SAND2020-7639C

UNDERSTANDING ALEATORY AND EPISTEMIC PARAMETER UNCERTAINTY IN STATISTICAL MODELS

N. W. Porter and V. A. Mousseau
Sandia National Laboratories, PO Box 5800 Albuquerque NM 87185-0748

nwporte@sandia.gov, vamouss@sandia.gov

ABSTRACT

To make informed decisions about model prediction, different sources of uncertainty must be treated appropriately. In general, uncertainties can take two forms. The first is aleatory uncertainty, which accounts for the natural variation of inputs and parameters; it is irreducible and cannot be decreased with additional knowledge. The second type of uncertainty is epistemic, which results from lack of knowledge about the system of interest. Epistemic uncertainty can be reduced by obtaining additional information. The distinction between epistemic and aleatory uncertainty is often ignored or misunderstood during the formulation of statistical models for calibration. In this work, two common statistical models for calibration are analyzed to fully understand the distinction between the two types of uncertainties during the calibration process. The first statistical model treats only epistemic uncertainty in the parameters, and the second includes both epistemic and aleatory parameter uncertainty. These statistical models are used to create synthetic data and a simple empirical model is calibrated to that data. If the calibration method has assumptions consistent with the statistical model used to create the synthetic data, then the parameter distributions are accurately estimated. However, if a calibration method is inconsistent with the statistical model, the underlying parameter distributions are mischaracterized. When incorporated into a best estimate plus uncertainty framework, this process will indicate the wrong parameter distribution(s) to be propagated through a computational model.

1 INTRODUCTION

Best Estimate Plus Uncertainty (BEPU) methodologies are becoming increasingly common in nuclear engineering. The most common methods follow four basic steps: (1) characterize the statistical distributions of model parameters, (2) randomly sample the model parameters, (3) evaluate a computational model using those parameters, and then (4) use the results to assess the distribution of some quantity of interest. For example of such methodologies, see [2, 6].

In this work, we focus on the first step in the BEPU process. In many studies, the model parameters are defined using expert elicitation, however this method is susceptible to biases and inaccuracies. Instead, calibration can be used, which optimally fits an empirical model to match experimental data. Unlike expert elicitation, calibration is easily defended and can be improved as more experimental data is gathered. A wide variety of calibration methods have been developed throughout the literature [12]. In this work, we analyze the assumptions which underlie these methods.

The first assumption of most calibration methods is that experimental data has some *measurement uncertainty*, related to the hardware and experimental process, which can be treated as random. In addition to measurement uncertainty, some statistical models used for calibration attribute uncertainty to the model parameters themselves. In the most common paradigm, the model parameter uncertainty is assumed to be *epistemic*. In other words, all parameter variation is due to a lack of knowledge or data. This implies that the parameter distribution would converge to a point value as more knowledge is obtained. In addition to epistemic parameter uncertainty, some calibration methods attribute *aleatory* uncertainty to the parameters. Aleatory uncertainty is due to natural variation, and is not reduced as more data is gathered. With these methods, the parameter distribution converges to the aleatory uncertainty distribution as more knowledge is gathered.

The two types of statistical models have very different physical interpretations. The first, with only epistemic uncertainty, implies that each model parameter has a true but unknown value. This idealized case would be exemplified if an input parameter were some physical constant, such as the acceleration of gravity or speed of light. If an experiment is assumed to have both aleatory and epistemic parameter uncertainty, the input parameter is acknowledged to have an unknown *distribution*. This parameter treatment could potentially represent any experiment where lack of knowledge prevents parameters from being known precisely. For example, low quality data or physics that are not represented in the empirical model.

In Section 2, the methods used in this work are outlined. Results are presented in Section 3 for (1) verification of the calibration methods when they are applied to consistent data, and (2) demonstration of the consequences when statistical models are misapplied. Section 4 provides a discussion of the results and possible impacts on BEPU analyses.

2 METHODS

The vector $y = [y_1, y_2, \ldots, y_N]$ represents N experimental measurements corresponding to state settings $x = [x_1, x_2, \ldots, x_N]$. An empirical model $f(x, \theta)$ is formulated to relate y and x. Here, θ is a vector of parameters which will be calibrated. The statistical model is formed as

$$y = f(x, \theta) + \varepsilon. \tag{1}$$

The observational error ε represents measurement uncertainties and other errors associated with the experiment. In this work, the observational error σ is assumed to be identically and independently distributed about zero with a constant variance: $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. The two statistical models discussed in Section 1 have the following treatment of model parameters: (1) they each have a single fixed unknown value, or (2) they are assumed to have joint distribution. The aleatory uncertainty distribution is assumed to be a joint Gaussian distribution $\theta \sim \mathcal{N}(\theta_0, \Psi)$, where the parameter means θ_0 and covariance Ψ have fixed unknown values. We employ an empirical model that has a single model parameter; therefore, the covariance simplifies to a single standard deviation which will be denoted as ψ.

2.1 Bayesian Calibration

Bayesian analysis assumes a joint prior parameter distribution $\pi_o(\theta)$, which characterizes previous knowledge about the distribution. The goal of Bayesian calibration is to estimate the posterior distribution $\pi(\theta|y)$, which incorporates information about both the prior and the experimental data [12]. The posterior and prior functions are related by Bayes' Formula using the likelihood function $\pi(y|\theta)$:

$$\pi(\theta|y) = \frac{\pi(y|\theta)\pi_o(\theta)}{\int \pi(y|\theta)\pi_o(\theta)d\theta}. \tag{2}$$

Greater likelihood values correspond to combinations of parameters that are most likely to have generated the observed experimental data. In most engineering applications of Bayes' formula, direct solution of Equation 2 is impossible because (1) the numerator has no closed form solution and cannot be directly sampled, or (2) the denominator involves an integration over the—often high-dimensional—parameter space. In these cases, Markov Chain Monte Carlo (MCMC) methods can be used to solve Bayesian Calibration problems. These are a general class of methods which are used to construct a sampling-based chain whose stationary distribution is the desired posterior.

In this work, a different Bayesian method is employed for each statistical model. Detailed algorithms for each method are contained in the citations in this section (particularly [7] and [8]). The conclusions of this work are independent of the particulars of the calibration methods, so extraneous details are not described here. This allows the reader to focus on the simple delineated example.

The first Bayesian method, which treats only epistemic uncertainty, is Delayed Rejection Adaptive Metropolis (DRAM) [3]. DRAM improves upon the most basic MCMC method, Metropolis-Hastings [5], in two ways. First, the covariance matrix is updated throughout the calibration, which makes the algorithm *adaptive* [4]. This increases the efficiency of the algorithm because the high-probability parameter space is preferentially sampled. The second improvement is that second stage candidates are proposed, which is termed *delayed rejection* and encourages sampling of the entire parameter space [9].

For Bayesian calibration, estimating both epistemic and aleatory parameter uncertainty requires a hierarchical algorithm. In this work, the method used is a hierarchical Metropolis-within-Gibbs algorithm. The epistemic variation is estimated using the Metropolis algorithm, and the aleatory uncertainty is estimated via a nested Gibbs algorithm. For more details of Gibbs sampling, see [1]. The implementation used in this work is detailed in [7, 8, 13].

2.2 Synthetic Data

To verify that the selected calibration methods are accurate, this work utilizes simple synthetic data. The synthetic data is created using a linear function which relates x and y: $f(x,\theta) = \theta x$. This linear function is both the underlying physical basis of the synthetic data and the empirical model to be calibrated. Note that this model is selected so that the calibration is tractable and easy to

understand. Real physics models can be much more complex and can exhibit mismatches between the physics and the empirical model.

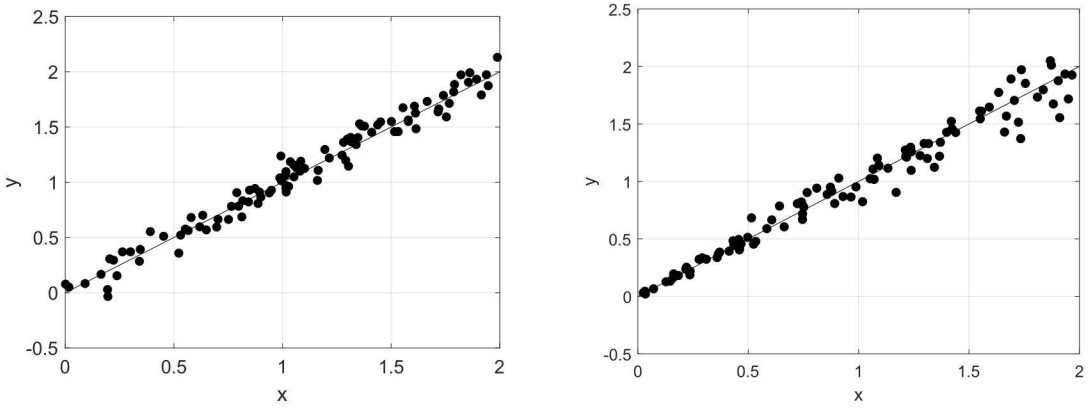
Each set of data uses the statistical model in Equation 1 with different assumptions. The statistical model with only epistemic uncertainty will be consistent with data sampled from:

$$y = \theta x + \mathcal{N}(0, \sigma^2). \tag{3}$$

For this model, $\theta = 1$ and $\sigma = 0.1$. The statistical model with both epistemic and aleatory uncertainty will be calibrated to data sampled from:

$$y = \mathcal{N}(\theta_0, \psi^2)x + \mathcal{N}(0, \sigma^2). \tag{4}$$

Here, ψ is the standard deviation of the aleatory distribution of θ. For this model, $\theta_0 = 1$, $\psi = 0.1$, and $\sigma = 0.01$. One hundred synthetic data points are shown for each model in Figure 1. The base linear model, $f(x, \theta) = \theta x$, is also shown.



(a) Only epistemic uncertainty (Equation 3) (b) Epistemic and aleatory (Equation 4)

Figure 1: Example synthetic data for each statistical model

2.3 Area Metric

In this work, convergence to the expected value or distribution will be quantified using the so-called *area metric*. This metric, which is widely used in engineering disciplines, quantifies the area between two cumulative distribution functions (CDFs) [11]. So, if $F(x)$ and $G(x)$ characterize the known and calibrated CDFs, the area metric is defined as

$$d_{area} = \int_{-\infty}^{\infty} |F(x) - G(x)| dx. \tag{5}$$

Note that this metric was chosen because it can be calculated when either $F(x)$ or $G(X)$ (or both) are characterized by a point value. Therefore, the same metric can be used when θ is expected to converge to a single value or when it converges to the underlying aleatory distribution.

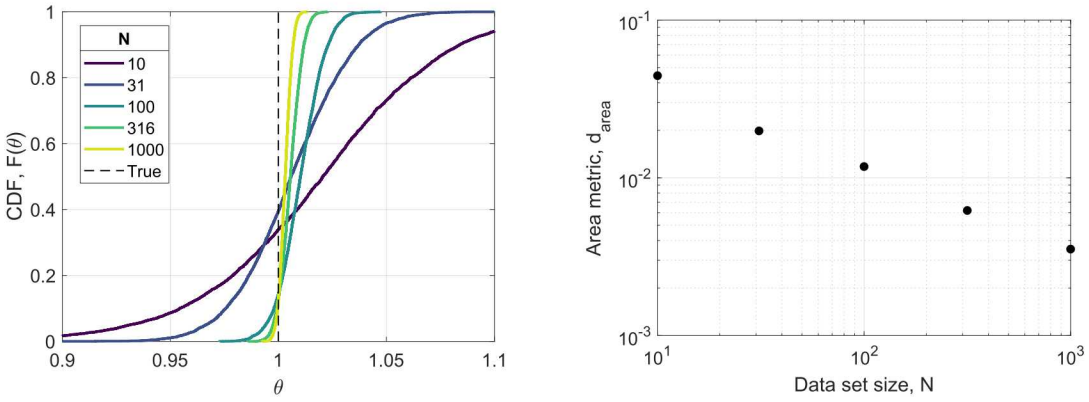
3 RESULTS

In this section, the results of calibration to synthetic data are reported as a function of the number of synthetic data points N. The first example is a traditional calibration to data with only epistemic uncertainty. The second example includes both aleatory and epistemic parameter uncertainty. The final example calibrates data with both sources of uncertainty using a model which only accounts for epistemic uncertainty.

3.1 Only Epistemic Uncertainty

In this section, synthetic data is created using Equation 3 and a consistent statistical model is used to estimate the underlying parameters. The DRAM method is used with 10^4 burnin samples. By comparing the calibrated θ and σ to the values used to compute the synthetic data, we verify that the calibration functions properly.

The results for θ and σ are respectively shown in Figure 2 and Figure 3. Note that the synthetic data is sampled without replacement to reduce statistical noise in the results. As the number of data points increases, epistemic uncertainty in θ is reduced. As the epistemic uncertainty is eliminated, θ converges to the expected point value. This is shown qualitatively by the CDF and quantitatively by the area metric in Figure 2. The observational error variance σ is also shown to converge to the correct point value as data set size increases in Figure 3.



(a) θ converges to known value

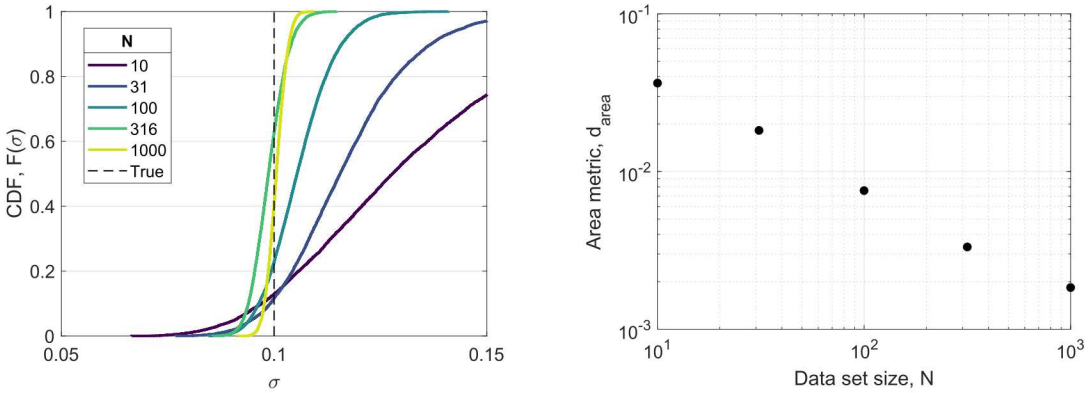
(b) d_{area} decreases as N increases

Figure 2: Convergence of θ for data and statistical model with only epistemic uncertainty

The assumptions of the statistical model are consistent with the data, so the calibration parameters converge to the correct values.

3.2 Epistemic and Aleatory Uncertainty

Now, synthetic data is created using Equation 4, which includes both epistemic and aleatory sources of parameter uncertainty. The hierarchical Metropolis-within-Gibbs algorithm is employed in this section with 10^5 burnin samples. Through comparison of the calibrated θ_0, ψ, and σ to the

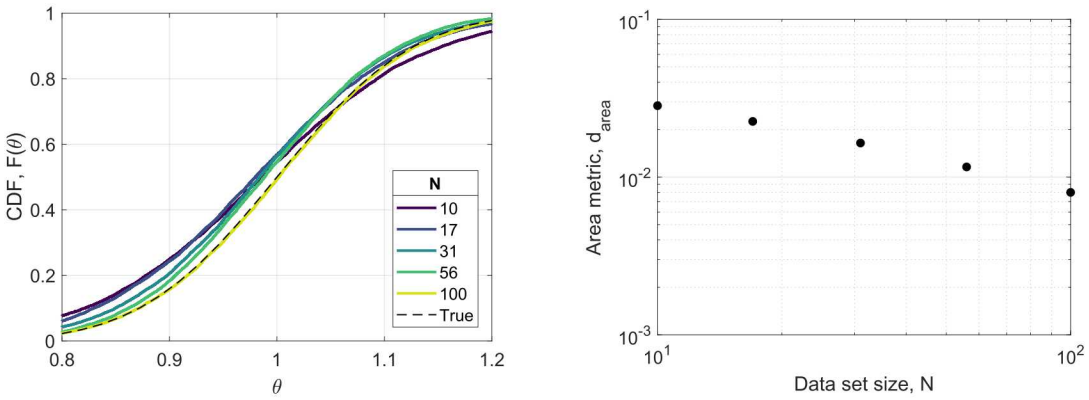


(a) σ converges to known value (b) d_{area} decreases as N increases

Figure 3: Convergence of σ for data and statistical model with only epistemic uncertainty

synthetic values, the calibration accuracy can be verified. Smaller data set sizes are used in this section due to the computational requirements of the hierarchical Bayesian calibration.

The results for θ and ψ are summarized by the distributions in Figure 4. To maintain continuity between the results in each section, these distributions represent the total uncertainty in θ, including both aleatory and epistemic. This type of distribution is known as an unconditional distribution; it is constructed by statistically sampling from a family of epistemic realizations of the aleatory uncertainty [10]. It is shown that, as the data set size is increased, the distribution of θ converges to the expected aleatory distribution. The results for σ are shown in Figure 5. As the data set size increases, d_{area} decreases; therefore, the observational error converges to the expected distribution.

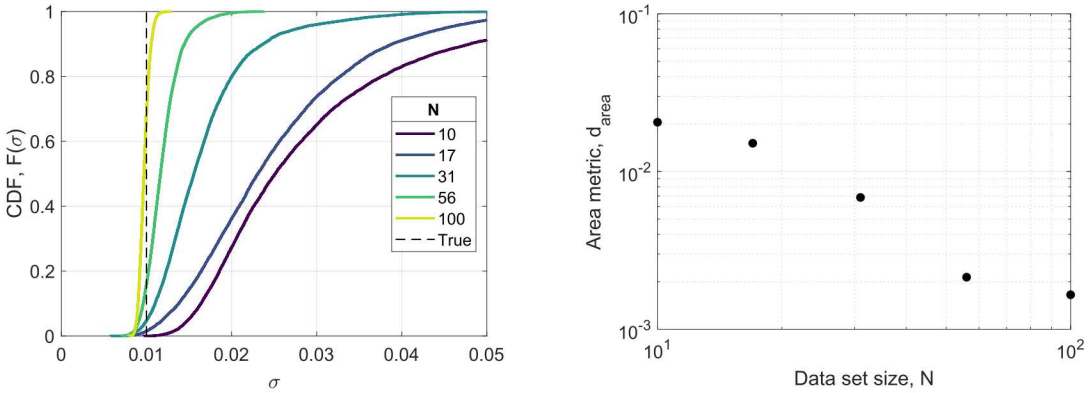


(a) θ converges to known distribution (b) d_{area} decreases as N increases

Figure 4: Convergence of θ for data and statistical model with both sources of uncertainty

3.3 Mischaracterization of Data

In this final exercise, we employ the incorrect statistical model for a set of synthetic data. Synthetic data with both sources of uncertainty is sampled from Equation 4 and the statistical model with only epistemic uncertainty is applied. The DRAM algorithm is used for the calibration with 10^4 burnin samples.

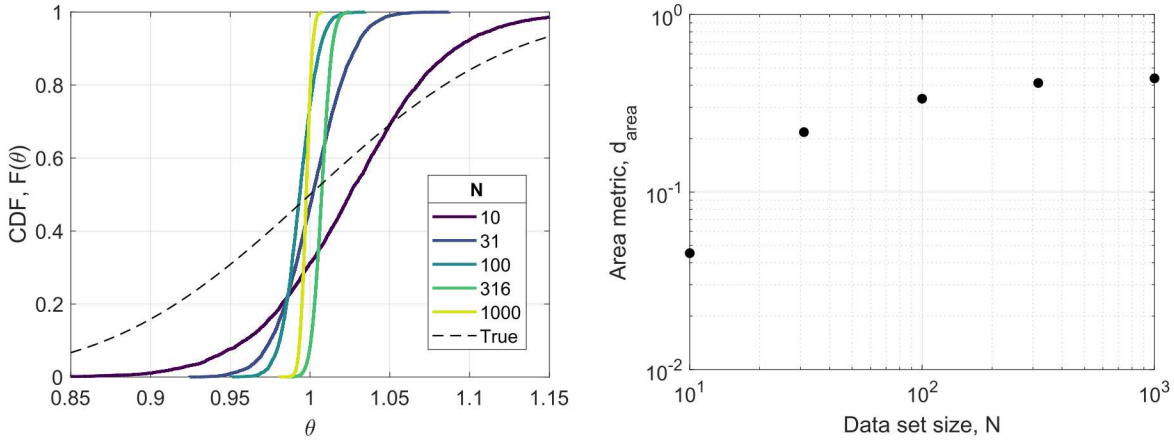


(a) σ converges to known value (b) d_{area} decrease as N increases

Figure 5: Convergence of σ for data and statistical model with both sources of uncertainty

The parameter distribution results are shown in Figure 6. Because the statistical model does not include aleatory parameter uncertainty, the calibration incorrectly converges to a point value. Therefore, the area metric increases and starts to approach a constant value.

The standard deviation of the observational error is shown in Figure 7 as data set size increases. Since the statistical model fails to account for the aleatory uncertainty in θ (see Figure 6), all uncertainty in the data gets apportioned to the observational error. This results in an overprediction of the observational error variance by an order of magnitude.

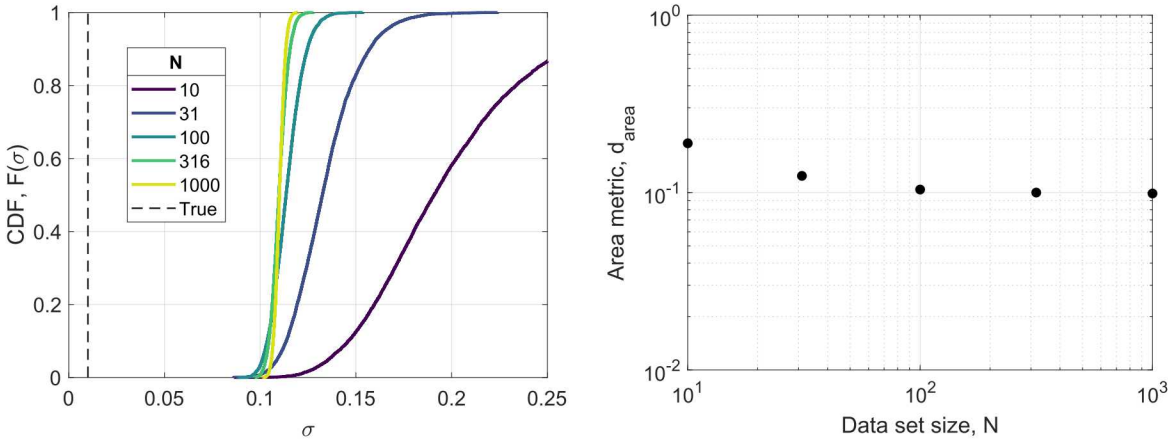


(a) θ converges to wrong value (b) d_{area} increases with N

Figure 6: Non-convergence of θ for the mismatched data and model

4 CONCLUSION

In this work, the underlying assumptions of two statistical models were analyzed for a linear empirical model with a single model parameter. By using simple synthetic data, relationships between the statistical models, parameter distributions, and data were easily identified. First, it was shown that calibration is accurate when statistical model assumptions are consistent with underlying data.



(a) σ converges to the wrong value (b) d_{area} converges a non-zero value

Figure 7: Non-convergence of σ for the mismatched data and model

This was demonstrated for both types of statistical model: when only epistemic uncertainty is present, and when both epistemic and aleatory uncertainty are considered.

Because the experimental data is synthetic, there are no unknown sources of uncertainty. Therefore, the epistemic uncertainty is eliminated as the number of data samples is increased. For statistical models without aleatory uncertainty, this means that the parameter distribution converges to a point value. For statistical models with aleatory uncertainty, parameter distributions converge to the aleatory uncertainty. The observational error, which is included in both statistical models to quantify measurement uncertainty and other sources of error, is also better characterized as more data is collected. This was demonstrated by the first two calibration exercises. This can be viewed as an additional source of epistemic uncertainty that is reduced with more information.

In the final exercise, a calibration to data with aleatory uncertainty was performed using a statistical model which assumed no aleatory uncertainty. Because the statistical model mischaracterized the underlying parameter distribution, all error was apportioned to the observational error term. This resulted in a drastic overestimation of the observational error by about an order of magnitude. Additionally, the shape of the data is mischaracterized by the calibrated model. The observational error is constant, whereas the parameter uncertainty scales with x. Since all error is apportioned to the observational error, the calibrated model approximates data with the shape of Figure 1(b) using the data shape of Figure 1(a). The error is spread evenly over the state space, and therefore it is mischaracterized.

The parameter distribution was completely mischaracterized during the final exercise. This is an important result because, in general, BEPU analyses propagate the model parameter distribution through a computational tool. If epistemic and aleatory uncertainty are not separated in a real analysis, the uncertainty of the quantity of interest can be underestimated. Moreover, an underestimation of uncertainty is often nonconservative, meaning that it can have drastic real-world consequences. Therefore, it is imperative that all sources of both epistemic and aleatory uncertainty are understood, fully characterized, and included in statistical models.

To compensate for underestimating the parameter uncertainty, an analyst might propagate the observational error as well. Since this would include all sources of error in the propagation, the uncertainty magnitude would be less drastically underpredicted. However, this would mischaracterize the uncertainty shape, since the parameter error scales with x whereas the observational error does not. Therefore, in this case, propagating the observational error through a simulation tool would result in uncertainty distributions shaped like Figure 1(a) when the underlying data is characterized by Figure 1(b).

This methodology can be applied to real problems with experimental data; however this introduces many complexities that are difficult to address: heteroskedastic observational error, unknown human errors, and model discrepancies between experimental data and assumptions made by statistical models. For clarity, these effects are left as topics for future study. In addition, future work can include additional statistical models and calibration methodologies, analysis of multiple model parameters, and nonlinearity between model parameters.

5 ACKNOWLEDGMENTS

We would like to thank Dusty Brooks and Aaron Krueger of Sandia National Laboratories for their thoughtful reviews. This work was supported by the Laboratory Directed Research and Development program at Sandia National Laboratories. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

6 REFERENCES

[1] D. Gamerman. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall, Boca Raton, FL, 1997.

[2] H. Glaeser. GRS method for uncertainty and sensitivity evaluation of code results and applications. *Science and Technology of Nuclear Installations*, 2008(798901), 2008. doi:10.1155/2008/798901.

[3] H. Haario, M. Laine, A. Mira, and E. Saksman. DRAM: Efficient adaptive MCMC. *Statistics and Computing*, 16(4):339–354, 2006. doi:10.1007/s11222-006-9438-0.

[4] H. Haario, E. Saksman, and J. Tamminen. An adapative metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001. doi:10.2307/3318737.

[5] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. doi:10.1093/biomet/57.1.97.

[6] J. Hou et al. Bechmark for uncertainty analysis in modeling (UAM) for design, operation, and safety analysis of LWRs. Technical Report NEA/NSC/DOC, 2017.

[7] M. Laine. *Adaptive MCMC Methods With Applications in Environmental and Geophysical Models*. PhD thesis, Lappeenranta University of Technology, 2008.

[8] M. Laine. MCMC toolbox for Matlab. `mjlaine.github.io/mcmcstat/`, 2017. Accessed: 2019-07-13.

[9] A. Mira. On Metropolis-Hastings algorithm with delayed rejection. *Metron*, 59(3):231–241, 2001.

[10] J. Mullins and S. Mahadevan. Bayesian uncertainty integration for model calibration, validation, and prediction. *Journal of Verification, Validation and Uncertainty Quantification*, 1, 2016. doi:10.1115/1.4032371.

[11] W. L. Oberkampf and C. J. Roy. *Verification and Validation in Scientific Computing*. Cambridge University Press, New York, NY, first edition, 2010. doi:10.1017/CBO9780511760396.

[12] R. C. Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*. SIAM, Philadelphia, PA, 2014.

[13] J. Susiluoto et al. Calibrating the sqHIMMELI v1.0 wetland methane emission model with hierarchical modeling and adaptive MCMC. *Geoscientific Model Development*, 11(8):1199–1228, 2018. doi:10.5194/gmd-11-1199-2018.