Response surfaces and sensitivity analyses for an environmental model of dose calculations

Submitted to SAMO 2004 Conference

Authors

B. Iooss, N. Devictor and F. Van Dorpe

Bertrand Iooss
CEA Cadarache, DEN/CAD/DER/STR/LCFR, 13108 Saint Paul lez Durance, FRANCE
Phone : 33 4 42 25 72 73 ; Fax : 33 4 42 25 24 08
Email : bertrand.iooss@cea.fr

Nicolas Devictor
CEA Cadarache, DEN/CAD/DER/STR/LCFR, 13108 Saint Paul lez Durance, FRANCE
Phone : 33 4 42 25 30 05 ; Fax : 33 4 42 25 24 08
Email : nicolas.devictor@cea.fr

François Van Dorpe
CEA Cadarache, DEN/CAD/DED/SAMRA/LTCR, 13108 Saint Paul lez Durance, FRANCE
Phone : 33 4 42 25 79 18 ; Fax : 33 4 42 25 62 72
Email : francois.vandorpe@cea.fr

ABSTRACT

The goal of this study is to perform a global sensitivity analysis on a computer code of radiological impact, describing the transfer of radionuclides in the food chains to the man following a chronic gas release of a nuclear installation. An effective dose received by age group can thus be calculated according to the duration of the release. Three age groups and three duration of release have to be considered. Due to the large number of output variables (nine for each radionuclide) and input parameters (more than fifty for each output variable) an efficient methodology have to be developed.

We choose to carry out an approach by Monte-Carlo simulation. A probability distribution has to be given for each input parameter. For simplicity, we do not take into account the possible correlations between input parameters. The computation time of the code is approximately thirty seconds. Therefore, we can generate one thousand different simulations of the code in a reasonable lapse of time (eight hours). This makes it possible to examine the distributions of the output variables. One can in particular note the spreading out of the response, the thickness of the distribution tails, the response skewness compared to the peak of distribution, the multiplicity of the modes. This last point is particularly important: if an output has several modes, it is appropriate to carry out separate sensitivity analyses on the variation interval of each mode. Another important observation is that some output variables give similar results. So we can restrict the number of output variables studied.

The ultimate goal of this study is to quantify the influence of the parameters on the variability of the output variables. First, from Monte-Carlo simulations, we can easily calculate the Pearson and Spearman correlation coefficients between input parameters and output variables. This gives fastly, under the linear relationship hypothesis, a first overview of important factors. Second, by a linear regression analysis, we observe if the model is linear or not. In our case, the model is not linear and linear sensitivity indices like correlation coefficients, standardized regression coefficients or partial correlation coefficients are not correct measures of sensitivity.

Therefore, we decide to calculate Sobol sensitivity indices which require much more simulations. Typically, we have to do ten thousand simulations to obtain an estimate of the Sobol indice for one parameter. As we want a confidence interval on this indice, we make one hundred times this estimation. We have approximately fifty input parameters and eighteen output variables. Therefore the total number of calculations we have to compute is approximately one billion. For computation time reason, the use of our computer code is not possible to calculate these Sobol indices and we decide to simplify the model by the construction of a response surface.

A response surface is a model fitting approximately the results of the code, having good prediction capacity, and requiring a negligible computation time. Classical response surfaces are simple linear models, polynomial forms, spline models, generalized linear or additive models, neural networks, ... We have seen previously that a linear model is not satisfactory, then we search to construct a response surface in a more complex form. Bu using our knowledge of the physical phenomena modeled in the computer code, we suspect that polynomial response surfaces can give good results. We do a least-squares multiple regression using our previous simulations for each output variable, and develop a regression strategy to choose between different possible response surfaces. Statistics on residuals (mean bias, root mean square, coefficient of determination R^2, relative standard deviation, ...), statistics on relative residuals, cross validation technique and bootstrap method are used to validate the response surfaces. The agreement between the computer code and the polynomials are rather good, and we conclude that we can use these response surfaces for the sensitivity analysis. Finally, we calculate the total sensitivity indices of Sobol and their confidence interval by the Monte-Carlo method.

We show the application of this methodology to a specific example defined by the couple (nuclear site, localized population), for two radionuclides, namely iodine 129 and uranium 238. It is thus shown that the most influential parameters for this example are all related to the food chain of the milk, in the order of importance decreasing according to: dose coefficient "effective ingestion", milk ration of the individuals of the reference group, fodder ration of the animals, dry deposit rate of the radionuclide, and transfer factor to the milk.