Multi-fidelity sensitivity analysis

Loic Le Gratiet[1,2], Claire Cannamela[2] & Bertrand Iooss[3]

[1] Université Denis-Diderot, Paris, France

[2] CEA, DAM, DIF, F-91297 Arpajon, France

[3] EDF R&D, 6 quai Watier, 78401 Chatou, France

7th International Conference on Sensitivity Analysis of Model Output

July 4th 2013

Introduction

- Perform a global sensitivity analysis of a time-consuming numerical model output by calculating variance-based measures of model input parameters.
- Interest in the so-called Sobol indices.
- Surrogate the code output by a metamodel for estimating the Sobol indices.
- Focus on a multi-fidelity cokriging metamodel.
- Propose a Sobol index estimator taking into account the metamodel error and the sampling error introduced during the estimation.

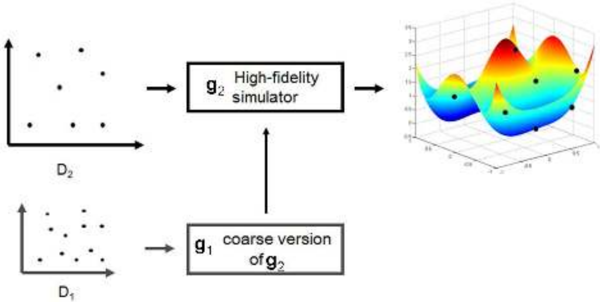
Outline

Motivations

- Objective : replace the output of a code, called $g_2(x)$, by a metamodel.

$$g_2(x) : x \in Q \subset \mathbb{R}^d \mapsto \mathbb{R}$$

- Framework : a coarse version g_1 of g_2 is available.



Principle : build a metamodel of $g_2(x)$ which integrates as well observations of the coarse code output.

$\longrightarrow$ Multi-fidelity co-kriging model

Recursive formulation of the model

- Multi-fidelity co-kriging model :[Kennedy & O'Hagan (2000), Le Gratiet (2012)]

$$\begin{cases} Z_2(x) = \rho Z_1^*(x) + \delta(x) \\ Z_1^*(x) \perp \delta(x) \end{cases}$$

where $Z_1^*(x) \sim [Z_1(x) | \mathbf{Z}_1 = \mathbf{g}_1, \boldsymbol{\beta}_1, \sigma_1^2, \boldsymbol{\theta}_1]$, with $\mathbf{g}_1 = g_1(x), x \in \mathbf{D}_1$

and $Z_1(x) \sim \mathrm{GP}\left(\mathbf{f}_1^t(x)\boldsymbol{\beta}_1, \sigma_1^2 r_1(x, \tilde{x}; \boldsymbol{\theta}_1)\right)$, $\delta(x) \sim \mathrm{GP}\left(\mathbf{f}_\delta^t(x)\boldsymbol{\beta}_\delta, \sigma_\delta^2 r_\delta(x, \tilde{x}; \boldsymbol{\theta}_\delta)\right)$

- Parameters estimation :
  - $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_\delta$, σ_1^2, σ_δ^2 : maximum likelihood method
  - $\boldsymbol{\beta}_1$, $\begin{pmatrix} \boldsymbol{\beta}_\delta \\ \rho \end{pmatrix}$: analytical posterior distribution (Bayesian inference)

- Finally, $Z_2^*(x) \sim [Z_2(x) | \mathbf{Z}_2 = \mathbf{g}_2, \mathbf{Z}_1 = \mathbf{g}_1]$ with $\mathbf{g}_2 = g_2(x), x \in \mathbf{D}_2$

  We suppose that $\mathbf{D}_2 \subset \mathbf{D}_1$ and $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_\delta$, σ_1^2, σ_δ^2 are known.

Recursive formulation of the model

- Multi-fidelity co-kriging model :[Kennedy & O'Hagan (2000), Le Gratiet (2012)]

$$\begin{cases} Z_2(x) = \rho Z_1^*(x) + \delta(x) \\ Z_1^*(x) \perp \delta(x) \end{cases}$$

where $Z_1^*(x) \sim [Z_1(x) | \mathbf{Z}_1 = \mathbf{g}_1, \boldsymbol{\beta}_1, \sigma_1^2, \boldsymbol{\theta}_1]$, with $\mathbf{g}_1 = g_1(x), x \in \mathbf{D}_1$

and $Z_1(x) \sim \mathrm{GP}\left(\mathbf{f}_1^t(x)\boldsymbol{\beta}_1, \sigma_1^2 r_1(x, \tilde{x}; \boldsymbol{\theta}_1)\right)$, $\delta(x) \sim \mathrm{GP}\left(\mathbf{f}_\delta^t(x)\boldsymbol{\beta}_\delta, \sigma_\delta^2 r_\delta(x, \tilde{x}; \boldsymbol{\theta}_\delta)\right)$

- Parameters estimation :
  - $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_\delta$,$\sigma_1^2$, σ_δ^2 : maximum likelihood method
  - $\boldsymbol{\beta}_1, \begin{pmatrix} \boldsymbol{\beta}_\delta \\ \rho \end{pmatrix}$: analytical posterior distribution (Bayesian inference)

- Finally, $Z_2^*(x) \sim [Z_2(x) | \mathbf{Z}_2 = \mathbf{g}_2, \mathbf{Z}_1 = \mathbf{g}_1]$ with $\mathbf{g}_2 = g_2(x), x \in \mathbf{D}_2$

We suppose that $\mathbf{D}_2 \subset \mathbf{D}_1$ and $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_\delta$,$\sigma_1^2$, σ_δ^2 are known.

Recursive formulation of the model

- Multi-fidelity co-kriging model :[Kennedy & O'Hagan (2000), Le Gratiet (2012)]

$$\begin{cases} Z_2(x) = \rho Z_1^*(x) + \delta(x) \\ Z_1^*(x) \perp \delta(x) \end{cases}$$

where $Z_1^*(x) \sim [Z_1(x)|\mathbf{Z}_1 = \mathbf{g}_1, \boldsymbol{\beta}_1, \sigma_1^2, \boldsymbol{\theta}_1]$, with $\mathbf{g}_1 = g_1(x), x \in \mathbf{D}_1$

and $Z_1(x) \sim \mathrm{GP}\left(\mathbf{f}_1^t(x)\boldsymbol{\beta}_1, \sigma_1^2 r_1(x, \tilde{x}; \boldsymbol{\theta}_1)\right)$, $\delta(x) \sim \mathrm{GP}\left(\mathbf{f}_\delta^t(x)\boldsymbol{\beta}_\delta, \sigma_\delta^2 r_\delta(x, \tilde{x}; \boldsymbol{\theta}_\delta)\right)$

- Parameters estimation :
  - $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_\delta$,$\sigma_1^2$, σ_δ^2 : maximum likelihood method
  - $\boldsymbol{\beta}_1, \begin{pmatrix} \boldsymbol{\beta}_\delta \\ \rho \end{pmatrix}$: analytical posterior distribution (Bayesian inference)

- Finally, $Z_2^*(x) \sim [Z_2(x)|\mathbf{Z}_2 = \mathbf{g}_2, \mathbf{Z}_1 = \mathbf{g}_1]$ with $\mathbf{g}_2 = g_2(x), x \in \mathbf{D}_2$

We suppose that $\mathbf{D}_2 \subset \mathbf{D}_1$ and $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_\delta$,$\sigma_1^2$, σ_δ^2 are known.

Predictive distribution

- In Universal Cokriging, the predictive distribution of $Z_2^*(x)$ is not Gaussian.

  The predictive mean and variance can be decomposed as :

$$\begin{aligned}\mu_{Z_2}(x) &= \mathbb{E}\left[Z_2(x)|\mathbf{Z}_2 = \mathbf{g}_2, \mathbf{Z}_1 = \mathbf{g}_1\right] \\ &= \hat{\rho}\mu_{Z_1}(x) + \mu_\delta(x)\end{aligned}$$

$$\begin{aligned}\sigma^2_{Z_2}(x) &= \mathrm{var}\left(Z_2(x)|\mathbf{Z}_2 = \mathbf{g}_2, \mathbf{Z}_1 = \mathbf{g}_1]\right) \\ &= \hat{\rho}^2\sigma^2_{Z_1}(x) + \sigma^2_\delta(x)\end{aligned}$$

- Remarks :
  - in $\mu_{Z_2}(x)$: $\boldsymbol{\beta}$ and ρ are replaced by their posterior means.
  - in $\sigma^2_{Z_2}(x)$: we infer from the posterior distributions of $\boldsymbol{\beta}$ and ρ.

Outline

Sobol index definition

- Context :
  - The input is a random vector $X = (X^{d_1}, X^{d_2})$ defined on $(\Omega_X, \mathcal{F}_X, \mathbb{P}_X)$, with probability measure $\mu = \mu^{d_1} \otimes \mu^{d_2}$, $d = d_1 + d_2$.
  - We replace the code output $g_2(x)$ with a multi-fidelity co-kriging model $Z_2^*(x)$ defined on $(\Omega_Z, \mathcal{F}_Z, \mathbb{P}_Z)$.
- Objective : we are interested in evaluating the closed Sobol indices defined by

$$\mathrm{S}^{d_1} = \frac{\mathrm{V}^{d_1}}{\mathrm{V}} = \frac{\mathrm{var}_X \left(\mathbb{E}_X \left[g_2(X)|X^{d_1}\right]\right)}{\mathrm{var}_X \left(g_2(X)\right)}$$

- Classical approach estimation : replace $g_2(x)$ by $\mu_{Z_2^*}(x)$.
  - \+ Computationally cheap
  - − do not infer from the meta-model error

Sobol index definition

- Context :
  - The input is a random vector $X = (X^{d_1}, X^{d_2})$ defined on $(\Omega_X, \mathcal{F}_X, \mathbb{P}_X)$, with probability measure $\mu = \mu^{d_1} \otimes \mu^{d_2}$, $d = d_1 + d_2$.
  - We replace the code output $g_2(x)$ with a multi-fidelity co-kriging model $Z_2^*(x)$ defined on $(\Omega_Z, \mathcal{F}_Z, \mathbb{P}_Z)$.
- Objective : we are interested in evaluating the closed Sobol indices defined by

$$\mathrm{S}^{d_1} = \frac{\mathrm{V}^{d_1}}{\mathrm{V}} = \frac{\mathrm{var}_X\left(\mathbb{E}_X\left[g_2(X)|X^{d_1}\right]\right)}{\mathrm{var}_X\left(g_2(X)\right)}$$

- Classical approach estimation : replace $g_2(x)$ by $\mu_{Z_2^*}(x)$.
  - + Computationally cheap
  - − do not infer from the meta-model error

The meta-model error

- Another approach :[Oakley & O'Hagan (2004), Marrel & al. (2009)]

  Replace $g_2(x)$ by $Z_2^*(x)$ and consider the term :

$$S_{Z_2^*}^{d_1} = \frac{\mathrm{var}_X\left(\mathbb{E}_X\left[Z_2^*(X)|X^{d_1}\right]\right)}{\mathrm{var}_X\left(Z_2^*(X)\right)} \quad \text{(on } (\Omega_Z, \mathcal{F}_Z, \mathbb{P}_Z))$$

  They suggest the following quantity :

$$\tilde{S}_{Z_2^*}^{d_1} = \frac{\mathbb{E}_Z\left[\mathrm{var}_X\left(\mathbb{E}_X\left[Z_2^*(X)|X^{d_1}\right]\right)\right]}{\mathbb{E}_Z\left[\mathrm{var}_X\left(Z_2^*(X)\right)\right]}$$

  They evaluate its uncertainty with :

$$\widetilde{\Sigma}_{Z_2^*}^2 = \frac{\mathrm{var}_Z\left(\mathrm{var}_X\left(\mathbb{E}_X\left[Z_2^*(X)|X^{d_1}\right]\right)\right)}{\mathbb{E}_Z\left[\mathrm{var}_X\left(Z_2^*(X)\right)\right]^2}$$

  + infer from the meta-model error
  − Computationally expensive (numerical integrations)
  − $\tilde{S}_{Z_2^*}^{d_1}$ is not the expectation of $S_{Z_2^*}^{d_1}$ and $\widetilde{\Sigma}_{Z_2^*}^2$ is different from the variance of $S_{Z_2^*}^{d_1}$

Sobol index estimation - objective

- Recall : $S^{d_1}_{Z_2^*}$ is a random variable

$$S^{d_1}_{Z_2^*} = \frac{\mathrm{var}_X\left(\mathbb{E}_X\left[Z_2^*(X)|X^{d_1}\right]\right)}{\mathrm{var}_X\left(Z_2^*(X)\right)}$$

- Objective : We want to build
  - A unbiased Monte Carlo estimator $\hat{\mathrm{S}}^{d_1}_{Z_2^*,m}$ of $S^{d_1}_{Z_2^*}$, where m represents the number of Monte-Carlo particles.
  - An estimator of the variance of $\hat{\mathrm{S}}^{d_1}_{Z_2^*,m}$.

Sobol index estimation - sampling

- Generate a $2m$-sample $\mathbf{x} = (x_i, \tilde{x}_i)_{i=1,\dots,m}$ of the random vector $(X, \tilde{X})$ [Sobol (1993)]

such that $X = (X^{d_1}, X^{d_2}) \quad \tilde{X} = (X^{d_1}, \tilde{X}^{d_2}) \quad X^{d_2} \perp \tilde{X}^{d_2}$

- For $k = 1, \dots, N_Z$:
  1. Sample a realization $z^*_{2,k}(\mathbf{x})$ of $Z^*_2(\mathbf{x})$ at points in $\mathbf{x}$.
  2. Compute $\hat{\mathrm{S}}^{d_1}_{Z^*_2,m,k,1}$ from $z^*_{2,k}(\mathbf{x})$ with : [Janon & al. (2012)]

$$\hat{\mathrm{S}}^{d_1}_{Z^*_2,m,k,1} = \frac{\frac{1}{m}\sum_{i=1}^m z^*_{2,k}(x_i) z^*_{2,k}(\tilde{x}_i) - \left(\frac{1}{2m}\sum_{i=1}^m z^*_{2,k}(x_i) + z^*_{2,k}(\tilde{x}_i)\right)^2}{\frac{1}{m}\sum_{i=1}^m z^*_{2,k}(x_i)^2 - \left(\frac{1}{2m}\sum_{i=1}^m z^*_{2,k}(x_i) + z^*_{2,k}(\tilde{x}_i)\right)^2}$$

  3. For $l = 2, \dots, B$:
     - Sample with replacements $\mathbf{x}^{B_l}$ from $\mathbf{x}$
     - Compute $\hat{\mathrm{S}}^{d_1}_{Z^*_2,m,k,l}$ from $z^*_{2,k}(\mathbf{x}^{B_l})$.

Sobol index estimation - sampling

- Generate a $2m$-sample $\mathbf{x} = (x_i, \tilde{x}_i)_{i=1,\dots,m}$ of the random vector $(X, \tilde{X})$ [Sobol (1993)]

  such that $X = (X^{d_1}, X^{d_2}) \quad \tilde{X} = (X^{d_1}, \tilde{X}^{d_2}) \quad X^{d_2} \perp \tilde{X}^{d_2}$

- For $k = 1, \dots, N_Z$:

  1. Sample a realization $z_{2,k}^*(\mathbf{x})$ of $Z_2^*(\mathbf{x})$ at points in $\mathbf{x}$.
  2. Compute $\hat{\mathrm{S}}_{Z_2^*,m,k,1}^{d_1}$ from $z_{2,k}^*(\mathbf{x})$ with : [Janon & al. (2012)]

$$\hat{\mathrm{S}}_{Z_2^*,m,k,1}^{d_1} = \frac{\frac{1}{m}\sum_{i=1}^m z_{2,k}^*(x_i) z_{2,k}^*(\tilde{x}_i) - \left(\frac{1}{2m}\sum_{i=1}^m z_{2,k}^*(x_i) + z_{2,k}^*(\tilde{x}_i)\right)^2}{\frac{1}{m}\sum_{i=1}^m z_{2,k}^*(x_i)^2 - \left(\frac{1}{2m}\sum_{i=1}^m z_{2,k}^*(x_i) + z_{2,k}^*(\tilde{x}_i)\right)^2}$$

  3. For $l = 2, \dots, B$:
     - Sample with replacements $\mathbf{x}^{B_l}$ from $\mathbf{x}$
     - Compute $\hat{\mathrm{S}}_{Z_2^*,m,k,l}^{d_1}$ from $z_{2,k}^*(\mathbf{x}^{B_l})$.

Sobol index estimation - sampling

- Generate a $2m$-sample $\mathbf{x} = (x_i, \tilde{x}_i)_{i=1,\dots,m}$ of the random vector $(X, \tilde{X})$ [Sobol (1993)]

  such that $X = (X^{d_1}, X^{d_2}) \quad \tilde{X} = (X^{d_1}, \tilde{X}^{d_2}) \quad X^{d_2} \perp \tilde{X}^{d_2}$

- For $k = 1, \dots, N_Z$:

  1. Sample a realization $z^*_{2,k}(\mathbf{x})$ of $Z^*_2(\mathbf{x})$ at points in $\mathbf{x}$.
  2. Compute $\hat{\mathrm{S}}^{d_1}_{Z^*_2,m,k,1}$ from $z^*_{2,k}(\mathbf{x})$ with : [Janon & al. (2012)]

$$\hat{\mathrm{S}}^{d_1}_{Z^*_2,m,k,1} = \frac{\frac{1}{m}\sum_{i=1}^m z^*_{2,k}(x_i) z^*_{2,k}(\tilde{x}_i) - \left(\frac{1}{2m}\sum_{i=1}^m z^*_{2,k}(x_i) + z^*_{2,k}(\tilde{x}_i)\right)^2}{\frac{1}{m}\sum_{i=1}^m z^*_{2,k}(x_i)^2 - \left(\frac{1}{2m}\sum_{i=1}^m z^*_{2,k}(x_i) + z^*_{2,k}(\tilde{x}_i)\right)^2}$$

  3. For $l = 2, \dots, B$:
     - Sample with replacements $\mathbf{x}^{B_l}$ from $\mathbf{x}$
     - Compute $\hat{\mathrm{S}}^{d_1}_{Z^*_2,m,k,l}$ from $z^*_{2,k}(\mathbf{x}^{B_l})$.

Sobol index estimation - sampling

- Generate a $2m$-sample $\mathbf{x} = (x_i, \tilde{x}_i)_{i=1,\dots,m}$ of the random vector $(X, \tilde{X})$ [Sobol (1993)]

  such that $X = (X^{d_1}, X^{d_2}) \quad \tilde{X} = (X^{d_1}, \tilde{X}^{d_2}) \quad X^{d_2} \perp \tilde{X}^{d_2}$

- For $k = 1, \dots, N_Z$:

  1. Sample a realization $z^*_{2,k}(\mathbf{x})$ of $Z^*_2(\mathbf{x})$ at points in $\mathbf{x}$.
  2. Compute $\hat{\mathrm{S}}^{d_1}_{Z^*_2,m,k,1}$ from $z^*_{2,k}(\mathbf{x})$ with : [Janon & al. (2012)]

$$\hat{\mathrm{S}}^{d_1}_{Z^*_2,m,k,1} = \frac{\frac{1}{m}\sum_{i=1}^m z^*_{2,k}(x_i) z^*_{2,k}(\tilde{x}_i) - \left(\frac{1}{2m}\sum_{i=1}^m z^*_{2,k}(x_i) + z^*_{2,k}(\tilde{x}_i)\right)^2}{\frac{1}{m}\sum_{i=1}^m z^*_{2,k}(x_i)^2 - \left(\frac{1}{2m}\sum_{i=1}^m z^*_{2,k}(x_i) + z^*_{2,k}(\tilde{x}_i)\right)^2}$$

  3. For $l = 2, \dots, B$:
     - Sample with replacements $\mathbf{x}^{B_l}$ from $\mathbf{x}$
     - Compute $\hat{\mathrm{S}}^{d_1}_{Z^*_2,m,k,l}$ from $z^*_{2,k}(\mathbf{x}^{B_l})$.

Sobol index estimation - meta-model and MC errors

- The estimator of $S_{Z_2^*}^{d_1}$:

$\hat{\mathcal{S}}_{Z_2^*,m}^{d_1}$ is defined as the empirical mean of the sample $\left(\hat{\mathcal{S}}_{Z_2^*,m,k,l}^{d_1}\right)_{\substack{k=1,\dots,N_Z\\ l=1,\dots,B}}$:

$$\hat{\mathcal{S}}_{Z_2^*,m}^{d_1} = \frac{1}{N_Z \times B} \sum_{k=1}^{N_Z} \sum_{l=1}^{B} \hat{\mathcal{S}}_{Z_2^*,m,k,l}^{d_1}$$

- The variance of the estimator $\hat{\mathcal{S}}_{Z_2^*,m}^{d_1}$:

$\hat{\Sigma}^2_{\hat{\mathcal{S}}_{Z_2^*,m}^{d_1}}$ is estimated from the empirical variance of $\left(\hat{\mathcal{S}}_{Z_2^*,m,k,l}^{d_1}\right)_{\substack{k=1,\dots,N_Z\\ l=1,\dots,B}}$:

$$\hat{\Sigma}^2_{\hat{\mathcal{S}}_{Z_2^*,m}^{d_1}} = \frac{1}{N_Z \times B - 1} \sum_{k=1}^{N_Z} \sum_{l=1}^{B} \left(\hat{\mathcal{S}}_{Z_2^*,m,k,l}^{d_1} - \hat{\mathcal{S}}_{Z_2^*,m}^{d_1}\right)^2$$

It integrates both meta-modeling error and Monte-Carlo integrations error.

Sobol index estimation - minimal number of MC particles

- Variance due to the meta-modeling :

$\mathrm{var}_Z\left(\mathbb{E}_X\left[\hat{\mathcal{S}}^{d_1}_{Z_2^*,m}|Z_2^*(x)\right]\right)$ can be estimated by :

$$\widehat{\Sigma}^2_{Z_2^*} = \frac{1}{B}\sum_{l=1}^{B}\left[\frac{1}{N_Z-1}\sum_{k=1}^{N_Z}\left(\hat{\mathcal{S}}^{d_1}_{Z_2^*,m,k,l} - \overline{\hat{\mathcal{S}}}^{d_1}_{Z_2^*,m,l}\right)^2\right]$$

- Variance due to Monte-Carlo integrations :

$\mathrm{var}_X\left(\mathbb{E}_Z\left[\hat{\mathcal{S}}^{d_1}_{Z_2^*,m}|(X_i,\tilde{X}_i)_{i=1,\dots,m}\right]\right)$ can be estimated by :

$$\widehat{\Sigma}^2_{\mathrm{MC}} = \frac{1}{N_Z}\sum_{k=1}^{N_Z}\left[\frac{1}{B-1}\sum_{l=1}^{B}\left(\hat{\mathcal{S}}^{d_1}_{Z_2^*,m,k,l} - \overline{\hat{\mathcal{S}}}^{d_1}_{Z_2^*,m,k}\right)^2\right]$$

- Determining the minimal number of Monte-Carlo particles m :

Considering given the cokriging model $Z_2^*(x)$, the minimal value m is the one such that $\widehat{\Sigma}^2_{Z_2^*(x)} \approx \widehat{\Sigma}^2_{\mathrm{MC}}$

Sobol index estimation - minimal number of MC particles

- Variance due to the meta-modeling :

  $\mathrm{var}_Z\left(\mathbb{E}_X\left[\hat{\mathcal{S}}^{d_1}_{Z_2^*,m} | Z_2^*(x)\right]\right)$ can be estimated by :

$$\widehat{\Sigma}^2_{Z_2^*} = \frac{1}{B}\sum_{l=1}^{B}\left[\frac{1}{N_Z-1}\sum_{k=1}^{N_Z}\left(\hat{\mathcal{S}}^{d_1}_{Z_2^*,m,k,l} - \overline{\hat{\mathcal{S}}}^{d_1}_{Z_2^*,m,l}\right)^2\right]$$

- Variance due to Monte-Carlo integrations :

  $\mathrm{var}_X\left(\mathbb{E}_Z\left[\hat{\mathcal{S}}^{d_1}_{Z_2^*,m} | (X_i,\tilde{X}_i)_{i=1,\dots,m}\right]\right)$ can be estimated by :

$$\widehat{\Sigma}^2_{\mathsf{MC}} = \frac{1}{N_Z}\sum_{k=1}^{N_Z}\left[\frac{1}{B-1}\sum_{l=1}^{B}\left(\hat{\mathcal{S}}^{d_1}_{Z_2^*,m,k,l} - \overline{\hat{\mathcal{S}}}^{d_1}_{Z_2^*,m,k}\right)^2\right]$$

- Determining the minimal number of Monte-Carlo particles m :

  Considering given the cokriging model $Z_2^*(x)$, the minimal value m is the one such that $\widehat{\Sigma}^2_{Z_2^*(x)} \approx \widehat{\Sigma}^2_{\mathsf{MC}}$

Outline

Ishigami function - kriging case

First order Sobol index of the first input parameter when the size n of the learning sample increases.

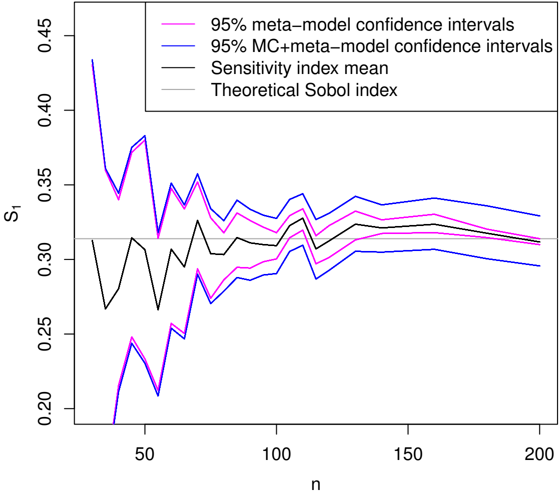
Learning sample :
space filling design

covariance structure :
tensorised $5/2$-Matérn kernel

$m = 10000$

$N_Z = 500$

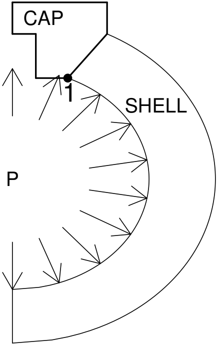
$B = 300$



Spherical tank under internal pressure - cokriging case

- Fine code : $g_2(x)$ is the Von Mises stress at point 1 provided by a finite elements code. $x = (P, R_{\text{int}}, T_{\text{shell}}, T_{\text{cap}}, E_{\text{shell}}, E_{\text{cap}}, \sigma_{\text{y,shell}}, \sigma_{\text{y,cap}}) \qquad (d = 8)$

P : value of the internal pressure.
R_{int} : length of the internal radius of the shell.
T_{shell} : thickness of the shell.
T_{cap} : thickness of the cap.
E_{shell} : Young's modulus of the shell material.
E_{cap} : Young's modulus of the cap material.
$\sigma_{\text{y,shell}}$: yield stress of the cap material.
$\sigma_{\text{y,cap}}$: yield stress of the cap material.



- Coarse code :

  g_1 is the 1D simplification corresponding to a perfectly spherical tank

$$g_1(x) = \frac{3}{2} \frac{(R_{int} + T_{shell})^3}{(R_{int} + T_{shell})^3 - R_{int}^3} P$$

- Multi-fidelity co-kriging model : built with $n_1 = 100$ and only $n_2 = 20$.

  We estimate the efficiency E_{ff} of the metamodel of $g_2(x)$ by the coefficient of determination calculated on an external set of 7000 points. $E_{ff} \approx 86\%$

Spherical tank under internal pressure - results

Kriging-based sensitivity analysis of the coarse code output $g_1(x)$

Learning sample :

space filling designs

$n_1 = 100$

covariance structure :

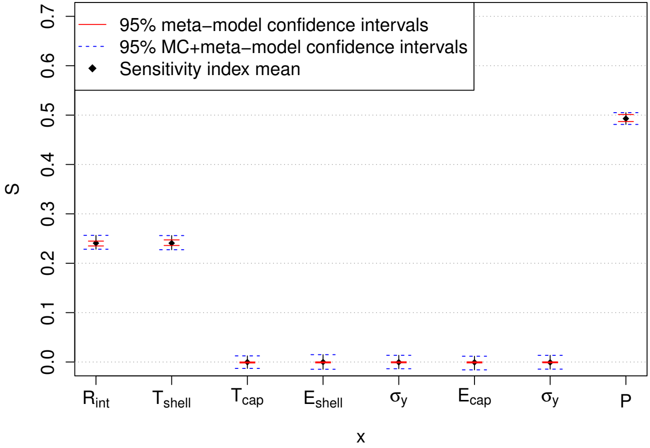
tensorised $5/2$-Matérn kernel

$m = 20000$

$N_Z = 300$

$B = 150$

$\sum_{i=1}^{8} S_i \approx 97\%$



Spherical tank under internal pressure - results

Cokriging-based sensitivity analysis of the fine code output $g_2(x)$

Learning sample :

space filling design

$n_1 = 100$

$n_2 = 20$

covariance structures :

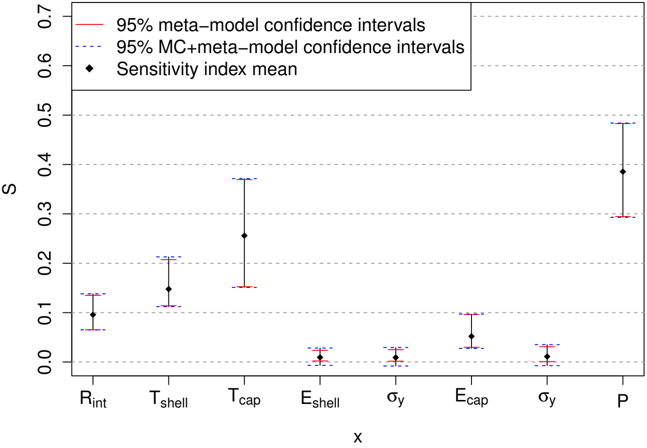
2 tensorised $5/2$-Matérn kernels

$m = 20000$

$N_Z = 300$

$B = 150$

$\sum_{i=1}^{8} S_i \approx 96.7\%$



Conclusion

- We have proposed a Sobol index estimator which uses realizations of a predictive process (built from a multi-fidelity cokriging model) at points in a large sample.

  $\longrightarrow$ we use the *conditioning kriging* [Chilès & Delfiner (1999)]

- In the computation of its variance, we distinguish among metamodel and Monte-Carlo variance contributions.

- We can also determine the minimal number of particles m such that the Monte-Carlo variance no longer dominates.

  $\longrightarrow$ The value m is different for each Sobol index estimation.

Bibliography

- SOBOL, I. M. (1993) , Sensitivity estimates for non linear mathematical models. *Mathematical modeling and Computational Experiments* **1**, 407–414.
- KENNEDY, M. C. & O'HAGAN, A. (2000) , Predicting the output from a complex computer code when fast approximations are available. *Biometrika* **87**, 1–13.
- LE GRATIET, L. & GARNIER, J. (2012), *Recursive co-kriging model for Design of Computer experiments with multiple levels of fidelity with an application to hydrodynamic*, submitted to International Journal of Uncertainty Quantification.
- J.E. OAKLEY & A O'HAGAN (2004), *Probabilistic sensitivity analysis of complex models : A bayesian approach*, *Journal of the Royal Statistical Society B*, **66**(3) :751–769.
- A. MARREL, B. IOOSS, B. LAURENT & O. ROUSTANT (2012), *Calculations of Sobol indices for the gaussian process metamodel*, *Reliability Engineering and System Safety*, **94** :742–751.
- A. JANON, T. KLEIN, A. LAGNOUX, M. NODET & C. PRIEUR (2013), *Asymptotic normality and efficiency of two Sobol index estimators*, to appear in ESAIM probability and statistics.
- J.P. CHILÈS & P. DELFINER (1999), *Geostatistics : modeling spatial uncertainty*, *Wiley series in probability and statistics (Applied probability and statistics section)*.
- L. LE GRATIET,C. CANNAMELA, & B.IOOSS (2013), *A Bayesian approach for global sensitivity analysis of (multi-fidelity) computer codes*, submitted to SIAM/ASA UQ.

R CRAN package : `http://cran.r-project.org/web/packages/MuFiCokriging`

Sampling from the predictive distribution

- Two issues :
  1. How to sample from $Z_2^*(x) \sim [Z_2(x)|\mathbf{Z}_2 = \mathbf{g}_2, \mathbf{Z}_1 = \mathbf{g}_1]$?
  2. How to deal with large m ?

- Sampling from $Z_2^*(x) \sim [Z_2(x)|\mathbf{Z}_2 = \mathbf{g}_2, \mathbf{Z}_1 = \mathbf{g}_1]$:

  $Z_2^*(x)$ can be written by : $Z_2^*(x) = \rho Z_1^*(x) + \delta^*_{\rho,\boldsymbol{\beta}_\delta}(x)$

  where $Z_1^*(x)$ and $\delta^*_{\rho,\boldsymbol{\beta}_\delta}(x) = [\delta(x)|\rho, \boldsymbol{\beta}_\delta]$ are Gaussian processes

  and $(\rho, \boldsymbol{\beta}_\delta)$ have a known distribution.

  1. Sample a realization $z^*_{1,k}(x)$ of $Z_1^*(x)$
  2. Sample a realization $(\rho^k, \boldsymbol{\beta}^k_\delta)$ of $(\rho, \boldsymbol{\beta}_\delta)$
  3. Sample a realization $\delta^*_k(x)$ of $\delta^*_{\rho^k,\boldsymbol{\beta}^k_\delta}(x)$
  4. Set $z^*_{2,k}(x) = \rho^k \tilde{z}^*_{1,k}(x) + \delta^*_k(x)$

Sampling from the predictive distribution

- Two issues :
  1. How to sample from $Z_2^*(x) \sim [Z_2(x)|\mathbf{Z}_2 = \mathbf{g}_2, \mathbf{Z}_1 = \mathbf{g}_1]$?
  2. How to deal with large m ?

- Sampling from $Z_2^*(x) \sim [Z_2(x)|\mathbf{Z}_2 = \mathbf{g}_2, \mathbf{Z}_1 = \mathbf{g}_1]$:

  $Z_2^*(x)$ can be written by : $Z_2^*(x) = \rho Z_1^*(x) + \delta^*_{\rho,\boldsymbol{\beta}_\delta}(x)$

  where $Z_1^*(x)$ and $\delta^*_{\rho,\boldsymbol{\beta}_\delta}(x) = [\delta(x)|\rho, \boldsymbol{\beta}_\delta]$ are Gaussian processes

  and $(\rho, \boldsymbol{\beta}_\delta)$ have a known distribution.

  1. Sample a realization $z^*_{1,k}(x)$ of $Z_1^*(x)$
  2. Sample a realization $(\rho^k, \boldsymbol{\beta}_\delta^k)$ of $(\rho, \boldsymbol{\beta}_\delta)$
  3. Sample a realization $\delta_k^*(x)$ of $\delta^*_{\rho^k,\boldsymbol{\beta}_\delta^k}(x)$
  4. Set $z^*_{2,k}(x) = \rho^k \tilde{z}^*_{1,k}(x) + \delta_k^*(x)$

Sampling from the predictive distribution

- Two issues :
  1. How to sample from $Z_2^*(x) \sim [Z_2(x) | \mathbf{Z}_2 = \mathbf{g}_2, \mathbf{Z}_1 = \mathbf{g}_1]$?
  2. How to deal with large m ?

- Sampling with large m :

  Conditioning kriging result :[Chilès & Delfiner (1999)]

  We can sample from the conditional distributions of $Z_1^*(x)$ and $\delta^*(x)$ by sampling from unconditioned distributions and then by applying a linear transformation on them.

  We can use efficient algorithms to compute realizations of processes with unconditioned distributions :

  - Stationary kernel : spectral decomposition
  - Tensorised covariance kernel : Karhunen-Loeve decomposition

    $\longrightarrow$ sequentially adding new points to $\mathbf{x}$ without re-estimating the decomposition.