Utah State University

DigitalCommons@USU

A Bayesian Markov Chain Monte Carlo Approach to Uncertainty Quantification

Matthew Isaac
Utah State University

A BAYESIAN MARKOV CHAIN MONTE CARLO APPROACH TO UNCERTAINTY

QUANTIFICATION

by

Matthew Isaac

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Statistics

Approved:

_____   _____
Daniel Coster, Ph.D.   Mark Ewing, Ph.D.
Major Professor   Committee Member

_____   _____
Geordie Richards, Ph.D.   Richard S. Inouye, Ph.D.
Committee Member   Vice Provost for Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2020

ABSTRACT

A Bayesian Markov Chain Monte Carlo Approach to Uncertainty Quantification

by

Matthew Isaac, Master of Science

Utah State University, 2020

Major Professor: Daniel Coster, Ph.D.
Department: Mathematics and Statistics

Uncertainty quantification (UQ) is a methodological framework used frequently in engineering analyses to understand how uncertainty in inputs propagate through a system and lead to uncertainty in the output. An instability was observed in a UQ method proposed by Roy and Oberkampf (2011) when the assignment of aleatory and epistemic uncertainties was perturbed. A modification to this method is proposed. A Bayesian Markov Chain Monte Carlo approach to UQ is offered as an alternative to the Roy and Oberkampf method. This approach allows analysts to incorporate information from various available sources and to update the analysis and results as more information becomes available. An illustrative example involving a simply supported beam is provided as a platform to demonstrate the Bayesian UQ approach and to compare it with the Roy and Oberkampf method. Several options for visualizing the results from the Bayesian UQ approach are explored. Methods for interpreting the results are discussed. This research is expected to produce a new approach to UQ that will be useful to engineers and other practitioners as they quantify the uncertainty in a system, visualize the uncertainty, and interpret the results of that UQ analysis to inform decisions.

(60 pages)

PUBLIC ABSTRACT

A Bayesian Markov Chain Monte Carlo Approach to Uncertainty Quantification

Matthew Isaac

Uncertainty quantification (UQ) is a framework used frequently in engineering analyses to understand how uncertainty in system inputs lead to uncertainty in the system output. An instability is observed in a UQ method proposed by Roy and Oberkampf and a Bayesian Markov Chain Monte Carlo approach to UQ is offered as an alternative. The Bayesian approach allows analysts to incorporate information from various available sources including observed measurements and expert opinion and to update the analysis and results as more information becomes available. An illustrative engineering example is provided as a platform to demonstrate the Bayesian UQ approach and to compare it with the Roy and Oberkampf method. Methods for visualizing and interpreting the results from the Bayesian UQ approach are explored. This research is expected to produce a new approach to UQ that will be useful to engineers and other practitioners as they quantify the uncertainty in a system, visualize the uncertainty, and interpret the results to inform decisions.

ACKNOWLEDGMENTS

I would first like to thank Dr. Dan Coster for helping me to navigate this challenging experience of completing a thesis. I would also like to thank Dr. Mark Ewing and Dr. Geordie Richards for serving on my thesis committee and for the time they invested to offer helpful guidance, insights, and suggestions. A special thanks also needs to be given to Todd Remund, for taking the time to patiently explain concepts and to teach me much of what I understand about this topic. He has provided vital direction during this research process and has been (and continues to be) an invaluable mentor. Thanks also to Jared McRory, for his assistance in the data collection and engineering analysis portions of this research. I would also like to thank my parents and siblings, who have always been an incredible blessing to me. Lastly, I would like to express my love to my wife, Vicki, and my daughter, Zoey, and to thank them for their wonderful support, patience, and love. In short, there are many people for whom I am extremely grateful and to whom I owe much of who I am and what I have been able to accomplish.

Matthew A. Isaac

CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Within the field of engineering, more and more emphasis is being placed on uncertainty quantification (UQ). It is used frequently to help engineers and decision makers understand how uncertainty in the inputs of a system will propagate through that system and produce uncertainty in the output of the system. The understanding gained from UQ analyses can give insight to the system's reliability, risk, need for redesign, and various other consid- erations. UQ is particularly useful in fields where full-scale prototypes, testing, and data collection are difficult, dangerous, or otherwise prohibitive. This has led the design and development phases to be more dependent on analysis (Ewing et al., 2018).

One method for performing UQ, introduced by Roy and Oberkampf (2011), has been observed to have issues with stability and reproducibility (section 3.1.1). Because decisions based on UQ analyses may have high-stakes consequences regarding expenses and safety (particularly in the aerospace industry), we wish to suggest a stabilizing modification to the Roy and Oberkampf (2011) method as well as an alternative Bayesian approach to UQ (sections 3.1.2 and 3.2). The modification to the Roy and Oberkampf (2011) method involves adding a third loop around the existing double-loop structure in which a weighted coin toss determines the aleatory and epistemic uncertainty assignments. However, the new Bayesian UQ approach is preferred over modifying the unstable method. The proposed Bayesian approach uses Markov Chain Monte Carlo (MCMC) methods to incorporate all available information into the analysis. As was previously noted, experimental data may be limited due to the expensive or otherwise difficult nature of data collection. The small amount of available data renders a frequentist approach unfeasible. In cases like this, it becomes important to incorporate all other sources of information into the analysis. Some information may come in the form of observed data; other information may come in the form of expert opinion. A Bayesian approach can naturally incorporate both observed data

and prior beliefs or expert opinion. In addition, previously obtained results can be updated as newly acquired data becomes available; as the available information matures, so does the prediction. Techniques for visualizing and interpreting the results are also explored (section 3.4).

The Bayesian UQ approach described herein is demonstrated with an applied engineering example involving a simply supported beam (chapter 4). This example is not meant to be a rigorous engineering analysis but rather to illustrate the proposed method. Data was collected and used in the analysis to make this more realistic than simply using computer generated data. Following the analysis with the Bayesian UQ approach, an analogous UQ analysis is performed for this same example using the original Roy and Oberkampf (2011) method. The results from these two methods are compared and discussed in chapter 5.1. Lastly, further work and useful extensions relating to the Bayesian UQ approach are discussed (section 5.2).

We hope that this research will provide engineers with a flexible framework to apply a Bayesian approach to UQ to a variety of problems.

CHAPTER 2

LITERATURE REVIEW

2.1 Uncertainty Quantification

Uncertainty quantification (UQ) has been defined as "the science of identifying, quantifying, and reducing uncertainties associated with models, numerical algorithms, experiments, and predicted outcomes or quantities of interest" (Smith, 2013). UQ methods are used to propagate input uncertainty through a system to determine output uncertainty. UQ is a broad field that has applications in many disciplines, including computational fluid dynamics (Maitre and Knio, 2010), hydrology (Anagnostou and Krajewski, 1999), chemical engineering (Najm et al., 2009), structural engineering, (Sankararaman et al., 2011), and aerospace engineering (Ewing et al., 2018; Zhang, 2013). Uncertainty quantification is a field that is under development and is an active area of research (Sullivan, 2015). The demand for further advances in UQ methodology continues to grow due to the increasing use of UQ in analyzing physical systems to ensure proper function, safety, and reliability. UQ has proved to be particularly valuable when experimental data is not available due to the difficult, expensive, or impractical nature of data collection (Jiang and Mahadevan, 2009).

Several types of problems can be undertaken with UQ. Methods to solve inverse problems (Malinverno and Briggs, 2004; Nagel and Sudret, 2015) as well as classical forward problems (Ewing et al., 2018; Roy and Oberkampf, 2011) have been developed.

2.2 Uncertainty

Roy and Oberkampf (2011) state that uncertainty can come from various sources throughout a problem, including:

1. Uncertainty in model input values (material properties, loads, environmental conditions, etc.)

2. Uncertainty of the underlying mathematical model (model form uncertainty)

3. Uncertainty from numerical approximation

Uncertainty has been divided into two main categories: aleatory and epistemic (Der Kiureghian and Ditlevsen, 2009; Ewing et al., 2018; Wang et al., 2009). The word "aleatory" is derived from the Latin word *alea* meaning "game of chance" or "die" (Dezfuli et al., 2009). Thus, aleatory uncertainty relates to uncertainty associated with random events. Aleatory uncertainty has been described as "inherent random behavior" (Ewing et al., 2018) and "the intrinsic randomness of a phenomenon" (Der Kiureghian and Ditlevsen, 2009). The outcomes of aleatory events are generally described using probabilities (Dezfuli et al., 2009). Because aleatory uncertainty is inherent to the system in question, it cannot be reduced through further experimentation or data collection. Epistemic uncertainty has been described as uncertainty that arises from "a lack of knowledge (or data)" and "a lack of knowledge about the system that is conceptually resolvable" (Wang et al., 2009). The word "epistemic" is derived from the Greek word "episteme", meaning "knowledge". Accordingly, epistemic uncertainty relates to "the degree of knowledge of models and their parameters" (Dezfuli et al., 2009). Epistemic uncertainty can be reduced (or eventually eliminated) if more information is acquired (e.g. future experimentation, data collection, or improved modeling). Theoretically, if an infinite amount of data were available, there would be no epistemic uncertainty. In other words, if an infinite amount of data were available parameters could be determined exactly.

It is not always obvious whether a certain source of uncertainty should be categorized as aleatory or epistemic. One variable may be classified correctly as aleatory in one context and epistemic in another (Der Kiureghian and Ditlevsen, 2009; Roy and Oberkampf, 2011). Aleatory uncertainties are generally characterized using a probability distribution, while epistemic uncertainties have historically been characterized with an interval (Ewing et al., 2018; Lockwood et al., 2012). When taking a Bayesian approach, epistemic uncertainty has been characterized as a probability density function that "represents the degree of belief of the true value on the part of the analyst"(Roy and Oberkampf, 2011).

2.3 Probability Elicitation

Probability elicitation is "the practice and methods of encoding the judgments of experts into probabilities or probability distributions" (Hora, 2016). There is a large amount of literature available on the subject of probability elicitation (Cooke, 1991; Hora, 2016; O'Hagan, 2019; Spetzler and Stael Von Holstein, 1975). Having a structured and systematic way to extract expert opinion is an important aspect of conducting uncertainty quantification analyses, particularly in situations where data sets are limited or expensive to obtain. It allows analysts to extract expert knowledge and represent it as a probability distribution. The information obtained from the expert can then be used to improve the quality and relevance of the analysis, thus fully leveraging all available sources of information. This is a particularly useful step when generating informative prior distributions for a Bayesian analysis (O'Hagan, 2019).

Wang et al. (2009) recommends conducting "probability encoding interviews" with subject matter experts. Interview questions are framed in such a way that responses to the questions are essentially points on a cumulative probability distribution. As described in Wang et al. (2009), there are three main methods of conducting the probability encoding interviews:

1. P-methods ask the subject to specify points on the probability scale while the values are fixed.

2. V-methods ask the subject to specify points on the value scale while the probabilities remain fixed.

3. PV-methods ask questions that must be answered on both scales jointly; the subject essentially describes points on the cumulative distribution.

O'Hagan (2019) discusses the importance of specific "protocols" that have been developed to avoid introducing biases (e.g. anchoring, availability, range-frequency, overconfidence) in the elicitation process, and recommends three leading elicitation protocols for extracting opinions from multiple experts. The recommended protocols include the Cooke protocol

(Cooke, 1991), the SHELF protocol (Oakley and O'Hagan, 2019), and the Delphi protocol (Rowe and Wright, 1999).

The Cooke protocol (Cooke, 1991) employs a facilitator to obtain judgements from multiple experts. Some of these judgements relate to the uncertain variables of interest, while other judgements relate to "seed variables". These seed variables are similar to the actual variables, but the true quantity is known to the researcher. If an expert makes good judgements on the seed variables, it is assumed that he will also make good judgements on the variables of interest. Multiple expert opinions are pooled together and are weighted according to the quality of predictions made on the seed variables. This is referred to as "mathematical aggregation", where opinions are collected separately and combined via a mathematical pooling rule (O'Hagan, 2019).

The Sheffield Elicitation Framework, or SHELF, (Oakley, 2019) is designed for use with multiple experts but can be adapted for use with a single expert. Experts first make judgements in private about the variables of interest. After individual private judgements have been made, the experts are gathered and are then shown all individual judgements. The group of experts must then come to an agreement on one "consensus judgement". This type of protocol is referred to as "behavioral aggregation", where a group of experts are asked to come to an agreement (O'Hagan, 2019). According to the SHELF protocol, a facilitator is required for this group discussion. Documentation and an e-learning course exist to train facilitators to conduct effective elicitation (Oakley and O'Hagan, 2019). In addition, a software package implementing SHELF has been created in the R programming language (Oakley, 2019).

Lastly, the Delphi method (Rowe and Wright, 1999) is a combination of the previous two methods, incorporating both mathematical and behavioral aggregation. Experts make individual repeated judgements, with opinions aggregated and shown to experts between rounds. After the last round of expert judgement collection, the final judgements are aggregated mathematically (O'Hagan, 2019).

2.4 Bayesian Analysis

The basis for Bayesian data analysis is Bayes' Theorem. It is stated as follows: For two events, A and B,

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}. \tag{2.1}$$

Bayes' Theorem is frequently applied to "observable" random variables to calculate conditional probabilities of events, as one might encounter in an introductory statistics course. However, the Bayesian statistics framework extends Bayes' Theorem to be used in Bayesian inference for unknown parameters. It is at this point that frequentist and Bayesian views differ. The frequentist approach only allows distributions on data, but the Bayesian approach considers both data and parameters to have distributions (Lunn et al., 2013). In the formulation of Bayes' Theorem given in Eq. (2.2), x represents observed data and θ is a random variable (or vector of random variables) representing the parameters of the distribution of the data.

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} \tag{2.2}$$

$$p(x) = \int p(x|\theta)p(\theta)d\theta \tag{2.3}$$

Thus, $p(\theta)$ represents the prior distribution, or the analyst's knowledge of θ *before* observing the data. On the other hand, $p(\theta|x)$ is the posterior distribution, or the analyst's knowledge of θ *after* observing the data. In other words, the prior knowledge of θ, $p(\theta)$, is updated by the data that has been observed, resulting in the posterior distribution of θ, $p(\theta|x)$. In addition, $p(x|\theta)$ is the sampling density for the data, proportional to the likelihood function, and $p(x)$ (Eq. (2.3)) is the marginal density of the data (Lynch, 2007). Lynch (2007) also points out that since the denominator of Eq. (2.2) only scales the posterior density, Bayes' Theorem can be represented as

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}. \tag{2.4}$$

"Conjugacy" occurs when the prior distribution and the posterior distribution belong to the same family of distributions. Under conditions of conjugacy the analysis to compute the posterior distribution is simplified and the posterior distribution will be one that is well understood. Any subsequent analyses based on the posterior distribution will be straightforward. An analytic evaluation of the posterior distribution is available under conjugacy (Gamerman and Hedibert, 2006).

However, conjugacy is the exception rather than the rule. Gamerman and Hedibert (2006) state that, "In general, the expression of the posterior even in simple models is complex enough to not allow the analytic evaluation". These authors also say that it is usually unlikely that the conjugate prior is a sufficient characterization of the prior state of uncertainty. In cases where conjugacy is not attainable, approximate methods of inference may be used, some of which involve methods based on stochastic sampling. This includes Gibbs sampling and the Metropolis-Hastings algorithm, two popular Markov Chain Monte Carlo methods (section 2.5).

Bayesian inference, if properly performed, "uses all of the available information and leads to better parameter estimates and to better decisions" (Dezfuli et al., 2009).

2.5   Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) algorithms are comprised of a Markov Chain component and a Monte Carlo component. MCMC algorithms were preceded by ordinary Monte Carlo algorithms (Brooks et al., 2011). MCMC is a general algorithm structure into which many simulations can be arranged.

A Markov chain is a sequence of random variables, $X^{(0)}, X^{(1)}, X^{(2)}, ...$, with the $t + 1^{th}$ random variable in the sequence conditional *only* on the random variable immediately preceding it, $X^{(t)}$, and independent from $X^{(t-1)}, ..., X^{(0)}$ (Lunn et al., 2013). That is, the distribution of $X^{(t+1)}$ is given by

$$X^{(t+1)} \sim p_{trans}(x|X^{(t)} = x^{(t)}). \tag{2.5}$$

In Eq. (2.5), p_{trans} is called the transition distribution and defines "the conditional probability of moving to any particular new value given the current value of the chain" (Lunn et al., 2013). The idea of a Markov chain is that as t approaches infinity, the unconditional distribution of $X^{(t+1)}$ reaches a unique stationary distribution. This stationary distribution does not depend on the initial value of the chain, $X^{(0)}$ (Lunn et al., 2013).

As was explained in section 2.4, many cases arise when performing Bayesian inference where the necessary calculations to obtain an analytical solution of the posterior distribution are not feasible. In situations like these, stochastic simulation methods can be used. In these cases, the X random variables in the Markov chain are replaced by θ, a random variable representing the unknown parameter of a distribution. Although these methods cannot find a closed-form solution for the posterior distribution, they can sample from the posterior distribution, $p(\theta|x)$. The Gibbs sampling method (Geman and Geman, 1984) and the Metropolis-Hastings algorithm (Hastings, 1970; Metropolis et al., 1953) are two popular MCMC methods that can draw samples from the posterior distribution. The samples from the posterior distribution can then be used to calculate credible intervals, summary statistics, and various percentiles, all of which provide information that can be used in decision making.

2.6   Roy and Oberkampf Method

The method implemented by Roy and Oberkampf (2011) will be reviewed in detail in this section. This method uses a nested sampling approach to propagate uncertainties through a mathematical model and keeps aleatory and epistemic uncertainties separated throughout the analysis.

In this method a mathematical model is first identified. The output of this model is termed the System Response Quantity (SRQ). The SRQ is the value for which uncertainty is to be quantified. The uncertainty for each model input is classified as being either epistemic or aleatory, and these uncertainties are separately propagated through the model. Aleatory and epistemic uncertainties are defined respectively as uncertainty from inherent randomness and uncertainty from a lack of knowledge of models and parameters (see section 2.2

for a detailed discussion). Roy and Oberkampf (2011) acknowledge that the classification of input quantities as either aleatory or epistemic is not always straightforward or obvious, and in some cases an input could be a mixture of both aleatory and epistemic uncertainty. Within this methodology, aleatory uncertainty is characterized by a probability distribution function (PDF) or a cumulative distribution function (CDF). Epistemic uncertainty is characterized by "an interval with no associated PDF" or by "a PDF which represents [the] degree of belief of the analyst" (Roy and Oberkampf, 2011). Mixed aleatory and epistemic uncertainty is characterized by a distribution where the distributional parameters are not scalars but are described by intervals or distributions. The framework for the Roy and Oberkampf (2011) method includes the following steps:

1. **Identify all sources of uncertainty**

   Potential sources of uncertainty may include model inputs, numerical approximation, and model form.

2. **Characterize uncertainties**

   This step involves "assigning a mathematical structure to describe the uncertainty and determining the numerical values of all the needed parameters of the structure" (Roy and Oberkampf, 2011). Analysts need to first determine whether the source of uncertainty will be classified as aleatory, epistemic, or a mixture of the two. The authors note that information for determining the probability distributions or intervals that characterize the uncertainty may come from experimental data, other models, or expert opinion. However, it is left to the reader to research methods to actually encode information from these sources into some mathematical structure representing the uncertainty.

3. **Estimate uncertainty due to numerical approximations**

   All numerical approximation uncertainty (discretization error, iterative error, round-off error, coding mistakes) is considered epistemic. Uncertainty from unknown coding errors is difficult to quantify, so the authors recommend minimizing errors as much as possible by using good coding practices. Error estimates for the remaining three

types of numerical uncertainty are characterized with intervals and added together to obtain a single interval representing numerical approximation uncertainty.

4. **Propagate input uncertainties throughout the model**

   In this step, if there are both aleatory and epistemic model input uncertainties present, the epistemic and aleatory uncertainties are kept separate. This approach is called double-loop sampling or nested sampling. The authors describe this approach as follows: "For each sample of all of the epistemic uncertainties, the aleatory uncertainties are propagated through the model ... to produce a single CDF of the SRQ. After all of the epistemic samples have been chosen and the resulting CDFs have been computed, one has an ensemble of M CDFs" (Roy and Oberkampf, 2011). The widest points of this ensemble of CDFs constitutes a probability box, or p-box. Alternatively, a certain percentage of the largest and smallest CDFs may be discarded. For example, if the largest 2.5% and the smallest 2.5% of the CDFs were discarded, this would leave the 95% in the center. This has been called a "95% credible p-box" (Ewing et al., 2018). The p-box is used to visualize and interpret the uncertainty associated with the SRQ.

5. **Estimate model form uncertainty through model validation**

   Uncertainty from model form is estimated by model validation. The authors suggest using a validation metric. This metric is ideally computed via a validation experiment, but it is often necessary to use pre-obtained experimental data. The model validation process reveals how the experimental values of the SRQ differ from the model predicted values of the SRQ. Various validation metrics can be selected, but the area validation metric is used by the authors. This measures the area between a CDF from simulation samples (i.e. from the predictive model) and a CDF from experimental measurements.

6. **Determine total uncertainty in the system response quantity**

   The p-box constructed in step 4 is modified by appending the area validation metric to each side of the p-box. This makes the p-box wider, accounting for the uncertainty

associated with model form uncertainty. Because the epistemic and aleatory uncertainty have been kept separate as they were propagated through the model, they each contribute to different attributes of the resulting p-box. The epistemic uncertainty is indicated by the width of the p-box and the aleatory uncertainty is indicated by the range of SRQs that are covered by the p-box.

CHAPTER 3

METHODS

3.1 Roy and Oberkampf Method

In this section, the motivation for a new approach to UQ is given. The instability observed in the Roy and Oberkampf (2011) approach will be demonstrated, which raises the need to either develop a stabilizing modification to the existing method or to offer a new method. One possible stabilization is given in section 3.1.2. The proposed alternative method is given in section 3.2.

3.1.1 Observed Instability

The purpose of the example given here is to demonstrate an instability that has been observed in the method described by Roy and Oberkampf (2011). This instability may not occur in all cases, but the fact that it has been found raises a concern that an instability may exist but remain unknown to the analyst.

A simple example, borrowed from Ewing et al. (2018), is used for this demonstration. A mathematical model (not associated with any physical context) is given as

$$Q = E_1^{E_2} A_3 A_2^{A_1} + E_3, \tag{3.1}$$

where Q is the system response quantity, or quantity of interest. The model inputs E_1, E_2, and E_3 are inputs which we will assume are subject to epistemic uncertainty. We will assume that the remaining model inputs, A_1, A_2, and A_3, are subject to aleatory uncertainty. The characterizations of model input uncertainty are given below.

$$E_1 \sim Uniform(0.5, 0.7)$$

$$E_2 \sim Uniform(0.5, 0.7)$$

$$E_3 \sim Uniform(0.5, 0.7)$$

$$A_1 \sim Normal(2.0, 0.13)$$

$$A_2 \sim Normal(7.0, 0.55)$$

$$A_3 \sim Normal(50, 1.2)$$

Note that a and b represent the lower and upper bounds of a $Uniform(a, b)$ distribution and μ and σ^2 represent the mean and variance of a $Normal(\mu, \sigma^2)$ distribution.

The double loop or nested sampling approach described in section 2.6 was performed to propagate the uncertainty of the model inputs through the model to quantify the uncertainty of Q. Roy and Oberkampf (2011) note that determining whether a model input should be classified as having aleatory or epistemic uncertainty is not always an obvious choice. In other words, there is some ambiguity in these uncertainty class assignments, and different analysts may make different assignments if performing the same analysis independently. In this simulation, the original Uniform or Normal distribution assignments remained the same while epistemic and aleatory assignments were switched one input at a time. As can be seen in Fig. 3.1, changing A_1 or A_2 to epistemic uncertainties widened the p-boxes considerably, while changing E_1 to be an aleatory uncertainty led to a much narrower p-box. Note that the p-boxes resulting from changing A_3 to epistemic and E_2 and E_3 to aleatory aren't shown on the plot as they resulted in essentially the same p-box as the original uncertainty assignments. Because results from UQ analyses are frequently used to make decisions that may have a large impact on the financial well-being of a company or the physical well-being of employees and product users, this instability is an undesirable characteristic of this method.

3.1.2   Proposed Modification

The instability demonstrated in section 3.1.1 occurs when the classification of the uncertainty associated with a model input changes from aleatory to epistemic or vice versa. If there is no disagreement between analysts on the correct classification this may not be

an issue. However, this modification may help if there is difficulty determining whether a model input should be classified as having aleatory or epistemic uncertainty. Instead of framing this classification as binary, analysts might determine a degree of membership, or a weighted membership, to both uncertainty classes. These membership weights should sum to 1. For example, a model input θ_i might be determined to have an aleatory membership weight of 0.7 and an epistemic membership weight of 0.3. A third outer loop could be added around the existing double loop in the Roy and Oberkampf (2011) method (section 2.6). In this third loop, a "weighted coin toss" (may be referred to statistically as a Bernoulli trial) is performed for each model input that has been assigned weighted memberships to each uncertainty class. For a model input, θ_i, with an aleatory membership weight of p and an epistemic weight of q, $P(\text{aleatory}) = p$ and $P(\text{epistemic}) = 1 - p = q$. If the weighted coin toss results in aleatory, the model input is treated as aleatory for the subsequent nested loops. Otherwise, it is treated as epistemic. After the two inner loops are completed, the next iteration of the third outer loop begins, and uncertainty classifications are reassigned given the new outcomes of the weighted coin tosses. If the original method proposed by Roy and Oberkampf is termed "2D" UQ (Ewing et al., 2018), this modification can reasonably be called "3D" UQ.
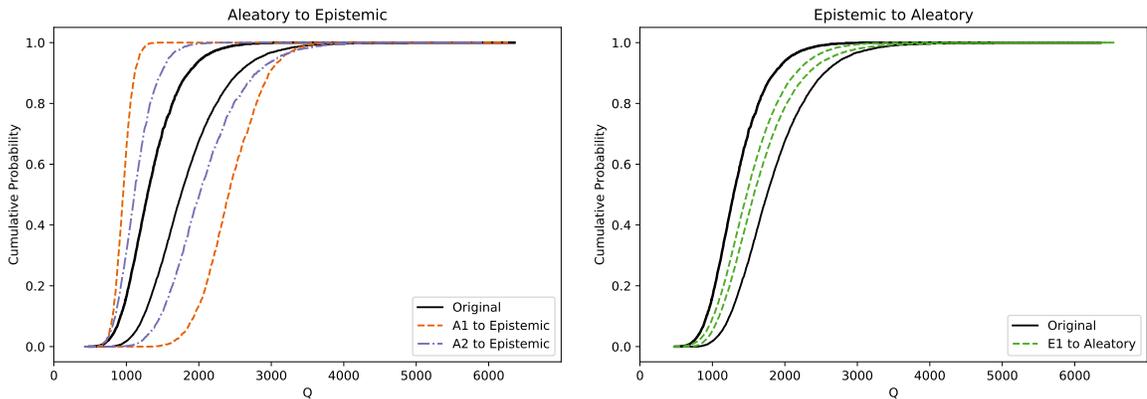

Fig. 3.1: Results from the Roy and Oberkampf method when perturbing the aleatory and epistemic assignments of the model inputs. Changing A_1 and A_2 to epistemic uncertainties resulted in a large change in the resulting p-box, as did changing E_1 to an aleatory uncertainty.

To demonstrate this modification, the example from section 3.1.1 is used. As we saw an instability in the results when changing the uncertainty classification of input A_1 to epistemic and input E_1 to aleatory, we assign membership weights to these inputs. A_1 is given an aleatory membership of $p = 0.7$ and an epistemic membership assignment of $q = 1 - p = 0.3$. In a separate simulation, E_1 is given an aleatory membership of $p = 0.4$ and an epistemic membership assignment of $q = 1 - p = 0.6$. The results are shown in Fig. 3.2. The orange (dashed) trace on the left plot shows the p-box when changing A_1 completely to epistemic uncertainty, resulting in a drastically different p-box from the original aleatory and epistemic assignments. The blue (dash-dotted) trace on the left plot shows that a middle ground is reached when A_1 is given a degree of membership to both the aleatory and epistemic uncertainty classes. While not as dramatic, we see on the right plot that the magenta (dash-dotted) trace finds a middle ground between the original uncertainty class assignments and changing E_1 completely to aleatory.

If the Roy and Oberkampf (2011) method must be used, a modification like this may help to stabilize the results. In the next section, however, we offer a Bayesian approach to UQ as a preferred alternative.
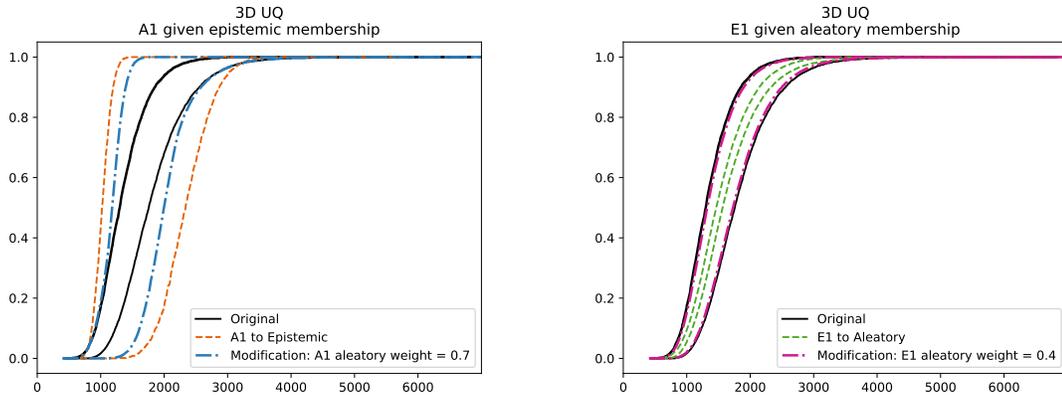

Fig. 3.2: Modification to the Roy and Oberkampf method (3D UQ). The result of the modification (the dash-dotted blue trace and the dash-dotted magenta trace) yields a middle ground between the original aleatory and epistemic uncertainty classifications and changing the uncertainty classification of A_1 or E_1 completely.

3.2 A Bayesian Approach to Uncertainty Quantification

The proposed algorithm for a Bayesian approach to UQ is outlined below. A diagram of the algorithm is given in Fig. 3.3.

1. **Specify the engineering model and quantity of interest**

   The engineering model, f, mathematically relates a vector of system inputs, $\boldsymbol{\Theta} = \begin{bmatrix} \theta_1 & \theta_2 & \cdots & \theta_k \end{bmatrix}^T$ (geometry, material properties, loads, environmental conditions, etc.), to the to the quantity of interest, Q. Note that Q here is analogous to the SRQ in the Roy and Oberkampf (2011) method. We wish to learn something of the uncertainty associated with Q. Thus, the engineering model is of the form

$$f(\boldsymbol{\Theta}) = Q. \tag{3.2}$$

2. **Determine a distribution family for each model input**

   For each input, $\theta_i \in \boldsymbol{\Theta}, i = 1, ..., k$, specify a distribution family, D_i. That is,

$$\theta_1 \sim D_1(\beta_1, \eta_1)$$
$$\theta_2 \sim D_2(\beta_2, \eta_2)$$
$$\vdots$$
$$\theta_i \sim D_i(\beta_i, \eta_i)$$
$$\vdots$$
$$\theta_k \sim D_k(\beta_k, \eta_k).$$

   Without loss of generality, the notation above assumes that each distribution has two parameters. In practice the distributions that are used need not have exactly two parameters. The parameters β_i and η_i are not directly specified in this approach.

3. **For each distribution specified in step 2 determine prior distributions of the distributional parameters**

Each distributional parameter, β_i and η_i, $i = 1, ..., k$ should be given a prior distribution. This prior may be uninformative if necessary, but if possible should be informed by expert opinion or legacy data. Thus, for the parameters of the distribution of θ_i, $i = 1, ..., k$,

$$\beta_i \sim A_i(h_i, g_i) \tag{3.3}$$

$$\eta_i \sim B_i(s_i, t_i). \tag{3.4}$$

The values h_i, g_i, s_i, and t_i are parameters of the prior distributions. In Bayesian inference they would be called hyperparameters; parameters for the distributions of parameters. All hyperparameters must be specified in this step.

4. **Draw samples from the posterior distributions of β_i and η_i using MCMC methods**

   Given some observed data, $\mathbf{x}_i$, where $\mathbf{x}_i$ is a vector of realized observations corresponding to θ_i, use Bayesian MCMC methods (e.g. Metropolis-Hastings) to estimate and sample from the posterior distributions of parameters. Because determining the distribution $p(\beta_i, \eta_i | \mathbf{x}_i)$ is often not analytically tractable, MCMC methods are used to sample from this distribution. From the MCMC sampling, $j = 1, ..., m$ samples are obtained from posterior distributions of β_i and η_i. Note that the m samples do not include the burn-in period for the MCMC simulation. The number of MCMC iterations, m, should be large enough to ensure convergence. Convergence can be assessed using trace plots and other convergence diagnostics. We denote these posterior distribution samples as

$$\boldsymbol{\beta}_i^* = \begin{bmatrix} \beta_i^{(1)*} \\ \beta_i^{(2)*} \\ \vdots \\ \beta_i^{(m)*} \end{bmatrix}, \quad \boldsymbol{\eta}_i^* = \begin{bmatrix} \eta_i^{(1)*} \\ \eta_i^{(2)*} \\ \vdots \\ \eta_i^{(m)*} \end{bmatrix}. \tag{3.5}$$

In other words, we obtain $\beta_i^{(j)*}$ and $\eta_i^{(j)*}$ for $j = 1, ..., m$ and $i = 1, ..., k$.

5. **Sample values from the distributions comprised of pairs of parameters from MCMC simulation**

   Next, $r = 1, ..., n$ samples are taken from the distribution $D_i(\beta_i^{(j)*}, \eta_i^{(j)*})$, for $j = 1, ..., m$ and $i = 1, ..., k$. The value $\theta_i^{(r,j)*}$ is the rth sample from distribution $D_i(\beta_i^{(j)*}, \eta_i^{(j)*})$. If this is performed for each θ_i, $i = 1, ..., k$, we can represent the k samples for each unique (r, j) in a vector:

$$\Theta^{(r,j)*} = \begin{bmatrix} \theta_1^{(r,j)*} & \cdots & \theta_i^{(r,j)*} & \cdots & \theta_k^{(r,j)*} \end{bmatrix}^T. \tag{3.6}$$

6. **Evaluate engineering model over each vector obtained in step 5**

   Because each $\Theta^{(r,j)*}$ contains a value for each of the k model inputs, $f(\Theta^{(r,j)*}) = Q^{(r,j)*}$ can be evaluated. This should be done for all $\Theta^{(r,j)*}$, $r = 1, ..., n$ and $j = 1, ..., m$. The resulting $Q^{(r,j)*}$ values can be compiled into an $n \times m$ matrix like the one shown in Eq. (3.7)

$$\mathbf{Q}^* = \begin{bmatrix} Q^{(1,1)*} & Q^{(1,2)*} & \cdots & Q^{(1,j)*} & \cdots & Q^{(1,m)*} \\ Q^{(2,1)*} & Q^{(2,2)*} & \cdots & Q^{(2,j)*} & \cdots & Q^{(2,m)*} \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ Q^{(r,1)*} & Q^{(r,2)*} & \cdots & Q^{(r,j)*} & \cdots & Q^{(r,m)*} \\ \vdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ Q^{(n,1)*} & Q^{(n,2)*} & \cdots & Q^{(n,j)*} & \cdots & Q^{(n,m)*} \end{bmatrix} \tag{3.7}$$

7. **Visualize and interpret results**

   The m columns of $\mathbf{Q}^*$ correspond to the m samples from the posterior distributions of β_i and η_i for $i = 1, ..., k$. Stated another way, each column of $\mathbf{Q}^*$ represents samples from a different set of potential parameters for distributions D_i, $i = 1, ..., k$. Empirical CDFs can be calculated for each column in $\mathbf{Q}^*$, and p-boxes constructed

and plotted. A more detailed discussion about visualizing and interpreting results is given in section 3.4.

8. **Update analysis as more data becomes available:** As more data becomes available, the new information that this data contains should be incorporated into the UQ analysis. This can be done by rerunning the MCMC simulations with both the newly available data, $\mathbf{x}_{i(new)}$, and the previously available data, $\mathbf{x}_i$. Note that if new data is only available for one model input, it is only necessary to re-run the MCMC simulation for that model input. In other words, new samples only need to be taken from posterior distributions of the model inputs for which new data has been incorporated. After new samples are obtained from the posterior distribution, $p(\beta_i, \eta_i | \mathbf{x}_{i(new)}, \mathbf{x}_i)$, perform step 5, inserting newly obtained samples into the $\mathbf{\Theta}^{(r,j)*}$ vectors. Next, perform step 6 again, resulting in an updated $\mathbf{Q}^*$. Visualizations can then be recreated and interpretations revised.

3.3 Accounting for Aleatory and Epistemic Uncertainties

First, we define three terms: deterministic, epistemic, and aleatory. Epistemic and aleatory uncertainties are also discussed in section 2.2, but are briefly revisited here due to their importance in this section. After definitions are given, the role of these three terms in the Bayesian UQ method will be explained.

Deterministic is defined as, "pertaining to exactly predictable (or precise) events, the outcome of which is known with certainty if the inputs are known with certainty" (Dezfuli et al., 2009). Models such as $F = ma$ or $E = mc^2$ would be classified as deterministic models. Next, aleatory is defined as, "pertaining to stochastic (non-deterministic) events, the outcome of which is described by a probability" (Dezfuli et al., 2009). Anything that is inherently random contains aleatory uncertainty. Lastly, epistemic is defined as, "pertaining to the degree of knowledge of models and their parameters" (Dezfuli et al., 2009). Imprecise knowledge of model parameters is epistemic uncertainty.

21
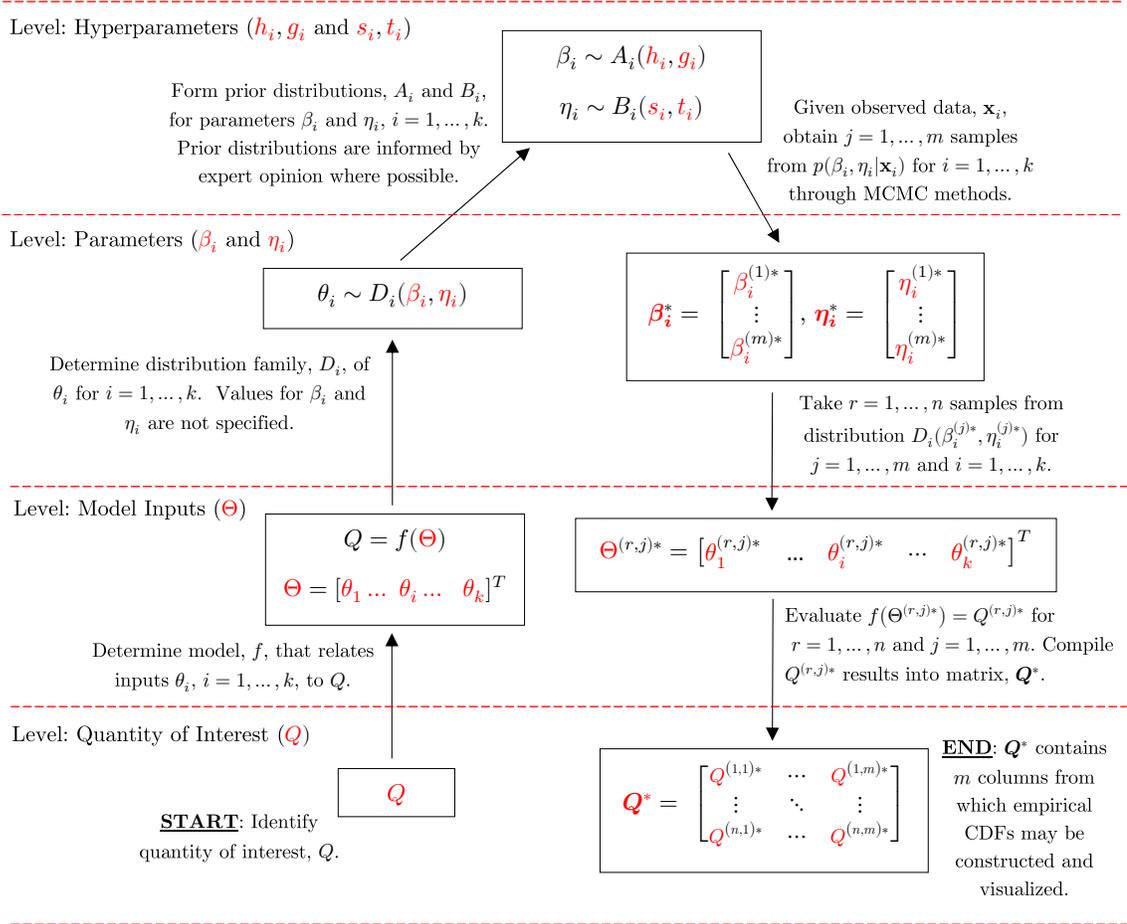
Fig. 3.3: Diagram of the Bayesian UQ algorithm. This diagram supplements the written description of the Bayesian UQ method. It emphasizes how the algorithm starts at the level of the quantity of interest and works up to the hyperparameter level and back down again. The algorithm begins in the bottom left corner at **START**.

We assume that the model relating the model inputs to the quantity of interest is deterministic. However, we usually do not know the precise values of all model inputs. Thus, even though our model is deterministic, we are uncertain of the precise value of the model output (quantity of interest). The uncertainty must be characterized for each model input, θ_i. A distribution family, D_i, is placed on each θ_i. This probability distribution represents the aleatory uncertainty associated with model input θ_i. However, a probability distribution is not complete without specifying the distributional parameters. Without loss of generality, we assume that distribution D_i has two parameters, β_i and η_i. As was

noted previously in this section, epistemic uncertainty stems from a lack of knowledge of parameters. Thus, if the distributional parameters are not known precisely, epistemic uncertainty is introduced and needs to be accounted for. This is done by placing prior distributions on the distributional parameters, allowing them to vary according to our prior beliefs. As new data is obtained and incorporated, the Bayesian framework updates these prior distributions, thus reducing the epistemic uncertainty. In this way, uncertainty for a model input with some amount of both aleatory and epistemic uncertainty may be characterized.

Much of the time, model inputs may have both epistemic and aleatory uncertainty associated with them. However, there may be instances where a model input can be determined to have purely aleatory or purely epistemic uncertainty. A purely aleatory model input would still have a distributional family assigned. However, since no epistemic uncertainty is present, constants would be used for β_i and η_i, and no prior distributions used for these distributional parameters. Consequently, observed data would not impact the uncertainty of this model input since aleatory uncertainty is not reducible. A purely epistemic model input would involve setting the distribution family, D_i to a single parameter, β_i. This parameter would have a prior distribution placed on it. The prior distribution would be updated with observed data in the Bayesian framework, thus reducing the epistemic uncertainty.

3.4   Visualizing and Interpreting Results

Useful and informative visualizations are at the core of interpreting results from this Bayesian approach to UQ. Although the underlying methods that generate $\mathbf{Q}^*$ are different than the Roy and Oberkampf (2011) method, the results of both methods have similar structures. The resulting matrices from both methods contain columns from which empirical CDFs can be constructed. The resulting ensemble of empirical CDFs can be visualized with a p-box (section 2.6). The p-boxes used in this and following sections are 95% p-boxes. That is, the p-box contains the middle 95% of the ensemble of CDFs.

3.4.1 Visualizations

Here, we suggest several possible ways to enhance the p-box plots described by Roy and Oberkampf (2011) and Ewing et al. (2018). Plotting the p-box alone can be informative, but further insight may be gained by visualizing the underlying ensemble of CDFs. However, simply plotting each CDF in the ensemble is not a satisfactory solution. First, as the number of CDFs (the number of columns, m, in $\mathbf{Q}^*$) increases, the plotting area quickly becomes over-saturated with plot traces, resulting in a solid band of CDFs, which gives no insight about how the CDF density varies over the plotting area. In data visualization, this is referred to as overplotting. Usually, overplotting can be handled by using narrower lines and allowing the lines to be semi-transparent. This would allow the CDFs to "stack" on top of each other and areas with a higher density of CDFs will appear darker than areas of the plot with a lighter density. This may help initially, but will still be inadequate if there is a large number of CDFs to be plotted. In addition, as the dimensions of the $\mathbf{Q}^*$ matrix increase (as n and m increase), plotting each empirical CDF in the ensemble becomes quite computationally expensive and generating a single plot can take a significant amount of time.

There are ways that the density of the CDFs can be visualized without plotting each individual CDF. Two possible visualization techniques are heatmap plots and hexagon-binning plots. Both of these plots map the density of points over the plotting area to a color scale. A challenge is encountered when using these kinds of plots to visualize the ensemble of CDFs. Because every CDF in the ensemble begins at a cumulative probability of 0 and ends at a probability of 1, there is an extremely high point density at the bottom left and top right corners of the plot. These high density areas wash out much of the structure throughout the rest of the plot. In order to better visualize the density over the entire plot, it is helpful to use a transformation on the density color scale. A square root transformation, $\log_2$ transformation, and $\log_{10}$ transformation are all effective in improving the plot. Examples of both the heatmap and the hexagon-binning plots are shown in Fig. 3.4 along with the impact of a square root transformation. These plots, along with all subsequent plots, were

generated using the ggplot2 package (Wickham, 2016). The p-box may also be overlaid
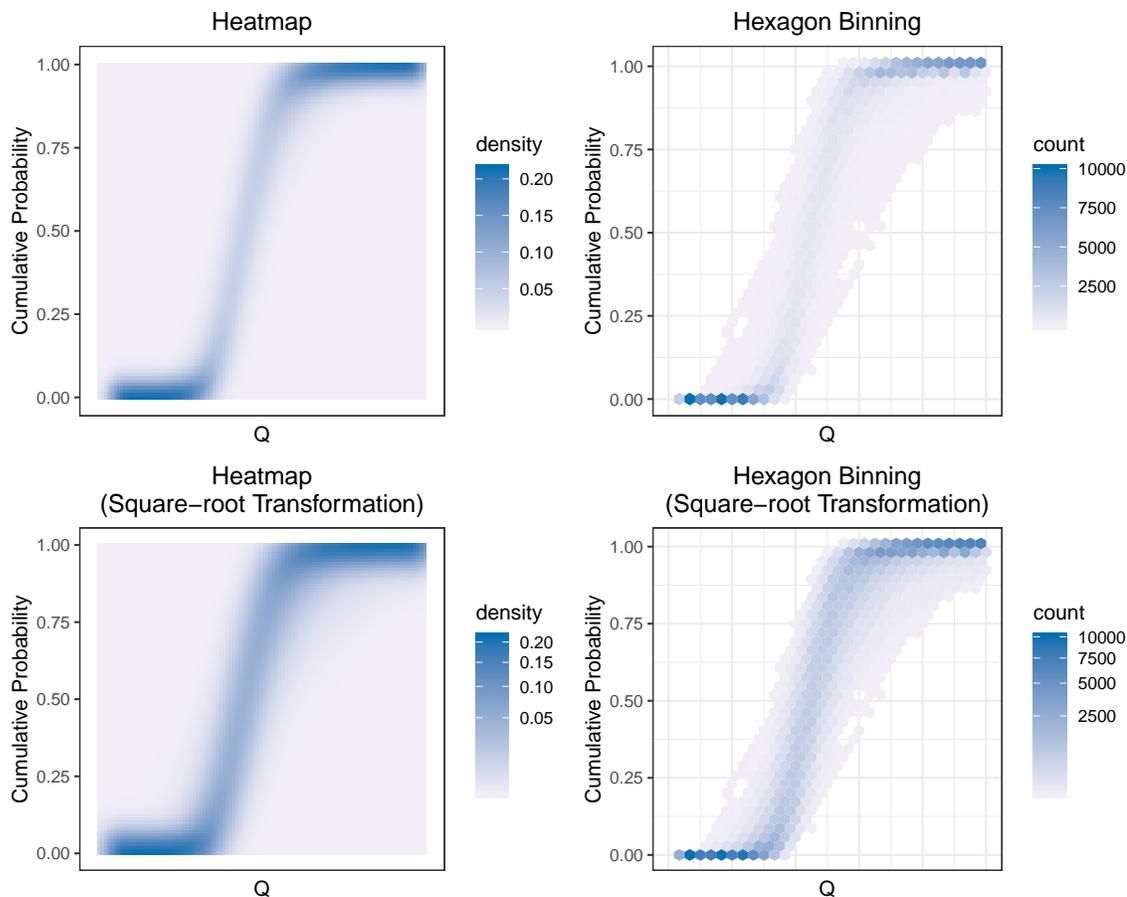
Fig. 3.4: Heatmap and hexagon-binning plots. These plots demonstrate the use of heatmap and hexagon-binning plots to visualize the underlying ensemble of CDFs that are obtained from the Bayesian UQ method.

on the heatmap or hexagon-binning plot to assist in interpretability. An example of this is given in Fig. 3.5. Incorporating both the CDF density (through a heatmap or hexagon-binning plot) and a p-box on the same graphic results in a more informative visualization than either plot on its own. Another possible method for visualizing the density of CDFs is to plot the p-box on one plot and then, on an accompanying plot, show a histogram of the probability values at a given value of Q. The ensemble of CDFs is essentially "sliced" vertically, and the distribution of the probabilities at that value are shown in the histogram. An example of this is given in Fig. 3.6.
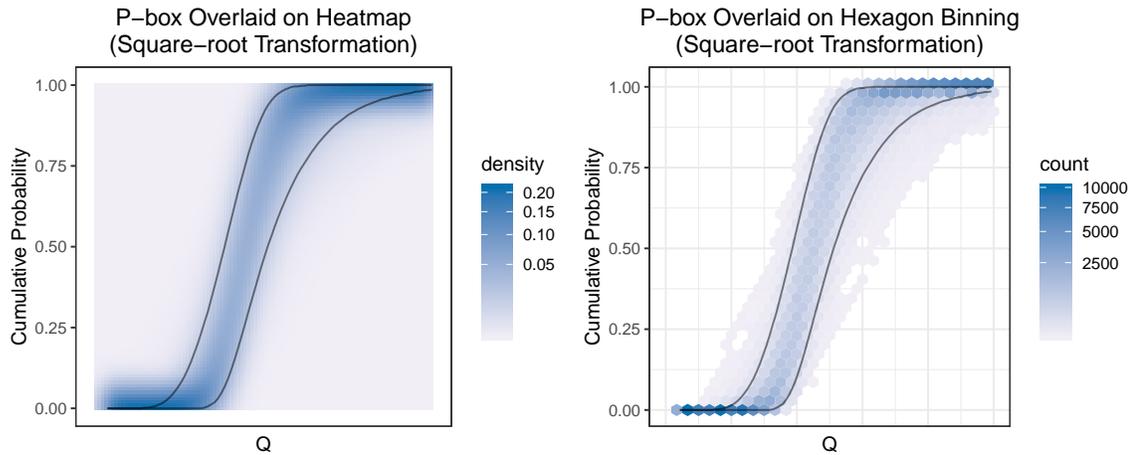
Fig. 3.5: Heatmap and hexagon-binning plots with an overlaid p-box. This plot demonstrates how a combination of a heatmap or hexagon-binning plot with an overlaid p-box is more informative than either plot on its own.
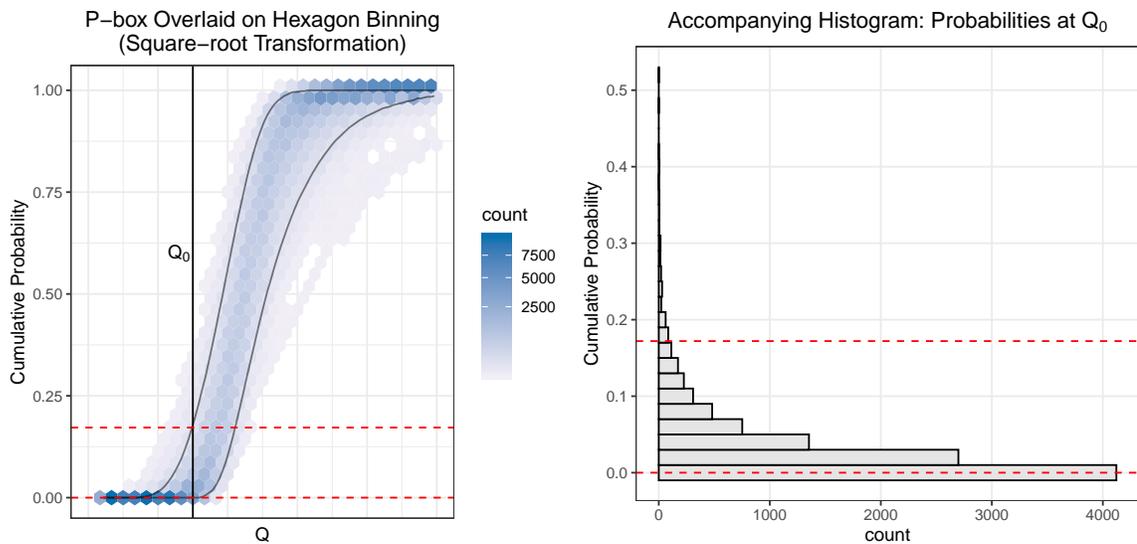
Fig. 3.6: P-box and accompanying histogram. The distribution shown in the histogram (right) corresponds to the values in the CDF ensemble at the vertical solid line shown on the p-box (left). The histogram helps visualize the distribution of probabilities at a selected value of Q.

Both the heatmap and the hexagon-binning plots yield similar visualizations, but the hexagon-binning plots had a render time that was about one-third that of the heatmap plots (Fig. 3.7). It was observed that randomly sampling a certain number of columns from $\mathbf{Q}^*$ and plotting only those resulted in a heatmap with a similar relative density as would
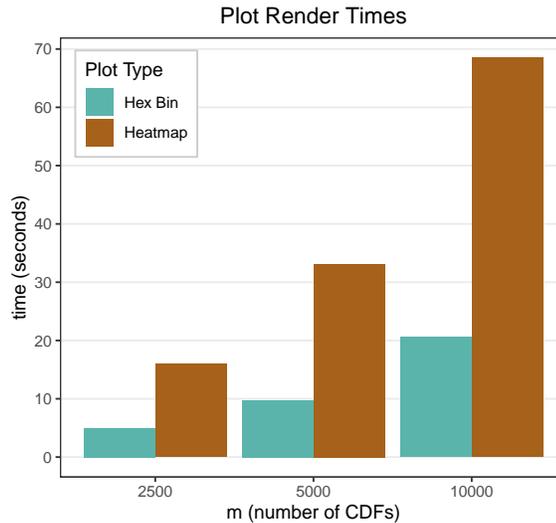
Fig. 3.7: Render times of heatmap and hexagon-binning plots. The hexagon-binning plots had a render time that was about one-third that of the heatmap plots.

be obtained when plotting all the CDFs in an ensemble. Figure 3.8 shows a heatmap plot where 1,000 CDFs were randomly sampled and used to generate the plot and a plot where all 10,000 CDFs were used. When the number of CDFs was changed from 10,000 to 1,000, the plot rendering time dropped from 68.53 seconds to 6.62 seconds (a difference of almost 62 seconds). The differences between the two plots are subtle, and it is likely that similar interpretations would result from both plots. However, if time and computing power permit, it is recommended to plot all the data.

3.4.2 Interpretation

The visualizations described in the section above will aid greatly in understanding and interpreting results of the UQ analysis. The p-box, heatmap, or hexagon-binning plots alone give a rough picture of the UQ results. It allows the analyst to see approximately where the body of CDFs lie in relation to a value of Q that is of particular interest for the current application. It will also allow the analyst to get a feel for how much uncertainty is present in the quantity of interest. High uncertainty corresponds to a horizontally wide p-box, while low uncertainty corresponds to a narrow p-box.

In order to get a more detailed understanding of the UQ results, the visualization
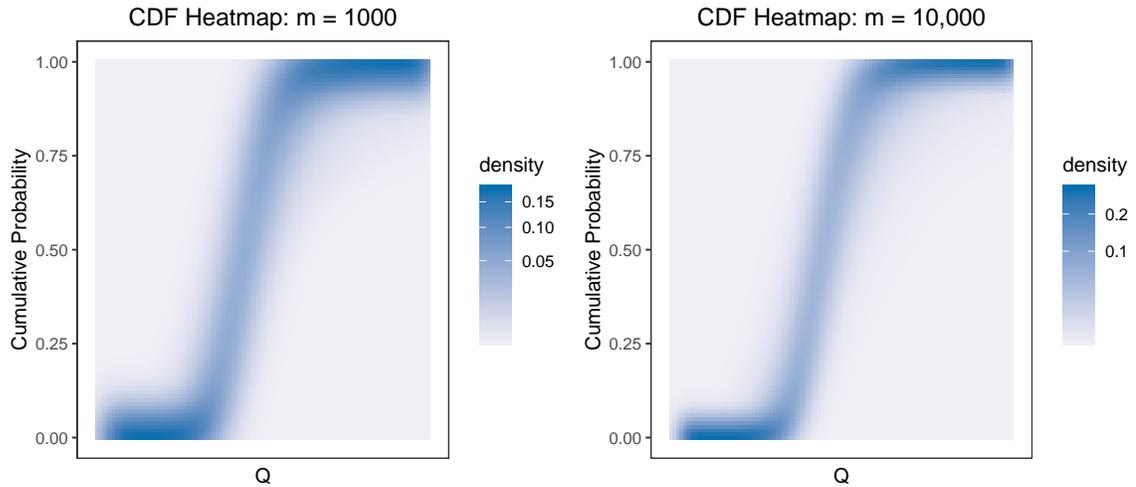
Fig. 3.8: Heatmap plot with $m = 1,000$ and $m = 10,000$. The relative density remains approximately the same between the two plots, but the rendering time decreases from 68.53 seconds to 6.62 seconds when $m = 1,000$, a difference of almost 62 seconds.

technique using an accompanying histogram (Fig. 3.6) will be more useful than a heatmap or p-box plot alone. The analyst should "slice" the p-box vertically at the value(s) of Q that is of particular interest for the current application. The values at which the vertical line intersects the p-box should be reported. This interval can be interpreted as a credible interval. For example, if Q is desired to be greater than some value Q_0, and the interval resulting from slicing the p-box vertically at $Q = Q_0$ is $(0, 0.005)$, we could say that, with 95% credibility, the probability that Q is less than Q_0 is between 0 and 0.005 (between 0 and 0.5%). Those not familiar with statistics and UQ (oftentimes the decision makers) may find that this approach does not immediately make clear the correct decision or course of action. The 95% credible interval of probabilities may be confusing and not as explicit and straightforward as desired. Thus, more research and work remain to develop methods for creating detailed interpretations for results from the Bayesian UQ approach. This is further discussed in section 5.2.1.

CHAPTER 4

APPLICATION

An illustrative example will now be given to demonstrate the Bayesian approach to UQ as described in section 3. Note that the example given in this section is not meant to be a rigorous engineering analysis. Rather, it is meant to demonstrate the utility of the Bayesian UQ approach and to provide a platform for comparison between this approach and the Roy and Oberkampf (2011) method. The calculations for this example were performed in the R (R Core Team, 2019) and Python (PYT, 2018) programming languages, and visualizations created with the ggplot2 (Wickham, 2016) package.

4.1 Simply Supported Beam - Engineering Background

A simply supported beam, as shown in Fig. 4.1, has a pin support at one end and a roller support at the other. Simply supported beams with point loads are used in various applications, such as beams in bridges, buildings, and beds of machine tools (Bird and Chivers, 1993).
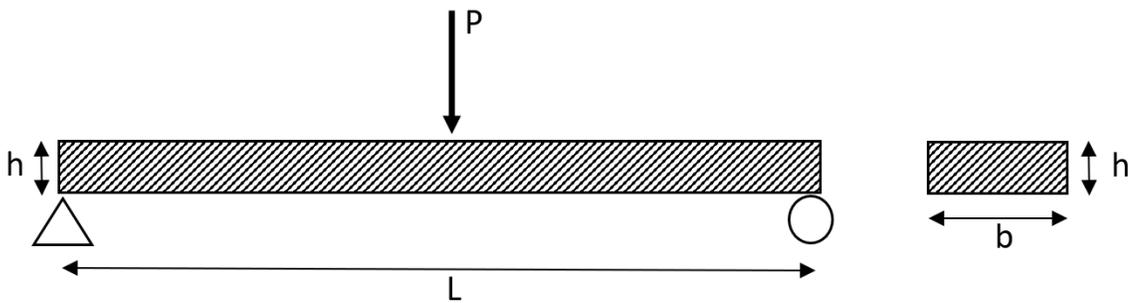
Fig. 4.1: Diagram of a simply supported beam. This diagram shows a simply supported beam with a midspan point load.

An assessment of the design of a simply supported beam might include an evaluation of the safety factor, SF. As shown in Eq. 4.1, the safety factor is a ratio of the strength of

the beam, S, to the burden of the beam, B.

$$SF = \frac{S}{B} \tag{4.1}$$

A safety factor greater than 1 indicates that the strength is greater than the burden, and the beam will be capable of supporting the burden. A safety factor less than 1 indicates that the burden surpasses the strength of the beam and the beam will fail under the burden.

For a wooden beam, the strength, S, is given by the modulus of rupture, MOR. The burden is given by

$$B = \frac{My}{I}, \tag{4.2}$$

where M is the maximum moment, y is the distance from neutral axis to the extreme tensile fiber, and I is the moment of inertia. This gives a safety factor of

$$SF = \frac{S}{B} = \frac{S}{\frac{My}{I}}. \tag{4.3}$$

L, b, and h represent the length, width, and height of the beam respectively, and P is the point load applied to the midspan of the beam. For a simply supported beam experiencing three point flexure (simply supported beam with a center point load) and a rectangular cross section, we have Eq. (4.4) - (4.6).

$$M = \frac{PL}{4} \tag{4.4}$$

$$y = \frac{h}{2} \tag{4.5}$$

$$I = \frac{bh^3}{12} \tag{4.6}$$

Substituting these into the safety factor equation gives us the following:

$$SF = \frac{S}{\frac{My}{I}}$$

$$SF = S\frac{I}{My}$$

$$SF = S\frac{\frac{bh^3}{12}}{\frac{PL}{4}\frac{h}{2}}$$

$$SF = S\frac{8bh^3}{12PLh}$$

$$SF = S\frac{2bh^2}{3PL}.$$

Thus, the engineering model is

$$SF = f(\mathbf{\Theta}) = S\frac{2bh^2}{3PL}, \tag{4.7}$$

$$\text{where } \mathbf{\Theta} = \begin{bmatrix} h & b & L & S & P \end{bmatrix}^T.$$

For this illustrative example, assume that a simply supported beam needs to be designed and installed in predetermined locations to support existing fixtures. The beams are to be constructed from Douglas Fir Larch lumber that is approximately 1.5 inches (3.81 cm) in height and 3.5 inches (8.89 cm) wide (commonly referred to as a "2 × 4"). The beam will need to span approximately 15 inches (50.8 cm). For standard grade Douglas Fir-Larch lumber, the modulus of rupture is tabulated by the American Wood Council (NDS, 2015) to be 575 psi (3964.49 kPa). The engineers have determined that a beam with a span of 15 inches should be able to support the expected load of about 150 lbs (68.04 kg). The calculations below support this conclusion.

$$SF = S\frac{2bh^2}{3PL}$$

$$SF = 575\frac{2(3.5)(1.5)^2}{3(150)(15)}$$

$$SF = 575\frac{15.75}{6750}$$

$$SF = 1.34$$

Thus, with a safety factor of $1.34 > 1$, the burden does not exceed the strength.

4.2   Data Collection

Computer-generated data could have been used to demonstrate the use of the Bayesian UQ method. For demonstration purposes measurements were taken on physical beams to give a realistic example of how these methods might be used in engineering applications.

Nine 2 ft (60.96 cm) Doug Fir-Larch 2×4 beams were obtained from a local lumber distributor. They were marked as "std & btr" or "standard and better" grade lumber, meaning that all lumber was guaranteed to be at least standard grade and some percentage of the lumber may be a higher grade (e.g. No. 2 or No. 1 grades).

Measurements of height (h) and width (b) were taken on these 9 specimens (Fig. 4.2) with digital calipers.

Fig. 4.2: Taking height and width measurements. Measurements were taken for height (left) and width (right) using digital calipers.

The strength (modulus of rupture) was measured on all 9 specimens. This was done

Fig. 4.3: Compression-tension universal testing machine. This machine was used to apply a midspan point load to measure the modulus of rupture.

by using a tension-compression universal testing machine (Fig. 4.3). Each specimen was placed on supports with a 20 inch (50.8 cm) span and a monotonically increasing point load was applied at midspan. The load was increased at a rate of 0.1 inches (2.54 mm) per minute until the beam ruptured (Fig. 4.5 and 4.6). Figure 4.4 shows four examples of strength tests results. The plot traces indicate the increasing load over time, and the sharp drop in load indicates when the beam ruptured. The peak load was recorded and the MOR was calculated with Eq. (4.8) and recorded.

$$MOR = \frac{3PL}{2bh^2} \tag{4.8}$$

Although the beam is being designed to span specific gaps in existing locations and to bear the weight of existing fixtures, we assume that in this scenario no measurements have been obtained for either the gap span or the weight of the fixtures. This will allow us to

demonstrate the use of posterior predictive distributions in the Bayesian UQ approach to handle situations where no measurements have been observed for one or more parameters.

Strength Test of Beam Segments
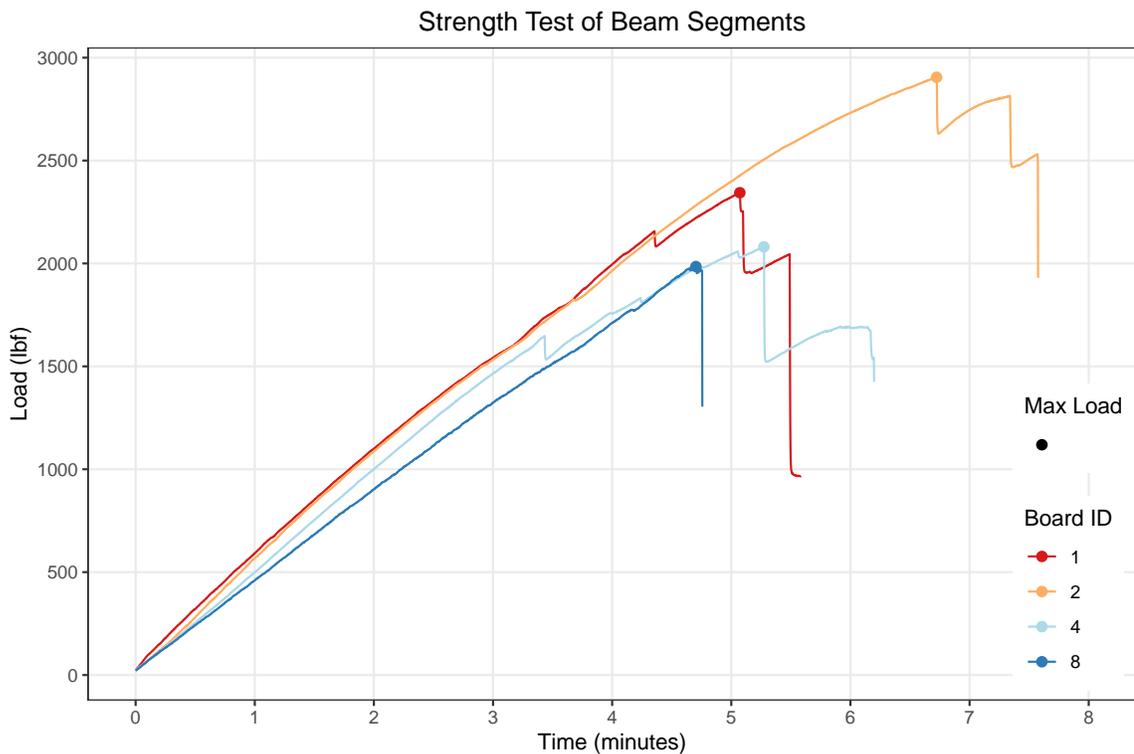
Fig. 4.4: Plot of the load over time during strength testing. The sharp drop in load indicates the time and load at which the beam ruptured. The point on each trace shows the maximum load.

4.3   UQ Analysis

In the next two sections, a UQ analysis will be performed using both the Bayesian UQ approach and the original Roy and Oberkampf (2011) approach. Results from these approaches will be compared in chapter 5.

4.3.1   Roy and Oberkampf Method

Only model input uncertainty is considered for this application of the Roy and Oberkampf (2011) method because the Bayesian UQ method has only been developed to the point of

(a) Start of strength test

(b) Partially completed strength test

(c) End of strength test (beam has ruptured)
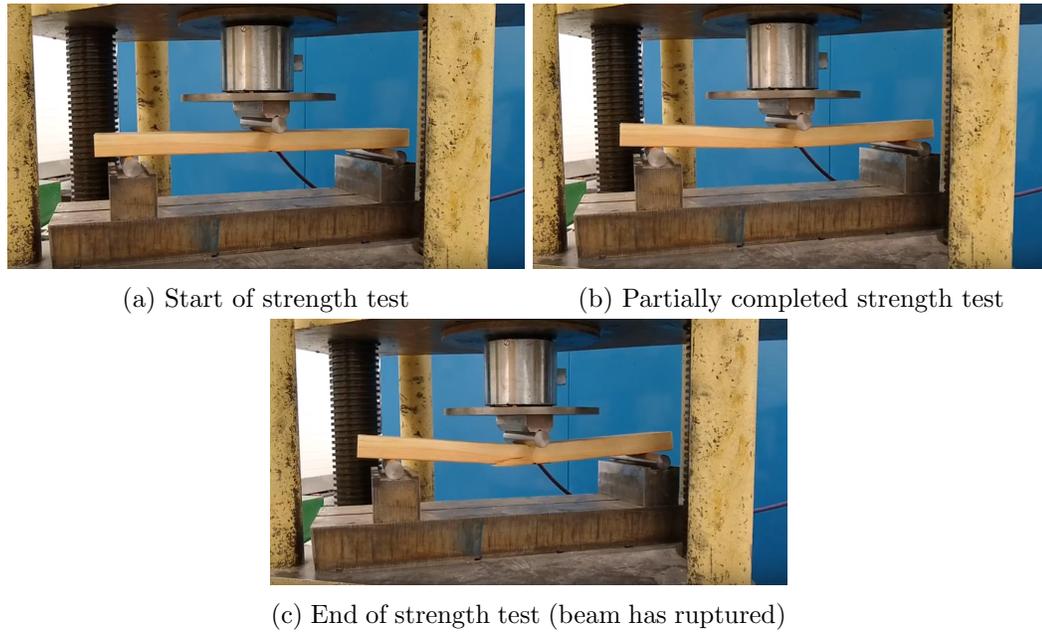
Fig. 4.5: Taking strength measurements for the simply supported beam. These pictures show the beam before a load has been applied (a), the beam flexing under the applied load (b), and the beam after it has ruptured (c).

accounting for these uncertainties. Model form and numerical approximation uncertainty will not be considered here. In order to keep the analyses performed by the two methods as comparable as possible, the same probability distribution families were assigned to the model inputs for both this method and the Bayesian UQ method. Distribution parameters were informed by the available experimental data. Note that when applying the Roy and Oberkampf method, observed data does not have as natural of a place to feed into the method as it does with the Bayesian approach. The authors state that "information for characterizing input uncertainties typically derives from experimentally measured data from the actual system or similar systems of interest" (Roy and Oberkampf, 2011). Thus, after fitting the respective distributions to the available data, the uncertainty characterizations shown below were constructed.

$$b \sim Normal(3.5, 0.02)$$

$$h \sim Normal(1.5, 0.01)$$
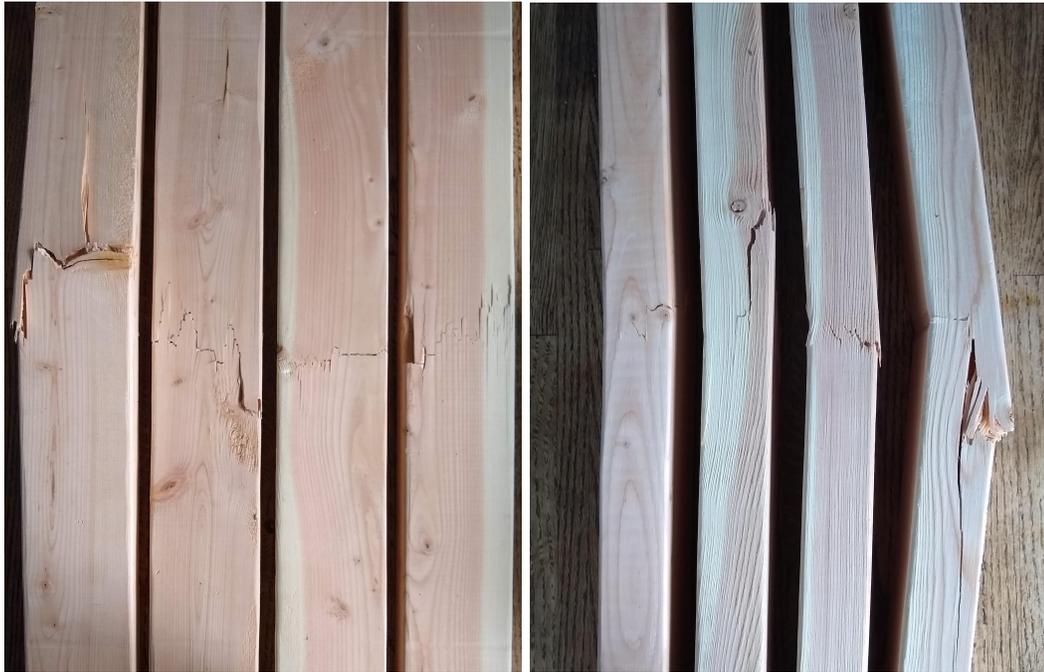
Fig. 4.6: Ruptured beams after strength testing. Four of the nine tested beams are shown after they ruptured during the strength testing. The beams are shown from bottom (left) and from the side (right).

$$L \sim Normal(15, 0.3)$$

$$S \sim LogNormal(9, 0.19)$$

$$P \sim Uniform(147.9, 152.1)$$

Uncertainties relating to b, h, and S were characterized as aleatory. In other words, the width and height of the 2×4's used to construct a beam have naturally occurring variability. Additionally, the strength varies randomly from beam to beam. Additional data collection may help us better understand the probability distributions of these variables, but will not decrease our uncertainty. Uncertainties relating to L and P were characterized as epistemic uncertainties. In this scenario the beams are being designed to be installed in predetermined locations implying that there are predetermined spanning distances. In addition, one of several existing fixtures may be supported by the beam. The approximate spanning distances and fixture weight (corresponding to inputs L and P) have not been
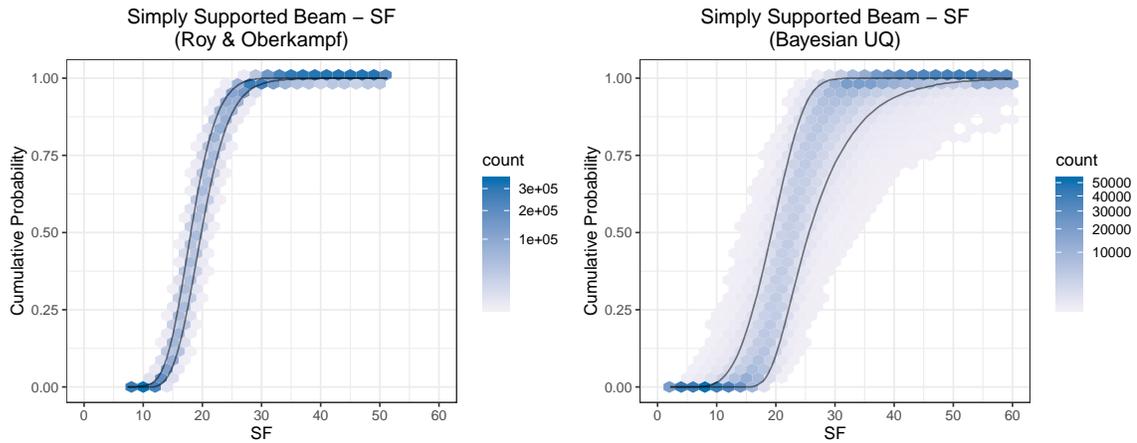
Fig. 4.7: Resulting p-boxes from the UQ analyses on the simply supported beam example. Results are shown from both the Roy and Oberkampf method (left) and results from the Bayesian UQ approach (right).

measured yet, but if or when they are measured, uncertainty regarding these variables can be reduced. Roy and Oberkampf (2011) characterize epistemic uncertainty by assigning an interval with no associated probability. They also state that "the Bayesian approach to uncertainty quantification characterizes epistemic uncertainty as a PDF that represents the degree of belief of the true value on the part of the analyst." For this application to the simply supported beam example, we allow some flexibility to the method described by Roy and Oberkampf (2011) and assign a PDF to each of the epistemic uncertainties.

The Roy and Oberkampf (2011) double-loop method (described in section 2.6) is used to propagate the uncertainty through the engineering model given in Eq. (4.7). 15,000 outer and 500 inner loops are performed. Empirical CDFs are then constructed from the results and summarized using a p-pox containing the middle 95% of the resulting CDFs. Figure 4.7 shows the p-box overlaid on a hexagon-binning plot. Similar to the results from the Bayesian UQ approach, all simulated safety factor values are much greater than 1. In this case, no safety factors less than 7 were calculated. A more detailed interpretation and discussion is given in chapter 5.

4.3.2   Bayesian UQ

Figure 4.8 outlines the Bayesian UQ approach as pertaining to this example. It is a diagram of the same form as Fig. 3.3, but has been applied to the simply supported beam application.

Level: Hyperparameters $(1.5, 0.09, \dots, 152, 0.1)$

Form prior distributions, for all parameters $\mu_h, \sigma_h^2, \dots, a$, and b. Prior distributions are informed by expert opinion where possible.

$$\mu_h \sim Normal(1.5, 0.09)$$
$$\sigma_h^2 \sim Gamma(0.0001, 0.0001)$$
$$\vdots$$
$$a \sim Normal(148, 0.1)$$
$$b \sim Normal(152, 0.1)$$

Given observed data, $\mathbf{x}_h, \mathbf{x}_b, \mathbf{x}_S$, obtain $j = 1, \dots, 15{,}000$ samples from $p(\mu_h, \sigma_h^2 | \mathbf{x}_h)$ through MCMC methods. Sample directly from prior distributions for parameters of L and P since no data is available.

Level: Parameters $(\mu_h, \sigma_h^2, \dots a, b)$

Determine distribution family of h, b, L, S, and P. Parameter values are not specified.

$$h \sim Normal(\mu_h, \sigma_h^2)$$
$$b \sim Normal(\mu_b, \sigma_b^2)$$
$$L \sim Normal(\mu_L, \sigma_L^2)$$
$$S \sim LogNormal(\mu_S, \sigma_S^2)$$
$$P \sim Uniform(a, b)$$

$$\boldsymbol{\mu_h^*} = \begin{bmatrix} \mu_h^{(1)*} \\ \vdots \\ \mu_h^{(15000)*} \end{bmatrix}, \boldsymbol{\sigma_h^{2^*}} = \begin{bmatrix} \sigma_h^{2(1)*} \\ \vdots \\ \sigma_h^{2(15000)*} \end{bmatrix}$$

Take $r = 1, \dots, 500$ samples from distribution $Normal(\mu_h^{(1)*}, \sigma_h^{2(1)*})$ for $j = 1, \dots, 15{,}000$. Perform same action for samples from the parameters of b, L, S, and P.

(Note that only samples for parameters of h are shown above.)

Level: Model Inputs (Θ)

Determine model, f, that relates inputs, h, b, L, S, and P to SF.

$$SF = f(\Theta) = S\frac{2bh^2}{3PL}$$
$$\Theta = [h \;\; b \;\; L \;\; S \;\; P]^T$$

$$\Theta^{(r,j)*} = [h^{(r,j)*} \quad b^{(r,j)*} \quad L^{(r,j)*} \quad S^{(r,j)*} \quad P^{(r,j)*}]^T$$

Evaluate $f(\Theta^{(r,j)*}) = SF^{(r,j)*}$ for $r = 1, \dots, 500$ and $j = 1, \dots, 15{,}000$. Compile $SF^{(r,j)*}$ results into matrix, $\boldsymbol{SF^*}$.

Level: Quantity of Interest (SF)

START: Identify quantity of interest. Here, it is the safety factor, SF.

$$SF$$

$$\boldsymbol{SF^*} = \begin{bmatrix} SF^{(1,1)*} & \cdots & SF^{(1,15000)*} \\ \vdots & \ddots & \vdots \\ SF^{(500,1)*} & \cdots & SF^{(500,15000)*} \end{bmatrix}$$

END: $\boldsymbol{SF^*}$ contains 15,000 columns from which empirical CDFs may be constructed and visualized.
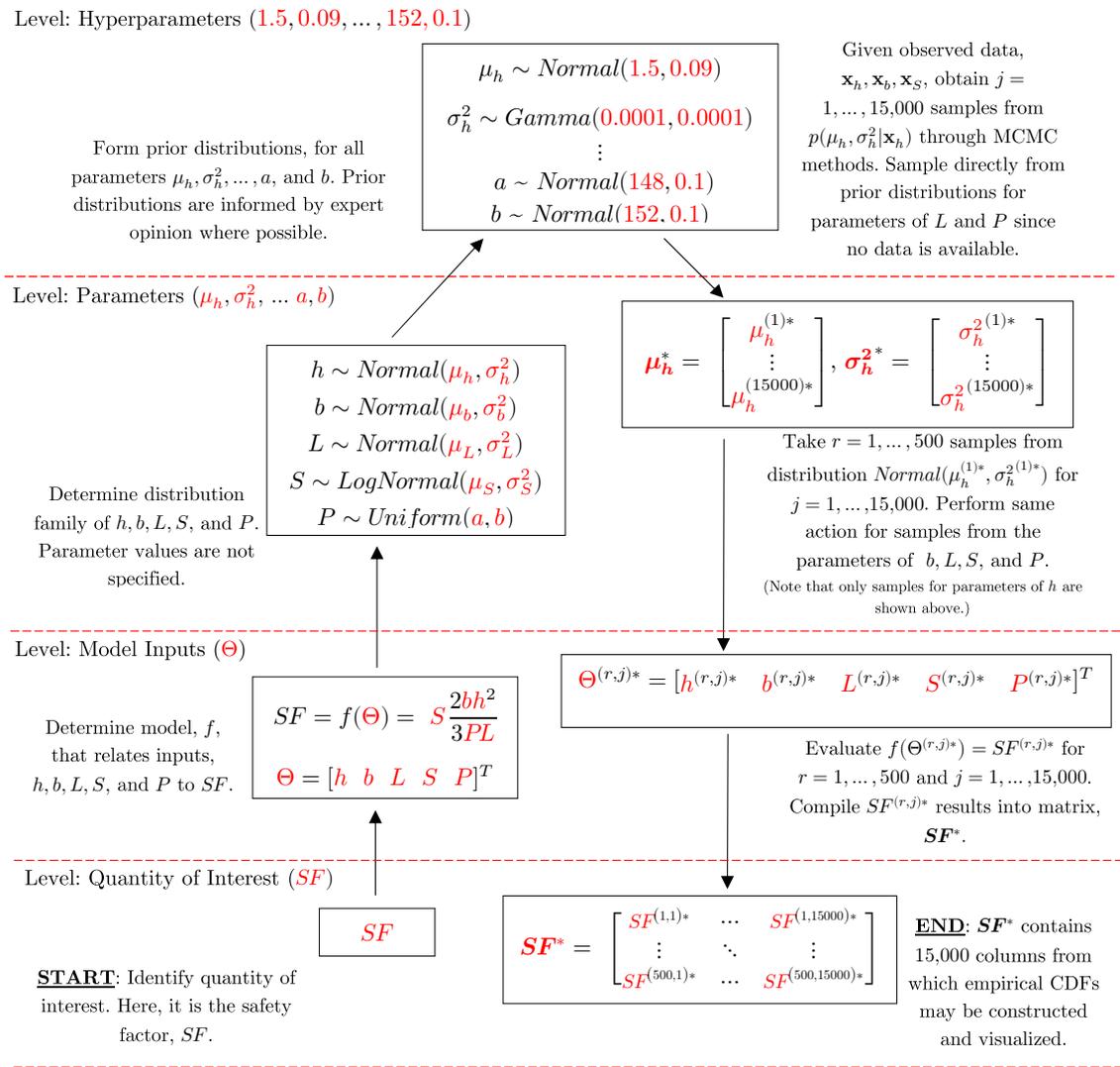
Fig. 4.8: Diagram of the Bayesian UQ algorithm applied to the simply supported beam example. The algorithm begins in the bottom left corner at **START**.

Step 1 of the algorithm is to specify the engineering model and value of interest. The engineering model, shown in Eq. (4.7), has been specified, and the safety factor, SF, is our

value of interest. In other words, we wish to learn about the uncertainty associated with SF.

Next, each model input must be given a distribution (step 2). The following distribution families are assigned to each of the model inputs:

$$h \sim Normal(\mu_h, \sigma_h^2)$$
$$b \sim Normal(\mu_b, \sigma_b^2)$$
$$L \sim Normal(\mu_L, \sigma_L^2)$$
$$S \sim LogNormal(\mu_S, \sigma_S^2)$$
$$P \sim Uniform(a, b)$$

Prior distributions are next assigned to each parameter in the above distributions (step 3). Expert opinion elicitation was not performed to inform these priors, but ideally some of the methods described in section 2.3 would be used to generate the values used in these prior distributions. Some of these prior distributions were chosen based on some well-known values (e.g. average height and width of a 2×4, expected point load) and were made uninformative otherwise (e.g. variance of beam strength).

$$\mu_h \sim Normal(1.5, 0.09)$$
$$\sigma_h \sim Gamma(0.0001, 0.0001)$$
$$\mu_b \sim Normal(3.5, 0.09)$$
$$\sigma_b \sim Gamma(0.0001, 0.0001)$$
$$\mu_L \sim Normal(15, 0.01)$$
$$\sigma_L \sim Gamma(0.0001, 0.0001)$$
$$\mu_S \sim Normal(7.2, 0.01)$$
$$\sigma_S^2 \sim Gamma(0.0001, 0.0001)$$

$$a \sim Normal(148, 0.1)$$

$$b \sim Normal(152, 0.1)$$

Given the measurements taken on the height, width, and strength ($\mathbf{x}_h, \mathbf{x}_b$, and $\mathbf{x}_S$ respectively), Bayesian MCMC methods are employed to sample from the following posterior distributions (step 4):

$$p(\mu_h, \sigma_h^2 | \mathbf{x}_h) = \frac{f(\mathbf{x}_h | \mu_h, \sigma_h^2) p(\mu_h, \sigma_h^2)}{\iint f(\mathbf{x}_h | \mu_h, \sigma_h^2) p(\mu_h, \sigma_h^2) d\mu_h d\sigma_h^2},$$

$$p(\mu_b, \sigma_b^2 | \mathbf{x}_b) = \frac{f(\mathbf{x}_b | \mu_b, \sigma_b^2) p(\mu_b, \sigma_b^2)}{\iint f(\mathbf{x}_b | \mu_b, \sigma_b^2) p(\mu_b, \sigma_b^2) d\mu_b d\sigma_b^2},$$

$$p(\mu_S, \sigma_S | \mathbf{x}_S) = \frac{f(\mathbf{x}_S | \mu_S, \sigma_S) p(\mu_S, \sigma_S)}{\iint f(\mathbf{x}_S | \mu_S, \sigma_S) p(\mu_S, \sigma_S) d\mu_S d\sigma_S}.$$

The Bayesian MCMC sampling was performed via the OpenBUGS software, run in R through the R2OpenBUGS package (Lunn et al., 2009; R Core Team, 2019; Sturtz et al., 2017). 10,000 iterations were run over 3 chains for a total of $t = 30,000$ iterations. The "burn-in" length (number of iterations to discard from the beginning) is $t/2 = 15,000$. This leaves us with $m = 15,000$ samples from each posterior distribution (above).

Convergence is confirmed via trace plots, shown in Fig. 4.9. We now have samples from the posterior distributions of our model input parameters.

Recall that data were not available for P or L. This is a scenario that is likely to be encountered in practice when certain measurements are difficult or expensive to obtain. A prior predictive distribution is used to obtain values of these model inputs to be used in this analysis. This involves making $m = 15,000$ draws from the prior distributions of a and b (the parameters of P) and the prior distributions of μ_L and σ_L (the parameters of L). We add the draws from these prior distributions to the posterior distribution draws obtained from the MCMC simulation.

Step 5 is performed next. $r = 1, ..., 500$ draws are sampled from the distributions created by each pair of parameters, $j = 1, ..., 15,000$. These values are compiled into
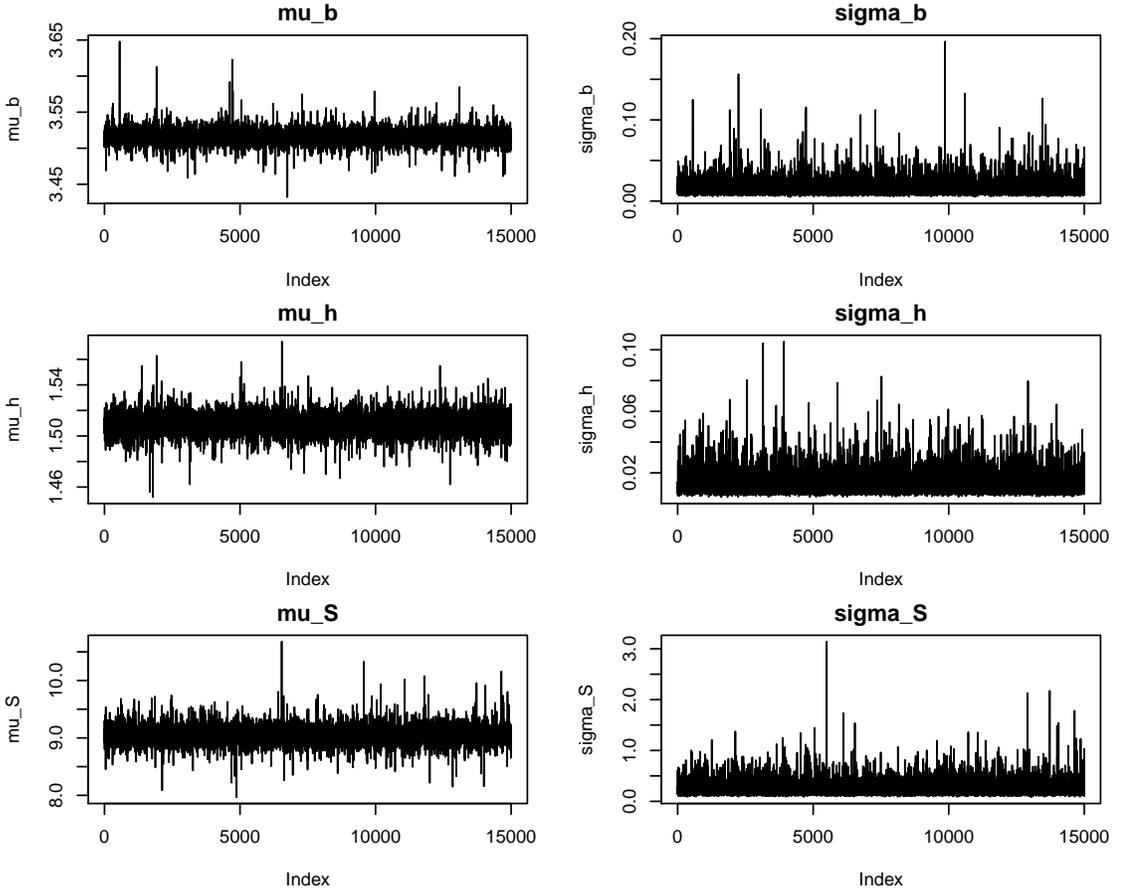
Fig. 4.9: Trace plots from the MCMC simulation for the simply supported beam example. These plots are used to determine if the MCMC algorithm has converged to the posterior distribution.

vectors of the following form:

$$\Theta^{(r,j)*} = \begin{bmatrix} h^{(r,j)*} & b^{(r,j)*} & L^{(r,j)*} & S^{(r,j)*} & P^{(r,j)*} \end{bmatrix}^T, \quad (4.9)$$

for $r = 1, ..., 500$ and $j = 1, ..., 15,000$. (4.10)

Finally, step 6 consists of calculating $f(\theta^{(r,j)*}) = SF^{(r,j)*}$ for $r = 1, ..., 500$ and $j = 1, ..., 15,000$. The results are compiled into a $500 \times 15,000$ matrix, $\mathbf{SF}^*$:

$$\mathbf{SF}^* = \begin{bmatrix} SF^{(1,1)*} & SF^{(1,2)*} & \cdots & SF^{(1,j)*} & \cdots & SF^{(1,15000)*} \\ SF^{(2,1)*} & SF^{(2,2)*} & \cdots & SF^{(2,j)*} & \cdots & SF^{(2,15000)*} \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ SF^{(r,1)*} & SF^{(r,2)*} & \cdots & SF^{(r,j)*} & \cdots & SF^{(r,15000)*} \\ \vdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ SF^{(500,1)*} & SF^{(500,2)*} & \cdots & SF^{(500,j)*} & \cdots & SF^{(500,15000)*} \end{bmatrix}. \tag{4.11}$$

An empirical CDF was calculated for each of the 15,000 columns of $\mathbf{SF}^*$. The resulting ensemble of CDFs was used to construct a 95% p-box. The CDFs and the p-box are then visualized according to the visualization techniques described in section 3.4 (Fig. 4.7). A preliminary inspection of the p-box and underlying CDFs indicate that the safety factors resulting from the UQ analysis were much greater than 1. In fact, the probability that a safety factor less than 5 is obtained appears to be approximately 0. A more detailed interpretation will be given in chapter 5.

Updating With New Observations

Here, we demonstrate how the UQ analysis can be updated if more measurements for a model input become available. In practice, it is likely that as more time passes additional data will be collected. If this data is added to the original data available for that model input and MCMC simulations are repeated for this parameter, the remainder of the Bayesian UQ method can be re-run using the samples from updated posterior distributions (see step 8 in section 3.2). For the simply supported beam example, 9 new observations of the beam strength, S, were simulated. The UQ analysis was then updated as has been described. Visualizations of the updated results compared with the original results are given in Fig. 4.10 and Fig. 4.11. More data results in less uncertainty about distributional parameters, which leads to an overall decrease in uncertainty in the safety factor. This is

seen by the decrease in spread of the ensemble of CDFs and a narrower p-box when these 9 additional strength observations were added.
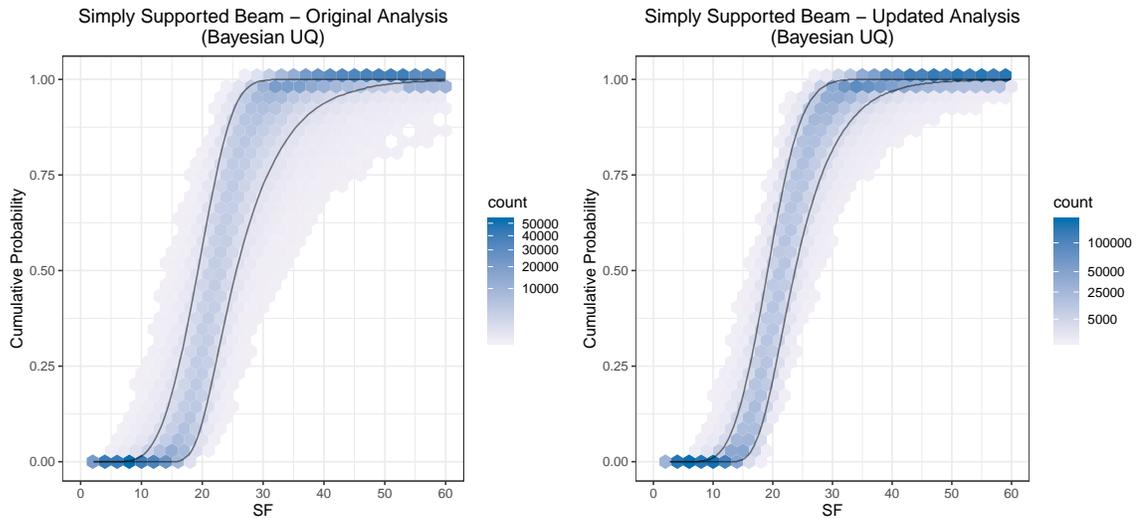

Fig. 4.10: Results from the original and updated Bayesian UQ analysis. The p-boxes and hexagon binning plots in this figure show the impact that an additional 9 strength observations had on the UQ analysis. The original results are shown on the left and the updated results (after the additional 9 strength observations were incorporated) are shown on the right.
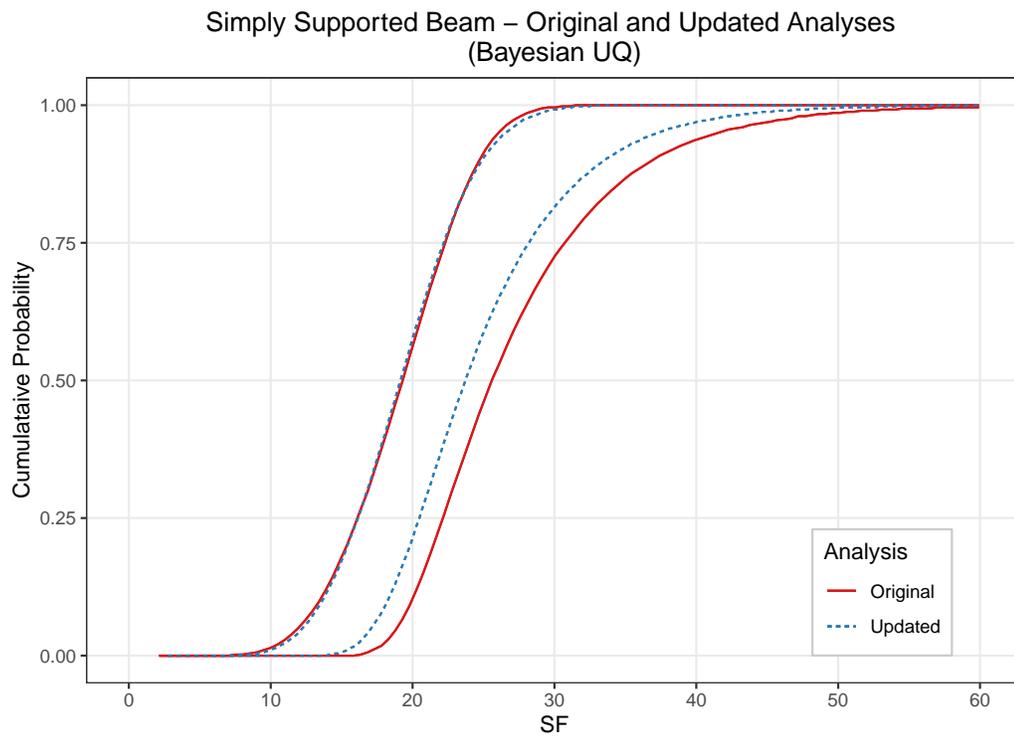
Fig. 4.11: Overlaid *p*-boxes from the original and updated Bayesian UQ analysis. The narrow *p*-box from the updated analysis shows that less uncertainty exists as the amount of available data increases.

CHAPTER 5

DISCUSSION

5.1   Comparison of Bayesian and Roy and Oberkampf Results

Comparison between the Bayesian UQ and Roy and Oberkampf (2011) results is rather natural because the results have a similar structure and can be summarized and visualized in similar ways. First, we inspect the resulting p-boxes, shown in Fig. 5.1. The most striking difference between the two is that the p-box from the Roy and Oberkampf (2011) method is much narrower than the p-box from the Bayesian UQ method. This indicates that a lower amount of uncertainty is represented in the Roy and Oberkampf (2011) method. In addition, the CDFs from the Roy and Oberkampf (2011) method are slightly steeper, which also indicates less uncertainty. We also notice that the Roy and Oberkampf p-box is centered around lower values of SF than the Bayesian UQ p-box.

We hypothesize that the increased variability in the Bayesian UQ method comes from the prior distributions that are placed on the parameters of the distributions of the model inputs. The Roy and Oberkampf (2011) method does not place distributions on the parameters of model input distributions. Instead, constants are used for model input distribution parameters which implies that these values are known precisely. Although observed data does inform the distributions that are used in the double-loop method used by Roy and Oberkampf, the assumption cannot be made that we know all of these parameters precisely. Thus, there is some uncertainty unaccounted for in the Roy and Oberkampf (2011) method. The wider p-box generated through the Bayesian approach automatically takes into account the uncertainty in the model input distributional parameters.

Due to the strength of the beams being much higher than anticipated, the resulting interpretation and decisions from these two methods would likely be quite similar. Both support the conclusion that the existing beam design is adequate for the intended use. In
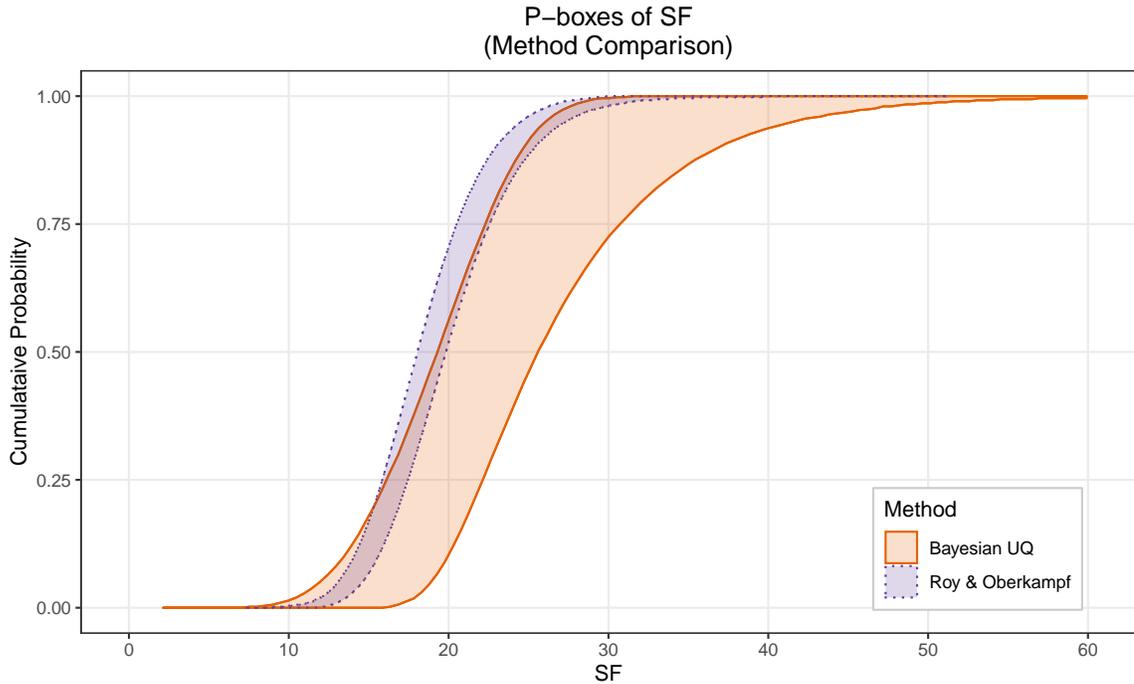
Fig. 5.1: Comparison of p-boxes from the Bayesian UQ and the Roy and Oberkampf methods. The most striking feature is the much narrower p-box resulting from the Roy and Oberkampf method.

fact, the safety factors obtained from both methods are so high that we might consider the proposed beam to be over designed. If it was made from expensive materials, engineers might choose to redesign the beam so the strength of the beam is more appropriate for the expected application.

In order to demonstrate an interpretation of these results, despite the safety factor being much greater than 1, let us assume that the engineers designing this beam are interested in the probability of the safety factor being less than 10. The graphics in Fig. 5.3 and Fig. 5.2 were generated to aid in this interpretation. For the Bayesian UQ approach, the vertical line at $SF = 10$ intersects the p-box at a probability level of 0 and 0.014. These values define a 95% credible interval for the specified safety factor value. We might interpret this interval by saying, "With 95% credibility, the probability that the safety factor is less than 10 is between 0 and 0.014 (0 and 1.4%)." For the Roy and Oberkampf (2011) method, a vertical slice of the p-box yields an interval of 0 and 0.002. This might be interpreted similarly

by saying, "With 95% credibility, the probability that the safety factor is less than 10 is between 0 and 0.002 (0 and 0.2%)." Given these intervals and interpretations, we again see by the wider interval that the Bayesian UQ method predicts more uncertainty than the Roy and Oberkampf (2011) method. We believe that the Bayesian UQ method more accurately represents the true amount of uncertainty present by not assuming model input distributional parameters to be known precisely and by naturally incorporating aleatory and epistemic uncertainties in the algorithm.
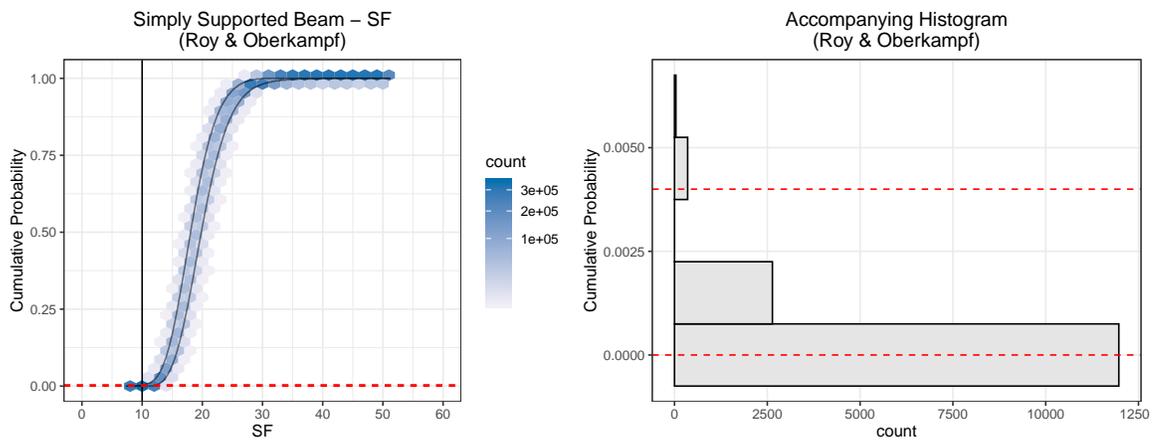

Fig. 5.2: *P*-box and accompanying histogram from the Roy and Oberkampf method on the simply supported beam example. The histogram shows the distribution of probabilities at a safety factor of 10. The horizontal red dashed lines are drawn at the same probability values on both plots. Note that the y-axis scales for cumulative probability differ between the two plots.

5.2   Future Work

While the results from the proposed Bayesian UQ method are promising, there are still extensions that will need to be added to improve the utility and versatility. Some of the future developments are discussed in the following sections.
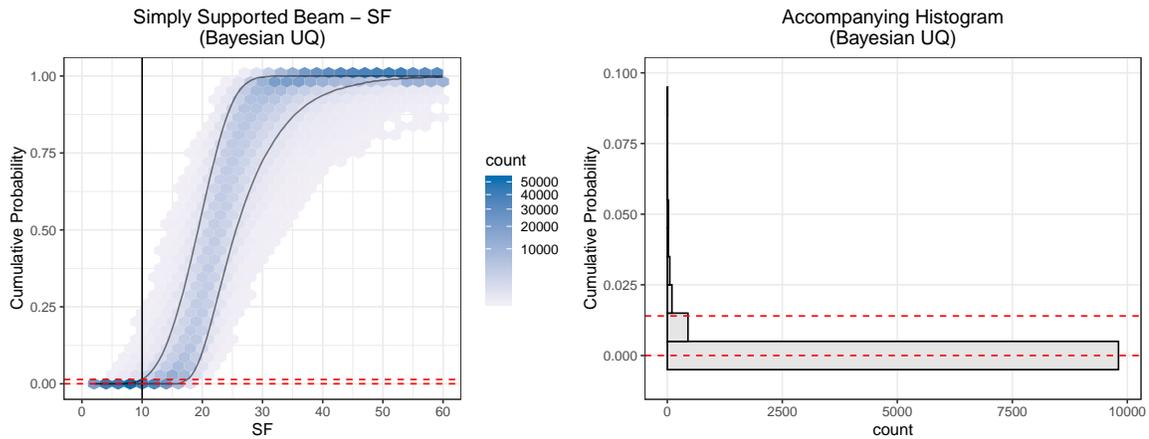
5.2.1   Interpretation

Fig. 5.3: P-box and accompanying histogram from the Bayesian UQ approach on the simply supported beam example. The histogram shown the distribution of probabilities at a safety factor of 10. The horizontal red dashed lines are drawn at the same probability values on both plots. Note that the y-axis scales for cumulative probability differ between the two plots.

More research is needed to refine the interpretation process of the Bayesian UQ approach. If this method is to be successfully used in industry, the final results and interpretation of the method must be clear and unambiguous, even to those who are not familiar with statistical methods, Bayesian theory, or UQ techniques. Decision-makers must be able to use the results from the UQ analysis to reach an agreement on future actions. It is not immediately clear how to use a credible interval obtained by slicing the p-box vertically to make decisions. Thus, future work will involve summarizing the results of Bayesian UQ differently, condensing the results in a way that more easily informs decisions. This may also involve a localized secondary analysis with higher resolution in a particular area of interest. For example, if a safety factor of 10 was an important threshold, a secondary analysis might be performed that would give a more detailed depiction of the uncertainty in a small region around $SF = 10$.

5.2.2 Model Averaging

As described in section 3.2 in step 2, the analyst must assign a distribution, D_i to each model input, θ_i for $i = 1, ..., k$. Cases may arise where multiple distributions may

fit existing data equally well and it may not be clear which distribution is best suited to characterize a particular model input. The choice of distribution family D_i, corresponding to model input θ_i, introduces additional epistemic uncertainty because the analyst likely lacks knowledge of the true distribution family. Thus, instead of using a single D_i, it may be useful to use some criteria, such as the Akaike Information Criteria (AIC) to determine which probability distribution best fits the data. Burnham and Anderson (2002) describe the use of AIC to calculate AIC Differences and eventually Akaike Weights for candidate distributions. These weights can be thought of as "evidence in favor of [a model] being the actual ... best model for the situation at hand *given* that one of the [candidate] models must be the ... best model of that set of models" (Burnham and Anderson, 2002). If the Akaike Weights indicate that one distribution is a clearly the best choice, then that single distribution may be used to describe the data. However, if no single candidate distribution seems adequate alone, then a model average weighted by the Akaike Weights may be used. We believe that these concepts may be useful when integrated into the proposed Bayesian UQ algorithm, particularly in the selection of D_i for a given θ_i.

5.2.3   Expert Opinion Elicitation

This topic was discussed in section 2.3, and it was shown that many approaches exist for eliciting or extracting expert opinion and representing it in a mathematical way (e.g. a probability distribution). Thus, the future work that is required in this area is not necessarily to produce new techniques for eliciting expert opinion, but rather to continue researching existing methods to determine the strengths and weaknesses of the various methods. Some methods may be more appropriate for certain circumstances, and it will be useful to understand the advantages and disadvantages of using a particular method. Eliciting expert opinion in a systematic way is an important component of the Bayesian UQ method as it will assist analysts in constructing informative prior distributions that leverage the valuable existing knowledge as much as possible. If done correctly, this will lead to a more informative UQ analysis, resulting in more informed decisions.

5.2.4   Theory

As noted in section 5.1, the Bayesian UQ method estimates more uncertainty than the Roy and Oberkampf (2011) method on the simply supported beam application. This was indicated by the much wider p-box resulting from the Bayesian UQ. However, as shown in Fig. 4.11, after adding 9 additional strength measurements, the uncertainty was decreased resulting in a narrower p-box. This suggests that the more data we incorporate in the analysis, the narrower our posterior distributions become. This in turn results in more specific distributional parameter estimates. It may be that if infinite data were available, the posterior distributions of the distributional parameters would converge to a single value. As more data leads to a narrower p-box, it would be interesting to perform the theoretical work to determine if having access to infinite data would lead the Bayesian UQ results converge to the Roy and Oberkampf (2011) results.

The 9 strength measurements collected from the simply supported beam example (section 4.2) were obtained from 9 segments originating from 9 different boards. However, measurements were actually obtained on 36 segments, as the 9 boards were originally 8 ft (243.84 cm) long and were cut into 4 pieces, each piece 2 ft (60.96 cm) long. Because every 4 specimens were cut from the same board, the 36 measurements are not independent. Dependent measurements are likely to be encountered in practice when measurements are taken on material from the same lot or batch. It remains to be determined whether the assumption of independence is necessary for the proposed Bayesian UQ method. If it is found that the assumption of independence is needed for the proposed Bayesian approach, adjusting the Bayesian UQ method to account for lot or batch dependence will be an important extension to add in future research.

REFERENCES

(2015). *NDS Supplement: National design specification design values for wood construction.* American Wood Council. Available at (https://www.awc.org/pdf/codes-standards/publications/nds/AWC-NDS2015-Supplement-ViewOnly-1411.pdf).

(2018). *Python Language Reference, version 3.7.* Python Software Foundation. Available at (https://www.python.org).

Anagnostou, E. N. and Krajewski, W. F. (1999). Uncertainty quantification of mean-areal radar-rainfall estimates. *Journal of Atmospheric and Oceanic Technology*, 16(2).

Bird, J. O. and Chivers, P. J. (1993). *Newnes Engineering and Physical Science Pocket Book.* Newnes.

Brooks, S., Gelman, A., Jones, G., and Meng, X. (2011). *Handbook of Markov Chain Monte Carlo.* CRC Press.

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* Springer.

Cooke, R. M. (1991). *Experts in Uncertainty: Opinion and Subjective Probability in Science.* Oxford University Press.

Der Kiureghian, A. D. and Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter? *Structural Safety*, 31(2):105–112.

Dezfuli, H., Kelly, D., Smith, C., Vedros, K., and Galyean, W. (2009). *Bayesian Inference for NASA Probabilistic Risk and Reliability Analysis.* National Aeronautics and Space Administration. Retrieved from https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20090023159.pdf.

Ewing, M., Liechty, B. C., and Black, D. L. (2018). A general methodology for uncertainty quantification in engineering analyses using a credible probability box. *Journal of Verification, Validation, and Uncertainty Quantification*, 3(2). https://doi.org/10.1115/1.4041490.

Gamerman, D. and Hedibert, F. L. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference.* CRC Press.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6). https://doi.org/10.1109/TPAMI.1984.4767596.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1). https://www.jstor.org/stable/2334940.

Hora, S. (2016). Probability elicitation. *The Oxford Handbook of Probability and Philosophy.* https://doi.org/10.1093/oxfordhb/9780199607617.013.30.

Jiang, X. and Mahadevan, S. (2009). Bayesian hierarchical uncertainty quantification by structural equation modeling. *International Journal in Numerical Methods in Engineering*, 80(6–7):717–737. https://doi.org/10.1002/nme.2550.

Lockwood, B. A., Anitescu, M., and Mavriplis, D. J. (2012). Mixed aleatory/epistemic uncertainty quantification for hypersonic flows via gradient-based optimization and surrogate models. *American Institute of Aeronautics and Astronautics.* https://doi.org/10.2514/6.2012-1254.

Lunn, D., Jackson, C., Best, N., Thomas, A., and Spiegelhalter, D. (2013). *The BUGS Book: A Practical Introduction to Bayesian Analysis.* CRC Press.

Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N. (2009). The bugs project: Evolution, critique, and future directions. *Statistics in Medicine*, 28:3049–3067.

Lynch, S. M. (2007). *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists.* Springer.

Maitre, O. L. and Knio, O. M. (2010). *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*. Springer Science & Business Media.

Malinverno, A. and Briggs, V. A. (2004). Expanded uncertainty quantification in inverse problems: Hierarchical bayes and empirical bayes. *Geophysics*, 69(4):1005–1016. `http://doi.org/10.1190/1.1778243`.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092. `https://doi.org/10.1063/1.1699114`.

Nagel, J. B. and Sudret, B. B. (2015). A unified framework for multilevel uncertainty quantification in bayesian inverse problems. *Probabilistic Engineering Mechanics*, 43:68–84. `http://doi.org/10.1016/j.probengmech.2015.09.007`.

Najm, H. N., Debusschere, B. J., Marzouk, Y. M., Widmer, S., and Le Maitre, O. P. (2009). Uncertainty quantification in chemical systems. *International Journal for Numerical Methods in Engineering*, 80(6-7):789–814. `https://www.doi.org/10.1002/nme.2551`.

Oakley, J. (2019). *SHELF*. R package version 1.6.0. (`https://cran.r-project.org/web/packages/SHELF/SHELF.pdf`).

Oakley, J. E. and O'Hagan, A. (2019). *SHELF: the Sheffield Elicitation Framework (version 4.0)*. School of Mathematics and Statistics, University of Sheffield, UK. `http://www.tonyohagan.co.uk/shelf/`.

O'Hagan, A. (2019). Expert knowledge elicitation: Subjective but scientific. *The American Statistician*, 73. `https://doi.org/10.1080/00031305.2018.1518265`.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. (`https://www.R-project.org/`).

Rowe, G. and Wright, G. (1999). The Delphi technique as a forecasting tool: Issues and analysis. *International Journal of Forecasting*, 15.

Roy, C. J. and Oberkampf, W. L. (2011). A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing. *Computer Methods in Applied Mechanics and Engineering*, 200(25-28). https://doi.org/10.1016/j.cma.2011.03.016.

Sankararaman, S., Ling, Y., and Mahadevan, S. (2011). Uncertainty quantification and model validation of fatigue crack growth prediction. *Engineering Fracture Mechanics*, 78(7):1487–1504. http://doi.org/10.1016/j.engfracmech.2011.02.017.

Smith, R. C. (2013). *Uncertainty Quantification: Theory, Implementation, and Applications*. SIAM.

Spetzler, C. S. and Stael Von Holstein, C. (1975). Probability encoding in decision analysis. *Management Science*, 22(3):340–358. http://doi.org/10.1287/mnsc.22.3.340.

Sturtz, S., Ligges, U., and Gelman, A. (2017). *R2OpenBUGS: Running OpenBUGS from R*. R package version 3.2 (https://cran.r-project.org/web/packages/R2OpenBUGS/index.html).

Sullivan, T. J. (2015). *Introduction to Uncertainty Quantification*. Retrieved from https://userpage.fu-berlin.de/sullivan/pdf/bms_uq_notes_00-05.pdf.

Wang, P., Youn, B. D., Z., X., and Kloess, A. (2009). Bayesian reliability analysis with evolving, insufficient, and subjective data sets. *Journal of Mechanical Design*, 131(11). https://doi.org/10.1115/1.4000251.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. (http://ggplot2.org).

Zhang, Y. (2013). Efficient uncertainty quantification in aerospace analysis and design. *Doctoral Dissertations*, 2162. https://scholarsmine.mst.edu/doctoral_dissertations/2162.