ON THE QUANTIFICATION OF MODELING UNCERTAINTIES

by

N. Siu

Pickard, Lowe and Garrick, Inc.
2260 University Drive
Newport Beach, CA 92660
USA

G. Apostolakis

University of California
School of Engineering and
Applied Science
5532 Boelter Hall
Los Angeles, CA 90024
USA

Paper M2 1/5

8th International Conference on
Structural Mechanics in Reactor Technology
Brussels, Belgium
August 19-23, 1985

INTRODUCTION

Probabilistic risk assessment (PRA) studies, as currently performed, employ evidence which consists of statistical data and expert judgments. As a special case of the latter, it may be in the form of predictions given by a physical model, the model being a formalized encoding of a system, e.g., the behavior of fire in a room. Because judgment is involved both in the construction and in the application of the model, as it is involved in the entire risk assessment procedure, of which the model is just a part, the uncertainties in the model's predictions must be quantified.

The existence of uncertainties in the predictions of physical models has, of course, long been acknowledged. The comparison, or "benchmarking", of an analytical or computer-based model with experimental data is generally viewed as a highly desirable step in model development and documentation. Such exercises may lead to rough estimates of the model's accuracy, perhaps in terms of maximum percent deviation from the experimentally observed values. What has been lacking, until fairly recently, however, is a formal quantification of the modeling uncertainty, an expression of the model's output not in terms of a single point (or set of points) but rather in terms of a probability distribution.

Two characteristics relevant to the analysis of uncertainties in a physical model's predictions are: a) the degree of inherent randomness of the process being modeled, and b) the existence of uncertainties in the values of the model's input parameters. If the process is stochastic, i.e., if it varies randomly in time, the analysis must distinguish between the uncertainties arising from this inherent randomness and the uncertainties arising due to imperfect knowledge. Since the stochastic distribution governing the random variability can usually be parameterized, the analysis can focus on developing probability distributions for the stochastic distribution's parameters.

The propagation of input parameter uncertainties through complex models is fairly straightforward in concept and is widely used to estimate uncertainties in model output. The treatment of modeling uncertainties in the presence of input parameter uncertainties is, on the other hand, not as easily dealt with. By modeling uncertainties, we mean those uncertainties stemming from imperfections in the model's structure, due to its approximate nature. Consider, for example, the computer code COMPBRN [1], which can be used to predict the growth rate of a fire. The code requires as input numerous empirical [illegible] the fuel bed supporting the fire. If the values of these parameters are not precisely known, the growth rate predictions obtained will reflect this uncertainty; the formal assessment of the code's accuracy under these circumstances is not obvious, even when experimental results are available [2].

To illustrate the analytical approach that is applied to COMPBRN, we define τ_G to be the mean fire growth time. Then, Ref. 1 uses

$$\tau_G = E_\tau \ \tau_{G,DRM} \tag{1}$$

where E_τ is the error factor representing the analyst's confidence in the prediction of the code (called the deterministic reference model, or DRM) and $\tau_{G,DRM}$ is the mean growth time predicted by COMPBRN. Both E_τ and $\tau_{G,\ DRM}$ are uncertain variables and thus are characterized by probability distributions. The distribution of $\tau_{G,\ DRM}$ is determined by propagating input parameter uncertainties through COMPBRN (actually, through a response surface representation of COMPBRN). The distribution of E_τ is assessed subjectively on the basis of comparisons of the code's predictions with: a) experimental data for a vertical cable tray fire scenario, and b) an expert's estimate for a horizontal tray scenario. In both of these comparisons, parameter uncertainties are propagated through the code and have an impact on the quantification of E_τ. In applications of COMPBRN to scenarios for which growth data do not exist, the distribution of $\tau_{G,\ DRM}$ is reassessed (because the input parameters are different) but the distribution for E_τ is not modified unless the new scenarios are greatly different.

In Reference 3, a slightly different approach is used to estimate the impact of modeling uncertainties on the output of the CUT code, which models the transient behavior of a pressurized water reactor during a small LOCA. In that analysis, one intermediate result, the mass flow rate through a relief valve, can be computed using a number of different sub-models. Reference 3 uses one sub-model to estimate the flow rate, and accounts for the uncertainty arising from model-to-model variability by multiplying that flow rate with an uncertainty factor. The distribution for the uncertainty factor is assessed subjectively, using the different predictions of the various models to indicate the possible range of variation. The resulting modeling uncertainty is propagated through the CUT code in the same manner that parameter uncertainties are treated, i.e., using Monte Carlo sampling off of a response surface. Thus, the primary difference between this approach and what is done for COMPBRN is that in the latter the modeling uncertainty analysis is performed at the level of COMPBRN's final results, rather than at an intermediate level, as is done in Ref. 3.

A CONCEPTUAL FRAMEWORK

In the preceding examples, evidence in the form of sensitivity analyses results, alternate model predictions, experimental data, and expert judgment are loosely employed to quantify the modeling uncertainty factor distributions. A natural tool to formalize the use of this evidence is Bayes' Theorem [4]. Before we outline a conceptual approach for using Bayes' Theorem in this context, however, we first discuss some of the issues which may affect the analysis.

One of the key issues involves the form of the modeling uncertainty approach to be problems where modeling uncertainty can be expressed in terms of a continous parameter. While additional modeling uncertainty parameters could be introduced, this would complicate the analysis a great deal.

In general, the formal analysis of modeling uncertainties appears to be a complex problem, as will be illustrated by an example in the following section. A practical issue, therefore, is the allocation of limited resources to the detailed quantification of modeling uncertainties versus allocation towards reducing these uncertainties.

Another important issue is the availability and quantity of data to benchmark the model. Without relevant information, a detailed uncertainty analysis could be wasteful. On the other hand, such an analysis could be useful in guiding the conduct of experiments, since the analysis would indicate exactly what data are needed for unambiguous benchmarking. The data would include, for example, measurements of the physical parameters required as input for the model; any differences between model predictions and experimental outcomes could then be directly attributed to modeling uncertainties. It is because such complete data are generally not available that the treatment of modeling uncertainties is so complicated.

To outline a conceptual approach for handling modeling uncertainties in the presence of parameter uncertainties, consider the following situation arising in the course of a fire risk analysis. Suppose we wish to use COMPBRN to develop the distribution $p(\tau_G)$ for the mean growth time τ_G for a very specific fire scenario, e.g., the spread of fire from one horizontal cable tray to one immediately above. Further suppose that the evidence for this distribution is:

1. A distribution $p(\tau_{G,\ DRM})$ for the predicted mean growth time specific to the scenario being modeled. This distribution is obtained by propagating the input parameter uncertainties through our COMPBRN model for the two trays.
2. An actual mean growth time, τ_G^*, estimated from growth time data for an experiment similar in configuration to the scenario being modeled.
3. A distribution $p\ (\tau^*_{G,\ DRM})$ for the predicted mean growth time of the experiment. This distribution, similar to $p\ (\tau_{G,\ DRM})$, arises from uncertainties in the COMPBRN input parameters relevant to the experiment.

The framework for our approach is provided by Equation (1). We note that while our scenario and the experiment are similar, they are not identical (else we would probably be using COMPBRN to develop the distribution of τ_G). In general, therefore, the uncertainty factor for the scenario, E_τ, and that for the experiment, E_τ^*, will differ by an unknown amount. To model the scenario-to-scenario variability in E_τ, we assume that E_τ is a random variable governed by a two-parameter stochastic (frequency) distribution $f(E_\tau|\mu,\sigma)$. Our approach, then, is to use the available evidence to assess the joint distribution of μ and σ, and then to use this intermediate result in developing a probability distribution for E. It can be seen that this approach is similar to the first stage of the two-stage Bayesian methodology developed in Reference 5 to treat plant-to-plant variability in equipment failure rates.

If there are no uncertainties in $\tau^*_{G,\ DRM}$ (i.e, if $p(\tau^*_{G,\ DRM})$ is a delta function), the joint posterior distribution for μ and σ follows directly from Bayes Theorem:

$$\pi_1(\mu,\sigma|E_\tau^*) = k^{-1}\ f(E_\tau^*,|\mu,\sigma)\ \pi_0(\mu,\sigma) \qquad (2)$$

where

$$E_\tau^* = \frac{\tau_G^*}{\tau_{G,\ DRM}^*}\ ,\quad k = \iint f(E_\tau^* \mid \mu,\sigma)\ \pi_o(\mu\ \sigma)\ d\mu\ d\sigma$$

To account for the uncertainty in $\tau^*_{G,\ DRM}$, we observe that $\pi_1(\mu,\sigma \mid E_\tau^*)$ is conditione on E_τ^*, and hence, on $\tau^*_{G,\ DRM}$. This conditioning can be removed simply by taking the expectation of $\pi_1(\mu,\sigma \mid E_\tau^*)$:

$$\pi_1(\mu,\sigma \mid p(E_\tau^*)) = \int \pi_1(\mu,\sigma \mid E_\tau^*)\ \ p(E_\tau^*)\ \ dE_\tau^* \tag{3}$$

where $p(E_\tau^*)$ is determined directly using $p(\tau^*_{G,\ DRM})$. This use of uncertain, or "fuzzy" data, is developed more thoroughly (in a different context) in Reference 6.

To complete the analysis, we follow Reference 5 and find the average frequency distribution for E_τ, which we equate with $p(E_\tau)$:

$$p(E_\tau) = \iint f(E_\tau \mid \mu,\sigma)\ \ \pi_1(\mu,\sigma \mid p(E_\tau))\ \ d\mu\ \ d\sigma \tag{4}$$

This analysis for E_τ incorporates both statistical data obtained from an experiment and an uncertain prediction (caused by input parameter uncertainties) from a simulation of that experiment. The extension of this approach to many experiments is straightforward in principle, but may be difficult to actually perform. In fact, difficulties may easily arise even for the simple one experiment case presented. If, for example, certain combinations of input parameter values lead to a COMPBRN prediction that fire damage is impossible, although damage was actually observed in the experiment, the scaling factor approach used loses applicability. Work on the actual implementation of the formal approa and on generalizations of the approach is continuing.

REFERENCES

1. Siu, N.O. and G.E. Apostolakis, "Probabilistic Models for Cable Tray Fires," Reliability Engineering, 3, 213-227(1982).

2. Chung, G., N. Siu, and G. Apostolakis. "Improvements in Compartment Fire Modeling and Simulation of Experiments." Nuclear Technology, to appear in 1985.

3. Chu, T.L. and G.E. Apostolakis, "Assessment of the Uncertainties Associated with the Core Uncovery Time in TMI-Type Accidents," Reliability Engineering, 8, 23-56 (1984).

4. Apostolakis, G., "Bayesian Methods in Risk Assessment," in Advances in Nuclear Science and Technology, 13, J. Lewins and M. Becker (eds.), Plenum Press, 415-465(1981).

5. Kaplan, S., "On a 'Two-Stage' Bayesian Procedure for Determining Failure Rates from Experiential Data," IEEE Transactions on Power Apparatus and Systems, PAS-102, No. 1, 1983.

6. Siu, N. and G. Apostolakis, "Models for the Detection and Suppression of Fires in Nuclear Power Plants and Some Related Statistical Problems," Draft Report UCLA-ENG-8440, 1984.