LECTURE 5.3

UNCERTAINTY ANALYSIS IN BEST ESTIMATE AND COUPLED CALCULATIONS

Rafael Macián-Juan

Institute for Nuclear Engineering
Technische Universität München

INTRODUCTION

Considerable progress has been made in recent years in improving the quality and depth of the safety analysis of nuclear systems with the development of increasingly more sophisticated analytical tools for the simulation of the behavior of nuclear systems. There has been, in particular, a move away from over-conservatism in safety analysis towards the application of so-called Best Estimate methodologies. The result has been, in general, a better definition of the safety limits applied to nuclear power plant (NPP) operation, because Best Estimate calculations employ more accurate physical models, numerical solution procedures, and closer to actual system boundary conditions which, as a result, offer more realistic plant behavior descriptions to decision makers.

The application of best estimate safety analysis has opened new prospects, but more realistic evaluations make it crucial to assess their degree of reliability, especially when, for instance, they have to be compared against safety limits that must not be crossed. For this reason, it is necessary to identify and quantify all the sources of possible uncertainties that could affect the accuracy of the results, and then apply mathematically sound and justifiable methodologies that can propagate them through the nuclear safety analysis procedure to the predictions of the systems variables of interest (the outputs of the Best Estimate codes).

UNCERTAINTY METHODOLOGY AND SOURCES OF UNCERTAINTY

A series of such uncertainty propagation methodologies have been developed and tested in the last 20 years, and most of them have proved their value by meeting some of the requirements that any methodology must fulfill. However, there is not yet a consensus as to which methodology is the best to satisfy all the needs that a complete uncertainty propagation analysis demands. Nevertheless, two main approaches have been adopted by a large number of researchers and users of these methodologies, namely, one based on statistical propagation and processing, and another based on the application, when computationally feasible, of *variational* methods that quantify analytical sensitivities, usually first order, of the physical-numerical model of a nuclear system, and combine them with the uncertainty measures of models and boundary conditions.

Before a Best Estimate code can be confidently applied to safety analysis, a series of important issues should be addressed (Boyack, 1990). First, it is necessary to demonstrate the capability of the code models and solution procedure to simulate the physical processes expected during the scenario of interest. Thus, simplifications, assumptions and nodalization schemes may affect the behavior of the code physical models and the solution procedures. For instance, a coarse geometric representation of a NPP may miss local important local effects that could control a transient's behavior.

Second, the scalability of the results produced by physical models, usually developed to predict phenomena in separate effect test and scaled integral facilities, to a full size nuclear power plant (NPP) should be determined. The predictive capability of the code physical models and experimental correlations is neither perfect nor valid for all applications. For this reason, validation and assessment of the code models can provide useful information about their performance under a wide range of physical conditions. Such information can then be quantified as an uncertainty associated with the code physical models, and then propagated to the code results by means of appropriate techniques. Unfortunately, many of the integral and most of the separate effect test facilities are scaled down with respect to full size NPPs. Thus, a good performance by the models in these assessments is no guarantee that they will do as well when applied to full scale NPPs, since different physical phenomena may dominate the transients' development at different scales.

Finally, the initial state of the NPP may not be fully known, and the value of some important parameters or boundary conditions may be known only approximately. The results calculated by the code under such conditions will necessarily reflect this uncertainty.

THEORETICAL BASIS

This lecture introduces one of the most used uncertainty propagation methodologies today, which is based on a statistical the propagation of the sources of uncertainty in best estimate analyses.

This methodology has been established on the theoretical developments made at Los Alamos National Laboratory (McKay 1988) and at the Gesellschaft für Reaktorsicherheit (GRS) (Glaeser 1994, 1998). Its main feature is the propagation of uncertainties affecting the computer simulation using the statistical estimation of uncertainty measures by means of non-parametric probability density functions (*pdf*). Uncertainties in the code are treated as *stochastic variables*, so that a deterministic code transforms a stochastic input into a stochastic output: the uncertainty in code inputs is propagated to the output through a series of code executions.

If we assume that the vector $\boldsymbol{X} = (X_1, X_2,\ldots, X_K)$ can be considered to behave statistically as a random variable, we can create the set $(\boldsymbol{X})_1, \ldots, (\boldsymbol{X})_N$ as a sample of size N of $\boldsymbol{X}$. If each of the elements of the vector $\boldsymbol{X}$ is considered as an input variable for a given computer code, then the sample of size N of $\boldsymbol{X}$ contains N different values for the $X_{j=1\ldots K}$ obtained by a statistical sampling procedure of each one of the $X_{j=1\ldots K}$. The N sets of values for the K-components of the vector $\boldsymbol{X}$ represent the uncertainty in the determination of the values for each X_k. By executing the computer code N times, one for each set of $(\boldsymbol{X})_{j=1\ldots N}$ values, a set of N values $(\boldsymbol{Y})_1, \ldots, (\boldsymbol{Y})_N$ for, say, M output variables $\boldsymbol{Y} = (Y_1, \ldots, Y_M)$ can be obtained, which will also behave stochastically: a deterministic function transforms a random variable into another random variable.

$$\begin{pmatrix} (\boldsymbol{X})_1 \\ (\boldsymbol{X})_2 \\ \vdots \\ (\boldsymbol{X})_N \end{pmatrix} \equiv \begin{pmatrix} (X_1,\ldots,X_k)_1 \\ (X_1,\ldots,X_k)_2 \\ \vdots \\ (X_1,\ldots,X_k)_N \end{pmatrix} \xrightarrow[\text{to Input}]{\text{Assign Values}} \begin{pmatrix} (x_1,\ldots,x_k)_1 \\ (x_1,\ldots,x_k)_2 \\ \vdots \\ (x_1,\ldots,x_k)_N \end{pmatrix} \rightarrow \begin{pmatrix} Code\left[(x_1,\ldots,x_k)_1\right] \\ Code\left[(x_1,\ldots,x_k)_2\right] \\ \vdots \\ Code\left[(x_1,\ldots,x_k)_N\right] \end{pmatrix} \equiv \begin{pmatrix} (y_1,\ldots,y_m)_1 \\ (y_1,\ldots,y_m)_2 \\ \vdots \\ (y_1,\ldots,y_m)_N \end{pmatrix} \equiv \begin{pmatrix} Code\left[(\boldsymbol{X})_1\right] \\ Code\left[(\boldsymbol{X})_2\right] \\ \vdots \\ Code\left[(\boldsymbol{X})_N\right] \end{pmatrix} \equiv \begin{pmatrix} (\boldsymbol{Y})_1 \\ (\boldsymbol{Y})_2 \\ \vdots \\ (\boldsymbol{Y})_N \end{pmatrix}$$

Figure 1: Uncertainty Propagation Process.

Figure 1 represents the process of uncertainty propagation, with the function $Code\ [(\boldsymbol{X})_i]$ representing a code execution for the sample element i, formed by the sampled values of K input variables. Given the stochastic character of the probability distribution of ***Y*** (the probability distribution determines the uncertainty of ***Y***), statistical methods can be applied to quantify the uncertainty information contained in the sample of outputs $(\boldsymbol{Y})_1, \ldots, (\boldsymbol{Y})_N$.

The extraction of statistically meaningful information from the sample of outputs $(\boldsymbol{Y})_{i=1\ldots N}$ is accomplished by applying appropriate techniques that use the information in $(\boldsymbol{Y})_{i=1\ldots N}$ to generate for the code results.

One of the most useful uncertainty estimators is the *statistic* called **quantile**: a point in the sample space of a random variable such that the probability of the variable being **less than or equal to** the quantile is a given value $p < 1$. In mathematical terms, a $p = 0.95$ quantile for a random variable ***X*** is expressed as:

$$P(X \le x_{.95}) = 0.95, \tag{1}$$

and, in general, the p-quantile, x_p is:

$$P(X \le x_p) = p \tag{2}$$

Quantiles are only known precisely when the actual *pdf* is known. In case this is not possible, as mentioned above, it is important to obtain an estimate of their value. This is achieved by calculating the Tolerance Interval for a given stochastic variable, which is a more qualitative and conservative measure of uncertainty (McKay 1988). The probability by which they overestimate a given quantile can be determined precisely by using non-parametric statistical methods applied to a sample of values of the random variable. The Tolerance Intervals thus represent the range of variation expected in a given variable, with a certain level of confidence. In the case of uncertainty propagation, this variability is induced by the uncertainty in the input variables and in the code models, thus making the tolerance interval a very useful estimator of uncertainty when the actual *pfd* is unknown. In practice, a Tolerance Interval (L, U) is an *statistic* of a random variable that contains a specified fraction of the variable's probability p, with a prescribed level of confidence, γ(Crow 1960). For the above case, $p = 0.9$ and $\gamma = 0.95$. Tolerance Intervals are constructed from sampled data so as to enclose p% of the population of a random variable ***Y*** with a given confidence γ. They show where most of the population of the random variable ***Y*** can be expected to lie.

Uncertainty in input variables and code models is typically quantified by determining *pdfs* or any other appropriate *statistical measures* that quantify the uncertainty in the knowledge of their true values. After these *measures*, e.g. *pdfs*, and their probable ranges of variation have been assigned to the input variables and code models, the space of those random variables has to be sampled. The important decision at this stage is to determine the size of the sample, which is equal to the number of code executions needed.

A rigorous approach to determining the minimum sample size was introduced in the statistical uncertainty propagation methodology by researchers at the GRS in Germany by making use of the Noether´s (Noether 1967, Conover 1980) expression obtained by defining the upper and lower limits of the tolerance interval as the estimates of the upper and lower quantiles for the desired probability content of the interval. Details about the derivation of this formulation can be found in (Conover 1980):

$$\sum_{j=0}^{m-1} \binom{N}{j} q^{N-j} (1-q)^j \leq \alpha \text{ for the upper tolerance limit.} \quad (3)$$

$$\sum_{i=0}^{r-1} \binom{N}{i} q^{N-i} (1-q)^i < \alpha \text{ for the lower tolerance limit.} \quad (4)$$

$$\sum_{i=0}^{r+m-1} \binom{N}{i} (1-q)^i q^{N-i} \leq \alpha \text{ for upper and lower tolerance limits simultaneously.} \quad (5)$$

Where $(1\text{-}\alpha)$ is the confidence of the tolerance interval, q is the proportion of the population within the interval, N is the size of the sample (the unknown), and r and m ($r<m$)are numbers related to the elements in an ordered sample: e.g. $r=0 \rightarrow X^{(0)}=-\infty$, $m=0 \rightarrow X^{(N+1-m)} = +\infty$, and $r=1$, $X^{(1)}$ is the smallest element in the ordered sample, $m=1$, $X^{(N)}$ is the largest element in the ordered sample.

The most used values are $r+m=1$ (one-side tolerance limits) and $r+m=2$ (two-sided tolerance limits). Solution of this equation for N, when q and $1\text{-}\alpha$ are known, yields the minimum sample size. In (Conover 1980) tables are given with the solution for various pairs $(1\text{-}\alpha, q)$ and for the cases $r+m=1$ and $r+m=2$. For $r+m=2$, Eq. (5) is known as the Wilks Formula (Wilks 1962):

$$1-\gamma \geq (1-N)\beta^n + \beta^{n-1} \text{ with } 1\text{-}\gamma = \alpha \text{ and } q = \beta. \quad (6)$$

For instance, the two-sided tolerance limits for 95% of the population $(q=\beta=0.95)$, with 95% confidence $(\gamma = 1\text{-}\alpha = 0.95)$ the minimum sample size, N, is 93.

PROCESS OF UNCERTAINTY ANALYSIS

The practical application of the statistically based uncertainty propagation methodology has five main steps:

1. Selection of uncertain Input Variables and Code Models.
2. Assignment of Uncertainty Information and Sampling Procedure.
3. Determination of the Sample Size for the statistical significance of the uncertainty measures of the output variables.
4. Execution of the Code and Assembly of the Output Sample File.
5. Statistical Processing of the Output Sample File. Uncertainty and Sensitivity Quantification.

Of them, the most crucial step is number 2, because the quality of the final uncertainty estimate of the code´s results depends on the quality of the uncertainty information. Uncertainty in initial and boundary conditions is usually obtained based on plant measurements. For the estimation of uncertainty in code models, however, a variety of methods have been devised to produce uncertainty information (Vinai 2007a, 2007b). Most of them rely on statistical techniques that try to reduce the need for so-called “expert opinion” in the process. One of them is based on the determination of an estimate for the *pdf* of model´s predictions with a partition of the physical space, on which the model is applied depending on the variation of its uncertainty when the model makes predictions inside this space. This lecture introduces the main concepts on which this method is based.

APPLICATION TO COUPLED CALCULATIONS

The statistical approach explained in this lecture, can also be applied to coupled thermal-hydraulic-neutronics calculations, or to any coupled code analysis, by following a coherent approach for each of the codes involved in the coupling. The development of the uncertainty methodology for the coupling of System codes with CFD codes, for instance, is a complex process that involves the transfer of uncertainty in input variables that specify the initial state and boundary conditions of the system modelled with CFD in a similar manner as that followed for the System codes input variables, by a specially developed script that modifies the specific uncertain variables at the beginning of every coupling execution.

The introduction of the uncertainty in the relevant CFD code models, however, is a more complex process that must take into account:

- The local models used by CFD codes are refined if we compare to the 1-D averaged correlations of the System codes. If accurate uncertainty information is already difficult to draw from the models used in the System codes, it becomes even more complicated to draw in the case of CFD physical models, since the validation and assessment databases in this case are considerably sparse.
- The accessibility to the source of CFD codes. For this reason, even if uncertainty information for some of the CFD code physical models were available, its introduction in the uncertainty propagation procedure would be limited to those models which are directly accessible from the input files.
- After the uncertainty information has been introduced to the CFD codes in any of the manners described above, the coupled calculation procedure is similar to launching the System code alone in terms of code run management. The procedures already developed and tested for the Best Estimate System Codes, in order to manage the code execution process, must then be adapted to take into account the particularities of the coupled calculation procedure. That is, the launching of parallel CFD code processes synchronized with the System code.

Once the coupled executions have been completed, there is also the need to adapt the scripts already developed for the preparation of the System output sample files according to the System output characteristics to the CFD codes features. In order to process statistically the relevant outputs, it is necessary to assemble the coupled System/CDF output sample files. Such a procedure is relatively simple in the case of System codes only, but it may require the development of special scripts to process the CFD output information appropriately: e.g. contour plots, vector field plots which are two-dimensional or three-dimensional in nature and will require special statistical processing.

Figure 7 describes the flow of uncertainty information in a coupled calculation procedure. Due to the statistical nature of the uncertainty propagation procedure that we follow, the information which is passed between the two codes contains the effect of the uncertainties in the input variables and physical models assigned in a similar manner to both the System code and to the CFD code. As mentioned above, it may not be possible to take into account properly the uncertainty in the physical models of the CFD code. The only output uncertainty of interest may be that of the CFD part of the coupled calculation. In either case the methodology will also provide the output uncertainty for the System code output variables of interest.

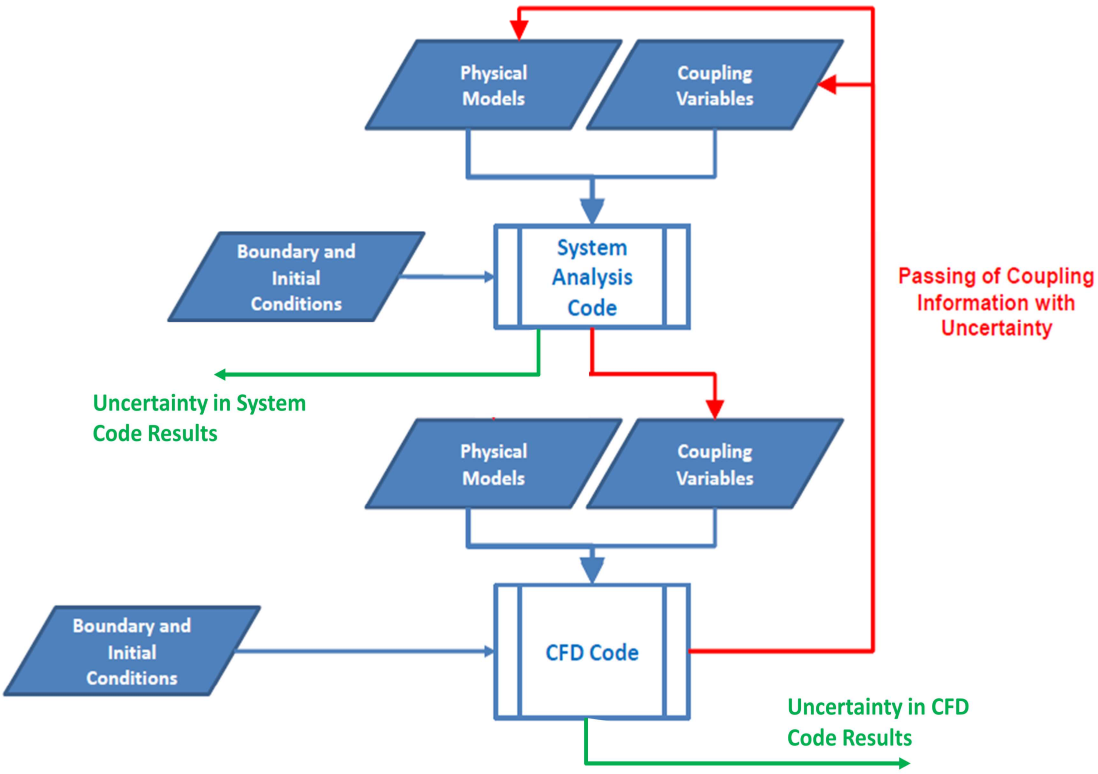

Figure 7: Scheme of the flow of uncertainty information in System-CFD coupled calculations

REFERENCES

McKAY, M.D., 1988, Uncertainty Analysis, Ch. 4, Ronen, Y. Editor, CRC Press, Florida, USA.

GLAESER, H., et al., 1998, in Uncertainty Methods Study for Advanced Best Estimate Thermal-hydraulic Code Applications, Vol. 2, Report from the Uncertainty Analysis Group, CSNI.

GLAESER, H., Hofer, M., Kloos, K. and Skorek, T., 1994, Unsicherheits- und Sesitivitätsuntersuchungen von Analysen mit Thermohydraulik-Recehenprogramme, GRS Abschlussbericht, GRS-A-2177.

NOETHER, G.E., 1967, Elements of Nonparametric Statistics, Wiley Ed., New York, (2.4, 3.3, 5.1, 5.7, 5.8, 5.9).

CONOVER, W.J., 1980, Practical Nonparametric Statistics, 2nd Ed., Wiley Ed., New York.

VINAI, P., 2007a, Development and application of objective uncertainty measures for nuclear power plant transient analysis, EPFL, PhD Thesis No. 3897.

VINAI, P., Macian-Juan, R. and Chawla, R., 2007b, ´A statistical methodology for quantification of uncertainty in best estimate code physical models´, *Annals of Nuclear Energy*, **34**-8, 628-640.