GAUSSIAN PROCESS METAMODEL AND WAVELET DECOMPOSITION FOR SENSITIVITY ANALYSIS OF A FUNCTIONAL OUTPUT OF COMPUTER CODE

A. Marrel

IFP, Reservoir engineering division, Rueil-Malmaison, amandine.marrel@ifp.fr

Introduction

In the simulations studies of flow and transfer in porous media like reservoir oil simulations, hydrogeological environmental transfer or risk assessment for CO2 geological storage, numerical models are implemented to simulate, understand and predict the transfer phenomenon. These computer codes can depend on a high number of uncertain input parameters (geophysical variables, chemical parameters, etc.). Therefore, to provide guidance to a better understanding of this numerical modelling and in order to reduce the response uncertainties most efficiently, sensitivity measures of the input importance on the response variability can be useful (Sobol [9], Saltelli et al. [8]). However, this global sensitivity analysis (based on Monte Carlo method) requires a large number of numerical model simulations, which is untractable for time expensive computer codes.

To overcome the problem of huge calculation time in sensitivity analysis, the computer code can be approximated by a metamodel also called response surface. The metamodel is built on an acceptable number of simulations of the code and requires a negligible calculation time. In our environmental transfer problems, the metamodel has to be efficiently built in the case of a high number of inputs (from 10 to 50 inputs) and with a small number of simulations available. We focused our research work on the use and the implementation of Gaussian process metamodel (Sacks et al. [7]) and its use to make the sensitivity analysis of the code.

Moreover, in the case of flow and transfer studies in porous media, the output of the numerical simulator is often functional. Indeed, the response of the simulator can be a function of time or 2D space (for example : spatial concentration of pollutant). A metamodel cannot be reasonably built to fit the output for each point of the time or space discretization. To overcome this problem of high dimensional output, a strategy to reduce the dimension of the output should be found.

Gaussian process metamodel and sensitivity analysis

The Gaussian process metamodel presents several advantages like its interpolating property or its analytical formulation which, under certain hypothesis, is very useful for sensitivity analysis. For the mean and the covariance functions which characterize the Gaussian process metamodel, we respectively choose a polynomial and a generalized exponential function. The crucial point is then to accurately estimate the hyperparameters of the covariance function. We propose a methodology combining an estimation procedure with a double input selection in order to build the Gaussian process metamodel in the case of a high number of inputs and with a small number of simulations available (Marrel et al. [4]). All the developed methodology is described. Then, to make the global sensitivity analysis, the Sobol indices are computed, using the Gaussian process metamodel (Marrel et al. [3], Oakley and O'Hagan [6]). The methodology and the computation of Sobol indices are illustrated by their application on the real case of a complex hydrogeological computer code, simulating radionuclide transport in groundwater. In this first part, only scalar output of simulators are considered.

Extension to functional output : strategy for reduction of high dimensional problem

In this case, a functional metamodel has to be defined (Fang et al. [1], Higdon et al. [2]). As a Gaussian process metamodel cannot be reasonably built to fit each point of the discretization, we propose a methodology to reduce the high dimension of the functional output. This methodology combines a decomposition on a wavelet basis, a judicious selection of the main coefficients of the decomposition and a Gaussian process modelling (Marrel [5]). Thus, in the case of 2D-spatial output, functional sensitivity indices are computed in order to obtain maps of Sobol indices. All the proposed methodology and the sensitivity maps are illustrated on the real case of hydrogeological computer code, previously mentioned.

References

[1] Fang K-T, Li R., Sudjianto A.: Design and modelling for computer experiments, *Chapman & Hall/CRC* (2006).

[2] Higdon D., Gattiker J., Williams B., Rightley M.: Computer model calibration using high dimensional output, *Journal of the American Statistical Association*, **103**, 570-583 (2008).

[3] Marrel A, Iooss B, Laurent B, Roustant O: Calculations of Sobol indices for the Gaussian process metamodel, *Reliability Engineering and System Safety*, **94**, 742-751 (2009).

[4] Marrel A, Iooss B, Van Dorpe F, Volkova E: An efficient methodology for modeling complex computer codes with Gaussian processes, *Computational Statistics and Data Analysis,* **52**, 4731-4744 (2008).

[5] Marrel A, Mise en oeuvre et utilisation du métamodèle processus gaussien pour l'analyse de sensibilité de modèles numériques, *Thèse de l'INSA de Toulouse* (2008).

[6] Oakley JE, O'Hagan A: Probabilistic sensitivity analysis of complex models: A Bayesian approach, *Journal of the Royal Statistical Society, Series B*, **66**, 751-769 (2004).

[7] Sacks J, Welch WJ, Mitchell TJ, Wynn HP: Design and analysis of computer experiments, *Statistical Science*, **4**, 409-435 (1989).

[8] Saltelli A, Chan K, Scott EM, editors: *Sensitivity Analysis*, Wiley, 2000.

[9] Sobol I.M., Sensitivity estimates for non linear mathematical models, *Mathematical Modelling and Computational Experiements*, 1:407 414, 1993.