



## Sensitivity analysis of spatial models using geostatistical simulation

Nathalie Saint-Geours, Christian Lavergne, Jean-Stéphane Bailly, Frédéric Grelot

### ► To cite this version:

Nathalie Saint-Geours, Christian Lavergne, Jean-Stéphane Bailly, Frédéric Grelot. Sensitivity analysis of spatial models using geostatistical simulation. Robert Marschallinger, Fritz Zobl. Mathematical Geosciences at the Crossroads of Theory and Practice - IAMG 2011 Conference, Sep 2011, Salzburg, Austria. pp.178-189, 2012, <10.5242/iamg.2011.0172>. <hal-00666755>

**HAL Id: hal-00666755**

**<https://hal.archives-ouvertes.fr/hal-00666755>**

Submitted on 6 Feb 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Sensitivity analysis of spatial models using geostatistical simulation

Nathalie SAINT-GEOURS<sup>1</sup>, Christian LAVERGNE<sup>2</sup>, Jean-Stéphane BAILLY<sup>3</sup>  
& Frédéric GRELOT<sup>4</sup>

<sup>1</sup> AgroParisTech, UMR TETIS, I3M, France, [saintge@teledetection.fr](mailto:saintge@teledetection.fr)

<sup>2</sup> I3M, Université Montpellier III, France, [christian.lavergne@univ-montp3.fr](mailto:christian.lavergne@univ-montp3.fr)

<sup>3</sup> AgroParisTech, UMR TETIS, France, [bailly@teledetection.fr](mailto:bailly@teledetection.fr)

<sup>4</sup> Cemagref, UMR G-EAU, France, [grelot@cemagref.fr](mailto:grelot@cemagref.fr)

## Abstract

Geostatistical simulations are used to perform a global sensitivity analysis on a model  $Y = f(X_1 \dots X_k)$  where one of the model inputs  $X_i$  is a continuous 2D-field. Geostatistics allow specifying uncertainty on  $X_i$  with a spatial covariance model and generating random realizations of  $X_i$ . These random realizations are used to propagate uncertainty through model  $f$  and estimate global sensitivity indices. Focusing on variance-based global sensitivity analysis (GSA), we assess in this paper how sensitivity indices vary with covariance parameters (range, sill, nugget). Results give a better understanding on how and when to use geostatistical simulations for sensitivity analysis of spatially distributed models.

## 1 Introduction

Numerous spatial models are developed to support decision making in various fields of environmental management. These models use environmental data that is spatially distributed, including maps derived from sampled data (e.g. digital elevation model, soil map, etc.). These spatial inputs are always partly uncertain, due to measurement errors, lack of knowledge, aleatory variability (see Refsgaard et al., 2007 for a discussion on the various sources of uncertainty in model inputs). In order to provide confidence in these models, uncertainty analysis (UA) and sensitivity analysis (SA) are increasingly recognized as important steps in the modelling process. They allow robustness of model predictions to be checked and help identifying the input factors that account for most of model output variability (Saltelli et al., 2008).

Geostatistical simulation has an important role to play in UA/SA of models  $Y = f(X_1 \dots X_k)$  when some model input  $X_i$  is a continuous 2D-field. Geostatistics first offers a way to describe the uncertainty on spatial input  $X_i$  with a spatial covariance model. Then, random realizations of  $X_i$  can be generated through geostatistical simulation (Journel and Huijbregts, 1978). These random realizations can be used to propagate uncertainty through model  $f$  and discuss the resulting uncertainty on model output  $Y$  (Aerts et al., 2003 - on a problem of optimal location of a ski run; Ruffo et al., 2006 - on hydrocarbon exploration risk evaluation). Within variance-based global sensitivity analysis (GSA) framework, these random realizations can also be sampled alongside with other scalar model inputs to estimate sensitivity indices for each model input (Lilburne & Tarantola, 2009).

Still, a practical problem remains for modellers who intend to use geostatistical simulations in UA/SA of a spatially distributed model: covariance parameters which describe uncertainty on input 2D-field  $X_i$  must be estimated carefully, but there is usually few data to support this estimation. At the same time, UA/SA results are known to depend heavily on the specification of uncertainty on model inputs. Thus, the following questions arise: to what extent are UA/SA results influenced by spatial covariance parameters? In which cases the uncertainty on input 2D-field  $X_i$  accounts for a large or a small part of total variability of model output?

To answer these questions, this article aims at determining, in the context of spatial GSA, how sensitivity indices depend on the covariance parameters which describe uncertainty on spatially distributed model inputs. We first describe a simple spatial model  $Y = f(X, Z)$  with two inputs: a scalar input  $X$  and a 2D spatially distributed input  $Z(u)$  (section 2). Then we present variance-based global sensitivity analysis (section 3), and show into details how to estimate sensitivity indices on model  $M$  using geostatistical simulations of 2D-field  $Z(u)$  (section 4). We finally assess the impact of the three usual covariance parameters (range, sill, nugget) on sensitivity indices in model  $M$  (section 5). Our results might well prove useful in better understanding the results of a spatial GSA and in deciding whether it is necessary to carefully estimate spatial covariance parameters to describe uncertainty on input 2D-fields.

## 2 A simple spatially distributed model $M$

For sake of clarity, we will base our paper on a simple case-study. We describe in this section an example of a spatially distributed model  $M$ .

### 4.1 Description of model $M$

Consider a spatial domain  $D \subset \mathbb{R}^2$ . For numerical application, we represent domain  $D$  by a regular square grid  $G$  of size  $50 \times 50$ . We will study in the following sections a model  $M$  with two inputs:

$$Y = M(X, Z)$$

where:

- $X = (X_1, X_2)$  is a vector of two scalars
- $Z(u)$  is a 2D continuous field defined on domain  $D$ .
- model output  $Y(u)$  is also a 2D continuous field defined by :

$$\forall u \in D, \quad Y(u) = f(X, Z(u))$$

Function  $f(.,.)$  can be any mapping from  $\mathbb{R}^3$  to  $\mathbb{R}$ . For numerical application, we arbitrarily choose the following mapping:

$$f(x_1, x_2, z) = 10^{-3} \cdot (x_1^2 + x_2 \cdot e^{0.036 \cdot Z(u)} + 40 \cdot Z(u))$$

Model  $M$  is a “point-based model”: the value of model output  $Y(u)$  at any point  $u \in D$  only depends on the scalar inputs  $(X_1, X_2)$  and on the value of  $Z(u)$  at the same point  $u$ . Point-based models are encountered in many environmental applications. For example,  $M$  could be a spatially distributed model used for economic assessment of flood risk: in this case, model input  $Z(u)$  could be a map of the maximal water levels reached during a flood event over a given area  $D$ ,

$X = (X_1, X_2)$  would be a set of economic parameters, and model output  $Y(u)$  would be the map of expected damages due to the flood over the area.

## 4.2 Output of interest

In order to perform sensitivity analysis of model  $M$ , we need to consider a single scalar quantity of interest derived from model output  $Y(u)$ . In most applications, the output of interest is either the value of 2D-field  $Y(u)$  at some specific point  $u$  of the study area, or the mean (or total) value of  $Y(u)$  over a given zone within the study area. Here we define the output of interest  $Y_D$  as the mean value of field  $Y(u)$  over spatial domain  $D$ :

$$Y_D = \frac{1}{|D|} \cdot \int_{u \in D} Y(u) \cdot du$$

In the following sections we will use variance-based global sensitivity analysis to assess the variability of  $Y_D$  due to the uncertainty on model inputs  $X$  and  $Z(u)$ .

## 3 Variance-based global sensitivity analysis

Sensitivity analysis (SA) aims at a studying how uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model inputs. Among the various available SA techniques (see Helton and Davis, 2006 for a review), variance-based global sensitivity analysis (GSA) has several advantages: it explores widely the space of uncertain input factors and is suitable for complex models with non-linear effects and interactions among factors.

GSA is based on the decomposition of the variance of model output  $Y$  in conditional variances. It leads to the definition of two importance measures for each input factor  $X_i$  of a model: first-order sensitivity index  $S_i$  and total-order sensitivity index  $ST_i$ . First-order sensitivity index of input factor  $X_i$  is defined by:

$$S_i = \frac{\text{Var}(E[Y|X_i])}{\text{Var}(Y)}$$

$S_i$  measures the main effect contribution of input factor  $X_i$  to the variance of model output  $Y$ . It is the expected part of output variance  $\text{Var}(Y)$  that could be reduced if input factor  $X_i$  was perfectly known. Total order sensitivity index  $ST_i$  of input factor  $X_i$  is defined as:

$$ST_i = \frac{E(\text{Var}[Y|X_{-i}])}{\text{Var}(Y)}$$

where  $X_{-i}$  denotes all input factors but  $X_i$ .  $ST_i$  measures the contribution of input factor  $X_i$  and all its interactions with other input factors  $X_j$  to the variance of model output  $Y$ . It is the expected part of output variance  $\text{Var}(Y)$  that would remain if all input factors but  $X_i$  were perfectly known.

Sensitivity indices can be used to identify the model inputs that account for most of model output variability (input factors  $X_i$  with high first order indices  $S_i$ ); it may lead to model simplification by identifying model inputs that have little influence on model output variance (input factors  $X_i$  with low total order sensitivity indices  $ST_i$ ); it also allows discussing the contribution of interactions between input factors to the model output variance (comparison between first and total order sensitivity indices). For more details on GSA basics, see Saltelli et al., 2008.

## 4 Estimating sensitivity indices using geostatistical simulations

GSA was initially designed to study models with scalar inputs only. Some authors have suggested solutions to handle spatially distributed inputs as well (Volkova et al., 2008 ; Iooss & Ribatet, 2009; Ruffo et al., 2006; Lilburne & Tarantola, 2009). We describe in this section how to estimate sensitivity indices in model  $M$  by associating randomly generated realizations of uncertain 2D-field  $Z(u)$  to scalar values, according to the approach developed by Lilburne and Tarantola.

Three steps are needed to apply GSA on model  $M$  (Figure 1):

1. modelling uncertainty on model inputs  $X$  and  $Z(u)$
2. propagating input uncertainty through model  $M$
3. estimating sensitivity indices

Each step is described in details in the following subsections.

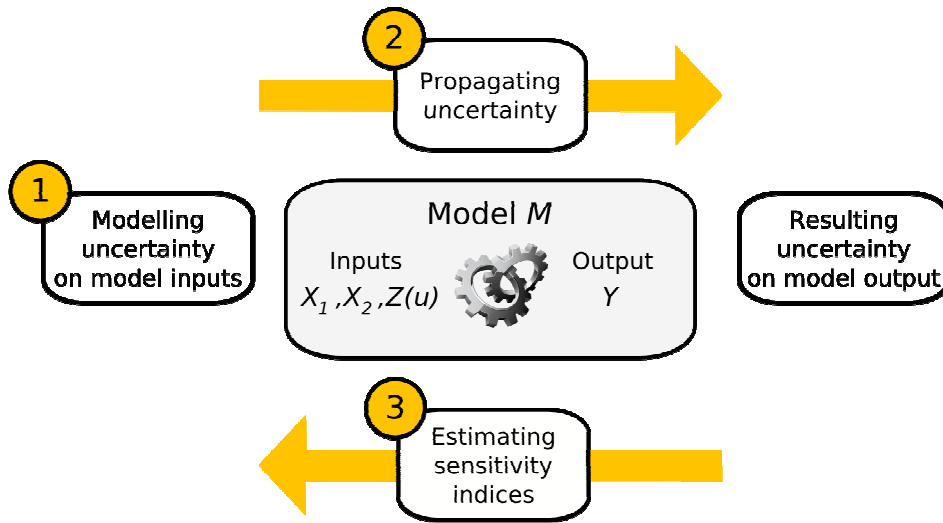


Figure 1: Steps of sensitivity analysis of model  $M$

### 4.1 Modelling uncertainty using geostatistical simulations

The values of model inputs are always partly uncertain, due to measurement errors, lack of knowledge, natural variability, modelling errors... Within the GSA method, uncertainty on model inputs is described using a probabilistic framework (Table 1).

Table 1: Specification of uncertainty on model inputs

Model input	Model of uncertainty
$X=(X_1, X_2)$	$X_1$ and $X_2$ independent random variables: $X_1 \sim \mathcal{N}(12,24)$ and $X_2 \sim \mathcal{N}(1,20)$
$Z(u)$	Gaussian Random Field of mean $\mu=1$ and covariance $\rho_\theta(h)$

#### 4.1.1 Uncertainty on model input $X$

Model input  $X=(X_1, X_2)$  is a vector of two scalar factors.  $X_1$  and  $X_2$  are supposed to be independent random variables following Gaussian distribution  $\mathcal{N}(12,24)$  and  $\mathcal{N}(1,20)$  respectively (Gaussian distribution parameters were chosen arbitrarily).

### 4.1.2 Uncertainty on model input $Z(u)$

2D-field  $Z(u)$  is supposed to be a Gaussian random field. It is assumed to be order 2 stationary with mean  $\mu=1$ . Its covariance function is denoted by  $\rho_\theta(h)$ :

$$\forall u \in D, \quad \forall h \geq 0, \quad \text{cov}[Z(u), Z(u+h)] = \rho_\theta(h)$$

For numerical application, covariance function  $\rho_\theta(h)$  is supposed to be exponential:

$$\forall h \geq 0, \quad \rho_\theta(h) = \sigma^2 \cdot \left[ \eta \cdot \delta_0(h) + (1-\eta) \cdot e^{-\frac{h}{l}} \right]$$

Parameter  $\theta = (l, \sigma^2, \eta)$  describes the covariance parameters:  $l$  is the practical range of covariance,  $\sigma^2$  the sill and  $\eta$  the nugget.

In order to represent the uncertainty on 2D-field  $Z(u)$ , a set of  $n=100$  random realizations is sampled. These random realizations are generated with Simple Random Sampling using LU decomposition of the covariance matrix (Journel and Huijbregts, 1978). These  $n$  random realizations are considered as equiprobable, and each realization is labelled with a unique integer in the set  $\{1, \dots, n\}$  (Figure 2).

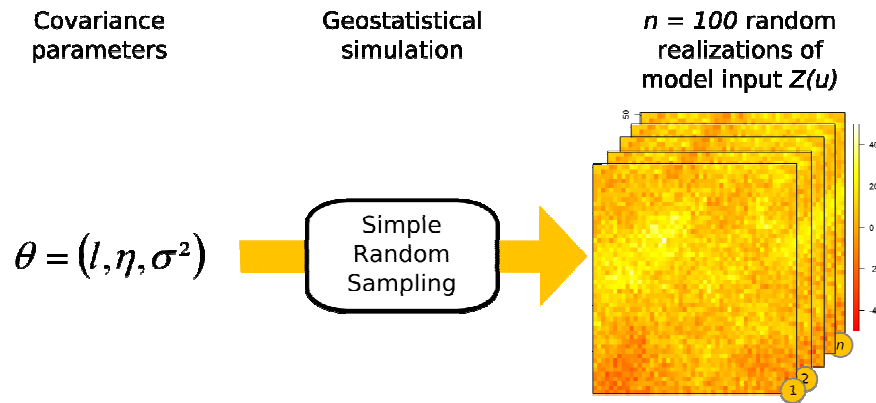


Figure 2: Modelling uncertainty on model input  $Z(u)$

## 4.2 Propagating uncertainty through model $M$

Input uncertainty is propagated through model  $M$  using a sampling-based approach, according to “spatial GSA, method (Lilburne and Tarantola, 2009).

### 4.2.1 Sampling of model inputs

Spatial GSA uses two quasi-random independent samples  $A$  and  $B$  of length  $N=4096$ , combined through several permutations, to explore the uncertainty domain of input factors  $X$  and  $Z(u)$ . The  $i^{\text{th}}$  line of sample  $A$  or  $B$  is a set  $(X_1^{(i)}, X_2^{(i)}, z^{(i)})$  where:

- $X_1^{(i)}$  is a random value drawn from pdf of input factor  $X_1$
- $X_2^{(i)}$  is a random value drawn from pdf of input factor  $X_2$
- $z^{(i)}$  is a random integer sampled from a discrete uniform distribution in  $\{1, \dots, n\}$ . Each discrete level in  $\{1, \dots, n\}$  is associated with a single random realization of  $Z(u)$  from the

set of  $n$  maps previously generated (see 4.1.2). The value of  $z^{(i)}$  indicates which random realization of  $Z(u)$  could be used to evaluate model  $M$  for the  $i^{\text{th}}$  line of the sample.

#### 4.2.2 Permutations

In order to estimate sensitivity indices for model inputs  $X=(X_1, X_2)$  and  $Z(u)$ , we must evaluate model  $M$  at points  $(X_1^{(i)}, X_2^{(i)}, z^{(i)})$  where only one of the three factors changes from a previous line  $(X_1^{(j)}, X_2^{(j)}, z^{(j)})$  where model  $M$  has already been evaluated. Thus, new samples are created by combining original samples  $A$  and  $B$ . For  $j = 1$  to 3, a new sample  $A_B^{(j)}$  is created: it is equal to sample  $A$ , except for the  $j^{\text{th}}$  column which is taken from sample  $B$  (Figure 3).

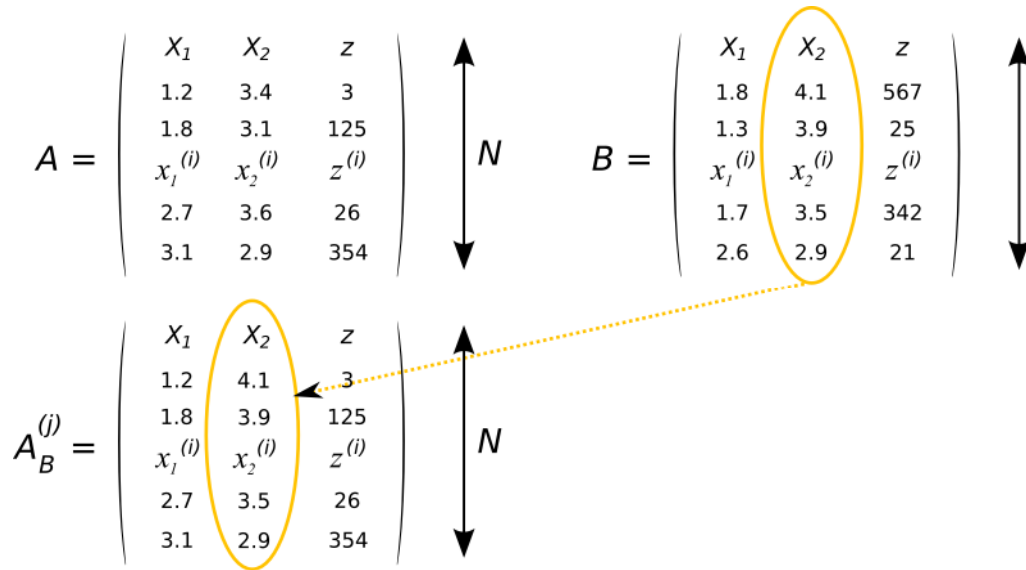


Figure 3: Creating sample  $A_B^{(j)}$

#### 4.2.3 Model runs

Model  $M$  is finally evaluated for each line of samples  $A$ ,  $B$  and  $A_B^{(j)}$  for  $j=1..3$ . Total number of model runs is  $C = 5 \cdot N = 20480$ . Each model run gives a value for the output of interest  $Y_D$ . We denote by  $Y_A$ ,  $Y_B$  and  $Y_{A_B^{(j)}}$  the vectors of length  $N$  giving the value of  $Y_D$  for each line of samples  $A$ ,  $B$  and  $A_B^{(j)}$ .

#### 4.3 Estimating sensitivity indices

First and total order sensitivity indices of the  $j^{\text{th}}$  input factor are estimated using expressions (1) and (2) given in (Saltelli et al., 2008). Model input  $X=(X_1, X_2)$  is treated as a “group of factors”; components  $X_1$  and  $X_2$  were sampled independently from their pdf, but first order and total order sensitivity indices are estimated globally for the group  $X=(X_1, X_2)$  (see section 1.2.15 of Saltelli et al., 2008 for a complete discussion on grouping model inputs in GSA).

$$S_j = \frac{\frac{1}{N} \cdot \sum_{i=1}^N Y_B^{(i)} \cdot Y_{A_B^{(j)}}^{(i)} - \frac{1}{N} \cdot \sum_{i=1}^N Y_B^{(i)} \cdot Y_A^{(i)}}{\frac{1}{N} \cdot \sum_{i=1}^N Y_A^{(i)} \cdot Y_A^{(i)} - \left( \frac{1}{N} \cdot \sum_{i=1}^N Y_A^{(i)} \right) \cdot \left( \frac{1}{N} \cdot \sum_{i=1}^N Y_B^{(i)} \right)} \quad (1)$$

$$ST_j = \frac{\frac{1}{2N} \cdot \sum_{i=1}^N \left( Y_A^{(i)} - Y_{A_B^{(j)}}^{(i)} \right)^2}{\frac{1}{N} \cdot \sum_{i=1}^N Y_A^{(i)} \cdot Y_A^{(i)} - \left( \frac{1}{N} \cdot \sum_{i=1}^N Y_A^{(i)} \right) \cdot \left( \frac{1}{N} \cdot \sum_{i=1}^N Y_B^{(i)} \right)} \quad (2)$$

We finally obtain four different sensitivity indices: first and total order sensitivity indices of model input  $X=(X_1, X_2)$ , denoted by  $S_X$  and  $ST_X$ ; first and total order sensitivity indices of model input  $Z(u)$ , denoted by  $S_Z$  and  $ST_Z$ . In the current case of a model with only two inputs ( $X$  and  $Z(u)$ ), the following properties hold:

$$ST_X = S_X + S_{X,Z} \quad \text{and} \quad ST_Z = S_Z + S_{X,Z}$$

where  $S_{X,Z} = 1 - S_X - S_Z$  is a second order sensitivity index which accounts for the contribution of the interaction between  $X$  and  $Z(u)$  to the variance of model output  $Y_D$ . Thus, we will only pay attention in the following sections to first order indices  $S_X$  and  $S_Z$ .

## 5 Influence of covariance parameters on sensitivity indices

In this section, we want to assess how GSA results on model  $M$  are influenced by covariance parameters  $(l, \sigma^2, \eta)$ . 26 different sets  $\theta_k = (l_k, \sigma_k^2, \eta_k)$  of covariance range, sill and nugget are defined (Table 2). For each set  $\theta_k$  of covariance parameters, GSA is performed as follows:

- a set of  $n=100$  random realizations of input random field  $Z(u)$  is generated using geostatistical simulation as described in 4.1
- uncertainty is propagated through model  $M$  as described in 4.2
- total variance of model output  $Y_D$  is computed
- first order sensitivity indices  $S_X$  and  $S_Z$  are estimated as described in 4.3.

The whole procedure is replicated 100 times. Then, for each set of covariance parameters, mean value of  $Var(Y_D)$ ,  $S_X$  and  $S_Z$  and their 95% confidence interval over the 100 replicas are computed.

**Table 2: Sets of covariance parameters**

Set name	Covariance parameters		
	Range $l$	Sill $\sigma^2$ (square root)	Nugget $\eta$
$\theta_1$ to $\theta_8$	5 to 40 (step 5)	70	0.1
$\theta_9$ to $\theta_{16}$	60	20 to 55 (step 5)	0.1
$\theta_{17}$ to $\theta_{26}$	60	70	0.1 to 1 (step 0.1)



## 5.1 Influence of the ratio covariance range $l$ / size of domain $D$

Fig 4. shows output variance  $Var(Y_D)$  and sensitivity indices  $S_X$  and  $S_Z$  for increasing covariance range  $l$  (sets  $\theta_1$  to  $\theta_8$ ). It appears that the absolute contribution of model input  $Z(u)$  to total output variance  $Var(Y_D)$  increases with covariance range  $l$ , while absolute contribution of model input  $X$  remains constant. Accordingly, sensitivity index of model input  $Z(u)$  increases with covariance range  $l$ , while sensitivity index of  $X$  decreases when covariance range  $l$  increases.

Let define the ratio  $r$  of covariance range  $l$  compared to the size of domain  $D$ :  $r = l/|D|$ . This numerical case-study illustrates the following property: the larger the ratio  $r$ , the larger the part of output variance  $Var(Y_D)$  explained by the uncertainty on  $Z(u)$ . For a low ratio (i.e. when range  $l$  is small compared to the size of domain  $D$ ), variability of  $Z(u)$  is mainly “local”, and spatial correlation of  $Z(u)$  variability over domain  $D$  is weak. This “local” variability averages over domain  $D$  when model output  $Y_D$  is computed. Thus the uncertainty on input 2D-field  $Z(u)$  has a small influence on output variance  $Var(Y_D)$ .

On the contrary, for a greater ratio  $r$  (i.e. when range  $l$  is large compared to the size of domain  $D$ ), spatial correlation of  $Z(u)$  variability over domain  $D$  is strong. The averaging effect of “local” variability of  $Z(u)$  over domain  $D$  is weaker. Thus the uncertainty on input 2D-field  $Z(u)$  has a larger influence on output variance  $Var(Y_D)$ .

## 5.2 Influence of covariance sill

Fig 5. shows output variance  $Var(Y_D)$  and sensitivity indices  $S_X$  and  $S_Z$  for increasing covariance sill  $\sigma^2$  (sets  $\theta_9$  to  $\theta_{16}$ ). It appears that the absolute contribution of model input  $Z(u)$  to total output variance  $Var(Y_D)$  increases with covariance sill  $\sigma^2$ , while absolute contribution of model input  $X$  remains constant. Accordingly, sensitivity index of model input  $Z(u)$  increases with covariance sill  $\sigma^2$ , while sensitivity index of  $X$  decreases when covariance sill  $\sigma^2$  increases.

This numerical case-study illustrates the following straightforward property: the larger the covariance sill  $\sigma^2$  in random field  $Z(u)$ , the larger the part of output variance  $Var(Y_D)$  explained by the uncertainty on  $Z(u)$ . Covariance sill  $\sigma^2$  controls the overall variability of model input  $Z(u)$ , thus sensitivity index of  $Z(u)$  with respect to model output  $Y_D$  is a monotonically increasing function of sill  $\sigma^2$ .

## 5.3 Influence of covariance nugget

Fig 6. shows output variance  $Var(Y_D)$  and sensitivity indices  $S_X$  and  $S_Z$  for increasing covariance nugget  $\eta$  (sets  $\theta_{17}$  to  $\theta_{26}$ ). It appears that the absolute contribution of model input  $Z(u)$  to total output variance  $Var(Y_D)$  decreases when covariance nugget  $\eta$  increases, while absolute contribution of model input  $X$  remains constant. Accordingly, sensitivity index of model input  $Z(u)$  decreases when covariance nugget  $\eta$  increases.

Nugget parameter  $\eta$  controls the intensity of “noise” in Gaussian random field  $Z(u)$ . When  $\eta$  is close to 1, the largest part of  $Z(u)$  variability is due to the “nugget effect”, i.e. to “local” noise at each point  $u \in D$  with no spatial correlation. This local noise averages over domain  $D$  when model output  $Y_D$  is computed. Thus the uncertainty on input 2D-field  $Z(u)$  has a small influence on output variance  $Var(Y_D)$ . On the contrary, for a lower value of nugget parameter, ( $\eta$  close to 0), most of the uncertainty in random field  $Z(u)$  is spatially correlated, and local noise plays a small part. The averaging effect of uncorrelated variability of  $Z(u)$  over domain  $D$  is weaker. Thus the uncertainty on input 2D-field  $Z(u)$  has a larger influence on output variance  $Var(Y_D)$ .

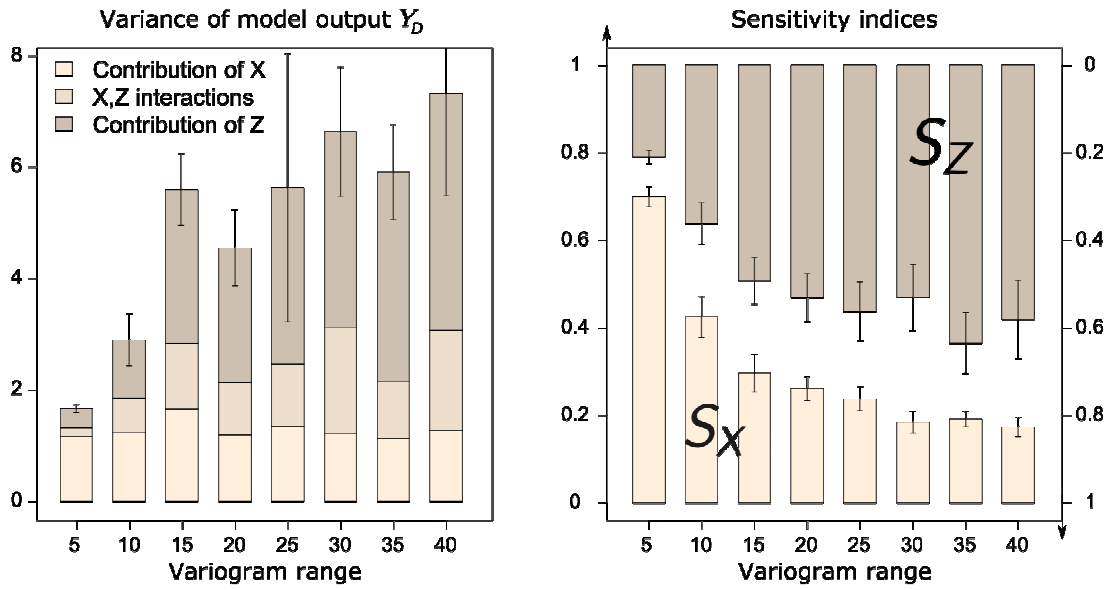


Figure 4: Influence of covariance range  $l$  on GSA results. (left) Total variance of model output  $Y_D$  and contribution of model inputs. (right) First order sensitivity indices  $S_X$  and  $S_Z$ . (error bars show 95% confidence interval over 100 replicas).

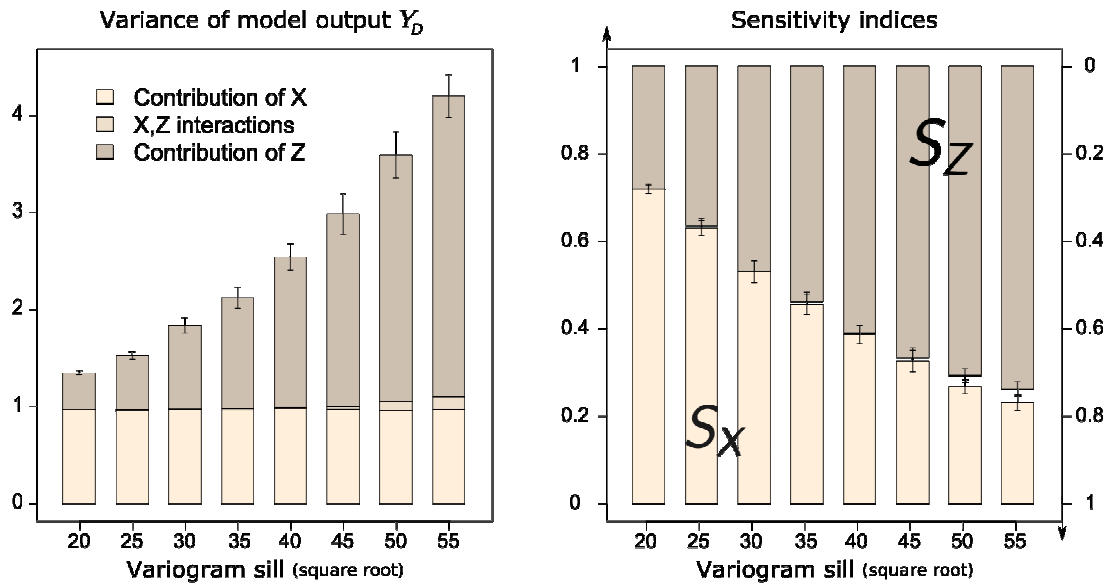
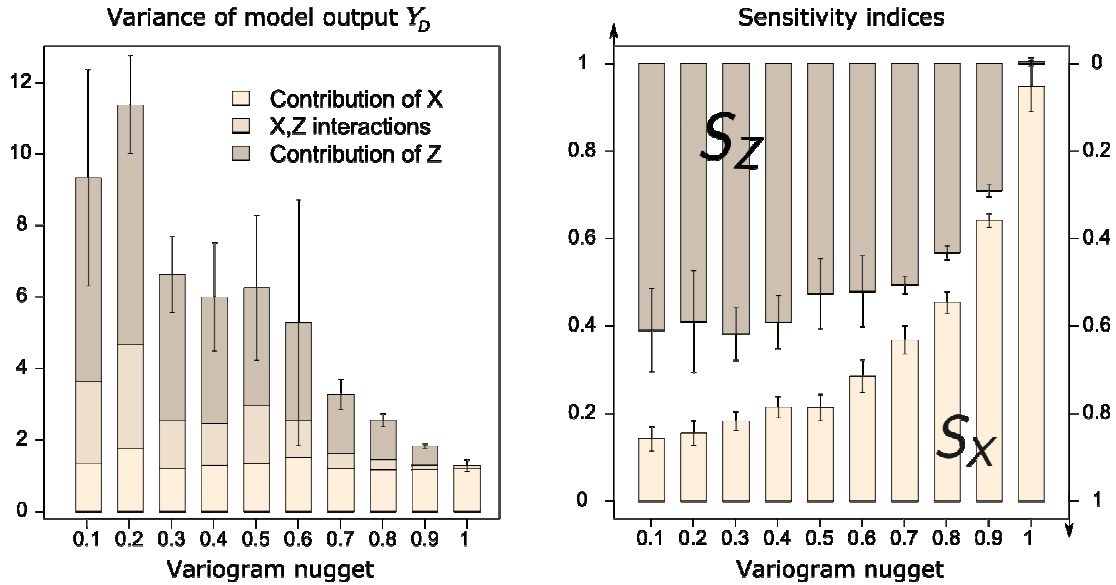


Figure 5: Influence of covariance sill  $\sigma^2$  on GSA results. (left) Total variance of model output  $Y_D$  and contribution of model inputs. (right) First order sensitivity indices  $S_X$  and  $S_Z$  (error bars show 95% confidence interval over 100 replicas).



**Figure 6: Influence of covariance nugget  $\eta$  on GSA results. (left) Total variance of model output  $Y_D$  and contribution of model inputs. (right) First order sensitivity indices  $S_X$  and  $S_Z$  (error bars show 95% confidence interval over 100 replicas).**

## 6 Discussion

This research sought to illustrate on a simple case-study how to use geostatistical simulation to perform variance-based global sensitivity analysis (GSA) on a spatially distributed model. We also aimed at exploring how GSA results depend on covariance parameters chosen to describe uncertainty on spatially distributed model inputs.

### 6.1 Using geostatistical simulation for spatial GSA

We demonstrated on a simple case-study the suitability of “spatial GSA” approach (Lilburne & Tarantola, 2009) to perform sensitivity analysis on a spatially distributed model with continuous 2D-fields inputs. Geostatistical simulation was used to generate a set of  $n$  random realizations of continuous 2D-field  $Z(u)$  and estimate sensitivity indices of uncertain model inputs  $Z(u)$  and  $X$  through a sampling-based approach. Spatial GSA makes it possible to account for the relative contribution of each uncertain model input to the total variance of model output. It helps assessing model robustness and should be systematically performed when developing a model with uncertain spatial inputs. Nevertheless, two limits of this approach must be highlighted:

- spatial GSA is a sampling-based approach which needs lots of model runs to estimate sensitivity indices. As a consequence, it is limited to models with low CPU-cost. For high CPU-cost models, other sensitivity analysis methods such as Elementary Effects or One-At-a-Time should be applied (see Saltelli et al., 2008).
- spatial GSA uses a set of  $n$  random realizations to represent the uncertainty on spatial input  $Z(u)$  (assumed to be a Gaussian Random Field). When  $n$  is too low, the small set of map simulations fails to capture the overall variability of  $Z(u)$ , and sensitivity indices estimates  $S_X$  and  $S_Z$  are biased. Previous work had been carried out to compare the use of two different geostatistical simulation algorithms (Simple Random Sampling and Latin Hypercube Sampling) to generate realizations of spatial input  $Z(u)$  for GSA (Kyriakidis, 2005; Saint-Geours et al., 2010), but no optimal sampling strategy was found to reduce this bias.

## 6.2 Impact of spatial covariance parameters on spatial GSA

The influence of covariance range, sill and nugget on sensitivity indices was assessed on a simple case-study. It was initially suggested that covariance parameters chosen to describe uncertainty on spatial input  $Z(u)$  would influence GSA results. Our results prove such to be the case. On our case-study, it appears that first order sensitivity index  $S_Z$  of model input  $Z(u)$  is a monotonically increasing function of both covariance range  $l$  and covariance sill  $\sigma^2$ , and a decreasing function of covariance nugget  $\eta$ .

These properties were only illustrated on a simple case-study with a specific model  $M$  and an exponential covariance function. Nevertheless, it can be analytically shown (on-going work) that these properties are actually verified for any monotonically increasing covariance function and for any point-based model  $M$  where mapping  $f$  is square-integrable.

Results of this study may well help modellers when estimating spatial covariance parameters to describe uncertainty on a spatial input  $Z(u)$  for sensitivity analysis of a spatially distributed model. When field data is lacking to carefully estimate covariance parameters, at least the *a-priori* impact of giving wrong values to these parameters will be known: over-estimating covariance range  $l$  or covariance sill  $\sigma^2$  will result in over-estimating sensitivity indices of spatial input  $Z(u)$  and under-estimating sensitivity indices of scalar inputs  $X_i$ . On the contrary, over-estimating covariance nugget  $\eta$  will result in under-estimating sensitivity indices of  $Z(u)$ .

## 7 Conclusion

Variance-based global sensitivity analysis (GSA) was performed on a simple example of a spatially distributed model  $Y=M(X,Z)$  with two inputs: a scalar input  $X$  and a spatial input  $Z(u)$ . In order to represent the variability on uncertain spatial input  $Z(u)$ , it was assumed to be a Gaussian Random Field, and random realizations were generated using geostatistical simulation. These random realizations were used to propagate input uncertainty through model  $M$ . Sensitivity indices of model inputs  $X$  and  $Z(u)$  were estimated with a sampling-based approach. The influence of spatial covariance parameters on GSA results was assessed by estimating sensitivity indices for different sets of covariance range, sill and nugget.

Results show that (1) first order sensitivity index  $S_Z$  of spatial input  $Z(u)$  is a monotonically increasing function of covariance range  $l$  (2) first order sensitivity index  $S_Z$  of spatial input  $Z(u)$  is a monotonically increasing function of covariance sill  $\sigma^2$  (3) first order sensitivity index  $S_Z$  of spatial input  $Z(u)$  is a monotonically decreasing function of covariance nugget  $\eta$ .

These empirical results may be of importance when setting covariance parameters to describe uncertainty in spatial inputs for sensitivity analysis of a spatial model. Yet further research is needed to prove analytically that these properties hold for a large range of point-based models and monotonic covariance functions. Such study may help promoting the use of geostatistical simulation to perform sensitivity analysis of spatially distributed models.

## References

- AERTS, J. C. J. H., HEUVELINK, G. B. M., GOODCHILD, M. F. (2003): Accounting for spatial uncertainty in optimization with spatial decision support systems. *Transactions in GIS*, Vol. 7, 211-230
- HELTON, J. C., DAVIS, F. J. (2006): Sampling-based methods for uncertainty and sensitivity analysis. *Multimedia Environmental Models*, Vol. 32, 135-154
- IOOSS, B., RIBATET, M. (2009): Global sensitivity analysis of computer models with functional inputs. *Reliability Engineering and System Safety*, Vol. 94, 1194 - 1204
- JOURNAL, A.G., HUIJBREGTS, C.J. (1978): *Mining geostatistics*. The Blackburn Press
- KYRIAKIDIS, P. C. (2005): Sequential spatial simulation using Latin Hypercube Sampling. *Geostatistics Banff 2004*, 65-74
- LILBURNE, L., TARANTOLA, S. (2009): Sensitivity analysis of spatial models. *International Journal of Geographical Information Science*, Vol. 23, No. 2, 151-168
- REFSGAARD, J. C., VAN DER SLUIJS, J. P., HØJBERG, A. L., VANROLLEGHEM, P. A. (2007): Uncertainty in the environmental modelling process: a framework and guidance. *Environmental Modelling and Software*, Vol. 22, 1543-1556
- RUFFO, P., BAZZANA, L., CONSONNI, A., CORRADI, A., SALTELLI, A., TARANTOLA, S. (2006): Hydrocarbon exploration risk evaluation through uncertainty and sensitivity analyses techniques. *Reliability Engineering and System Safety*, Vol. 91, 1155-1162
- SAINT-GEOURS, N., BAILLY, J.-S., GRELOT, F., LAVERGNE, C. (2010): Is there room to optimise the use of geostatistical simulations for sensitivity analysis of spatially distributed models ? *Accuracy 2010 : proceedings of the ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, 81-84
- SALTELLI, A., RATTO, M., ANDRES, T., CAMPOLONGO, F., CARIBONI, J., GATELLI, D., SAISANA, M., TARANTOLA, S. (2008): *Global Sensitivity Analysis - The Primer*. Wiley, New-York.
- VOLKOVA, E., IOOSS, B., VAN DORPE, F. (2008): Global sensitivity analysis for a numerical model of radionuclide migration from the RRC Kurchatov Institute radwaste disposal site. *Stochastic Environmental Research and Risk Assessment*, Vol. 22, 17-31