

Modeling of computer experiments for uncertainty propagation and sensitivity analysis

Anestis Antoniadis · Alberto Pasanisi

Received: 21 July 2011 / Accepted: 27 July 2011 / Published online: 21 September 2011
© Springer Science+Business Media, LLC 2011

Nowadays, computer simulation has undoubtedly become one of the major tools to investigate and predict the behavior of complex systems in more and more disciplinary fields. A crucial question is how to account for uncertainties, tainting both input variables and (possibly) the computer model itself, in the results provided by the numerical code. That is particularly interesting in industrial practice, when results of computer simulations are eventually used to guide decisions, which can involve important financial, societal and safety stakes. For instance, in structural reliability problems, one is interested in assessing the probability for some state variables of a system (e.g. pressure, temperature, . . .) to be inside, or outside, a given domain, which is associated to safe operating conditions.

While computer simulations are faster and cheaper than physical experiments, computer models generate data (often large amounts) that must be analyzed and care is needed at the design stage to determine more cost-effective and informative simulation settings. Moreover, determining the most influential variables for the numerical code and the relevant parameter ranges within which to set up a computer experimental design is a critical and difficult step in the practical use of any formal statistical experimental planning, be it for screening or optimization purposes.

This short list of problems is, of course, not exhaustive but one can easily understand that proper statistical and

probabilistic tools are needed to address such issues. These problems become more and more difficult to deal with as far as the complexity of the computer model (and the CPU time needed for a single run) increases. In these cases, standard brute force sampling approaches are unfeasible and classical solutions involve advanced Monte Carlo methods and/or meta-modeling techniques. The first class of methods provides sampling strategies for building small variance estimators of the quantities of interest, the last strategy aims to build a surrogate model (also called emulator or response surface) which fits the real CPU time consuming code in a satisfactory way, whilst being very much cheaper. Following both strategies, the algorithmic challenge is the same: how to use a given computational budget (i.e. a certain number of expensive model runs) in the best way possible.

The interest about uncertainty analysis and computer experiments has noticeably grown during these last years. Several working groups put nowadays together industrial and academic researchers and a very interesting dynamics is engaged: the first come with real and complex problems to cope with, the last find in these problems a source of inspiration for their works, as well as the opportunity to test their innovative computing techniques on real-world industrial case-studies. Among these working groups, one can cite the MUCM project (Managing Uncertainty in Complex Models) as well as the French research group MASCOTNUM (Stochastic Analysis Methods for Codes and Numerical treatments) and the consortium DICE (Deep Inside Computer Experiments). Their works demonstrate that this domain remains a flourishing and developing field.

The idea and stimulus of this special number of *Statistics and Computing* germinated within the framework of the OPUS project (Open source Platform for Uncertainty treatment in Simulation), funded by the French *Agence Nationale de la Recherche* (ref. ANR-07-CIS7-010 and ANR-

A. Antoniadis (✉)
Lab. Jean Kuntzmann, University Joseph Fourier, Tour IRMA,
B.P. 53, 38041 Grenoble, France
e-mail: Anestis.Antoniadis@imag.fr

A. Pasanisi
Industrial Risk Management Dept., EDF R&D, 6 quai Watier,
78401 Chatou, France
e-mail: Alberto.Pasanisi@edf.fr

07-TLOG-015), which aims to build, structure and maintain Open Source software modules implementing advanced method for uncertainty analysis. OPUS (2008–2011) gathered ten partners coming from different communities (industrial, academia, information technology) and disciplinary fields. This special issue therefore presents a variety of papers on statistical problems that arise in the domains of uncertainty analysis and computer experiments modeling and reflect several points of view. We believe it will be of interest for different kind of readers: some will be most interested in the theoretical aspects of the proposed algorithms, others will focus on the computational issues and on the perspectives of use in real engineering studies. All of the articles presented here were accepted after *Statistics and Computing* regular review process.

Through a series of 11 articles, this special issue highlights the wide range of areas where uncertainty analysis and computer experiments are applied today. We give below, a short overview of the different papers which demonstrates how they cover a wide array of variations on the general theme.

A first group of papers deals with the design of numerical experiments and response surfaces.

Challenging problems arise in the design and analysis of experiments with methods which aim to uniformly cover the space of the model's inputs when the goal of the study is mainly to explore the numerical code, without making any particular assumption about the emulator to be built. The paper of L. Pronzato and W. Müller gives a quite exhaustive review of the so-called space-filling methods to build experimental design. The authors also extend the general case of space-filling to the case where the experimental design is used to build a kriging emulator of a computer model.

Y. Auffray et al. also study an experimental design problem, and in particular deal with the case where the domain of inputs is not hyper-cubic. Here, the classical strategy consisting of choosing the maximin design within the class of Latin hypercube designs may lead to poor exploratory properties. After a justification of the maximin criterion with respect to kernel interpolations, the authors use simulated annealing algorithms to determine maximin designs in bounded connected domains and provide a numerical example, coming from aerospace industry.

Gaussian process models (kriging) are one of the most popular approaches for generating emulators of physical experiments. R. Gramacy and H. Lee discuss in their paper the role played by the so-called nugget when fitting a kriging emulator. The point of view of the authors is that explicitly considering a non-zero nugget term in the covariance of the Gaussian process is valuable to build emulators with good statistical properties in case of sparse data or when some classical assumptions about the computer model, as stationarity for example, are not valid.

T. Mühlentädt et al. propose a data-driven methodology to build effective kriging emulators, the covariance of which explicitly takes into account the interaction structure of the data. The interaction structure is analyzed by using FANOVA based methods and visualized over a graph, called FANOVA graph, the clique structure of which is used for defining the covariance kernels.

V. Sambucini paper is concerned with an important problem in response surface methodology (RSM): what are the relevance and the reliability of a stationary point, in other words can we construct a “correct” confidence region for this point? She offers a comparison between traditional Box-Hunter exact confidence intervals for the stationary point and asymptotic intervals based upon a clever re-parametrization of the quadratic model making the dependence upon the stationary point explicit.

Finally, the paper by S. Cohen et al. develop a new algorithm for the construction of sequential designs for nonparametric regression models and performs a model selection in a multi-resolution setting when competing models are considered. The availability of such an algorithm is particularly useful when building response surfaces for sensitivity analysis.

Some other articles cope specifically with the problem of the sampling-based estimation of the (low) probability for the output of a numerical code to be greater than a given threshold value, which is commonly interpreted as the probability of failure of the system under investigation.

For estimating this probability, J. Bect et al. propose to build a Gaussian process emulator of the code following on a sequential experimental design (SUR, stepwise uncertainty reduction) based on the Bayesian decision-theory formulation of the estimation problem. The response surface, so obtained, is thus optimal for the particular problem to be solved.

F. Cérou et al. also deal with the problem of computation of the probability of rare events. They develop an original way for simulating rare events and present an interesting Monte Carlo based method for estimation of tail probabilities. To achieve this they adopt a particular system of interacting particles whose fluctuations are studied by using a particular Feynman-Kac representation.

This problem is also addressed in the paper by M. Muñoz-Zuniga et al., in which the authors are using for this purpose an algorithm that combines stratification and directional sampling strategies to fully take advantage of some useful features of both techniques. Their paper gives a very useful addition into the reliability literature on the robust estimation of small failure probabilities for high computational time computer codes.

Finally, two papers are specifically concerned with sensitivity analysis, which is an important field in the analysis of physical and computer experiments (i.e., understanding

how sensitive models and decisions are to uncertain knowledge of inputs and identifying which inputs contribute most to the uncertainty).

One way to perform such an analysis is by computing the so-called Sobol's indices. The latter task is especially challenging in high-dimensional settings. A. Jourdan suggests an interesting methodology, based on decomposition of the response surface using orthogonal basis of complex functions (Yates's basis) to overcome the computational burden associated to traditional methods like Monte Carlo estimation.

Finally, the paper of A. Marrel et al. proposes an appropriate methodology to perform global sensitivity analysis on stochastic computer codes. As it is usual in sensitivity analysis a surrogate model approach is followed. Here, the authors use joint meta-modeling techniques, which consist of building two separate meta-models: one for the mean and one for the dispersion of the computer code. Different families

of joint meta-models (Generalized Linear Models, Generalized Additive Models and Gaussian processes) are discussed and some numerical results are given.

The papers in the current issue often illustrate a special combination of practical potential and depth of theoretical interest. We hope that this special issue will stimulate a new wave of interest in the field of uncertainty and sensitivity analysis and computer experiments modeling and lead to exciting new results in the domain.

Acknowledgements We gratefully thank Prof. Gilles Celeux, editor-in-chief of this journal, for having invited us as guest editors of this special issue and for his helpful advices all along its preparation. We would also like to thank Melissa Fearon, Springer Senior Editor, for her help in preparing the final version of this special issue. We also thank all the anonymous reviewers for their accurate and valuable work and all the authors for their contributions.