

DE LA RECHERCHE À L'INDUSTRIE



**ESTIMATION
OF THE DERIVATIVE-BASED
GLOBAL SENSITIVITY MEASURES
FROM A GAUSSIAN PROCESS METAMODEL**

Matthias De Lozzo Amandine Marrel

CEA Cadarache

Journée de Statistique de la SFdS

- Lille -

Tuesday 2nd June 2015

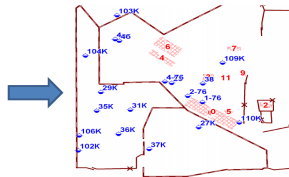
www.cea.fr



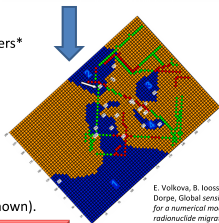
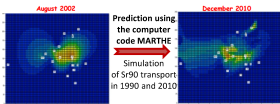
The MARTHE case study



Temporary storage site for radioactive wastes in Kurchatov, near Moscow



- Storage from 1943 to 1974
- Site characterization in 1990 : network of 20 piezometers*
- Strontium-90 contamination of the superior aquifer



20 Marthe input parameters are uncertain (but laws are known).

GOAL : identify the more influential input parameters for the prediction of the concentration for a given piezometer.

* A piezometer is a porous probe enabling to extract deep water and measure the water level.

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

We consider the computer code

$$y = f(x_1, \dots, x_d) = f(x).$$

- Output: $y \in \mathcal{Y} \subset \mathbb{R}$.
- Inputs: $x_1 \in \mathcal{X}_1 \subset \mathbb{R}, \dots, x_d \in \mathcal{X}_d \subset \mathbb{R}$.

Input uncertainties & probabilistic modelization

- x_k is a realization of the r.v. X_k with known law \mathcal{L}_k .
- X_1, \dots, X_d are independent r.v.

\Rightarrow y is a realization of the r.v. Y with unknown law.

★ **Contribution of each X_k to the uncertainty over Y ?** ★

Tool: The dataset $\mathcal{D} = (\mathbf{x}, \mathbf{y}) = (x^{(i)}, f(x^{(i)}))_{1 \leq i \leq n}$ where $(x^{(i)})_{1 \leq i \leq n}$ is a space-filling design or a n -sample.




- 1 Introduction and motivations
- 2 Sensitivity analysis and DGSM
- 3 DGSM estimation from a Gaussian process metamodel
- 4 Significance test for screening
- 5 Numerical implementation
- 6 Numerical illustrations
- 7 Application to a hydrogeologic transport model
- 8 Conclusions & perspectives
- 9 Références



- 1 Introduction and motivations
- 2 Sensitivity analysis and DGSM
- 3 DGSM estimation from a Gaussian process metamodel
- 4 Significance test for screening
- 5 Numerical implementation
- 6 Numerical illustrations
- 7 Application to a hydrogeologic transport model
- 8 Conclusions & perspectives
- 9 Références

If $\mathbb{E}[f^2(X)] < +\infty$, orthogonal decomposition:


$$\begin{aligned} f(X) &= f_{\emptyset} + \sum_{i=1}^d f_j(X_j) + \sum_{i=1}^d \sum_{i < j}^d f_{ij}(X_{ij}) + \dots + f_{1\dots d}(X_{1\dots d}) \\ &= \sum_{u \subset \{1, \dots, d\}} f_u(X_u) \end{aligned}$$

where $f_{\emptyset} = \mathbb{E}[f(X)]$ and $f_u(X_u) = \mathbb{E}[f(X)|X_u] - \sum_{v \subset u} f_v(X_v)$.

Sobol' indices associated to X_u :

- 1st order index: $S_u = \frac{\text{Var}[f_u(X_u)]}{\text{Var}[f(X)]}$
- total index: $S_u^T = \sum_{v \supset u} S_v$

Properties: $S_u \in [0, 1]$ and $\sum_{u \subset \{1, \dots, d\}} S_u = 1$

Definition (Sobol and Gresham 1995), $\forall k \in \{1, \dots, d\}$

$$D_k = \mathbb{E}_X \left[\left(\frac{\partial f(X)}{\partial x_k} \right)^2 \right]$$

Relation to the Morris method (Kucherenko et al. 2009)

Let $d_k^{(r)} = \frac{f(X_1^{(r)}, \dots, X_{k-1}^{(r)}, X_k^{(r)} + \Delta, X_{k+1}^{(r)}, \dots, X_d^{(r)}) - f(X^{(r)})}{\Delta}$, $r \in \{1, \dots, R\}$.

1. Morris : $m_k = \frac{1}{R} \sum_{r=1}^R d_k^{(r)}$ et $\sigma_k^2 = \frac{1}{R-1} \sum_{r=1}^R (d_k^{(r)} - m_k)^2$

2. DGSM : $M_k^2 + \sigma_k^2 \rightarrow D_k$ when $\Delta \rightarrow 0$

2nd order index (Roustant et al. 2014)

$$D_{ij} = \mathbb{E}_X \left[\left(\frac{\partial^2 f(X)}{\partial x_i \partial x_j} \right)^2 \right], \forall i, j \in \{1, \dots, d\}$$



DGSMs for bounding Sobol' indices

- For $X_k \sim \mathcal{U}(a_k, b_k)$ (Sobol and Kucherenko 2009)

$$S_k^T \leq \frac{(b_k - a_k)^2}{\pi^2 \text{Var}[f(X)]} D_k$$

- For $X_k \sim \mathcal{N}(a_k, b_k^2)$ (Sobol and Kucherenko 2009)

$$S_k^T \leq \frac{b_k^2}{\pi^2 \text{Var}[f(X)]} D_k$$

- For Boltzmann probability measure¹

$$S_k^T \leq \frac{4C_k^2}{\text{Var}[f(X)]} D_k$$

where $C_k = \sup_{x \in \mathbb{R}} \min\{\mathbb{P}(X_k > x), \mathbb{P}(X_k < x)\} / \mu_{X_k}(x)$ is the Cheeger constant (Lamboni et al. 2013).

¹ Absolutely continuous measure w.r.t the Lebesgue measure, with density function $d\mu(x) = c \exp(-v(x))dx$ where v is a continuous function and c a normalization constant.



- Lower bounds for Sobol' indices (Kucherenko and Song 2014, submitted)
- Survey of recent advances in lower and upper bounds: (Kucherenko and Iooss 2014)
- Applications:
 - aquatic prey-predator chain (Iooss et al. 2012),
 - biological system model (Kiparissides et al. 2009),
 - flood model (Lamboni et al. 2013),
 - reservoir simulator (Touzani and Busby 2014),
 - ...

DGSM estimation and computational barriers

For the estimation of $D_k = \mathbb{E}_X \left[\left(\frac{\partial f(X)}{\partial x_k} \right)^2 \right]$, $k \in \{1, \dots, d\}$:

- Important computational cost for $\nabla_x f(X)$.
- Computer code without adjoint code.

Proposition

- 1 Build a metamodel \hat{f} of f from the dataset \mathcal{D} .
- 2 Build a M -sample $\mathcal{S} = \left(x^{(i)}, \nabla_x \hat{f}(x^{(i)}) \right)_{1 \leq i \leq M}$.
- 3 Build an estimator $\hat{D}_k(\mathcal{S}) = \mathbb{E}_X \left[\left(\frac{\partial \hat{f}(X)}{\partial x_k} \right)^2 \right]$.

Extra: Supply a measure of the error $\left(\hat{D}_k(\mathcal{S}) - D_k \right)^2$.



- 1 Introduction and motivations
- 2 Sensitivity analysis and DGSM
- 3 DGSM estimation from a Gaussian process metamodel**
- 4 Significance test for screening
- 5 Numerical implementation
- 6 Numerical illustrations
- 7 Application to a hydrogeologic transport model
- 8 Conclusions & perspectives
- 9 Références

Assumption: f is a realization of the Gaussian process

$$\mathbf{Z}(\mathbf{x}; \omega) \sim \mathcal{GP} \left(\boldsymbol{\beta}, \sigma^2 \left(\mathbf{r}(\mathbf{x}, \mathbf{x}') + \tau^2 \delta_{\mathbf{x}\mathbf{x}'} \right) \right)$$

where $r(x, x') = \prod_{k=1}^d r_k(x_k, x'_k)$ is a kernel function parametrized by $\theta \in \Theta$ and $k(x) = (r(x, x^{(1)}), \dots, r(x, x^{(n)}))^2$.

The **Gaussian process metamodel** takes the form

$$Z_C(x; \omega) = \left[Z(x; \omega) | \mathcal{D}, \boldsymbol{\beta}, \sigma^2, \tau^2, \theta \right] \sim \mathcal{GP}(\hat{f}(x), s^2(x, x')).$$

with $\boldsymbol{\gamma} = R^{-1}(\mathbf{Y} - \boldsymbol{\beta}\mathbf{1})$ and $R = (r(x^{(i)}, x^{(j)}) + \tau^2 \delta_{ij})_{1 \leq i, j \leq n}$.

- $\hat{f}(x) = \boldsymbol{\beta} + k(x)^T \boldsymbol{\gamma}$ is the **predictor** of f ,
- $s^2(x, x') = \sigma^2 (r(x, x') - k(x)^T R^{-1} k(x'))$,
- $s^2(x) := s^2(x, x)$ is the **quadratic error** of $\hat{f}(x)$.



Plug-in estimator

$$\hat{D}_k^{(1)} = \mathbb{E}_X \left[\left(\frac{\partial \hat{f}(X)}{\partial x_k} \right)^2 \right] = \gamma^T \mathbb{E}_X \left[\frac{\partial k(X)}{\partial x_k} \frac{\partial k(X)^T}{\partial x_k} \right] \gamma$$

Stochastic estimator

$$\widehat{RD}_k^{(2)}(\omega) = \mathbb{E}_X \left[\left(\frac{\partial Z_C(X; \omega)}{\partial x_k} \right)^2 \right]$$

Full-GPM estimator

$$\hat{D}_k^{(2)} = \mathbb{E}_\omega \left[\widehat{RD}_k^{(2)}(\omega) \right] = \hat{D}_k^{(1)} + \mathbb{E}_X \left[\mathbb{V}_\omega \left[\frac{\partial Z_C(X; \omega)}{\partial x_k} \right] \right]$$

Extra from $\widehat{RD}_k^{(2)}(\omega)$

- confidence intervals
- significance tests





- 1 Introduction and motivations
- 2 Sensitivity analysis and DGSM
- 3 DGSM estimation from a Gaussian process metamodel
- 4 Significance test for screening**
- 5 Numerical implementation
- 6 Numerical illustrations
- 7 Application to a hydrogeologic transport model
- 8 Conclusions & perspectives
- 9 Références

Significance test for screening

Test: " $\mathcal{H}_0: D_k = 0$ " against " $\mathcal{H}_1: D_k \neq 0$ ".

Under \mathcal{H}_0 , we have $\mathbb{E}_X \left[\frac{\partial Z_C(\mathbf{X})}{\partial x_k} \right] = m_M = 0_{\mathbb{R}^M}$.

p -value of the test applied to $\hat{D}_k^{(2)} = \mathbb{E}_\omega \left[\widehat{RD}_k^{(2)}(\omega) \right]$

$$p_{\text{val},k} = \mathbb{P} \left[\widehat{RD}_k^{(2)}(\omega) > \hat{D}_k^{(2)} \mid \mathcal{H}_0 \right]$$

Use for screening with $\alpha \in (0, 1)$

- Designate as non-influential the input parameters with p -values lower than α .
- Sort the input parameters with their p -values.





- 1 Introduction and motivations
- 2 Sensitivity analysis and DGSM
- 3 DGSM estimation from a Gaussian process metamodel
- 4 Significance test for screening
- 5 Numerical implementation
- 6 Numerical illustrations
- 7 Application to a hydrogeologic transport model
- 8 Conclusions & perspectives
- 9 Références

Elementary decomposition of the estimators

1. For $\hat{D}_k^{(1)} = \gamma^T \mathbb{E}_X \left[\frac{\partial k(X)}{\partial x_k} \frac{\partial k(X)}{\partial x_k}^T \right] \gamma = \gamma^T A^{[k]} \gamma$, we have:

$$A^{[k]} = B^{[k]} \bigodot_{\substack{l=1 \\ l \neq k}}^d C^{[l]} \text{ where}$$

- $B_{ij}^{[k]} = \mathbb{E}_{X_k} \left[\frac{\partial r_k(X_k, X_k^{(i)})}{\partial x_k} \frac{\partial r_k(X_k, X_k^{(j)})}{\partial x_k} \right], \forall i, j \in \{1, \dots, n\}$
- $C_{ij}^{[l]} = \mathbb{E}_{X_k} \left[r_k(X_k, X_k^{(i)}) r_k(X_k, X_k^{(j)}) \right], \forall i, j \in \{1, \dots, n\}$

2. For $\hat{D}_k^{(2)} = \hat{D}_k^{(1)} + \mathbb{E}_X \left[\text{Var}_\omega \left[\frac{\partial Z_C(X; \omega)}{\partial x_k} \right] \right]$, we have:

$$\mathbb{E}_X [\dots] = \sigma^2 \left\{ \mathbb{E}_X \left[\frac{\partial^2 r(U, V)}{\partial u_k \partial v_k} \Big|_{(X, X)} \right] - \mathbf{1}^T \left(R^{-1} \bigodot A^{[k]} \right) \mathbf{1} \right\}$$



Uniform laws and Gaussian kernels

Particular case: $r_k(x_k, x'_k) = \exp\left(-\frac{(x_k - x'_k)^2}{2\theta_k^2}\right)$ and $X_k \sim \mathcal{U}([m_k, M_k])$, $k \in \{1, \dots, d\}$

- With $W_{i,j}^{[k]} \sim \mathcal{N}\left(\frac{x_k^{(i)} + x_k^{(j)}}{2}, \frac{\theta_k^2}{2}\right)$, we have:

$$C_{ij}^{[k]} = \sqrt{\pi} \frac{\mathbb{P}[W_{i,j}^{[k]} \in [m_k, M_k]]}{M_k - m_k} \theta_k e^{-\frac{(x_k^{(i)} - x_k^{(j)})^2}{4\theta_k^2}}$$

- With $Z_{i,j}^{[k]} \sim \mathcal{N}(0, 1)$, $a_{i,j}^{[k]} = \frac{2m_k - x_k^{(i)} - x_k^{(j)}}{\sqrt{2}\theta_k}$, $b_{i,j}^{[k]} = \frac{2M_k - x_k^{(i)} - x_k^{(j)}}{\sqrt{2}\theta_k}$ and $\varphi(\cdot)$ the p.d.f. of $\mathcal{N}(0, 1)$, we have:

$$B_{ij}^{[k]} = -\frac{(x_k^{(i)} - x_k^{(j)})^2}{4\theta_k^2} C_{ij}^{[k]} + \frac{\sqrt{\pi}}{2\theta_k(M_k - m_k)} \left\{ a_{i,j}^{[k]} \varphi(a_{i,j}^{[k]}) - b_{i,j}^{[k]} \varphi(b_{i,j}^{[k]}) + \mathbb{P}[Z_{i,j}^{[k]} \in [a_{i,j}^{[k]}, b_{i,j}^{[k]}]] \right\} e^{-\frac{(x_k^{(i)} - x_k^{(j)})^2}{4\theta_k^2}}$$

- $\mathbb{E}_X \left[\frac{\partial^2 r(U, V; \theta)}{\partial u_k \partial v_k} \Big|_{(U, V) = (X, X)} \right] = \frac{1}{\theta_k^2}$.



Monte-Carlo approximation

$$\hat{D}_k^{(1)} = \mathbb{E}_X \left[\left(\frac{\partial \hat{f}(X)}{\partial x_k} \right)^2 \right] \text{ et } \hat{D}_k^{(2)} = \mathbb{E}_\omega \left[\mathbb{E}_X \left[\left(\frac{\partial Z_C(X; \omega)}{\partial x_k} \right)^2 \right] \right]$$

For the plug-in estimator

$$\text{Let } \mathcal{S} = \left(x^{(i)}, \nabla_x \hat{f}(x^{(i)}) \right)_{i=1}^M. \text{ Then } \hat{D}_k^{(1)} \approx \frac{1}{M} \sum_{i=1}^M \left(\frac{\partial \hat{f}(x^{(i)})}{\partial x_k} \right)^2$$

For the full-GPM estimator

$$\text{Let } \mathcal{S} = \left(x^{(i)}, \nabla_x \hat{f}(x^{(i)}) \right)_{i=1}^M. \text{ Then } \star \text{ what do we do? } \star$$

\mathbb{E}_X is approximated from \mathcal{S} but \mathbb{E}_ω ?

Idea: simulate N realizations of $(x^{(i)}, \nabla_x Z_C(x^{(i)}; \omega))_{i=1}^M$
and then compute $\hat{D}_k^{(2)} \approx \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \left(\frac{\partial Z_C(x^{(i)}; \omega^{(j)})}{\partial x_k} \right)^2$

Propagative Gibbs sampler²

Objective

Fast simulation of $\frac{\partial Z_C(\mathbf{X};\omega)}{\partial x_k} := \left(\frac{\partial Z_C(X^{(l)};\omega)}{\partial x_k} \right)_{l=1}^M \sim \mathcal{N}(b_M, C_M)$

when M is important.

Method

Use of a propagative version of the Gibbs sampler which does not require inversion of the matrix C_M .

Idea: apply the Gibbs sampler to the vector $W := C_M^{-1}(Y - b_M)$, then compute $Y := b_M + C_M W$. This requires C_M , the inverse of C_M^{-1} , which is available.

²L. Lantuéjoul and N. Desassis. "Simulation of a Gaussian random vector: A Propagative version of the Gibbs sampler". In: *Ninth International Geostatistics Congress. 2012*.



Approximation using a chi-squared trick³

We approximate $\widehat{RD}_k^{(2)}(\omega) = \mathbb{E}_X \left[\left(\frac{\partial Z_C(X; \omega)}{\partial x_k} \right)^2 \right]$ by:

$$\widehat{RD}_k^{(2),M}(\omega) = \frac{1}{M} \sum_{i=1}^M \left(\frac{\partial Z_C(X^{(i)}; \omega)}{\partial x_k} \right)^2 = \left(\frac{\partial Z_C(\mathbf{X}; \omega)}{\partial x_k} \right)^T A \frac{\partial Z_C(\mathbf{X}; \omega)}{\partial x_k}$$

where $A = M I_M \in \mathcal{M}_M(\mathbb{R})$.

Method

Approximate the distribution of $\widehat{RD}_k^{(2),M}(\omega)$ using a noncentral chi-squared distribution $\chi_l^2(\delta)$, $(l, \delta) \in \mathbb{N}^* \times \mathbb{R}_+$.

- No matrix inversion or spectral decomposition
- Estimation of l and δ in order to:
 - equalize the skewnesses of $\widehat{D}_k^{(2),M}(\omega)$ and $\chi_l^2(\delta)$.
 - minimize the difference between their kurtoses.

³H. Liu, Tang Y., and H. H. Zhang. "A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables". In: *Computational Statistics & Data Analysis* 53.4 (2009), pp. 853–856.





- 1 Introduction and motivations
- 2 Sensitivity analysis and DGSM
- 3 DGSM estimation from a Gaussian process metamodel
- 4 Significance test for screening
- 5 Numerical implementation
- 6 Numerical illustrations**
- 7 Application to a hydrogeologic transport model
- 8 Conclusions & perspectives
- 9 Références

Ishigami function

We consider the model

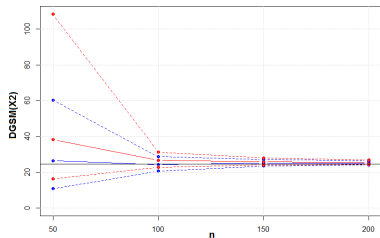
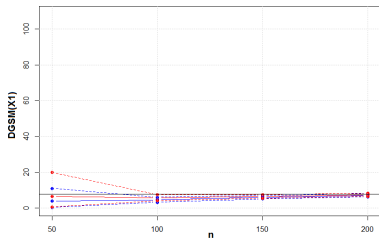
$$f(X) = \sin(X_1) + 7 \sin^2(X_2) + 0.1 X_3^4 \sin(X_1)$$

where X_1, X_2, X_3 are i.i.d. $\mathcal{U}([-\pi, \pi])$.

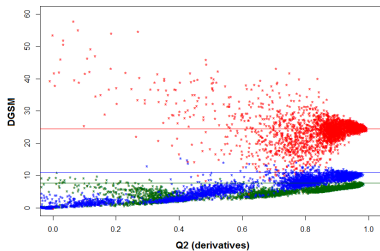
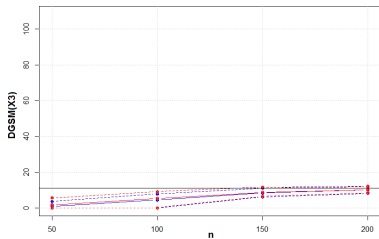
	X_1	X_2	X_3
1st order Sobol	0.31	0.44	0
Total Sobol	0.56	0.44	0.24
DGSM	7.7	24.5	11.0



Estimator convergence rates



95% IC (- -) and Mean (-) ; **plug-in** and **full-GPM**



Monte-Carlo method comparison

- 1 analytical results for uniform laws and Gaussian kernels
- 2 Monte-Carlo approximation based on a SVD
- 3 propagative version of the Gibbs sampler
- 4 approximation with the chi-squared trick

Method	$N = 500$				$N = 1'00$			
	1	2	3	4	1	2	3	4
$n = 50$	0.01	5.9	0.9	2.1	0.01	6.8	1.3	2.0
$n = 100$	0.04	6.1	1.1	2.2	0.03	6.6	1.4	2.0
$n = 200$	0.13	6.3	1.3	2.5	0.13	6.8	1.7	2.3
Method	$N = 5'000$				$N = 10'000$			
	1	2	3	4	1	2	3	4
$n = 50$	0.01	14.4	4.6	1.9	0.01	24.1	8.7	1.9
$n = 100$	0.03	14.5	4.7	2.0	0.03	24.2	8.7	2.0
$n = 200$	0.13	14.8	5.0	2.3	0.13	24.5	9.0	2.3

Table: Computational time (in seconds) for different dataset sizes n and different numbers N of ω realizations.

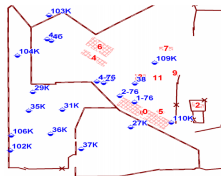


- 1 Introduction and motivations
- 2 Sensitivity analysis and DGSM
- 3 DGSM estimation from a Gaussian process metamodel
- 4 Significance test for screening
- 5 Numerical implementation
- 6 Numerical illustrations
- 7 Application to a hydrogeologic transport model**
- 8 Conclusions & perspectives
- 9 Références

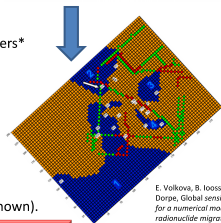
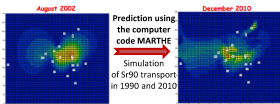
Reminder on the MARTHE model



Temporary storage site for radioactive wastes in Kurchatov, near Moscow



- Storage from 1943 to 1974
- Site characterization in 1990 : network of 20 piezometers*
- Strontium-90 contamination of the superior aquifer



20 Marthe input parameters are uncertain (but laws are known).

GOAL : identify the more influential input parameters for the prediction of the concentration for a given piezometer.

* A piezometer is a porous probe enabling to extract deep water and measure the water level.

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

Input parameters of the MARTHE model



	Input parameters	Notation	Distribution type	Distribution parameters
1	Hydraulic conductivity layer 1	per1	Uniform	1, 15
2	Hydraulic conductivity layer 2	per2	Uniform	5, 0
3	Hydraulic conductivity layer 3	per3	Uniform	1, 15
4	Hydraulic conductivity zone 1	perz1	Uniform	1, 15
5	Hydraulic conductivity zone 2	perz2	Uniform	1, 15
6	Hydraulic conductivity zone 3	perz3	Uniform	1, 15
7	Hydraulic conductivity zone 4	perz4	Uniform	1, 15
8	Longitudinal dispersivity layer 1	d1	Uniform	0.05, 2
9	Longitudinal dispersivity layer 2	d2	Uniform	0.05, 2
10	Longitudinal dispersivity layer 3	d3	Uniform	0.05, 2
11	Ratio of transversal to longitudinal dispersivity layer 1	dt1/d1	Uniform	0.01, 0.1
12	Ratio of transversal to longitudinal dispersivity layer 2	dt2/d2	Uniform	0.01, 0.1
13	Ratio of transversal to longitudinal dispersivity layer 3	dt3/d3	Uniform	0.01, 0.1
14	Volumetric distribution coefficient I.1	kd1	Weibull	1.1597, 19.9875
15	Volumetric distribution coefficient I.2	kd2	Weibull	0.891597, 24.4455
16	Volumetric distribution coefficient I.3	kd3	Weibull	1.27363, 22.4986
17	Porosity	poros	Uniform	0.3, 0.37
18	Infiltration type 1	i1	Uniform	0, 0.0001
19	Infiltration type 2 - type 1	$\delta i2$	Uniform	0, 0.0092
20	Infiltration type 3 - type 2	$\delta i3$	Uniform	0, 0.096

Goal: define a DGSM whose physical dimension does not depend on the input parameter ones.

Definition

$$\hat{\Delta}_k(\omega) = \sigma_k^2 \mathbb{E}_X \left[\left(\frac{\partial Z_C(X; \omega)}{\partial x_k} \right)^2 \right], \quad k \in \{1, \dots, d\}$$

Mean variance of a linear estimator

Decomposition of the mean variance of a one-order Taylor series expansion for the function $f(X)$:

- 1 $f(x) \approx f^*(x) = f(x^*) + \sum_{k=1}^d \frac{\partial f(x)}{\partial x_k} \Big|_{x^*} (x_k - x_k^*)$
- 2 $\text{Var} [f^*(X)] = \sum_{k=1}^d \left(\frac{\partial f(x)}{\partial x_k} \Big|_{x^*} \right)^2 \sigma_k^2$
- 3 $\mathbb{E}_{X^*} [\text{Var} [f^*(X)]] = \sum_{k=1}^d \sigma_k^2 D_k = \sum_{k=1}^d \Delta_k$



Results for the MARTHE model



X_k	Stochastic DGSM $\widehat{RD}_k^{(2)}(\omega)$		Weighted DGSM		Sobol 1st order
	Mean $\hat{D}_k^{(2)}$	Screening using $p_{val,k}$	Mean $\hat{\Delta}_k$	Mean $\hat{\Delta}_k$	
kd1	2.2e-4	0.000 ***	5e-2	88%	0.75
$\delta i3$	5.6e+0	0.000 ***	4e-3	7%	0.16
perz2	5.8e-5	0.000 ***	9e-4	2%	0.03
d1	2.5e-4	0.006 **	8e-5	0%	0.03
kd2	4.2e-8	0.113	4e-5	0%	0.03
per1	7.2e-5	0.000 ***	1e-3	2%	0.06
perz1	5.1e-6	0.001 **	9e-5	0%	0.03
dt1/d1	1.2e-2	0.001 **	8e-5	0%	0.03
per3	1.0e-5	0.000 ***	2e-4	0%	0.03
d3	2.7e-4	0.000 ***	8e-5	0%	0.03
per2	3.7e-6	0.022 *	7e-5	0%	0.03
perz4	4.0e-6	0.040 *	7e-5	0%	0.03
perz3	5.3e-6	0.000 ***	8e-5	0%	0.03
i1	8.0e+4	0.040 *	7e-5	0%	0.03
dt2/d2	1.1e-1	0.011 *	7e-5	0%	0.03
$\delta i2$	1.0e+1	0.000 ***	9e-5	0%	0.03
kd3	1.1e-7	0.215	3e-5	0%	0.03
d2	2.2e-4	0.037 *	7e-5	0%	0.03
dt3/d3	1.0e-1	0.034 *	7e-5	0%	0.03
poros	1.7e-1	0.028 *	7e-5	0%	0.03



- 1 Introduction and motivations
- 2 Sensitivity analysis and DGSM
- 3 DGSM estimation from a Gaussian process metamodel
- 4 Significance test for screening
- 5 Numerical implementation
- 6 Numerical illustrations
- 7 Application to a hydrogeologic transport model
- 8 Conclusions & perspectives
- 9 Références

Conclusions

- DGSM estimation without adjoint code.
- Surrogate modeling with a Gaussian process metamodel.
- Plug-in and full-GPM estimators.
- Confidence intervals and significance tests using the full-GPM estimator.
- Analytical results for uniform laws and Gaussian kernels.
- Monte-Carlo methods otherwise (proagative Gibbs sampler or chi-squared trick).
- Numerical tests:
 - Convergence rates function of the input parameter.
 - MC method comparison for different discretizations.
- Weighted DSGMs.
- Application to an hygrogeological transport model.





- Adaptive sampling for the gradient approximation enhancement.
→ New input vector $x^{(n+1)}$ maximizing the trace of the GPM gradient covariance or its reduction
- Integration of some computer code gradient evaluations.
- Combination with lower and upper bounds for Sobol' indices.

Références

- I. Sobol and A. Gresham. "On an alternative global sensitivity estimators". In: *Proceedings of SAMO 1995*. Belgirate, 1995, pp. 40–42
- A. Kiparissides et al. "Global Sensitivity Analysis Challenges in Biological Systems Modeling". In: *Industrial & Engineering Chemistry Research* 48.15 (2009), pp. 7168–7180
- S. Kucherenko et al. "Monte Carlo evaluation of derivative-based global sensitivity measures". In: *Reliability Engineering & System Safety* 94.7 (2009). Special Issue on Sensitivity Analysis, pp. 1135–1148
- B. Iooss et al. "Some new insights in derivative-based global sensitivity measures". In: *Proceedings of SPAM11 & ESREL 2012 Conference*. ESRA, 2012
- M. Lamboni et al. "Derivative-based global sensitivity measures: General links with Sobol' indices and numerical tests". In: *Mathematics and Computers in Simulation* 87 (2013), pp. 45–54
- S. Touzani and D. Busby. "Screening Method Using the Derivative-based Global Sensitivity Indices with Application to Reservoir Simulator". In: *Oil Gas Sci. Technol. - Rev. IFP Energies nouvelles* (2014)
- O. Roustant et al. "Crossed-derivative based sensitivity measures for interaction screening". In: *Mathematics and Computers in Simulation* 105 (2014), pp. 105–118
- S. Kucherenko and S. Song. "Derivative-based global sensitivity measures as bounds on variance-based global sensitivity indices". In: *Mathematics and Computers in Simulation* (2014, submitted)
- S. Kucherenko and B. Iooss. "Derivative based global sensitivity measures". In: *ArXiv e-prints* (Dec. 2014). arXiv: 1412.2619 [math.ST]

Références

A. Saltelli et al. *Global sensitivity analysis: the primer*. John Wiley, 2008

C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005

E. Volkova, B. looss, and F. Van Dorpe. "Global sensitivity analysis for a numerical model of radionuclide migration from the RRC Kurchatov Institute radwaste disposal site". In: *Stochastic Environmental Research and Risk Assessment* 22.1 (2008), pp. 17–31. ISSN: 1436-3240

H. Liu, Tangn Y., and H. H. Zhang. "A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables". In: *Computational Statistics & Data Analysis* 53.4 (2009), pp. 853–856

L. Lantuéjoul and N. Desassis. "Simulation of a Gaussian random vector: A Propagative version of the Gibbs sampler". In: *Ninth International Geostatistics Congress*. 2012

