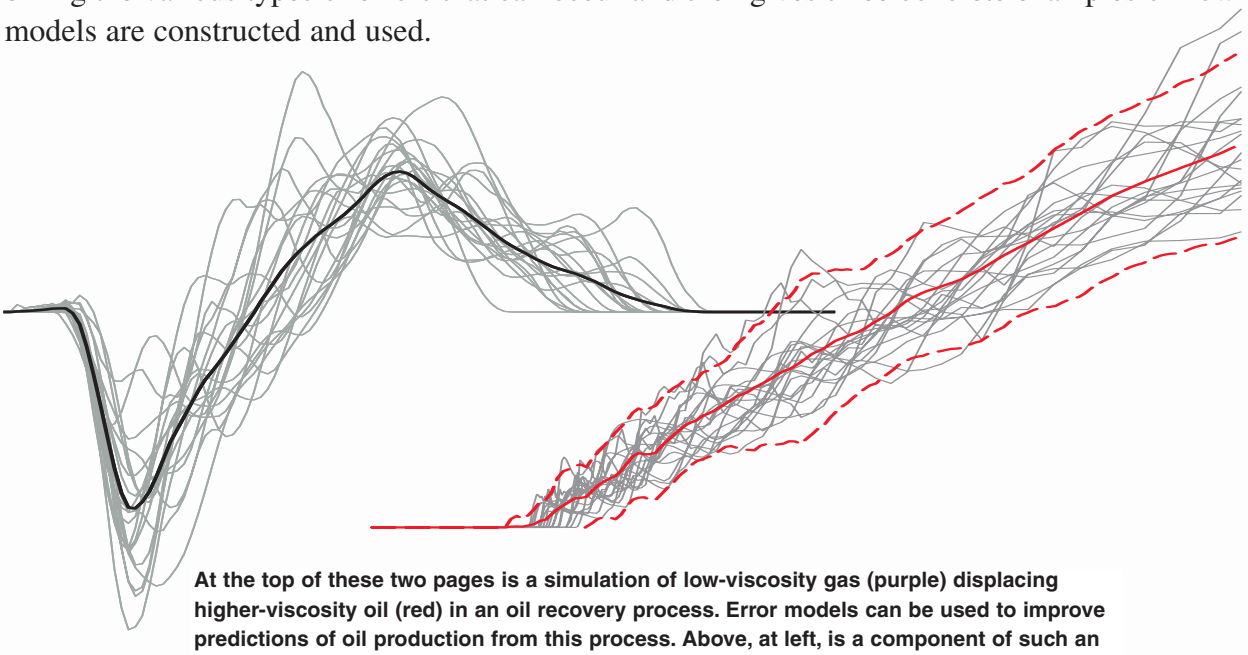
Error Analysis and Simulations of Complex Phenomena

Michael A. Christie, James Glimm, John W. Grove, David M. Higdon, David H. Sharp, and Merri M. Wood-Schultz

Large-scale computer-based simulations are being used increasingly to predict the behavior of complex systems. Prime examples include the weather, global climate change, the performance of nuclear weapons, the flow through an oil reservoir, and the performance of advanced aircraft. Simulations invariably involve theory, experimental data, and numerical modeling, all with their attendant errors. It is thus natural to ask, "Are the simulations believable?" "How does one assess the accuracy and reliability of the results?" This article lays out methodologies for analyzing and combining the various types of errors that can occur and then gives three concrete examples of how error models are constructed and used.



At the top of these two pages is a simulation of low-viscosity gas (purple) displacing higher-viscosity oil (red) in an oil recovery process. Error models can be used to improve predictions of oil production from this process. Above, at left, is a component of such an error model, and at right is a prediction of future oil production for a particular oil reservoir obtained from a simple empirical model in combination with the full error model.

Reliable Predictions of Complex Phenomena

There is an increasing demand for reliable predictions of complex phenomena encompassing, where possible, accurate predictions of full-system behavior. This requirement is driven by the needs of science itself, as in modeling of supernovae or protein interactions, and by the need for scientifically informed assessments in support of high-consequence decisions affecting the environment, national security, and health and safety. For example, decisions must be made about the amount by which greenhouse gases released into the atmosphere should be reduced, whether and for what conditions a nuclear weapon can be certified (Sharp and Wood-Schulz 2003), or whether development of an oil field is economically sound. Large-scale computer-based simulations provide the only feasible method of producing quantitative, *predictive* information about such matters, both now and for the foreseeable future. However, the cost of a mistake can be very high. It is therefore vitally important that simulation results come with a high level of confidence when used to guide high-consequence decisions.

Confidence in expectations about the behavior of real-world phenomena is typically based on repeated experience covering a range of conditions. But for the phenomena we consider here, sufficient data for high confidence is often not available for a variety of reasons. Thus, obtaining the needed data may be too hazardous or expensive, it may be forbidden as a matter of policy, as in the case of nuclear testing, or it just may not be feasible. Confidence must then be sought through understanding of the scientific foundations on which the predictions rest, including limitations on the experimental and calculational data and numerical methods used to make the prediction. This understanding must be sufficient to allow quantitative estimates of the level of accuracy and limits of applicability of the simulation, including evidence that any factors that have been ignored in making the predictions actually have a small effect on the answer. If, as sometimes happens, high-confidence predictions cannot be made, this fact must also be known, and a thorough and accurate uncertainty analysis is essential to identify measures that could reduce uncertainties to a tolerable level, or mitigate their impact.

Our goal in this paper is to provide an overview of how the accuracy and reliability of large-scale simulations of complex phenomena are assessed, and to highlight the role of what is known as an error model in this process.

Why Is It Hard to Make Accurate Predictions of Complex Phenomena?

We begin with a couple of examples that illustrate some of the uncertainties that can make accurate predictions difficult. In the oil industry, predictions of fluid flow through oil reservoirs are difficult to make with confidence because, although the fluid properties can be determined with reasonable accuracy, the fluid flow is controlled by the poorly known rock permeability and porosity. The rock properties can be measured by taking samples at wells, but these samples represent only a tiny fraction of the total reservoir volume, leading to significant uncertainties in fluid flow predictions. As an analogy of the difficulties faced in predicting fluid flow in reservoirs, imagine drawing a street map of London and then predicting traffic flows based on what you see from twelve street corners in a thick fog!

In nuclear weapons certification, a different problem arises. The physical processes in an operating nuclear weapon are not all accessible to laboratory experiments (O'Nions et al. 2002). Since underground testing is excluded by the Comprehensive Test Ban Treaty (CTBT), full system predictions can only be compared with limited archived test data.

The need for reliable predictions is not confined to the two areas above. Weather forecasting, global climate modeling, and complex engineering projects, such as aircraft design, all generate requirements for reliable, quantitative predictions—see, for example, Palmer (2000) for a study of predictability in weather and climate simulations. These often depend on features that are hard to model at the required level of detail—especially if many simulations are required in a design-test-redesign cycle.

More generally, because we are

dealing with complex phenomena, knowledge about the state of a system and the governing physical processes is often incomplete, inaccurate, or both. Furthermore, the strongly nonlinear character of many physical processes of interest can result in the dramatic amplification of even small uncertainties in the input so that they produce large uncertainties in the system behavior. The effects of this sensitivity will be exacerbated if experimental data are not available for model selection and validation. Another factor that makes prediction of complex phenomena very difficult is the need to integrate large amounts of experimental, theoretical, and computational information about a complex problem into a coherent whole. Finally, if the important physical processes couple multiple scales of length and time, very fast and very high memory capacity computers and sophisticated numerical methods are required to produce a high-fidelity simulation. The examples discussed in this article exhibit many of these difficulties, as well as the uncertainties in prediction to which they lead.

To account for such uncertainties, models of complex systems and their predictions are often formulated probabilistically. But the accuracy of predictions of complex phenomena, whether deterministic or probabilistic, varies widely in practice. For example, estimates of the amount of oil in a reservoir that is at an early stage of development are very uncertain. Large capital investments are made on the basis of probabilistic estimates of oil in place, so that the oil industry is fundamentally a risk-based business. The estimates are usually given at three confidence levels: $p90$, $p50$, and $p10$, meaning that there is a 90 percent, 50 percent, and 10 percent chance, respectively, that the amount of oil in place will be greater than the specified reserve level. Figure 1 shows a schematic plot (based on a real North Sea example) of estimated reserves as a function of time. The plot clearly shows that, as more information about the reservoir was acquired during the course of field development, estimates of the range of reserves changed out-

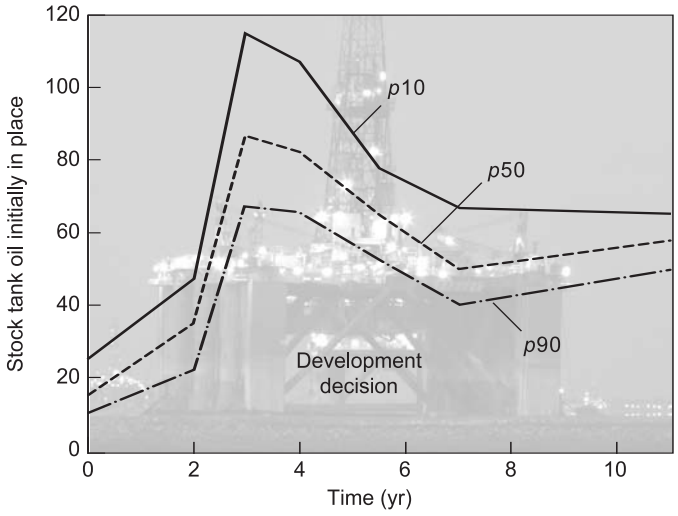


Figure 1. Oil-in-Place Uncertainty Estimate Variation with Time
This figure shows estimates of *p*90, *p*50, and *p*10 probabilities that the amount of oil in a reservoir is greater than the number shown. The estimated probabilities are plotted as a function of time. The variations shown indicate the difficulties involved in accurate probability estimations. [Photo courtesy of Terrington (York) Ltd.]

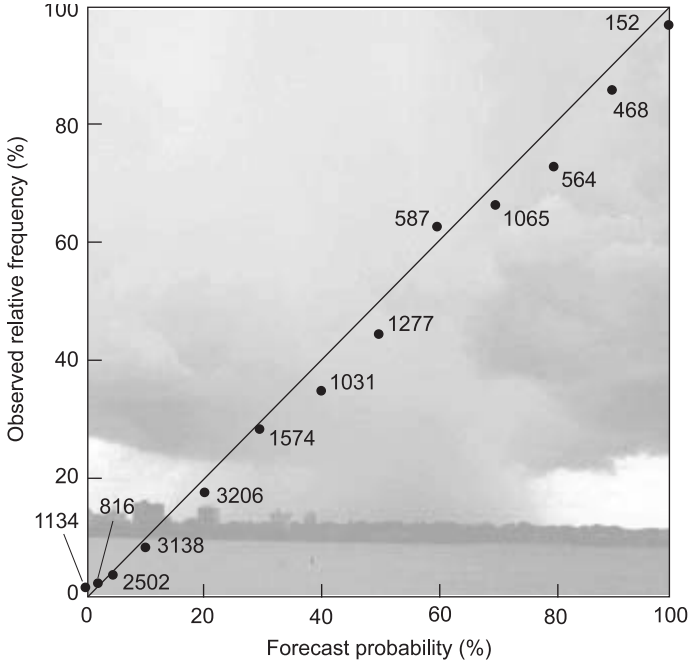


Figure 2. Calibration Curve for Weather Forecasts
This plot shows estimates of the probability of precipitation from simulation forecasts vs the observed frequency of precipitation for a large number of observations. Next to each data point is the number of observations for that forecast.

side the initial prediction. In other words, the initial estimates of reserves, although probabilistic, did not capture the full range of uncertainty and were thus unreliable. This situation was obviously a cause for concern for a company with billions of dollars in investments on the line.

Probabilistic predictions are also used in weather forecasting. If the probabilistic forecast "20 percent chance of rain" were correct, then on average it would have rained on 1 in 5 days that received that forecast. Data on whether or not it rained are easily obtained. This rapid and repeated feedback on weather predictions has resulted in significantly improved reliability of forecasts compared with predictions of uncertainty in oil reserves. The comparison between the observed frequency of precipitation and a probabilistic forecast for a locality in the United States shown in Figure 2 confirms the accuracy of the forecasts.

This accuracy did not come easily, and so we next briefly describe two of the principal methods currently used to improve the accuracy of predictions of complex phenomena: calibration and data assimilation.

Calibration is a procedure whereby a simulation is matched to a particular set of experimental data by performing a number of runs in which uncertain model parameters are varied to obtain agreement with the selected data set. This procedure is sometimes called "tuning," and in the oil industry it is known as history matching. Calibration is useful when codes are to be used for interpolation, but it is of limited help for extrapolation outside the data set that was used for tuning. One reason for this lack of predictability is that calibration only ensures that unknown errors from different sources, say inaccurate physics and numerics, have been adjusted to compensate one another, so that the net error in some observable is small. Because different physical processes and numerical errors are unlikely to scale in the same way, a calibrated simulation is reliable only for the regime for which it has been shown to match experimental data.

In one variant of calibration, multiple simultaneous simulations are performed with different models. The "best" prediction is defined as a weighted average over the results obtained with the different models. As additional observations become available, the more successful models are revealed, and their predictions are weighted more heavily. If the models used reflect the range of modeling uncertainty, then the range of results will indicate the variance of the prediction due to those uncertainties.

Data assimilation, while basically a form of calibration, has important distinctive features. One of the most important is that it enables real-time utilization of data to improve predictions. The need for this capability comes from the fact that, in operational weather forecasting, for example, there is insufficient time to restart a run from the beginning with new data, so that this information must be incorporated on the fly. In data assimilation, one makes repeated corrections to model parameters during a single run, to bring the code output into agreement with the latest data. The corrections are typically determined using a time series analysis of the discrepancies between the simulation and the current observations. Data assimilation is widely used in weather forecasting. See Kao et al. (2004) for a recent application to shock-wave dynamics.

Sources of Error and How to Analyze Them

Introducing Error Models. The role of a thorough error analysis in establishing confidence in predictions has been mentioned. But evaluating the error in a prediction is often more difficult than making the prediction in the first place, and when confidence in the answer is an issue, it is just as important.

A systematic approach for determining and managing error in simulations is to try to represent the effects of inaccurate models, neglected phenomena, and limited solution accuracy using an error model.

Unlike the calibration and data assimilation methods discussed above, an error model is not primarily a method of increasing the accuracy of a simulation. Error modeling aims to provide an independent estimate of the *known* inadequacies in the simulation. An error model does not purport to provide a complete and precise explanation of observed discrepancies between simulation and experiment or, more generally, of the differences between the simulation model and the real world. In practice, an error model helps one achieve a scientific understanding of the knowable sources of error in the simulation and put quantitative bounds on as much of the error as possible.

Simulation Errors. Computer codes used for calculating complex phenomena combine models for diverse physical processes with algorithms for solving the governing equations. Large databases containing material properties such as cross sections or equations of state that tie the simulation to a real-world system must be integrated into the simulation at the lowest level of aggregation. These components and, significantly, input from the user of the code must be linked by a sophisticated computer science infrastructure, with the result that a simulation code for complex phenomena is an exceedingly elaborate piece of software. Such codes, while elaborate, still provide only an approximate representation of reality.

Simulation errors come from three

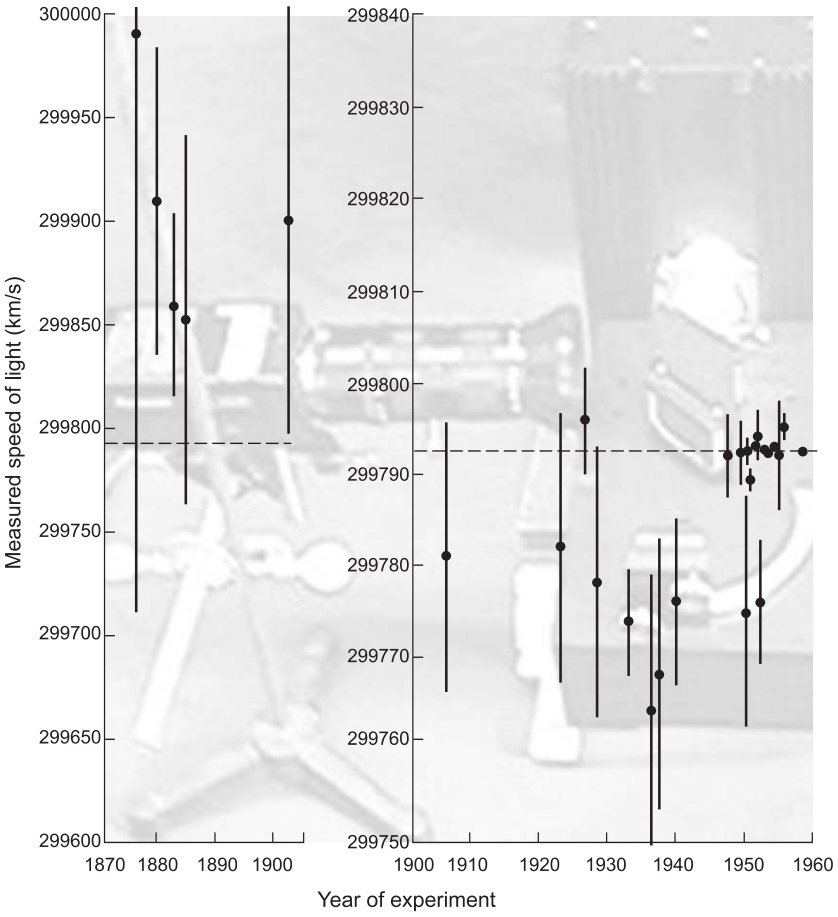


Figure 3. Uncertainties in Reported Measurements of the Speed of Light (1870–1960)

This figure shows measured values of the speed of light along with estimates of the uncertainties in the measured values up until 1960. The error bars correspond to the estimated 90% confidence intervals. The currently accepted value lies outside the error bars of more measurements than would be expected, indicating the difficulty of truly assessing the uncertainty in an experimental measurement. Refer to the article by Henrion and Fischoff on pp. 666–677 in *Heuristics and Biases* (2002) for more details on this and other examples of uncertainties in physical constants.

(Photo courtesy of Department of Physics, Carnegie Mellon University.)

main sources: inaccurate input data, inaccurate physics models, and limited accuracy of the solutions of the governing equations. Clearly, each of these generic sources of error is potentially important. A perfect physics model with perfect input data will give wrong answers if the equations are solved poorly. Likewise, a perfect solution of the wrong equations will also give incorrect answers. The relative importance of errors from each source is problem dependent, but each source of error must be evaluated. Our discussion of error models will reflect the above comments by categorizing simulation inadequacies as due to input, solution, and physics errors.

Input errors refer to errors in data used to specify the problem, and they include errors in material properties, the description of geometrical configurations, boundary and initial conditions, and others. Solution error is the difference between the exact mathematical solution of the governing equations for the model and the approximate solution of the equations obtained with the numerical algorithms used in the simulation. Physics error includes the effects of phenomena that are inadequately represented in the simulation, for example, the unknown details of subscale physics, such as the microscopic details of material that is treated macroscopically in the simulation. Evaluations of the effects of these details are typically based on statistical descriptions. The physics component of an error model is thus based on knowledge of aspects of the nominal model that need or might need correction.

Experimental Errors and Solution Errors. Much of our understanding of how to analyze errors comes from studies of experimental error. We will also see below that experimental and solution errors play a similar role in an uncertainty analysis. We therefore start by discussing experimental errors.

Experimental errors play an important role in building error models for simulations. First, they can bias conclusions that are drawn when simulation results are compared with measured data. Second, experimental errors affect the accuracy of simulations indirectly through their effects on databases and input data used in a simulation. Experimental errors are classified as random or systematic. Typically, both types of error are present in any particular application. A familiar example of a random error is the statistical sampling error quoted along with the results of opinion polls. Another type of random error is the result of variations in random physical processes, such as the number of radioactive decays in a sample per unit time. The signals from measuring instruments usually contain a compo-

nent that either is or appears to be random whether the process that is the subject of the measurement is random or not. This component is the ubiquitous "noise" that arises from a wide variety of unwanted or uncharacterized processes occurring in the measurement apparatus. The way in which noise affects a measurement must be taken into consideration to attain valid conclusions based on that data. Noise is typically treated probabilistically, either separately or included with a statistical treatment of other random error. However, systematic error is often both more important and more difficult to deal with than random error. It is also frequently overlooked, or even ignored.

To see how a systematic error can occur, imagine that an opinion poll on the importance of education was conducted by questioning people on street corners "at random"—not knowing that many of them were coming and going from a major library that happened to be located nearby. It is virtually certain that those questioned would on average place a higher importance on education than the population in general. Even if a very large number of those individuals were questioned, an activity that would result in a small statistical sampling error, conclusions about the importance of higher education drawn from this data could be incorrect for the population at large. This is why carefully conducted polls seek to avoid systematic errors, or biases, by ensuring that the population sampled is representative.

As a second example, suppose that 10 measurements of the distance from the Earth to the Sun gave a mean value of 95,000,000 miles due, say, to flaws in an electric cable used in making these measurements. How would someone know that 95,000,000 miles is the wrong answer? This error could not be revealed by a statistical analysis of only those 10 measurements. Additional, independent measurements made with independent measuring equipment would suggest that something was wrong if they were inconsistent with these results. However, the cause of the systematic error could only be identified through a physical understanding of how the instruments work, including an analysis of the experimental procedures and the experimental environment. In this example, the additional measurements should show that the electrical characteristics of the cable were not as expected. To reiterate, the point of both examples is that an understanding of the systematic error in a measured quantity requires an analysis that is independent of the instrument used for the original measurement.

An example of how difficult it can be to determine uncertainties correctly is shown in Figure 3, a plot of estimates of the speed of light vs the date of the measurement. The dotted line shows the accepted value, and the published experimental uncertainties are shown as error bars. The length of the error bars—1.48 times the standard deviation—is the "90 percent confidence interval" for a normally distributed uncertainty for the experimental error; that is, the experimental error bars will include the correct value 90 percent of the time if the uncertainty were assessed correctly. It is evident from the figure, however, that many of the analyses were inadequate: The true value lies outside the error bars far more often than 10 percent of the time. This situation is not uncommon, and it provides an example of the degree of caution appropriate when using experimental results.

The analysis of experimental error is often quite arduous, and the rigor with which it is done varies in practice, depending on the importance of the result, the accuracy required, whether the measurement technique is standard or novel, and whether the result is controversial. Often, the best way to judge the adequacy of an analysis of uncertainty in a complex experiment is to repeat the experiment with an independent method and an independent team.

Solution errors enter an analysis of simulation error in several ways. In addition to being a direct source of error in predictions made with a given model, solution errors can bias the conclusions one draws from comparing a model to data in exactly the same way that experimental errors do. Solution errors also can affect a simulation almost covertly: It is common for the data or the code output to need further processing before the two can be directly compared. When this processing requires modeling or simulation with a different code, then the solution error from that calculation can affect the comparison. As with experimental errors, solution errors must be determined independently of the simulations that are being used for prediction.

Using Data to Constrain Models

The scientific method uses a cycle of comparison of model results with data, alternating with model modification. A thorough and accurate error analysis is necessary to validate improvements. The availability of data is a significant issue for complex systems, and data limitations permeate efforts to improve simulation-based predictions. It is therefore important to use all relevant data in spite of differences in experiment design and measurement technique. This means that it is important to have a procedure to combine data from diverse sources and to understand the significance of the various errors that are responsible for limitations on predictability.

The way in which the various categories of error can affect comparison with experimental data and the steps

to be taken if the errors are too large are discussed in the next section. The comparison of model predictions with experimental data is often called the forward step in this cycle and is a key component in uncertainty assessments of a predicted result. The backward step of the cycle for model improvement, which is discussed next, is the statistical inference of an improved model from the experimental data. The Bayesian framework provides a systematic procedure for inference of an improved model from observations; lastly, we describe the use of hierarchical Bayesian models to integrate data from many sources.

Some discussion of the use of the terms "uncertainty" and "error" is in order. In general, any physical quantity, whether random or not, has a specific value—such as the number of radioactive decays in a sample of tritiated paint in a given 5-minute period. The difference between that actual number and an estimate determined from knowledge of the number of tritium nuclei present and the tritium lifetime is the error in that estimate. If the experiment were repeated many times, a distribution of errors would arise, and the probability density function for those errors is the uncertainty in the estimate.

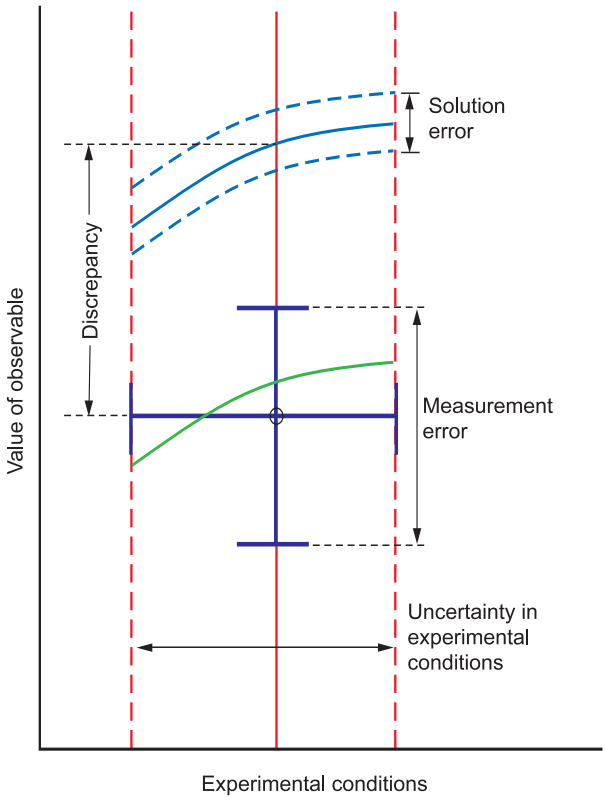


Figure 4. Comparing Experimental Measurements with Simulations
The green line shows the true, unknown value of an observable over the range of uncertainty in the experimental conditions, and the purple cross indicates the uncertainty in the observation. The discrepancy measures the difference between observation and simulation.

Decomposition of Errors. Our ability to predict any physical phenomenon is determined by the accuracy of our input data and our modeling approach. When the modeling input data are obtained by analysis of experiments, the experimental error and modeling error (solution error plus physics approximations) terms control the accuracy of our estimation of those data, and hence our ability to predict. Because a full uncertainty-quantification study is in itself a complex process, it is important to ensure that those errors whose size can be controlled—either by experimental technique or by modeling/simulation choices—are small enough to ensure that predictions of the phenomena of interest can be made with sufficient precision for the task at hand. This means that simpler techniques are often appropriate at the start of a study to ensure that we are operating with the required level of precision.

The discrepancy between simulation results and experimental data is illustrated in Figure 4, which shows the way in which this discrepancy can be related to measurement errors and solution errors. Note that the experimental conditions are also subject to uncertainties. This means that the observed value may be associated with a slightly different condition than the one for which the experiment was designed, as shown in Figure 4.

The three steps below could serve as an initial, deterministic assessment of the discrepancy between simulation and experiment.

Step 1. Compare Simulated and Experimental Results. The size of the measurement error will obviously affect the conclusions drawn from the comparison. Those conclusions can also be affected by the degree of knowledge of the actual as opposed to the designed experimental conditions. For example, the as-built composition of the physical parts of the system

under investigation may differ slightly from the original design. The effects of both of these errors are typically reported together, but they are explicitly separated here because error in the experimental conditions affects the simulated result, as well as the measured result, as can be seen in Figure 4.

Step 2. Evaluate Solution Errors. If the error is a simple matter of numerical accuracy—for example, spatial or temporal resolution—then the error is a fixed, determinable number *in principle*. In other cases—for example, subgrid stochastic processes—the error may be knowable in only a statistical sense.

Step 3. Determine Impact on Predictability. If the discrepancy is large compared with the solution error and experimental uncertainty, then the model must be improved. If not, the model may be correct, but in either case, the data can be used to define a range of modeling parameters that is consistent with the observations. If that range leads to an uncertainty in prediction that is too large for the decision being taken, the experimental errors or solution errors must be reduced.

A significant discrepancy in step 1 indicates the presence of errors in the simulation and/or experiment, and steps 2 and 3 are necessary, but not sufficient, to pinpoint the source(s) of error. However, these simple steps do not capture the true complexity of analyzing input or modeling errors. In practice, the system must be subdivided into pieces for which the errors can be isolated (see below) and independently determined. The different errors must then be carefully recombined to determine the uncertainties in integral quantities, such as the yield of a nuclear weapon or the production of an oil well, that are measured in full system tests. A potential drawback of this paradigm is that experiments on subsystems may not be able to probe the entire parameter space encountered in full system operation. Nevertheless, because the need to predict integral quantities motivates the development and use of simulation, a crucial test of the "correctness" of a simulation is that it consistently and accurately matches all available data.

Statistical Prediction

A major challenge of statistical prediction is assessing the uncertainty in a predicted result. Given a simulation model, this problem reduces to the propagation of errors from the simulation input to the simulated result. One major problem in examining the impact of uncertainties in input data on simulation results is the "curse of dimensionality." If the problem is described by a large number of input parameters and the response surface is anything other than a smooth quasilinear function of the input variables, computing the shape of the response surface can be intractable even with large parallel machines. For example, if we have identified 8 critical parameters in a specific problem and can afford to run 1 million simulations, we can resolve the response surface to an accuracy of fewer than 7 equally spaced points per axis.

Various methods exist to assess the most important input parameters. Sensitivities to partial derivatives can be computed either numerically or through adjoint methods. Adjoint methods allow computation of sensitivities in a reasonable time and are widely used.

Experimental design techniques can be used to improve efficiency. Here, the response surface is assumed to be a simple low-order polynomial in the input variables, and then statistical techniques are used to extract the maximum amount of information for a given number of runs. Principal component analysis can also be used to find combinations of parameters that capture most of the variability.

The principle that underlies many of these techniques is that, for a complex engineering system to be reliable, it should not depend sensitively on the values of, for example, 10^4 or more parameters. This is as true for a weapon system that is required to operate reliably as it is for an oil field that is developed with billions of dollars of investment funds.

Statistical Inference—The Bayesian Framework. The Bayesian framework for statistical inference provides a systematic procedure for updating current knowledge of a system on the basis of new information. In engineering and natural science applications, we represent the system by a simulation model m, which is intended to be a complete specification of all information needed to solve a given problem. Thus m includes the governing evolution equations (typically, partial differential equations) for the physical model, initial and boundary conditions, and various model parameters, but it would not generally include the parameters used to specify the numerical solution procedure itself. Any or all of the information in m may be uncertain to some degree. To represent the uncertainty that may be present in the initial specification of the system, we introduce an ensemble of models $\mathcal{M}$, with $m \in \mathcal{M}$, and define a probability distribution on $\mathcal{M}$. This is called the prior distribution and is denoted by $p(m)$.

If additional information about the system is supplied by an observation $\mathcal{O}$, one can determine an updated estimate of the probability for m, called the posterior distribution and denoted by $p(m|\mathcal{O})$, by using Bayes' formula

$$p(m|\mathcal{O}) = \frac{p(\mathcal{O}|m)\,p(m)}{\int_{\mathcal{M}} p(\mathcal{O}|m)\,p(m)\,dm} \quad . \qquad (1)$$

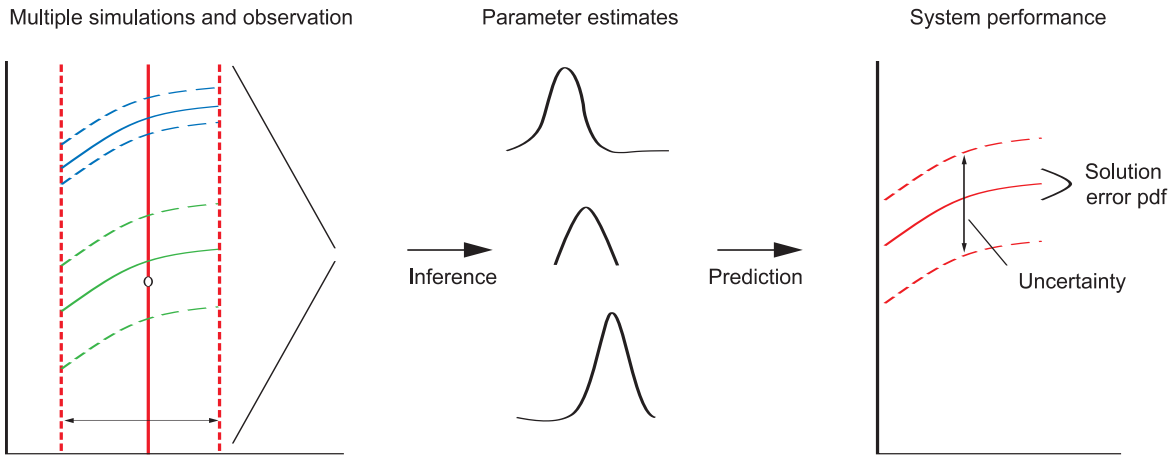


Figure 5. Bayesian Framework for Predicting System Performance with Relevant Uncertainties
Multiple simulations are performed using the full physical range of parameters. The discrepancies between the observation and the simulated values are used in a statistical inference procedure to update estimates of modeling and input uncertainties. The update involves computing the likelihood of the model parameters by using Bayes' theorem. The likelihood is computed from a probability model for the discrepancy, taking into account the measurement errors (shown schematically by the green dotted lines) and the solution errors (blue dotted lines). The updated parameter values are then used to predict system performance, and a decision is taken on whether the accuracy of the predictions is adequate.

It is important to realize that the Bayesian procedure does not determine the choice of $p(m)$. Thus, in using Bayesian analysis, one must supply the prior from an independent data source or a more fundamental theory, or otherwise, one must use a noninformative "flat" prior.

The factor $p(\mathcal{O}|m)$ in Equation (1) is called the likelihood. The likelihood is the (unnormalized) conditional probability for the observation $\mathcal{O}$, given the model m. In the cases of interest here, model predictions are determined by solutions $s(m)$ of the governing equations. The simulated observables are functionals $\mathcal{O}(s(m))$ of $s(m)$. If both the experimentally measured observables $\mathcal{O}$ and the solution $s(m)$, hence $\mathcal{O}(s(m))$, are exact, the likelihood $p(\mathcal{O}|m)$ is a delta function concentrated on the hypersurface in $\mathcal{M}$ defined by the equation

$$\mathcal{O} = \mathcal{O}\big(s(m)\big) \quad . \tag{2}$$

Real-world observations and simulations contain errors, of course, so that a discrepancy will invariably be observed between $\mathcal{O}$ and $\mathcal{O}(s(m))$. Because the likelihood is evaluated subject to the hypothesis that the model $m \in \mathcal{M}$ is correct, any such discrepancy can be attributed to errors either in the solution or in the measurements. The likelihood is defined by assigning probabilities to solution and/or measurement errors of different sizes. The required probability models for both types of errors must be supplied by an independent analysis.

This discussion shows that the role of the likelihood in simulation-based prediction is to assign a weight to a model m based on a probabilistic measure of the quality of the fit of the model predictions to data. Probability models for solution and measurement errors play a similar role in determining the likelihood.

This point is so fundamental and sufficiently removed from common approaches to error analysis that we repeat it for emphasis: *Numerical and observation errors are the leading terms in the determination of the Bayesian likelihood*. They supply critical information needed for uncertainty quantification.

Alternative approaches to inference include the use of interval analysis, possibility theory, fuzzy sets, theories of evidence, and others. We do not survey these alternatives here, but simply mention that they are based on different assumptions about what is known and what can be concluded. For example, interval analysis assumes that unknown

parameters vary within an interval (known exactly), but that the distribution of possible values of the parameter within the interval is not known even in a probabilistic sense. This method yields error bars but not confidence intervals.

An illustration of the Bayesian framework we follow to compute the impact of solution error and experimental uncertainty is shown in Figure 5. Multiple simulations are performed with the full physical range of parameters. The discrepancies (between simulation and observation) are used in a statistical inference procedure to update estimates of modeling and input uncertainties. These updated values are then used to predict system performance, and a decision is taken on whether the accuracy of the predictions is adequate.

Combining Information from Diverse Sources

Bayesian inference can be extended to include multiple sources of information about the details of a physical process that is being simulated (Gaver 1992). This information may come from "off-line" experiments on separate components of the simulation model m, expert judgment, measurements of the actual physical process being simulated, and measurements of a physical process that is related, but not identical, to the process being simulated. Such information can be incorporated into the inference process by using Bayesian hierarchical models, which can account for the nature and strength of these various sources of information. This capability is very important since data directly bearing on the process being modeled is often in short supply and expensive to acquire. Therefore, it is essential to make full use of all possible sources of information—even those that provide only indirect information.

In principle, an analysis can utilize any experimental data that can be compared with some part of the output of a simulation. To understand this point, let us make the simple and often useful assumption that the family of possible models $\mathcal{M}$ can be indexed by a set of parameters . In this case, the somewhat abstract specification of the prior as a probability distribution $p(m)$ on models can be thought of simply as a probability distribution $p(\theta)$ on the parameters θ. Depending on the application, θ may include parameters that describe the physical properties of a system, such as its equation of state, or that specify the initial and boundary conditions for the system, to mention just a few examples. In any of these cases, uncertainty in θ affects prediction uncertainty. Typically, different data sources will give information about different parameters.

Multiple sources of experimental data can be included in a Bayesian analysis by generalizing the likelihood term. If, for example, the experimental observations $\mathcal{O}$ decompose into three components ($\mathcal{O}_1$, $\mathcal{O}_2$, $\mathcal{O}_3$), the likelihood can be written as

$$p\left(\mathcal{O}\middle|m(\theta)\right) = p\left(\mathcal{O}_1\middle|m_1(\theta)\right) \times p\left(\mathcal{O}_2\middle|m_2(\theta)\right) \times p\left(\mathcal{O}_3\middle|m_3(\theta)\right)$$

if we assume that each component of the data gives information about an independent parameter θ. The subscripts on the models are there to remind us that, although the same simulation model is used for each of the likelihood components, different subroutines within the simulation code are likely to be used to simulate the different components of the output. This means that each of the likelihood terms will have its own solution error, as well as its own observation error. The relative sizes of these errors greatly affect how these various data sources constrain θ. For example, if it is known that $m_2(\theta)$ does not reliably simulate $\mathcal{O}_2$, then the likelihood should reflect this fact. Note that a danger here is that a misspecification of a likelihood term may give some data sources undue influence in constraining possible values of one of the parameters θ.

In some cases, one (or more) component (components) of the observed data is (are) not from the actual physical system of interest, but from a related system. In such cases, Bayesian hierarchical models can be used to borrow strength from that data by specifying a prior model that incorporates information from the different systems. See Johnson et al. (2003) for an example.

Finally, expert judgment usually plays a significant role in the formulation and use of models of complex phenomena—whether or not the models are probabilistic. Sometimes, expert judgment is exercised in an indirect way, through selection of a likelihood model or through the choice of the data sources to be included in an analysis. Expert judgment is also used to help with the choice of the probability distribution for $p(\theta)$, or to constrain the range of possible outcomes in an experiment, and such information is often invoked in applications for which experimental or observational data are scarce or nonexistent. However, the use of expert judgment is fraught with its own set of difficulties. For example, the choice of a prior can leave a strong "imprint" on results inferred from subsequent experiments. See *Heuristics and Biases* (2002) for enlightening discussions of this topic.

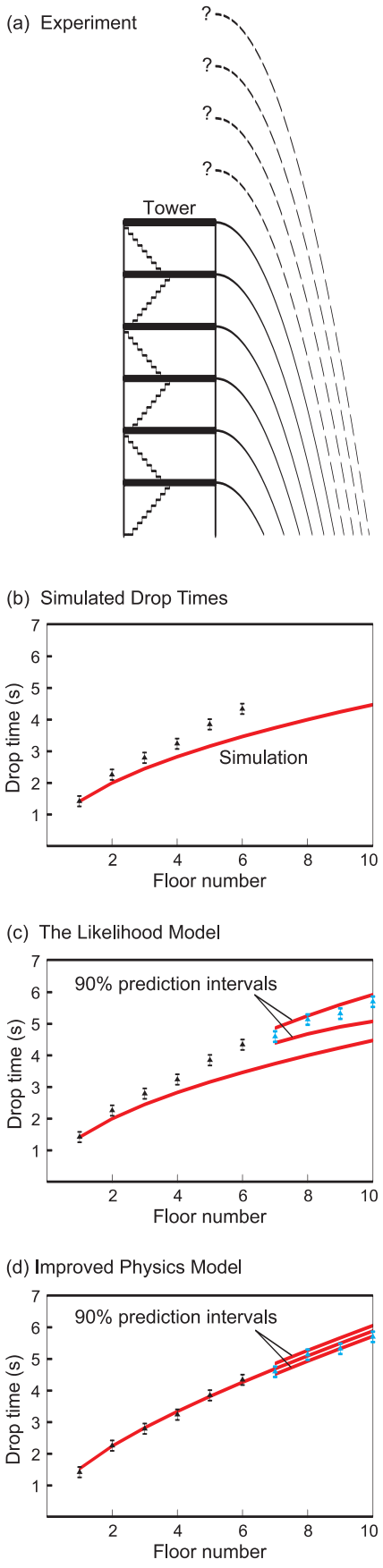


Figure 6. Dropping an Object from a Tower

(a) The time it takes an object to drop from each of 6 floors of a tower is recorded. There is an uncertainty in the measured drop times of about ±0.2 s. Predictions for times are desired for drops from floors 7 through 10, but they do not yet exist.

(b) A mathematical model is developed to predict the drop times as a function of drop height. The simulated drop times (red line) are systematically too low when compared with the experimental data (triangles). The error bars around the observed drop times show the observation uncertainty.

(c) This systematic deviation between the mathematical model and the experimental data is accounted for in the likelihood model. A fitted correction term adjusts the model-based predictions to better match the data. The resulting 90% prediction intervals for floors 7 through 10 are shown in this figure. Note that the prediction intervals become wider as the drop level moves away from the floors with experimental data. The cyan triangles corresponding to floors 7 through 10 show experimental observations taken later only for validation of the predictions.

(d) An improved simulation model was constructed that accounts for air resistance. A parameter controlling the strength of the resistance must be estimated from the data, resulting in some prediction uncertainty (90% prediction intervals are shown for floors 7 through 10). The improved model captures more of the physics, giving reduced prediction uncertainty.

Building Error Models—Examples

Dropping Objects from a Tower. Some of the basic ideas used in building error models are illustrated in Figure 6. In this example, experimental observations are combined with a simple physics model to predict how long it takes an object to fall to the ground when it is dropped from a tower. The experimental data are drop times recorded when the object is dropped from each of six floors of the tower. The actual drop time is measured with an observation error, which we assume for illustrative purposes to be Gaussian (normal), with mean 0 and a standard deviation of 0.2 second. The physics model is based solely on the acceleration due to gravity. We observe that the predicted drop times are too short and that this discrepancy apparently grows with the height from which the object is dropped.

Even though this model shows a substantial error, which is apparent from the discrepancy between the experimental data and the model predictions (Figure 6(b)), it can still be made useful for predicting drop times from heights that are greater than the height of the tower. As a first step, we account for the discrepancy by including an additional unknown correction in the initial specification of the model, namely, in the prior. This term represents the discrepancy as an unknown, smooth function of drop height that is estimated (with uncertainty) in the analysis. The results are applied to give predictions of drop times for heights that would correspond to the seventh through tenth floors of the tower. These predictions have a fair amount of uncertainty because the discrepancy term has to be extrapolated to drop heights that are beyond the range of the experimental data. Note also that the prediction uncertainty increases with drop height (refer to Figure 6(c)).

This strictly phenomenological

modeling of the error leads to results that can be extrapolated over a very limited range only, because predictions of drop times from just a few floors above the sixth have unacceptably large uncertainties. But an improved physics model can greatly extend the range over which useful predictions can be made. Thus, we next carry out an analysis using a model that incorporates a physically motivated term for air resistance. This model requires estimation of a single additional parameter appearing as a coefficient in the air resistance term. But when this parameter is constrained by experimental data, much better agreement with the measured drop times is obtained (see Figure 6(d)). In fact, in this case, the discrepancy is estimated to be nearly zero. The remaining uncertainty in this improved prediction results from uncertainties in the measured data and in the value of the air resistance parameter.

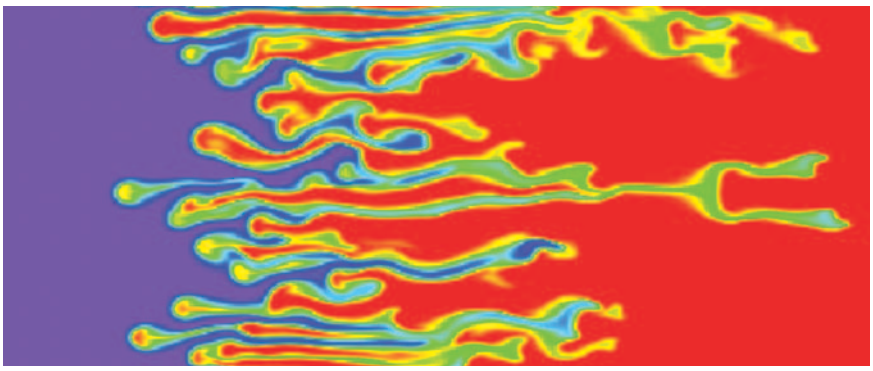


Figure 7. Viscous Fingering in a Realization of Porous Media Flow
Low-viscosity gas (purple) is injected into a reservoir to displace higher-viscosity oil (red). The displacement is unstable and the gas fingers into the oil, reducing recovery efficiency.

Using an Error Model to Improve Predictions of Oil Production. In most oil reservoirs, the oil is recovered by injecting a fluid to displace the oil toward the production wells. The efficiency of the oil recovery depends, in part, on the physical properties of the displacing fluid. The example in this section concerns estimation of the viscosity (typically poorly known) of an injected gas displacing oil in a porous medium. We will show how an error model for such estimates allows improved estimates of the uncertainty in future oil production using this method of recovery.

Because the injected gas has lower viscosity than the oil, the displacement process is unstable and viscous fingers develop (see Figure 7). The phenomenon is similar to the Rayleigh-Taylor instability of a dense fluid on top of a less dense fluid. The fingers have a reasonably predictable average behavior, but there is some randomness in their formation and evolution associated with the lack of knowledge of the initial conditions and with unknown small-scale fluctuations in rock properties.

The oil industry has a simple empirical model that accounts for the effects of fingering. This model, called the Todd and Longstaff model, fits an expansion wave (rarefaction fan) to the average behavior. Although the model is good, it is not perfect, and in particular, when applied to cases with a correlated permeability field, it tends to underestimate the speed with which the leading edge of the gas moves through the medium. If we compare results from the Todd and Longstaff model with observed data in order to estimate physical parameters such as viscosity, we will introduce errors into the parameter estimates because of the errors in the solution method. To compensate for these errors, we create a statistical model for the solution errors.[1]

For this example, we assume that the primary unknown in the Todd and Longstaff model is the ratio of gas viscosity to oil viscosity, which determines the rate at which instabilities grow. This ratio will be determined by comparing simulation and observation (in practice, oil and gas viscosities would be measured, although there would still be uncertainties associated with amounts of gas dissolved in the oil). To construct a solution error model for the average gas concentration in the reservoir, we run a number of fine-grid simulations at discrete values of the viscosity ratio, which we refer to as calibration points. Then, for each value of the viscosity ratio, we compute the difference between the Todd and Longstaff model and the fine-grid simulations as a function of scaled distance along the flow (x) and dimensionless time (t) (time divided by the time for gas to break through in the absence of fingering). The mean error computed in this way for the viscosity ratio 10 is shown in Figure 8 as a function of the similarity variable x/t. We also compute the standard deviation of the error at each time, as well as the correlation between errors at different times. This information is represented as a "covariance matrix."

We will show that the solution error model (the mean error and the covariance matrix), when used in conjunction with predictions of the Todd and Longstaff model at different viscosity ratios, can yield good estimates of the viscosity ratio for a given production data set. Figure 9 shows the

[1] All the results cited in this section are from Alannah O'Sullivan's Ph.D. thesis on error modeling (O'Sullivan 2004). We are grateful to her for permission to use these unpublished results in this article.

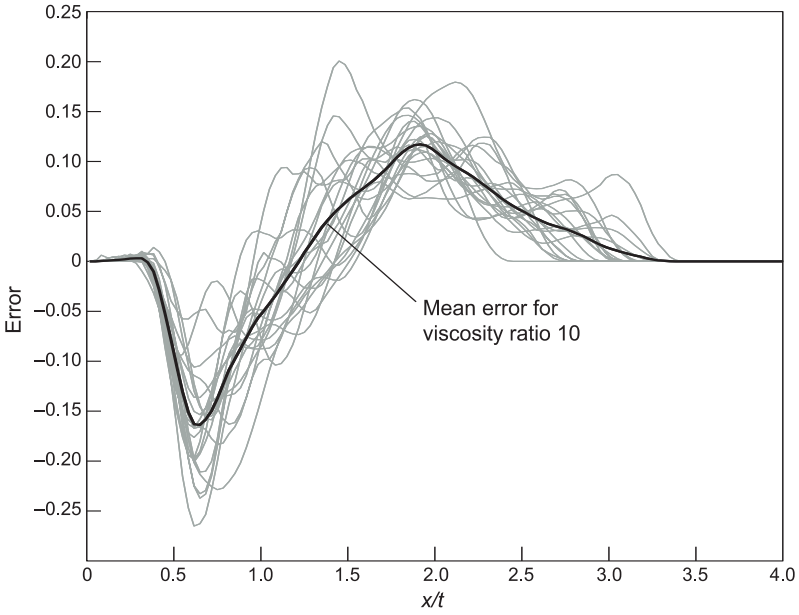


Figure 8. Mean Error and Data to Compute Mean Error

The black curve is the mean error in the gas concentration for viscosity ratio 10. The data to compute the mean error (gray curves) come from the differences between a single coarse-grid or approximate solution (in this case, the Todd and Longstaff model) and multiple fine-grid realizations, all computed at viscosity ratio 10. The variability in the fine-grid realizations reflects random fluctuations in the permeability field, which create different finger locations and growth paths. In this case, the gas concentration averaged across the flow from the fine-grid solution is subtracted from the coarse Todd and Longstaff prediction as a function of *x* (distance along the flow) divided by *t* (time). In the example discussed in the text, we compute the mean error and covariance matrix at viscosity ratios 5, 10, and 15, and interpolation is used to predict the behavior in between these values.

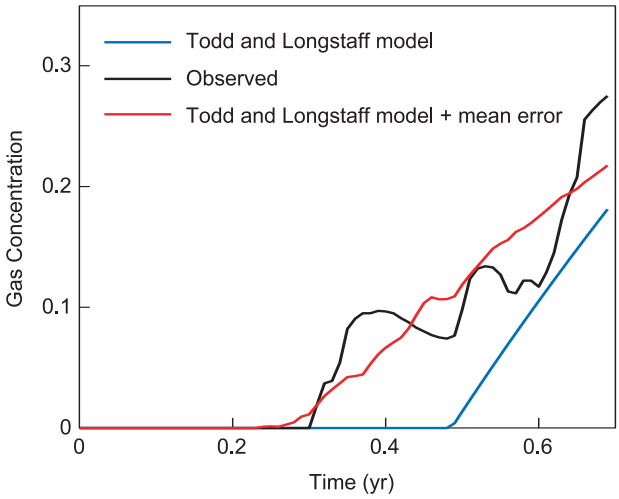


Figure 9. Observed Production Compared with Predictions

The mean error from the error model is added to the coarse-grid result (blue curve) at each time to generate an improved estimate of the gas concentration produced (red curve). The black curve is observed data (actually synthetic data calculated using the fine-grid model with oil-gas viscosity ratio equal to 13).

observed production data (black curve) for which we wish to determine the unknown viscosity ratio. We first run the Todd and Longstaff model at different viscosity ratios from a prior range of 5 to 25 and then correct each prediction by adding to it the mean error at that specific viscosity ratio. The mean error at each viscosity ratio is calculated by interpolating between the mean error at the known calibration points for each value of the similarity variable *x/t*. The blue curve in Figure 9 gives an example of a Todd and Longstaff prediction, and the red curve gives the corrected curve obtained by adding the mean error to the blue curve. To apply the error model, we have converted from the similarity variable *x/t* to time using the known length of the system.

After calculating the corrected predictions for each viscosity ratio, the next step is to compare the corrected prediction (an example is shown in red) for each viscosity ratio with the observed data (shown in black) and compute the misfit M between the simulation and the data. The misfit is given by

$$M = \frac{1}{2}\left(o - s + \bar{e}\right)^T C^{-1}\left(o - s + \bar{e}\right) , \quad (3)$$

where o is the observed value, s is the simulated value, $\bar{e}$ is the mean error, and the covariance matrix is given by $C = \sigma_d^2 I + C_{sem}$. That is, for the covariance matrix, we assume that the data errors are Gaussian, independent, and identically distributed and that therefore they have a standard deviation of σ_d^2, and we estimate the solution error model covariance matrix C_{sem} from the fine-scale simulations performed at the calibration points. The red curve in Figure 10 shows the misfits as a function of viscosity ratio computed using the full error model as in Equation (3). The other misfit statis-

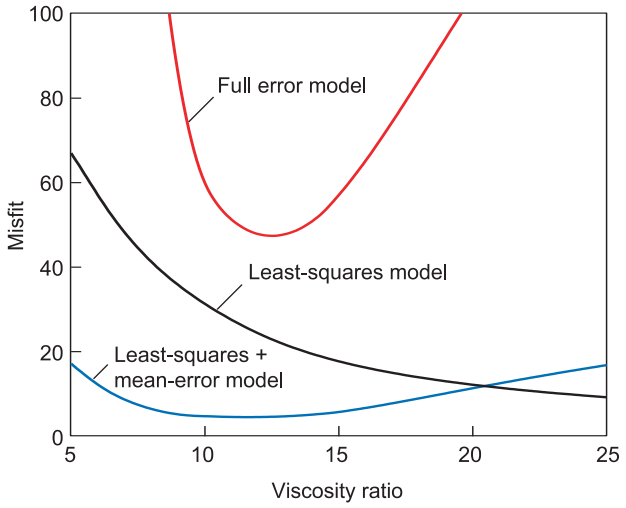


Figure 10. Misfit Statistic vs Viscosity Ratio Calculated in Three Ways
This figure shows a plot of misfit as a function of the viscosity ratio. The misfit is computed using a standard least-squares approach (black curve), least squares with mean error added (blue curve), and the full error model. The misfit measures the quality of the fit to the observed data with low misfits indicating a good fit.

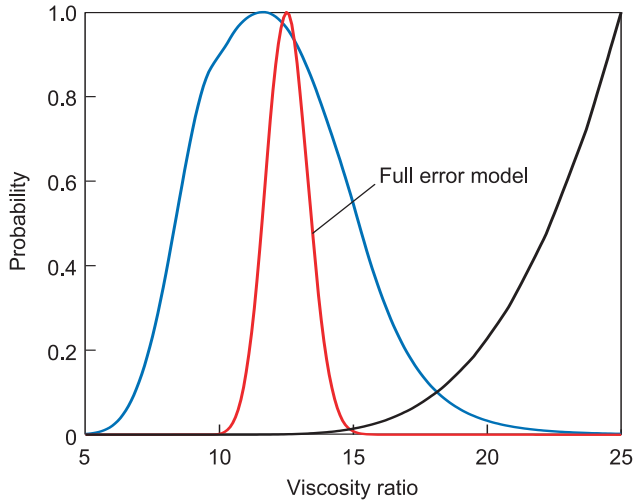


Figure 11. Posterior Probability Distribution Functions for the Viscosity Ratio Calculated in Three Ways
This figure shows the estimated posterior probability (assuming a uniform prior probability in the range 5–25) of the viscosity ratio obtained from three different methods for matching the Todd and Longstaff predictions to observed data. The black curve is obtained from the Todd and Longstaff predictions and a standard least-squares approach. The probability density rises to a maximum at the upper end of the viscosity range specified in the prior model. The blue curve shows the effect of adding the mean error to the predictions. The bias in the coarse model has been removed, but the uncertainty is still large. The red curve shows the estimated viscosity ratio from a full error model treatment—refer to Equation (3)—indicating that it is possible to use a statistical model of solution error to get a good estimate of a physical parameter. The true value of the viscosity ratio in this example was 13.

tics in Figure 10 were computed using

$$M = \sum (o-s)^2 / \sigma_d^2$$

for the least-squares model and

$$M = \sum (o-s+\bar{e})^2 / \sigma_d^2$$

for least-squares plus mean-error model.

The likelihood function L for the viscosity ratio is then given by $L = \exp(-M)$. Notice that the exponential is a signal that the probabilities are sensitive to the method used for computing the misfit. The likelihoods are converted to probability distribution functions by being normalized so that they integrate to 1.

To illustrate the improvement in parameter estimation that results from using an error model, we computed estimates of the probability distribution function for the unknown viscosity ratio using the three different misfit curves in Figure 10, which were calculated with the three different methods: standard least squares, least squares modified by the addition of a mean error term, and least squares with the inclusion of the mean error plus the full covariance matrix. The range of possible values for the viscosity ratio and their posterior probabilities are shown in Figure 11.

The true value of the viscosity ratio used to generate the "observed" (synthetic) production data in Figure 9 was 13, and one can see that this value has been accurately identified by the full error model. The standard least-squares method has not identified this value because of the underlying bias in the Todd and Longstaff model.

We sample from the estimated probability distribution for the viscosity ratio to generate a forecast of uncertainty in future production. Figure 12 is a plot of the maximum likelihood prediction from the Todd and Longstaff model, along with the

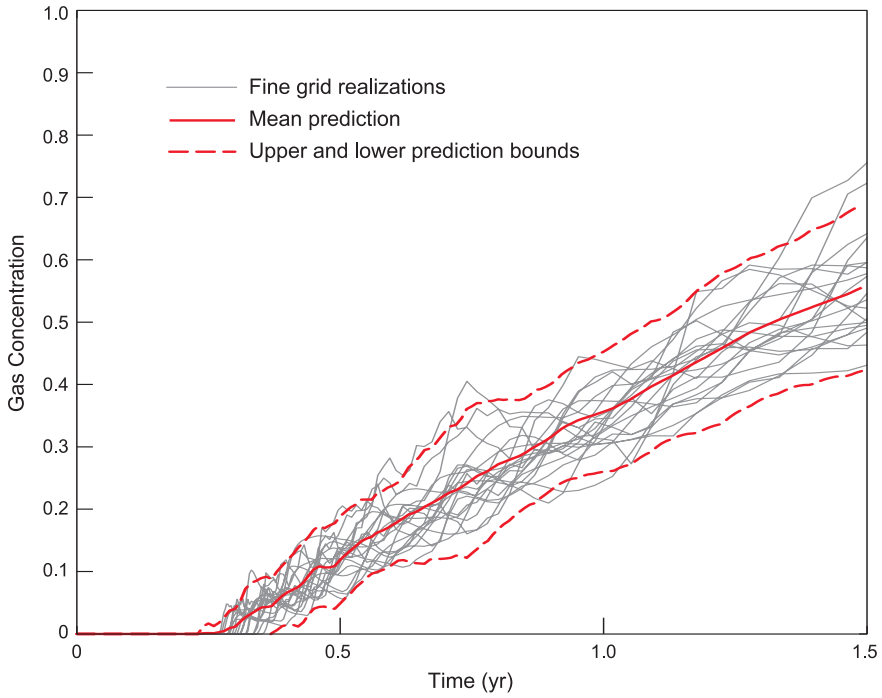


Figure 12. Prediction of Future Oil Production Using Error Model
The solid red line shows the mean (maximum likelihood) prediction from the Todd and Longstaff model and the full error model. The dashed red lines show the 95% confidence interval, and the fine gray curves show the results from 20 fine-grid simulations using the exact viscosity ratio of 13.

95 percent confidence limits obtained by sampling for different values of viscosity. In addition, 20 predictions from fine-grid simulation are shown. They use the exact viscosity ratio 13. The uncertainty in the evolution of the fingers gives rise to the uncertainty in prediction shown by the multiple light-gray curves. It is clear from the figure that use of an error model has allowed us to produce well-calibrated predictions.

Fluid Dynamics—Error Models for Reverberating Shock Waves. Compressible flow exhibits remarkable phenomena, one of the most striking being shock waves, which are propagating disturbances characterized by sudden and often large jumps in the flow variables across the wave front (Courant and Friedrichs 1967). In fact, for inviscid flows, these jumps are represented as mathematical discontinuities. Shock waves play a prominent role in explosions, supersonic aerodynamics, inertial confinement fusion, and numerous other problems. Most problems of practical importance involve two- or three-dimensional (2-D or 3-D) flows, complex wave interactions, and other complications, so that a quantitative description of the flow can be obtained only by solving the fluid-flow equations numerically. The ability to numerically simulate complex flows is a triumph of modern science, but such simulations, like all numerical solutions, are only approximate. The errors in the numerical solution can be significant, especially when the computations use moderate to coarse computational grids as is often necessary for real-world problems. In this section, we sketch an approach to estimating these errors.

Our approach makes heavy use of the fact that shock waves are persistent, highly localized wave disturbances. In this case, "persistent" means that shock waves propagate as locally steady-state wave fronts that can be modified only by interactions with other waves or unsteady flows. Generally, interactions consist of collisions with other shock waves, boundaries, or material interfaces. The phrase "highly localized" refers to shock fronts being sharp and their interactions occurring in limited regions of space and time and possibly being characterized by the refraction of shock fronts into multiple wave fronts of different families. These properties are illustrated in Figure 13, which shows a sequence of wave interactions being initiated when a shock incident from the left collides with a contact located a short distance from a reflecting wall at the right boundary in the figure. Each collision event produces three outgoing waves: a transmitted shock, a contact discontinuity, and a reflected shock or rarefaction wave. The buildup of a complex space-time pattern due to the multiple wave interactions is evident.

Generally, solution errors are determined by comparison to a fiducial solution, that is, a solution that is accepted, not necessarily as perfect, but as "correct enough" for the problem being studied. But producing a fiducial solution may not be easy. In principle, one might obtain one using a very highly resolved computation. However, in real-world problems, this is generally not feasible. If it were, one would just do it and forget about solution errors. So, what do we do when we cannot compute a fiducial solution?

The development of models for error generation and propagation offers an approach for dealing with flows that are too complex for direct computation of a fiducial solution. For compressible flows, the key point is that the equations are hyperbolic, which implies that errors are largely

advected through smooth-flow regions and significant errors are only created when wave fronts collide. The flow shown in Figure 13 consists of a sequence of binary wave interactions, each of which is simple enough to be computed on an ultrafine grid. The basic idea is to determine the solution errors for an elementary wave interaction and to construct "composition laws" that give the error at any given point in terms of the error generated at each of the elementary wave interactions in its domain of influence.

A number of points need to be made here. First, there are a limited number of types of elementary wave interactions. One-dimensional (1-D) interactions occur as refractions of pairs of parallel wave fronts, 2-D interactions are refractions of two oblique wave fronts, and 3-D interactions correspond to triple points produced by three interacting waves. It is important to note that, in each spatial dimension, the elementary wave interactions occur at isolated points. Most of the types of wave interactions that can occur in 1-D flow appear in Figure 13. The coherent traveling wave interactions that occur in 2-D flows have been characterized (Glimm et al. 1985). However, substantial limitations are left on the refinement and thoroughness with which 3-D elementary wave interactions can be studied.

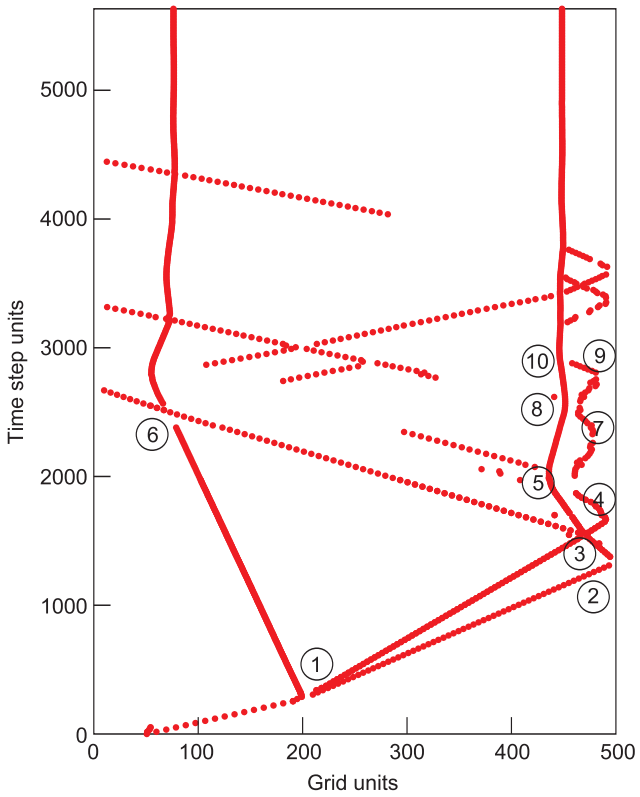


Figure 13. The Space-Time Interaction History of a Shock-Tube Refraction

This figure shows the interaction history as reconstructed from the simulated solution data from a shock-tube refraction problem. A planar shock is incident from the left on a contact discontinuity located near the middle of the test section of the shock tube. A reflecting wall is located on the right side of the tube. Event 1 corresponds to the initial refraction of the shock wave into reflected and transmitted waves, event 2 occurs when the transmitted shock produced by interaction 1 reflects at the right wall, and the events numbered 3–10 correspond to subsequent wave interactions between the various waves produced by earlier refractions or reflections. Our error model is applied at each interaction location to estimate the additional solution error produced by the interaction.

(This figure was supplied courtesy of Dr. Yan Yu, Stony Brook University.)

Event 1 in Figure 13 is a typical example of a 1-D wave interaction. Here, the "incoming waves" consist of an incident shock and a contact discontinuity, and the "outgoing state" is described by a reflected shock, a (moving) contact, and a transmitted shock. The interaction can be described as the solution to a Riemann problem with data given by the states behind the incoming wave fronts. A Riemann problem is defined as the initial value problem for a hyperbolic system of conservation laws with scale-invariant initial data. Riemann problems and their solutions are basic theoretical tools in the study of shock dynamics, the development of shock-capturing schemes to numerically compute flows, and they also play a key role in our study of solution errors. A key point in the use of Riemann problem solutions in our error model is that the solution of a 1-D Riemann problem for hydrodynamics reduces to solving a single, relatively simple algebraic equation. It is thus possible to solve large numbers of Riemann problems for a flow analysis quickly and efficiently. This observation is particularly important because our error model requires the solution of multiple Riemann problems whose data are drawn from statistical ensembles of initial data to represent uncertainties in the incoming waves.

A final point here is that a realistic solution error model must include the study of the size distribution of errors over an ensemble of problems, in which the variability of problem char-

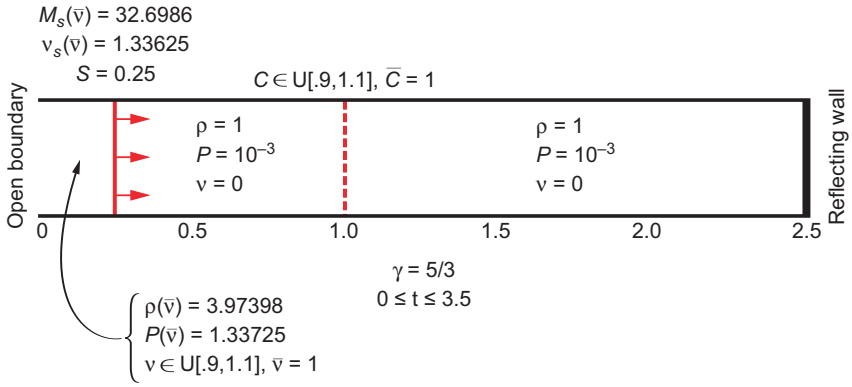


Figure 14. Initial Data for a 1-D Shock-Tube Refraction Problem

This schematic diagram is for the initial data used to conduct an ensemble of simulations of a 1-D shock tube refraction. Each simulation consisted of a shock wave incident from the left on a contact discontinuity between gases at the indicated pressures and densities. Each realization from the ensemble is obtained by selecting a shock strength consistent with a velocity v behind the incident shock taken from a 10% uniform distribution about the mean value $\bar{v}$ = 1, and an initial contact location C chosen from a 10% uniform distribution about the mean position $\bar{C}$ = 1. In the diagram, S is the shock position, M_s is the shock strength, and v_s is the velocity of the shock. The initial state behind the shock is set by using the Rankine-Hugoniot conditions for the realization shock strength and the specified state ahead of the shock.

acteristics is described probabilistically. Of course, one will often want to make as refined an error analysis as possible within a given realization from the ensemble (that is, a deterministic error analysis), but there are powerful reasons for a probabilistic analysis to be needed as well. First, you need probability to describe features of a problem that are too complex for feasible deterministic analysis. Thus, fine details of error generation in complex flows are modeled as random, just as are some details of the operation of measuring instruments. Second, a sensitivity analysis is needed to determine the robustness of the conclusions of a deterministic error analysis to parameter variation. To get an accurate picture, one needs to do sensitivity analysis probabilistically, to answer the question of how likely the parameter variations are that lead to computed changes in the errors. Third, to be a useful tool, the error model must be applicable to a reasonable range of conditions and problems. The only way we are aware of for achieving these goals is to base the error model on a study of an ensemble of problems that reflects the degree of variability one expects to encounter in practice. Of course, the choice of such an ensemble reflects scientific judgment and is an ongoing part of our effort.

Now, let us return to the analysis of solution errors in elementary wave interactions. Our work was motivated by a study of a shock-contact interaction—refer to event 1 in Figure 13. The basic setup is shown in Figure 14, which illustrates a classic shock-tube experiment. An ensemble of problems was generated by sampling from uniform probability distributions (±10 percent about nominal values) for the initial shock strength and the contact position. The solution errors were analyzed by computing the difference between coarse to moderate grid solutions and a very fine grid solution (1000 cells). Error statistics are shown in Figure 15 for a 100-cell grid (moderate grid) solution. Two facts about these solution errors are apparent. First, the solution errors follow the same pattern as the solution (the shock waves) itself; they are concentrated along the wave fronts, where steep gradients in the solution occur. Second, errors are generated at the location of wave interactions. The error generated by the interaction increments the error in the outgoing waves, which is inherited from errors in the incoming waves.

Comparable studies have been carried out for each of the types of wave interaction shown in Figure 13, as well as corresponding wave interactions that occur in spherical implosions or explosions (Dutta et al. 2004). An analysis of statistical ensembles of such interactions has led us to suggest the following scheme for estimating the solution errors. The key steps are (a) identification of the main wave fronts in a flow, (b) determination of the times and locations of wave interactions, and (c) approximate evaluation of the errors generated during the interactions. Wave fronts are most simply identified as regions of large flow gradients, and the distribution of the wave positions and velocities are found by solving Riemann problems whose data are taken from ensembles of state information near the detected wave fronts. The error generated during an interaction is fit by a linear expression in the uncertainties of the incoming wave's strength. The coefficients are computed using a least-squares fit to the distribution of outgoing wave strengths. This fitting procedure can be thought of as defining an input/output relation between errors in incoming and outgoing waves.

A linear relation of this kind, which amounts to treating the errors perturbatively, holds even for strong, and hence nonlinear, wave interactions. But there are limitations. Linearity works if the errors in the incoming waves are not too large, but it may break down for larger errors. In the latter case, higher order (for example, bilinear or rational) terms in the expansion may be needed. See

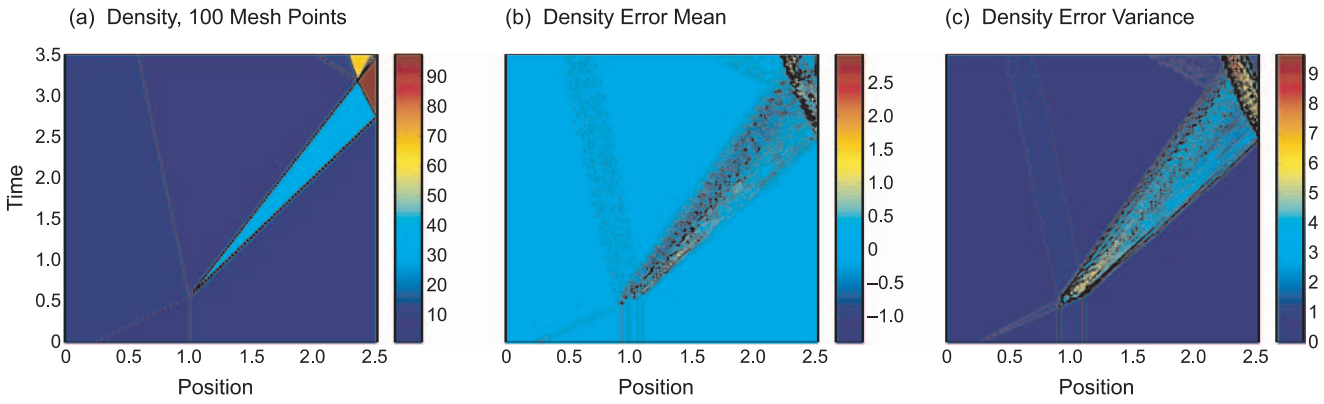


Figure 15. Space-Time Error Statistics for Shock-Tube Refraction Problems

Panel (a) shows the space-time 100-mesh-point density field for a single realization from the flow ensemble. The space-time error field for each realization is computed from the difference between a 100-mesh-zone calculation and a fiducial solution computed using 1000 mesh zones. Panels (b) and (c) show the mean and variance, respectively, over the ensemble as a function of space and time. Note that most errors are generated at the wave interactions and then move with the wave fronts.

Glimm et al. (2003) for details.

We can now explain how the composition law for solution errors actually works. The basic idea is that errors are introduced into the problem by two mechanisms: input errors that are present in waves that initiate the sequence of wave interactions—see the incoming waves for event 1 in Figure 13—and errors generated at each interaction site. However they are introduced, errors advect with the flow and are transferred at each interaction site by computable relations.

Generally, waves arrive at a given space-time point by more than one path. Referring again to Figure 13, suppose you want to find the errors in the output waves for event 3, where the shock reflected off the wall reshocks the contact. On path A, the error propagates directly from the output of interaction 1 along the path of the contact, where it forms part of the input error for event 3. On path B, the output error in the transmitted shock from event 1 follows the transmitted shock to the wall, where it is reflected and then re-crosses the contact. In this way, the error coming into event 3 is given as a sum of terms, with each term labeled by a sequence of wave interactions and of waves connecting these interactions. Moreover, each term can be computed on the basis of elementary wave interactions and does not require the full solution of the numerical problem. The final step in the process is to compute the errors in the output waves at event 3, by using the input/output relations developed for this type of wave interaction.

This procedure represents a substantial reduction in the difficulty of the error analysis problem, and we must ask whether it actually works. Full validation requires use in practice, of course. As a first validation step, we compute the error in two ways. First, we compute the error directly by comparing very fine and coarse-grid simulations for an entire wave pattern. Results are shown in Figure 15. Second, we compute the error using the composition law procedure shown in Figure 13. Comparing the errors computed in these two ways provides the basis for validation.

In Glimm et al. (2003) and Dutta et al. (2004), we carried out such validation studies for planar and spherical shock-wave reverberation problems. As an example, for events 1 to 3 in the planar problem in Figure 13, we considered three grid levels, the finest (5000 cells) defining the fiducial solution, and the other two representing "resolved" (500 cells) and "under-resolved" (100 cells) solutions for this problem. We introduced a 10 percent initial input uncertainty to define the ensemble of problems to be examined. The results can be summarized briefly as follows. For the resolved case, the composition law gave accurate results for the errors (as determined by direct fine-to-coarse grid comparisons) in all cases: wave strength, wave width, and wave position errors. This was not the case for the under-resolved simulation. Although the composition law gave good results for wave strength and wave width errors, it gave poor results for wave position errors. The nature of these results can be understood in terms of a breakdown in some of the modeling assumptions used in the analysis.

An interesting point of contrast emerged between the planar and spherical cases. For the planar case, the dominant source of error was from initial uncertainty, while for the spherical symmetry case, the dominant source of error arose in the simulation itself, and especially from shock

reflections off the center of symmetry.

We come now to the "so what?" question for error models. What are they good for? Our analysis shows that, with an error model, one can determine the relative importance of input and solution errors (thereby allocating resources effectively to their reduction), as well as the precise source of the solution error (for the same purpose), and, finally, one can assess the error in a far more efficient manner than by direct comparison with a highly refined computation of the full problem.

A significant limitation in our results to date is that they pertain mostly to 1-D flows, namely, to flows having planar, cylindrical, or spherical symmetry. Two-dimensional problems are currently under study, while full 3-D problems are to be solved in the future. Furthermore, errors in some important fluid flows lie outside the framework we have developed, and their analysis will require new ideas. One such problem—fluid mixing—was discussed in the previous subsection.

Conclusions

This paper started from the premise that predictive simulations of complex phenomena will increasingly be called upon to support high-consequence decisions, for which confidence in the answer is essential. Many factors limit the accuracy of simulations of complex phenomena, one of the most important being the sparsity of relevant, high-quality data. Other factors include incomplete or insufficiently accurate models, inaccurate solutions of the governing equations in the model, and the need to integrate the diverse and numerous components of a complex simulation into a coherent whole. Error analysis by itself does not circumvent these limitations. It is a way to estimate the level of confidence that can be placed in a simulation-based prediction on the basis of a careful analysis of the source and size of errors affecting this prediction. Thus, the metric of success of an error analysis is the confidence it gives that the errors are of a specific size—not necessarily that they are small (they might not be).

We have reviewed some of the ideas and methods that are used in the study of simulation errors and have presented three examples illustrating how these methods can be used. The examples show how an improved physics model can dramatically reduce the size of errors, how an improved error model can reduce uncertainty in prediction of future oil production, and how an error model for a complex shock-wave problem can be built up from an error analysis of its components.

Similar to models of natural phenomena, error models will never be perfect. Estimates of errors and uncertainties are always provisional because the data supporting these estimates are derived from a limited range of experience. Certainty is not in the picture. Nevertheless, confidence in predictions can be derived from the scope and power of the theory and solution methods that are being used. Scope refers to the number and variety of cases in which a theory has been tested. Scope is important in building confidence that one has identified the factors limiting the applicability of the theory. Power is judged by comparing what is put into the simulation with what comes out.

Error models contribute to confidence by clarifying what we do and do not understand. They also guide efforts to improve our understanding by focusing on factors that are the leading sources of error. Thus, in predictions of complex phenomena, an error analysis will form an indispensable part of the answer. ■

Further Reading

Courant, R., and K. O. Friedrichs. 1967. *Supersonic Flow and Shock Waves*. New York: Springer-Verlag.

Dutta, S., E. George, J. Glimm, J. W. Grove, H. Jin, T. Lee, et al. 2004. Shock Wave Interactions in Spherical and Perturbed Spherical Geometries, Los Alamos National Laboratory document LA-UR-04-2989. (Submitted to *Nonlinear Anal*.).

Gaver, D. P. 1992. *Combining Information: Statistical Issues and Opportunities for Research* Report by Panel on Statistical Issues and Opportunities for Research in the Combination of Information, Committee on Applied and Theoretical Statistics, National Research Council. Washington, DC: National Academic Press.

Gilovich, T., D. Griffin, and D. Kahneman, eds. 2002. *Heuristics and Biases: The Psychology of Intuitive Judgment.* Cambridge, UK: Cambridge University Press.

*G*limm, J., J. W. Grove, Y. Kang, T. W. Lee, X. Li, D. H. Sharp, et al. *2003.* Statistical Riemann Problems and a Composition Law for Errors in Numerical Solutions of Shock Physics Problems, Los Alamos National Laboratory document LA-UR-03-2921. *SIAM J. Sci. Comput.* (in press).

*G*limm, J., C. Klingenberg, O. McBryan, B. Plohr, S. Yaniv, and D. H. Sharp. 1985. Front Tracking and Two-Dimensional Riemann Problems. *Adv. Appl. Math.* **6**: 259.

Johnson, V. E., T. L. Graves, M. S. Hamada, and C. S. Reese. 2003. A Hierarchical Model for Estimating the Reliability of Complex Systems. In *Bayesian Statistics 7*: *Proceedings of the Seventh Valencia International Meeting*. p. 199. UK: Oxford University Press.

Kao, J., D. Flicker, R. Henninger, S. Frey, M. Ghil, and K. Ide. 2004. Data Assimilation with an extended Kalman Filter for Impact-Produced Shock-Wave Dynamics. *J. Comp. Phys.* **196** (2): 705.

O'Nions, K., R. Pitman, and C. Marsh. 2002. The Science of Nuclear Warheads. *Nature* **415**: 853.

O'Sullivan, A. E. Modelling Simulation Error for Improved Reservoir Prediction Ph.D. thesis, Heriot-Watt University, Edinburgh.

Palmer, T. N. 2000. Predicting Uncertainty in Forecasts of Weather and Climate. *Rep. Prog. Phys.* **63**: 71.

Sharp, D. H., and M. Wood-Schulz. 2003. QMU and Nuclear Weapons Certification. *Los Alamos Science* **28**: 47.

For further information, contact David H. Sharp (505) 667-5266 (dcso@lanl.gov).