Nuclear Engineering and Design 262 (2013) 306–318

Contents lists available at SciVerse ScienceDirect

Nuclear Engineering and Design

journal homepage: www.elsevier.com/locate/nucengdes

Using finite mixture models in thermal-hydraulics system code uncertainty analysis

S. Carlos[a,*], A. Sánchez[b], D. Ginestar[c], S. Martorell[a]

[a] Department d'Enginyeria Química i Nuclear, Universitat Politècnica de València, Camí de Vera s.n, 46022 València, Spain
[b] Department d'Estadística Aplicada i Qualitat, Universitat Politècnica de València, Camí de Vera s.n, 46022 València, Spain
[c] Department de Matemàtica Aplicada, Universitat Politècnica de València, Camí de Vera s.n, 46022 València, Spain

HIGHLIGHTS

- Best estimate codes simulation needs uncertainty quantification.
- The output variables can present multimodal probability distributions.
- The analysis of multimodal distribution is performed using finite mixture models.
- Two methods to reconstruct output variable probability distribution are used.

ARTICLE INFO

Article history:
Received 26 October 2012
Received in revised form 17 April 2013
Accepted 19 April 2013

ABSTRACT

Nuclear Power Plant safety analysis is mainly based on the use of best estimate (BE) codes that predict the plant behavior under normal or accidental conditions. As the BE codes introduce uncertainties due to uncertainty in input parameters and modeling, it is necessary to perform uncertainty assessment (UA), and eventually sensitivity analysis (SA), of the results obtained. These analyses are part of the appropriate treatment of uncertainties imposed by current regulation based on the adoption of the best estimate plus uncertainty (BEPU) approach. The most popular approach for uncertainty assessment, based on Wilks' method, obtains a tolerance/confidence interval, but it does not completely characterize the output variable behavior, which is required for an extended UA and SA. However, the development of standard UA and SA impose high computational cost due to the large number of simulations needed. In order to obtain more information about the output variable and, at the same time, to keep computational cost as low as possible, there has been a recent shift toward developing metamodels (model of model), or surrogate models, that approximate or emulate complex computer codes. In this way, there exist different techniques to reconstruct the probability distribution using the information provided by a sample of values as, for example, the finite mixture models. In this paper, the Expectation Maximization and the k-means algorithms are used to obtain a finite mixture model that reconstructs the output variable probability distribution from data obtained with RELAP-5 simulations. Both methodologies have been applied to a separated effects experiment, and to an integral effects simulation.

1. Introduction

Nuclear Power Plant safety analysis is mainly based on the use of neutronic and thermal-hydraulic models to predict the plant behavior under normal or accidental conditions. Thermal hydraulic calculations can be performed using either conservative or best estimate (BE) codes, providing this latter option more realistic results. The general BE modeling methodology process comprises different steps, from capturing reality to conceptual models to convert those models to computerized BE codes. In this process there are always numerous simplifications, model approximations, round-off errors, numerical techniques, and so on, which cause uncertainties in the calculation (Pourgol-Mohammad, 2009). Thus, the uncertainty of the results obtained with the code (code output), which mainly depends on the uncertainty of code inputs and the modeling itself, has to be quantified in order to give credit to the predictions obtained (Pourgol-Mohammad et al., 2011).

From the last decade, the regulatory bodies allow the use of thermal-hydraulic BE simulation codes to guarantee the safe operation of nuclear installations, but requiring, at the same time, an appropriate treatment of uncertainties (Boyack et al., 1990; Wilson et al., 1990; Wulf et al., 1990). The treatment of uncertainties should

* Corresponding author. Tel.: +34 3877000x7617; fax: +34 3877639.
E-mail address: scarlos@iqn.upv.es (S. Carlos).

0029-5493/$ – see front matter
http://dx.doi.org/10.1016/j.nucengdes.2013.04.030

include not only uncertainty quantification but also uncertainty control and reduction. The plant simulations undertaken using best estimate codes combined with uncertainty analysis is known as best estimate Plus Uncertainty (BEPU) approach (de Crécy et al., 2008).

In the process of thermal hydraulic modeling using BE codes the initial plant state is represented by the initial and boundary conditions of the plant model, which are input variables in the BE simulation. In many cases their values are unkown or uncertain, an such uncertainty is transmitted through the code to the output variable of interest. The uncetainty analysis (UA) using BE codes may require a large number of simulations, depending on the UA approach adopted, to produce a sample of the output variable what results in a large computational cost.

In the literature, there exist different approaches to quantify uncertainty in BE codes. For example, in Cacuci and Ionescu-Bujor (2000a,b) a deterministic approach is followed using the adjoint sensitivity analysis method for RELAP code. This approach needs to implement the uncertainty quantification method in a new code, which has to be coupled with the thermal-hydraulic code. This is known as an intrusive method as the original thermal hydraulic code has to be modified. But the most popular approaches to quantify code uncertainty are the non intrusive methods. Such approaches use the thermal-hydraulic code to produce output variables linked to the input variables (Guba et al., 2003). That is, given and input variable vector, $\vec{x}$, the computer code transforms it into a vector $\vec{y}$ of output variables,

$$\vec{y}(t) = f(\vec{x}, t). \tag{1}$$

In practical, this link is very complex but it is assumed to be deterministic, that is, once the input variables are fixed the same output is obtained from the code within the computation accuracy of each run. Some of the non intrusive methodologies developed to quantify best estimate codes uncertainty are the CSAU (Boyack et al., 1990; Wilson et al., 1990; Wulf et al., 1990), the GRS methodologies (Glaeser et al., 1994) ASTRUM and IMTHUA (Pourgol-Mohammad, 2009). Reference Pourgol-Mohammad (2009) provides a detailed comparison of the uncertainty methodologies developed, and applied to the thermal hydraulic calculations. In these methodologies, it is assumed that the input variables are uncertain, and follow a statistical distribution. In this way, fixing the time of the transient, N random samples of an output variable are obtained, which are due to varied input parameter values and computer code model parameters.

An approach for performing UA of thermal-hydraulic BE code results, which has been accepted by regulatory authorities to evaluate the safety of a nuclear plant for selected accident scenarios (e.g. LOCA scenarios), consists of the use BE code results coupled with order statistics to determine a one-side tolerance limit or an uncertainty interval in order to be compared with an acceptance limit given by regulation (NRC, 1989). Thus, uncertainty in the output variables can be quantified by obtaining a tolerance/confidence interval, making use of the advantage of order statistics (Guba et al., 2003). Thus, assuming there is one output variable, y, with a probability distribution $g(y)$. If we carry on N runs with varying inputs, we obtain a sample $\{y_1, y_2, \ldots, y_N\}$ of the output variable y. The usual approach is to construct two random functions $L = L(y_1, y_2, \ldots, y_N)$ and $U = U(y_1, y_2, \ldots, y_N)$, called tolerance limits, such that

$$P\left\{\int_L^U g(y)dy > \gamma\right\} = \beta, \tag{2}$$

where

$$\int_L^U g(y)dy = A(y_1, y_2, \ldots y_N), \tag{3}$$

is a random variable, called probability content, which measures the portion of the distribution included in the random interval [L, U]. Probability β is the confidence level, and γ is a non-negative real number not grater that 1. It is desirable to have values of β and γ as large as possible inside the interval [0, 1]. Having fixed β and γ, it becomes possible to determine the number of runs N necessary to determine an appropriate interval [L, U]. The first works that discussed the problem of setting tolerance/confidence intervals based on samples where developed by Wilks (1941), and they are the basis of uncertainty methodologies for quantifying best estimate codes uncertainty (Glaeser et al., 1994; Guba et al., 2003). This approach has the advantage that the number of runs, N, necessary to determine the tolerance limits, is much lower than the runs necessary in a Monte Carlo approach.

However, the information provided by the tolerance limits methodology does not completely characterize the output variable behavior. In addition, it tends to provide conservative estimates of the tolerance limits when the statistic of first order is adopted, which is the usual case as this imposes the lowest N. More precise results can be obtained adopting statistics of order two, three and higher, but at the expenses of increasing N and therefore the computational cost (Hong et al., 2011).

By increasing N one can obtain more information about the output variable of the BE code, even the shape of its probability distribution function. For example, in Chang et al. (2004), to analyze the pressurized thermal shock a thermal-hydraulic uncertainty analysis is performed reconstructing the probability density function and the cumulative distribution function of the downcomer temperature, using RELAP-5 best estimate code. The advantage of this UA approach is that it allows obtaining a more precise estimate of the tolerance limits and at the same time provides the information required to perform Standard Sensitivity Analysis (SA) (Langewisch, 2010). SA refers to a collection of tools whose aim is to elucidate the dependency of the model output on the set of model inputs, in particular but not only, to assess how the uncertainty in output depends on the uncertainty in inputs. There are different SA methods (Borgonovo, 2006; Saltelli et al., 2008) that provide a better understanding of how various parameters affect the response of a system.

The predominant limiting factor in most of the UA and SA methods referred is the very large computational burden due to the large number of simulations required. As a result, there has been a recent shift in research efforts toward developing methods for approximating, or emulating, complex computer models. These approximate models are referred to as metamodels, or surrogate models. In brief, a metamodel is a model of a model, i.e. a metamodel is a simplified model that is capable of approximating the output from the underlying computer model. Once constructed, the metamodel serves as a fast-running surrogate to the computer model and is used to quickly predict outputs from a Monte Carlo simulation. Thus, the UA and SA are performed through the approximate model, thereby circumventing the initial computational burden (Strolie et al., 2009). For example Neural networks are used to estimate safety margins on the maximum fuel cladding temperature reached during a complete group distribution header blockage scenario in a RBMK-1500 nuclear reactor (Secchi et al., 2008). This approach has the advantage of its fast performance but the results obtained can be poor depending on the output variable behavior (Zio, 2006; Cadini et al., 2008).

On the other hand, polynomial chaos expansion methods have been recently used to reconstruct the probability distribution and to estimate its parameters, as a lower number of runs are needed compared with Monte Carlo approaches (Sundret, 2008; Eaton and Williams, 2010; Gilli et al., 2012).

So assuming that the input variables are random and follow a certain probability distribution, if its variance or range of variation