model.

7. Conclusions

In this paper, a BME framework is developed for imprecise model updating in presence of measurement data (i.e. point data) and statistical information (e.g. moment and reliability information). The idea behind this approach is by transforming heterogeneous information into multiple constraints, thus the optimal posterior distribution can be derived by resolving the corresponding optimization problem with entropy terms under the proposed BME framework. To facilitate the whole process, a novel WD-based metric is further developed and embedded into the framework to tackle the cumbersome construction of likelihood function resulting from information fusion. Two case studies are demonstrated to validate its effectiveness and efficiency. Based on the derived results, the following remarks are drawn:

- The proposed method has the capacity of aggregating multiple information (including point measurements, reliability information and moment information) in the BME framework and producing better calibration results with less uncertainty.
- The TMCMC algorithm is used to facilitate the approximate Bayesian computation in inverse problem

solving, which is effective in dealing with the complicated (implicit) likelihood function.

- The WD is employed to build a novel UQ metric to capture higher-order information in imprecise probabilistic models and is proved to be more efficient than some other existing metrics.

Acknowledgement

The authors gratefully acknowledge the support of the National Natural Science Foundation of China under Grant 52005032, 72271025, the Aeronautical Science Foundation of China under Grant 2018ZC74001, the Hong Kong Scholars Program under Grant No. XJ2021003, the Fundamental Research Funds for the Central Universities of China under Grant QNXM20220032, the Research Grant Council of Hong Kong under Grant 11203519, 11200621 and the Hong Kong Innovation and Technology Commission (InnoHK Project CIMDA).

Appendix

For a general case with direct measurements (point data), reliability information and moment information, its corresponding Lagrange function is given as

$$\begin{aligned} F(p,q,\alpha_1,\alpha_2,\beta_3,\beta_4) = S(p,q) + \alpha_1\left[\int_{\mathrm{Y}\times\mathrm{X}} p(y,x)dydx - 1\right] + \int_{\mathrm{Y}} \alpha_2(y)\left[\int_{\mathrm{X}} p(y,x)dx - \delta(y-y_i)\right]dy \\ +\beta_3\left[\int_{\mathrm{Y}\times\mathrm{X}} p(y,x)R(x)dydx - R^{'}\right] + \beta_4\left[\int_{\mathrm{Y}\times\mathrm{X}} p(y,x)M(x)dydx - \bar{y}\right] \end{aligned} \tag{A-1}$$

where $\alpha_1,\alpha_2,\beta_3,\beta_4$ are the Lagrange multipliers corresponding to normalization constraint, point data constraint, reliability data constraint and moment data constraint. The optimal posterior joint distribution is obtained by applying the calculus of variations by imposing $\frac{\partial F}{\partial p} = 0$:

$$p(y,x) = p_0(y,x)e^{1+\alpha_1+\alpha_2(y)+\beta_3 R(x)+\beta_4 M(x)} = p_0(y,x)e^{\alpha_2(y)+\beta_3 R(x)+\beta_4 M(x)}\frac{1}{z} \tag{A-2}$$

where $z = \int_{Y\times X} p_0(y,x)e^{\alpha_2(y)+\beta_3 R(x)+\beta_4 M(x)}dydx$ is the integration constant, and $e^{\alpha_2(y)}$ is derived from point data constraint $\int_X p(y,x)dx = \delta(y-y_i)$:

$$\frac{e^{\alpha_2(y)}}{z} = \frac{\delta(y-y_i)}{p_0(y)\int_{\mathrm{X}} p_0(x\,|\,y_i)e^{\beta_3 R(x)+\beta_4 M(x)}dx} \tag{A-3}$$

The joint posterior is expressed as:

$$\begin{aligned} p(y,x) &= p_0(y,x)e^{\beta_3 R(x)+\beta_4 M(x)}\frac{\delta(y-y_i)}{p_0(y)\int_{\mathrm{X}} p_0(x\,|\,y_i)e^{\beta_3 R(x)+\beta_4 M(x)}dx} \\ &= p_0(x\,|\,y_i)e^{\beta_3 R(x)+\beta_4 M(x)}\frac{\delta(y-y_i)}{\int_{\mathrm{X}} p_0(x\,|\,y_i)e^{\beta_3 R(x)+\beta_4 M(x)}dx} \end{aligned} \tag{A-4}$$

Integrating out y yields (Considering all the point data $\mathrm{Y}_{\mathrm{obs}}=[y_1, y_2, \ldots, y_{Nobs}]^{\mathrm{T}}$):

$$p(x) = \frac{p_0(x\,|\,\mathrm{Y}_{obs})e^{\beta_3 R(x)+\beta_4 M(x)}}{\int_{\mathrm{X}} p_0(x\,|\,\mathrm{Y}_{obs})e^{\beta_3 R(x)+\beta_4 M(x)}dx} = p_0(x)\prod_{i=1}^{N_{obs}} p_0(y_i\,|\,x)e^{\beta_3 R(x)+\beta_4 M(x)} / Z_4 \tag{A-5}$$

where Lagrange multiplier β_3 , β_4 is determined by Eq. (A-6):

$$
\begin{cases} \dfrac{\partial \ln Z_4}{\partial(\beta_3)} = R' \\ \dfrac{\partial \ln Z_4}{\partial(\beta_4)} = \overline{y} \end{cases} \tag{A-6}
$$

References

[1] B.M. Ayyub, G.J. Klir, Uncertainty modeling and analysis in engineering and the sciences, (2006).

[2] M. Grigoriu, Stochastic Systems: Uncertainty Quantification and Propagation, (2012).

[3] A. Lye, M. Kitahara, M. Broggi, E. Patelli, Robust optimization of a dynamic Black-box system under severe uncertainty: A distribution-free framework, Mechanical Systems and Signal Processing 167 (2022) 108522.

[4] O. Sedehi, C. Papadimitriou, L.S. Katafygiotis, Data-driven uncertainty quantification and propagation in structural dynamics through a hierarchical Bayesian framework, Probabilistic Engineering Mechanics 60 (2020) 103047.

[5] O. Sedehi, L.S. Katafygiotis, C. Papadimitriou, Hierarchical Bayesian operational modal analysis: Theory and computations, Mechanical Systems and Signal Processing 140 (2020) 106663.

[6] G. Mircea, Stochastic Calculus: Applications in Science and Engineering, 2002.

[7] J. Ching, J.L. Beck, K.A. Porter, Bayesian state and parameter estimation of uncertain dynamical systems, Probabilistic Engineering Mechanics 21(1) (2006) 81-96.

[8] D. Gamerman, H.F. Lopes, Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Second Edition, Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Second Edition2006.

[9] J.M. Marin, K.L. Mengersen, C.P.J.H.o.S. Robert, Bayesian Modelling and Inference on Mixtures of Distributions, 25 (2005).

[10] D. Sivia, J.J.C.P. Skilling, Oxford University Press, Data analysis : a Bayesian tutorial, (2006).

[11] J.B. Tenenbaum, T.L. Griffiths, C.J.T.i.C.S. Kemp, Theory-Based Bayesian Models of Inductive Learning and Reasoning, 10(7) (2006) 309-318.

[12] O. Sedehi, C. Papadimitriou, D. Teymouri, L.S. Katafygiotis, Sequential Bayesian estimation of state and input in dynamical systems using output-only measurements, Mechanical Systems and Signal Processing 131 (2019) 659-688.

[13] E. VanDerHorn, S. Mahadevan, Bayesian model updating with summarized statistical and reliability data, Reliability Engineering & System Safety 172 (2018) 12-24.

[14] S. Sankararaman, S.J.R.E. Mahadevan, S. Safety, Likelihood-based representation of epistemic uncertainty due to sparse point data and/or interval data, 96(7) (2011) 814-824.

[15] D.O. Defence, Military Handbook Reliability Prediction of Electronic Equipment (MIL-HDBK-217E), (1991).

[16] E. T. Jaynes, Information Theory and Statistical Mechanics. Physical Review. Series II. 106 (4) (1957): 620–630.

[17] E. T. Jaynes, Information Theory and Statistical Mechanics II. Physical Review. Series II. 108 (2) (1957): 171–190.

[18] A. Giffin, A.J.A.C.P. Caticha, Updating Probabilities with Data and Moments, (2007).

[19] A.J.P. Giffin, Maximum Entropy: The Universal Method for Inference, (2009).

[20] A.J.P.A.S.M. Giffin, I. Applications, From Physics to Economics: An Econometric Example Using Maximum Relative Entropy, 388(8) (2009) 1610-1620.

[21] D. Zhou, J. He, Y.-M. Du, C.P. Sun, X. Guan, Probabilistic information fusion with point, moment and interval data in reliability assessment, Reliability Engineering & System Safety 213 (2021) 107790.

[22] J.K. Pritchard, M.T. Seielstad, A. Perez-lezaun, Population growth of human Y chromosomes: a study of Y chromosome microsatellites %J Molecular biology and evolution %J, 16(12) (1999) 1791-1798.

[23] E. Bernton, P.E. Jacob, M. Gerber, C.P.J.a.e.-p. Robert, Approximate Bayesian computation with the Wasserstein distance, (2019).

[24] S. Bi, M. Broggi, M. Beer, The role of the Bhattacharyya distance in stochastic model updating, Mechanical Systems and Signal Processing 117 (2019) 437-452.
[25] L. Yang, S. Bi, M.G.R. Faes, M. Broggi, M. Beer, Bayesian inversion for imprecise probabilistic models using a novel entropy-based uncertainty quantification metric, Mechanical Systems and Signal Processing 162 (2022) 107954.
[26] Robin, A. W., Information theory and statistical mechanics, (1981).
[27] B.D. Youn, K.K.J.J.o.M.D. Choi, An Investigation of Nonlinearity of Reliability-Based Design Optimization Approaches, 126(3) (2004) 881-890.
[28] H.H. Khodaparast, J.E. Mottershead, Efficient methods in stochastic model updating, (2008).
[29] L. Sun, L. Zhu, W. Li, C. Zhang, T. Balezentis, Interval-valued functional clustering based on the Wasserstein distance with application to stock data, Information Sciences 606 (2022) 910-926.
[30] S. Marelli, B. Sudret, UQLab: A Framework for Uncertainty Quantification in Matlab, Int Conf on Vulnerability, Risk Analysis & Management, 2014, pp. 2554-2563.
[31] J. Ching, Y.C.J.J.o.E.M. Chen, Transitional Markov Chain Monte Carlo Method for Bayesian Model Updating, Model Class Selection, and Model Averaging, 133(7) (2007) 816-832.
[32] P.C. Paris, F. Erdogan, A critical analysis of crack propagation laws, (1960).
[33] A. Coppe, R.T. Haftka, N.H. Kim, F.G.J.J.o.A. Yuan, Uncertainty Reduction of Damage Growth Properties Using Structural Health Monitoring, 47(6) (2012) 2030-2038.
[34] Z.S.J.S.S. Mahadevan, Model uncertainty and Bayesian updating in reliability-based inspection, (2000).
[35] Z. Zhao, A. Haldar, RELIABILITY-BASED STRUCTURAL FATIGUE DAMAGE EVALUATION AND MAINTENANCE USING NON-DESTRUCTIVE INSPECTIONS, Uncertainty Modeling in Finite Element, Fatigue and Stability of Systems1997.
[36] J.M. Karandikar, N.H. Kim, T.L.J.E.F.M. Schmitz, Prediction of remaining useful life for fatigue-damaged structures using Bayesian inference, 96 (2012) 588-605.
[37] X. Guan, A. Giffin, R. Jha, Y.J.P.E.M. Liu, Maximum relative entropy-based probabilistic inference in fatigue crack damage prognostics, 29(none) (2012) 157-166.
[38] D.A. Virkler, B.M. Hillberry, P.K.J.J.o.E.M. Goel, Technology, The Statistical Nature of Fatigue Crack Propagation, 101(2) (1979) 241.