

Sixth International Conference on Sensitivity Analysis of Model Output

Joint metamodeling for sensitivity analysis of continuous and stochastic inputs of a computer code

A. Marrel*

IFP, 1 & 4, avenue de Bois Préau, 92852Rueil-Malmaison, France

Abstract

To perform the global sensitivity analysis of a complex and cpu time expensive code, a mathematical function built from a small number of simulations referred to as a metamodel can be used to approximate the code. In some applications like oil reservoir simulations, the code output can depend on complex stochastic inputs such as random permeability fields. This paper proposes a new metamodeling approach to perform global sensitivity analysis of both scalar and such type of stochastic input. A joint metamodeling based on two Gaussian process metamodels is proposed to model the mean and the variance. Then, the sensitivity indices of scalar and stochastic inputs are estimated from this joint metamodeling. An application on a reservoir simulator illustrates the overall methodology.

Keywords: Gaussian process; Joint modeling; Metamodel, Sensitivity analysis; Sobol' index; Stochastic input.

1. Introduction

Complex computer simulations studies such as multiphase fluid flow simulation in porous media for oil reservoir forecasting can involve a high number of uncertain input parameters (geophysical variables, chemical parameters, etc.). To provide guidance to a better understanding of these numerical models and to reduce the response uncertainties most efficiently, sensitivity measures of the input importance on the output variability can be used (Saltelli et al., 2000). However, such simulators usually require several hours or days for a single run, therefore direct sampling methods for sensitivity analysis (Monte Carlo) are usually impractical. To overcome this problem of huge computation, the computer code can be approximated by a metamodel, i.e. a mathematical function built from a small number of simulations. In this work, we focus on Gaussian process (Gp) metamodel (Sacks et al., 1984, Oakley and O'Hagan, 2004); this approach was already used in a previous work (Marrel et al. 2009) to make sensitivity analysis of continuous scalar inputs. In oil reservoir simulations, the computer code can also depend on stochastic inputs like a geostatistical parameter (a porosity map for example), a fracture network or a group of inputs with relatively low but non negligible influence. Because of the stochastic behavior that they involved on the computer code output and their huge dimension, these stochastic inputs can not be handled with classical metamodel. A possible solution is to combine a modeling of the mean component with one of the variance component (variance related to the stochastic input). Zabalza-Mezghani et al. (1998) have proposed a joint modeling

* Corresponding author. Tel.: +33-147-525-320; fax: +33-147-526-031.

E-mail address: amandine.marrel@ifp.fr.

based on two interlinked generalized linear models, and requiring several repetitions of the scalar input experimental design (one for each realization of the stochastic input), thus implying a high number of simulations. Iooss & Ribatet (2009) have recently proposed a joint modeling based on generalized additive models without repetitions. Our work focuses on extending the joint modeling to a more flexible Gp metamodel with a fitting process avoiding the repetitions. Then, this joint modeling is used to perform a sensitivity analysis.

2. Joint modeling with two Gp metamodels and sensitivity analysis

The d scalar inputs are denoted as $X=(X_1, \dots, X_d)$ and supposed independent, the stochastic input is denoted as ε and the simulator as $Y(X, \varepsilon)$. The joint modeling consists in modeling simultaneously the mean component $Y_m(X) = E[Y|X]$ and the dispersion component $Y_d(X) = Var[Y|X]$. To do this, we use two Gp metamodels with a one-degree polynomial trend, a generalized exponential correlation function (Marrel et al., 2008) and a nugget effect to relax the interpolation property. This work proposes a methodology to estimate, in the case without repetitions, the two Gp metamodels, one for the mean and one for the dispersion component. After estimating the first Gp fitting Y_m , several approaches are foreseen to estimate the residuals which are necessary for estimating the Gp fitting Y_d . After comparing the results obtained with the direct residuals on the learning sample and the ones estimated by cross validation, we propose an alternative estimator of the dispersion component based on the cross validation residuals but taking also into account the prediction error of the Gp. The goal is to make out only the dispersion component in the cross validation residuals. A better estimate of the dispersion component is so obtained. Then, estimators for Sobol' indices based on the two estimated Gp are built. All the Sobol' indices for scalar inputs are estimated with the following formulas:

$$S_i \Leftarrow \frac{Var[E[Y_m|X_i]]}{Var[Y]}, S_{ik} \Leftarrow \frac{Var[E[Y_m|X_i, X_k]] - Var[E[Y_m|X_i]] - Var[E[Y_m|X_k]]}{Var[Y]}, \dots \quad (1)$$

Without repetitions, the total Sobol' index is estimated for the stochastic input with the proposed formula, deduced from the decomposition of the variance of $Y(X, \varepsilon)$:

$$S_{T_\varepsilon} = \frac{E_X[Y_d(X)]}{Var[Y]}. \quad (2)$$

On standard analytic test functions, the several approaches to estimate the two Gp are compared in terms of predictivity for the dispersion component and accuracy for the Sobol' indices estimation. A comparison is also made, at equal number of simulations, with a dual modeling built on an experimental design with repetitions. An application on a reservoir simulator will also be presented. In practice, one reservoir simulation can require several hours while all the proposed joint modeling requires few minutes for ten uncertain inputs.

3. References

- Iooss B. and Ribatet M., 2009: Global sensitivity analysis of computer models with functional inputs, *Reliability Engineering & System Safety*, 94, pp 1194-1204.
- Marrel A., Iooss B., Laurent B. and Roustant O., 2009: Calculations of Sobol indices for the Gaussian process metamodel, *Reliability Engineering and System Safety*, 94, pp 742-751.
- Marrel A., Iooss B., Van Dorpe F. and Volkova E., 2008: An efficient methodology for modeling complex computer codes with Gaussian processes, *Computational Statistics and Data Analysis*, 52, pp 4731-4744.
- Oakley JE. and O'Hagan A., 2004: Probabilistic sensitivity analysis of complex models: A Bayesian approach, *Journal of the Royal Statistical Society, Series B*, 66, 751-769
- Sacks J., Welch WJ., Mitchell TJ. and Wynn HP., 1989: Design and analysis of computer experiments, *Statistical Science*, 4, 409-435.
- Saltelli A., Chan K. and Scott E.M. (Eds.), 2000: *Sensitivity Analysis*, Wiley.
- Zabalza I., Dejean J. and Collombier D., 1998: Prediction and density estimation of a horizontal well productivity index using generalized linear models, in *ECMOR VI*, Peebles.