MASCOT-NUM 2022
Clermont Ferrand, France
7-9 June 2022

Statistical methods for the study of computer experiments failures: Application to a fuel-coolant interaction simulation code

F. Hakimi
CEA cadarache - Institut de mathématiques de Toulouse (IMT)

Supervisor(s): C. Brayer (CEA), A. Marrel (CEA), F. Gamboa (IMT) and B. Habert (CEA)

PhD expected duration: Novembre. 2019 - Octobre. 2022

Address: Cadarache, 13115 Saint-Paul-lez-Durance, bâtiment 238

E-mail: faouzi.hakimi@cea.fr

Abstract:

In the framework of risk assessment in nuclear accident, some calculation tools are now widely used to understand, model and predict physical phenomena. In particular, the characterisation of severe accident involves the use of a large amount of models linked together in a global numerical simulator. Each of these models and the way they are linked together require a large set of input parameters. These parameters may be subject to uncertainties due to the partial characterisation of the phenomenon, the state of knowledge or modelling uncertainties. Exploring the whole range of variation of these inputs is then crucial in order to understand the overall behaviour of the simulation code.

To achieve this exploration, Monte-Carlo type methods are often used: the inputs space is randomly sampled. A code run is performed on each sampled point. These type of exploration methods have the advantage of allowing the use of statistical inference tools: central limit theorem, estimation theory, statistical test, confidence interval theory... Monte-Carlo methods thus enable to get quantitative results with associated degrees of confidence (e.g via confidence intervals and statistical tests). However, the codes under study are expensive in terms of computing time. This means that we have access to only a limited number of simulations. Hence, this issue limit guides the choice of statistical tools and methods to be used.

In addition, sometimes the execution of the code fails. This may be due to some numerical problems or inadequacy of the physical models or the way they are implemented in the simulator, in the range of variation of the input parameters. The analysis of these code failures and their relationship to the input parameters can provide valuable information to better understand the code behaviour and its validity domain. More generally, this analysis intends to improve the robustness of the simulation software and modelling process.

Ideally, classification algorithms such as random forest [2] or logistic regression [3] could be used to predict the code failures according to the inputs. Nevertheless, as previously mentioned, the severe accident codes considered here are complex, computationally expensive and take many variables as inputs. These drawbacks, and in particular the limited number of possible simulations, make it difficult to obtain satisfactory results with this kind of algorithms.

To overcome it, we propose statistic-based tools to analyse code failures. More precisely, failure occurrence can be considered as a binary output and sensitivity analysis techniques adapted to this type of variable can be applied. Sensitivity Analysis methods aim at determining how the variability of model's inputs affects the fluctuation of the output. Many methods have been developed for this purpose [4]. However, most of classic tools used for sensitivity analysis, such as Sobol' indices [8] or the Elementary Effect method [7] are not well tailored for code failures. Furthermore, the estimation of total Sobol' indices requires a large number of simulations. Recently, some new tools based on dependence measures have been proposed for global sensitivity analysis. They allow to remove some of these limitations [9]. Among them, the Hilbert-Schmidt Independence Criterion [1] denoted by HSIC, generalises the notion of covariance between two random variables. Relying on the embedding of probability distributions in Reproducing Kernel Hilbert Spaces, the HSIC dependence measure allows to fully characterise the independence between two

variables (with certain parametric choices, namely the use of characteristic kernels). Statistical hypothesis tests of independence can thus be built upon HSIC measures. HSIC has also the advantage that it can be used with many types of variables, including binary variables. Last but not least, the HSIC estimation often requires only a limited budget of simulations (of the order of a few hundred) to obtain reliable estimates. Hence, we propose two methods to perform a sensitivity analysis on code failures:

- a first approach based on the HSIC measures and associated statistical tests, to assess the dependence between the inputs and the occurrence of code failures;
- a second approach based on Goodness-of-fit tests that compares the conditional probability distributions of the inputs knowing the code failures.

The development of these methods is carried out in the context of severe nuclear accidents. More precisely, the simulation code studied is MC3D [5, 6]. This code models the fuel-coolant interaction (FCI). This interaction can lead to the steam explosion phenomenon. When exploring the input space of this code using a Monte-Carlo method, we found a fairly high failure ratio (around 35%). The two proposed approaches (HSIC-based and Goodness-of-fit tests) were applied and resulted in the consistent identification of inputs that significantly influence the code failures. It has also paved the way for a physical interpretation of some of the input/failure relationships, while others that are more difficult to explain and will require further investigation.

References

[1] A. Gretton, O. Bousquet, A. Smola, and B. Schölkopf. Measuring Statistical Dependence with Hilbert-Schmidt Norms. *Algorithmic Learning Theory 3734*, pages 63–77, 2005.

[2] T.K. Ho. Random decision forests. In *Proceedings of the Third International Conference on Document Analysis and Recognition - Volume 1*, pages 278–282, USA, 1995.

[3] D.W. Hosmer and S.Lemeshow. *Applied logistic regression.* John Wiley and Sons, 2000.

[4] B. Iooss and P. Lemaître. A review on global sensitivity analysis methods. In *Uncertainty management in Simulation-Optimization of Complex Systems: Algorithms and Applications.* Springer, 2015.

[5] R. Meignen, S. Picchi, J. Lamome, B. Raverdyand S.C Escobar, and G. Nicaise. The challenge of modeling fuel–coolant interaction: Part I – Premixing. *Nuclear Engineering and Design 280*, pages 511–527, 2014.

[6] R. Meignen, B. Raverdy, S. Picchi, and J. Lamome. The challenge of modeling fuel–coolant interaction: Part II – Steam explosion. *Nuclear Engineering and Design 280*, pages 528–541, 2014.

[7] M.D Morris. Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics 33*, pages 161–174, 1991.

[8] M. Sobol'. Sensitivity Estimates for Nonlinear Mathematical Models (english translation). *Mathematical Modeling and Computational Experiment 4*, pages 407–414, 1993.

[9] S. Da Veiga. Global sensitivity analysis with dependence measures. *Journal of Statistical Computation and Simulation 85*, pages 1283–1305, 2015.

Short biography – After studying statistical signal processing at Centrale Marseille, Faouzi Hakimi started his PhD at the CEA of Cadarache in November 2019. The general objective of his thesis is to develop and implement statistical methods in order to improve reliability of complex simulation codes in the context of a severe nuclear accident.