# Change of support effects in spatial variance-based sensitivity analysis

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

# Mathematical Geosciences

## Change of support in spatial variance-based sensitivity analysis

--Manuscript Draft--

| | |
|---|---|
| **Manuscript Number:** | MATG731R1 |
| **Full Title:** | Change of support in spatial variance-based sensitivity analysis |
| **Article Type:** | Original Research |
| **Keywords:** | sensitivity analysis; Sobol' indices; model upscaling; change of support; regularization theory; spatial model |
| **Corresponding Author:** | Nathalie Saint-Geours, Ph.D.<br>AgroParisTech - UMR TETIS<br>MONTPELLIER CEDEX 5, FRANCE |
| **Corresponding Author Secondary Information:** | |
| **Corresponding Author's Institution:** | AgroParisTech - UMR TETIS |
| **Corresponding Author's Secondary Institution:** | |
| **First Author:** | Nathalie Saint-Geours, Ph.D. |
| **First Author Secondary Information:** | |
| **Order of Authors:** | Nathalie Saint-Geours, Ph.D. |
| | Christian Lavergne, Professor |
| | Jean-Stéphane Bailly, Professor |
| | Frédéric Grelot, Researcher |
| **Order of Authors Secondary Information:** | |
| **Abstract:** | E-mail address: saintge@teledetection.fr<br><br>Variance-based global sensitivity analysis (GSA) is used to study how the variance of the output of a model can be apportioned to different sources of uncertainty in its inputs. GSA is an essential component of model building: it helps to identify model inputs that account for most of the model output variance. However, this approach is seldom applied to spatial models because it cannot describe how uncertainty interacts with another key issue in spatial modeling: the issue of model upscaling, i.e., a change of spatial support of model output. In many environmental models, the end user is interested in the spatial average or the sum of the model output over a given spatial unit (e.g., the average porosity of a geological block). Under a change of spatial support, the relative contribution of uncertain model inputs to the variance of aggregated model output may change. We propose a simple formalism to discuss this issue within a GSA framework by defining point and block sensitivity indices. We show that the relative contribution of an uncertain spatially distributed model input increases with its covariance range and decreases with the size of the spatial unit considered for model output aggregation. The results are briefly illustrated by a simple example. |

# Change of support in spatial variance-based sensitivity analysis[1]

by Nathalie Saint-Geours[2,3], Christian Lavergne[2], Jean-Stéphane Bailly[3] and Frédéric Grelot[4]

[1] Received ; accepted .

[2] I3M, Université Montpellier 3, Montpellier, France

[3] UMR TETIS, AgroParisTech, Montpellier, France; saintge@teledetection.fr

[4] UMR G-EAU, Irstea, Montpellier, France

Corresponding Author:

Nathalie Saint-Geours

UMR TETIS

AgroParisTech - Université Montpellier 2

Montpellier, 34090 FRANCE

Phone + 33 4 67 54 87 45

fax + 33 803 777 6610

e-mail saintge@teledetection.fr

**Abstract** Variance-based global sensitivity analysis (GSA) is used to study how the variance of the output of a model can be apportioned to different sources of uncertainty in its inputs. GSA is an essential component of model building: it helps to identify model inputs that account for most of the model output variance. However, this approach is seldom applied to spatial models because it cannot describe how uncertainty propagation interacts with another key issue in spatial modeling: the issue of model upscaling, i.e., a change of spatial support of model output. In many environmental models, the end user is interested in the spatial average or the sum of the model output over a given spatial unit (e.g., the average porosity of a geological block). Under a change of spatial support, the relative contribution of uncertain model inputs to the variance of aggregated model output may change. We propose a simple formalism to discuss this issue within a GSA framework by defining *point* and *block* sensitivity indices. We show that the relative contribution of an uncertain spatially distributed model input increases with its correlation length and decreases with the size of the spatial unit considered for model output aggregation. The results are briefly illustrated by a simple example.

**Keywords** Sensitivity analysis · Sobol' indices · Model upscaling · Change of support · Regularization theory · Spatial model

## 1 Introduction

Variance-based global sensitivity analysis (GSA) is used to study how the variance of the output of a model can be apportioned to different sources of uncertainty in its inputs. Here, the term *model* denotes any computer code in which a response variable is calculated as a deterministic function of input variables. Originally developed in the 1990s (Sobol' 1993), GSA is now recognized as an essential component of model building (European Commission 2009; US Environmental Protection Agency 2009) and is widely used in different fields (Cariboni et al. 2007; Tarantola et al. 2002). GSA is based on the decomposition of a model output variance into conditional variances. So-called *first-order sensitivity indices* measure the main effect contribution of each uncertain model input to the model output variance. Based on these sensitivity indices, ranking the model inputs helps to identify inputs that should be better scrutinized first. Reducing the uncertainty on the inputs with the largest sensitivity indices (e.g., by collecting additional data or changing the geographical pattern of data locations) will often result in a reduction in the variance of the model output. More generally, GSA helps to explore the response surface of a *black box* computer code and to prioritize the possibly numerous processes that are involved in it.

Although GSA was initially designed for models where both inputs and output can be described as real valued random variables, some recent work has extended GSA to environmental models for which both the inputs and output are spatially distributed over a two-dimensional domain and can be described as random fields (Lilburne and Tarantola 2009, for a review). In these works, the computer code under study uses maps derived from field data (e.g., digital elevation models and land use maps). These maps are uncertain due to measurement errors, lack of knowledge or aleatory variability (Brown and Heuvelink 2007; Refsgaard et al. 2007). The uncertainty of these spatial inputs is usually modeled using random fields. Model output is also spatially distributed (e.g., a flood map or a pollution map). Authors use geostatistical simulation to incorporate spatially distributed model inputs into the GSA approach (Ruffo et al. 2006; Saint-Geours et al. 2010) and they display estimation procedures to compute sensitivity indices in a spatial context, either with respect to the spatial average of the model output (Lilburne and Tarantola 2009) or with respect to the values of the model output at each site of a study area (Heuvelink et al. 2010; Marrel et al. 2011; Pettit and Wilson 2010).

Nevertheless, to date, none of these studies has reported on a key issue: the link between uncertainty propagation and model upscaling/downscaling.

*Model upscaling* is the problem of translating knowledge from smaller scales to larger (Heuvelink 1998). In many environmental models, the physical quantities considered are spatially additive (e.g., porosity or evapotranspiration), i.e., their large-scale properties derive from small-scale properties by simple averaging (Chilès and Delfiner 1999, p.593). In this case, the model end user is usually interested in the spatial linear average or the sum of spatial output over a given spatial unit (e.g., the average porosity of a block or the total evapotranspiration over a plot of land) and model upscaling is thus reduced to a *change of support* problem (namely, a change of support of the end user's output of interest). Heuvelink (1998) observed that under a change of spatial support of the model output, the relative contribution of uncertain model inputs to the variance of the aggregated model output may change. Exploring how sensitivity analysis results interact with such a change of support is thus of great importance. It would allow the modeler to check the robustness of model-based environmental impact assessment studies and better assess the confidence of their results. Knowledge of this interaction would also allow the modeler to answer the following questions: What are the model inputs that explain the largest fraction of the variance of the output over a given spatial support? For which output support size does a given spatially distributed model input contribute to the largest fraction of the variance of the model output? How does the contribution of a spatially distributed input to the variance of the model output depend on the parameters of its covariance function?

The change of support effect has been extensively discussed in geostatistics in the context of regularization theory (Journel and Huijbregts 1978, p.77). Hence, we attempt in this paper to integrate regularization theory with variance-based GSA framework. Our idea is to define *site sensitivity indices* and *block sensitivity indices* to i) provide a simple formalism that extends variance-based GSA to spatial models when the modeler's interest is in the spatial average or the sum of model output over a given spatial support (Sect. 2) and ii) discuss how the relative contribution of uncertain model inputs to the variance of model output changes under model upscaling (Sect. 3). We limit our study to point-based models, i.e., models for which the computation of the model output at some location uses the values of spatial inputs at that same location only (Heuvelink, Brus, and Reinds 2010). An example is used throughout this paper to illustrate formal definitions and properties. Finally, we discuss the limits of our approach and its connections to related works in Sect. 4.

## 2 Variance-based sensitivity indices for a spatial model

### 2.1 Description of spatial model $\mathcal{M}$

We want to study a computer code $\mathcal{M}$ whose output is a map and whose inputs are a map and a set of $n$ real valued variables. Both inputs and output are *uncertain* and are described as random variables or random fields. More precisely, we use the following notations: let $\mathcal{D} \subset \mathbb{R}^2$ denote a 2D spatial domain, $\mathbf{x} \in \mathcal{D}$ a site, $\mathbf{h}$ the lag vector between two sites $\mathbf{x}$ and $\mathbf{x}'$, and $v \subset \mathcal{D}$ some spatial support (block) of area $|v|$. We consider the model $Y = \mathcal{M}(\mathbf{U}, Z)$ where $\mathbf{U} = (U_1, \ldots, U_n)$ is a random vector and $\{Z(\mathbf{x}) : \mathbf{x} \in \mathcal{D}\}$ is a second-order stationary random field (SRF) — that we will often simply denote by $Z(\mathbf{x})$. $\mathbf{U}$ and $Z(\mathbf{x})$ are supposed to be independent. Covariance function $C(\cdot)$ of $Z(\mathbf{x})$ is assumed to be isotropic, characterized by correlation length $a \in \mathbb{R}$, nugget parameter $\eta \in [0; 1[$ and of the form:

$$C(\mathbf{h}) = \begin{cases} C(0) & \text{if } \mathbf{h} = \mathbf{0} \\ (1-\eta) \cdot C(0) \cdot \rho_a(\|\mathbf{h}\|) & \text{if } \mathbf{h} \neq \mathbf{0} \end{cases} \tag{1}$$

where $\rho_a(\cdot)$ is some valid correlogram (Cressie 1993, p.67). The model output is a 2D random field $\{Y(\mathbf{x}) : \mathbf{x} \in \mathcal{D}\}$ that we will simply denote by $Y(\mathbf{x})$. We assume that the first two moments of $Y(\mathbf{x})$ exist. Finally, as discussed in the introduction, we limit our study to *point-based* models; hence, we assume that there exists a mapping $\psi : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ such that:

$$\forall \mathbf{x} \in \mathcal{D}\,, \quad Y(\mathbf{x}) = \psi\left[\mathbf{U}, Z(\mathbf{x})\right] \tag{2}$$

A sensitivity analysis of the model $\mathcal{M}$ must be performed with respect to a scalar quantity of interest derived from spatially distributed model output $Y(\mathbf{x})$. Here, we consider two different outputs of interest: the value $Y(\mathbf{x}^*)$ at some specific site $\mathbf{x}^* \in \mathcal{D}$ and the aggregated value $Y_v = 1/|v| \int_v Y(\mathbf{x})\, d\mathbf{x}$ over support $v$. Because model inputs $\mathbf{U}$ and $Z(\mathbf{x})$ are uncertain, $Y(\mathbf{x}^*)$ and $Y_v$ are both random variables; the sensitivity analysis will describe the relative contribution of uncertain model inputs $\mathbf{U}$ and $Z(\mathbf{x})$ to their respective variances.

### 2.2 Site sensitivity indices and block sensitivity indices

Before defining sensitivity indices for spatial model $\mathcal{M}$, we briefly review the mathematical basis of variance-based GSA. Let us consider a model $Y = G(X_1, \ldots, X_n)$, where $X_i$ are independent random variables and where the first two moments of $Y$ exist. The first-order sensitivity index $S_i$ of model input $X_i$ is defined by:

$$S_i = \frac{\mathrm{Var}\left[\mathbb{E}(Y|X_i)\right]}{\mathrm{Var}\,(Y)} \tag{3}$$

$S_i \in [0;1]$ measures the main effect contribution of the uncertain model input $X_i$ to the variance of model output $Y$. Sensitivity indices can be used to identify the model inputs that account for most of the variance of the model output (model inputs $X_i$ with high first-order indices $S_i$). Sum of $S_i$ is always less than 1 and the difference $1 - \sum_i S_i$ accounts for the contribution of the interactions between model inputs $X_i$ to model output variance $\mathrm{Var}(Y)$. Please refer to Saltelli et al. (2008) for more details on GSA theory and on the estimation of sensitivity indices.

To extend GSA to spatial model $\mathcal{M}$, we propose to use different types of sensitivity indices to describe the relative contribution of the uncertain model inputs $\mathbf{U}$ and $Z(\mathbf{x})$ to the variance of the model output: an index on a point support (i.e., with respect to output of interest $Y(\mathbf{x}^*)$) and an index on a larger support (i.e., with respect to output of interest $Y_v$). First-order sensitivity indices of model inputs with respect to $Y(\mathbf{x}^*)$ are called *site sensitivity indices*. Under the stationary hypothesis on SRF $Z(\mathbf{x})$, these indices do not depend on site $\mathbf{x}^*$ and thus will simply be denoted by $S_{\mathbf{U}}$ and $S_Z$:

$$S_{\mathbf{U}} = \frac{\mathrm{Var}\left[\mathbb{E}(Y(\mathbf{x}^*) \mid \mathbf{U})\right]}{\mathrm{Var}\left[Y(\mathbf{x}^*)\right]} \quad ; \quad S_Z = \frac{\mathrm{Var}\left[\mathbb{E}(Y(\mathbf{x}^*) \mid Z(\mathbf{x}))\right]}{\mathrm{Var}\left[Y(\mathbf{x}^*)\right]} \tag{4}$$

First-order sensitivity indices of model inputs with respect to the block average $Y_v$ are called *block sensitivity indices* and are denoted by $S_{\mathbf{U}}(v)$ and

$S_Z(v)$:

$$S_{\mathbf{U}}(v) = \frac{\mathrm{Var}\left[\mathbb{E}(Y_v \mid \mathbf{U})\right]}{\mathrm{Var}\left[Y_v\right]} \quad ; \quad S_Z(v) = \frac{\mathrm{Var}\left[\mathbb{E}(Y_v \mid Z(\mathbf{x}))\right]}{\mathrm{Var}\left[Y_v\right]} \tag{5}$$

The ratio $S_Z(v)/S_{\mathbf{U}}(v)$ gives the relative contribution of model inputs $Z(\mathbf{x})$ and $\mathbf{U}$ to the variance of the output of interest $Y_v$. When $S_Z(v)/S_{\mathbf{U}}(v)$ is greater than 1, the variance of $Y_v$ is mainly explained by the variability of the 2D input field $Z(\mathbf{x})$; when $S_Z(v)/S_{\mathbf{U}}(v)$ is less than 1, it is the non spatial input $\mathbf{U}$ that accounts for most of $\mathrm{Var}\,(Y_v)$.

### 2.3 Illustrative example

The proposed formalism for spatial GSA is illustrated by the following example. A model $Y = \mathcal{M}\,(\mathbf{U}, Z)$ is used for the economic assessment of flood risk over a given floodplain $\mathcal{D}$. $Z(\mathbf{x})$ is the map of maximal water levels (m) reached during a flood event. $Z(\mathbf{x})$ is assumed to be a Gaussian random field with mean $\mu = 50$ and exponential covariance $C(\mathbf{h})$ with $C(0) = 100$, correlation length $a = 5$ and nugget parameter $\eta = 0.1$. $\mathbf{U}$ is a set of three economic parameters $U_1$, $U_2$ and $U_3$ that determine a so-called *damage function* that links water levels to monetary costs. $U_1$, $U_2$ and $U_3$ are assumed to be independent random variables following Gaussian distributions $\mathcal{N}(1.5, 0.5)$, $\mathcal{N}(55, 5)$ and $\mathcal{N}(10, 10)$, respectively. Random field $Z(\mathbf{x})$ and random vector $\mathbf{U}$ are supposed to be independent. Model output $Y(\mathbf{x})$ is the map of expected economic damages due to the flood over the area; these damages depend on $\mathbf{U}$ and $Z(\mathbf{x})$ through the mapping $\psi$:

$$\forall \mathbf{x} \in \mathcal{D}, \quad Y(\mathbf{x}) = \psi\left[\mathbf{U}, Z(\mathbf{x})\right] = U_1 \cdot Z(\mathbf{x}) - U_2 \cdot e^{-0.036 \cdot Z(\mathbf{x})} - U_3 \tag{6}$$

Stakeholders are interested in two outputs: the flood damage $Y(\mathbf{x}^*)$ on a specific building $\mathbf{x}^* \in \mathcal{D}$ and the total damage $|v| \cdot Y_v$ over a district $v$ (here, a disc of radius $r = 50$). Here, the expression of mapping $\psi$ and the statistical characterization of model inputs may be simple enough that exact values of sensitivity indices could be derived, but this is usually not the case in real applications in which the model is very complex. A usual alternative is to consider model $\mathcal{M}$ as a *black box* and estimate sensivity indices with Monte-Carlo simulation. We chose to use the estimators and the computational procedure described by Lilburne and Tarantola (2009, Sect. 3.2), based on a quasi-random sampling design, using $N = 4096$ model runs (Table 1). It appears that on a given site $\mathbf{x}^*$, the variability of the water level map explains most of the variance of $Y(\mathbf{x}^*)$: $S_Z = 0.89$. On a larger spatial support, the variance of the total flood damage $|v| \cdot Y_v$ is mainly due to the economic parameters $U_1$, $U_2$ and $U_3$: $S_{\mathbf{U}}(v) = 0.86$. Thus, to improve the accuracy of damage estimation for a specific building, the uncertainty should first be reduced on the water level map $Z(\mathbf{x})$; however, to improve the accuracy of total damage estimation over a large district $v$, the modeler should focus on reducing the uncertainty of economic parameters $U_1$, $U_2$ and $U_3$.

## 3 Change of support effect on block sensitivity indices

In this section, we assess how the ranking of uncertain model inputs based on their block sensitivity indices vary under a change of support $v$ of model output.

### 3.1 Relation between site sensitivity indices and block sensitivity indices

Site sensitivity indices and block sensitivity indices are related. Let $\mathbb{E}_Z Y(\mathbf{x})$ denote the conditional expectation of $Y(\mathbf{x})$ given $Z(\mathbf{x})$, that is:

$$\forall \mathbf{x} \in \mathcal{D}, \quad \mathbb{E}_Z Y(\mathbf{x}) = \mathbb{E}\left[Y(\mathbf{x}) \mid Z(\mathbf{x})\right] \tag{7}$$

$\mathbb{E}_Z Y(\mathbf{x})$ is the transform of the input SRF $Z(\mathbf{x})$ via the function $\bar{\psi}(z) = \int_{\mathbb{R}^n} \psi(\mathbf{u}, z) f_U(\mathbf{u}) d\mathbf{u}$ (Eq. (2)) where $f_U(\cdot)$ is the multivariate pdf of random vector $\mathbf{U}$. Under our assumptions concerning $Y(\mathbf{x})$, $\mathbb{E}_Z Y(\mathbf{x})$ is a second-order SRF. Let $C^*(\cdot)$ denote its covariance function, $\sigma^2 = C^*(0)$ its variance and $\sigma_v^2$ its block variance over support $v$, that is, the variance of block average $1/|v| \int_v \mathbb{E}_Z Y(\mathbf{x}) d\mathbf{x}$. Block variance $\sigma_v^2$ is equal to the mean value of $C^*(\mathbf{h})$ when the two extremities of lag vector $\mathbf{h}$ describe support $v$, which we denote by $\overline{C^*}(v, v)$ (Journel and Huijbregts 1978, p.78). Using these notations, it follows from Eq. (4) and (5) that site sensitivity indices and block sensitivity indices are related by (see Appendix A for a proof):

$$\frac{S_Z(v)}{S_{\mathbf{U}}(v)} = \frac{S_Z}{S_{\mathbf{U}}} \cdot \frac{\sigma_v^2}{\sigma^2} = \frac{S_Z}{S_{\mathbf{U}}} \cdot \frac{\overline{C^*}(v,v)}{C^*(0)} \tag{8}$$

### 3.2 Change of support effect

Consider now that model $\mathcal{M}$ was initially developed to study the spatial average $Y_v$ over the support $v$, and that after model upscaling the modeler is interested in the spatial average $Y_V$ over the support $V$, where $V \gg v$. We know from Krige's relation (Journel and Huijbregts 1978, p.67) that the block variance $\sigma_v^2$ decreases with increasing size of support: $\sigma_V^2 \leq \sigma_v^2$. It follows from Eq. (8) that:

$$\frac{S_Z(V)}{S_{\mathbf{U}}(V)} \leq \frac{S_Z(v)}{S_{\mathbf{U}}(v)} \tag{9}$$

The fraction of the variance of the aggregated model output explained by the input random field $Z(\mathbf{x})$—compared to the fraction explained by $\mathbf{U}$—is thus smaller on support $V$ than on support $v$. More specifically, let us suppose that the covariance function $C^*(\cdot)$ of the random field $\mathbb{E}_Z Y(\mathbf{x})$ has a finite effective range and that the support $v$ is large with respect to this range. To a first approximation, the block variance $\sigma_v^2$ is of the form $\sigma_v^2 \simeq \sigma^2 A/|v|$, where $A$ is the so-called *integral range* of $C^*(\cdot)$ and is defined by

$A = 1/\sigma^2 \int C^*(\mathbf{h})d\mathbf{h}$ (Chilès and Delfiner 1999, p.73). It follows from Eq. (8) that:

$$\frac{S_Z(v)}{S_{\mathbf{U}}(v)} \simeq \frac{|v|_{\mathrm{c}}}{|v|} \quad \text{with} \quad |v|_{\mathrm{c}} = A \cdot \frac{S_Z}{S_{\mathbf{U}}} \tag{10}$$

Equation (10) shows that the ratio $|v|_{\mathrm{c}}/|v|$ determines the relative contribution of the model inputs $Z(\mathbf{x})$ and $\mathbf{U}$ to the output variance $\mathrm{Var}(Y_v)$. The larger that this ratio is, the larger the part of the output variance $\mathrm{Var}(Y_v)$ is that is explained by the input random field $Z(\mathbf{x})$. For a small ratio (i.e., when the area of the support $v$ is large compared with the critical size $|v|_{\mathrm{c}}$), the variability of $Z(\mathbf{x})$ is mainly *local*, and the spatial correlation of $Z(\mathbf{x})$ over $v$ is weak. This local variability averages over the support $v$ when the aggregated model output $Y_v$ is computed; hence, input 2D random field $Z(\mathbf{x})$ explains a small fraction of the output variance $\mathrm{Var}(Y_v)$. However, for a greater ratio (i.e., when the area of the support $v$ is small compared with the critical size $|v|_{\mathrm{c}}$), the spatial correlation of $Z(\mathbf{x})$ over $v$ is strong. The averaging-out effect is weaker; hence, model input $Z(\mathbf{x})$ explains a larger fraction of the output variance $\mathrm{Var}(Y_v)$.

### 3.3 Link between covariance function and block sensitivity indices

Critical size $|v|_{\mathrm{c}} = A \cdot \frac{S_Z}{S_{\mathbf{U}}}$ depends on the covariance function $C^*(\cdot)$ of the random field $\mathbb{E}_Z Y(\mathbf{x})$, which is itself driven by the covariance function $C(\cdot)$

of the input SRF $Z(\mathbf{x})$. Let us now assume that $Z(\mathbf{x})$ is a Gaussian random field (GRF). $\mathbb{E}_Z Y(\mathbf{x})$ is then square-integrable with respect to the standard normal density. It can be decomposed into an Hermitian expansion and its covariance function $C^*(\cdot)$ can be written as (Chilès and Delfiner 1999, p.396-399; see Appendix B for a proof):

$$C^*(\mathbf{h}) = \sum_{k=0}^{\infty} \lambda_k^2 \cdot [C(\mathbf{h})]^k \tag{11}$$

For most of the usual transition covariance functions (e.g., spherical, exponential and Gaussian models), the covariance $C(\mathbf{h})$ is a monotically increasing function of correlation length $a$. In this case, it follows from Eq. (11) that the integral range $A = 1/\sigma^2 \int C^*(\mathbf{h}) d\mathbf{h}$ also increases with correlation length $a$. An increase in correlation length $a$ thus leads to an increase in the critical size $|v|_\mathrm{c}$, and the ratio of block sensitivity indices $S_Z(v)$ and $S_\mathbf{U}(v)$ satisfies (Eq. (10)):

$$\frac{\partial}{\partial a} \left[ \frac{S_Z(v)}{S_\mathbf{U}(v)} \right] \geq 0 \tag{12}$$

The relative contribution of the uncertain model input $Z(\mathbf{x})$ to the variance of the output of interest $Y_v$ increases when the correlation length of $Z(\mathbf{x})$ increases. Indeed, when correlation length $a$ increases, the averaging-out effect that occurs when the model output is aggregated over spatial support $v$ weakens; thus, the fraction of the output variance $\mathrm{Var}(Y_v)$ which is explained by the input random field $Z(\mathbf{x})$ increases.

Nugget parameter's impact on the block sensitivity indices can be interpreted in the same manner. The nugget parameter $\eta$ controls the relative part of *pure noise* in the input random field $Z(\mathbf{x})$ (Eq. (1)). The smaller $\eta$ is, the weaker the averaging-out effect will be when the block average $Y_v$ is computed over the support $v$, and the larger the part of output variance $\mathrm{Var}(Y_v)$ will be that is explained by $Z(\mathbf{x})$. The critical size $|v|_{\mathrm{c}}$ is thus a decreasing function of nugget parameter $\eta$, and the ratio of block sensitivity indices $S_Z(v)$ and $S_{\mathbf{U}}(v)$ satisfies (Eq. (1), (8), (11)):

$$\frac{\partial}{\partial \eta}\left[\frac{S_Z(v)}{S_{\mathbf{U}}(v)}\right] \leq 0 \tag{13}$$

### 3.4 Illustrative example

To illustrate the change of support effects on sensitivity analysis results, we performed spatial GSA on our numerical example in the following settings: varying disc-shaped support $v$ of increasing size (Fig. 2); varying correlation length from $a = 1$ to $a = 10$ (Fig. 3); varying nugget parameter from $\eta = 0$ to $\eta = 0.9$ (Fig. 4). For each setting, we computed estimates of the output variance $\mathrm{Var}(Y_v)$, the block sensitivity indices $S_{\mathbf{U}}(v)$, $S_Z(v)$ and the ratio $S_Z(v)/S_{\mathbf{U}}(v)$ over $N = 4096$ model runs. Mean values with a 95% confidence interval were then computed for each estimate using bootstrapping (100 replicas). In accordance with Eq. (9), (12) and (13), it appears that the block sensitivity index $S_Z(v)$ (i) decreases when the support $v$ increases (Fig. 2(b)), (ii) increases with the correlation length $a$ (Fig. 3(b)), and (iii) decreases with the nugget parameter $\eta$ (Fig. 4(b)). The opposite trends are observed for sensitivity index $S_{\mathbf{U}}(v)$. The change of support effect is clearly highlighted in Fig. 2(b): the contribution of the economic parameters $U_1$, $U_2$ and $U_3$ to the variance of total flood damage $|v| \cdot Y_v$ exceeds the contribution of the water level map $Z(\mathbf{x})$ when the radius $r$ of $v$ is greater than $r_c \simeq 18$; for radius $r < r_c$, the variance of total flood damage over the support $v$ is mainly explained by the variability of the water levels $Z(\mathbf{x})$. Finally, Figure 2(c) shows that the ratio $S_Z(v)/S_{\mathbf{U}}(v)$ is proportional to $1/|v|$ when the support $v$ is large enough. The theoretical curve $S_Z(v)/S_{\mathbf{U}}(v) = |v|_{\mathrm{c}}/|v|$ (Eq. (10)) was fitted (least squares - $R^2 = 0.99$) on data points (for $r \geq 20$ only), yielding an estimate of the critical size $|v|_{\mathrm{c}} \simeq 1,068$. All calculations and figures were realized in R (R Development Core Team 2009): random realizations of $Z(\mathbf{x})$ were generated with the *GaussRF()* function from the *RandomFields* package (Schlather 2001), while computation of sensitivity indices was based on a modified version of the *sobol()* function from the *sensitivity* package.

## 4 Discussion

Our first goal was to provide a formalism that extends the variance-based GSA approach to spatial models when the modeler is mainly interested in the linear average or the sum of a point-based model output $Y(\mathbf{x})$ over some spatial unit $v$. Our approach is strongly motivated by various prior publications. Other authors had already computed site sensitivity indices (Marrel et al. 2011; Pettit and Wilson 2010) and block sensitivity indices (Lilburne and Tarantola 2009), but did so without naming them or exploring their analytical properties or their relationship. Our work is an attempt to do so. Equation (8) provides an exact relation between the site and block sensitivity indices, it may prove useful in the case of a model with a simple enough analytical expression.

Our research also sought to account for the change of support effects in the propagation of uncertainty through spatial models, within a variance-based GSA framework. We proved that the fraction of the variance of the model output that is explained by a spatially distributed model input $Z(\mathbf{x})$ decreases under model upscaling; when the support $v$ is large enough, the ratio of the block sensitivity index of spatially distributed input to the block sensitivity index of non-spatial inputs is proportional to $|v|_c/|v|$. The critical size

$|v|_{\mathrm{c}}$ depends on the covariance function of the input SRF $Z(\mathbf{x})$; it usually increases with an increase of the correlation length $a$ or a decrease of the nugget parameter $\eta$. These findings are a translation into GSA formalism of the averaging-out effect clearly exhibited by Journel and Huijbregts (1978) in the regularization theory. Our contribution is to discuss this issue from the point of view of GSA practitioners. Formalizing the effect of a change of support on sensitivity analysis results may help modelers when they consider model upscaling; it will orientate future data gathering by identifying model inputs that will explain the largest fraction of the variance of the model output over a new spatial support. Our contribution also promotes an increased awareness of the issue of sharing out efficiently, among the various inputs used by a complex computer code, the cost of collecting field data. At some point of the model building process, the modeller will usually aim at reducing the variance of the output below a given threshold, that will depend on the model use. To do so, the modeller may have to improve his knowledge on the *real* value of some of the model inputs, usually by collecting extra data. In this case, gathering extra field data on inputs maps that have small sensitivity indices ($S_Z(v) < 0.1$) would be unefficient, as it would be costly but could not reduce the variance of the model output by a large fraction. Saint-Geours et al. (2011) discuss this issue on a flood risk assessment case study.

It should be noted that our approach is based on conditions that may not be met in some practical cases. First, we considered a model $\mathcal{M}$ with a single spatially distributed input $Z(\mathbf{x})$. In real applications, modelers may have to deal with several spatial inputs $Z_1(\mathbf{x}), \ldots, Z_m(\mathbf{x})$, with different covariance functions $C_i(\cdot)$, correlation lengths $a_i$ and nugget parameters $\eta_i$. In this case, it can be shown that Eq. (8) still holds separately for each spatial input $Z_i(\mathbf{x})$. However, no conclusion can be drawn *a priori* regarding how a change of support affects the relative ranking of two spatial inputs $Z_i(\mathbf{x})$ and $Z_j(\mathbf{x})$; the ratio of their block sensitivity indices $S_{Z_i}(v)/S_{Z_j}(v)$ will depend on the ratio of block variances $\sigma^2_{v,i}/\sigma^2_{v,j}$. Second, some environmental models are not point-based and involve spatial interactions (e.g., erosion and groundwater flow models). In this case, it still may be possible to build a point-based surrogate model as a coarse approximation of the original model; if not, then the change of support properties discussed in Sect. 3 may not hold. Third, we assumed the input random field $Z(\mathbf{x})$ to be stationary; if it is not, site sensitivity indices depend on site $\mathbf{x}^*$ (Eq. (4)). It is then possible to compute maps of these indices (Marrel et al. 2011; Pettit and Wilson 2010) to discuss the spatial variability of model inputs sensitivities.

Finally, we focused on the case in which the modeler's interest is in the spatial linear average or the sum of model output $Y(\mathbf{x})$ over the support

$v$. As discussed by Lilburne and Tarantola (2009), other outputs of interest may be considered, such as the maximum value of $Y(\mathbf{x})$ over $v$ (e.g., maximal pollutant concentration over a zone), some quantile of $Y(\mathbf{x})$ over $v$ (Heuvelink et al. 2010), or the percentage of $v$ for which $Y(\mathbf{x})$ exceeds a certain threshold. To our knowledge, no study has investigated the properties of sensitivity indices computed with respect to such outputs of interest.

## 5 Conclusion

This paper provides a formalism to apply variance-based global sensitivity analysis to spatial models when the modeler's interest is in the average or the sum of the model output $Y(\mathbf{x})$ over a given spatial unit $v$. Site sensitivity indices and block sensitivity indices allow us to discuss how a change of support modifies the relative contribution of uncertain model inputs to the variance of the output of interest. We demonstrate an analytical relationship between these two types of sensitivity indices. Our results show that the block sensitivity index of an input random field $Z(\mathbf{x})$ increases with the ratio $|v|_{\mathrm{c}}/|v|$, where $|v|$ is the area of the spatial support $v$ and the critical size $|v|_{\mathrm{c}}$ depends on the covariance function of $Z(\mathbf{x})$. Our formalization is made with a view toward promoting the use of sensitivity analysis in model-based spatial decision support systems. Nevertheless, further research is needed to explore the case of non-point-based models and extend our study to outputs of interest other than the average value of model output over support $v$.

## References


Brown JD, Heuvelink GBM (2007) The Data Uncertainty Engine (DUE): a software tool for assessing and simulating uncertain environmental variables. Comput Geosci 33(2):172–190

Cariboni J, Gatelli D, Liska R, Saltelli A (2007) The role of sensitivity analysis in ecological modelling. Ecol Model 203(1-2):167–182

Chilès J-P, Delfiner P (1999) Geostatistics, modeling spatial uncertainty. Wiley, New York

Cressie N (1993) Statistics for spatial data, revised edn. Wiley, New York

European Commission (2009) Impact assessment guidelines. Guideline #SEC(2009) 92

Heuvelink GBM (1998) Uncertainty analysis in environmental modelling under a change of spatial scale. Nutr Cycl Agroecosystems 50:255–264

Heuvelink GBM, Brus DJ, Reinds G (2010) Accounting for spatial sampling effects in regional uncertainty propagation analysis. In: Tate NJ, Fisher PF (eds) Proceedings of Accuracy2010 - The ninth international symposium on spatial accuracy assessment in natural resources and environmental sciences, pp 85–88

Heuvelink GBM, Burgers SLGE, Tiktak A, Van Den Berg F (2010) Uncertainty and stochastic sensitivity analysis of the GeoPearl pesticide leaching model. Geoderma 155(3-4):186–192

Journel AG, Huijbregts CJ (1978) Mining geostatistics. The Blackburn Press, Caldwell

Lilburne L, Tarantola S (2009) Sensitivity analysis of spatial models. Int J Geogr Inf Sci 23(2):151–168

Marrel A, Iooss B, Jullien M, Laurent B, Volkova E (2011) Global sensitivity analysis for models with spatially dependent outputs. Environmetrics 22(3):383–397

Pettit CL, Wilson DK (2010) Full-field sensitivity analysis through dimension reduction and probabilistic surrogate models. Probab Eng Mech 25(4):380–392

R Development Core Team (2009) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienne. ISBN 3-900051-07-0, URL http://www.R-project.org

Refsgaard JC, Van Der Sluijs JP, Højberg AL, Vanrolleghem PA (2007) Uncertainty in the environmental modelling process: a framework and guidance. Environ Model Softw 22:1543–1556

Rohmer J, Foerster E (2011) Global sensitivity analysis of large-scale numerical landslide models based on Gaussian-Process meta-modeling. Comput Geosci, in press. Doi:10.1016/j.cageo.2011.02.020

Ruffo P, Bazzana L, Consonni A, Corradi A, Saltelli A, Tarantola S (2006) Hydrocarbon exploration risk evaluation through uncertainty and sensitivity analyses techniques. Reliab Eng Syst Saf 91(10-11):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? In:

Tate NJ, Fisher PF (eds) Proceedings of Accuracy2010 - The ninth international symposium on spatial accuracy assessment in natural resources and environmental sciences, pp 81–84

Saint-Geours N, Bailly J-S, Grelot F, Lavergne C (2011) Analyse de sensibilité de Sobol d'un modèle spatialisé pour l'évaluation économique du risque d'inondation. J Soc Fr Stat 152:24–46

Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S, eds. (2008) Global sensitivity analysis, the primer. Wiley, New York

Schlather M (2001) Simulation of stationary and isotropic random fields. R-News 1(2):18–20

Sobol' I (1993) Sensitivity analysis for non-linear mathematical model. Model Comput Exp 1:407–414

Tarantola S, Giglioli N, Jesinghaus J, Saltelli A (2002) Can global sensitivity analysis steer the implementation of models for environmental assessments and decision-making? Stoch Env Res Risk Assess 16:63–76

US Environmental Protection Agency (2009) Guidance on the development, evaluation, and application of environmental models. Council for Regulatory Environmental Modeling. http://www.epa.gov/crem/library/cred_guidance_0309.pdf. Accessed 17 Apr 2012

## Appendix A: Proof of the relation between site sensitivity indices and block sensitivity indices

As mentioned in Sect. 2, we assume that the first two moments of $Y(\mathbf{x})$ exist. The ratio of block sensitivity indices gives (Eq. (5)):

$$\frac{S_Z(v)}{S_{\mathbf{U}}(v)} = \frac{\mathrm{Var}\left(\mathbb{E}\left[Y_v \mid Z(\mathbf{x})\right]\right)}{\mathrm{Var}\left(\mathbb{E}\left[Y_v \mid \mathbf{U}\right]\right)} \tag{14}$$

The conditional expectation of block average $Y_v$ given $Z(\mathbf{x})$ gives:

$$\begin{aligned}
\mathbb{E}\left[Y_v \mid Z\right] &= \mathbb{E}\left[\left(1/|v| \int_v Y(\mathbf{x}) d\mathbf{x}\right) \mid Z(\mathbf{x})\right] \quad \text{(definition of } Y_v\text{)} \\
&= 1/|v| \int_v \mathbb{E}\left[Y(\mathbf{x}) \mid Z(\mathbf{x})\right] d\mathbf{x} \quad \text{(for a point-based model)} \\
&= 1/|v| \int_v \mathbb{E}_Z Y(\mathbf{x}) d\mathbf{x} \quad \text{(definition of } \mathbb{E}_Z Y(\mathbf{x})\text{)}
\end{aligned}$$

Thus we have $\mathrm{Var}\left(\mathbb{E}\left[Y_v \mid Z\right]\right) = \mathrm{Var}\left(1/|v| \int_v \mathbb{E}_Z Y(\mathbf{x}) d\mathbf{x}\right) = \sigma_v^2$ (definition of $\sigma_v^2$). Moreover, the conditional expectation of block average $Y_v$ given input $\mathbf{U}$ gives:

$$\begin{aligned}
\mathbb{E}\left[Y_v \mid \mathbf{U}\right] &= \mathbb{E}\left[\left(1/|v| \int_v Y(\mathbf{x})\, d\mathbf{x}\right) \mid \mathbf{U}\right] \qquad \text{(definition of } Y_v\text{)} \\
&= 1/|v| \int_v \mathbb{E}\left[Y(\mathbf{x}) \mid \mathbf{U}\right]\, d\mathbf{x} \qquad \text{(Fubini's theorem)}
\end{aligned}$$

$\mathbb{E}\left[Y(\mathbf{x}) \mid \mathbf{U}\right]$ does not depend on site $\mathbf{x}$ under the stationarity of SRF $Z(\mathbf{x})$; thus, we have in particular $\mathbb{E}\left[Y_v \mid \mathbf{U}\right] = \mathbb{E}\left[Y(\mathbf{x}^*) \mid \mathbf{U}\right]$, and $\mathrm{Var}\left(\mathbb{E}\left[Y_v \mid \mathbf{U}\right]\right) = \mathrm{Var}\left(\mathbb{E}\left[Y(\mathbf{x}^*) \mid \mathbf{U}\right]\right)$. Combining these expressions with Eq. (14) yields:

$$\frac{S_Z(v)}{S_{\mathbf{U}}(v)} = \frac{\sigma_v^2}{\mathrm{Var}(\mathbb{E}\left[Y(\mathbf{x}^*) \mid \mathbf{U}\right])} \tag{15}$$

The ratio of site sensitivity indices gives (Eq. (4)):

$$\cdot\frac{S_Z}{S_{\mathbf{U}}} = \frac{\mathrm{Var}(\mathbb{E}\left[Y(\mathbf{x}^*) \mid Z(\mathbf{x})\right])}{\mathrm{Var}(\mathbb{E}\left[Y(\mathbf{x}^*) \mid \mathbf{U}\right])} \tag{16}$$

We notice that for point-based models $\mathrm{Var}\left[\mathbb{E}(Y(\mathbf{x}^*) \mid Z(\mathbf{x})\right] = \mathrm{Var}\left[\mathbb{E}_Z Y(\mathbf{x}^*)\right] = \sigma^2$ (definition of $\mathbb{E}_Z Y(\mathbf{x})$ (Eq. (7))). Finally, it follows from Eq. (15) and (16) that:

$$\frac{S_Z(v)}{S_{\mathbf{U}}(v)} = \frac{S_Z}{S_{\mathbf{U}}} \cdot \frac{\sigma_v^2}{\sigma^2}$$

## Appendix B: Hermitian expansion of random field $\mathbb{E}_Z Y(\mathbf{x})$

The random field $\mathbb{E}_Z Y(\mathbf{x})$ can be written (Eq. (2), (7)) as a transformation of the Gaussian random field $Z(\mathbf{x})$ through the function $\bar{\psi} : z \mapsto \int_{\mathbb{R}^n} \psi(\mathbf{u}, z) \cdot f_U(\mathbf{u})\, d\mathbf{u}$:

$$\mathbb{E}_Z Y = \bar{\psi}\,(Z)$$

where $f_U(\cdot)$ is the multivariate pdf of random vector $\mathbf{U}$. Under the hypothesis that the first two moments of $Y(\mathbf{x})$ exist, random field $\mathbb{E}_Z Y(\mathbf{x})$ has finite expected value and finite variance. Thus, $\bar{\psi}$ belongs to the Hilbert space $L^2(\mathcal{G})$ of functions $\phi : \mathbb{R} \to \mathbb{R}$, which are square-integrable with respect to Gaussian density $g(.)$. Hence, $\bar{\psi}$ can be expanded on the sequence of Hermite polynomials $(\chi_k)_{k\in\mathbb{N}}$, which forms an orthonormal basis of $L^2(\mathcal{G})$ (Chilès and

Delfiner 1999, p.399):

$$\bar{\psi} = \sum_{k=0}^{\infty} \alpha_k \cdot \chi_k \quad \text{with} \quad \chi_k(z) = \frac{1}{\sqrt{k!}} \cdot \frac{1}{g(z)} \cdot \frac{\partial^k}{\partial z^k} g(z)$$

where coefficients $\alpha_k$ are given by: $\alpha_k = \int_{\mathbb{R}} \chi_k(z)\bar{\psi}(z)g(z)\,dz$. It follows that $\mathbb{E}_Z Y(\mathbf{x})$ can be written as an infinite expansion of polynomials of $Z(\mathbf{x})$:

$$\forall \mathbf{x} \in \mathcal{D}, \quad \mathbb{E}_Z Y(\mathbf{x}) = \sum_{k=0}^{\infty} \alpha_k \cdot \chi_k \left[Z(\mathbf{x})\right]$$

Its covariance then gives (Chilès and Delfiner 1999, p.396, Eq. (6.23) and p.399, Eq. (6.25)):

$$\mathrm{Cov}\left(\mathbb{E}_Z Y(\mathbf{x}), \mathbb{E}_Z Y(\mathbf{x}+\mathbf{h})\right) = \sum_{k=0}^{\infty} \alpha_k^2 \cdot \left[\frac{C(\mathbf{h})}{C(0)}\right]^k = \sum_{k=0}^{\infty} \lambda_k^2 \cdot \left[C(\mathbf{h})\right]^k$$

where $C(\mathbf{h})$ is the covariance function of GRF $Z(\mathbf{x})$ and $\lambda_k = \alpha_k \cdot C(0)^{-k/2}$.

## Captions of tables

**Tab. 1** Sensitivity analysis results over $N = 4096$ model runs with respect to the outputs of interest $Y(\mathbf{x}^*)$ and $|v| \cdot Y_v$. Mean values with $\pm$s.d. computed by bootstrapping (100 replicas).

TABLE 1

| **Support** | **Site $\mathbf{x}^*$** | **Block $v$** |
|---|---|---|
| Output of interest | $Y(\mathbf{x}^*)$ | $\lvert v \rvert \cdot Y_v$ |
| Mean of output | $66.5 \pm 4.2$ | $539 \cdot 10^3 \pm 3.6 \cdot 10^3$ |
| Variance of output | $1393 \pm 188$ | $9 \cdot 10^9 \pm 0.2 \cdot 10^9$ |
| Sensitivity indices | *Site indices:*<br>$S_{\mathbf{U}} = 0.09 \pm 0.03$<br>$S_Z = 0.89 \pm 0.02$ | *Block indices:*<br>$S_{\mathbf{U}}(v) = 0.86 \pm 0.02$<br>$S_Z(v) = 0.12 \pm 0.02$ |

**Captions of figures**

**Fig. 1** Spatial model with uncertain inputs $\mathbf{U}$ and $Z(\mathbf{x})$ and spatial output $Y(\mathbf{x})$. The modeler is interested in the block average of $Y(\mathbf{x})$ over some spatial unit $v$

**Fig. 2** GSA results depending on the size of disc-shaped support $\nu$ (with radius $r$ and area $|\nu| = \pi r^2$), for $a = 5$, $\eta = 0.1$: **(a)** total variance of $Y_v$, **(b)** block sensitivity indices $S_{\mathbf{U}}(v)$ (solid line) and $S_Z(v)$ (dashed line), **(c)** ratio $S_Z(v)/S_{\mathbf{U}}(v)$ with fitted curve $S_Z(v)/S_{\mathbf{U}}(v) = |v|_{\mathrm{c}}/|v|$ (dashed line). Error bars show 95 % confidence interval computed by bootstrapping (100 replicas)

**Fig. 3** GSA results depending on correlation length $a$, for $\eta = 0.1$ and a disc-shaped support $v$ of radius $r = 50$: **(a)** total variance of $Y_v$, **(b)** block sensitivity indices $S_{\mathbf{U}}(v)$ (solid line) and $S_Z(v)$ (dashed line). Error bars show 95 % confidence interval computed by bootstrapping (100 replicas)

**Fig. 4** GSA results depending on covariance nugget parameter $\eta$, for $a = 5$ and a disc-shaped support $v$ of radius $r = 50$: **(a)** total variance of $Y_v$, **(b)** block sensitivity indices $S_{\mathbf{U}}(v)$ (solid line) and $S_Z(v)$ (dashed line). Error bars show 95 % confidence interval computed by bootstrapping (100 replicas)

Figure 1
Click here to download high resolution image

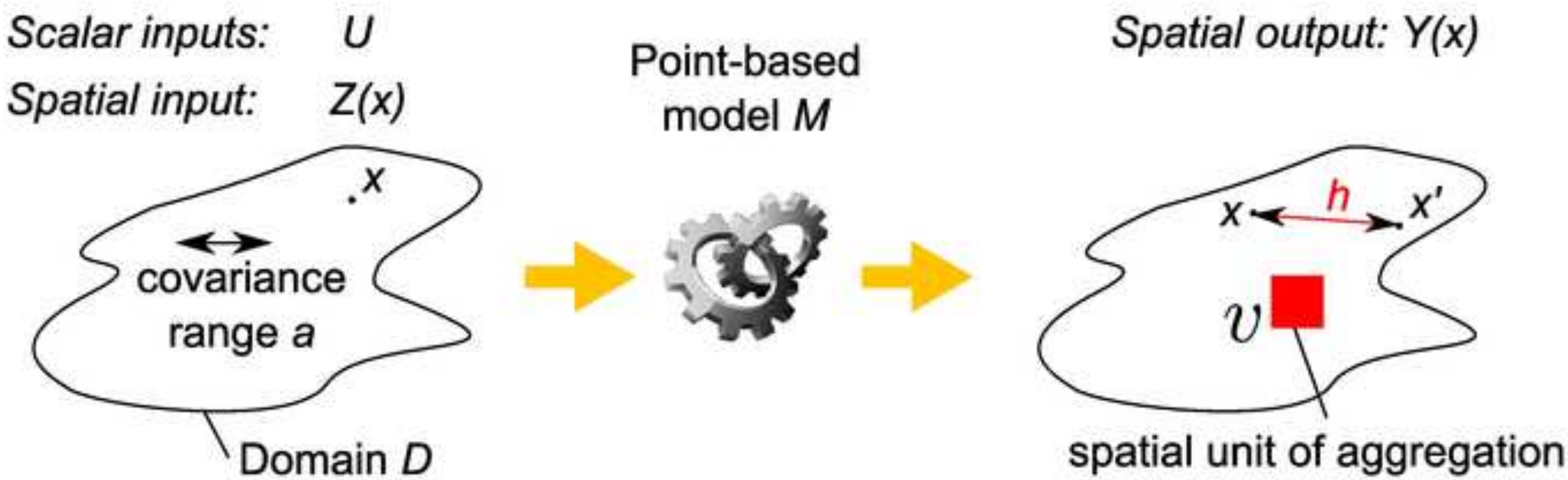


Figure 2
Click here to download high resolution image

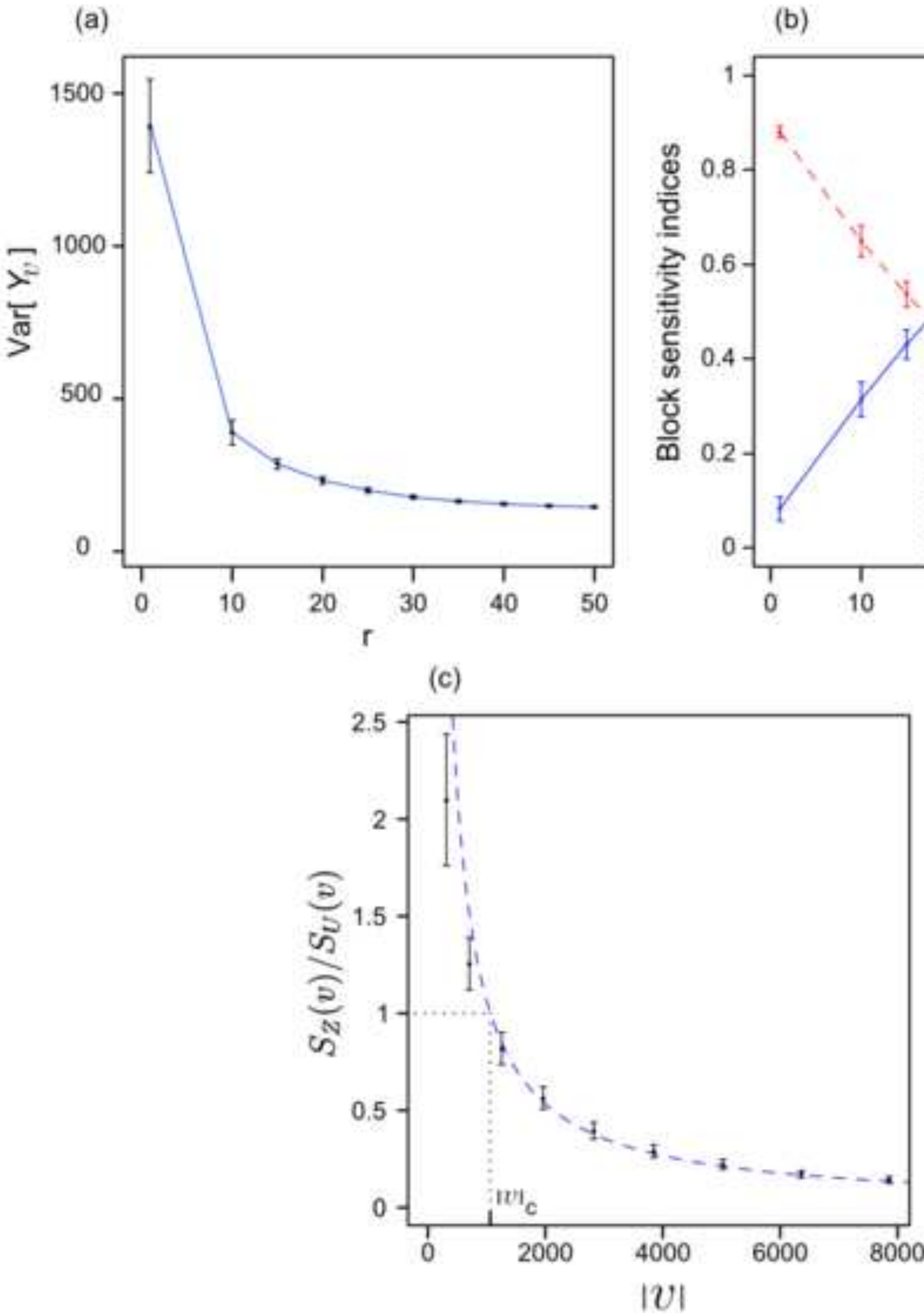


Figure 3
Click here to download high resolution image

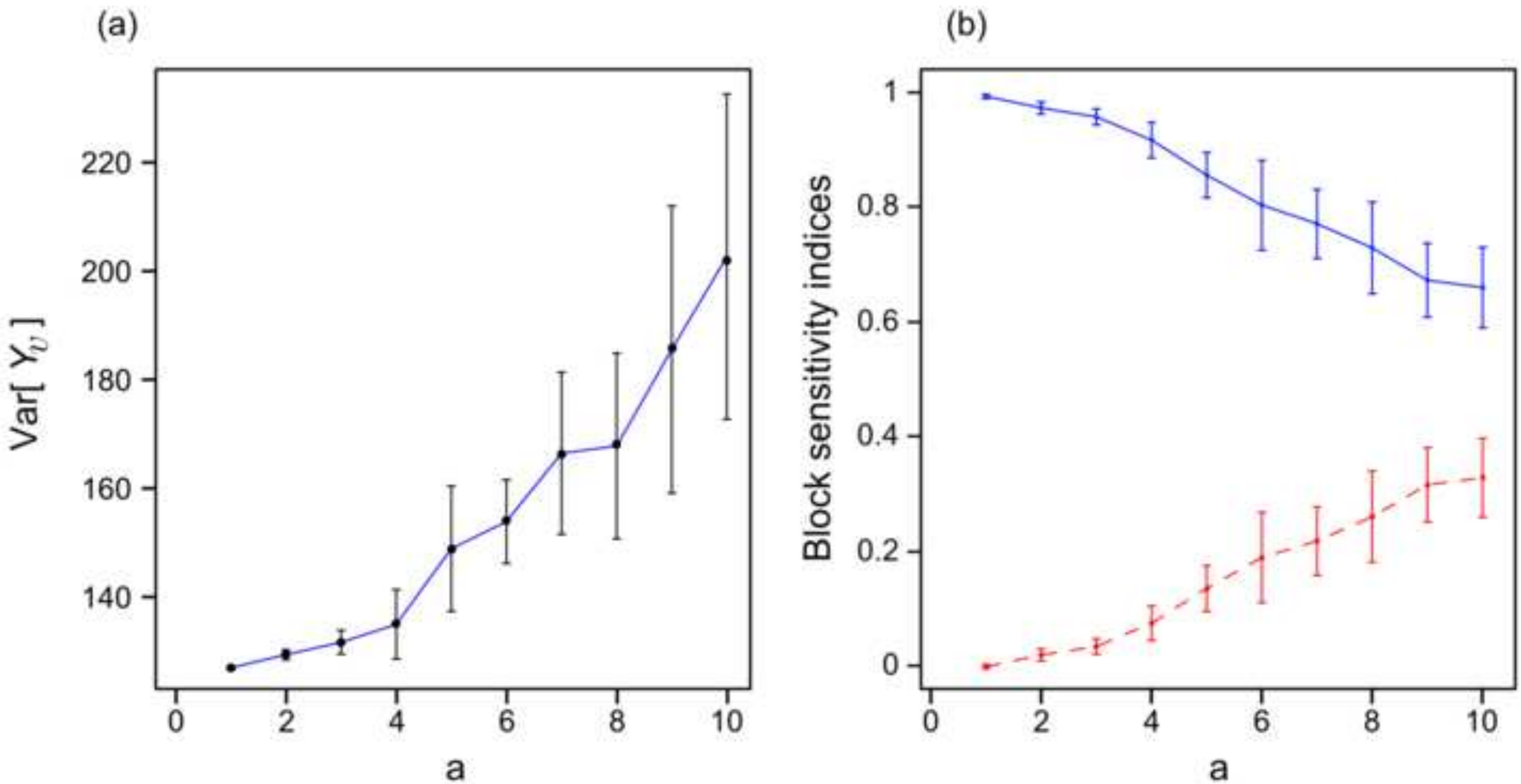
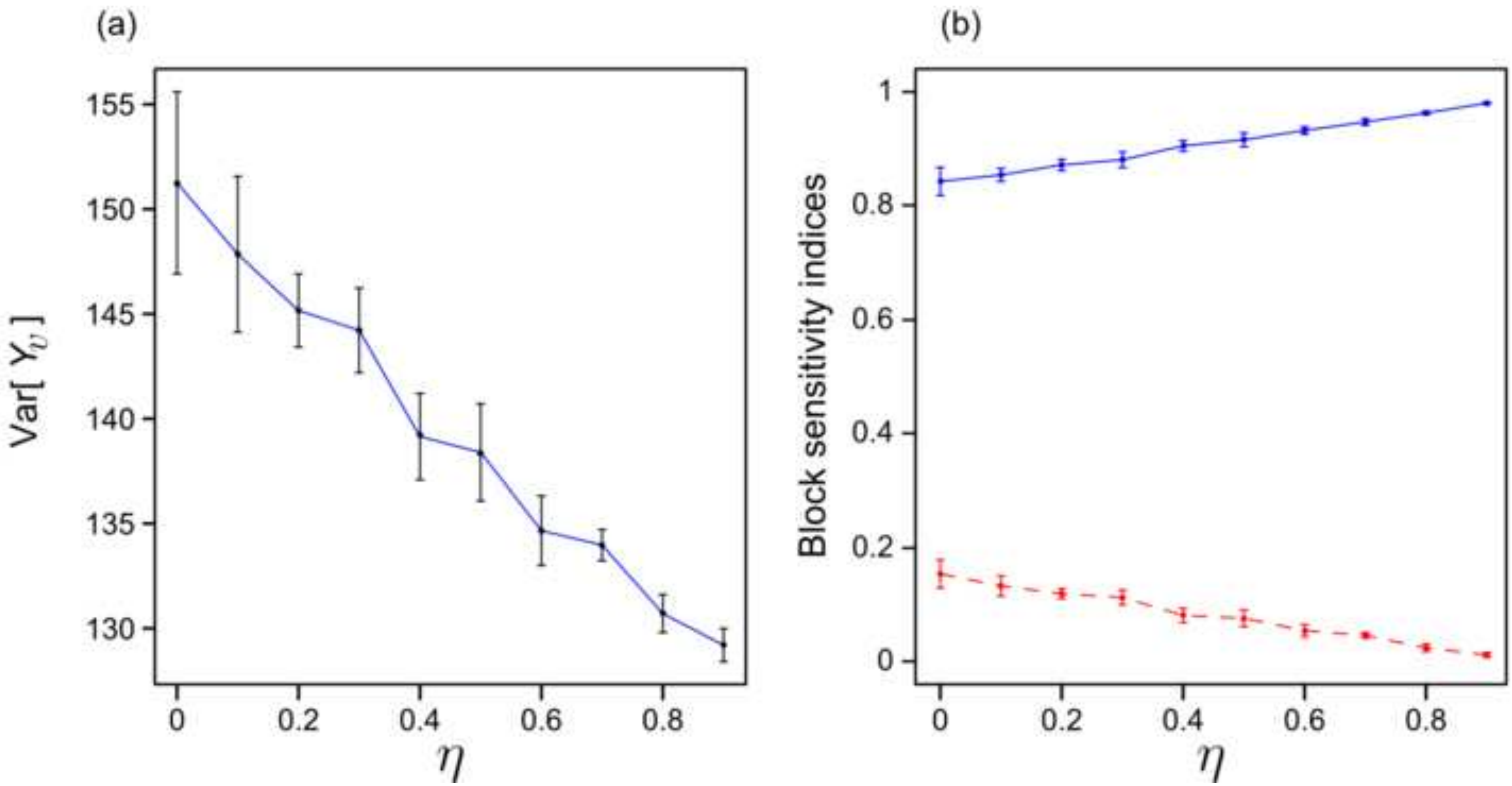


Figure 4
Click here to download high resolution image



April 2012, 16th

MATG731
"Change of Support in spatial variance-based sensitivity analysis"
N. Saint-Geours, C. Lavergne, J.-S. Bailly and F. Grelot

**Response to reviewer's comments**

We would like to thank the reviewer for his comments and suggestions. Our comments and answers are provided in the response below (in blue).

**-------------**

**Conditional Expectation**
Let $f$ X,Y(x,y) be the joint probability density function for random variables X, Y. To compute E[X |Y ] we must first consider the conditional probability density function for X (given Y = y). This
is
$f$ X | Y(x | y) = $f$ X,Y(x,y)/ $f$ Y(y) for $f$ Y(y) ≠ 0
= 0 for $f$ Y(y) = 0
Where $f$ Y(y) = ∫ $f$ X,Y(x,y)dx (the marginal probability density for Y)
Then E[X | Y = y] = ∫ x$f$ X | Y(x |y)dx provided ∫| x|$f$ X | Y(x |y)dx is finite.
While some properties of conditional expectation can be presented/discussed without explicitly mentioning the joint probability density it is always implicitly involved. In particular a proof of the "Total Law of Variance", i.e. the decomposition of the variance of X does not require a particular joint probability density function. But to compute Var{E[X |Y]} for a particular pair of random variables does require knowing the joint density function and both marginals.
Var{E[X | Y = y]} = EY{(E[X | Y])2 - (EY{E[X | Y]})2
The subscripting "EY" is a reminder that to compute this expectation the marginal density for Y must be used,
While (EY{E[X | Y]})2 = (E[X])2 , EY{(E[X | Y])2 ≠ E[X2] (unless X, Y are independent)

--> Thank you for this clarification. Some notations and explanations have been modified in the revised manuscript. *U* and *Z(x)* are supposed to be independent thoughout the manuscript (we forgot to make it clear in the initial submission). See the answers below for more details.

**VARIOUS CONCERNS**

- Page 2, lines 9-14 "Variance-based global sensitivity analysis" is an inanimate object or concept, hence it can't "aim" to do anything. The user of this methodology might have a particular goal or object in mind but the method itself does not. While variance is somewhat related to "uncertainty" it is not the same. "Uncertainty" would have to be quantified in terms of probabilities.

--> Thank you. We replaced "*aim*" by "*is used to*" in the revised manuscript. We also changed most occurrences of the term "*uncertainty*" -- that was indeed too vague -- by "*variance*" when it was appropriate.

- Page 3, line 29 The referenced document from EPA is no longer available at that URL. Elsewhere on the EPA site it indicates that this was a document put out to solicit public comment. It did not represent adopted policy or guidance.

--> Thank you. On April, 2012, 12th, we found the referenced document still available on the Web, at the URL which is given in the manuscript: http://www.epa.gov/crem/library/cred_guidance_0309.pdf

This document is a final version approved by EPA, after they collected public comments:
"*A draft version of this document was reviewed by an independent panel of experts established by EPA's Science Advisory Board and revised by CREM in response to the panel's comments. This final document is available in printed and electronic form."* (extract from the document preface)

The referenced document does not represent any legal requirement, yet it provides official guidance to develop and evaluate environmental models:
*"This document provides guidance to those who develop, evaluate, and apply environmental models. It does*

*not impose legally binding requirements; depending on the circumstances, it may not apply to a particular situation. The U.S. Environmental Protection Agency (EPA) retains the discretion to adopt, on a case-by-case basis, approaches that differ from this guidance."* (extract from the document disclaimer)

- Page 3, lines 45-47 Since the authors are using the terms “model”, “input”, “output” in only an abstract, vague way there is no assurance that collecting additional data will reduce the fraction of the variance of the output that is attributable to a particular input. It may also be a question of the nature of the”data”. Since the authors are specifically concerned with spatial models, the geographical pattern of the data locations may be very important and changing that pattern may result in a reduction in the variance of the output.

--> Thank you. Indeed, there is no assurance that collecting additional data will reduce the fraction of the variance of the output that is attributable to a particular uncertain input -- even if it is the case in many applications. The question of the spatial pattern of the data locations is also very important, and exploring how sensitivity indices vary under a change of these data locations would be interesting.

Yet our goal in this paragraph was just to give a general -- and probably too vague as you mentioned it -- overview of what sensitivity analysis is and what it is usually used for. In most cases, sensitivity analysis is used to identify the main processes involved in a complex computer code, and to select the inputs that contribute the most to the variance of the model outuput. Then the modeller will try to get a more accurate estimation of the 'real' value of these inputs. In most cases (yet not always) reducing the variance of these uncertain inputs will result in a reduction of the variance of the model output. We changed this sentence in the revised manuscript, to try and be more precise:

*"Based on these sensitivity indices, ranking the model inputs helps to identify inputs that should be better scrutinized first. Reducing the uncertainty on the inputs with the largest sensitivity indices (e.g, by collecting additional data or changing the geographical pattern of data locations) will often result in a reduction in the variance of the model output."*

- Page 3, line 50 “influence” is a vague term. Presumably the authors mean that the presence/absence of the inputs in question don’t cause the variance of the output to increase or decrease. They should be much more precise.

--> Thank you. Here we wanted to explain that if a model input has a very small sensitivity index compared to the other inputs, then trying to reduce the uncertainty on this input is of small interest : in any case it wouldn't cause the variance of the output to decrease much. As it was not clear enough and not really important, we removed this sentence in the revised manuscript.

- Page 3, line 52 “Model behavior” is a very vague term and can have a variety of meanings. “Understanding behavior” is at least partially a characteristic of the person(s) using the model.

--> Thank you. Our sentence was too vague. What we tried to say briefly is the following: global sensitivity analysis (GSA) is useful to explore the "response surface" of a complex computer code with many inputs, no simple analytical expression, that may represent various physical or anthropic processes. By performing GSA, the modeller will usually get a better idea of what are the main processes in his model, which processes strongly interact with others, which process may be simplified or even removed from the model, etc. We changed this sentence in the revised manuscript :

*"More generally, GSA helps to explore the response surface of a black box computer code and to prioritize the possibly numerous processes that are involved in it."*

- Page 3, line 56. Presumably in using the term “scalar inputs” the authors are contrasting those with “vector inputs”. In the example on page 7 the inputs are either random variables or a random field hence the term “scalar inputs” is not appropriate. However the inputs in this example are presumably “scalar valued”, i.e. each of the random variables is real valued and the random field is real valued. Also see line 9 on page 7

--> Thank you. What we wanted to oppose is: on one side, uncertain inputs that are modeled as real valued random variables; on the other side, uncertain inputs that are modeled as random functions or random fields (real valued or not). We changed the sentence in the revised manuscript:

*"Although GSA was initially designed for models where both inputs and output can be described as real valued random variables, some recent work has extended GSA to environmental models for which both the*

*inputs and output are spatially distributed over a two-dimensional domain and can be described as random fields."*

We also changed the description of inputs **U** and **Z**(x) on section 2.1.

- Page 4, line 14 Delete “sampled”. The field data is likely already a sample. Note that digital elevation models and land use maps might refer to physical presentations, e.g. printed on paper or they might refer to a visual representation in electronic form. In either case the maps are not the actual inputs. To actually determine/choose a random field model will likely require collecting data and then estimating parameters. Moreover there are different types of random field models and those different types might result in different variances for the output, this could result in a significant amount of “uncertainty” that is not quantified.

-->Thank you. We removed the word *"sampled"* in the revised manuscript, and rephrased the whole sentence:

*"In these works, the computer code under study uses maps derived from field data (e.g., digital elevation models and land use maps)."*

We couldn't make a clear distinction in the manuscript (to try keep things simple and short enough) between the *actual computer code inputs* (which may be numbers or maps), and *the description of the uncertainty on these inputs* (using random variables or random fields).
From a very practical point of view, maps (digital elevation model, landuse map) are stored as layers in a GIS software. These layers are the *actual inputs* that the computer code uses to calculate its output. Then, for real valued inputs, uncertainty is described by representing the input with a real valued random variable with a given pdf. For spatially distributed inputs (maps), the uncertainty can be described by representing the input with a given random field model. In this case, describing properly the random field model will indeed require collecting data and estimating parameters. The choice of the random field model will also result in additional "uncertainty" that is not quantified.

We rephrased section section 2.1 in the revised manuscript to make a clearer distinction between these two different notions.

- Page 4, line 35 Does “*efficient* estimation procedures” pertain to the meaning used in the statistical literature or does it have some other meaning? In the statistical literature one might refer to an “efficient estimator”, i.e., one which has a smaller variance than other estimators but that would not make sense with respect to “procedures”.

--> Thank you. We misused the word "*efficient*" in this sentence. We didn't intend to talk about the "*efficiency*" of estimators with the meaning used in the statistical litterature. Thus we just removed the word "*efficient*" from this sentence in the revised manuscript.

- Page 4, line 54 Note that the term “scale” has multiple (and sometimes contradictory) meanings. That is one of the reasons why the term “support” is used in the geostatistical literature as opposed to scale.

--> Thank you. We replaced most occurrences of the term "*scale*" by the term "*support*" in the revised manuscript. We kept the terms "*scale*" and "*model upscaling/dowscaling*" on p.4, line 54, because we are citing a paper (Heuvelink 1998, p.1) and we don't want to modify the expression used by the original author. We also kept the terms "*small-scale properties*" and "*large-scale properties*" on p.4, line 55, because we are citing a definition from Chilès & Delfiner (1999, section 8.1, p.593) that we don't want to modify either.

- Page 5, lines 43 & 53. What do “influential” and “influence” mean in this context? Also see line 22 on page 6 and many other places. “Influence” could mean deterministic, i.e. actual cause-effect or simply correlated or some other vague idea. The authors are using the wrong word

--> Thank you. Manuscript was modified according to this recommendation: we removed all occurrences of the terms "influential" and "influence" in the revised manuscript.

- Page 5, lines 48-50. Change”most to model output uncertainty” to “largest fraction of the variance of the output of the model”

--> Thank you, manuscript was modified according to this recommendation.

- Page 5, line 50 What is “spatial structure of uncertainty of a model input”? Moreover the authors never actually consider “uncertainty” they only consider the variance of the model output. All references to “uncertainty” should be omitted or appropriately changed.

--> Thank you. Our sentence was too vague. What we wanted to denote by the expression *"spatial structure of uncertainty of a model input"* is more precisely the shape and parameters of the covariance function *C(.)* of the random field model *Z(x)* that is used to describe an uncertain spatial input. We modified this expression in the revised manuscript to be more precise:

*"How does the contribution of a spatial input to the variance of the model output depend on the parameters of its covariance function?"*

Moreover, we replaced all occurences of the term "*uncertainty*" by "*variance*" when it was appropriate to do so (not in the very begining of the introduction which we want to be very general).

- Page 6, lines 27-33. The description of “point-based” models is incorrect.

--> Thank you. We took the term "*point-based model*" from the following paper (p.1):
Heuvelink GBM, Brus DJ, Reinds G (2010) *Accounting for spatial sampling effects in regional uncertainty propagation analysis.* In: Tate NJ, Fisher PF (eds) Proceedings of the 9th international symposium on spatial accuracy assessment in natural resources and environmental sciences (Accuracy2010), pp 85--88

where the authors say:
*"Many environmental models involve spatial interactions. Examples are erosion, groundwater flow and plant dispersal models. However, there are also many environmental models that are essentially point-based. For instance, models that predict crop growth, greenhouse gas emission, soil acidification or evapotranspiration at some location typically use soil, landuse, management and climate input data at that same location only."*

We want to limit our study to models that can be described by Eq.(2) in our manuscript:

$$\text{for all } x \text{ in } D,\ Y(x) = f\,[\,U, Z(x)\,]$$

We don't want to discuss the case where $Y(x^*)$ (model output at some location $x^*$) would be a function of $Z(x^*)$ (model input at that same location) and $Z(y)$ with $y \neq x^*$ (model input at some other location). Thus we want to limit our study to the kind of models where "the computation of the model output at some location $x^*$ uses the values of spatial inputs at that same location $x^*$ only".

We don't know if that family of models has a name in the geostatistical litterature. We would be happy to change the term "*point-based*" to a more appropriate term if it exists. As we don't know which other term to use so far, we didn't modify this term in the revised manuscript.

- Page 7, lines 14-17. Although there is some variation in the statistical literature of the definition of strict stationarity, it always includes the condition that the joint distribution of Z(x1), …, Z(xn), for any finite collection of points in D, is translation invariant. It may or may not include the requirement that Z(x) have finite expected value and finite variance. If it does then it is also second order stationary. In most applications the statistical characteristics of the random field will not be known, e.g. the multivariate probability distribution and its parameters. Since it is unlikely that one can collect data from multiple realizations of the random field, one can’t use the data to determine characteristics of the multivariate probability distribution.

--> Thank you for this clarification. We want to use the notion of "*second-order stationary random field*" as it is defined in Chilès & Delfiner (1999, p.17 section 1.1.4.). If we understood properly:
– "strict stationarity" with finite first-order and second-order moments implies "second-order stationarity"
– "second-order stationarity" does not necessarily implies "strict stationarity"
– under the hypothesis of second-order stationarity, it is possible to use the data of a single realization of the random field to determine its mean and the characteristics of its covariance function. Without this hypothesis, one can’t use the data to determine characteristics of the multivariate probability distribution.

We rephrased the whole section 2.1 in the revised manuscript.

- Page 7, lines 17-49. Since Y(x) is purportedly a function of **U** and Z(x), in order to consider statistical properties of Y(x) one must assume that **U** and Z(x) are defined on the same probability space. Are the authors assuming that U and Z(x) are independent or is there some unstated joint probability distribution? In any case, if Y(x) has finite expected value then both of the conditional expectations must exist. To actually compute any of these expectations one needs to know the probability densities and the conditional densities. All of this material needs to be re-written in a more careful way (and statistically correct).

--> Thank you for this question. Random vector **U** and random field **Z**(x) are supposed to be independent throughout the paper, but we unfortunately forgot to make it appear in the initial manuscript. We added the following sentence in the revised manuscript (section 2.1):

*"U and Z(x) are supposed to be independent."*

Under this hypothesis, the conditional density of U given Z(x) is the marginal density for U and the conditional density of Z(x) given U is the marginal density for Z(x).

Also, when we wrote the initial manuscript, we were not sure that when Y(x) has finite expected value and variance, then E[ Y(x) | Z(x)] and E[Y(x) | U] (which are both random variables) necessarily both have finite expected value and finite variance. After some more careful examination, it appears that this property simply derives from conditional Cauchy-Schwarz inequality. Thus we removed the additionnal conditions on E[ Y(x) | Z(x)] and E[Y(x) | U] in the revised manuscript.

Finally, we are not really interested in discussing the expression of the conditional densities of Y(x) knowing U or Z(x), so we don't think there is a need to explicitly mention these conditional densities in the manuscript.

- Page 7, line 17. Change "covariogram" to "covariance function", see page 70 of Chiles and Delfiner for the definition of "covariogram" which is not the same as "covariance function". Also see the same problem on page 12, line 28 and page 13, line 37, page 14, line 54.

--> Thank you. Manuscript was modified according to this recommendation: all occurrences of the term "*covariogram*" were replaced by "*covariance function*".

For information, we took the definition of "*covariogram*" from Cressie (1993, p.53, eqn. 2.3.3) :
*" The function C(.) is called a covariogram or a stationary covariance function:*
*cov(Z(s1),Z(s2)) = C(s1 – s2) for all s1, s2 in D"*

and also p.67:
*" Call the function C(.) given by (2.3.3) (provided it is well defined), a covariogram. (Notice that it has also been called an autocovariance function by time-series analysts.)"*

- Page 8, line 31. It is essential to restrict the possible choices for the function *f,* for example if *f(*X1, X2) = X1/X2 where X1, X2 are independent standard normal random variables then *f*(X1, X2) does NOT have either an expected value nor a variance. Thus all of the discussion that follows is erroneous.

--> Thank you. The discussion holds if *Y* has finite expected value and finite variance. The manuscript was modified according to this recommendation:

*"Let us consider a model $Y = f(X_1, ..., X_n)$ where $X_i$ are independent random variables and where the first two moments of Y exist."*

- Page 8, lines 43-45. The statement "It is the expected part of the output variance Var(Y) that could be reduced" is incorrect The sensitivity index is really based on the following Var(Y) = E(Var (Y | X)) + Var(E(Y |X)) The two terms on the right hand side are the "fraction of variance unexplained" and "fraction of variance explained". This relationship is sometimes called the "Total law of variance". See most text books on probability theory for more details.

--> Thank you. We took most of the material of this paragraph from the following book, which is very often cited if the field of sensitivity analysis:

Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S, eds (2008) *Global sensitivity analysis, the primer.* Wiley, New York

In the literature of sensitivity analysis, the definition of sensivity indices is usually built from the decomposition of the function *f(.)* into what is often called "*high dimensionnal model representation*" (HDMR) (that was originally suggested by Hoeffding (1948), see also Sobol (2001, 2003), references below):

$$f(x_1,...x_n) = f_0 + f_1(x_1) + ... + f_n(x_n) + f_{1,2}(x_1,x_2) + ... + f_{1,...n}(x_1,....,x_n)$$

where the integral of each function $f_{i1,...,ik}()$ is equal to 0. The decomposition of the variance of *f(X)* that derives from this functional decomposition is very close to the ANOVA framework:

$$Var[\ f(X)\ ]= Var[\ f_1(X_1)\ ] + ... + Var[\ f_n(X_n)\ ] + Var[\ f_{1,2}(X_1,X_2)\ ] + ... + Var[\ f_{1,...n}(X_1,....,X_n)\ ].$$

First-order sensitivity index of input $X_i$ is then defined as the ratio $Var[\ f_i(X_i)\ ] / Var\ [\ f(X)\ ]$, which is the same as the Equation (3) in our manuscript, written in terms of conditional expectations.

Higher-order sensitivity indices can also be defined from this decomposition, to characterize the contribution of the interactions between several uncertain inputs $X_{i1},...,X_{ik}$ to the variance of the output *f(X)*.

The statement *"It is the expected part of the output variance Var(Y) that could be reduced"* may be incorrect. In the revised manuscript, we just removed it.

References :

W. Hoeffding. *A class of statistics with asymptotically normal distribution*. Annals of Mathematical Statistics, 19 :293–325, 1948.

Homma, T. & Saltelli, A. *Importance measures in global sensitivity analysis of nonlinear models* Reliability Engineering & System Safety, 1996, 52, 1 - 17

Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S, eds (2008) *Global sensitivity analysis, the primer.* Wiley, New York

Sobol, I. *Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates* Mathematics and Computers in Simulation, 2001, 55, 271 - 280

Sobol', I. M. *Theorems and examples on high dimensional model representation* Reliability Engineering & System Safety, 2003, 79, 187 - 193

- Page 9, lines 47-50. See the comment just above.

--> OK.

- Page 9, eqs 4 & 5. Delete the ":x' ε D", that is logically incorrect

--> Thank you, the manuscript was modified according to this recommendation.

- Page 10, lines 24-26. The exponential covariance function does not have a true range in the sense that there is no spatial dependence beyond that distance. However in the geostatistical literature it is common to refer to the "effective range", i.e. the distance at which the value of covariance function has decreased to .05C(0). If *a* is the parameter in the functional form of the covariance function then it is not the effective range, however it is what is called the *correlation length*. The authors need to clarify which it is, "effective range" or "correlation length/integral range". In any case it appears that the authors have chosen a range that is quite small in comparison to the scale of the "block" used later.

--> Thank you. We misused the term "range", denoted by '*a*', to refer to the parameter in the functional form of the covariance function *C(.)*. In our illustrative example, the "*effective range*" is thus close to *3*a* (as we use an exponential covariance function:

$$C(h) = C(0)\ .\ [\ \eta\ \delta(|\mathbf{h}|) + (1-\eta)exp(-|\mathbf{h}|/a)\ )]$$

We changed the manuscript according to the recommendation, replacing all occurrences of the term "*range*" by "*correlation length*" or "*effective range*" depending on the intended meaning.

The term "*integral range*", denoted by *A* in our manuscript, was taken from Chilès and Delfiner (1999, p.73):

$$A = 1/\sigma^2 \int C(h)\, dh$$

where *C(.)* is a covariance function. If we understood properly this definition, for a given covariance function *C(.)*, the integral range *A* is not necessarily equal to the parameter *a* in the functional form of *C(.)* (that we now call *"correlation length"*). If *C(.)* is an exponential covariance function with correlation length *a*, then the integral range is equal to $A = \pi a^2$. We didn't change the occurrences of the term "*integral range*" in the revised manuscript.

In the illustrative example, we chose a correlation length varying from *a=1* to *a=10,* corresponding to an effective range varying from *3* to *30*. The radius of the disc-shaped "block" was varying from *r = 1* to *r = 50*. The ratio between the range and the size of the block is of the same order of magnitude than in the real case-study that we work on (not described in the manuscript), where *Z(x)* is a map of water levels and *M* is a model that computes total flood damage over a large study area.

- Page 10, section 2.3. Are the random variables and the random field dependent or independent? They have not sufficiently characterized the random field Y(x)

--> Thank you for this question. Random vector **U** and random field **Z**(x) are supposed to be independent throughout the paper, but we unfortunately forgot to make it appear in the manuscript. We added the following sentence in the revised manuscript (section 2.1):

*"U and Z(x) are supposed to be independent."*

- Page 10, line 29 Change “describe” to “determine”, very different meanings

--> Thank you. Manuscript was modified according to this recommendation.

- Page 10, line 31 Change “expected monetized costs” to “monetary costs”. The “expected” term is clearly incorrect here.

--> Thank you. Manuscript was modified according to this recommendation.

- Page 10, lines 53-55. What does “analytical expression” mean here? Does it only mean the formula given in eq(6) or does it mean a complete statistical characterization? In any case the authors have left out the key question of whether the two inputs are dependent or independent? If they are dependent then it is essential to give the actual joint probability distribution. This would be necessary even if Monte Carlo simulation is used.

--> Thank you for this question. By *"analytical expression"* we mean eq(6) + the knowledge of the pdf of random variables $U_1$, $U_2$, and $U_3$ + the knowledge of random field model of Z(x). Random vector **U** and random field **Z**(x) are supposed to be independent throughout the paper. In the revised manuscript we changed this sentence to:

*"Here, the expression of mapping ψ and the statistical characterization of model inputs may be simple enough that exact values of sensitivity indices could be derived"*

- Page 10, line 60. Change “model complexity is high” to “model is very complex”

--> Thank you. Manuscript was modified according to this recommendation.

- Page 11. line 9. The authors should provide more information about how they used a “Monte Carlo approach” and also the “sampling-based method”. Lilburne and Tarantola (2009) do not use the terminology “sampling-based method”, they do consider various “sampling designs”.

--> Thank you. We estimated sensitivity indices using the computational procedure described in section 3.2 of Lilburne and Tarantola (2009), with a quasi-random sampling design, and with the estimators given in formulas (8), (9) (11)-(16). We changed the sentence in the revised manuscript:

*"A usual alternative is to consider model M as a black box and estimate sensivity indices with Monte-Carlo simulation. We chose to use the estimators and the computational procedure described by Lilburne and Tarantola (2009, Sect. 3.2), based on a quasi-random sampling design, using N = 4096 model runs."*

- Page 12, line 26 The expression apparently refers back to page 7 but "p(**u**)" is not identified either on page 7 or on this page (nor on page 25). The integral should be ∫y *f* Y | Z (y |z) dy where *f* Y | Z (y |z) = *f* Y , Z (y , z)/ *f* Z (z), i.e. the conditional density of Y given Z is the quotient of the joint density of Y, Z and the marginal density of Z. To determine the joint density of Z, Y one must know the joint density of **U** and Z. The authors have not given that density nor said anything about the joint distribution of U and Z. It is also possible to use a change of variables in the integral to write the above integral in terms of the conditional density of U given Z. Perhaps the authors are assuming (but without ever saying so) that U and Z are independent. In that case the joint density would be the product and then the conditional density of U given Z would be the marginal density for U.

--> Thank you for this clarification. *p(u)* is supposed to denote the multivariate pdf of the random vector **U**, as it was (not very clearly) said on page 7, line 12 in the initial manuscript.
Random vector **U** and random field **Z**(x) are supposed to be independent throughout the paper, but we unfortunately forgot to make it appear in the initial manuscript. We added the following sentence in the revised manuscript (section 2.1):

*"U and Z(x) are supposed to be independent."*

Under this hypothesis, the conditional density of **U** given **Z** is the marginal density for **U**, i.e. the multivariate pdf *p(u)*. To make notations more explicit, we replaced the notation $p(\mathbf{u})$ by $f_U(\mathbf{u})$ in the revised manuscript. We also put the definition of $f_U(\mathbf{u})$ on p.12 in the revised manuscript:

*"$E_Z Y(x)$ is the transform of the input SRF Z(x) via the function $\underline{\Psi}(z) = \int \Psi(\mathbf{u}, z)\, f_U(\mathbf{u})\, d\mathbf{u}$ [Eq. (2)] --- where $f_U(.)$ is the multivariate pdf of random vector **U**."*

- Page 13, line 32 Again the term "influence", also the comparative word should be "lesser", Also see page 14, line 22., page 14, line 30, page 14, line 54. Also see the title of section 3.3, page 15 line 39

--> Thank you. Manuscript was modified according to this recommendation: we replaced the term "*influence*" by more appropriate terms throughout the manuscript. We also replaced "lower" by "lesser".

- Page 13, line 35 Again "uncertainty" when the authors really mean variance of the output, also see line 9 page 14., line 19 page 14

--> Thank you. Manuscript was modified according to this recommendation: we replaced the term "*uncertainty*" by "*variance*" when it was more appropriate.

- Page 13, line 55 Change "drives" to "determines"

--> Thank you. Manuscript was modified according to this recommendation.

- Page 14, change "low" to "small"

--> Thank you. Manuscript was modified according to this recommendation.

- Page 14, lines 49-54. The real reason that EZ[Y(x] can be represented as an infinite series of Hermite polynomials is that EZ[Y(x] is square integrable with respect to the standard normal density function. Note this is "mean-square convergence, not "point-wise"

--> Thank you for this clarification. This is the proof that we try to develop in more details in Appendix B. The sentence page 14, lines 49-54 has been modified to make this reason clearer:

*"EZ[Y(x)] is then square-integrable with respect to the standard normal density function. It can be decomposed into an Hermitian decomposition and its covariogram can be written as: (...)"*

- Page 15, line 22 Change "verifies" to "satisfies", also see page 16, line 9

--> Thank you. Manuscript was modified according to this recommendation.

- Page 15, line 50 Change “lower” to “smaller”

--> Thank you. Manuscript was modified according to this recommendation.

- Page 16, line 40. See comment with respect to page 29 concerning the confidence intervals

--> Thank you. Confidence intervals were computed using bootstraping, with n=100 replicas. A clarification was added on this point in the revised manuscript: *"Mean values with a 95% confidence interval were then computed for each estimate using bootstrapping (100 replicas)."*

- Page 17, line 12 “uncertainty” again

--> Thank you. Manuscript was modified according to this recommendation: we replaced the term "*uncertainty*" by "*variance*".

- Page 17, line 14 Change “illustrates” to “shows”

--> Thank you. Manuscript was modified according to this recommendation.

- Page 17, line 25. The R software system incorporates a great many “packages”, some are part of the standard download but others are not. The authors should be more explicit about how they used R, i.e., which “packages” and which functions within those packages..

--> Thank you for this question. We used the *GaussRF()* function from the *RandomFields* package to generate realizations of random field Z(x). We also modified the function *sobol()* from package *sensitivity* and the S3 method "*tell*" for objects of class "*sobol*", to estimate variance-based sensitivity indices with the estimators given by Lilburne and Tarantola (2009, section 3.2, eq. (8), (9), (11)-(16)). To be more explicit, we added a sentence in the revised manuscript:

*"All calculations and figures were realized in R: random realizations of Z(x) were generated with the 'GaussRF()' function from the 'RandomFields' package (Schlather 2001), while computation of sensitivity indices was based on a modified version of the 'sobol()' function from the 'sensitivity' package."*

We also added a reference on the *RandomFields* package (Schlather 2001) in the list of references.

- Page 17, line 42 Change “previous related works” to “various prior publications”

--> Thank you. Manuscript was modified according to this recommendation.

- Page 18, line 14 “influence”, also line 45

--> Thank you. Manuscript was modified according to this recommendation: we replaced the term "*influence*" by more appropriate terms throughout the manuscript.

- Page 18, lines 53 & 58 How is “accuracy” quantified? In what sense is accuracy “appropriate”or not appropriate?

--> Thank you for these questions. By "accuracy", we mean a measure of how much the input maps (e.g. a digital elevation model, a landuse map) that are used by the computer code under study, differ from the "real" spatially distributed physical variable they represent. Measuring this "accuracy" is part of the quality assessment of the input map. It could be measured for example from an extra set of validation field points, by computing the root mean square error between the value of the physical variable at validation field points and the value of the variable taken from the input map at the same locations.

The precise meaning of the term "*appropriate*" depends on the goal of the modeller. In a given context, the modeller will usually fix a limit on the variance of the output of his computer code: under this threshold, he will consider that the model output is "precise enough" for what it will be used for (decision making for example). To reach such a variance, the modeller may have to improve his knowledge on the 'real' value of some the model inputs. In that case, he'd better start working on the model inputs with the largest sensitivity indices, because reducing the variance of these inputs will likely reduce the variance of the output by a large fraction.

We rephrased this sentence in the revised manuscript:

*"Our contribution also promotes an increased awareness of the issue of sharing out efficiently, among the various inputs used by a complex computer code, the cost of collecting field data. At some point of the model building process, the modeller will usually aim at reducing the variance of the output below a given threshold, that will depend on the model use. To do so, the modeller may have to improve his knowledge on the real value of some of the model inputs, usually by collecting extra data. In this case, gathering extra field data on inputs maps that have small sensitivity indices ($S_Z(v) < 0.1$) would be unefficient, as it would be costly but could not reduce the variance of the model output by a large fraction."*

- Page 24, lines 31 & 34 Delete the ":x' ε D", that is logically incorrect. The block average is not conditioned on the random field but rather on a random variable, i.e. Z at a particular location. Also see eqs 4 & 5 on page 9

--> Thank you, the manuscript was modified according to this recommendation.

- Page 29 In the captions for Figures 3a, 4a it refers to "$Y_v$" but on the figures themselves the titles seem to use a different identification

--> Both Figures 3a and 4a show the total variance of $Y_v$ on the y-axis. The label on the figures was "Variance of $Y_v$", while the two captions said "total variance of $Y_v$". To make things clearer, we changed the labels of the y-axis on the figures to "Var[$Y_v$]".

- Page 29 In the captions for Figures 2,3,4 the authors claim that the error bars show 95% confidence intervals. How are the confidence intervals computed, did the authors use "bootstrapping" or did they assume a normal distribution to do so but with no justification. In any case they should provide clarification.

--> Confidence intervals were computed using bootstrapping, with n=100 replicas. A clarification was added on this point in the captions for Figures 2,3,4: *"Error bars show 95 % confidence interval computed by bootstrapping (100 replicas)"*