Mechanical Systems and Signal Processing 98 (2018) 465–483

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

On the quantification and efficient propagation of imprecise probabilities resulting from small datasets

Jiaxin Zhang, Michael D. Shields *

Department of Civil Engineering, Johns Hopkins University, Baltimore, MD 21218, USA

ARTICLE INFO

Article history:
Received 20 September 2016
Received in revised form 25 April 2017
Accepted 27 April 2017
Available online 19 May 2017

Keywords:
Uncertainty quantification
Uncertainty propagation
Epistemic uncertainty
Imprecise probability
Bayesian inference
Importance sampling
Akaike information criterion
Kullback-Leibler information theory

ABSTRACT

This paper addresses the problem of uncertainty quantification and propagation when data for characterizing probability distributions are scarce. We propose a methodology wherein the full uncertainty associated with probability model form and parameter estimation are retained and efficiently propagated. This is achieved by applying the information-theoretic multimodel inference method to identify plausible candidate probability densities and associated probabilities that each method is the best model in the Kullback-Leibler sense. The joint parameter densities for each plausible model are then estimated using Bayes' rule. We then propagate this full set of probability models by estimating an optimal importance sampling density that is representative of all plausible models, propagating this density, and reweighting the samples according to each of the candidate probability models. This is in contrast with conventional methods that try to identify a single probability model that encapsulates the full uncertainty caused by lack of data and consequently underestimate uncertainty. The result is a complete probabilistic description of both aleatory and epistemic uncertainty achieved with several orders of magnitude reduction in computational cost. It is shown how the model can be updated to adaptively accommodate added data and added candidate probability models. The method is applied for uncertainty analysis of plate buckling strength where it is demonstrated how dataset size affects the confidence (or lack thereof) we can place in statistical estimates of response when data are lacking.

1. Introduction

Uncertainty quantification (UQ) and propagation play a critical role in computationally evaluating the performance of engineering systems. Generally speaking, uncertainty can be categorized as either *aleatory*, resulting from inherent randomness, or *epistemic*, resulting from a lack of complete knowledge [1]. It has been argued that epistemic uncertainties require a different mathematical treatment than aleatory uncertainties that are naturally stochastic [2]. It remains an open, and sometimes contentious, debate as to what that mathematical treatment should be.

Arguments have been made for a variety of probabilistic and non-probabilistic treatments of epistemic uncertainties. Non-probabilistic uncertainty theories include Dempster-Schafer evidence theory [3,4], fuzzy sets [5], interval methods [6,7], convex models [8], possibility theory [9], and information theory [10] among others. Probabilistic approaches meanwhile include methods based on random sets [11], probability boxes (p-boxes) [12,13], Bayesian [14], and frequentist [15] theories. Moreover, several variations of and connections between both probabilistic and non-probabilistic theories have

* Corresponding author.
E-mail addresses: jiaxin.zhang@jhu.edu (J. Zhang), michael.shields@jhu.edu (M.D. Shields).

http://dx.doi.org/10.1016/j.ymssp.2017.04.042
0888-3270/

been made [16–18] in the interest of constructing a unified theory of imprecise probabilities [19,20]. In other words, to think of these theories as necessarily distinct and mutually exclusive is incorrect. Common formalisms can be applied across many of these approaches such that, for example, fuzzy sets can be interpreted as generalizations of both random sets and intervals and can also be interpreted to loosen the conditions on probability measures. For the interested reader, an extensive review of many of these approaches for engineering applications can be found in [21].

While the intricacies, interconnections, and nuances of these theories are critical to a comprehensive understanding of uncertainty, this work is motivated by epistemic uncertainty that specifically stems from having only small datasets available to characterize otherwise aleatory uncertainties. This is a problem that has been studied by several researchers (e.g. in the context of reliability analysis [22–24]), and one for which we propose a probabilistic approach that addresses its primary challenges: namely probability model selection, model parameter uncertainties, and uncertainty propagation. For clarity, herein the term *model* is used synonymously with a probability model, which more specifically assigns a probability measure to the probability space of the underlying aleatory uncertainty. Practically speaking, this is uncertainty in the form and parameter values of a probability distribution.

2. Uncertainty caused by lack of data

Statistical theories, both frequentist [15] and Bayesian [14], are well-suited for problems with large sample size. However, large datasets are rarely available for many engineering applications. Small sample size creates a specific form of epistemic uncertainty (sometimes referred to as second-order uncertainty [25]) that manifests in the inability to identify a unique model for the underlying probabilities. This section briefly discusses the challenges created by this type of uncertainty.

2.1. Model-form uncertainty

The desire to assign a specific model form (i.e. probability distribution type) is understandable from many perspectives. Certainly, it simplifies the uncertainty analysis. Going further, the inability to assign a model-form instills a sense of uneasiness in an analysts' mind because, among other reasons, tools for propagating multiple probability models are limited and computationally very expensive [25]. But one must always ask whether the assignment of a probability model from a small data set is justified. Undoubtedly, cases exist where assignment is justified such as when the data in the small set arise from, say, the sum of many other random variables such that the Central Limit Theorem applies. More commonly, distributions are assigned arbitrarily based on either analyst judgment or through some comparative down-selection or averaging process [26].

Through a comparative down-selection process, one begins with a number of candidate models and through statistical analysis selects the model that is "best" according to some criterion. The most commonly employed methods include statistical goodness-of-fit tests (e.g. [27]), Bayesian hypothesis testing and model selection [28,14], and information-theoretic model selection [29]. For small datasets, statistical goodness-of-fit tests are known to be inappropriate [30]. Bayesian and information-theoretic approaches, meanwhile, remain widely used when data is scarce (justifiably or not). The two methods differ in the nature of their selection criteria. Bayesian model selection methods use ratios of the integrated likelihood functions for different models to compute the Bayes factor. The information-theoretic process, on the other hand, bases the selection process on the Kullback-Leibler mean information loss. These methods are further elaborated in Section 3.

Another popular process is to average the various models that are being considered by assigning relative weights to them. This process, elaborated nicely for the lack of data uncertainties in [31], is perhaps a more comprehensive means of model selection in that it aims to include each of the possible candidate models through the averaging process.

Our primary disagreement with both the down-selection and averaging methods is not a methodological one, but a philosophical one. Given lack of data, we believe and attempt to show that the goal of identifying a single probability model capable of including both stochastic and small dataset uncertainty is an unrealistic one. Selecting a single model (averaged or not) disregards a portion of the uncertainty, and for this reason we prefer a multimodel inference approach, such as that presented by Burnham and Anderson [32] and outlined in Section 3.

2.2. Parameter uncertainty

For a given model, estimation of the model parameters introduces additional uncertainties. In the frequentist approach, the parameters are estimated as deterministic values, perhaps with confidence bounds established through methods such as bootstrapping [33]. Again, these methods can be problematic for small datasets. In the Bayesian approach, the parameters are treated as random variables with their joint distribution estimated through Bayes' rule as discussed in Section 3.1. For small datasets, these joint distributions can be heavily biased by the prior model and even the use of so-called noninformative priors can have important implications that will not be explored in more detail here.

In this work, we adopt the Bayesian approach with a noninformative prior (accepting that this perhaps has unintended consequences). The primary difference between the approach presented here and most Bayesian approaches is that, rather than estimating the joint parameter density and promptly discarding it by selecting a single set of maximum likelihood parameters for the model or integrating out its variability [34], we retain the full joint parameter densities for each model and propagate them through the function along with their parent models from the multimodel inference.

2.3. Uncertainty propagation

The primary argument against a multimodel approach that retains full parametric uncertainty is clearly computational expense [25]. As pointed out in [31] and illustrated in Fig. 1(a), the Monte Carlo propagation of this type of uncertainty involves multiple loops that come at very large computational cost. Our ability to retain and propagate the full model-form and parametric uncertainties is predicated on a new methodology, shown in Fig. 1(b) and presented in Sections 4 and 5, to collapse these multiple loops to a single Monte Carlo loop on a surrogate distribution obtained through optimization. By propagating this surrogate distribution and reweighting the samples according to importance sampling, it becomes possible to simultaneously propagate uncertainty associated with a full set of model families each having random variable parameters.

3. Bayesian inference and probability model selection from limited data

In this section, we review the principles of Bayesian inference and multimodel selection to assess the viability of different probability models to represent a dataset. Probability model selection is addressed through an information-theoretic approach employing the Akaike Information Criterion to assess the viability of each model and estimate the probability that the given model is the "best" model. For a given probability model, Bayes' rule is used to quantify uncertainties associated with the model parameters.

3.1. Bayesian inference

Consider the random variable $\boldsymbol{X}$ with probability model M and uncertain parameters $\boldsymbol{\theta}$. Bayes' rule provides a means of updating our knowledge of the parameters $\boldsymbol{\theta}$ for M using collected data. Bayes' rule treats $\boldsymbol{\theta}$ probabilistically by assigning a prior probability density function (pdf), denoted $p(\boldsymbol{\theta};M)$, that reflects our current belief or knowledge. The pdf of the parameters is updated, given data $\boldsymbol{d}$, to define a posterior distribution $p^*(\boldsymbol{\theta}|\boldsymbol{d},M)$ reflecting our updated knowledge by:

$$p^*(\boldsymbol{\theta}|\boldsymbol{d},M) = \frac{p(\boldsymbol{d}|\boldsymbol{\theta},M)p(\boldsymbol{\theta};M)}{p(\boldsymbol{d};M)} \propto \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{d},M)p(\boldsymbol{\theta};M) \tag{1}$$

where $\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{d},M) = p(\boldsymbol{d}|\boldsymbol{\theta},M)$ is the likelihood function and the normalizing factor $p(\boldsymbol{d};M)$, referred to as the evidence, is computed by marginalizing $\mathcal{L}(\cdot)$ over the parameters $\boldsymbol{\theta}$

$$p(\boldsymbol{d};M) = \int \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{d},M)p(\boldsymbol{\theta};M)d\boldsymbol{\theta} \tag{2}$$

This is a non-trivial task since the integral in Eq. (2) is often analytically intractable. When the prior and posterior distributions belong to the same family, they are referred to as conjugate distributions. For certain distribution families, closed-form relations exist to relate the conjugate distributions that make the evaluation of Bayes' rule mathematically convenient. More

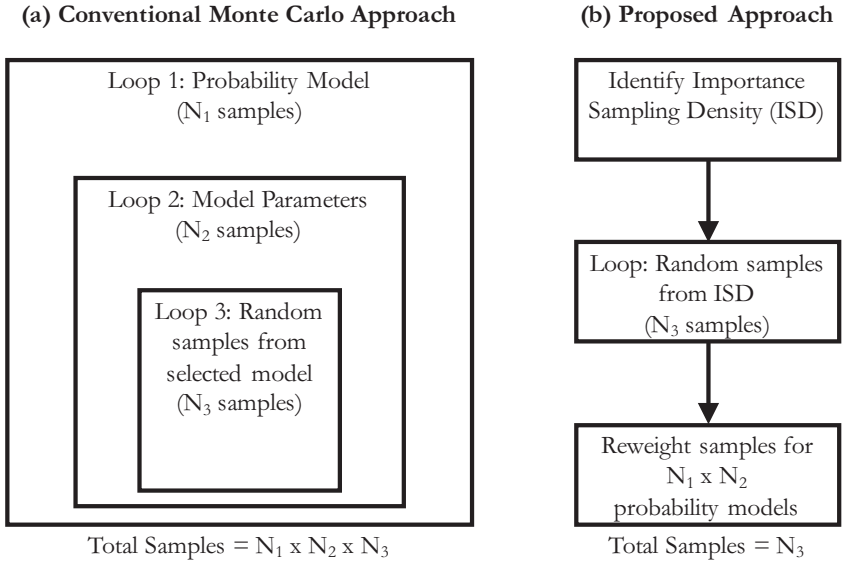

Fig. 1. Conceptual comparison of (a) the conventional Monte Carlo approach for propagating model-form and parametric uncertainty, and (b) the proposed approach.

generally, it is necessary to employ the Markov Chain Monte Carlo (MCMC) method to draw samples from $p^*(\boldsymbol{\theta}|\boldsymbol{d}, M)$. In the present study, we use the affine invariant ensemble sampler for MCMC proposed by Goodman and Weare [35] and packaged as emcee [36].

In the limited data case, Bayesian inference will improve the parameter estimates over prior knowledge but to a limited extent. As a result, the posterior parameter distribution will likely possess large variance which does not instill confidence in the selection of a single parameter set (such as the maximum likelihood estimate). Instead, we propose in this work that the entire joint parameter density should be retained and propagated as discussed in Sections 4 and 5.

3.2. Multimodel selection and model uncertainty

Classically, it is often assumed that there are a set of models and data, statistical inference is then applied to select a single "best" model, and that model is the sole model used for making inference from data. Since a single best model has been found, any uncertainty associated with model selection is ignored. As is well-known and confirmed herein, such selection methods require very large datasets. In the limited data case, it is difficult (if not impossible) to identify a unique best model. It therefore becomes necessary to quantify the model uncertainty and compare the validity of multiple candidate models – a process referred to as multimodel inference [32]. Note that we avoid the use of the term "true" model as we adhere to the information-theoretic belief that models are, by definition, approximations of reality and therefore true models do not exist. The best model, by contrast, is the one that minimizes the difference between the observed reality and the model in some sense. To quantify and incorporate model selection uncertainty via multimodel inference, we employ a popular approach proposed by Burnham and Anderson [32] and summarized below.

Model selection requires establishing a well-justified criterion for determining the best model. Two well-known approaches utilize Bayesian model selection criteria based on Bayes factors/likelihood functions and information-theoretic selection criteria based on the Kullback-Leibler (K-L) information theory [37].

In the Bayesian setting, selection between two models is often conducted by estimating the Bayes factor, B_F, as the ratio of the likelihoods of the data for the two models, M_0 and M_1, integrated over the parameter space as:

$$\frac{p(M_0|\boldsymbol{d})}{p(M_1|\boldsymbol{d})} = \frac{p(M_0)p(\boldsymbol{d}|M_0)}{p(M_1)p(\boldsymbol{d}|M_1)} = B_F \cdot \frac{p(M_0)}{p(M_1)} \tag{3}$$

where

$$p(\boldsymbol{d}|M) = \int p(\boldsymbol{\theta}|M)p(\boldsymbol{d}|\boldsymbol{\theta}, M)d\boldsymbol{\theta} \tag{4}$$

This can be generally interpreted such that $B_F > 1$ implies that the data provides greater evidence in support of model M_0 and vice versa for $B_F < 1$. It is well known that Bayes factors are sensitive to prior probabilities (especially in the limited data case) [32] and theoretically infinite Bayes factors exist for a given dataset – only some of which are reasonable. A closely related metric for Bayesian model selection, the Bayesian Information Criterion (BIC, a misnomer since no connection to information theory has been established) is derived from the integrated likelihood function. Specifically, the BIC is derived by estimating $-2\log(p(\boldsymbol{d}|M))$ through a Taylor series expansion around the maximum likelihood estimate of the parameters, $\hat{\boldsymbol{\theta}}$, with some terms neglected as [14]:

$$\mathrm{BIC} = -2\log(\mathcal{L}(\hat{\boldsymbol{\theta}}|\boldsymbol{d}, M)) + K\log n \tag{5}$$

where $\mathcal{L}(\hat{\boldsymbol{\theta}}|\boldsymbol{d}, M) = p(\boldsymbol{d}|\boldsymbol{\theta}, M)$ is the likelihood function given the maximum likelihood estimate of the parameters $\hat{\boldsymbol{\theta}}$, K is the dimension of the parameter vector $\boldsymbol{\theta}$, and n is the sample size of the dataset. The BIC also provides a convenient means of estimating a Bayes factor for model comparison with some implicitly defined prior distribution that may or may not be reasonable [38].

It is important in model selection to establish a relative scale for the BIC as follows:

$$\Delta_B^{(i)} = \mathrm{BIC}^{(i)} - \mathrm{BIC}^{\min} \tag{6}$$

where $\mathrm{BIC}^{(i)}$ is the BIC for candidate model M_i and $\mathrm{BIC}^{\min} = \min_i(\mathrm{BIC}^{(i)})$. This normalizes the best model to a value $\Delta_B^{(i)} = 0$. By doing so, it is possible to establish posterior model probabilities, $\mathfrak{p}_i$, as [14]:

$$\mathfrak{p}_i = p(M_i|\boldsymbol{d}) = \frac{\exp\left(-\frac{1}{2}\Delta_B^{(i)}\right)}{\sum_{i=1}^{N}\exp\left(-\frac{1}{2}\Delta_B^{(i)}\right)} \tag{7}$$

under the assumption that the prior models, M_i, have equal probability $\frac{1}{N}$. This, as noted by Burnham and Anderson [32], does not imply that a true model is among the candidate models M_i or that such a model even exists – even if $\mathfrak{p}_i = 1$. Instead, these probabilities can be interpreted as the probability that M_i is the model that the BIC would select with $\mathfrak{p}_i = 1$ given $n \to \infty$ (referred to as the BIC target model).

In the information-theoretic setting, model selection is based on establishing a criterion for the information loss resulting from approximating truth with a model. An appropriate model selection criterion is therefore one that minimizes the information loss. Akaike [29] showed that the expected relative K-L information could be approximated by the maximized log-likelihood function with a bias correction. Using this relation, Akaike established the Akaike Information Criterion:

$$\mathrm{AIC} = -2\log(\mathcal{L}(\hat{\boldsymbol{\theta}}|\boldsymbol{d}, M)) + 2K \tag{8}$$

where $2K$ is a bias correction factor. This criterion is referred to as information-theoretic because it provides a simple estimate of the K-L information. However, it can also be conceived in a Bayesian context through the use of a class of *savvy priors* (see [32] for explanation).

As with the BIC, it is necessary to develop a relative scale for AIC values such that:

$$\Delta_A^{(i)} = \mathrm{AIC}^{(i)} - \mathrm{AIC}^{\min} \tag{9}$$

forces the best model to have $\Delta_A^{(i)} = 0$ and all other models to have positive $\Delta_A^{(i)}$ interpreted as the information lost relative to the best model. Moreover, the simple transformation $\exp\left(-\frac{\Delta_A^{(i)}}{2}\right)$ provides the likelihood for the model M_i given the data. Normalizing these likelihoods, we can estimate model probabilities analogous to those in Eq. (7) as:

$$\mathbb{p}_i = p(M_i|d) = \frac{\exp\left(-\frac{1}{2}\Delta_A^{(i)}\right)}{\sum_{i=1}^{N}\exp\left(-\frac{1}{2}\Delta_A^{(i)}\right)} \tag{10}$$

These probabilities can be interpreted as the probability that model M_i is, in fact, the K-L best model for the data.

Finally, both the AIC and BIC are asymptotic quantities that imply large datasets – each with their own distinct advantages and disadvantages [32,38]. For small datasets, a critical extension of the AIC has been developed [39,40] that introduces a second-order bias correction term yielding:

$$\mathrm{AIC}_c = -2\log(\mathcal{L}(\hat{\boldsymbol{\theta}}|\boldsymbol{d}, M)) + 2K + \frac{2K(K+1)}{n-K-1} \tag{11}$$

AIC_c should be used when $\frac{n}{K} <\sim 40$ and, because $\mathrm{AIC}_c \to \mathrm{AIC}$ as $n \to \infty$, it always makes sense to use the second-order version.

In this work, we utilize the AIC_c for multimodel selection given its information-theoretic and Bayesian interpretation and allowance for small sample size. Models are ranked according to $\Delta_A^{(i)}$ and model probabilities assigned according to Eq. (10).

4. Importance sampling and the selection of an optimal sampling density

Given the parametric and model-form uncertainties created by lack of data discussed in the previous section, we now turn our attention to the issue of uncertainty propagation. Methods for propagation of uncertainty characterized by exactly known probability models are ubiquitous. The authors are presently unaware of any methods capable of simultaneously propagating many possible/plausible probability models without explicitly propagating each distribution individually at great computational expense [25]. To achieve this simultaneous propagation of the many plausible models identified through multimodel inference with Bayesian parameter estimation, we employ the principles of importance sampling. In this section, we review the importance sampling concept and derive a means of identifying a single optimal sampling density for a set of plausible probability models.

4.1. Importance sampling

Consider the generic performance function $g(\boldsymbol{X})$ defining the response quantity of interest for a given system. The problem of uncertainty propagation is generally concerned with evaluating the expected value $E[g(\boldsymbol{X})]$ where $\boldsymbol{X} \in \Omega$ is a random vector having probability model $M_{\boldsymbol{X}}$ with density function $p(\boldsymbol{x})$. The classical Monte Carlo estimator is given by:

$$E_p[g(\boldsymbol{X})] = \int_\Omega g(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} \approx \frac{1}{N}\sum_{i=1}^{N} g(\boldsymbol{x}_i) \tag{12}$$

where $E_p[\cdot]$ is the expectation under $p(\boldsymbol{x})$, N is the number of samples, and $\boldsymbol{x}_i$ are independent random samples drawn from $p(\boldsymbol{x})$. In some cases, it may be difficult to draw samples directly from $p(\boldsymbol{x})$. Importance sampling allows samples to be drawn from an alternate density $q(\boldsymbol{x})$ that is easier to sample. The Monte Carlo estimator in Eq. (12) is then modified to correct the bias generated by sampling from an alternate distribution as:

$$E_q\left[g(\boldsymbol{X})\frac{p(\boldsymbol{X})}{q(\boldsymbol{X})}\right] = \int_\Omega g(\boldsymbol{x})\frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}q(\boldsymbol{x})d\boldsymbol{x} \approx \frac{1}{N}\sum_{i=1}^{N} g(\boldsymbol{x}_i)\frac{p(\boldsymbol{x}_i)}{q(\boldsymbol{x}_i)} \tag{13}$$

where $E_q[\cdot]$ is the expectation with respect to $q(\boldsymbol{x})$. We define the ratios $w(\boldsymbol{x}_i) = p(\boldsymbol{x}_i)/q(\boldsymbol{x}_i)$ as the importance weights, which will play an important role in the proposed approach.

4.2. Optimal sampling density for a single target density

A great deal of research has been conducted toward identifying an efficient proposal sampling density $q(\boldsymbol{x})$ when the target density $p(\boldsymbol{x})$ is known. Much of this attention has focused on achieving a variance reduction – in which case the true optimal sampling density is given by $q(\boldsymbol{x}) = \frac{g(\boldsymbol{x})p(\boldsymbol{x})}{E_p[g(\boldsymbol{X})]}$ which yields a zero variance estimator but is clearly infeasible.

For our purposes we are less interested in achieving a variance reduction (although it would be nice) and are instead interested in ensuring that our sampling density is as close as possible to the target density $p(\boldsymbol{x})$, given the difficulty of sampling from $p(\boldsymbol{x})$ itself. This is accomplished by minimizing the f-divergence [41–43] defining the difference between two distributions P and Q over a space Ω with measure μ as:

$$D_f(P\|Q) = \int_\Omega f\left(\frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}\right) q(\boldsymbol{x}) d\mu(\boldsymbol{x}) \tag{14}$$

Several different functions $f(\cdot)$ have been proposed yielding different common divergences such as Kullback-Leibler divergence [37], the Hellinger distance [44], and the total variation distance [45]. Here, we employ the Hellinger distance given by:

$$H^2(P\|Q) = \frac{1}{2}\int_\Omega \left(\sqrt{p(\boldsymbol{x})} - \sqrt{q(\boldsymbol{x})}\right)^2 d\boldsymbol{x} \tag{15}$$

Specifically, for a given family of distributions, we aim to find the density $q^*(\boldsymbol{x})$ that yields the square Minimum Hellinger Distance Estimator (MHDE) [46,47] relative to the target density $p(\boldsymbol{x})$ given by:

$$q^*(\boldsymbol{x}) = \arg\min \frac{1}{2}\int_\Omega \left(\sqrt{p(\boldsymbol{x})} - \sqrt{q(\boldsymbol{x})}\right)^2 d\boldsymbol{x} \tag{16}$$

The optimization is then conducted by identifying a nonparametric form for $q(\boldsymbol{x})$ that minimizes the objective function in Eq. (16).

4.3. Optimal sampling density for multiple distributions

In the case of imprecise probabilities of interest here, the target density $p(\boldsymbol{x})$ is not known precisely. Instead, multiple probability models, M_i, are plausible, each with uncertain (probabilistic) parameters $\boldsymbol{\theta} \subseteq \Theta$ where $\Theta \in \mathbb{R}^d$ is the set of all parameters for all models, quantified through Bayesian inference and density functions $p_i(\boldsymbol{x}|\boldsymbol{\theta})$ having model probabilities $\mathbb{p}_i$ identified through multimodel inference (Eq. (10)). We now aim to identify a *single* sampling density $q^*(\boldsymbol{x})$ that is representative of all $p_i(\boldsymbol{x}|\boldsymbol{\theta})$ and can therefore be used with importance sampling to simultaneously propagate the full set of plausible candidate probability models.

Consider that we have a finite set $\mathbb{M} = \{M_i\}, i = 1, 2, \ldots, N_d$ of candidate target probability models having densities $p_i(\boldsymbol{x}|\boldsymbol{\theta})$. The total Hellinger distance can be defined as:

$$\widehat{H}^2(\mathbb{M}\|Q) = \sum_{i=1}^{N_d} H^2(M_i\|Q) = \sum_{i=1}^{N_d} \frac{1}{2}\int_\Omega \left(\sqrt{p_i(\boldsymbol{x}|\boldsymbol{\theta})} - \sqrt{q(\boldsymbol{x})}\right)^2 d\boldsymbol{x} \tag{17}$$

The total Hellinger distance is therefore a random variable indexed on $\boldsymbol{\theta}$ with expected value given by:

$$\begin{aligned} E\left[\widehat{H}^2(\mathbb{M}\|Q)\right] &= \sum_{i=1}^{N_d} E_\theta\left[H^2(M_i\|Q)\right] = \sum_{i=1}^{N_d} \frac{1}{2} E_\theta\left[\int_\Omega \left(\sqrt{p_i(\boldsymbol{x}|\boldsymbol{\theta})} - \sqrt{q(\boldsymbol{x})}\right)^2 d\boldsymbol{x}\right] \\ &= E_\theta\left[\int_\Omega \sum_{i=1}^{N_d} \frac{1}{2}\left(\sqrt{p_i(\boldsymbol{x}|\boldsymbol{\theta})} - \sqrt{q(\boldsymbol{x})}\right)^2 d\boldsymbol{x}\right] \end{aligned} \tag{18}$$

referred to as the total expected squared Hellinger distance.

In order to identify an optimal sampling density $q(\boldsymbol{x})$, we need to minimize the total expected squared Hellinger distance expressed as a functional $\mathcal{T}[q]$ given the isoperimetric constraint $\mathcal{I}[q]$ as

$$\begin{aligned} &\underset{q}{\text{minimize}}\ \mathcal{T}(q) = E_\theta\left[\int_\Omega F(\boldsymbol{x}, \boldsymbol{\theta}, q(\boldsymbol{x}))d\boldsymbol{x}\right] \\ &\text{subject to}\ \mathcal{I}(q) = \int_\Omega q(\boldsymbol{x})d\boldsymbol{x} - 1 = 0 \end{aligned} \tag{19}$$

where

$$F(\boldsymbol{x}, \boldsymbol{\theta}, q(\boldsymbol{x})) = \frac{1}{2}\sum_{i=1}^{N_d}\left(\sqrt{p_i(\boldsymbol{x}|\boldsymbol{\theta})} - \sqrt{q(\boldsymbol{x})}\right)^2 \tag{20}$$

and the isoperimetric constraint ensures that $q(\boldsymbol{x})$ is a valid pdf. Formulating the optimization in this way allows us to use the method of Lagrange multipliers in conjunction with the calculus of variations. Let us define the action functional (Lagrangian) as:

$$\mathcal{L}[q,\lambda] = \mathcal{T}[q] + \lambda\mathcal{I}[q] \tag{21}$$

The function $q(\boldsymbol{x})$ that minimizes $\mathcal{T}[q]$ causes the variation of the action functional $\mathcal{L}[q,\lambda]$ with respect to q and λ to vanish.

The variation of the action functional $\mathcal{L}[q,\lambda]$ is given by:

$$\begin{aligned}\delta\mathcal{L}(q,\lambda) &= \delta\mathcal{T}(q) + \delta(\lambda\mathcal{I}(q)) \\ &= E_\theta\left[\int_\Omega \frac{\partial F}{\partial q}(\boldsymbol{x},\boldsymbol{\theta},q(\boldsymbol{x}))\delta q(\boldsymbol{x})d\boldsymbol{x}\right] + \lambda\left(\int_\Omega \delta q(\boldsymbol{x})d\boldsymbol{x}\right) + \mathcal{I}(q)\delta\lambda \\ &= E_\theta\left[\int_\Omega \left(\frac{\partial F}{\partial q}(\boldsymbol{x},\boldsymbol{\theta},q(\boldsymbol{x})) + \lambda\right)\delta q(\boldsymbol{x})d\boldsymbol{x}\right] + \left(\int_\Omega q(\boldsymbol{x})d\boldsymbol{x} - 1\right)\delta\lambda\end{aligned} \tag{22}$$

By the fundamental lemma of the calculus of variations, this expression has to vanish for all variations δq and $\delta\lambda$, which is equivalent to

$$\begin{aligned}E_\theta\left[\frac{\partial F}{\partial q}(\boldsymbol{x},\boldsymbol{\theta},q(\boldsymbol{x})) + \lambda\right] = 0, \\ \int_\Omega q(\boldsymbol{x})d\boldsymbol{x} - 1 = 0\end{aligned} \tag{23}$$

In the case of the total expected squared Hellinger distance,

$$\begin{aligned}\frac{\partial F}{\partial q}(\boldsymbol{x},\boldsymbol{\theta},q(\boldsymbol{x})) &= \frac{1}{2}\frac{\partial\left(\sum_{i=1}^{N_d}\left(p_i(\boldsymbol{x}|\boldsymbol{\theta}) - 2\sqrt{(p_i(\boldsymbol{x}|\boldsymbol{\theta})q(\boldsymbol{x}))} + q(\boldsymbol{x})\right)\right)}{\partial q} \\ &= \frac{1}{2}\sum_{i=1}^{N_d}\left(-\frac{\sqrt{p_i(\boldsymbol{x}|\boldsymbol{\theta})}}{\sqrt{q(\boldsymbol{x})}} + 1\right) \\ &= \frac{N_d}{2} - \frac{1}{2}\sum_{i=1}^{N_d}\frac{\sqrt{p_i(\boldsymbol{x}|\boldsymbol{\theta})}}{\sqrt{q(\boldsymbol{x})}}\end{aligned} \tag{24}$$

Applying the expectation with respect to $\boldsymbol{\theta}$ yields:

$$\begin{aligned}E_\theta\left[\frac{\partial F}{\partial q}(\boldsymbol{x},\boldsymbol{\theta},q(\boldsymbol{x})) + \lambda\right] &= E_\theta\left[\frac{N_d}{2} - \frac{1}{2}\sum_{i=1}^{N_d}\frac{\sqrt{p_i(\boldsymbol{x}|\boldsymbol{\theta})}}{\sqrt{q(\boldsymbol{x})}} + \lambda\right] \\ &= \frac{N_d + 2\lambda}{2} - \frac{1}{2}\sum_{i=1}^{N_d}E_\theta\left[\frac{\sqrt{p_i(\boldsymbol{x}|\boldsymbol{\theta})}}{\sqrt{q(\boldsymbol{x})}}\right]\end{aligned} \tag{25}$$

Setting Eq. (25) equal to 0 and solving for $q(\boldsymbol{x})$ gives the minimizer, which is the optimal sampling density

$$q^*(\boldsymbol{x}) = \left(\frac{1}{N_d + 2\lambda}\right)^2\left(\sum_{i=1}^{N_d}E_\theta\left[\sqrt{p_i(\boldsymbol{x}|\boldsymbol{\theta})}\right]\right)^2 \tag{26}$$

where λ is chosen to satisfy $\int_\Omega q^*(\boldsymbol{x})d\boldsymbol{x} = 1$. However, it is not necessarily straightforward to identify λ given the large (theoretically infinite) number of densities $p_i(\boldsymbol{x}|\boldsymbol{\theta})$.

Alternatively, let us define the optimization with respect to the mean square difference (MSD), given by

$$\mathcal{M}(P\|Q) = \frac{1}{2}\int_\Omega (p(\boldsymbol{x}) - q(\boldsymbol{x}))^2 d\boldsymbol{x} \tag{27}$$

The total expected mean squared difference can be written as

$$\begin{aligned}E[\mathcal{M}(\mathbb{M}\|Q)] = \sum_{i=1}^{N_d}E_\theta[\mathcal{M}(M_i\|Q)] &= \sum_{i=1}^{N_d}\frac{1}{2}E_\theta\left[\int_\Omega (p_i(\boldsymbol{x}|\boldsymbol{\theta}) - q(\boldsymbol{x}))^2 d\boldsymbol{x}\right] \\ &= E_\theta\left[\int_\Omega \sum_{i=1}^{N_d}\frac{1}{2}(p_i(\boldsymbol{x}|\boldsymbol{\theta}) - q(\boldsymbol{x}))^2 d\boldsymbol{x}\right]\end{aligned} \tag{28}$$

Again, we define the optimization problem in terms of the functional $\hat{\mathcal{T}}[q]$ and the isoperimetric constraint $\hat{\mathcal{I}}[q]$ as

$$\begin{aligned}&\underset{q}{\text{minimize}}\ \hat{\mathcal{T}}[q] = E_\theta\left[\int_\Omega \widehat{F}(\boldsymbol{x},\boldsymbol{\theta},q(\boldsymbol{x}))d\boldsymbol{x}\right] \\ &\text{subject to } \hat{\mathcal{I}}[q] = \int_\Omega q(\boldsymbol{x})d\boldsymbol{x} - 1 = 0\end{aligned} \tag{29}$$

where

$$\widehat{F}(\boldsymbol{x},\boldsymbol{\theta},q(\boldsymbol{x}))=\frac{1}{2}\sum_{i=1}^{N_d}(p_i(\boldsymbol{x}|\boldsymbol{\theta})-q(\boldsymbol{x}))^2 \tag{30}$$

Similarly, we define the action functional $\hat{\mathcal{L}}[q,\lambda]$ as in Eq. (21) and eliminate all variations with respect to q and λ, which leads to

$$\begin{aligned}E_\theta\left[\frac{\partial F}{\partial q}(\boldsymbol{x},\boldsymbol{\theta},q(\boldsymbol{x}))+\lambda\right]&=E_\theta\left[-\sum_{i=1}^{N_d}(p_i(\boldsymbol{x}|\boldsymbol{\theta})-q(\boldsymbol{x}))+\lambda\right]\\&=q(\boldsymbol{x})N_d+\lambda-\sum_{i=1}^{N_d}E_\theta[p_i(\boldsymbol{x}|\boldsymbol{\theta})]\\&=0\end{aligned} \tag{31}$$

Solving for $q(\boldsymbol{x})$ gives the minimizer

$$\hat{q}^*(\boldsymbol{x})=\frac{1}{N_d}\left(\sum_{i=1}^{N_d}E_\theta[p_i(\boldsymbol{x}|\boldsymbol{\theta})]-\lambda\right) \tag{32}$$

which leads to $\lambda=0$ such that $\int_\Omega q^*(\boldsymbol{x})d\boldsymbol{x}=1$. Hence, the optimal sampling density

$$\hat{q}^*(\boldsymbol{x})=\frac{1}{N_d}\sum_{i=1}^{N_d}E_\theta[p_i(\boldsymbol{x}|\boldsymbol{\theta})] \tag{33}$$

is a mixture distribution combining the candidate target densities and their parameter ranges. It assumes that each probability model M_i is equally probable (i.e. $\mathbb{p}_i=\frac{1}{\mathbb{N}_d}$). It is straightforward to show that this solution generalizes as:

$$\hat{q}^*(\boldsymbol{x})=\sum_{i=1}^{N_d}\mathbb{p}_i\mathbb{E}_\theta[p_i(\boldsymbol{x}|\boldsymbol{\theta})] \tag{34}$$

where $\mathbb{p}_i$ is the AIC_c model probability (see Eq. (10)) for model M_i satisfying $\sum_{i=1}^{N_d}\mathbb{p}_i=\mathbb{1}$.

5. Propagation of imprecise probabilities

With the constituents outlined in the previous sections, the proposed methodology is summarized here and a flowchart is provided in Fig. 2.

- *Step 1: Multimodel inference* – Given a limited dataset $\boldsymbol{d}$, first identify the set of candidate probability models $\mathbb{M}=\{M_i\}, i=1,2,\ldots,N_d$ that will be considered. Using multimodel inference, compute AIC_c and the associated model probabilities $\mathbb{p}_i$ for each candidate model M_i, using Eqs. (9)–(11). Models with extremely low probability (alternatively very large $\Delta_A^{(i)}$) may be discarded.
- *Step 2: Bayesian inference* – For each candidate probability model M_i, apply Bayes' rule to identify the joint probability density of the model parameters, $p(\boldsymbol{\theta}|\boldsymbol{d},M_i),\ i=1,2,\ldots,N_d$. In general, this is conducted using the MCMC method.
- *Step 3: Establish a finite model set* – In theory, Steps 1–2 yield an infinite set of parametrized probability models. This set is reduced to a finite (but large) set of models using a classical Monte Carlo sampling approach. For each model in the finite set, the model family M_i is selected randomly with probability $\mathbb{p}_i$. Next, the parameter values of the model are randomly drawn from $p(\boldsymbol{\theta}|\boldsymbol{d},M_i)$. Here it may be advantageous to sample directly from the previous MCMC draws conducted in Step 2. It is important to draw a sufficiently large model set here (typically several thousand density functions will make up this set) in order to span the full range of the candidate models. Undersampling this space will lead to artificially narrow probability bounds and adding additional densities comes only at the small cost of additional importance sample reweighting. No additional function evaluations are required in order to add additional densities. It is also possible to add densities if, after completing the analysis, the user has reason to believe the probability bounds may be artificially narrow (Section 6).
- *Step 4: Determine the optimal sampling density* – The importance sampling density is determined through optimization as discussed in Section 4.3. In this work, we utilize the analytically obtained optimal density $\hat{q}^*(\boldsymbol{x})$ from Eq. (34) that minimizes the total expected mean squared difference, which comes at little computational cost. Other optimization schemes may add significant expense. Other sampling densities may also be explored such as the parameter integrated and model averaged densities proposed in [34,31].

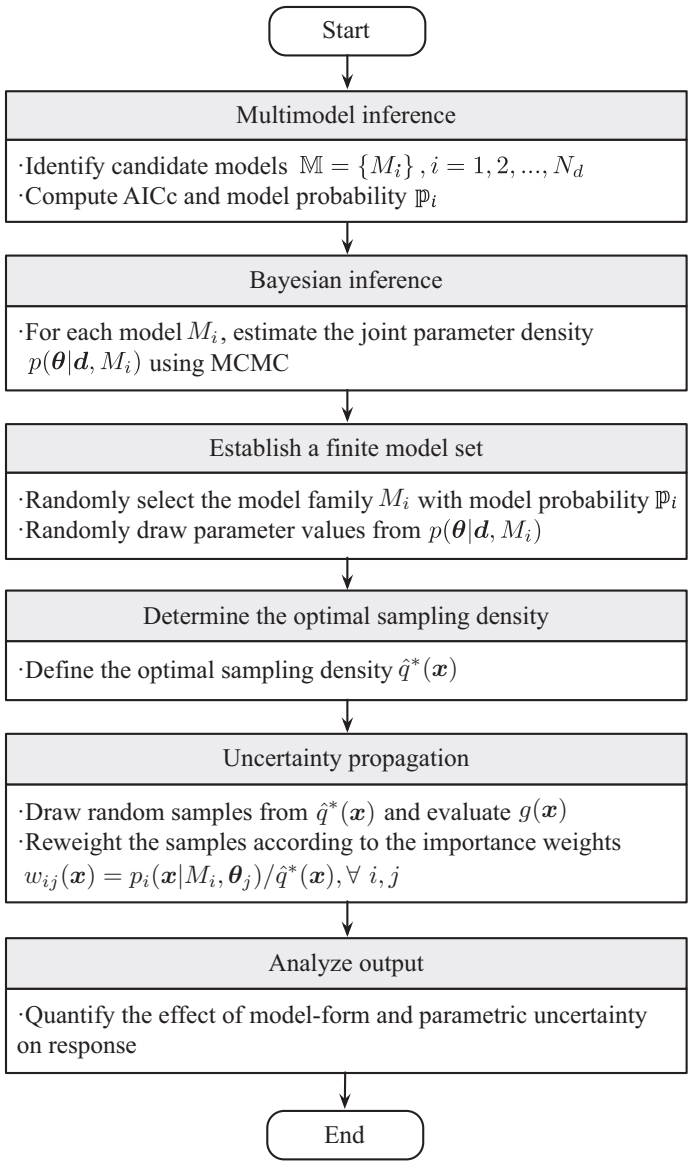

Fig. 2. Flowchart of the proposed method for propagation of imprecise probabilities.

- *Step 5: Uncertainty propagation* – Uncertainty is propagated using importance sampling with sampling density $\hat{q}^*(\boldsymbol{x})$. Samples are drawn from $\hat{q}^*(\boldsymbol{x})$ and are re-weighted according to the importance weights $w_{ij}(\boldsymbol{x}) = \frac{p_i(\boldsymbol{x}|M_i,\boldsymbol{\theta}_j)}{\hat{q}^*(\boldsymbol{x})}, \forall i, j$. That is, each sample drawn from $\hat{q}^*(\boldsymbol{x})$ is re-weighted a very large number of times (often several thousand) according to each of the plausible probability models. With these reweighted sets, statistical analysis is performed across all candidate models to obtain response statistics, distributions, etc. corresponding to each candidate model.

6. Model updating and notable limitations

The proposed approach offers a high degree of flexibility in that it may be easily updated to accommodate new data or the addition of new candidate probability models. The singular drawback in this updating is a loss of optimality in the importance sampling density.

6.1. Adding data

Consider the case where the original dataset $\boldsymbol{d}$ are found to produce undesirably large uncertainties in the output. To reduce these uncertainties, additional data $\boldsymbol{d}^*$ are collected using the same testing methods such that they can be reasonably

assumed to be drawn from the same statistical distribution as the original data. Using the proposed approach, the multi-model inference can be directly updated by computing the new AIC_c values and corresponding model probabilities and the parameter distributions can be updated using classical Bayesian updating. This will likely provide increased confidence in the model form while also narrowing the parameter densities. However, this updating comes at the expense of optimality of the importance sampling density, the effect of which might be an increase in the variance of statistical estimates from the importance sampling method. Note this does not necessitate recomputing the optimal sampling density and resampling. In certain cases, the variance increase may warrant additional samples. In this case, a new optimal sampling density may be determined, additional samples drawn from this new density, and the aggregate sample set (old and new samples) combined using a method like multiple importance sampling [48]. This is not explored further here.

6.2. Adding probability models

If the uncertainty ranges in the output appear artificially narrow, it may be the result of undersampling the candidate model space. Once again, additional models can be produced from the model-form and parameter probabilities without necessarily determining a new optimal sampling density at the expense of an increase in variance of statistical estimators. Additional samples may be drawn as described in the previous section if needed.

6.3. Notable limitations

The proposed method, even with the flexibility discussed herein is not without its limitations. Most notably, at this juncture the set of candidate probability models has been constrained to those well-established distributions with a parametric form. However, more general non-parametric models are not currently considered and pose several challenges (e.g. how to identify a non-parametric model from small data without overfitting) that are not addressed here. Moreover, the method is limited by the candidate probability models supplied by the user. Given a sufficiently diverse set of candidate models, we believe that the set of models is sufficient to span the epistemic uncertainty but unfortunately cannot prove nor test this hypothesis at this point. Fortunately, as discussed in Section 6.2, the method is robust enough that additional probability models can be added a posteriori at very little additional cost if the user has reason to believe the candidate models do not sufficiently span the epistemic uncertainty.

7. Application to uncertainty quantification in plate buckling strength

The proposed methodology is illustrated through uncertainty quantification in the prediction of the buckling strength of a simply supported rectangular plate under uniaxial compression considering certain geometric and material uncertainties. Faulkner [49] first proposed an analytical expression for the normalized buckling strength for a pristine plate as

$$\psi = \frac{\sigma_u}{\sigma_0} = \frac{2}{\lambda} - \frac{1}{\lambda^2} \tag{35}$$

where σ_0 is the yield stress, σ_u is the stress at failure and λ is the slenderness of the plate with width b, thickness t, and elastic modulus E given by

$$\lambda = \frac{b}{t}\sqrt{\frac{\sigma_0}{E}} \tag{36}$$

Carlsen [50] modified this equation to consider the effects of non-dimensional initial deflections δ_0 and residual stresses associated with welding

$$\psi = \left(\frac{2.1}{\lambda} - \frac{0.9}{\lambda^2}\right)\left(1 - \frac{0.75\delta_0}{\lambda}\right)\left(1 - \frac{2\eta t}{b}\right) \tag{37}$$

where ηt is the width of the zone of tension residual stress.

Uncertainty analysis of such plate structures has attracted much attention, especially within the naval community [51,52]. We are interested in studying the influence of the six uncertain variables listed in Table 1 on the buckling strength of simply supported mild steel plates. The geometric and material variabilities for these plates have been estimated from data by Hess et al. [51] and Guedes Soares [52], as presented in Table 1.

Although extension to the full 6-variable case is straightforward and has been conducted, we will focus on the influence of a single material parameter, the yield stress σ_0, for clarity and brevity in demonstration. Global sensitivity analysis (GSA) is used to identify the relative contributions of the input variabilities to the buckling strength variability [53]. In this work, we employ the GSAT tool for GSA [54] based on Sobol's approach [55] and Fourier Amplitude Sensitivity Test (FAST) [56] to identify the contribution of each variable. The sensitivity indices are shown in Table 1 where it is observed that the buckling strength is most sensitive to the yield stress σ_0 and least sensitive to the plate width b. We start with only 10 yield stress data, and then consider the cases with added data.

Table 1
Statistical properties of plate material, geometry and imperfection variables from Hess [51] and Guedes Soares [52].

Variables	Physical Meaning	Nominal Value	Mean	COV	Global Sensitivity
b	Width	24	0.992∗24	0.028	0.017
t	Thickness	0.5	1.05∗0.5	0.044	0.045
σ_0	Yield stress	34	1.3∗34	0.1235	0.482
E	Young modulus	29,000	0.987∗29,000	0.076	0.194
δ_0	Initial deflection	0.35	1.0∗0.35	0.05	0.043
η	Residual stress	5.25	1.0∗5.25	0.07	0.233

The nominal design value for yield stress of mild steel used for ship structures is $\sigma_0 = 34$ ksi so that we define a random variable $\hat{\sigma}_0$ as the deviation from this nominal value as $\hat{\sigma}_0 = \sigma_0 - 34$. Data collected by Hess et al. [51] suggest that the "true" mean yield stress is actually closer to $\mu_{\sigma_0} \approx 1.3 * 34 = 44.2$ with coefficient of variation 0.1235 and following a Lognormal distribution. We assume this to be the truth (while recognizing that it is, in fact, an approximation) and generate 10 random yield stress values from $\sigma_0 = 34 + \hat{\sigma}_0$ with $\hat{\sigma}_0 \sim Lognormal(\mu_{\hat{\sigma}_0} = 1.3 * 34 - 34, \sigma_{\hat{\sigma}_0} = 0.1235 * 1.3 * 34)$, as shown in Fig. 3. These serve as the initial data from which uncertainty will be quantified and propagated. Clearly, a probability model form cannot be precisely identified from these data.

The nominal design yield stress provides a lower-bound which we assume is strictly enforced. In practice this may not be the case but is considered a reasonable assumption given sufficient quality control. This enables us to consider only probability models with support on the positive real line $(0, \infty)$. The selected candidate models are given in Table 2. In this table, the AIC_c model selection criteria is employed to rank the candidate distributions. The top seven probability models have very similar AIC_c values (i.e. very small $\Delta_A^{(i)}$). Correspondingly, there is little difference among the AIC_c probabilities for these seven models. This is a classic small data case where a precise "best" model is impossible to identify. The last three distribution models have much larger AIC_c values and very low probabilities. We consider this sufficient evidence for their removal from the model set. Therefore, we consider the top seven probability models to represent this limited dataset.

Next, Bayesian inference is employed to estimate the joint parameter probability densities for the seven models. We assume little a priori knowledge and start with a uniform prior $p(\boldsymbol{\theta})$ with bounds $(0, 10^6]$. Fig. 4 shows the joint posterior parameter distributions for six of the models in Table 2. The upper-left and lower-right plots in each group show kernel density estimates of the posterior marginal densities for each parameter, and the lower-left plot gives contours of the joint posterior density function estimated from the MCMC samples. Note that the Rayleigh distribution has only one parameter, so uncertainty is quantified in terms of its marginal density, which is not shown.

The seven probability models and their joint parameter densities are discretized to create a finite set of models by Monte Carlo sampling. For each sample, the probability model is randomly drawn according to the AIC_c probabilities $\mathbb{p}_i$ in Table 2 and the parameters of that model are then drawn randomly from the corresponding joint parameter distribution in Fig. 4. Fig. 5(a) shows 5000 candidate densities for the yield stress drawn from the model set along with two optimal sampling densities. Optimal sampling density A, $q^*(\boldsymbol{x})$, is identified by minimizing the total expected squared Hellinger distance using Eq. (26). Optimal sampling density B, $\hat{q}^*(\boldsymbol{x})$, is determined by minimizing the total expected mean square distance using Eq. (34). Optimal sampling density B is used throughout the remainder of this work given that it is comparatively easy to identify from the optimization and it samples more densely in the lower tail. Using this optimal sampling density, the "cloud" of distributions in Fig. 5(a) are propagated using importance sampling reweighting of 50,000 samples drawn

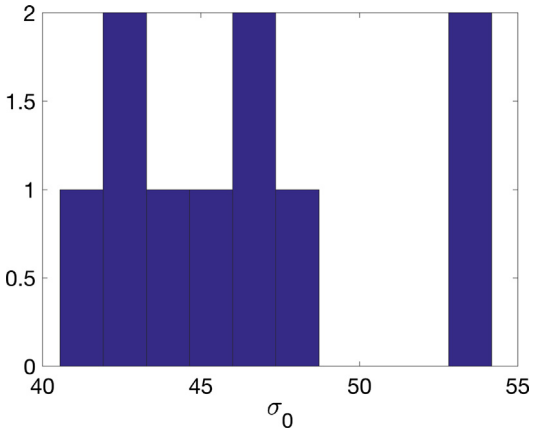

Fig. 3. Ten randomly generated yield stress values that serve as the initial dataset for uncertainty quantification and propagation in plate buckling strength.

Table 2
Ranked candidate probability models based on AIC_c given 10 yield stress values

Rank	Distribution	AIC_c	$\Delta_A^{(i)}$	$\mathbb{P}_i$
1	Inverse Gaussian	61.615	0.000	0.185
2	Lognormal	61.753	0.138	0.173
3	Gamma	61.954	0.338	0.156
4	Log-logistic	62.280	0.664	0.133
5	Rayleigh	62.357	0.742	0.128
6	Nakagmai	62.381	0.765	0.126
7	Weibull	62.956	1.341	0.095
8	Levy	69.952	8.337	0.003
9	Exponential	72.750	11.134	0.001
10	F	98.389	36.774	0.000

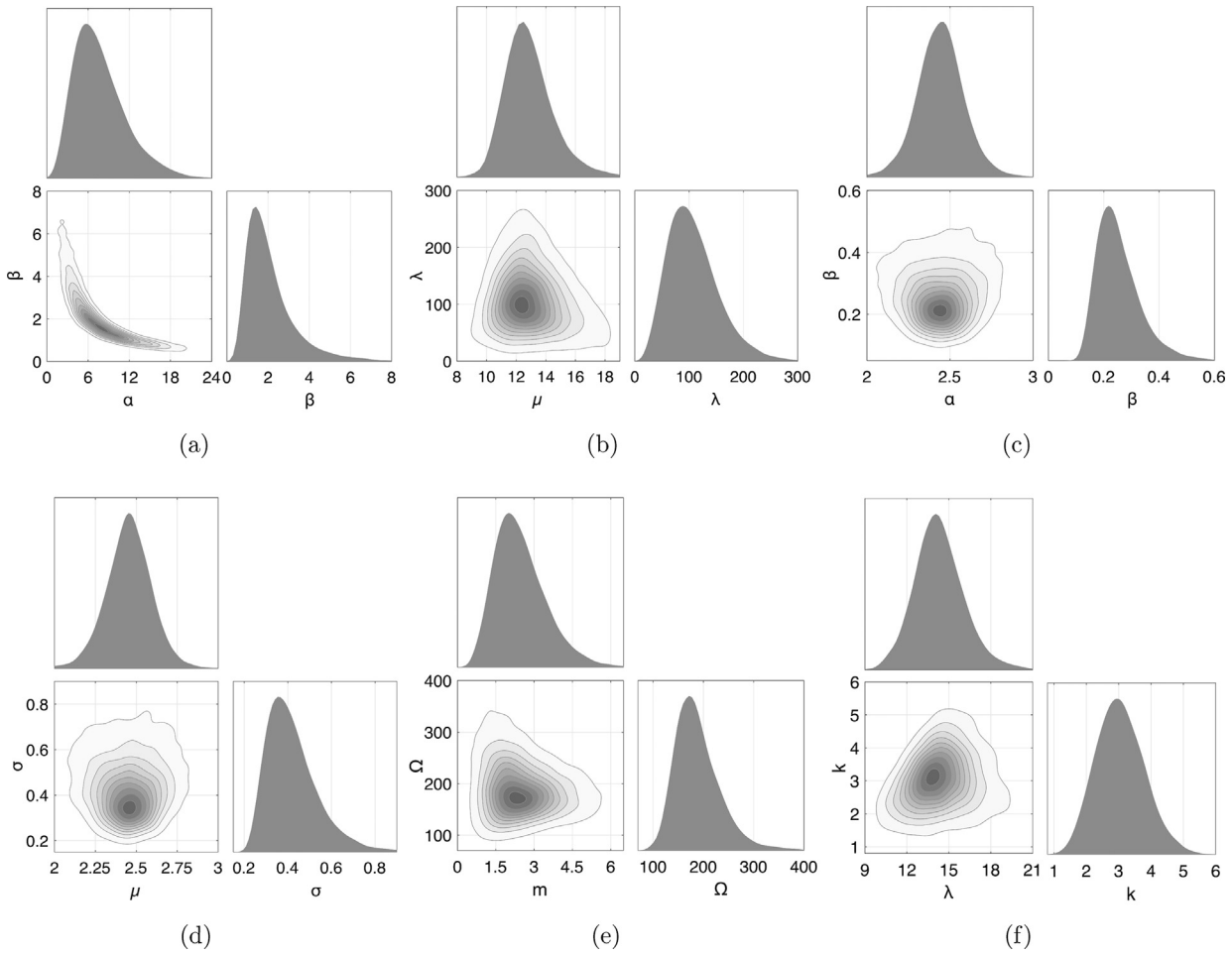

Fig. 4. MCMC posterior joint parameter densities for the following probability models: (a) Gamma, (b) Inverse Gaussian, (c) Loglogistic, (d) Lognormal, (e) Nakagami, and (f) Weibull.

randomly from $\hat{q}^*(\boldsymbol{x})$. This yields a corresponding cloud of cumulative distribution functions for the buckling strength in Fig. 5(b) that shows the range of probabilistic response resulting from both model-form and parametric uncertainties caused by the small dataset for the yield stress. Notice that the strength range is very wide at each fixed probability level.

7.1. Separating model-form and parametric uncertainties

It is further possible to delineate the effects of model-form and parametric uncertainties. This can be seen by the color coding of the CDFs in Fig. 5 according to their underlying model form. The variations between colors represents model-

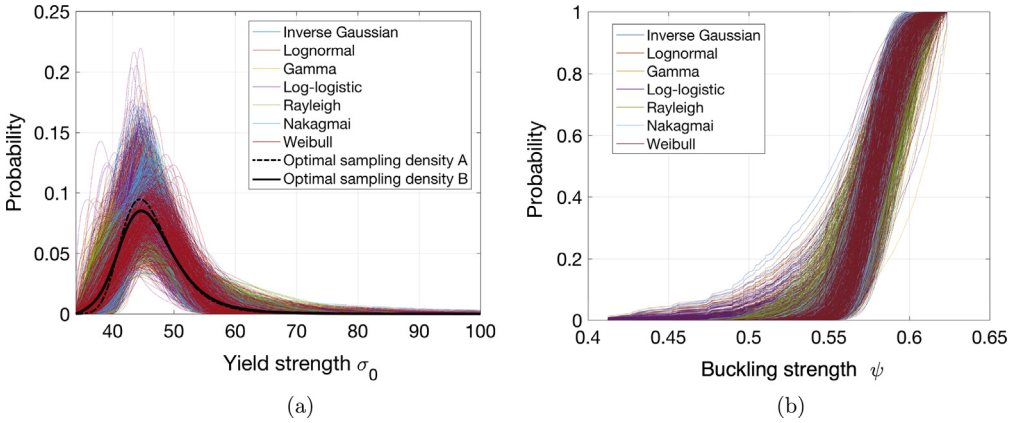

Fig. 5. (a) Candidate pdfs and the optimal sampling densities from ten yield stress values, and (b) collection of candidate empirical CDFs for buckling strength ψ.

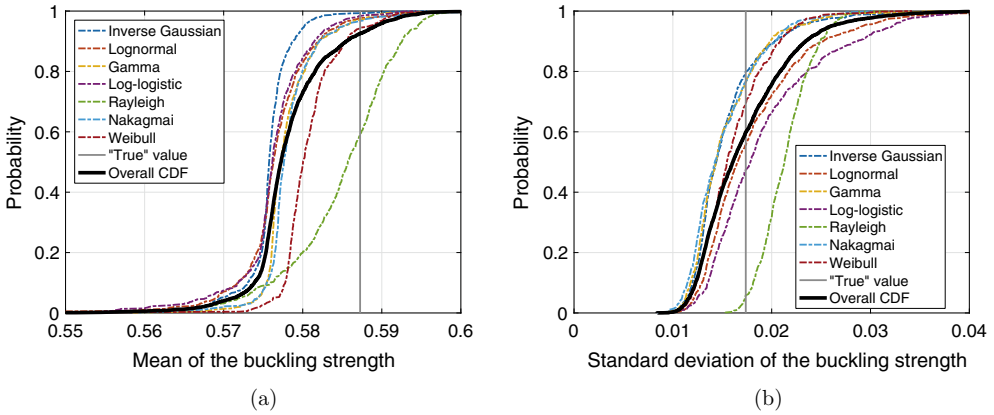

Fig. 6. Empirical CDFs for (a) mean of buckling strength, and (b) standard deviation of buckling strength ψ.

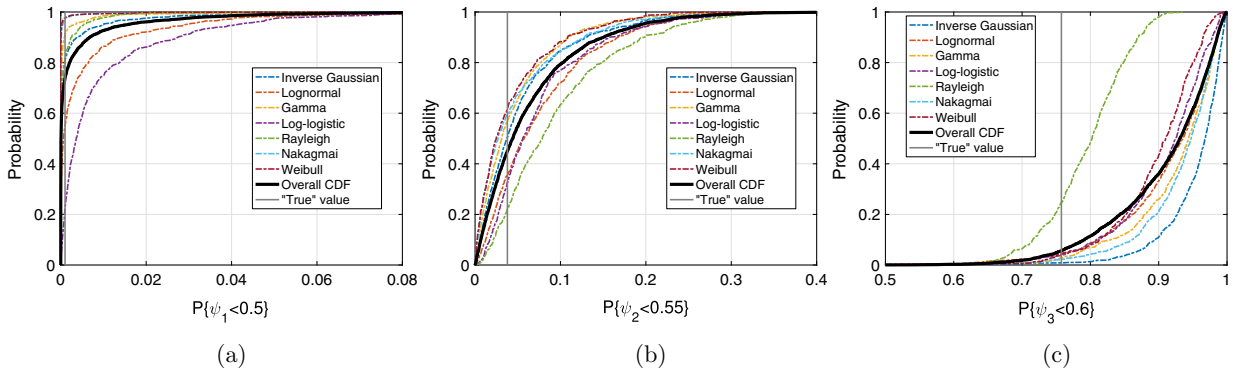

Fig. 7. Empirical CDFs for the probability of failure occurs when (a) $\psi_1 < 0.5$, (b) $\psi_2 < 0.55$, and (c) $\psi_3 < 0.6$.

form uncertainty while the variations with a single color correspond to the parametric uncertainties for a given model-form. Perhaps a more illustrative approach is to view CDFs of the statistical response according to the different models. Fig. 6 shows CDFs of the mean and standard deviation of the buckling strength. This figure provides the overall CDF considering all probability models (thick black line) as well as the conditional CDFs for each candidate probability model. These conditional CDFs include only the effects of parametric uncertainties given the assignment of a specific model form. Note that the conditional CDFs do not occur with equal probability but rather occur with the AIC_c model probabilities in Table 2.

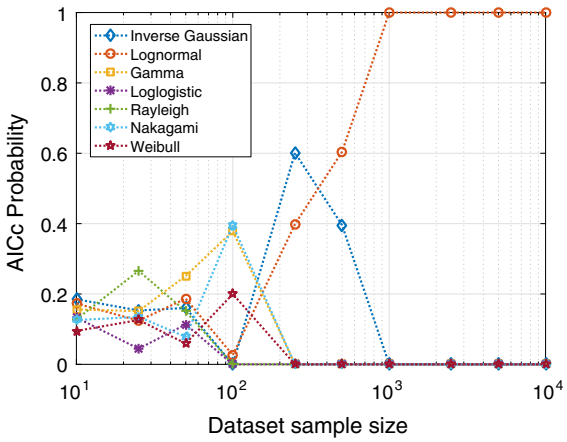

Fig. 8. AIC_c probability as a function of dataset size.

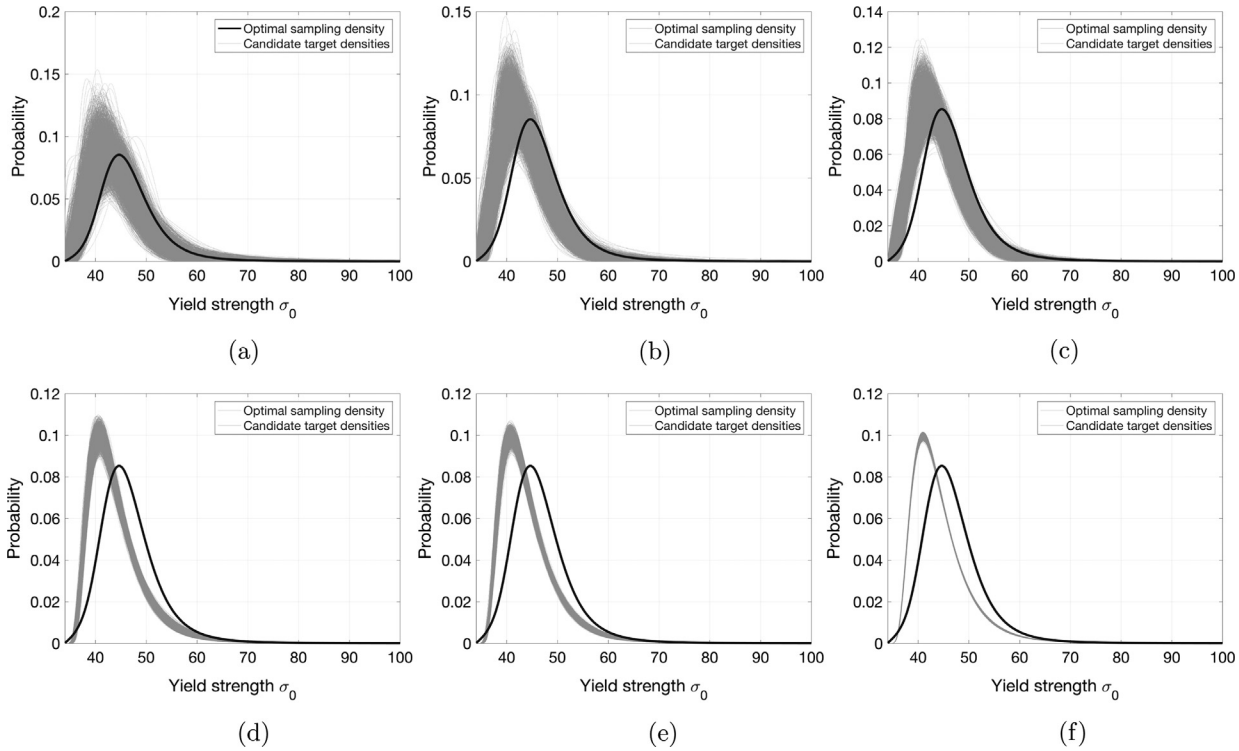

Fig. 9. Optimal sampling density with candidate target densities based on: (a) 25 data, (b) 50 data, (c) 100 data, (d) 500 data, (e) 1000 data, and (f) 10,000 data.

A similar investigation has been conducted to study the effects of model-form and parametric uncertainty on probability of failure. Consider three cases where "failure" is assumed to occur when $\psi_1 < 0.5$, $\psi_2 < 0.55$ and $\psi_3 < 0.6$. The empirical CDFs for $P\{\psi_1 < 0.5\}$, $P\{\psi_2 < 0.55\}$ and $P\{\psi_3 < 0.6\}$ are given in Fig. 7. Again these figures show the overall CDF as well as the conditional CDFs assuming each specific probability model.

The CDFs in Figs. 6 and 7 highlight the large degree of uncertainty that results from having such little data. Even the mean normalized buckling strength, which is straightforward to assess when probabilities are precise, ranges from 0.55 to 0.6. The probability of failure ranges, meanwhile, are so large that one cannot assume any real confidence in the estimates. Furthermore, while we can see that the dominant effects are the parametric uncertainties, it is clear that the model-form uncertainties are important as well since there exist significant differences between the model specific CDFs in Figs. 6 and 7.

7.2. Influence of dataset size

Here, we are interested in studying the convergence of the uncertain response as a function of the amount of data collected. The previous section highlighted how very small datasets led to very large uncertainties (model-form and parametric) in the buckling strength. This begs the question: "How much data is necessary to reduce this uncertainty?".

Let us begin with model-form uncertainty. Fig. 8 shows the AIC_c probabilities for each candidate model as a function of dataset size. Notice that the multimodel inference does not pick out the correct Lognormal model conclusively until 1000 yield stress measurements are collected. In fact, prior to that point the inference placed a high level of probability on the incorrect Inverse Gaussian model. How, then does this influence the uncertainty in buckling strength?

Figs. 9–13 show the evolution of various quantities for increasing dataset sizes. Fig. 9 shows the candidate target densities along with the original optimal sampling density (based on only 10 data). Here we emphasize the loss of optimality as data are added but note that this density may still be used for propagation (at the potential expense of increased solution variance). As expected, the cloud of candidate densities narrows as data are added. Fig. 10 shows the corresponding set of CDFs for the buckling strength. An important feature here is that no new function evaluations were necessary to produce these CDFs. These are quantified directly from the 50,000 samples drawn from the original optimal sampling density reweighted to reflect the updated candidate densities. Again, the cloud of response CDFs narrows toward the "true" as would be expected as additional data are collected.

Fig. 11–13 show the mean, standard deviation, and probability of failure $P\{\psi < 0.5\}$ CDFs for various dataset sizes. Both the overall CDFs and conditional CDFs for each candidate model are shown along with the "true" statistical values obtained using Monte Carlo simulation with 10^6 samples from the aforementioned Lognormal distribution. We notice that the model-form uncertainty for the mean and standard deviation diminishes rapidly and becomes largely insignificant after $\sim$ 100 yield stress data are available. Probability of failure, meanwhile, retains significant model-form uncertainty until $\sim$ 1000 yield stress data are collected, at which point the Inverse Gaussian model can be eliminated as a candidate.

Parametric uncertainty accounts for the greater proportion of the total uncertainty and is reflected by the breadth of the CDFs for the statistical quantities in Figs. 11–13. A lack of parametric uncertainty therefore corresponds to a CDF that is a simple step function. We notice that the parametric uncertainty diminishes more slowly than the model-form uncertainty and can remain noticeable even for very large data sets. Nonetheless, as expected, the CDFs gradually converge toward the true statistical values as the sample size grows.

For decision makers, it is useful to either establish a confidence threshold and determine how much data must be collected to meet this threshold or to determine the amount of data you can afford to collect and estimate the level of

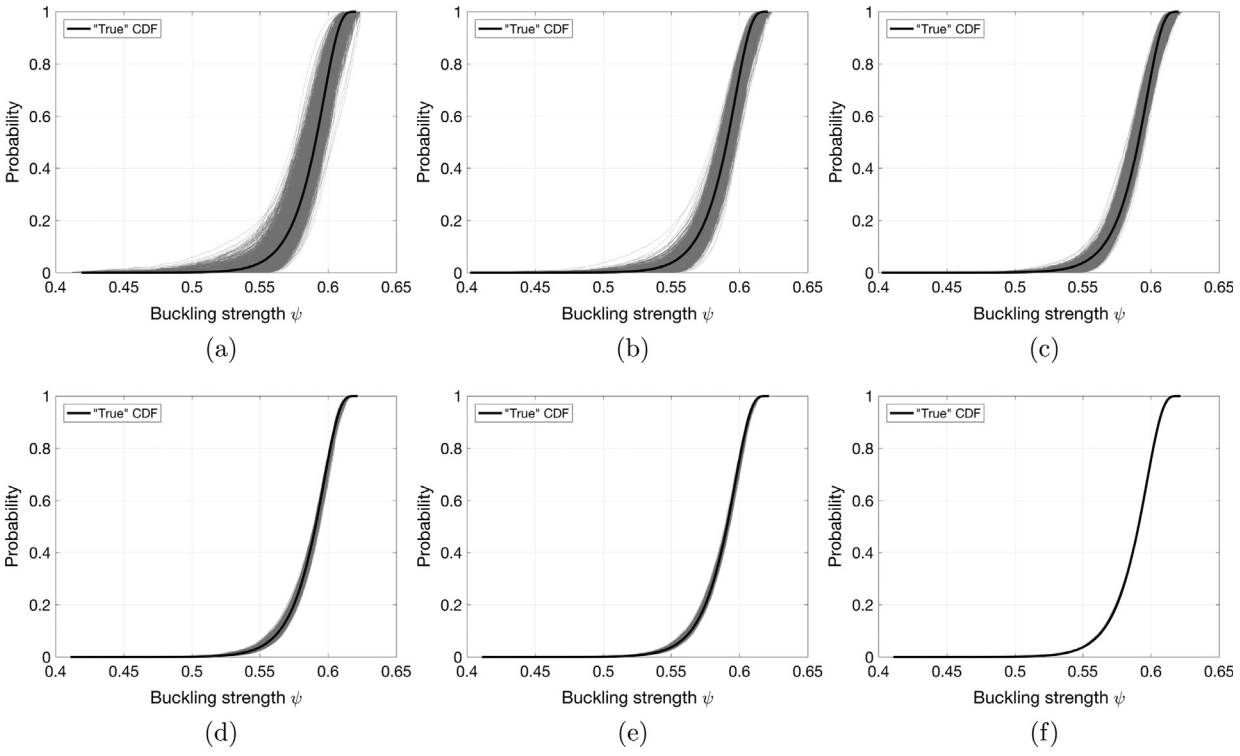

Fig. 10. CDFs for the buckling strength based on: (a) 25 data, (b) 50 data, (c) 100 data, (d) 500 data, (e) 1000 data, and (f) 10,000 data.

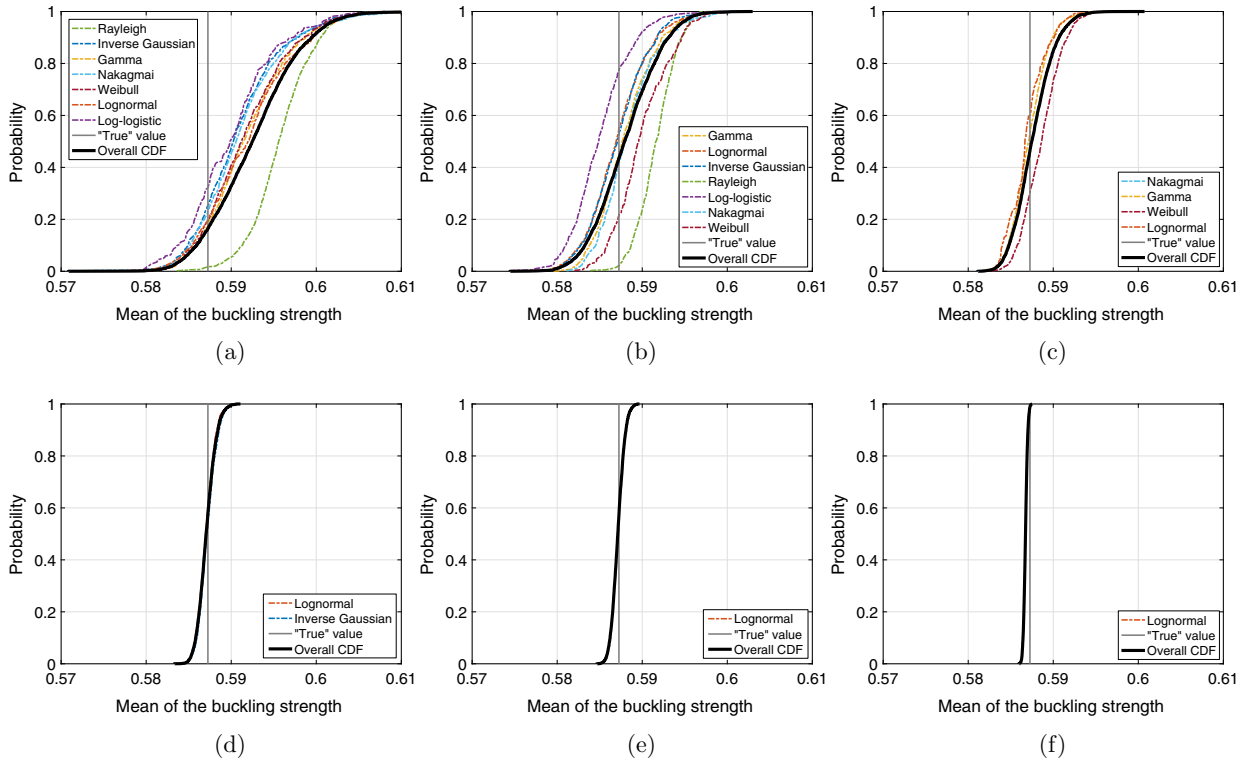

Fig. 11. CDFs for the mean buckling strength based on: (a) 25 data, (b) 50 data, (c) 100 data, (d) 500 data, (e) 1000 data, and (f) 10,000 data.

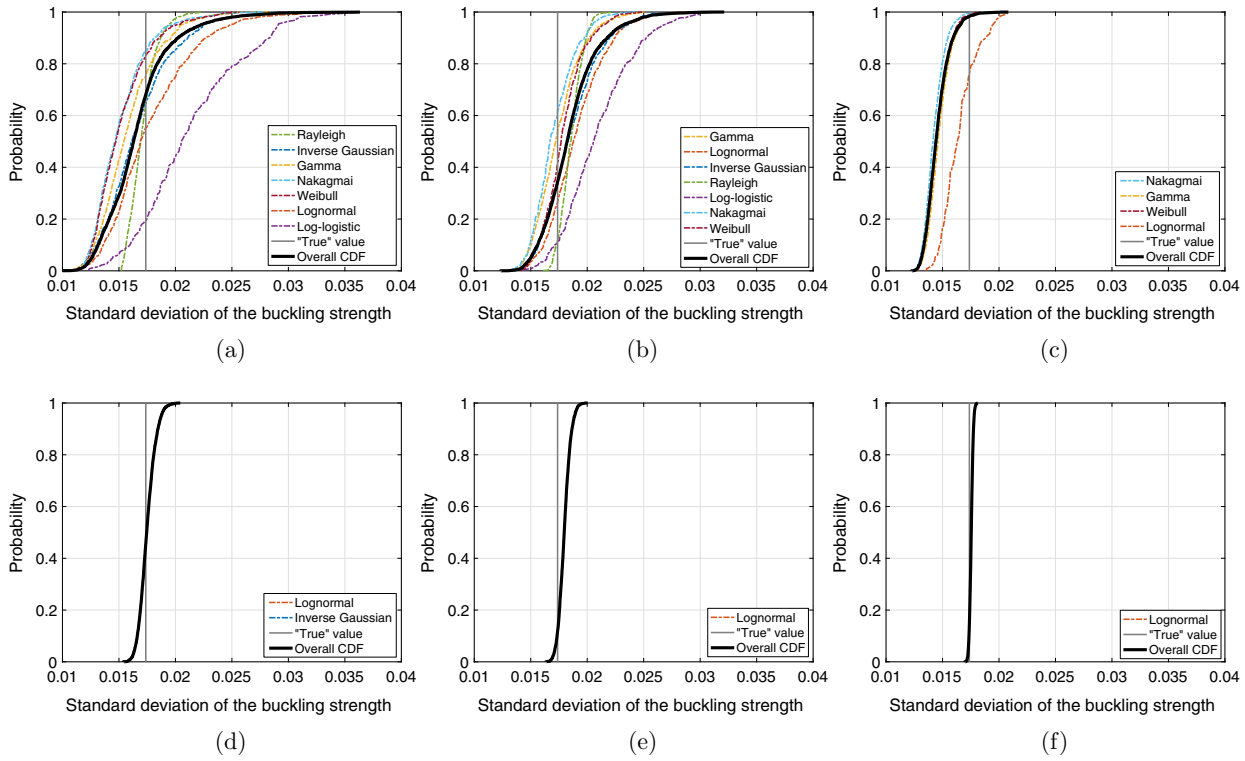

Fig. 12. CDFs for the standard deviation of the buckling strength based on: (a) 25 data, (b) 50 data, (c) 100 data, (d) 500 data, (e) 1000 data, and (f) 10,000 data.

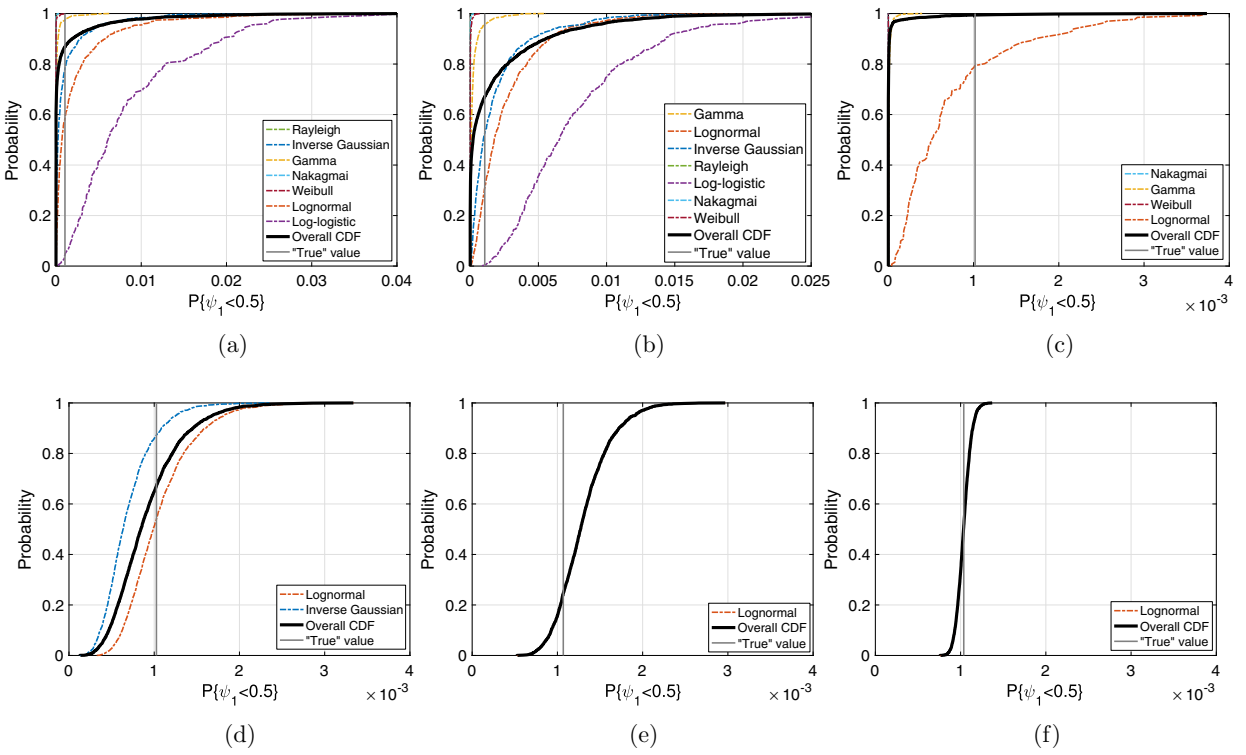

Fig. 13. CDFs for the probability of failure $P\{\psi < 0.5\}$ based on: (a) 25 data, (b) 50 data, (c) 100 data, (d) 500 data, (e) 1000 data, and (f) 10,000 data.

confidence that can be achieved. For simplicity, let us define a simple confidence metric as the range of the upper and lower quantiles of width 0.025 for statistical quantity Y considering both parametric and model-form uncertainties given n data by:

$$\delta_Y^{(n)} = Q_{0.975}(Y^{(n)}) - Q_{0.025}(Y^{(n)}) \tag{38}$$

The ranges for the mean, standard deviation, and probability of failure are therefore denoted $\delta_\mu^{(n)}$, $\delta_\sigma^{(n)}$, and $\delta_{(\psi<\psi^*)}^{(n)}$. Fig. 14 plots these bounds relative to dataset size. Plots of this type are very useful for decision makers because they provide easy-to-interpret information to aid in determining the appropriate amount of data to collect (e.g. number of experiments to conduct) and the confidence that can be achieved. Observations here imply that probability ranges (as defined by Eq. (38)) for the various statistics converge at a rate of $\frac{1}{n^2}$. If this is found to hold more generally, it may be possible to collect small datasets initially and project the level of confidence that will be achievable in order to determine how much additional data to collect in order to meet specified uncertainty tolerances. Plots of this nature can be easily generated for other measures. The quantile range in Eq. (38) is used simply for demonstration.

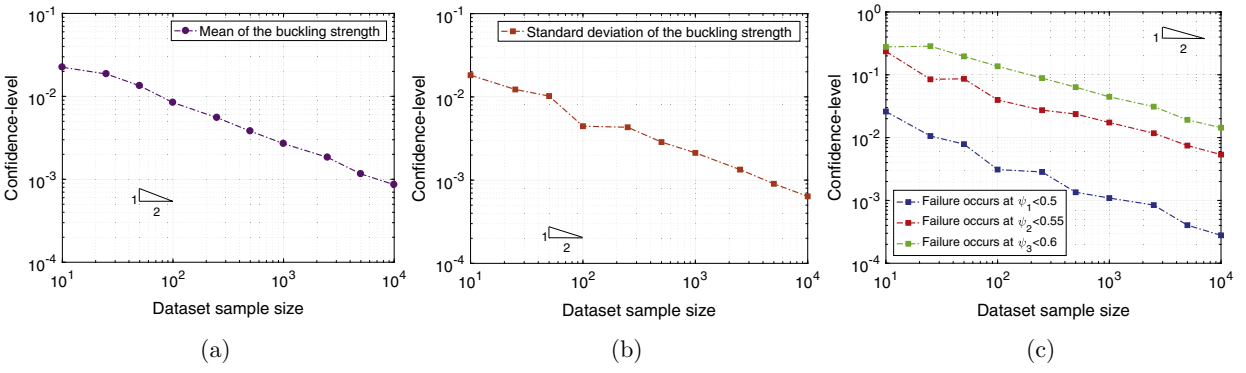

Fig. 14. Convergence of the quantile range for (a) mean, (b) standard deviation, and (c) probability of failure.

Conclusion

A methodology has been proposed for quantifying and propagating uncertainty created by a lack of sufficient statistical data. The methodology employs the concepts of multimodel inference to identify a set of candidate probability models that are plausibly representative of the given small dataset. It is shown that each of these candidate probability models can be assigned a probability that it is the "best" model for the data. For each plausible model, the joint probability density of the parameters is estimated using Bayesian inference. The candidate probability models (and their associated probabilities) along with the joint parameter densities thus quantify the full epistemic uncertainty associated with the small dataset.

Unlike prior methods that reduce this full probabilistic description to a single probability model, we retain and propagate all candidate models and their joint parameter densities. This is achieved by identifying, through an analytical optimization procedure, an optimal importance sampling density that best fits the full suite of models. This optimal sampling density is propagated using the Monte Carlo method and the samples are reweighted (using the importance weights) according to each of the candidate probability models. The result is a suite of possible probabilistic models for the response along with their associated probabilities (i.e. probabilities of probabilities) obtained at a fraction of the cost of conventional Monte Carlo methods. In effect, the procedure reduces a multi-loop Monte Carlo requiring n^3 samples to a single-loop Monte Carlo requiring n samples.

The effects of small dataset size in the assessment of plate buckling strength are analyzed. It is demonstrated that, when datasets are very small (i.e. $N < 100$), both model-form and parametric uncertainties are large and statistical analysis of response quantities shows large variability – highlighting our lack of confidence in the response. The effect of dataset size is carefully studied and the rate of convergence of statistics is shown to be on the order of $\frac{1}{N^2}$ with dataset size, N.

The proposed methodology effectively treats uncertainties associated with small datasets. Future work along these lines will serve to place the work more broadly within the context of imprecise probabilities such that it can be applied in cases where uncertainties are treated, for example, as intervals or fuzzy sets. Also, the uncertainty associated with dependencies between random variables will be explored as well as uncertainties associated with the selection of a prior probability model in the Bayesian inference.

Acknowledgements

The work presented herein has been supported by the Office of Naval Research under Award Number N00014-15-1-2754 with Dr. Paul Hess as program officer.

References

[1] A. Der Kiureghian, O. Ditlevsen, Aleatory or epistemic? Does it matter?, Struct Saf. 31 (2) (2009) 105–112.
[2] S. Ferson, L.R. Ginzburg, Different methods are needed to propagate ignorance and variability, Reliab. Eng. Syst. Saf. 54 (2–3) (1996) 133–144.
[3] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Stat. (1967) 325–339.
[4] G. Shafer, A Mathematical Theory of Evidence, vol. 1, Princeton University Press, 1976.
[5] L.A. Zadeh, Fuzzy sets, Inf. Control 8 (3) (1965) 338–353.
[6] R.E. Moore, Methods and Applications of Interval Analysis, vol. 2, SIAM, 1979.
[7] K. Weichselberger, The theory of interval-probability as a unifying concept for uncertainty, Int. J. Approx. Reason. 24 (2) (2000) 149–170.
[8] Y. Ben-Haim, I. Elishakoff, Convex Models of Uncertainty in Applied Mechanics, vol. 25, Elsevier, 2013.
[9] D. Dubois, H. Prade, Possibility Theory: An Approach to Computerized Processing of Uncertainty, Springer Science & Business Media, 2012.
[10] G.J. Klir, Uncertainty and Information: Foundations of Generalized Information Theory, John Wiley & Sons, 2005.
[11] I. Molchanov, Theory of Random Sets, vol. 53, Springer-Verlag, London, 2005.
[12] S. Ferson, V. Kreinovich, L. Ginzburg, D.S. Myers, K. Sentz, Constructing Probability Boxes and Dempster-Shafer Structures, vol. 835, Sandia National Laboratories Albuquerque, 2002.
[13] S. Ferson, J.G. Hajagos, Arithmetic with uncertain numbers: rigorous and (often) best possible answers, Reliab. Eng. Syst. Saf. 85 (1) (2004) 135–152.
[14] A.E. Raftery, Bayesian model selection in social research, Sociol. Methodol. (1995) 111–163.
[15] Y. Pawitan, In All Likelihood: Statistical Modelling and Inference using Likelihood, Oxford University Press, 2001.
[16] D. Dubois, H. Prade, Random sets and fuzzy interval analysis, Fuzzy Sets Syst. 42 (1) (1991) 87–101.
[17] D. Dubois, H. Prade, Interval-valued fuzzy sets, possibility theory and imprecise probability, in: EUSFLAT Conf., 2005, pp. 314–319.
[18] T. Fetz, M. Oberguggenberger, Propagation of uncertainty through multivariate functions in the framework of sets of probability measures, Reliab. Eng. Syst. Saf. 85 (1–3) (2004) 73–87.
[19] P. Walley, Statistical Reasoning with Imprecise Probabilities, vol. 42, Chapman & Hall, 1991.
[20] P. Walley, Towards a unified theory of imprecise probability, Int. J. Approx. Reason. 24 (2–3) (2000) 125–148.
[21] M. Beer, S. Ferson, V. Kreinovich, Imprecise probabilities in engineering analyses, Mech. Syst. Signal Process. 37 (1) (2013) 4–29.
[22] A. Der Kiureghian, Measures of structural safety under imperfect states of knowledge, J. Struct. Eng. 115 (5) (1989) 1119–1140.
[23] R. Zhang, S. Mahadevan, Integration of computation and testing for reliability estimation, Reliab. Eng. Syst. Saf. 74 (1) (2001) 13–21.
[24] A. Der Kiureghian, Analysis of structural reliability under parameter uncertainties, Probab. Eng. Mech. 23 (4) (2008) 351–358.
[25] E.F. Halpern, M.C. Weinstein, M.G.M. Hunink, G.S. Gazelle, Representing both first-and second-order uncertainties by Monte Carlo simulation for groups of patients, Med. Decis. Making 20 (3) (2000) 314–322.
[26] J.A. Hoeting, D. Madigan, A.E. Raftery, C.T. Volinsky, Bayesian model averaging: a tutorial, Stat. Sci. (1999) 382–401.
[27] F.J. Massey, The Kolmogorov-Smirnov test for goodness of fit, J. Am. Stat. Assoc. 46 (253) (1951) 68–78. <http://www.jstor.org/stable/228009>.
[28] J.M. Bernardo, R. Rueda, Bayesian hypothesis testing: a reference approach, Int. Stat. Rev. 70 (3) (2002) 351–372.
[29] H. Akaike, A new look at the statistical model identification, IEEE Trans. Autom. Control 19 (6) (1974) 716–723.
[30] M.W. Browne, R. Cudeck, K.A. Bollen, J.S. Long, Alternative ways of assessing model fit, Sage Focus Ed. 154 (1993) 136.
[31] S. Sankararaman, S. Mahadevan, Distribution type uncertainty due to sparse and imprecise data, Mech. Syst. Signal Process. 37 (1) (2013) 182–198.
[32] K.P. Burnham, D.R. Anderson, Multimodel inference understanding AIC and BIC in model selection, Sociol. Meth. Res. 33 (2) (2004) 261–304.
[33] B. Efron, Bootstrap methods: another look at the jackknife, Ann. Stat. 7 (1) (1979) 1–26.

[34] S. Sankararaman, S. Mahadevan, Likelihood-based representation of epistemic uncertainty due to sparse point data and/or interval data, Reliab. Eng. Syst. Saf. 96 (7) (2011) 814–824.
[35] J. Goodman, J. Weare, Ensemble samplers with affine invariance, Commun. Appl. Math. Comput. Sci. 5 (1) (2010) 65–80.
[36] D. Foreman-Mackey, D.W. Hogg, D. Lang, J. Goodman, emcee: The MCMC hammer, Publ. Astron. Soc. Pac. 125 (925) (2013) 306.
[37] S. Kullback, R.A. Leibler, On information and sufficiency, Ann. Math. Stat. 22 (1) (1951) 79–86.
[38] D.L. Weakliem, A critique of the Bayesian information criterion for model selection, Sociol. Meth. Res. 27 (3) (1999) 359–397.
[39] C.M. Hurvich, C.-L. Tsai, Regression and time series model selection in small samples, Biometrika 76 (2) (1989) 297–307.
[40] C.M. Hurvich, C.-L. Tsai, Model selection for extended quasi-likelihood models in small samples, Biometrics (1995) 1077–1084.
[41] I. Csisz et al, Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten, Publ. Math. Inst. Hungar. Acad. 8 (1963) 95–108.
[42] T. Morimoto, Markov processes and the h-theorem, J. Phys. Soc. Jpn. 18 (3) (1963) 328–331.
[43] S.M. Ali, S.D. Silvey, A general class of coefficients of divergence of one distribution from another, J. Roy. Stat. Soc. Ser. B (Meth.) (1966) 131–142.
[44] E. Hellinger, Neue Begründung der Theorie quadratischer Formen von unendlich vielen Veränderlichen, Journal für die Reine und Angewandte Mathematik 136 (1909) 210–271.
[45] A.L. Gibbs, F.E. Su, On choosing and bounding probability metrics, Int. Stat. Rev. 70 (3) (2002) 419–435.
[46] R. Beran, Minimum Hellinger distance estimates for parametric models, Ann. Stat. (1977) 445–463.
[47] B.G. Lindsay, Efficiency versus robustness: the case for minimum Hellinger distance and related methods, Ann. Stat. (1994) 1081–1114.
[48] A. Owen, Y. Zhou, Safe and effective importance sampling, J. Am. Stat. Assoc. 95 (449) (2000) 135–143.
[49] D. Faulkner, A review of effective plating to be used in the analysis of stiffened plating in bending and compression, Tech. rep., 1973.
[50] C.A. Carlsen, Simplified collapse analysis of stiffened plates, Norw. Maritime Res. 5 (4) (1977).
[51] P.E. Hess, D. Bruchman, I.A. Assakkaf, B.M. Ayyub, Uncertainties in material and geometric strength and load variables, Naval Eng. J. 114 (2) (2002) 139–166.
[52] C.G. Soares, Uncertainty modelling in plate buckling, Struct. Saf. 5 (1) (1988) 17–34.
[53] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, Global Sensitivity Analysis: the Primer, John Wiley & Sons, 2008.
[54] F. Cannavó, Sensitivity analysis for volcanic source modeling quality assessment and model selection, Comput. Geosci. 44 (2012) 52–59.
[55] I.M. Sobol, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Math. Comput. Simul. 55 (1) (2001) 271–280.
[56] A. Saltelli, R. Bolado, An alternative way to compute fourier amplitude sensitivity test (fast), Comput. Stat. Data Anal. 26 (4) (1998) 445–460.