ICES REPORT 10-39

October 2010

Computer Predictions with Quantified Uncertainty

by

Tinsley Oden, Robert Moser, and Omar Ghattas

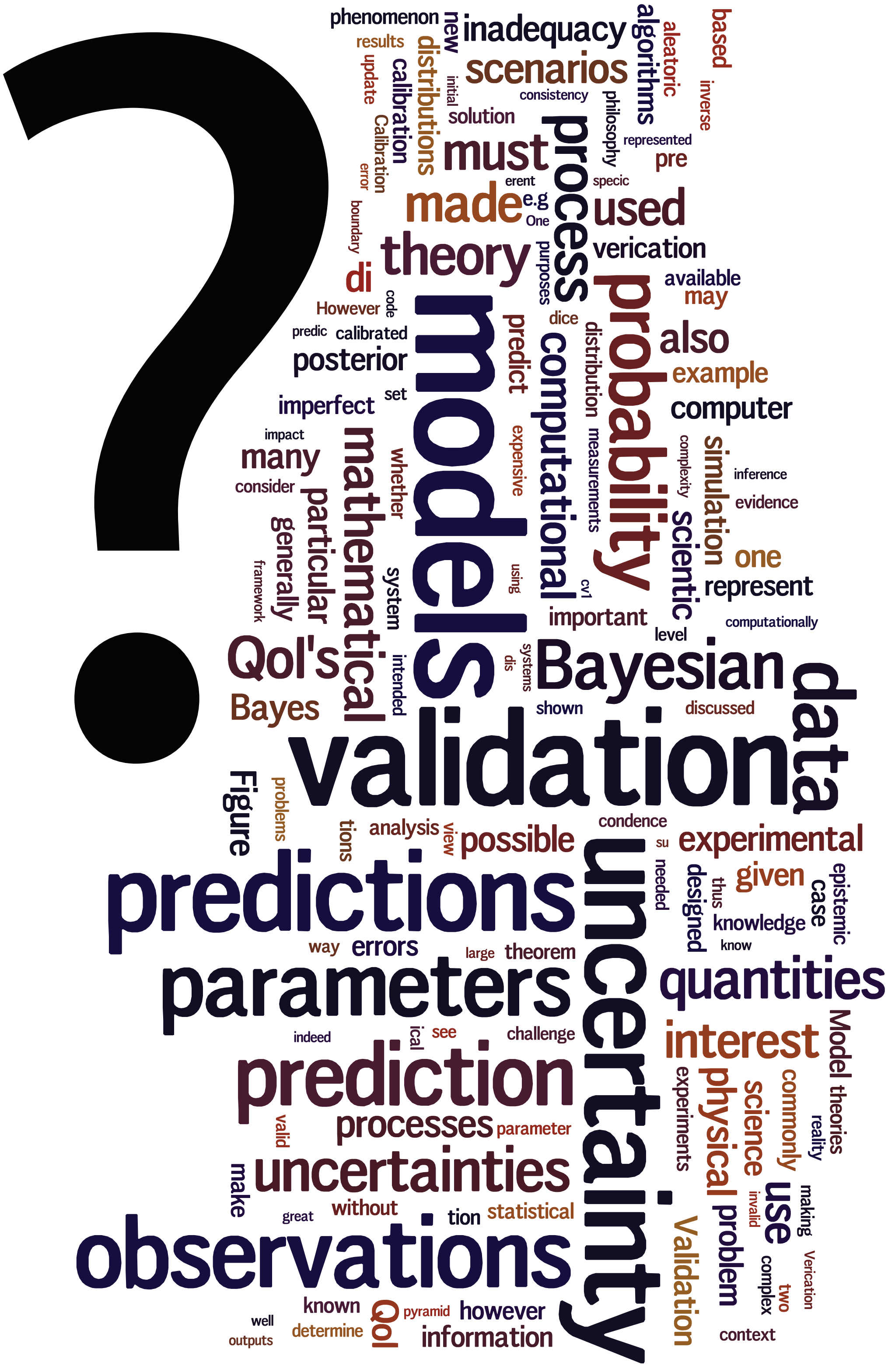
The Institute for Computational Engineering and Sciences
The University of Texas at Austin
Austin, Texas 78712

Reference: Tinsley Oden, Robert Moser, and Omar Ghattas, "Computer Predictions with Quantified Uncertainty", ICES REPORT 10-39, The Institute for Computational Engineering and Sciences, The University of Texas at Austin, October 2010.

PECOS

Computer Predictions with Quantified Uncertainty

by Tinsley Oden, Robert Moser, and Omar Ghattas



Computer Predictions with Quantified Uncertainty[1]

by Tinsley Oden, Robert Moser, and Omar Ghattas

Center for Predictive Engineering and Computational Sciences (PECOS)
Institute for Computational Engineering and Sciences
The University of Texas at Austin

August, 2010

Recently, a fresh look at how computer predictions are made and several old philosophical ideas are bringing about a revolution in computational science, which is creating a panorama of formidable new challenges and research opportunities. It has to do with what computer models were always intended to do: make predictions of physical reality. Now, however, we are asking computer models to predict events and processes of enormous importance in making critical decisions that affect our welfare and security, such as climate change, the performance of energy and defense systems, the biology of diseases, and the outcome of medical procedures. With such high stakes, we must insist that the predictions include concrete, quantifiable measures of uncertainty. In other words, we must know how good the predictions are. The term "predictive simulation" has thus taken on special meaning: the systemic treatment of model and data uncertainties and their propagation through a computational model to produce predictions of quantities of interest with *quantified uncertainty.*

Our goal is to use scientifically based predictions of physical reality to make informed decisions. So, what are "scientifically-based predictions?" They are forecasts of physical events and processes based on the methods of science: scientific theories (assertions about the underlying reality that brings about a phenomenon) and observations (knowledge received through the senses or the use of instruments).[2]

For our purposes, theory and observation, the fundamental pillars of science, can be cast as mathematical models: mathematical constructs that describe a system and represent knowledge of the system in a usable form. Mathematical models are thus abstractions of physical reality. They have been fundamental to all science and used successfully for millennia. However, mathematical models generally involve parameters that must be "tuned" so that the model best represents the particular system or phenomenon about which predictions are to be made. These are the moduli, coefficients, solution domains, boundary and initial data, etc. that distinguish one model from another within a class determined by the theory that has been selected to characterize the physical phenomenon. Unfortunately, these model parameters are commonly not known with great precision; they may vary from material to material, specimen to specimen, and case to case, or they may not be known at all. In short, they generally involve large uncertainties that can be resolved only with sufficient experimental evidence. On the other hand, the experimental observations themselves are often fraught with errors and uncertainties owing to imperfections in the instruments or the difficulty or impossibility of acquiring observational data relevant to the problem at hand.

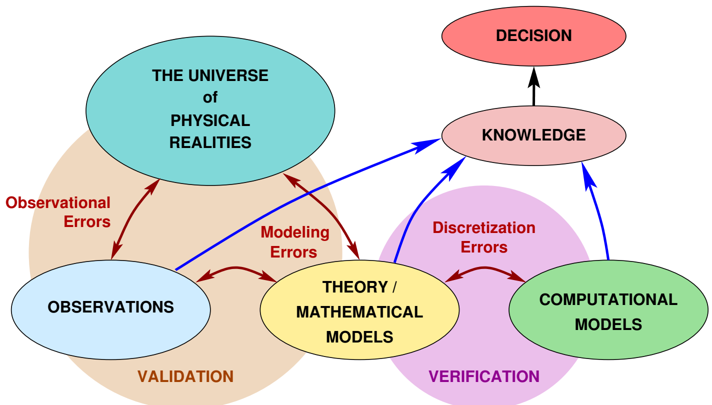


Figure 1: The imperfect paths to knowledge.

One of the greatest triumphs of modern science and technology is the digital computer, which makes possible the use of mathematical models of enormous complexity, leading to the recognition of computer modeling and simulation as the third indispensable pillar of science. But this comes at a cost. Mathematical models often must be corrupted to create computational models that render them amenable to solution via computer, and this corruption introduces more errors. Thus, a cascade of errors and uncertainties infect every aspect of scientifically-based predictions. The imperfect mathematical model of reality with the unknown or incomplete information on model parameters, with incomplete observational data or observations delivered by imperfect devices, and the corruption of the model itself through the discretization process necessary for computation, all lead to imperfect paths to knowledge, depicted symbolically in Figure 1.

How can we cope with these imperfections? It is here that the old ideas from philosopher Karl Popper and theologian and mathematician Thomas Bayes reemerge.

The Popperian view of scientific philosophy was that a hypothesis did not qualify as a legitimate scientific theory unless it could be refuted by experimental evidence. This is the *Principle of Falsification*, the answer to the *Problem of Induction*, which had prevailed as a great paradox in scientific philosophy from the times of David Hume in the 18th century. In this view, a scientific theory can never be validated; it can only be

[1]An edited version of this note is to be published in *SIAM News*

[2]This subject is the focus of the National Nuclear Security Administration's Predictive Science Academic Alliance Program (PSAAP). The mathematical foundations of Verification, Validation, and Uncertainty Quantification are also the focus of a new study organized by the Board of Mathematical Sciences and their Applications of the National Research Council.

invalidated by contrary experimental evidence. Experimental observations, then, are intrinsically interwoven into the scientific method: without a possible program of experiments, scientific theories are not legitimate. Analogously, regarding the mathematical models we use in computer simulations, experimental observations are necessary to be able to test them for invalidity.

But the Popperian view has been criticized for being a canon of *Objective Philosophy*, which is regarded by many as too rigid, not reflecting the way science is really done. Indeed, theories and their consistency with experiments, it is argued, must be judged in terms of probabilities in light of the evidence. This softer interpretation of validity more closely fits the philosophy of Thomas Bayes, from whose writings grew Bayesian statistics and the simple, but powerful, Bayes theorem. While laid down two and a half centuries ago, the Bayesian view of the scientific method and its role in prediction has only recently been recognized as a great unifying framework that elegantly ties all the components of predictivity together.[3] Before discussing this framework, we consider the *processes* of predictive science, which consists of several important stages:

1. **Identifying Quantities of Interest:** An idea not widely appreciated by computer modelers only a few years ago is that the entire exercise of simulating a physical phenomenon begins with a clear specification of the goals of the calculation: what are the target outputs, what are the principal "quantities of interest" (QoI's)? At the heart of the verification and validation processes discussed below is the realization that models may be perfectly acceptable for simulating some features of a phenomenon, while being entirely inadequate for modeling other features. So, to even ask whether a model is or is not invalid, one must specify in advance the particular quantities we wish to predict, and situations or scenarios in which they are to be predicted.

   Clearly, the quantities of interest for a predictive simulation are dictated by the larger purpose of the prediction, for example, to decide whether an aircraft design is safe, or whether a medical procedure will be successful. Often, such quantities of interest appear mathematically as numbers; that is, functionals on the solutions of the governing equations. In any case, the verification and validation process, and indeed the prediction itself, must be built around the calculation of specific QoI's.

2. **Verification:** Since computational models are corrupted (approximate) versions of the mathematical models on which they are based, it is critical that we confirm that a computation accurately approximates results for the quantities of interest produced by the underlying mathematical model. This is the process of verification, which is designed to detect and control errors brought about by the corruption of the mathematical model through discretization, and errors arising from the implementation of the model in software (bugs).

3. **Calibration:** Mathematical models generally have adjustable parameters whose values are to be assigned to bring predictions into (closer) agreement with experimental observations. Measurements are made of the response of components of a system and then the corresponding model parameters are inferred from the measurements in what amounts to an inverse problem (given the observed outputs, find the corresponding input parameters). Such inverse problems are commonly ill-posed or under-determined, requiring the modeler to provide further information regarding the nature of the parameters and/or the quality of the expected correspondence between model and observations. As we will see below, when considering uncertainty, this can be done through Bayesian inference.

4. **Validation:** The mathematical models used in predictive simulation are intended to represent the physical phenomena important to the prediction. The validation process is designed to build confidence that the model can indeed accurately predict the quantities of interest. Thus, validation addresses the question of legitimacy of the theoretical model for the purposes of the predictions that are to be made. Validation processes necessarily involve a carefully designed program of physical experiments that are intended to assess the degree with which the model can reproduce experimental observations. But, the process is complicated by the fact that the quantities of interest are commonly not accessible for observation. To determine whether a set of experimental observations invalidate a model for the purposes of a particular prediction, the modeler must thus consider the degree of disagreement between model and observations, the impact of these disagreements on predictions of the quantities of interest, and the tolerance for error in the predictions.

The use of the word "process" in the definitions of verification and validation is important: it is generally impossible to completely verify a computational model, and, in line with Popper's principle, a model can never actually be validated.

Uncertainty Formulations. The above descriptions of the predictive simulation processes are independent of considerations of uncertainty; however, how these processes are actually conducted depends critically on the treatment of uncertainty. In treating uncertainty in predictive simulation, we must first decide how to represent it mathematically. A number of different representations have been proposed, based on such formalisms as the Dempster-Shafer Theory of Evidence, the theory of fuzzy sets, interval analysis, worst-case scenarios, and many more. The Bayesian philosophy discussed above suggests representing uncertainty with probability, a particularly powerful approach that will be discussed in some detail here. However, the underlying concepts are more general, and are relevant regardless of the uncertainty formalism used.

Representing uncertainty probabilistically requires some explanation. Probability most commonly represents random

[3] see Howson & Urbach, *Scientific Reasoning: the Bayesian Approach*, 2006, for a modern account. The view that scientific theories must be judged in terms of their probabilities was shared by many of the foremost scientific minds of history, including Laplace (1749-1827), Poincaré (1854-1912), Jaynes (1922-1998), and others.

processes, and indeed one type of uncertainty in computational predictions arises from randomness. These are *aleatoric* uncertainties, which are irreducible, in that better data or improved models cannot reduce this uncertainty. However, in many cases, the dominant uncertainties arise from lack of knowledge, particularly lack of knowledge of physical model parameters and imperfections in the mathematical models themselves. These are *epistemic* uncertainties, which can in principle be reduced. In the Bayesian formulation, epistemic uncertainties are also represented probabilistically. In this case, probability represents our confidence in some proposition, given all of the currently available information. Fundamentally, Bayes' theorem gives us the formalism to update that confidence (probability) when new information becomes available.

In the Bayesian uncertainty formulation, no distinction is made between epistemic and aleatoric uncertainties. Instead, this distinction is one of interpretation, in the context of the prediction to be made, of what information could be obtained on which to base the prediction. As an example, consider an epitome of random processes, the throwing of dice ("aleatoric" is from the Latin *alea* meaning dice game or die). The outcome of a throw of fair dice is, in the context of games of chance, considered to be random. However, the physical processes controlling this outcome are themselves deterministic. To make a reliable prediction, all that is lacking is detailed knowledge of the initial conditions of the throw. At the casino, this information is interpreted to be unattainable, and so a dice game is, appropriately, ruled by aleatoric uncertainty. In another context, however, it might be possible to measure or control these initial conditions (perhaps with a dice throwing machine in the laboratory). In this case an imperfect knowledge of the initial conditions and therefore the outcome of a throw is an epistemic uncertainty.

While Bayesian analysis does not distinguish between aleatoric and epistemic uncertainty, the distinction is nonetheless useful. For example, consider simulations intended to predict the failure of widgets built at the widget factory. If the uncertainty in the simulations is primarily aleatoric (e.g., due to random variations in the precise dimensions of the parts), then the predicted probability of failure is the predicted percentage of widgets from the factory that will fail. On the other hand, if the uncertainty is primarily epistemic (e.g., from imperfections in the computational model), then the output of the factory will presumably all fail or all work, and the failure probability characterizes our confidence in predicting which will happen. Clearly the difference between these two scenarios would be important to the warranty claims office at the factory.

Calibration and Statistical Inverse Problems. In the Bayesian analysis of computational predictions, uncertainty from three sources is introduced, and, as discussed above, each is represented probabilistically.[4] We have:

- *Model parameters.* Instead of a list of parameters m, we begin with the prior joint probability density function (pdf) $\rho_M(m)$ describing what we know about the parameters before we begin the modeling exercise. They are gleaned from the literature, previous experiments or archival databases, or are specified to represent almost complete ignorance of the parameters. Calibration via Bayesian inference will be used to update these parameter distributions to make the theory (model) consistent with particular observations (data).
- *Experimental observations.* As noted, the observational data d also will have uncertainties, also represented by pdfs $\rho_D(d)$.
- *Theoretical model.* The mathematical model with which the predictions are to be made is a key component of the overall process—it defines the way given parameter sets m are mapped into the theoretical observations d. But the theory (model) may not be completely reliable, so our uncertainty or lack of confidence in the theory must be included in the analysis. We do this by expressing the theory as a conditional probability distribution that maps the model inputs to a probability distribution of the outputs. Let the conditional probability $\theta(d|m)$ ("θ" for "theory") be the probability that the model will predict the observations d given the parameters m. This is also known as the *likelihood*.

Various forms of Bayes' theorem pull these elements together to characterize exactly what is known about the model parameters, expressed as the posterior pdf $\sigma(m|d)$ of the parameters m conditioned on the data d. From Bayes' theorem,

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

This expression (1) is essentially a statistical calibration, as it statistically infers the (posterior) distribution of parameters $\sigma(m|d)$ that fit the theoretical model to the observations d.

Bayes' theorem and the statistical calibration process embodied in (1) are also the keys to validation, and ultimately prediction with quantifiable uncertainties.

The Validation Process. Having developed and calibrated a model to be used for a particular prediction, we are now ready to begin the validation process to assess the suitability of the calibrated model for the prediction. To this end, a new set of observations are made, that are designed to challenge the model. The new observational data and their uncertainties are then to 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 prediction quantities of interest or for the prediction scenarios; they are instead 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 answer to these questions depends critically on the character of the prediction problem at hand. To make the nature of the issues clear, first observe that the validation

[4]See, e.g., J. Kaipio & E. Somersalo *Statistical and Computational Inverse Problems*, Springer, 2005.

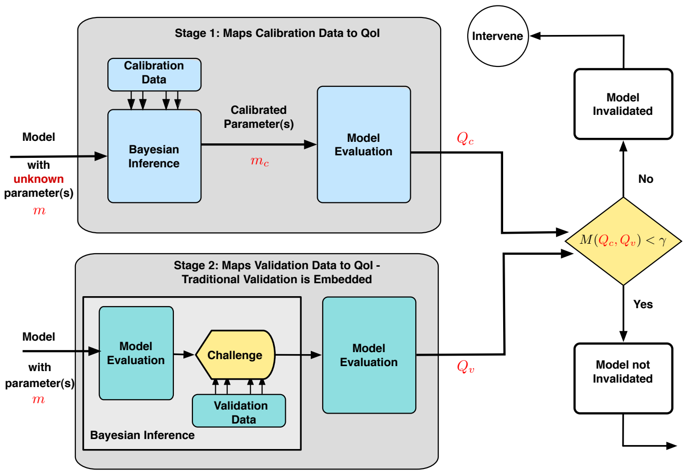


Figure 2: Schematic of the validation process.

questions posed here are somewhat different from the usual experimental falsification of scientific theories. In the latter, one aspires to discern the laws of nature (theories), so that a disagreement between theory and experiment falsifies the theory (in the context of uncertainty, a low probability that the experimental observations is consistent with the theory makes the theory improbable). For example, measurements taken in relativistic scenarios will, we know, falsify Newtonian mechanics as a scientific theory. Here, however, we seek useful models, which we acknowledge may be imperfect, and ask the narrower question of whether they are valid for the purposes of making particular predictions. For example, despite the fact that Newtonian mechanics has been falsified as a theory, it provides a valid model for making many predictions. Unfortunately, models used for making predictions in complex systems generally do not have well characterized domains of applicability (like Newtonian mechanics). Validation processes as discussed here are thus needed to detect whether the models proposed for use in a particular prediction are invalid for that purpose.

The validation process is centered on the question of what discrepancies between models and experimental observations imply about the reliability of the specific QoI predictions to be made using the models. To address this question, Bayes' 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 resulting change in the predictions due to the updates is too large in some appropriate measure (see below), then the model is inconsistent with the validation data in such a way as to influence the predictions, and the model and its calibration are invalid for predicting those QoI's.[5] A flow chart illustrating 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 sufficiently rich parametrization to represent most any observations. In this case, 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 that inconsistency has on the prediction of the QoI's. In situations where such a parametric representation is not possible, the model will need to be enriched to allow it to 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. There may be other ways to evaluate the consequences of the validation observations on the QoI's as well.

The impact of the validation observations on the predictions of 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 being asked about the QoI. One might, for example, need to know the most likely value of the QoI with a specified tolerance for error, in which case inconsistencies in the prediction of the most probable value should be evaluated. In other situations, it might be important to determine the probability that a presumably unlikely failure will occur. Then, the invalidation metric should assess the consistency of this prediction.

Another important issue in validation is the selection of validation scenarios and validation measurements to challenge models in a way that is similar to the prediction scenario. Generally, the modeler makes such selections using his understanding of the potential weaknesses in the models (e.g., what questionable assumptions were made), and his judgment regarding the important aspects of the models for the predictions to be made. It is also possible to inform this selection using the computational model by analyzing sensitivities of the QoI's to parameter variation and model variation, and by maximizing the information about the QoI that is expected to be gained. In complex multi-physics systems, designing a validation process is also complicated by the need to perform validation at multiple (many) levels of complexity, from simple scenarios involving models of a single physical phenomenon, to complex scenarios involving all of the models needed in the prediction scenario.

We represent these possible scenarios symbolically in a pyramid of complexity, such as that 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 ($\mathrm{S_c}$), and are inexpensive and copious. Multi-component and multi-physics validation is done at a higher level in the pyramid based on experiments with more complex scenarios ($\mathrm{S_v}$). They are designed to effectively challenge the ability of the model to predict the QoI's (the $\mathrm{Q_p}$'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 $\mathrm{S_p}$, and predictions q of $\mathrm{Q_p}$ are made. Commonly no observations are available for validation at this level.

[5]See Babuška *et al*, *Comput. Methods Appl. Mech. Engrg.* **197**, 2517-2539, 2008.

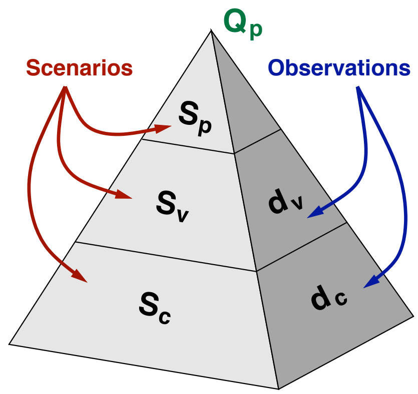


Figure 3: The prediction pyramid.

Let us suppose that a calibrated model has been judged likely to be valid through a validation process. It is then 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? This depends on the purpose of the predictions. Generally, one would be concerned about some statistical properties of q such as its moments. These statistical properties of q effectively quantify the uncertainty in the prediction.

The Verification Process. We have not yet taken up the problem that the computational model we use in a simulation is a corrupted version of the mathematical model on which the simulation is based. There are two levels of possible corruption that must be addressed by verification processes. First, there is the process of code verification that 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 aims at uncovering "bugs" in the code, checking computed results against benchmark problems with known solutions, assessing the performance of algorithms, and exploring a host of issues intended to verify that the code is properly functioning as designed. Second, solution verification is the assessment of the accuracy of a particular feature of the solution that was the focus of the simulation. Since 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 is often a challenging undertaking, and is a province of numerical mathematics. One may ask, is it possible to actually estimate such errors? The answer is yes, and indeed, a large literature exists on this subject. 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, determining the residual when equations are not perfectly satisfied. Remarkably, by processing this residual, it is possible to extract very good estimates of the actual errors in the quantities of interest 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, one identifies a set M of possible models of a phenomenon, and then constructs 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*[6]. Applying Bayes' theorem, the posterior probability or plausibility of the model M_j given the data is $\rho(M_j|d, M) \propto \rho(d|M_j)\rho(M_j|M)$. This allows one to compare all models in the set 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.[7]

Algorithms for Uncertainty Quantification. The innocent-looking Bayesian update formula given in (1) can actually be computationally formidable to evaluate, and a number of numerical algorithms have been developed to do so. The fundamental problem is to generate samples from, or other representations of, the posterior probability distribution in the calibration or inverse problem, for example. The most widely used algorithm is Markov chain Monte Carlo (MCMC), of which there are many variants.[8] However, 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 pdfs commonly suffer from the curse of dimensionality, in which the computational complexity increases catastrophically with the number of uncertain parameters. Once posterior pdfs of model parameters are known, the computational challenge is to calculate the probability distribution of output QoI's (the forward uncertainty propagation problem). Here classical sampling algorithms like Monte Carlo sampling are commonly used, but also converge slowly. As for inverse problems, each sample requires evaluation of the model, so this is computationally expensive. More advanced algorithms that converge more rapidly also commonly suffer from the curse of dimensionality. The development of rapidly 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 to address 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 that use derivatives of the QoI's with respect to the uncertain inputs are being developed, and in some cases can accelerate convergence by orders of magnitude. When there are many more input parameters than 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, advances in algorithms to provide predictions with quantified uncertainty that can scale to expensive models and large numbers of parameters are needed.

Example. To see how the process described here works in a simulation, we consider a model of a turbulent boundary

[6]This is not to be confused with Dempster-Shafer evidence, which is a quite different notion.

[7]See e.g., C.P. Robert, The Bayesian Choice, Springer, 2007.

[8]See, e.g., C.P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer, 2004.

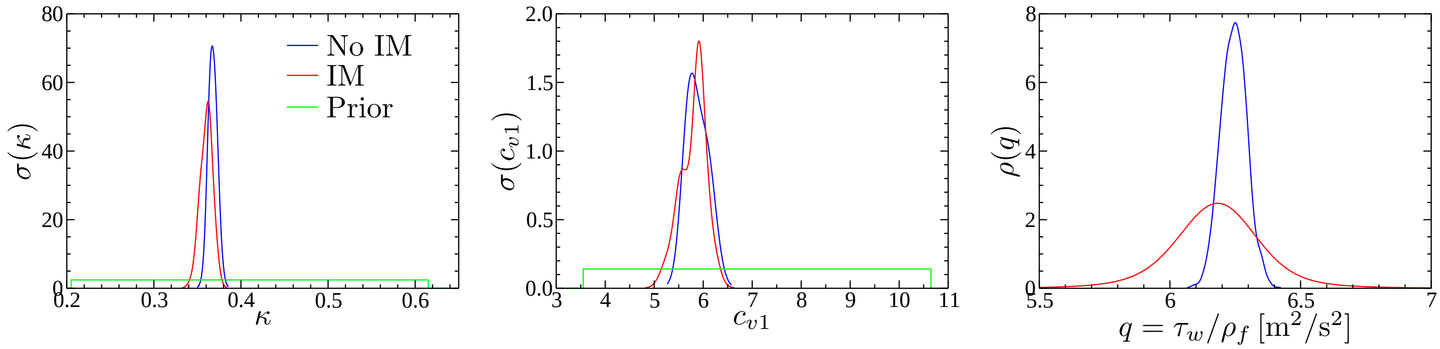


Figure 4: Posterior marginal probability distributions of the parameters κ and c_{v1} in the SA turbulence model and probability distributions of the wall shear stress ($q = \tau_w/\rho_f$ the QoI, where ρ_f is the fluid mass density) in the prediction scenario. Shown are pdf's for models with and without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

layer with several different pressure gradients.[9] A boundary layer is a thin region that forms near a solid surface when fluid flows over it, as, for example, when air flows 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 model, turbulence is represented with the Spalart-Allmaras (SA) model, which has 7 parameters that have to be calibrated. In addition, 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. Using Bayesian inference, these measurements along with estimates of the uncertainty in the data are used to determine the parameters for the SA model 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 the Karman constant) is well informed by the data, with rather narrow posterior distributions, while c_{v1} is less well informed (2.5 times wider distribution, relatively).

We wish to use the calibrated model to make predictions of 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. Further, as shown in Figure 4 where the predicted distributions of the QoI are plotted, the inadequacy model makes more than a 5% difference 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, note that 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.

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

Summary. We have seen that the consideration and quantification of uncertainty in computational predictions introduces 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 of or any 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 have argued that probability in the context of Bayesian statistics provides a particularly powerful representation. Many challenges remain, however, to develop new or improved conceptual and computational tools needed to treat uncertainty in computational simulations of complex systems. These include 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 to perform 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 Rev. Bayes and Sir Popper.

Acknowledgments. We have benefited from numerous discussions with Profs. 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 to Drs. 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 (NNSA) Predictive Science Academic Alliance Program (PSAAP) is gratefully acknowledged (award number DE-FC52-08NA28615).

The Institute for Computational Engineering and Sciences
ICES
THE UNIVERSITY OF TEXAS AT AUSTIN

www.ices.utexas.edu/
phone: (512) 471-3312