

Bayesian Uncertainty Integration for Model Calibration, Validation, and Prediction

Joshua Mullins

Department of Civil and
Environmental Engineering,
Vanderbilt University,
Nashville, TN 37235
e-mail: joshua.g.mullins@vanderbilt.edu

Sankaran Mahadevan

Department of Civil and
Environmental Engineering,
Vanderbilt University,
Nashville, TN 37235

This paper proposes a comprehensive approach to prediction under uncertainty by application to the Sandia National Laboratories verification and validation challenge problem. In this problem, legacy data and experimental measurements of different levels of fidelity and complexity (e.g., coupon tests, material and fluid characterizations, and full system tests/measurements) compose a hierarchy of information where fewer observations are available at higher levels of system complexity. This paper applies a Bayesian methodology in order to incorporate information at different levels of the hierarchy and include the impact of sparse data in the prediction uncertainty for the system of interest. Since separation of aleatory and epistemic uncertainty sources is a pervasive issue in calibration and validation, maintaining this separation in order to perform these activities correctly is the primary focus of this paper. Toward this goal, a Johnson distribution family approach to calibration is proposed in order to enable epistemic and aleatory uncertainty to be separated in the posterior parameter distributions. The model reliability metric approach to validation is then applied, and a novel method of handling combined aleatory and epistemic uncertainty is introduced. The quality of the validation assessment is used to modify the parameter uncertainty and add conservatism to the prediction of interest. Finally, this prediction with its associated uncertainty is used to assess system-level reliability (a prediction goal for the challenge problem). [DOI: 10.1115/1.4032371]

1 Introduction

Engineering decisions are often (if not always) made in the presence of significant uncertainty due to the insufficient quantity and quality of available data. Since information is typically especially sparse in the domain of interest (i.e., the usage condition of the system), data are commonly composed in the form of a hierarchy in which lower level data (e.g., simplified test configurations and conditions, material tests, and component tests) are used to inform computational models that will be used to predict system performance. Since the models are always imperfect, verification and validation (V&V) methods have been actively researched as an approach for accumulating evidence to support a prediction model. The results of these activities may then be integrated together in order to incorporate the known sources of uncertainty into the prediction.

V&V frameworks [1–3] for prediction typically involve the following activities: (1) uncertainty characterization, (2) model verification, (3) model calibration, (4) model validation, and (5) uncertainty propagation for prediction. Uncertainty characterization typically applies to natural variability in system inputs (configuration variables measured in corresponding experiments, e.g., load and temperature) and model parameters (system variables that are included in models but not directly measured in experiments, e.g., material properties), and this uncertainty is commonly described probabilistically by any of the many well-known parametric probability distributions. Characterizing this uncertainty informs model developers about which input ranges are important (i.e., what input settings are expected to be exercised during uncertainty propagation), which enables model verification (including both code verification and solution verification) to be conducted efficiently. Then, in model calibration [4–8], since the model parameters are not measured directly, they are inferred

from experimental observations of input and output quantities. These parameters may be deterministic values that can be estimated directly or naturally varying quantities for which distribution parameters that describe the variability may be estimated. Input and parameter uncertainty are then propagated through the calibrated models, and the result is compared against an independent set of experimental observations. This process, known as model validation [9–18], assesses the accuracy of the prediction of a computational model. Since the model is never completely accurate, the model form error should be accounted for. This adds an additional source of uncertainty when input and parameter uncertainty are propagated forward to make the prediction on system performance in the regime of interest. This paper presents a framework for model calibration, model validation, and prediction and discusses how the approach is applied differently in the presence of different types of uncertainty.

In order to make fair assessments of the models and informed decisions using the prediction, the sources of uncertainty should be considered separately and treated appropriately. An important distinction is the difference between aleatory and epistemic uncertainty sources. Aleatory uncertainty is the natural variation of inputs and parameters that impact outputs of interest. This uncertainty is irreducible, and it is commonly treated with probability theory. Epistemic uncertainty results from lack of knowledge about the system of interest, and it can be further separated into model uncertainty (e.g., parameter uncertainty, solution approximation errors, and model form uncertainty) and data uncertainty (e.g., measurement uncertainty and sparse or imprecise data). Since it stems from lack of knowledge, epistemic uncertainty can be reduced by obtaining additional information. In the literature, this uncertainty has been modeled in a number of different ways [19–24], but for any of these treatments, it is important to retain the separation between aleatory and epistemic sources [25–27] in order to support decisions about prediction uncertainty reduction. When aleatory and epistemic uncertainty sources are confounded, it is difficult to assess the benefit of collecting additional

Manuscript received February 25, 2015; final manuscript received December 22, 2015; published online February 19, 2016. Guest Editor: Kenneth Hu.

information (e.g., performing additional tests). For example, when the aleatory contribution to the overall uncertainty is large relative to the epistemic contribution and this information is not known, resources might be allocated inefficiently or consumed unnecessarily on testing even though the additional test data cannot reduce the aleatory uncertainty and therefore cannot significantly improve the precision of the prediction. Furthermore, when applying probabilistic model validation methods, the primary interest is the epistemic uncertainty (i.e., parameter uncertainty and model form uncertainty) in the prediction, and the separation of uncertainty sources enables these contributions to be isolated.

This paper applies a Bayesian subjective probability treatment of epistemic uncertainty and demonstrates how this approach can be used to distinguish aleatory and epistemic uncertainty sources in calibration, validation, and prediction. The proposed methodology was particularly motivated by the V&V challenge problem proposed by the Sandia National Laboratories [28]. This problem presents some significant challenges for uncertainty characterization and propagation. In particular, the limitations of the data add uncertainty to the characterization of a population of interest. Only a small number of samples are tested, yet the goal is to extract meaningful information about the overall population in order to make a reliability assessment. Physical intuition suggests that system properties and environments vary widely across the population, but only sparse observations are available to characterize this aleatory uncertainty. Thus, the data limitations introduce epistemic uncertainty about the form and magnitude of the aleatory uncertainty that is expected to exist. Separation of these sources of uncertainty is critical to a thorough reliability assessment because the predictive capability of the model cannot be assessed accurately when these uncertainty sources are combined.

The remainder of this paper details how uncertainty separation can be achieved and maintained during each phase of the assessment, and then, the methodology is demonstrated for the V&V challenge problem. In Sec. 2, Bayesian calibration is first described for the classical case of estimation of purely epistemic parameters, and it is then extended to a scenario in which the model parameters have both aleatory and epistemic uncertainty. The Johnson family of probability distributions is used to represent the aleatory variability, and the uncertainty about the Johnson distribution parameters represents the epistemic uncertainty. In Sec. 3, the model reliability metric approach to model validation is described. Since this metric is specifically targeted at epistemic uncertainty in the model (particularly epistemic parameter uncertainty and model bias), the approach is extended to account for the scenario in which the parameters are also affected by aleatory variability. In Sec. 4, a method for integrating the results of the calibration and validation activities is described. The Sandia V&V challenge problem prediction results and reliability analysis are presented and discussed in Sec. 5, and the paper is concluded in Sec. 6.

2 Bayesian Calibration Under Uncertainty

Bayesian calibration [5–7,29] is an approach for inferring unmeasured parameters by observing particular values of the outputs and corresponding inputs. As opposed to deterministic parameter estimation, which results in only a single value for the parameters, Bayesian calibration results in a posterior probability distribution that represents the subjective probability of each value in the domain. Note that the assumption implicit to this approach is that the parameter values are deterministic in reality, but the values cannot be inferred precisely due to data uncertainty in the observations as well as model errors that may bias the results. Therefore, the posterior distribution represents epistemic uncertainty and not aleatory uncertainty.

2.1 Calibration of Purely Epistemic Parameters. According to Bayes' theorem, the posterior probability of the parameters $f_{\Theta}(\theta|y_d)$ is proportional to the product of the likelihood function

$f_{Y_d}(y_d|\theta)$ (i.e., the probability of observing the data y_d given a particular parameter set θ), henceforth denoted $L(\theta)$, and the prior density $f_{\Theta}(\theta)$

$$f_{\Theta}(\theta|y_d) = \frac{L(\theta)f_{\Theta}(\theta)}{\int L(\theta)f_{\Theta}(\theta)d\theta} \quad (1)$$

To construct the likelihood function, a model that probabilistically describes the difference between prediction and observation is needed. A common model is to attribute the difference between a particular observation $y_{d_{ij}}$ and the prediction y_m at input x_j to zero-mean Gaussian measurement noise in the observation $e_{d_{ij}}$

$$y_{d_{ij}} = y_m(x_j, \theta) + e_{d_{ij}}, \quad i = 1, \dots, n_j; \quad j = 1, \dots, m \quad (2)$$

$$E_{d_{ij}} \sim N(0, \sigma_d), \quad \text{i.i.d.} \quad (3)$$

In Eq. (2), i can indicate a vector response of length n_j or n_j replicate experiments at state x_j , and m denotes the total number of states. To evaluate the likelihood in Eq. (1) at the data sample values according to the model given by Eqs. (2) and (3), observe that $\{y_{d_{ij}}\}$ is the sum of a constant and a vector of independent normal random variables. Thus,

$$L(\theta) = \prod_{j=1}^m \prod_{i=1}^{n_j} \frac{1}{\sigma_d \sqrt{2\pi}} \exp \left\{ -\frac{[y_m(x_j, \theta) - y_{d_{ij}}]^2}{2\sigma_d^2} \right\} \quad (4)$$

In practice, the product of Eq. (4) and the prior density (as in Eq. (1)) cannot be normalized or inverted easily in order to draw samples from the joint posterior distribution. Instead, a function that is proportional to the posterior density is sampled via Markov chain Monte Carlo (MCMC) sampling methods [30–33]. MCMC methods typically require a surrogate model [34–38] even when the model is relatively cheap because of the large number of evaluations that are needed. This is an important limitation of Bayesian methods; the number of function evaluations is usually on the order of 10^4 or 10^5 in order to achieve convergence in MCMC. Additionally, the evaluations are typically serial in MCMC algorithms, so parallel resources cannot improve the efficiency significantly unless they can be used to improve the efficiency of the model evaluation itself. In this paper, Gaussian process (GP) surrogate models [39] are used because of their ability to represent general forms of an output response and also provide a direct estimate of the uncertainty associated with the fit to the computational model. The required number of serial evaluations is unaffordable for almost any real computational model, but it is fairly negligible when using a GP model (on the order of seconds or minutes).

Note that the relationship given in Eq. (2) does not account for model form error. Since model form error is often a leading source of the difference between prediction and observation, many researchers [5,7,8] add a stochastic, input-dependent model discrepancy term to the model prediction. The goal of this approach, commonly referred to as the Kennedy–O'Hagan framework [29], is to reduce the bias in the parameter estimates; bias is introduced when parameters are used to fit an incorrect model form to the observed data. However, since the mathematical form of the model discrepancy is always unknown, an additional set of parameters must be introduced to define a stochastic model discrepancy function, and these parameters must be inferred jointly with θ . This expansion of the calibration problem leads to some additional difficulties, including selection of a proper discrepancy formulation [8] and unique identifiability of the expanded parameter set [7,8,29]. Therefore, in this paper, no model discrepancy term is included in the proposed methods, and the model form error is accounted for through model validation within the prediction framework that will be described in Sec. 4.

2.2 Calibration of Combined Aleatory and Epistemic Parameters. In some calibration problems, the available data are collected from multiple specimens, and the parameters cannot be measured directly for any specimen. In this situation, the model parameters are not simply unknown epistemic values; rather, the properties may also be varying across the tested specimens. If this variability is not explicitly considered (e.g., by applying the formulation of Sec. 2.1), all of the variation in the output must again be attributed to measurement noise. This treatment misrepresents the actual underlying parameter uncertainty by forcing the parameters to take the same values across all the experiments. In particular, by excluding the aleatory component of uncertainty, the posterior distribution is likely to underestimate and incorrectly characterize variability, and this effect is only exacerbated as more data are collected to reduce the epistemic component of uncertainty. Failure to adequately account for the variability leads to underestimates of uncertainty in future predictions of the model, which use the estimated parameter values, and these predictions will not be conservative to potential outcomes.

Therefore, the aleatory uncertainty should also be included in the calibration activity. This can be accomplished by assigning a probability distribution to represent the aleatory uncertainty and then applying Bayes' theorem to estimate its distribution parameters. When data are limited, the appropriate probability distribution to select may not be known, which forces an arbitrary choice to be made. To minimize the impact of this choice, it is desirable to select a distribution of the aleatory uncertainty that is capable of describing a wide range of potential forms for the variability. In this paper, the Johnson family of distributions is applied since it is able to reproduce any shape that can be described by a unique set of the first four statistical moments.

2.2.1 Johnson Distribution Family. The Johnson family of probability distributions [40] can represent a wide range of shapes of probability distributions. It is composed of four types of distributions: the normal, lognormal, bounded, and unbounded. Each of these forms is achieved by applying a specific normalizing transformation that depends on the four distribution parameters γ , δ , λ , and ξ . Since the normal and lognormal distributions are special cases that are rarely encountered when estimating the distributional form from observations, the focus of this paper is restricted to the bounded and unbounded forms. For the unbounded system, a general random variable X is transformed to a standard normal random variable Z by applying

$$Z = \gamma + \delta \ln \left\{ \left(\frac{X - \xi}{\lambda} \right) + \left[\left(\frac{X - \xi}{\lambda} \right)^2 + 1 \right]^{\frac{1}{2}} \right\}, \quad -\infty < X < \infty \quad (5)$$

Similarly, for the bounded system

$$Z = \gamma + \delta \ln \left(\frac{X - \xi}{\xi + \lambda - X} \right), \quad \xi < X < \xi + \lambda \quad (6)$$

Setting $X = \theta$ in either Eq. (5) or (6) transforms $\theta = \theta(z, \mathbf{p})$ where $\theta(\cdot)$ is the inverse mapping of Eq. (5) or (6), and $\mathbf{p} = [\xi, \lambda, \gamma, \delta]$ are hyperparameters that will be modeled as random variables to be estimated. This estimation will require that Eqs. (2) and (3) be restated in terms of θ and \mathbf{p} .

2.2.2 Bayesian Johnson Parameter Estimation From Direct Observations. While alternate methods exist to estimate the Johnson hyperparameters \mathbf{p} [41], a Bayesian estimation approach [42] is more suitable when observations of the random variable θ are sparse. The resulting probabilistic description of \mathbf{P} indicates that the set of observations can be supported by many probabilistic descriptions of θ , namely, those supported by any sample of the vector of random variables \mathbf{p} . For this Bayesian estimation problem, Eqs. (2) and (3) become

$$\theta = \theta(z, \mathbf{p}) \quad (7)$$

$$Z \sim N(0, 1), \quad \text{i.i.d} \quad (8)$$

where again, $\theta(\cdot)$ is the inverse of either Eq. (5) or (6). Then, it can be shown that the likelihood $L(\mathbf{p}) = f(\theta|\mathbf{p})$, given n independent samples of the random variable θ , can be expressed for the unbounded system and bounded system, respectively, as follows:

$$\begin{aligned} L(\gamma, \delta, \lambda, \xi) &= \prod_{i=1}^n \frac{\delta}{\lambda \sqrt{2\pi}} \times \frac{1}{\sqrt{1 + \left(\frac{\theta_i - \xi}{\lambda} \right)^2}} \\ &\times \exp \left\{ -\frac{1}{2} \left[\gamma + \delta \ln \left(\frac{\theta_i - \xi}{\lambda} + \sqrt{1 + \left(\frac{\theta_i - \xi}{\lambda} \right)^2} \right) \right]^2 \right\} \end{aligned} \quad (9)$$

$$\begin{aligned} L(\gamma, \delta, \lambda, \xi) &= \prod_{i=1}^n \frac{\delta}{\lambda \sqrt{2\pi} \left(\frac{\theta_i - \xi}{\lambda} \right) \left(1 - \frac{\theta_i - \xi}{\lambda} \right)} \\ &\times \exp \left\{ -\frac{1}{2} \left[\gamma + \delta \ln \left(\frac{\frac{\theta_i - \xi}{\lambda}}{1 - \frac{\theta_i - \xi}{\lambda}} \right) \right]^2 \right\} \end{aligned} \quad (10)$$

The choice of whether to use the form of Eq. (9) or (10) is determined from the statistics of the set of observations of θ . The second, third, and fourth central moments of the data set together uniquely specify the appropriate Johnson system [41]. By applying MCMC methods with the selected likelihood function, a set of samples can be generated from the joint posterior distribution of the Johnson distribution parameters, denoted as $f_{\mathbf{P}}(\mathbf{p})$. Each sample of the parameters corresponds to a particular realization of the distribution of the aleatory variability for the underlying random variable θ . Taken together, the posterior samples of the distribution parameters yield a family of distributions for the random variable. In some situations it may be useful to reduce the family of distributions to a single distribution that contains both aleatory and epistemic sources of uncertainty. This distribution is generated by integrating the distribution of θ conditioned on a particular parameter sample $f_{\Theta}(\theta|\mathbf{P} = \mathbf{p})$ over the domain D of the distribution parameters [43]

$$f_{\Theta}(\theta) = \int_D f_{\Theta}(\theta|\mathbf{P} = \mathbf{p}) f_{\mathbf{P}}(\mathbf{p}) d\mathbf{p} \quad (11)$$

The resulting distribution of θ is referred to here as the unconditional distribution. An example (estimation of Young's modulus E in the challenge problem solution in Sec. 5) of these two representations of combined aleatory and epistemic uncertainty is shown in Fig. 1. Note that the unconditional distribution is much more conservative to future observations than the maximum a posteriori (MAP) estimate of the distribution parameters since the MAP estimate does not account for the epistemic uncertainty in \mathbf{P} . However, after integration, the aleatory and epistemic sources of uncertainty can no longer be separated from each other.

2.2.3 Bayesian Johnson Parameter Calibration From Output Observations. When there are no direct observations of the parameter, the estimation problem is solved by inference from observations of the output, similar to Sec. 2.1. A separate Johnson distribution is assumed to describe the aleatory uncertainty of each unknown parameter. Again, an uncertainty model that relates prediction and observation must be assumed. For this scenario,

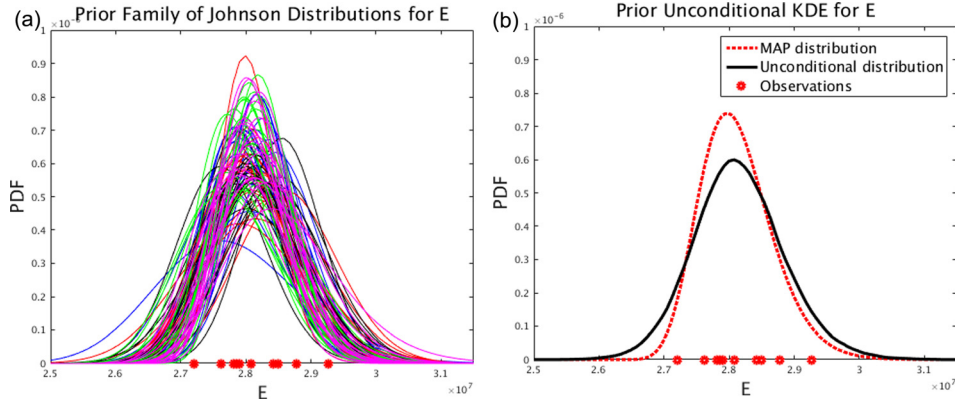


Fig. 1 Combined aleatory and epistemic uncertainty represented as a family of distributions and as an unconditional density

each individual parameter θ_k from the vector of parameters θ is described according to Eqs. (7) and (8). Thus, Eqs. (2) and (3) become

$$y_{dij} = y_m(\mathbf{x}_j, \theta(\mathbf{z}, \mathbf{p})) + e_{dij} \quad (12)$$

$$Z_k \sim N(0, 1), \quad \text{i.i.d.} \quad (13)$$

$$E_{dij} \sim N(0, \sigma_d), \quad \text{i.i.d.} \quad (14)$$

Here, \mathbf{Z} represents a vector of standard normal random variables, each corresponding to the description of a particular model parameter θ_k . The indices i and j hold the same meaning as in Sec. 2.1.

The likelihood derived from Eqs. (12)–(14) for a particular set of the Johnson hyperparameters can be evaluated via a nested Monte Carlo sampling procedure. The aleatory uncertainty of the unknown parameter vector θ must be sampled and propagated through the model. An arbitrary number of samples (each denoted with superscript i) of the aleatory variables can be drawn and propagated through the model at each input condition \mathbf{x}_j ; the appropriate number of samples depends on the allowable computational expense for the simulation

$$y_{mj}^i = y_m(\mathbf{x}_j, \theta^i) \quad (15)$$

The set of samples of Y_{mj} for a given vector \mathbf{p} of the Johnson parameters represents an empirical estimate of the conditional density of the model output $f_{Y_{mj}}(y_{mj}|\mathbf{P} = \mathbf{p})$. By assuming the relationship between model prediction and observation given by Eqs. (12)–(14), the likelihood of a particular \mathbf{p} is computed over a set of n_j independent observations at each input condition

$$L(\mathbf{p}) = \prod_{j=1}^m \prod_{i=1}^{n_j} f_{Y_{mj}}(y_{dij} - e_{dij}|\mathbf{P} = \mathbf{p}) \quad (16)$$

By applying Bayes' theorem as in Eq. (1) for \mathbf{P} , this likelihood is used to update the joint distribution $f_{\mathbf{P}}(\mathbf{p}|y_d)$. Note that if direct observations of θ are also available, this information can be used within the method described in Sec. 2.2.2 to obtain a prior distribution $f_{\mathbf{P}}(\mathbf{p})$.

The methods described in Sec. 2 provide a calibration framework for several different observation scenarios. It is important to recognize which method is consistent with the assumptions underlying the observed data. A Bayesian scheme is capable of handling any of these scenarios, but the analyst must first determine whether the calibration parameters are subject to only epistemic uncertainty or combined aleatory and epistemic sources and then apply the appropriate method. This decision comes from physical

knowledge of the tests and the corresponding specimens and expert judgment about their relationship to one another.

3 Probabilistic Model Validation

Before using a calibrated model for a prediction of interest, the model should be independently tested against an additional set of observations to assess its predictive capability. Typically, the validation experiments are closer to the domain of the application than the calibration experiments, such that they are more relevant to the prediction. In such a scenario, the assessment will provide stronger evidence of the adequacy of the model. However, since the model performance is never perfect, it is useful to compute a probability measure for the validation performance so that the validation quality can be directly incorporated into the prediction as will be shown in Sec. 4. The model reliability metric r [17,18] is one such probability measure, and it is the focus of the methods demonstrated in this paper. Treatments of this metric are demonstrated for situations with only epistemic uncertainty and for combined aleatory and epistemic uncertainty.

3.1 Model Reliability in the Presence of Epistemic Uncertainty. The model reliability metric directly measures the discrepancy between prediction and observation. It is particularly intended to penalize epistemic uncertainty in either the model or the observations that are used to assess the model. It is defined as the probability of the difference (Δ) between observed data (Y_d) and model prediction (Y_m) being less than a given tolerance limit ϵ

$$r = \Pr(-\epsilon < \Delta < \epsilon), \quad \Delta = Y_d - Y_m \quad (17)$$

In Eq. (17), the experimental observation at a given value of the input is treated as a random variable due to measurement error, and the model output at a given value of the input is a distribution resulting from the propagated posterior distribution of the model parameters. In this case, it is assumed that the parameter uncertainty is purely epistemic uncertainty due to insufficient calibration data. The difference between the prediction and observation Δ is also a random variable at particular input, and its distribution can be obtained from the probability distributions of Y_d and Y_m . Then, the model reliability metric (at a given value of the input) is computed by integration of the distribution of Δ

$$r = \int_{-\epsilon}^{\epsilon} f_{\Delta}(\omega) d\omega = F_{\Delta}(\epsilon) - F_{\Delta}(-\epsilon) \quad (18)$$

Since the distribution of Δ is obtained by independently sampling the prediction and observation, the metric only measures model bias. The sources of epistemic uncertainty (measurement noise in the data and parameter uncertainty in the model) are

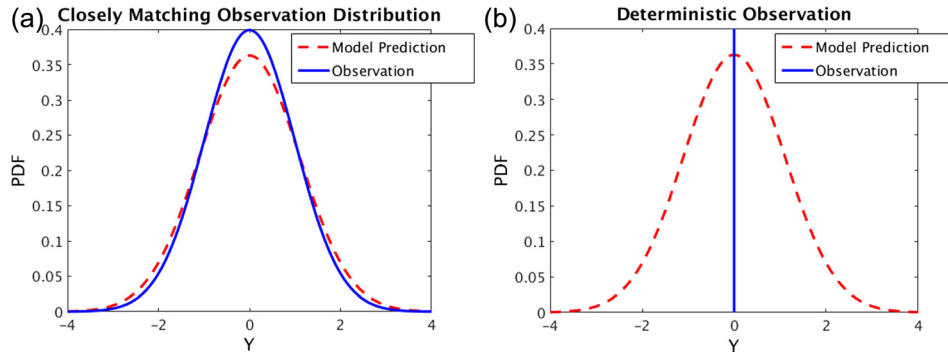


Fig. 2 For $\epsilon = 2$, the model reliability for the closely matching distributions Y_d and Y_m ($r = 0.86$) is lower than for the deterministic observation with no measurement noise and the same distribution of Y_m ($r = 0.95$) because the probability of large bias between the uncertain deterministic prediction and observation is greater when there is more uncertainty

independent and should not be expected to cause the distributions of Y_d and Y_m to take the same shape. Rather, only the expected discrepancy between two deterministic samples (i.e., the bias) is of interest. This behavior is illustrated in Fig. 2.

The model reliability metric can be improved by decreasing either the measurement uncertainty or the parameter uncertainty when the means are unbiased. This behavior is logical for epistemic uncertainty sources, but when there is also aleatory uncertainty that is common to the prediction and observation, this computation of model reliability is not suitable. In such a case, it is desirable for the model to be able to reproduce observed variability. Therefore, unless the dependency between the prediction and observation samples is known (it is rarely known in practice), the strategy given by Eqs. (17) and (18) cannot be applied to comparisons that include aleatory variability.

3.2 Model Reliability With Combined Aleatory and Epistemic Uncertainty. In order to incorporate aleatory variability into a formulation aimed at epistemic uncertainty, the formulation of Sec. 3.1 must be modified. In the presence of variability, the parameter uncertainty that leads to a stochastic model prediction has two components: an aleatory and an epistemic contribution. Section 2 demonstrated how aleatory and epistemic sources can be separated during calibration. Specifically, the uncertainty about distribution parameters is the epistemic contribution, and each sample of the distribution parameters represents a particular realization of the aleatory variability. When parameter uncertainty with combined aleatory and epistemic uncertainty is propagated through the model, the model prediction is similarly described by a family of distributions representing combined aleatory and epistemic uncertainty.

In this context, there are two primary criteria for model validation: (1) minimum model bias and (2) accurate prediction of observed variability. Since the model reliability metric is intended as a discrepancy (bias) criterion, it is best suited to handle the first of these two criteria. However, if the aleatory uncertainty is included in the model prediction, the prediction and observation are inherently correlated through their shared dependence on the underlying aleatory variables. Therefore, the prediction and observation must be sampled jointly to perform the computation correctly, but their dependence is unknown. If, by necessity, they are instead sampled independently, the reliability metric is artificially lowered since the prediction uncertainty is increased.

To address this issue, attention can be restricted to only the mean prediction of the model. If the model is unbiased, the observations, which also have both aleatory and epistemic uncertainty, are expected to scatter around the mean prediction of the model. Therefore, the model prediction distribution Y_m is integrated over the aleatory component of uncertainty in θ (represented by the vector of random variables \mathbf{Z}) to obtain the mean prediction

$$f_{M_{Y_m}}(\mu_{Y_m}) = \int_{-\infty}^{\infty} f_{Y_m}(y_m | \mathbf{Z} = \mathbf{z}) f_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z} \quad (19)$$

The distribution $f_{M_{Y_m}}(\mu_{Y_m})$ of the mean prediction is dependent on only epistemic uncertainty in the distribution parameters. Therefore, it can be analyzed in the same manner as the distribution of Y_m in Sec. 3.1 and compared against the set of observations

$$r = \Pr(-\epsilon < \Delta_M < \epsilon), \quad \Delta_M = Y_d - M_{Y_m} \quad (20)$$

However, the distribution Y_d also includes aleatory variability while M_{Y_m} does not, so the differences between observation and mean prediction may have large variance. To account for this issue, the second validation criterion (prediction of observed variability) can be included in the assessment through the choice of the tolerance ϵ .

One approach is to set the tolerance based on the average aleatory uncertainty in the family of model predictions (i.e., an expectation taken over the space of the epistemic distribution parameters). In this scenario, the outcome of the assessment demonstrates whether the predicted variability is a good predictor of the spread in the observations. For example, if $\epsilon = 2 * E_P(\sigma_{Y_m})$ as will be shown in the example of Sec. 5, 95% of the observations are expected to fall within $\pm\epsilon$ when the predicted variability is equivalent to the observed variability. Note that choosing ϵ based on the aleatory uncertainty in the prediction is also conservative. That is, the model attains high reliability when the predicted variability overestimates observed variability (high tolerance) and low reliability when the predicted variability underestimates observed variability (low tolerance). When the model has low reliability, additional conservatism is added in by expanding the range of the parameters as will be shown in Sec. 4. This conservatism is often desirable because it is preferable for the final prediction to predict a wider range of outcomes than will be encountered in reality. Unexpected scenarios (i.e., those outside the prediction uncertainty) are most likely to force a system outside its normal operating regime because they are not explicitly accounted for in design.

4 Including the Validation Result in Prediction

The calibration methodology of Sec. 2 and the validation approach of Sec. 3 demonstrate how epistemic and aleatory uncertainty can be separated for propagation and assessment. Since the validation result is interpreted as a probability measure, it can be used to modify the posterior parameter distributions in order to add additional conservatism to the prediction to account for model form error. The underlying assumption of the proposed approach is that parameters calibrated using imperfect models should not be fully trusted when they are propagated forward to the prediction

stage. Therefore, the probabilistic validation result is treated as a weight for the calibrated posterior distribution, and the remaining weight is given to an alternate distribution that may come from prior information or expert opinion.

Low model reliability signifies only partial support for the posterior parameters, which does not necessarily imply support for an alternate distribution. However, the posterior distribution always results from epistemic uncertainty reduction during the calibration phase. The alternate distribution should have a wider range of support that has not been updated from observation data. Attributing weight to this alternate distribution accounts for the possibility that the model form error caused the posterior parameter estimates to be biased and overconfident on a narrower range of values. In this situation, a wider range of parameter values should be included in the distribution that is propagated to the prediction of interest. Using the value of r obtained in the model reliability assessment as a weight on the posterior and the complementary probability $(1 - r)$ as a weight on the alternate distribution leads to the following formulation:

$$f_{\Theta}(\theta|y_d^C, y_d^V) = r f_{\Theta}(\theta|y_d^C) + (1 - r) \hat{f}_{\Theta}(\theta) \quad (21)$$

where $f_{\Theta}(\theta|y_d^C)$ is the calibrated posterior distribution resulting from the calibration observations y_d^C , $\hat{f}_{\Theta}(\theta)$ is the alternate distribution for the parameters, and $f_{\Theta}(\theta|y_d^C, y_d^V)$ is the predictive parameter distribution that is propagated to the prediction of interest. This predictive distribution depends on the calibration observations and the validation observations y_d^V , which are used to obtain r .

Once the predictive parameter distribution is obtained, it is propagated through the system model in the regime of interest to obtain the stochastic prediction. This distribution should always have at least as much uncertainty as the posterior distribution and no more uncertainty than the alternate distribution. The relative contributions of these two components depend on the quality of the model reliability assessment. Note that when the reliability assessment is conducted over a range of different input conditions, the value of r used in Eq. (21) is an aggregation across all input conditions. If all the validation input conditions are equally relevant to the prediction of interest, a simple averaging of the validation results is sufficient. However, in some cases, some validation experiments may be more relevant to the prediction regime, and the validation results from these experiments may be given higher weight.

5 Sandia Challenge Problem Results

The proposed methods of Secs. 2–4 are demonstrated in this section using data and models provided by the Sandia National Laboratories V&V challenge problem [28]. The information provided for the problem includes many heterogeneous sources that must be integrated within the proposed framework to make a prediction on system performance. In particular, the prediction goal addressed in this demonstration is reliability assessment of in-service storage tanks (shown in Fig. 3) of “mystery liquid” that are subject to pressure loads as well as loads from the liquid itself. Available information to make the assessment includes the

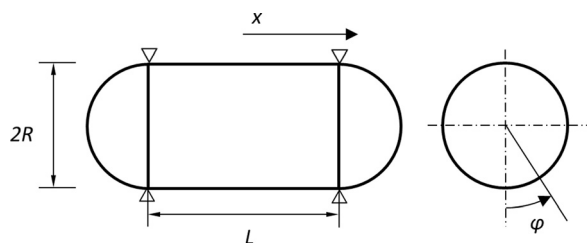


Fig. 3 Idealized diagram of the tanks [28]

following six data sets: (1) legacy data from the manufacturer, (2) material coupon tests in a lab environment, (3) liquid characterization tests in a lab environment, (4) full tank geometry measurements in a lab environment, (5) full tank pressure loading tests in a lab environment, and (6) full tank displacement measurements in the production environment with pressure and liquid. The provided computational model is capable of predicting the output for any of the test scenarios. The model predicts displacement w and stress σ and depends on the following quantities: axial location on the tank x , circumferential angle from the tank centerline ψ , gauge pressure P , liquid-specific weight γ , liquid height H , Young’s modulus E for the tank material, Poisson’s ratio ν for the tank material, length L , radius R , and wall thickness T .

The provided data provide a good approximation to the type of information that would be available in a realistic problem since it is subjected to many common sources of uncertainty. Only a limited number of tests were conducted, and all the data are also susceptible to measurement errors (approximate tolerances are known in some cases). Therefore, the available measurements are both uncertain and sparse (e.g., spatial variability along the tank is explored but unit-to-unit variability information is limited). Furthermore, the quantity of interest (stress) is not measured directly, and decision makers must rely instead on measurements and predictions of a related quantity (displacement) to develop confidence. In addition, direct measurements of some properties are made only for tanks that are pulled out of service. Since the goal is to make a prediction for the in-service tanks, additional assumptions are needed in order to extrapolate knowledge to these other tanks.

In order to pose this problem within the framework presented in Secs. 2–4 of this paper, the first important step is to establish the hierarchy of data and how it will be used for the calibration, validation, and prediction activities. Only data sets 5 and 6 have information about the tank response under an applied load. Therefore, only these two data sets can be used to calibrate and validate the input/output relationship in the model. Since data set 6 represents the full usage condition of the tank (pressure and liquid) while data set 5 contains only pressure loading information, data set 6 is considered more relevant to the prediction condition of interest. As a result, data set 5 is used to calibrate relevant parameters, and data set 6 is used to validate the predictive capability of the calibrated model. Note that raw data for data set 3 are not available, and the data set is only used to construct an empirical relationship between liquid composition and the specific weight γ . Therefore, other measurements of liquid composition are treated as equivalent to measurements of γ , except that the empirical model introduces an additional source of input measurement uncertainty. Since the loads from the pressure and/or liquid components (P , γ , and H) and the location variables (x and ψ) are measured in data sets 5 and 6, these variables are treated as input conditions (i.e., “ x ” in the formulation of Sec. 2). The remaining variables (E , ν , T , L , and R) are not measured for the calibration and validation scenarios, so they are treated as calibration parameters (i.e., “ θ ” in the formulation of Sec. 2). The direct measurements on material properties (data set 2) and tank geometry (data set 4) represent useful prior information for calibrating these parameters. The manufacturer data (data set 1) are used to formulate an alternate distribution for these parameters, which is applied within the framework of Sec. 4. The workflow for the proposed solution strategy is depicted in Fig. 4.

5.1 Parameter Calibration Results. As mentioned in the previous discussion, one complication (realistic for practical applications) is that measurements are taken on specific tanks, but the assessment needs to be made for other tanks. In such a scenario, it is expected that the calibration parameters vary naturally from tank-to-tank; however, no probabilistic model of this variability is known. Therefore, the model parameters are subject to both aleatory and epistemic uncertainty, and the calibration

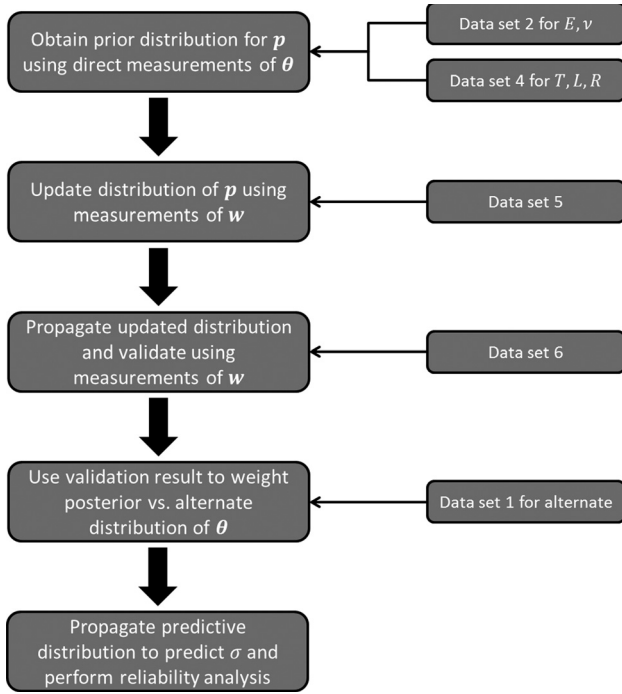


Fig. 4 Diagram of workflow and data usage for the proposed solution strategy

techniques of Sec. 2.2 should be applied rather than the technique of Sec. 2.1 (which would assume that the parameters are deterministic tank-to-tank). Since the Johnson family of distributions provides a flexible means of capturing the aleatory variability (as described in Sec. 2.2.1), the Johnson distribution model is applied to the aleatory uncertainty for each of the five calibration parameters. Note that this choice does increase the complexity of the problem since it expands the dimension of the parameter set to be calibrated. Other distribution types could also be chosen to represent the aleatory variability if there is some prior belief about the

form of the variability. If the Johnson family is chosen, the results must be carefully checked to ensure that they are converged because there are a potentially large number of parameter sets that lead to very similar descriptions of the uncertainty. Fortunately, MCMC methods are very well suited to handle correlated parameters since the samples are inherently taken jointly across the set of calibration parameters.

The calibration variables within the framework of Sec. 2.2 are the four Johnson distribution parameters that define the aleatory distribution for each of the five model parameters. Thus, there are a total of 20 calibration parameters. Both direct observations of the model parameters (data set 2 for E , ν , and T and data set 4 for L and R) and observations of the output that depend on the model parameters (data set 5) are available. Therefore, the direct observations can be used to obtain a prior family of distributions for each model parameter according to the approach of Sec. 2.2.2, and then the Johnson distribution parameters for each model parameter can be jointly updated with the output observations of data set 5 according to the approach of Sec. 2.2.3. Since this step requires a large number of model evaluations, a GP surrogate model is used in place of the underlying model to improve computational efficiency. The results of this procedure are shown in Fig. 5, along with the manufacturer estimates (not used in calibration) for a comparison benchmark. Only the Johnson distributions corresponding to the MAP estimates of the distribution parameters are shown in the figures for simplicity of illustration. The priors from data sets 2 and 4, as well as the posterior from the joint calibration, are actually families of distributions due to the epistemic uncertainty in the Johnson distribution parameters. The unconditional distributions, which are obtained from the joint posteriors according to Eq. (11), represent the combination of epistemic and aleatory sources of uncertainty.

Note that the direct observations on the model parameters are at different locations within a particular tank; they are not observations on different tanks. Therefore, in order to update with the output data from another tank, an assumption of ergodicity is needed. That is, it is assumed that the variability across locations of a particular tank follows approximately the same distribution as variability from tank-to-tank. This assumption is certainly not true in general, but even if the assumption is invalid, the result of the

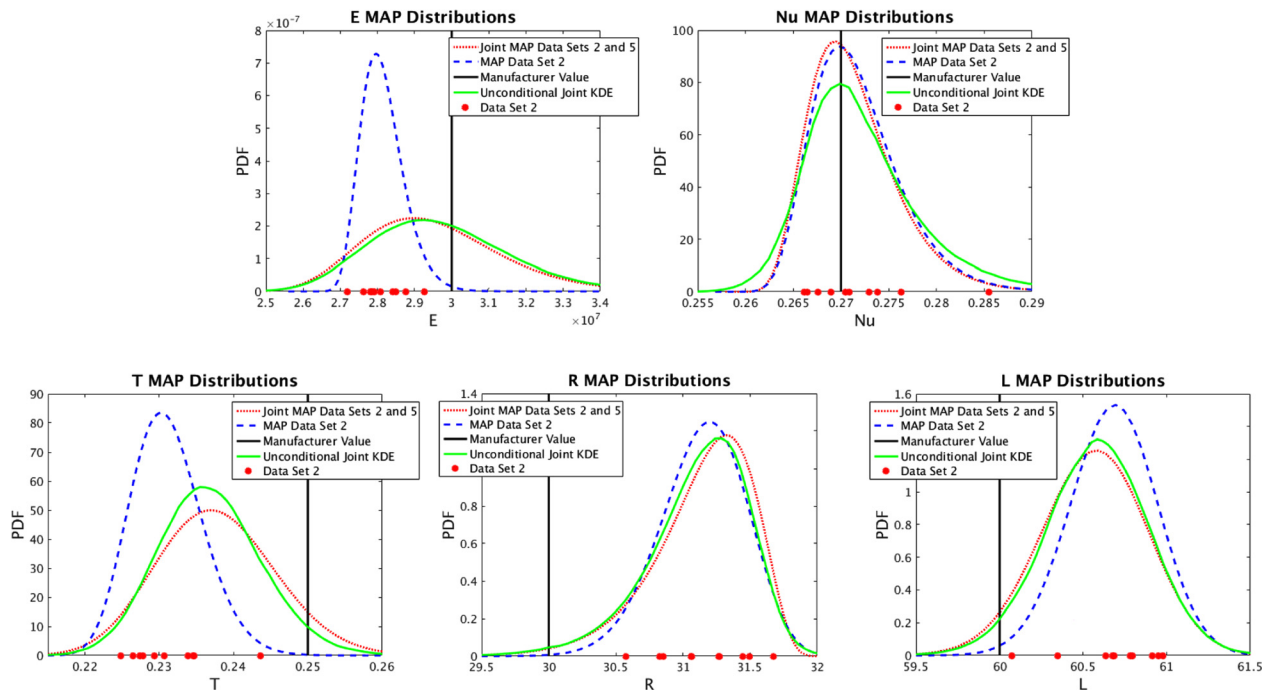


Fig. 5 Results of model parameter calibration

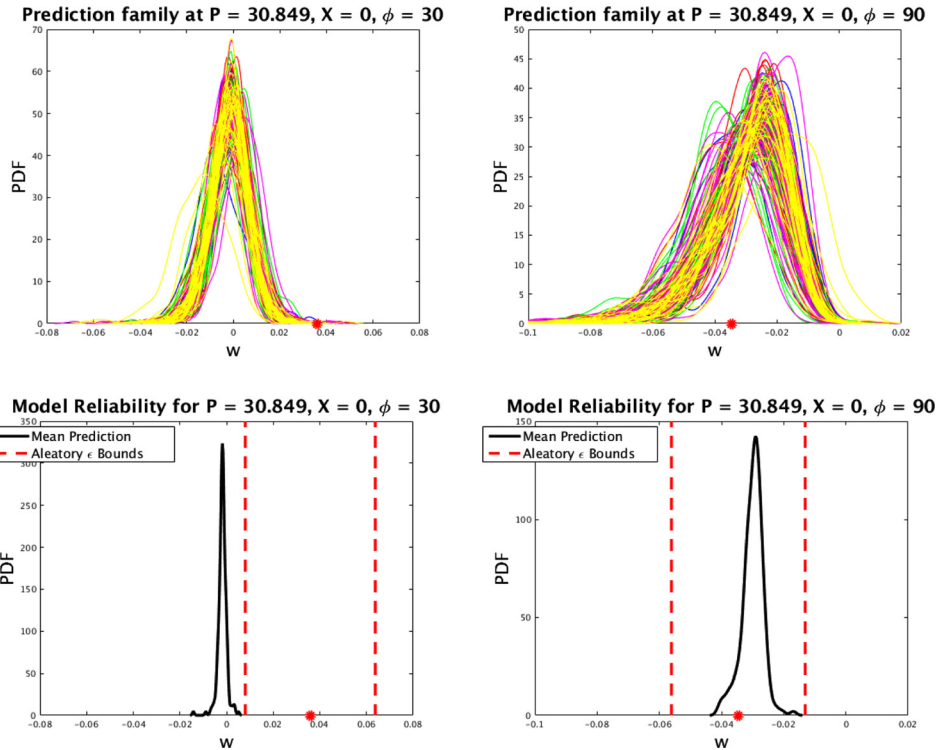


Fig. 6 Sample computations of model reliability in the presence of combined aleatory and epistemic uncertainty. For this particular tank prediction, $r = 0$ for $X = 0$ and $\psi = 30$ (left) and $r = 0.98$ for $X = 0$ and $\psi = 90$ (right).

calibration under this assumption will be a distribution of the aleatory variability that best represents the combination of variability within a tank and tank-to-tank variability. The negative effect of an invalid assumption of ergodicity is that the distribution may be biased toward the particular tank that has most measurements (since all observations are treated with equal weight in the likelihood function).

5.2 Model Validation Results. Once the model parameters have been calibrated, they are each represented by a posterior family of distributions that includes both aleatory and epistemic sources of uncertainty. These families of distributions for the model parameters can be propagated through the model to obtain a family of distributions for the model output Y_m . Since this distribution contains combined aleatory and epistemic uncertainty, the modified approach to computing the model reliability metric proposed in Sec. 3.2 must be applied. The displacement w is predicted at all locations where it is measured in data set 6, and a family of model predictions is obtained at each location for each of the four field-tested tanks. For each of the metric computations, the tolerance is set according to $\epsilon = 2 * E_P(\sigma_{Y_m})$ for the particular family of distributions for Y_m . Two examples of transforming a family of distributions for Y_m into a single distribution for M_{Y_m} are shown in Fig. 6. The distribution of the mean prediction is then integrated over the interval $[y_d - \epsilon, y_d + \epsilon]$ to obtain the reliability at each location.

At the two different tank locations depicted in Fig. 6, the model reliability is significantly different. In fact, the model reliability varies significantly from one location to another across the entire tank. An average of the results taken across the four tanks at each location is shown in Fig. 7. In general, the model reliability tends to increase as x increases (i.e., nearer the end caps of the tank), and it is maximized at a circumferential angle of about 60 deg. Understanding the predictive capability of the model as a function of location is useful for isolating physics inadequacies in the model and making improvements. However, if the model is used

in prediction as is, a global measure of reliability is useful to modify the parameter distributions according to the approach in Sec. 4. In this case, the mean model reliability across all the locations is approximately 0.4.

5.3 Prediction and Reliability Assessment Results. The validation assessment does not give high confidence in the predictive capability of the model across all the tank locations. Additional conservatism should be added to the prediction; one approach is to acknowledge a potentially wider range of uncertainty in the parameters. In some of the posterior distributions, particularly the posteriors of the geometry parameters, the inferred aleatory distributions are significantly biased from the manufacturer data. One potential explanation is that the ergodicity assumption is invalid, and the posterior distributions of the model

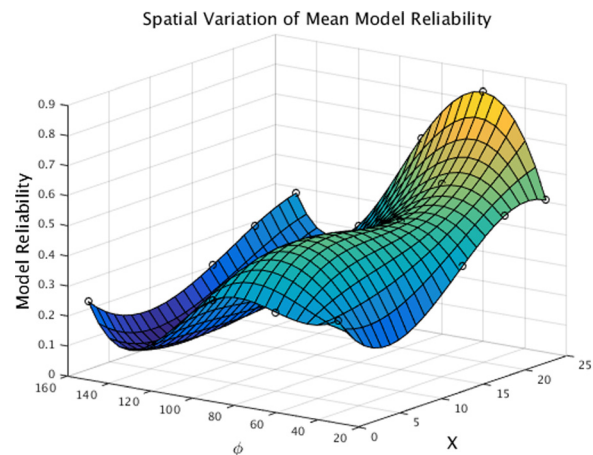


Fig. 7 Spatial variation of model reliability averaged across four tank predictions

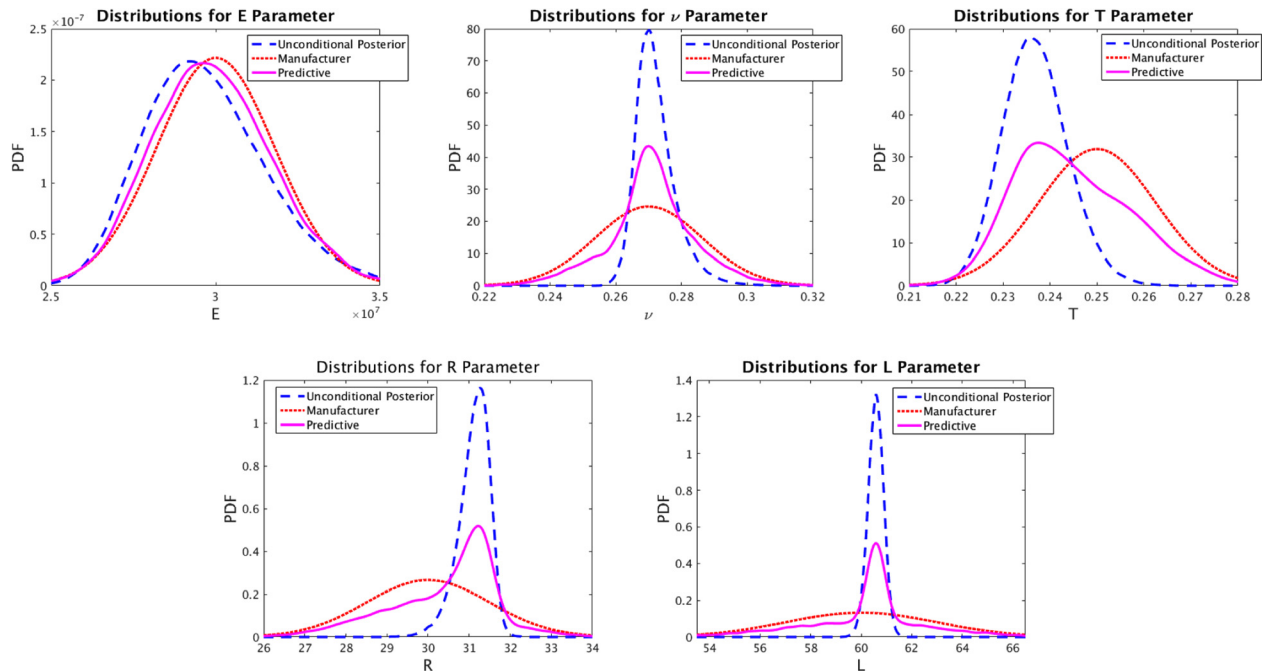


Fig. 8 Expansion of parameter uncertainty to account for model form error and insufficient variability information

parameters are the result of overweighting the observations for the particular tanks where measurements were available. In this scenario, other parameter values that have not been accounted for may be realized in other tanks. To account for this possibility, some weight should be given to the manufacturer specifications. However, it is challenging to incorporate this information because the manufacturer data include only deterministic estimates of the parameter.

Some additional expertise from past experience is needed in order to translate these deterministic estimates into an estimate of uncertainty. For example, past experience with other similar fabrication efforts might provide an estimate of expected variability. In the absence of specific experience, one option is to consult the literature for information about similar materials or assemblies. For example, U.S. Department of Commerce building codes [44] indicate that a coefficient of variation (COV) of 0.06 is typical for material properties, and a COV of 0.05 is typical for geometric parameters. For illustration, alternate distributions for the parameter uncertainty with mean equal to the manufacturer specification are assumed normal with variance obtained from these literature COV estimates. Following the approach in Eq. (21), a weight of 0.4 (the overall average model reliability) is given to the posterior parameter distributions, and a weight of 0.6 is given to the alternate distribution to achieve the more conservative parameter distributions that are shown in Fig. 8.

Once the uncertainty in the model parameters is expanded, the distributions must be propagated through the model in order to predict the behavior of the remaining tanks. The predictions of the maximum stress in the tank material can then be compared against the yield stress of the material in order to assess the tank reliability. However, the yield stress of the material is also variable with unknown aleatory distribution. By again applying the method of Sec. 2.2.2 and using the direct observations of yield stress from data set 2, a family of distributions is obtained for the yield stress of the material. By applying Eq. (11), this family is condensed to a single distribution that includes aleatory and epistemic uncertainty in the yield stress. The probability of maximum predicted stress exceeding the yield stress is then computed according to the Monte Carlo sampling from the distributions shown in Fig. 9.

The reliability analysis predicts a probability of 0.0075 that the maximum stress will exceed the yield stress of the material. This

probability may or may not be acceptable to the decision maker, but it should be noted that this value itself is very uncertain. This value is very sensitive to assumptions that were made in the analysis (in particular, the assumption of ergodicity and the assumption of an alternate distribution from the manufacturer data). This computation is primarily performed for the purpose of illustration, and the results should not be blindly trusted without first collecting some additional information to check the validity of these assumptions. In other words, the quality of these assumptions dramatically influences the credibility of the prediction, and in practice, it is very dangerous to make decisions based on an assessment without first testing the major assumptions. As a starting point, sensitivity analysis can provide some insight about how much a change in the assumptions impacts the final assessment. This knowledge may then help to guide additional activities aimed at reducing the sources of uncertainty to which the prediction is most sensitive.

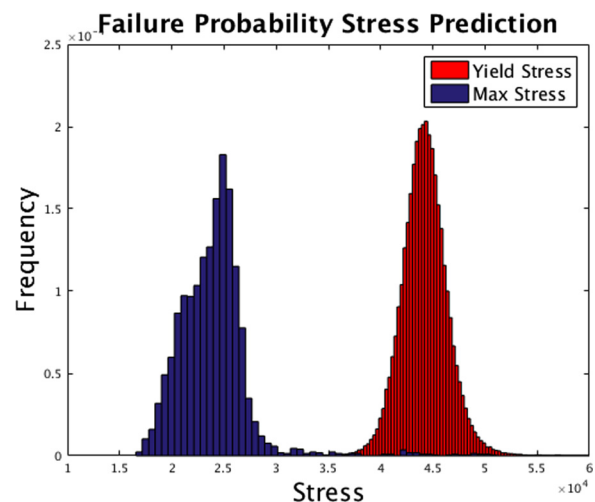


Fig. 9 Reliability assessment based on the distributions of maximum predicted stress and material yield stress

6 Conclusion

This paper proposes and demonstrates a comprehensive methodology for calibration, validation, and prediction under uncertainty. The proposed methods particularly focus on novel approaches to include and separate aleatory and epistemic sources of uncertainty in calibration and validation. For existing calibration and validation approaches (in particular, Bayesian calibration methods and the model reliability metric for validation) to be appropriate, they must be cast in a way that is consistent with the sources of uncertainty in the available data. A primary goal of the methodology is to be conservative when quantifying the prediction uncertainty. This paper adds conservatism by imposing a stricter validation tolerance on overconfident models (i.e., models that predict low uncertainty) and by expanding parameter uncertainty to account for model form error. The proposed approach incorporates heterogeneous data sources in a hierarchical fashion based on the relevance of the information to the prediction of interest. Decisions about how to handle the available information are always subject to some expert judgment and knowledge from prior experience, and this expertise may be qualitative. In problems where data are particularly limited and uncertain, some additional assumptions are required in order to integrate the available information. While the assumptions may be unavoidable, the assessment is not credible unless the major assumptions are tested, and the sensitivity of the predictions to these assumptions is understood.

Acknowledgment

This paper is based upon the research supported by the Sandia National Laboratories (Contract No. BG-7732) and the U.S. Department of Energy (National Nuclear Security Administration) Award No. DE-FC52-08NA28617 to the Purdue University (Principal Investigator: Professor Jayathi Murthy and Director: Professor Alejandro Strachan) and subaward to the Vanderbilt University. The support is gratefully acknowledged.

References

- [1] Roy, C., and Oberkampf, W., 2011, "A Comprehensive Framework for Verification, Validation, and Uncertainty Quantification in Scientific Computing," *Comput. Methods Appl. Mech. Eng.*, **200**(25–28), pp. 2131–2144.
- [2] Romero, V., Luketa, A., and Sherman, M., 2010, "Application of a Versatile 'Real-Space' Validation Methodology to a Fire Model," *J. Thermophys. Heat Transfer*, **24**(4), pp. 730–744.
- [3] Hills, R. G., and Leslie, I. H., 2003, "Statistical Validation of Engineering and Scientific Models: Validation Experiments to Application," *Sandia Technical Report No. SAND2003-0706*.
- [4] Trucano, T., Swiler, L., Igusa, T., Oberkampf, W., and Pilch, M., 2006, "Calibration, Validation, and Sensitivity Analysis: What's What," *Reliab. Eng. Syst. Saf.*, **91**(10–11), pp. 1331–1357.
- [5] Higdon, D., Kennedy, M., Cavendish, J., Cafeo, J., and Ryne, R., 2004, "Combining Field Data and Computer Simulations for Calibration and Prediction," *SIAM J. Sci. Comput.*, **26**(2), pp. 448–466.
- [6] Sankararaman, S., and Mahadevan, S., 2012, "Comprehensive Framework for Integration of Calibration, Verification and Validation," *AIAA Paper No. 2012-1366*.
- [7] Arendt, P. D., Apley, D. W., and Chen, W., 2012, "Quantification of Model Uncertainty: Calibration, Model Discrepancy, and Identifiability," *ASME J. Mech. Des.*, **134**(10), p. 100908.
- [8] Ling, Y., Mullins, J., and Mahadevan, S., 2014, "Selection of Model Discrepancy Priors in Bayesian Calibration," *J. Comput. Phys.*, **276**, pp. 665–680.
- [9] Hartmann, C., Smeyers-Verbeke, J., Penninckx, W., Heyden, Y. V., Vankeerberghen, P., and Massart, D., 1995, "Reappraisal of Hypothesis Testing for Method Validation: Detection of Systematic Error by Comparing the Means of Two Methods or of Two Laboratories," *Anal. Chem.*, **67**(24), pp. 4491–4499.
- [10] Hills, R. G., and Trucano, T. G., 1999, "Statistical Validation of Engineering and Scientific Models: Background," *Sandia Technical Report No. SAND99-1256*.
- [11] Rebba, R., and Mahadevan, S., 2006, "Validation and Error Estimation of Computational Models," *Reliab. Eng. Syst. Saf.*, **91**(10–11), pp. 1390–1397.

- [12] Rebba, R., and Mahadevan, S., 2006, "Validation of Models With Multivariate Output," *Reliab. Eng. Syst. Saf.*, **91**(8), pp. 861–871.
- [13] O'Hagan, A., 1995, "Fractional Bayes Factors for Model Comparison," *J. R. Stat. Soc., Ser. B (Methodological)*, **57**(1), pp. 99–138.
- [14] Wang, S., Chen, W., and Tsui, K.-L., 2009, "Bayesian Validation of Computer Models," *Technometrics*, **51**(4), pp. 439–451.
- [15] Ferson, S., Oberkampf, W., and Ginzburg, L., 2008, "Model Validation and Predictive Capability for the Thermal Challenge Problem," *Comput. Methods Appl. Mech. Eng.*, **197**(29–32), pp. 2408–2430.
- [16] Ferson, S., and Oberkampf, W., 2009, "Validation of Imprecise Probability Models," *Int. J. Reliab. Saf.*, **3**(1), pp. 3–22.
- [17] Rebba, R., and Mahadevan, S., 2008, "Computational Methods for Model Reliability Assessment," *Reliab. Eng. Syst. Saf.*, **93**(8), pp. 1197–1207.
- [18] Sankararaman, S., and Mahadevan, S., 2013, "Assessing the Reliability of Computational Models Under Uncertainty," *AIAA Paper No. 2013-1873*.
- [19] O'Hagan, A., and Oakley, J. E., 2004, "Probability is Perfect, but We Can't Elicit It Perfectly," *Reliab. Eng. Syst. Saf.*, **85**(1–3), pp. 239–248.
- [20] Jaulin, L., Kieffer, M., Didrit, O., and Walter, E., 2001, *Applied Interval Analysis*, Springer-Verlag, New York.
- [21] Shafer, G., 1976, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ.
- [22] Dubois, D., and Prade, H., 1986, *Possibility Theory: An Approach to Computerized Processing of Uncertainty*, Plenum Press, New York.
- [23] Ross, T. J., 1995, *Fuzzy Logic With Engineering Applications*, McGraw-Hill, New York.
- [24] Klir, G. J., and Wierman, M. J., 1998, *Uncertainty-Based Information: Elements of Generalized Information Theory*, 2nd ed., Vol. 15, Physica-Verlag, Heidelberg, DE.
- [25] Helton, J., and Sallaberry, C., 2012, "Uncertainty and Sensitivity Analysis: From Regulatory Requirements to Conceptual Structure and Computational Implementation," *Uncertainty Quantification in Scientific Computing, IFIP Advances in Information and Communication Technology*, Vol. 377, Springer, Berlin, pp. 60–77.
- [26] Oberkampf, W. L., Helton, J. C., Joslyn, C. A., Wojtkiewicz, S. F., and Ferson, S., 2004, "Challenge Problems: Uncertainty in System Response Given Uncertain Parameters," *Reliab. Eng. Syst. Saf.*, **85**(1–3), pp. 11–19.
- [27] Kiureghian, A., 2009, "Aleatory or Epistemic? Does It Matter?" *Struct. Saf.*, **31**(2), pp. 105–112.
- [28] Hu, K., 2014, "2014 V&V Challenge: Problem Statement," *Sandia Technical Report No. SAND2013-10486P*.
- [29] Kennedy, M. C., and O'Hagan, A., 2001, "Bayesian Calibration of Computer Models," *J. R. Stat. Soc., Ser. B (Stat. Methodol.)*, **63**(5), pp. 425–464.
- [30] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E., 1953, "Equation of State Calculations by Fast Computing Machines," *J. Chem. Phys.*, **21**(6), p. 1087.
- [31] Hastings, W., 1970, "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, **57**(1), pp. 97–109.
- [32] Gilks, W., and Wild, P., 1992, "Adaptive Rejection Sampling for Gibbs Sampling," *J. R. Stat. Soc., Ser. C (Appl. Stat.)*, **41**(2), pp. 337–348.
- [33] Neal, R., 2003, "Slice Sampling," *Ann. Stat.*, **31**(3), pp. 705–741.
- [34] Cressie, N. A. C., 1993, *Statistics for Spatial Data*, Revised edition, Wiley, New York.
- [35] Sacks, J., Schiller, S. B., and Welch, W., 1989, "Design of Computer Experiments," *Technometrics*, **31**(1), pp. 41–47.
- [36] Xiu, D., 2010, *Numerical Methods for Stochastic Computations: A Spectral Method Approach*, Princeton University Press, Princeton, NJ.
- [37] Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flanner, B. P., 2007, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed., Cambridge University Press, New York.
- [38] McCulloch, W., and Pitts, W., 1943, "A Logical Calculus of Ideas Immanent in Nervous Activity," *Bull. Math. Biophys.*, **5**(4), pp. 115–133.
- [39] Rasmussen, C. E., and Williams, C. K. I., 2006, *Gaussian Processes for Machine Learning*, The MIT Press, Cambridge, MA.
- [40] Johnson, N. L., 1949, "Systems of Frequency Curves Generated by Methods of Translation," *Biometrika*, **36**(1/2), pp. 149–176.
- [41] DeBrota, D. J., Roberts, S. D., Dittus, R. S., Wilson, J. R., Swain, J. J., and Venkataraman, S., 1988, "Input Modeling with the Johnson System of Distributions," *Winter Simulations Conference (WSC '88)*, M. Abrams, P. Haigh, and J. Comfort, eds., pp. 165–179.
- [42] Marhadi, K., Venkataraman, S., and Pai, S. S., 2012, "Quantifying Uncertainty in Statistical Distribution of Small Sample Data Using Bayesian Inference of Unbounded Johnson Distribution," *Int. J. Reliab. Saf.*, **6**(4), pp. 311–337.
- [43] Sankararaman, S., and Mahadevan, S., 2013, "Separating the Contributions of Variability and Parameter Uncertainty in Probability Distributions," *Reliab. Eng. Syst. Saf.*, **112**, pp. 187–199.
- [44] Ellingwood, B., Galambos, T., MacGregor, J., and Cornell, C. A., 1980, *Development of a Probability Based Load Criterion for American National Standard A58: Building Code Requirements for Minimum Design Loads in Buildings and Other Structures*, Vol. 577, National Bureau of Standards Publication, Gaithersburg, MD.