Benchmark :

CONSTRUCTION OF PREDICTIVE METAMODELS

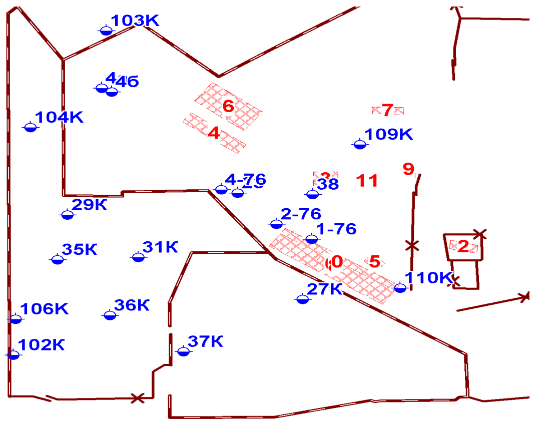
MARTHE data example

B. **Iooss**[◊] and A. Marrel[*]

[◊] CEA/DEN/CAD/DER/SESI/LCFR
[*] CEA/DEN/CAD/DTN/SMTM/LMTE

Data presentation (1)

- <u>STDR of Kurchatov (Moscow):</u>
  - Radioactive waste storage from 1943 to 1974
  - Sire characterization in 1990 : 20 piezometers network
  - Upper groundwater contamination in **^{90}Sr**



- <u>Objective :</u>

  Estimation of the environmental impact

- <u>Modelisation :</u>
  - Development of a generalized transport scenario of ^{90}Sr on this site between 2002 and 2010, with **MARTHE** software
  - Identification of model parameters

Data presentation (2)

Problem :

- Complex and time consuming computer code (~ 30mn /evaluation)
- Large number of input variables (> 10)

⇒ **Code non directly usable for fine sensitivity analysis**

Solution : Approximate the computer code by a simpler statistical function called **Metamodel**

⇒ **Approximation quality and very low cpu cost**

Objectives : Determination of the most influent input variables

Metamodel construction + Sensitivity analysis

Data presentation (3)

Distribution of the 20 model input parameters:

➞ nominal value, distribution type and et distribution interval/parameters

	Paramètres	Indicateur	Valeur du modèle	Type de distribution	Intervalle ou paramètres de distribution
1	Perméabilité couche 1	per1	8	Uniforme	1 - 15
2	Perméabilité couche 2	per2	15	Uniforme	5 - 20
3	Perméabilité couche 3	per3	8	Uniforme	1 - 15
4	Perméabilité zone 1	perz1	8	Uniforme	1 - 15
5	Perméabilité zone 2	perz2	8	Uniforme	1 - 15
6	Perméabilité zone 3	perz3	8	Uniforme	1 - 15
7	Perméabilité zone 4	perz4	8	Uniforme	1 - 15
8	Dispersivité longitudinale couche 1	d1	0,8	Uniforme	0,05 - 2
9	Dispersivité longitudinale couche 2	d2	0,8	Uniforme	0,05 - 2
10	Dispersivité longitudinale couche 3	d3	0,8	Uniforme	0,05 - 2
11	Dispersivité transversale couche 1	dt1	0,08	Uniforme	0,01*d1 - 0,1*d1
12	Dispersivité transversale couche 2	dt2	0,08	Uniforme	0,01*d2 - 0,1*d2
13	Dispersivité transversale couche 3	dt3	0,08	Uniforme	0,01*d3 - 0,1*d3
14	Coefficient de partage volumique c. 1	kd1	5,1	Weibull	1.1597, 19.9875
15	Coefficient de partage volumique c. 2	kd2	0,34	Weibull	0.891597, 24.4455
16	Coefficient de partage volumique c. 3	kd3	5,1	Weibull	1.27363, 22.4986
17	Porosité tous les couches	poros	0,3	Uniforme	0,3 - 0,37
18	Infiltration type 1	i1	0,0001	Uniforme	0 - 0,0001
19	Infiltration type 2	i2	0,004	Uniforme	i1 - 0,01
20	Infiltration type 3	i3	0,02	Uniforme	i2 - 0,1

Data presentation (4)

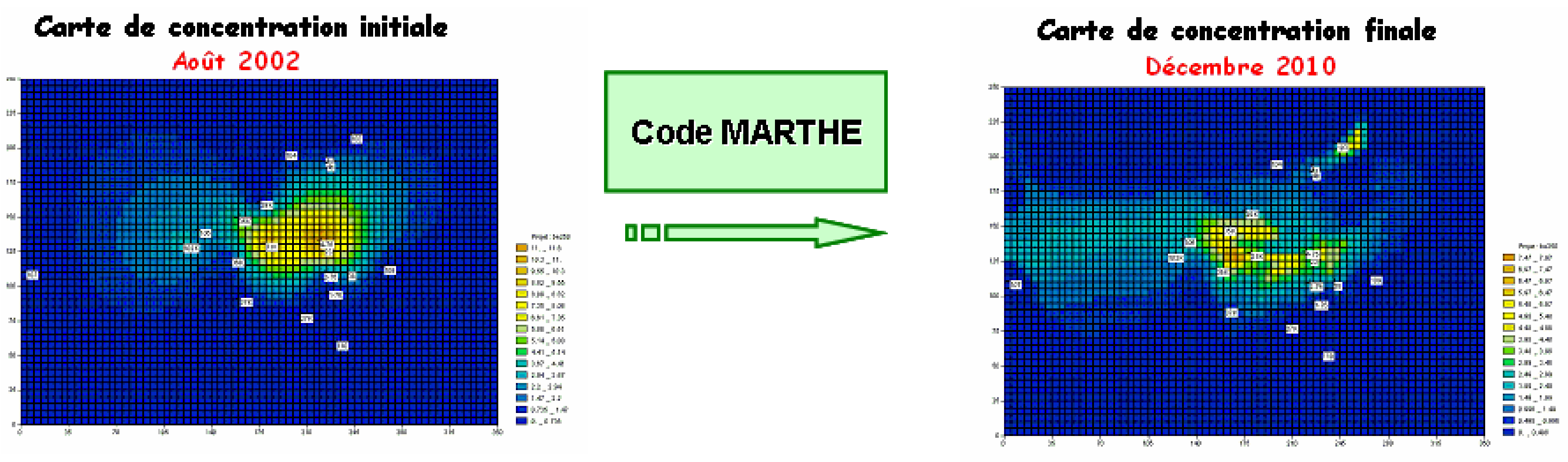
- Jeu de données MARTHE :
  - ➢ **20 input parameters :**

    permeability, dispersivities, Kd, porosity, infiltration intensities, …
  - ➢ **10 output variables :**

    Concentration in 10 piezometers

Simulation des 20 variables d'entrée (LHS)

300 LHS simulations



Benchmark objectives

- **Predictive metamodel construction**
  - Metamodel for each output variable
  - One multi-output metamodel
- **Validation**
  - Cross validation on a test basis of 1 to 50 points max.
- **Validation criteria**
  - Predictivity coefficient (called Q_2, which is a R^2 calculated ona test basis)

$$Q_2(Y,\hat{Y}) = 1 - \frac{\sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{N}\left(\overline{Y} - Y_i\right)^2}$$

  - MSE, MAE
  - Bias, maximum, distribution of test basis residuals
  - More robust (with respect to outliers) criteria

    ex : geometrical bias or geometrical variance

Bibliography

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

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