From *SIAM News*, Volume 43, Number 10, December 2010

Computer Predictions with Quantified Uncertainty, Part II

By Tinsley Oden, Robert Moser, and Omar Ghattas

Having developed and calibrated a model to be used for a particular prediction (see Part I of this article, *SIAM News*, November 2010, page 1), we are ready to begin the validation process—that is, to assess the suitability of the calibrated model for the prediction.

The Validation Process

To begin this process, we need a new set of observations, designed to challenge the model. The new observational data and their uncertainties will then be compared to the "predictions" of the model and its uncertainties for the physical scenarios in which the observations were made (the "validation scenarios"). Here, however, we encounter a problem: The validation observations are generally not of the quantities of interest (QoI's) or for the scenarios of the prediction, but rather for some validation quantities that are experimentally accessible in the validation scenarios. So how are the results of the comparison between model and validation experiments to be evaluated? What level of disagreement can be tolerated for the purposes of the predictions to be made?

The answers to these questions depend critically on the character of the prediction problem at hand. First, we note that the validation questions posed here are somewhat different from those arising in the usual experimental falsification of scientific theories. In the latter, the aim is to discern the laws of nature (theories); a disagreement between theory and experiment thus falsifies the theory. Or, in the context of uncertainty, a low probability that the experimental observations are consistent with the theory makes the theory improbable. We know, for example, that measurements taken in relativistic scenarios will falsify Newtonian mechanics as a scientific theory. But for our present purposes, we seek useful models, which may well be imperfect, and ask a narrower question: Are the models valid for making particular predictions? To return to our example: Newtonian mechanics, despite having been falsified as a theory, provides a valid model for making many predictions. Unfortunately, models used for making predictions in complex systems generally do not have the well-characterized domains of applicability of Newtonian mechanics. Validation processes, as discussed here, are thus needed to detect whether a model proposed for use in a particular prediction is invalid for that purpose.

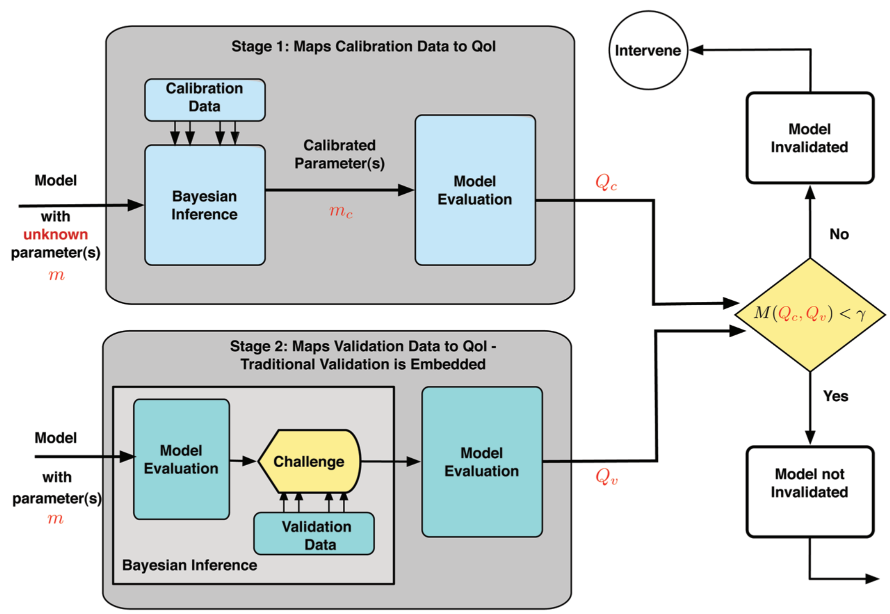

Figure 2. *Schematic of an example of the validation process.*

At the heart of the validation process is the question: What do discrepancies between models and experimental observations imply about the reliability of the specific QoI predictions to be made by the models? To address this question, Bayes's theorem can again be used, this time to update the model and its parameters in light of the validation observations. The effect of the updates on the predictions can then be determined by using the updated model to make predictions of the QoI's. If the change in the predictions is too large in some appropriate measure (as discussed below), then the inconsistency of the model with the validation data will influence the predictions, and the model and its calibration are invalid for predicting those QoI's.* A flow chart illustrating an example of this validation approach is shown in Figure 2.

In this process, the consequences of the validation observations must be representable within the structure of the model. In some situations, the models have a parametrization sufficiently rich to permit representation of almost any observation. In such cases, the validation data can be used to update the model parameters, and the invalidity of the model would be reflected in the inconsistency of the parameters needed to represent the validation data (relative to the calibration) and the impact of the inconsistency on the predicted QoI's. When such a parametric representation is not possible, the model will need to be enriched so that it can represent the validation data. One way to do this is to introduce a statistical model of the model error, and use it to represent the discrepancies between model and validation data. The impact of the discrepancy model on the QoI's, and thus the possible invalidity of the original model, can then be determined.

The impact of the validation observations on the predicted QoI's is expressed as the differences between two probability densities for the QoI's. Assessing the invalidity of the model then rests on an appropriate metric and tolerance for differences in these distributions. What this metric should be depends on the questions asked about the QoI's. Perhaps we need to know the most likely value of a QoI with a specified tolerance for error, in which

*See I. Babuška et al., Comput. Methods Appl. Mech. Engrg., 197 (2008), 2517–2539.

case inconsistencies in the prediction of the most probable value should be evaluated. Alternatively, it might be important to determine the probability that a presumably unlikely failure will occur. The invalidation metric should then assess the consistency of this prediction.

Another important issue is the selection of validation scenarios and measurements that challenge models in a way that is similar to the prediction scenario. Generally, such selections are based on the modeler's understanding of the potential weaknesses in the models (e.g., questionable assumptions made) and judgment regarding aspects of the models important for the predictions. The selection can also make use of the computational model, analyzing sensitivities of a QoI to parameter and model variation, and maximizing the information about the QoI to be gained. In complex multiphysics systems, designing a validation process is also complicated by the need to perform validation at many levels of complexity, from simple scenarios involving models of a single physical phenomenon, to complex scenarios involving all the models needed in the prediction scenario.

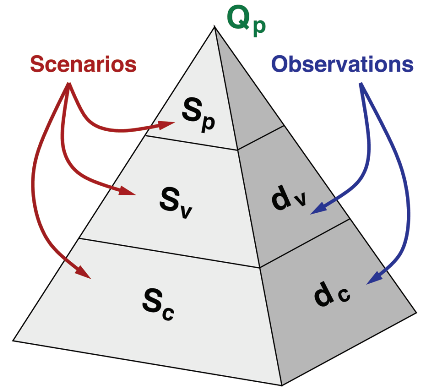

Figure 3. *The prediction pyramid.*

We represent these possible scenarios symbolically as a pyramid of complexity, as shown in Figure 3, in which the lowest level involves calibration and simple validation tests for components of the system or single physical phenomena. At this level experiments generally involve simple scenarios (S_c), and are inexpensive and copious. Multicomponent and multiphysics validation is done at a higher level in the pyramid, based on experiments with more complex scenarios (S_v). These experiments are designed to effectively challenge the ability of the model to predict the QoI's (the Q_p's in the diagram) and are generally more expensive and less numerous. Finally, at the top of the pyramid, the full system is used in the prediction scenarios S_p, and predictions q of Q_p are made. Commonly, no observations are available for validation at this level.

Once a calibrated model has been judged likely to be valid, it is used to compute a predictable quantity of interest Q_p, resulting in a distribution of predictions q. What do we now mean by quantified uncertainty? The answer depends on the purpose of the predictions. Generally, we would be concerned with some statistical properties of q, such as its moments, which effectively quantify the uncertainty in the prediction.

The Verification Process

As mentioned in Part I of this article, the computational model we use in a simulation is a corrupted version of the mathematical model on which the simulation is based. Two levels of possible corruption must be addressed by the verification processes. The first is the process of code verification, which addresses the fidelity of the computer code developed to implement the analysis of the class of models on which the simulation is based. Code verification is done to uncover "bugs" in the code, check computed results against benchmark problems with known solutions, assess the performance of algorithms, and explore a host of issues intended to verify that the code is functioning properly.

At the second level is solution verification—assessment of the accuracy of a feature of the solution that was the focus of the simulation. Because the goal is to predict specific quantities of interest, solution verification amounts to developing a posteriori error estimates for specific target outputs—the QoI's. This often challenging undertaking is a province of numerical mathematics. It is reasonable to ask whether such errors can actually be estimated. The answer is yes, and, indeed, the literature on this subject is extensive. The basic idea is to determine the degree to which the solution of the discrete problem fails to satisfy the governing equations of the mathematical model and, when the equations are not perfectly satisfied, to determine the residual. Remarkably, by processing this residual, it is possible to extract very good estimates of the actual errors in the QoI's targeted in the simulation.

Generalizations

The Bayesian framework allows a number of useful generalizations, one of which is the analysis of multiple models. In this framework, we identify a set M of possible models of a phenomenon, and then construct a conditional probability $\rho(d|M_j)$ that model M_j in the set produces the observations d. This is called the *evidence of plausibility*. (This is not to be confused with Dempster–Shafer evidence, which is quite a different notion.) By Bayes's theorem, the posterior probability or plausibility of the model M_j, given the data, is $\rho(M_j|d,M) \propto \rho(d|M_j)\rho(M_j|M)$.

In this way all models in the set can be compared for consistency with the data d. The idea of choosing the best models for given data among a class of models is an active area of study in modern Bayesian statistics.[†]

Algorithms for Uncertainty Quantification

The innocent-looking Bayesian update formula given in Part I,

$$\sigma(m|d) = \frac{\rho_M(m)\theta(d|m)}{\rho_D(d)}, \tag{1}$$

can actually be computationally formidable to evaluate, and a number of numerical algorithms have been developed for this purpose. The fundamental problem is to generate samples from, or other representations of, the posterior probability distribution in, say, the calibration or inverse problem. The most widely used algorithm is Markov chain Monte Carlo (MCMC), of which there are many variants.[‡] MCMC calculations generally converge slowly and require many evaluations of the physical model (evaluating the likelihood), making them computationally expensive. Alternative representations of the posterior pdf's commonly suffer from the curse of dimensionality—the computational complexity, in other words, increases catastrophically with the number of uncertain parameters. Once posterior pdf's of model parameters are known, the computational challenge is to calculate the probability distribution of output QoI's (the forward uncertainty propagation problem). Here, classic algorithms like Monte Carlo sampling are commonly used, but they too converge slowly. As for inverse problems, each sample requires ev-aluation of the model, which makes them computationally expensive. More advanced algorithms that converge more rapidly also commonly suffer from the curse of dimensionality. The development of rapidly

[‡]See, for example, C.P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer, 2004.

[†]See, for example, C.P. Robert, *The Bayesian Choice*, Springer, 2007.

converging algorithms for both forward and inverse uncertainty propagation problems that avoid the curse of dimensionality is a great need and a topic of current research.

A promising approach for addressing these computational issues is to use more information about the structure of the physical model than is available from simple point evaluations of the input–output map. In particular, algorithms now being developed use derivatives of the QoI's with respect to the uncertain inputs; in some cases these algorithms can accelerate convergence by orders of magnitude. When input parameters far outnumber QoI's, gradients and even Hessians of the QoI's can be efficiently computed from solutions of adjoint problems, which are also used to determine a posteriori estimates of numerical errors. Regardless of the specifics, algorithms are needed that can provide predictions with quantified uncertainty, scaling to expensive models and large numbers of parameters.

Example

To see how the process described here works in a simulation, we consider a model of a turbulent boundary layer with several different pressure gradients.[§] A boundary layer is a thin region that forms near a solid surface when fluid flows over it, as in the flow of air over a car or airplane. Turbulent boundary layers mediate the transfer of momentum from the fluid to the surface, resulting in viscous drag. In our example, turbulence is represented with the Spalart–Allmaras (SA) model, which has seven parameters that have to be calibrated. Because turbulence models are imperfect, a statistical model of the model inadequacy is also considered. Experimental measurements of the streamwise velocity and of the wall shear stress have been made. In a Bayesian inference framework, these measurements, along with estimates of the uncertainty in the data, are used to determine the parameters for the SA model, both alone and with the inadequacy model. The posterior probability distributions of two SA model parameters for cases with and without the inadequacy model are shown in Figure 4. The parameter κ (which is identified with

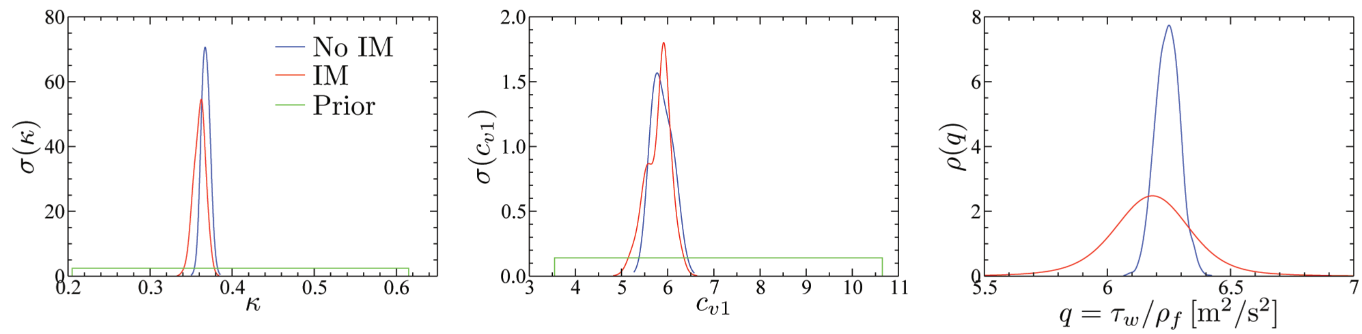

Figure 4. *Posterior marginal probability distributions of the parameters κ and c_{v1} in the Spalart–Allmaras turbulence model and probability distributions of the wall shear stress ($q = \tau_w/\rho_f$, the QoI, where ρ_f is the fluid mass density) in the prediction scenario. Shown are probability density functions for models with and without inadequacy models and the prior pdf's, which are uniform distributions over a selected range.*

the Karman constant) is well informed by the data, with rather narrow posterior distributions, while c_{v1} is less well informed (with a distribution 2.5 times wider).

We wish to use the calibrated model to predict the wall shear stress for a stronger pressure gradient than that for which data are available. This is the QoI, which we want to predict to within, say, 5%. To see if the model is valid for this purpose, we again use Bayesian inference to evaluate the posterior probabilities, given the data, of the model with and without the inadequacy model (this is the plausibility). The probability of the model with no inadequacy model is essentially zero (order 10^{-10}), indicating that the inadequacy model is necessary for consistency with the data. Furthermore, as shown in Figure 4, where the predicted distributions of the QoI are plotted, the inadequacy model makes a difference of more than 5% to the prediction. This indicates that, according to the available data, the model as calibrated (without the inadequacy model) is not valid for predicting the QoI within the required tolerance. In particular, without the inadequacy model, the uncertainty in the prediction is greatly underestimated. These results, however, say nothing about the validity of the model that includes the inadequacy model.

Summary

We have seen that the consideration and quantification of uncertainty in computational predictions introduce new challenges to the processes of calibration and validation of mathematical models and their use in making predictions. In calibration, uncertainties in experimental observations and uncertainties arising from inadequacies or randomness in the mathematical models must be accounted for to obtain estimates of input parameters and their uncertainties. The validation process is designed to challenge the models in their ability to predict the quantities of interest, and must account for the uncertainty of the observations and of the model and its parameters. Finally, issuing predictions of the quantities of interest involves estimating the QoI's and their uncertainties given the uncertainties in models and inputs. These uncertainty considerations require a mathematical representation of uncertainty, and we argue that probability in the context of Bayesian statistics provides a particularly powerful representation. Many challenges remain, however, for the development of new or improved conceptual and computational tools needed to treat uncertainty in computational simulations of complex systems. Among the challenges are representing model inadequacy, characterizing the usefulness of particular validation tests, determining the consequences of validation comparisons on the quantities of interest, posing validation metrics that are appropriate in various situations, and devising efficient algorithms for the required computations, especially when there are many uncertain parameters and the models are computationally expensive.

Challenges are to be expected in such a young field, but we build on a foundation in the philosophy of science that was laid long ago by the likes of the Reverend Thomas Bayes and Sir Karl Popper.

[§]For details, see Cheung et al., to appear in Reliab. Eng. Syst. Safety special issue on quantification of margins and uncertainty (QMU), 2010.

Acknowledgments

We have benefited from numerous discussions with I. Babuška and R. Tempone and from their published work on verification, validation, and uncertainty quantification. We thank the staff of the PECOS Center for their contributions to pursuing these ideas, and in particular T. Oliver, E. Prudencio, and S.H. Cheung for the results shown in Figure 4. The financial support of the PECOS Center by the Department of Energy under the National Nuclear Security Administration's Predictive Science Academic Alliance Program is gratefully acknowledged (DE–FC52–08NA28615).

Robert Moser is PI, and Tinsley Oden and Omar Ghattas are Co-PI's of the Center for Predictive Engineering and Computational Sciences (PECOS) of the Institute for Computational Engineering and Sciences (ICES) at the University of Texas at Austin.