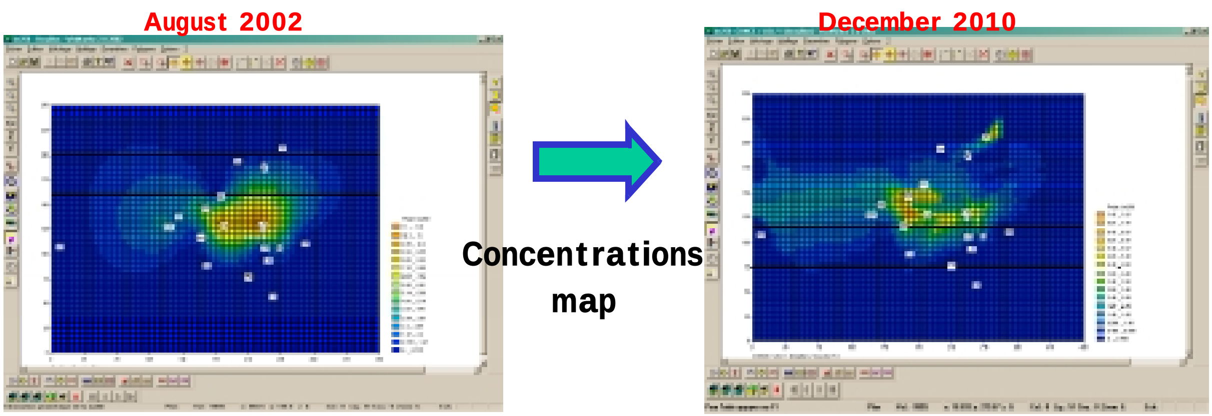
Benchmark « Predictive metamodel construction »
GdR MASCOT-NUM

CEA Cadarache example – MARTHE data

B. Iooss and A. Marrel
April 2008

In 2005, CEA (France) and Kurchatov Institute (Russia) developed a model of strontium 90 migration in a porous water-saturated medium. The scenario concerned the temporary storage of radioactive waste (STDR) in a site close to Moscow. The main purpose was to predict the transport of ^{90}Sr between 2002 and 2010, in order to determine the aquifer contamination. The numerical simulation of the ^{90}Sr transport in the upper aquifer of the site was realized via the MARTHE code (developed by BRGM, France). The figure below illustrates the ^{90}Sr concentration evolution inside the site. The left figure corresponds to the initial concentration field while the right figure corresponds to the calculated concentration field. The small white rectangles correspond to the piezometer locations.



To identify the most influential input parameters of the code on the calculated outputs, uncertainty and sensitivity analysis techniques were used. Due the complexity of the numerical model and its large calculation time, an intermediate metamodel fitting step on a restricted number of simulations has been performed.

20 input scalar parameters of the numerical model were considered as random in the uncertainty analysis. The following table synthesizes their distribution types and their distribution parameter values. The Weibull law which is used is the following one:

$$f(x) = \alpha\beta^{-\alpha} x^{\alpha-1} \exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right)$$

where α is the shape parameter and β is the scale parameter.

	Parameters	Notation	Model value	Distribution type	Interval or distribution parameters
1	Hydraulic conductivity layer 1	per1	8	Uniform	1 – 15
2	Hydraulic conductivity layer 2	per2	15	Uniform	5 – 20
3	Hydraulic conductivity layer 3	per3	8	Uniform	1 - 15
4	Hydraulic conductivity zone 1	perz1	8	Uniform	1 - 15
5	Hydraulic conductivity zone 2	perz2	8	Uniform	1 - 15
6	Hydraulic conductivity zone 3	perz3	8	Uniform	1 - 15
7	Hydraulic conductivity zone 4	perz4	8	Uniform	1 - 15
8	Longitudinal dispersivity layer 1	d1	0,8	Uniform	0,05 - 2
9	Longitudinal dispersivity layer 2	d2	0,8	Uniform	0,05 - 2
10	Longitudinal dispersivity layer 3	d3	0,8	Uniform	0,05 - 2
11	Transversal dispersivity layer 1	dt1	0,08	Uniform	0,01*d1 - 0,1*d1
12	Transversal dispersivity layer 2	dt2	0,08	Uniform	0,01*d2 - 0,1*d2
13	Transversal dispersivity layer 3	dt3	0,08	Uniform	0,01*d3 - 0,1*d3
14	Volumetric distribution coefficient l. 1	kd1	5,1	Weibull	1.1597, 19.9875
15	Volumetric distribution coefficient l. 2	kd2	0,34	Weibull	0.891597, 24.4455
16	Volumetric distribution coefficient l. 3	kd3	5,1	Weibull	1.27363, 22.4986
17	Porosity	poros	0,3	Uniform	0,3 - 0,37
18	Infiltration type 1	i1	0,0001	Uniform	0 - 0,0001
19	Infiltration type 2	i2	0,004	Uniform	i1 - 0,01
20	Infiltration type 3	i3	0,02	Uniform	i2 - 0,1

20 output variables were considered. They correspond to the concentrations calculated in the piezometer locations. The CPU time cost of the MARTHE model allowed us to make only 300 simulations. To compute variance-based sensitivity indices (called Sobol indices), a metamodel has been fitted for each output variable (Volkova et al., 2008).

As a conclusion of our first study, it turns out that some outputs have no physical interest or no statistical interest. We thus restrict the problem to the following outputs study: p102K, p104, p106, p2-76, p29K, p31K, p35K, p37K, p38, p4b

Benchmark objectives

The benchmark goal is to to build the most predictive metamodels for all these output variables, in all the variation range of the inputs. It is possible to build one metamodel per output variable, but also to build multiple metamodels (modelling several outputs).

The computer code is not accessible any more. Then, we restrict the study to the use of the 300 simulations. For the validation, we suggest to use a cross-validation technique. For each validation step, users will have to take a tests basis containing no more than 50 points.

We suggest the use of the following interesting validation criteria:

- The predictivity coefficient (called Q_2), corresponding to R^2 computed on a test basis;
- The mean absolute error (mean of the absolute values of the test basis residuals);
- The bias of the test basis residuals;
- The maximum of the test basis residuals;
- The distribution of the test basis residuals;
- More outliers-robust criteria (useful in case of a few extreme residuals): for example the geometrical bias $\mathrm{MG} = \exp\left[\mathrm{E}(\ln X) - \mathrm{E}(\ln Y)\right]$ and the geometrical variance $\mathrm{VG} = \exp\left\{\mathrm{E}\left[(\ln X - \ln Y)^2\right]\right\}$.

Other validation criteria can be proposed by benchmark participants.

Nota

This exercise has no judgment value on the global quality of the metamodels. However, it is a good representation of the industrial practice : computer experiments have not been performed in a particular metamodel fitting objective (here the first objective was to propagate uncertainties), the chosen design is far from an optimal one and the computer code is no more available

All the scenario details can be found in Volkova et al. (2008). A particular metamodel fitting methodology is given in Marrel et al. (2008). Please, contact B. Iooss (bertrand.iooss[at]cea.fr) for additional information.

Results

Each participant is invited to send its results to: bertrand.iooss[at]cea.fr

Acknowledgments

We are grateful to E. Volkova (Institut Kurchatov) who gives the authorization to use the MARTHE data.

Références

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

E. Volkova, B. Iooss and F. Van Dorpe, Global sensitivity analysis for a numerical model of radionuclide migration from the “RRC” Kurchatov Institute radwaste disposal site, *Stochastic Environmental Research and Risk Assessment*,22:17-31 , 2008