Available online at www.sciencedirect.com

ScienceDirect

Procedia Social and Behavioral Sciences 2 (2010) 7583–7584

Procedia
Social and Behavioral Sciences

Sixth International Conference on Sensitivity Analysis of Model Output

Computer experiments with functional outputs: Global sensitivity analysis and metamodeling

Benjamin Auder[a], Bertrand Iooss[b]*, Michel Marquès[a]

[a]*CEA, DEN, Cadarache, 13108 Saint-Paul-lez-Durance, France*
[b]*EDF R&D, 6 quai Watier, 78402 Chatou, France*

Abstract

To perform uncertainty, sensitivity or optimization analysis on scalar variables calculated by a cpu time expensive computer code, a widely accepted methodology consists in first identifying the most influential uncertain inputs (by screening techniques), and then in replacing the cpu time expensive model by a cpu inexpensive mathematical function, called a metamodel. This paper extends this methodology to the functional output case, for instance when the model output variables are curves. Our screening approach is based on the analysis of variance and principal component analysis of output curves. Our functional metamodeling consists in a curve classification step, a dimension reduction step, then a classical metamodeling step. An industrial nuclear reactor application (dealing with uncertainties in the pressurized thermal shock analysis) illustrates all these steps.

Keywords: Computer model; Curve classification; Functional data; Metamodel; sensitivity indices; Uncertainty analysis

1. Introduction

The global sensitivity analysis (GSA) process is a key step for the development and the use of predictive complex computer models. In practice, when dealing with GSA methods, four main problems can arise:

- Physical models can involve complex and irregular phenomena sometimes with strong interactions between physical variables. This problem is resolved by using variance-based measures (Saltelli et al., 2000);
- Computing variance-based measures can be infeasible for cpu time consuming code. Metamodel-based techniques (Fang et al., 2006) solve this problem and provide a deep exploration of the model behavior;
- Numerical models can take as inputs a large number of uncertain variables (typically $d > 10$). Applying a screening technique (Saltelli et al., 2000) allows to rapidly indentify the main influent input variables;
- Numerical models can produce functional output variables, for instance spatially or temporally dependent. This problem has paid only little attention in the GSA methodology (Campbell et al., 2006; Marrel et al., 2010).

In this communication, we present an overall methodology applicable to output curves of cpu time consuming computer models. During the presentation, we will also present some algorithms in the case of output spatial maps.

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
E-mail address: biooss@yahoo.fr

doi:10.1016/j.sbspro.2010.05.129

This methodology is applied to an industrial application: the pressurized thermal shock analysis. This quantitative analysis aims at calculating the vessel failure probability of a nuclear pressurized reactor. One major challenge in such calculations is to propagate input uncertainties in thermal-hydraulic models, by using Monte Carlo methods which require several thousands of runs. However, such models are cpu time consuming (5 hours per run), involve several tens of uncertain inputs (d = 31) and the output variables of interest are several time-dependent curves (temperature, primary pressure, exchange coefficient) discretized in T = 512 values.

2. Overall methodology

2.1. Screening with generalized sensitivity indices

The screening step is achieved by using a factorial fractional design of resolution IV (requiring N=64 runs). A principal component analysis is applied on the 64 output curves and all the first order variance-based measures are obtained for each retained component. Therefore, the generalized sensitivity index (Lamboni et al., 2009) is obtained for each input. By ordering the inputs, we reduce the number of input variables to d' =12 main variables.

In order to fit a metamodel depending on these 12 inputs, a specific space filling design (Fang et al., 2006) of size N' = 600 is then built and these N' new runs are performed with the computer code.

2.2. Curve classification

Building a functional metamodel on these 600 output curves first requires a clustering stage. Indeed, the model outputs can follow different physical behaviors, each being represented by several curves. The method implemented first computes the Euclidian commute-time distance (Yen et al., 2005) between every pair of curves. An ascendant hierarchical classification step then provides the clusters $\mathbf{C}_1,\ldots,\mathbf{C}_K$. The model inputs are classified accordingly in the same time with a non-parametric technique like k-nearest neighbors.

2.3. Dimension reduction and functional metamodeling

The focus is now on a specific cluster of curves $\mathbf{C}_k$. We only assume that the output curves are sampled from a connected smooth manifold. The goal is to obtain small finite dimensional representations $\mathbf{r}_i$ of the curves to handle them easily through calculations. A modified version of the "Riemaniann Manifold Learning" (RML) algorithm of Lin and Zha (2008) is used, combined with a global alignment method (Teh and Roweis, 2003). The relation between sampled inputs $\mathbf{x}_i$ and vectors $\mathbf{r}_i$ is learned with any metamodel (here we choose the projection pursuit regression algorithm). Thus, so far, the metamodel can predict an estimated representation $\mathbf{r^*}$ from a new input $\mathbf{x^*}$. The last step consists in expanding $\mathbf{r^*}$ into a curve, and is achieved using specific features of the RML algorithm.

3. References

Campbell K., McKay M.D. and Williams, B.J., 2006: Sensitivity analysis when model outputs are functions. *Reliability Engineering and System Safety*, 91, pp 1468-1472.

Fang K-T., Li R. and Sudjianto A., 2006: *Design and modeling for computer experiments,* Chapman & Hall.

Lamboni M., Makowski D., Lehuger S., Gabrielle B. and Monod H., 2009: Multivariate global sensitivity analysis for dynamic crops models. *Field Crops Research,* 113, pp 312-320.

Marrel A., Iooss B., Jullien M., Laurent B. and Volkova E., 2010: Global sensitivity analysis for models with spatially dependent ouputs, *Environmetrics,* submitted, URL: http://fr.arxiv.org/abs/0911.1189.

Lin T. and Zha H., 2008: Riemannian Manifold Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30, pp 796-809.

Saltelli A., Chan K. and Scott E.M. (Eds.), 2000: *Sensitivity Analysis*, Wiley.

Teh, Y. W. and Roweis, S., 2003: Automatic Alignment of Local Representations. *Advances in Neural Information Processing Systems*, 15, pp 865-872.

Yen L., Vanvyve D., Wouters F., Fouss F., Verleysen M. and Saerens M., 2005: Clustering using a random-walk based distance measure. *Proceedings of the 13th Symposium on Artificial Neural Networks*, pp 317-324.