Validation and Error Estimation of Computational Models

Ramesh Rebba, Sankaran Mahadevan and Shuping Huang

Department of Civil and Environmental Engineering,

Vanderbilt University, Box 1831-B, Nashville, 37235, USA

E-mail: rebba.ramesh@vanderbilt.edu, sankaran.mahadevan@vanderbilt.edu, shuping.huang@vanderbilt.edu

Abstract

This paper develops a Bayesian methodology for assessing the confidence in model prediction by comparing the model output with experimental data when both are stochastic. The prior distribution of the response is first computed, which is then updated based on experimental observation using Bayesian analysis to compute a validation metric. A model error estimation methodology is then developed to include model form error, discretization error, stochastic analysis error (UQ error), input data error and output measurement error. Sensitivity of the validation metric to various error components and model parameters in discussed. A numerical example is presented to illustrate the proposed methodology.

Keywords: Bayesian statistics, error estimation, sensitivity, uncertainty, validation, verification.

1. Introduction

Complex natural phenomena are increasingly sought to be modeled through sophisticated computational models, with very few or no full-scale experiments, thus reducing the time and cost of engineering development relying upon the understanding of these phenomena. However, such models incorporate many assumptions and approximations and hence need to be subjected to rigorous, quantitative verification and validation (V & V) before they can be applied to practical problems with confidence.

There are a number of physical, statistical and model uncertainties in the prediction apart from the various direct sources of numerical error. A probabilistic approach to V&V under uncertainty involves quantification of the statistical distribution of model prediction and then comparing it with experimental measurement that also follows a statistical distribution. Note that this could also be viewed as studying the joint distribution of the experiment and model. Various methods are available to carry out probabilistic analysis to quantify the uncertainty in the model output, given the statistical distributions of the input variables, such as Monte Carlo simulation [1] or response surface methods [2, 3]. The choice of method depends on the nature of model used for predicting the output, and the needs with respect to accuracy and efficiency.

Verification refers to the assessment of accuracy of the solution with respect to known solutions, or by some other means, such as a posteriori error estimation. This activity helps to identify, quantify and reduce the errors in the computational model [4, 5]. Several finite element discretization error estimators have been developed in the literature [6-8]. Error estimates for uncertainty quantification methods (Monte Carlo and response surface methods) are also available [2, 5].

Validation involves comparison of model prediction with experimental data [4]. The widely used method of "graphical validation" or viewgraph-based judgment (i.e., by plotting graphs of

Sensitivity Analysis of Model Output
Kenneth M. Hanson and François M. Hemez, eds.
Los Alamos National Laboratory, 2005; http://library.lanl.gov/

prediction and observation) is inadequate although it is better than a qualitative comparison [9, 10]. A rigorous quantitative model validation metric should include both prediction and measurement errors, and other uncertainties. Several metrics have been developed to include parametric uncertainty [11]. One such metric normalizes the difference between model predictions and experimental values and computes a relative error norm for discrete and continuous domain problems. Another metric includes the uncertainty in the experimental value due to limited data through statistical distributions and classical hypothesis testing [12, 13].

Two types of validation metrics are developed in this paper, based on the Bayesian approach. The first metric considers test data based on a simple fail/pass criterion, while the second metric compares model prediction with observed response measurement, both being continuous variables. The second approach explicitly incorporates the variability in the experimental data and the magnitude of its deviation from the model prediction. Once the model is validated, it may be calibrated to improve its predictive capability. A prediction error estimation methodology is developed for this purpose in this paper; this includes model form error, discretization error, stochastic analysis error (UQ error), input data error and output measurement error. The overall error is a nonlinear combination of these various errors. Sensitivity analysis of the validation metric to different physical and statistical parameters of the model output and measurement error variance can be very useful for model improvement or calibration and resource allocation. Section 2 develops the Bayesian validation metrics, and Section 3 describes the proposed methodology for model error estimation and sensitivity analysis of the validation metric. An illustrative numerical example is provided in Section 4.

2. Validation metric

2.1. Bayes factor

Consider two models (or hypotheses) M_i and M_j. Their prior probabilities of acceptance are denoted by $P(M_i)$ and $P(M_j)$. By Bayes' rule, when an event/data is observed, the relative posterior probabilities of two hypotheses are obtained as [14, 15]:

$$\frac{P\left(M_i \middle| \text{observation}\right)}{P\left(M_j \middle| \text{observation}\right)} = \left[\frac{P\left(\text{observation} \middle| M_i\right)}{P\left(\text{observation} \middle| M_j\right)}\right]\left[\frac{P\left(M_i\right)}{P\left(M_j\right)}\right] \tag{1}$$

The term in the first set of square brackets on the right hand side is called the "Bayes factor" [16]. If the Bayes factor is greater than 1.0 then it can be inferred that the data favors the model M_i more than model M_j. If only a single model M is proposed, then the model could be either accepted as correct or rejected as incorrect. Thus the Bayes factor in Eq. (1) may also be written as $P\left(\text{observation} \middle| M \text{ is correct}\right) / P\left(\text{observation} \middle| M \text{ is not correct}\right)$. When an observation is made, then the Bayes factor estimates the ratio of relative likelihoods of the null hypothesis (i.e., data supports the proposed model) and alternate hypothesis (i.e., data does not support the proposed model). The Bayes factor metric is further developed below for two situations: 1) reliability model 2) response computation.

2.1.1. Validation with pass/fail test data

Let x_o and x be the predicted failure probability and true failure probability respectively of an engineering system. The value x_o is predicted by model M. This can be considered as a point null

hypothesis (H_o: $x = x_o$). To estimate the Bayes factor in Eq. (1), we need to consider an alternative hypothesis (H_1: $x \neq x_o$).

If n experiments are undertaken, and k failures (e.g., stress greater than an allowable value) are observed out of n tests, then the probability of observing the data given that the true probability is equal to x comes from a binomial distribution as

$$P(k \mid x, n) = {}^{n}C_{k} x^{k} (1-x)^{n-k} \tag{2}$$

Under the null hypothesis, this probability, $P(data \mid H_o: x = x_o)$ can be exactly estimated by simply substituting x_o in Eq. (2). Assume that there is no prior information about x under the alternative hypothesis. Therefore, a uniform distribution in [0, 1] is assumed for $f(x \mid H_1)$, the prior density under the alternative hypothesis [17]. Then the Bayes factor is computed as

$$B(x_o) = \frac{P(data \mid H_0 : x = x_o)}{P(data \mid H_1 : x \neq x_o)} = \frac{C_k^n x_0^k (1-x_0)^{n-k}}{\int_0^1 C_k^n x^k (1-x)^{n-k} \, f(x \mid H_1) \, dx} = (n+1) C_k^n x_0^k (1-x_0)^{n-k} \tag{3}$$

It is easy to identify the above expression in Eq. (3) to be the probability density function (PDF) of a beta distribution with parameters $k + 1$, $n - k + 1$. It is well known that the posterior PDF of x follows the beta distribution, when the prior PDF has uniform distribution. In more detail, if the prior has a uniform PDF in [0, 1] i.e., $f(x) = 1$, then the posterior PDF is

$$f(x \mid data : n, k) = (n+1) C_k^n x^k (1-x)^{n-k} \tag{4}$$

Note that this result is the same as in Eq. (3), which is the Bayes factor $B(x)$ evaluated at the probability x (see Fig. 1). Therefore, the Bayes factor can be viewed here as the posterior density of x evaluated at the predicted value x_o.

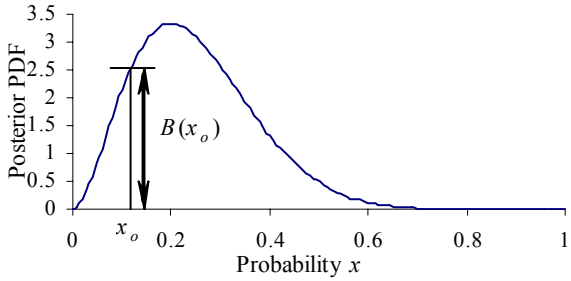

Fig. 1. Posterior density function for the probability of failure

The above result, that the Bayes factor is the posterior PDF at the predicted value x_0, is only for the case with uniform prior and binomial pass/fail data. For tests conducted in other situations, only a response quantity may be measured (such as deflection, strain etc.), but the specimen may not be loaded till failure. In such cases, it is valuable to derive a more general expression for the Bayes factor, by using prior and posterior PDF's of the predicted response.

2.1.2. Validation with response variable measurement

In Eq. (2), the probability of the data k for a given value of x, i.e., $P(k \mid x, n)$, is also the likelihood function of x, i.e., $L(x)$, where the failure probability x is the parameter of the binomial distribution. For a continuous distribution, the likelihood function is proportional to the

probability density of data y given the parameter x i.e., $L(x) \propto f(y|\, x)$ [18]. Consider x to be not the failure probability, but some general response quantity, with density function $f(x)$ and x_o the value predicted by the computational model under consideration. Then the probability of observing the data under the null hypothesis, $P(data|\, H_o\colon x = x_o)$ can be obtained from $L(x_o) = \varepsilon f(y|\, x_o)$ where ε is a positive constant [18]. Similarly, the probability of observing the data under the alternative hypothesis $P(data|\, H_1\colon x \neq x_o)$ can be obtained from $\int L(x)g(x)dx$ or $\int \varepsilon f(y\,|\,x)g(x)dx$, where $g(x)$ is the prior density of x under the alternative hypothesis. Since no information on $g(x)$ is likely to be available, one possibility is to assume $g(x) = f(x)$. Then, using Eq. (1) and Bayes theorem, the Bayes factor is computed as

$$B(x_o) = \frac{P\left(data\,\middle|\, H_0 : x = x_o\right)}{P\left(data\,\middle|\, H_1 : x \neq x_o\right)} = \frac{L(x_o)}{\int L(x) f(x)dx} = \frac{f(y\,|\,x_o)}{\int f(y\,|\,x) f(x)dx} = \left.\frac{f(x\,|\,y)}{f(x)}\right|_{x=x_0} \tag{5}$$

Thus, the Bayes factor simply becomes the ratio of posterior to prior PDFs of the predicted response when $g(x) = f(x)$. This result probabilistically quantifies the contribution to model validation of an experimental result that agrees with a given model prediction. If $g(x) \neq f(x)$, then the Bayes factor is computed using Eq. (5) with $g(x)$ instead of $f(x)$ in the denominator. Fig. 2 shows notional posterior and prior densities of model prediction x. Once again, $B > 1$ indicates data support for the model.

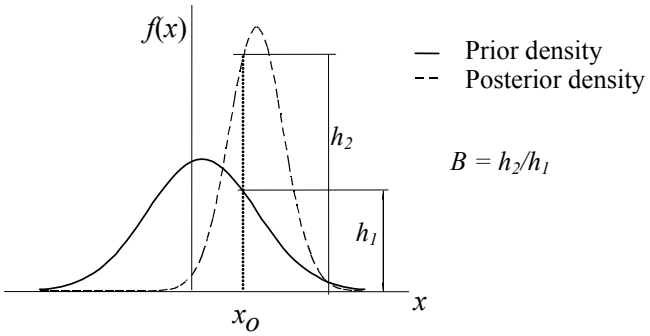

Fig. 2. Validation metric as a ratio of posterior and prior density values

If x_{true} is the true solution, x is the model output, and y is the experimental observation, then the following equations hold:

$$x_{true} = x + \varepsilon_{pred} \tag{6a}$$

$$x_{true} = y + \varepsilon_{exp} \tag{6b}$$

where ε_{pred} is the model prediction error and ε_{exp} is the measurement error. If we hypothesize that there is no prediction error, the observed value will simply be $y = x - \varepsilon_{exp}$. From this relation and a Gaussian experimental error assumption, we obtain $f(y|\, x) \sim N(x,\ \sigma^2_{\varepsilon_{\exp}})$. The likelihood function $L(x)$ in Eq. (5) can be created using $f(y|\, x)$. If there is only one observed value of y, then $L(x) = f(y|\, x)$. If multiple data are observed, the likelihood is constructed as a product of $f(y|\, x)$ values evaluated at each y. This can be used to test our hypothesis.

A Bayes factor less than unity denotes that ε_{pred} is significant and should not have been omitted and hence there is a need for estimating the total prediction error. It should be noted that the metric shown in Eq. (5) allows us to use non-Gaussian experimental errors also. Even when

$B > 1$, it is useful to quantify the prediction error, and to examine whether contributions from different errors cancel each other. Further the degree of confidence in the model prediction can be measured from the posterior probability of the null hypothesis being true i.e., $P(H_0|\ data)$ as $B/(B + 1)$ assuming that the prior probability $P(H_0)$ to be 0.5 in the absence of any prior knowledge. The following section presents methods for quantifying the errors and uncertainty in model prediction.

3. Error estimation

The total prediction error is a function of various error components that can be broadly divided into numerical solution errors and model form errors. Investigations on error combination are rarely available. This paper pursues a nonlinear combination method.

3.1 Numerical Error Components in Simulation

Several components of numerical errors in model prediction, such as data error, discretization error, stochastic analysis error (or UQ error), and measurement error are briefly discussed below.

3.1.1. Input data error (ε_d)

The measurement error in the input variables will be propagated to the prediction of the output. If the relationship between input and output is given by $u = f(x_1, x_2, ..x_m)$, then the error in the prediction of the output due to the measurement error in the input variables may be approximated using a first-order sensitivity analysis as

$$\varepsilon_d = \Delta u = \sum_{i=1}^{m} \left(\frac{\partial f}{\partial x_i} \right|_{x = \bar{x}} \delta x_i \tag{7}$$

in which δx_i is the measurement error in i^{th} input random variable x_i and $\left(\frac{\partial f}{\partial x_i} \right|_{x=\bar{x}}$ is the first order sensitivity coefficient of the model output u with respect to the i^{th} input random variable x_i. The measurement error in each input variable has been commonly quantified as a random variable with Gaussian distribution, with zero mean and a known or assumed variance, based on the instrument calibration.

3.1.2. Discretization error (ε_h)

Several methods to quantify the discretization error in finite element analysis are available in the literature. However, most of these methods do not quantify the actual error; instead, they quantify some surrogate measures to facilitate adaptive mesh refinement. The Richardson extrapolation (RE) method has been found to be suitable for model verification and validation, since it comes closest to quantifying the actual discretization error [19]. This method has been extended by the first author to stochastic finite element analysis [20]. It should be noted that RE requires that the model solution be asymptotically convergent and the domain is discretized uniformly (regular grid). The assumption of monotone truncation error convergence is not valid for very coarse models sometimes. In the Richardson extrapolation method, the error due to grid size (for a coarse mesh) is given by

$$\varepsilon_h = \frac{f_1 - f_2}{r^p - 1} \tag{8}$$

where the grid refinement ratio $r = h_2/ h_1$, and f_1 and f_2 are the solutions with coarse and fine meshes respectively. The order of convergence p can be obtained from the relation $p = \ln\left(\frac{f_3 - f_2}{f_2 - f_1}\right) / \ln(r)$ where f_3 is the solution with the finest grid size, and $r = h_2/ h_1 = h_3/ h_2$. Due to the input variable uncertainties, the finite element solutions f_1 and f_2 are stochastic. Therefore, based on Eq. (8), the discretization error (ε_h) is also a random variable.

3.1.3. Uncertainty quantification error (ε_{uq})

Errors in stochastic analysis are method-dependent, i.e. sampling error occurs (ε_{mc}) in Monte Carlo methods and truncation error (ε_{sm}) occurs in response surface methods (either conventional or polynomial chaos-based). For example, sampling error could be assumed to be a Gaussian random variable with zero mean and variance given by σ^2/N where N is the number of Monte Carlo runs and σ^2 is the original variance of the model output [21]. The truncation error (ε_{sm}) is simply the residual error in the response surface.

In this paper, due to the use of response surface techniques for uncertainty quantification, truncation error is used to represent ε_{uq}. A polynomial chaos-based response surface is used, which is found to have superior convergence characteristics than conventional response surface models [20]. The response surface is constructed by approximating both the input and output random variables through series expansions of independent standard random variables ξ_i. For example, a normal random variable can be expressed in terms of its parameters as $\mu + \sigma\xi$ where ξ is a standard normal variable. Similarly, a lognormal random variable with parameters λ and δ can be expressed as exp ($\lambda + \delta\xi$). The output response surface is expressed in terms of the input variables through a polynomial chaos expansion as

$$x = a_o + \sum_{i_1=1}^{n} a_{i_1} \Gamma_1(\xi_{i_1}) + \sum_{i_1=1}^{n} \sum_{i_2=1}^{i_1} a_{i_1 i_2} \Gamma_2(\xi_{i_1}, \xi_{i_2}) + \sum_{i_1=1}^{n} \sum_{i_2=1}^{i_1} \sum_{i_3=1}^{i_2} a_{i_1 i_2 i_3} \Gamma_3(\xi_{i_1}, \xi_{i_2}, \xi_{i_3}) + ... \quad (9)$$

where x is the output and $\Gamma_p(\xi_{i_1},...,\xi_{i_p})$ are multi-dimensional Hermite polynomials of degree p. The design points for the data used to construct the response surface are chosen such that they are the roots of the Hermite polynomial of the order p+1 where p is the order of the response surface [2].

The series in Eq. (9) could be truncated to a finite number of terms. Thus the accuracy of the stochastic computational model depends on the order of the expansion. The truncation error ε_{sm} in the response surface of order p can be estimated by constructing additional higher order response surfaces (i.e., order p +1 or p + 2), and using the Richardson extrapolation method, similar to Eq. (8).

3.1.4. Output measurement error (ε_{exp})

The measurement error in the output variable is a separate error component, whereas the measurement error in the input variables is compounded through propagation in the prediction model. Output measurement error is quantified commonly as a random variable with Gaussian distribution, with zero mean and a known or assumed variance.

3.2. Model Form Error (ε_{model})

If multiple models are considered, Bayesian model averaging (BMA) [22, 23] may be used to reduce the model form uncertainty and model errors, but not quantify them explicitly. In some practical cases, only one model may be available, in which case BMA may not be useful. If a single model is employed, this paper makes use of the observed data to express the overall prediction error through a regression model consisting of the individual error components. The residual of such a regression analysis should include the model form error (after subtracting the experimental error effects). From Eqs. (6a) and (6b) and by denoting ε_{obs} as the difference between the data and prediction, i.e., $\varepsilon_{obs} = y - x$, we can obtain the following relation:

$$\begin{aligned} \varepsilon_{obs} &= y - x = \varepsilon_{pred} - \varepsilon_{exp} \\ &= \varepsilon_{num} + \varepsilon_{model} - \varepsilon_{exp} \\ &= f(\varepsilon_h, \varepsilon_{uq}, \varepsilon_d) + \varepsilon_{model} - \varepsilon_{exp} \end{aligned} \tag{10}$$

In Eq. (10), overall numerical error ε_{num} is a nonlinear function of the error components ε_h, ε_{uq} and ε_d. Therefore, it is constructed as a response surface with respect to ε_h, ε_{uq}, and ε_d, using a polynomial chaos expansion, similar to Section 3.1.3. The quantity $\varepsilon_{model} - \varepsilon_{exp}$ is simply the residual $\varepsilon_{residual}$ of such a response surface. Thus the distribution of model error ε_{model} is quantified by knowing the distributions of $\varepsilon_{residual}$ and ε_{exp}. However in most practical situations, the validation data is very limited. From a single validation experiment, one has the numerical values of prediction and observation, and estimates of the numerical errors in prediction, but not the experimental error. In other words, values $f(\varepsilon_h, \varepsilon_{uq}, \varepsilon_d)$, and ε_{obs} are available but the exact value for experimental error ε_{exp} cannot be estimated. Only the distribution of ε_{exp} is available or assumed, if at all. If we have a sufficient number of validation data, we can compute the difference ($\varepsilon_{obs} - f(\varepsilon_h, \varepsilon_{uq}, \varepsilon_d)$) and add a randomly generated term ε_{exp} to it each time to obtain an estimate of model form error ε_{model}. Since the sample size (number of observations made) is limited, an empirical distribution for ε_{model} cannot be constructed with confidence. However, one can compute the statistics like mean and standard deviation of model error from a set of validation experiments. Bootstrapping [24] (sampling with replacement) can be done on the given data set to generate a large number of statistics for model form error, thus obtaining the distributions for mean and standard deviation of model form error. Bootstrapping assumes that the data set in hand is representative of the intended population and no prior assumptions are made regarding distribution of the samples. Further the observations are assumed to be independent and sampling is purely random.

3.3 Sensitivity Analysis

The Bayesian validation metric given in Eqs. (3) and (5) depends explicitly on model output and uncertainties arising from validation experiments like lack of sufficient data points and random measurement errors etc. Also, the statistical and physical model parameters affect the model output and hence a model may be accepted or rejected based on our prior assumption in a Bayesian analysis. Thus, there exists an implicit relation between the Bayes factor and each of the above model parameters, and curves may be fitted to depict this relation. The sensitivity of the Bayes factor to these variables may be estimated from the slopes (first order sense) of such plots. For example, with reference to Eq. (3), the uncertainty due to a limited number of data

points in a pass/fail type of test may be quantified as $\sigma_{x_o}^2 = \frac{x_o(1-x_o)}{n}$. For particular known values of k and x_0, one can generate a plot of $B(x_0)$ versus $\sigma_{x_o}^2$ for different values of n, based on Eq. (3), as shown in Fig. 3.

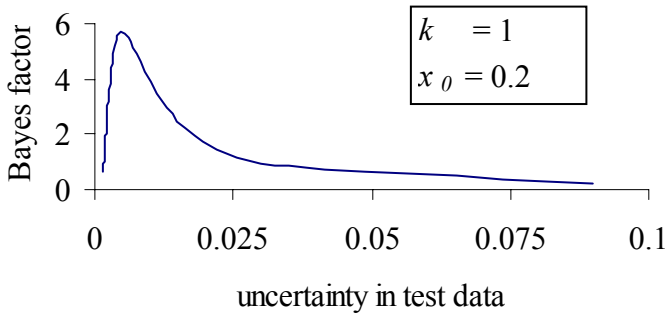

Fig. 3. Relation between $B(x_0)$ and $\sigma_{x_o}^2$

Similar relations between $B(x_0)$ and $\sigma_{\varepsilon_{\exp}}^2$, μ_x, σ_x etc in Eq. (5) can be derived to be used in a sensitivity analysis. A numerical example is provided to illustrate the proposed methodology.

4. Numerical example

The steady state heat transfer in a thin wire of length L, with thermal conductivity k, convective heat coefficient β is of interest. The temperature at midpoint of the wire needs to be predicted. We assume (acknowledging model form error) that this problem is essentially one dimensional and that the solution can be obtained from the boundary value problem

$$-k\frac{\partial T^2}{\partial x^2} + \beta T = Q(x) \qquad (11)$$

with known conditions $T(0)$ and $T(L)$

where $Q(x)$ is the heat source. Suppose k and β are assumed for the sake of illustration to be random variables, normally distributed with statistics N (5, 1) and N (0.5, 0.1) respectively. Also, the heat source $Q(x) = 25(2x\text{-}L)^2$ with $L = 4$. The wire is insulated at the ends, which are kept at zero temperature i.e., $T(0) = T(L) = 0$. It is required to predict $T(2.0)$.

The numerical solution $T(x)$ for Eq. (11) can be obtained using a finite-difference scheme with discretization step size h. Since k and β are random, the model prediction $T(2.0)$ is also random whose statistical distribution needs to be estimated. Since each computation of $T(2.0)$ using a finite-difference scheme could be computationally expensive, a response surface may be fitted to predict $T(2.0)$ as a function of input random variables k and β. A polynomial chaos-based response surface is used for this purpose, as mentioned in Section 3.1.3 earlier. Thus the random variables k and β are expressed as $(5+\xi_1)$ and $(0.5+0.1\xi_2)$ respectively, where ξ_1 and ξ_2 are standard normal variables. The design points for the data used to construct the response surface are chosen such that they are the roots of Hermite polynomial of the order p+1 where p is the order of the response surface. The corresponding values of k and β, with respect to these collocation points, are then used in the numerical model to compute the response $T(2.0)$. The

unknown coefficients of the stochastic response surface are then computed using standard regression techniques.

In this numerical example, a second order response surface in two variables is constructed for $T(2.0)$ and with step-size $h = 1$. The design points for k and β are selected at the collocation points obtained from the roots of 3rd order Hermite polynomials. The response surface is

$$T_{pred} = T(2.0) = 17.102 - 3.003\ \xi_1 - 0.698\ \xi_2 + 0.4964(\xi_1^2 - 1) + 0.0251(\xi_2^2 - 1) + 0.237\ \xi_1\xi_2 \quad (12)$$

where ξ_1 and ξ_2 are independent standard normal variables and $R^2 = 0.999$. The PDF of T_{pred} can be generated by simulating ξ_1 and ξ_2, and is found to have a lognormal distribution with mean 17.12 and variance 10.042. This is the prior density to be considered in the Bayesian model validation next.

4.1 Validation

Suppose for given values of k and β, the numerical model predicted a temperature of 18.5 degrees. A wire made of a material with properties k and β having the same measured values as input to the numerical model was tested three times repeatedly to measure the temperature at location $x = 2$.

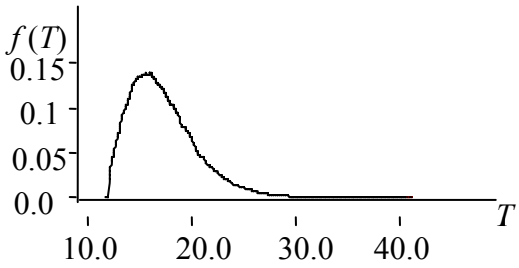

Fig. 4. PDF of $T(2.0)$

The measured temperature was different in each experiment i.e., 18.8, 18.2, 18.9 degrees. Assuming a Gaussian experimental error with zero mean, the true experimental value is assumed to be the mean of the three measurements, i.e., 18.633 degrees for the sake of illustration. Also, the experimental error is assumed to have a variance $\sigma^2_{\varepsilon\exp}$ estimated from the three measurements, again for the sake of illustration.

As described in Section 2.1.1, the likelihood function of the prediction is proportional to a normal density with mean $T(2.0)$ and variance $\sigma^2_{\varepsilon\exp} = 0.1433$. Also with the knowledge of $f(T(2.0))$, the validation metric is evaluated at $T = 18.5$ degrees. Using Eq. (5), the validation metric B is found to be 11.6 which is much greater than 1.0 indicating that the data matches very well with the prediction. However, one should be cautious in accepting this result since various errors like discretization error, input data error, truncation error and even model form error may be canceling each other to produce a result that is close to the measured value. Hence there is a need to estimate the various errors explicitly, as described below.

4.2 Error Estimation

The numerical model related to Eq. (11) was refined using $h = 0.5$ and $h = 0.25$ to estimate the convergence rate $p = 1.985 \approx 2$ as described in Section 3.1.2. The discretization error ε_h based on

the Richardson extrapolation method was obtained as a stochastic response surface in two variables as

$$\varepsilon_h = 5.9725 - 1.1\ \xi_1 - 0.1918\ \xi_2 + 0.1895(\xi_1^2 - 1) + 0.0057(\xi_2^2 - 1) + 0.0681\ \xi_1\xi_2 \qquad (13)$$

The discretization error was found to have a lognormal distribution with parameters $\lambda = 1.762$, and $\xi = 0.1923$ i.e., with mean 5.9725 degrees and a standard deviation of 1.15. Due to the use of the stochastic response surface, the uncertainty quantification error ε_{uq} is estimated by the truncation error $\varepsilon_{sm,}$ i.e., the residual error in the stochastic response surface for the model response $T(2.0)$, which was found to be a Gaussian variable with zero mean and a standard deviation of 0.2 degrees. This is much smaller than the FEM discretization error. The information on input data error ε_d was not available, and ε_d is assumed negligible in this example. Neglecting ε_{sm} and ε_d based on the above observations, the numerical error ε_{num} in Eq. (10) is approximated by ε_h. Thus Eq. (10) reduces to

$$\varepsilon_{obs} = \varepsilon_h + \varepsilon_{model} - \varepsilon_{exp} \qquad (14)$$

In Eq. (14), the distributions of ε_h and ε_{exp} are available from the above discussion. Specific values of ε_{obs} are available from each test. The number of tests, and thus the number of samples of ε_{obs}, is likely to be small in practical problems. In this example, nine values of T_{pred} (and hence ε_h) are calculated at nine collocation point values of ξ_1 and ξ_2 (using Eqs. 12 and 13 respectively), as shown in Table 1. Assume for the sake of illustration that nine corresponding values of T_{obs} (and hence ε_{obs}) are observed from nine tests, also shown in Table 1. A bootstrap resampling technique [24] (with replacement) can then be applied to generate ε_{model}, based on Eq. (14). Each time a value for $(\varepsilon_{obs} - \varepsilon_h)$ is resampled from the nine values shown in Table 1, a randomly generated value of ε_{exp} is added to it, and sample statistics of ε_{model} (mean and standard deviation) are calculated from nine such values in a resample. (Note that each resample contains the same number of data points as the original sample, i.e., nine in this case). The procedure is repeated and 10,000 values for $\mu_{\varepsilon model}$ and $\sigma_{\varepsilon model}$ are obtained, thus giving their distributions as shown in Table 2. This approach thus provides a measure of the uncertainties in the statistical parameters of model form error, since they are obtained by bootstrap resampling.

Table 1. Sample points for model form error

ε_h	T_{pred}	T_{obs}	ε_{obs}	$\varepsilon_{obs} - \varepsilon_h$
5.824	16.597	16.794	0.197	-5.627
8.126	12.902	12.997	0.095	-8.031
5.824	15.642	15.920	0.278	-5.546
4.174	23.222	23.310	0.088	-4.086
5.824	17.653	17.442	-0.211	-6.035
8.126	13.526	13.488	-0.038	-8.164
4.174	21.350	21.181	-0.169	-4.343
8.126	12.327	12.173	-0.154	-8.28
4.174	25.394	25.301	-0.093	-4.267

In this example, the mean model error $\mu_{\varepsilon model}$ was observed to follow a normal distribution (mean -6.03, see Table 2), and the standard deviation of model error $\sigma_{\varepsilon model}$ was observed to follow a Weibull distribution. The numerical error has a mean value of 5.9 degrees while the model form error has a mean value of -6.03 degrees.

Table 2. Distributions of model error statistics

Parameter	Mean	Variance	Type of distribution
$\mu_{\varepsilon model}$	-6.03	0.293	Normal
$\sigma_{\varepsilon model}$	1.6	0.066	Weibull

Looking at these numerical results, the two errors appear to have almost cancelled each other in Eq. (14), resulting in a small overall observed error ε_{obs} and hence a high Bayes factor in Section 4.1, indicating an optimistic model validation result. But the error quantification shows that the prediction model has large numerical and model form errors. Thus a model acceptance/ rejection criterion based on the overall ε_{obs} alone can lead to misleading conclusions, especially in other untested situations when the numerical and model form errors might become additive. This observation shows the importance of quantifying various error components, in addition to simply comparing prediction and observation.

Since the error components add up linearly in this example, the relative first order sensitivities of numerical, model and experimental errors to Bayes factor will simply be proportional to their respective standard deviations. The first order normalized sensitivities of the model parameters k and β to Bayes factor in Eq. (5) are found to be 0.978 and 0.208 respectively at their corresponding mean values. These values are computed by plotting the relation B vs. k and B vs. β separately, and normalizing the products of standard deviations
and slopes evaluated at their mean values.

5. Conclusion

V&V needs to quantify various errors under uncertainty and effectively compare them with imprecisely measured experimental data to assess the predictive capability of the model. A Bayesian approach was proposed for model validation in this paper, and developed for two situations. The first case performs model assessment using a pass/fail criterion and uses the Bayes factor as a metric. The second case includes the uncertainty in the experimental data explicitly and estimates the Bayes factor using prior and posterior distributions of the model output. The overall numerical error in prediction is expressed as a nonlinear response surface in terms of several errors such as discretization error, uncertainty quantification error and input data error, and compared with the observed error to estimate the statistics of the model form error. A bootstrapping technique is used to estimate the model form error from a limited number of experimental measurements. The simple numerical example resulted in a linear relation among various errors and the further work is needed to demonstrate the methodology for a more complex problem where the various errors are combined in a non-linear fashion. Also, the sensitivity analysis limited to a first-order evaluation showed that the validation metric is sensitive to the variance of each error component. Similarly model parameters that have an insignificant effect (small sensitivity) on Bayes factor can be omitted to reduce model complexity.

Acknowledgements

The research reported in this paper was supported by funds from Sandia National Laboratories, Albuquerque, NM (contract no. BG-7732), under the Sandia-NSF Life Cycle Engineering program (project monitors: Dr. Steve Wojtkiewicz and Dr. Thomas L Paez). The support is gratefully acknowledged.

References

1. R. L. Iman, and W. J. Conover. A distribution free approach to introducing rank correlation among input variables. *Communication on statistics- simulation and computation*, 11(3):311-334, 1982.
2. S. S. Isukapalli, and P. G. Georgopoulos. Computational Methods for the Efficient Sensitivity and Uncertainty Analysis of Models for Environmental and Biological Systems. *Technical Report,* State Univ. of New Jersey, 1999.
3. G. I. Schueller, C. G. Bucher, U. Bourgund, and W. Ouypornprasert. On efficient computational schemes to calculate failure probabilities. *Probabilistic engineering mechanics*, 4(1):10-18, 1989.
4. American Institute of Aeronautics and Astronautics, *Guide for the Verification and Validation of Computational Fluid Dynamics Simulations*, AIAA-G-077-1998, Reston, VA, 1998.
5. P. J. Roache. *Verification and Validation of in Computational Science and Engineering*, Hermosa, Albuquerque, USA, 1998.
6. M. Ainsworth, and J. T. Oden. A posteriori error estimation in finite element analysis. *Comput. Methods Appl. Mech. Engg*, 142:1-88, 1997.
7. O. C. Zienkiewicz, and R. L. Taylor. *The Finite Element Method. Volume 1-Basic Formulation and Linear Problems*, 4th edition, McGraw-Hill, 1989.
8. O. C. Zienkiewicz, and J. Z. Zhu. The super-convergent patch recovery and a posteriori estimates. Part 2: Error estimates and adaptivity. *Int. J. Numer. Methods Engg*, 33:1335-1382, 1992.
9. Defense Modeling and Simulation Office, *Verification, Validation, and Accreditation (VV&A) Recommended Practices Guide*, Office of the Director of Defense Research and Engg, www.dmso.mil/docslib, Alexandria, VA, April, 1996.
10. T. G. Trucano. Aspects of ASCI Code Verification and Validation. Sandia Technical Report No. SAND2000-0390C, Sandia National Laboratories, Albuquerque, NM, 2000.
11. W. L. Oberkampf, and T. G. Trucano. Verification and Validation in Computational Fluid Dynamics. *Progress in Aerospace Sciences*, 38: 209-272, 2002.
12. R. G. Hills, and I. H. Leslie. Statistical Validation of Engineering and Scientific Models: Validation Experiments to Application. Report No. SAND2003-0706, Sandia National Laboratories, Albuquerque, NM, 2003.
13. A. Urbina, T. L. Paez, T. K. Hasselman, G. W. Wathugala, K. Yap. Assessment of Model Accuracy Relative to Stochastic System Behavior. *Proceedings of 44th AIAA Structures, Structural Dynamics, Materials Conference*, April 7-10, Norfolk, VA, 2003.
14. J. O. Berger, and L. R. Pericchi. The intrinsic Bayes factor for model selection and prediction. *J. Amer. Statist. Assoc,* 91:109-122, 1996.
15. T. Leonard, and J. S. J. Hsu. *Bayesian Methods: An Analysis for Statisticians and Interdisciplinary Researchers*, Cambridge University Press, Cambridge, 1999.
16. H. Jeffreys. *Theory of Probability* (3rd. ed.), Oxford University Press, London, 1961.

17. R. Zhang, and S. Mahadevan. Bayesian Methodology for Reliability Model Acceptance. *Reliability Engineering and System Safety*, 80(1):95-103, 2003.
18. Y. Pawitan. *In All Likelihood: Statistical Modeling and Inference using Likelihood*, Oxford Science Publications, New York, 2001.
19. S. A. Richards. Completed Richardson extrapolation in space and time. *Comm. Numer. Methods Engg*, 13:573-58, 1997.
20. R. Rebba. *Computational Model Validation under Uncertainty.* Master's thesis, Vanderbilt University, Nashville, TN, 2002.
21. R. Y. Rubinstein. *Simulation and the Monte Carlo Method.* John-Wiley, New York, 1981.
22. D. Draper. Assessment and Propagation of Model Uncertainty. *J. of Royal Statistical Society Series B*, 57 (1):45-97, 1995.
23. J. A. Hoeting, D. Madigan, A. Raftery, and C. T. Volinsky. Bayesian Model Averaging: A Tutorial (with discussion). *Statistical Science*, 14(4):382 - 417, 1999.
24. B. Efron, B. Bootstrap method: Another look at the jackknife. *Ann. Statist*., 7:1-26, 1979.