Change of Support in Spatial Variance-Based Sensitivity Analysis

**Nathalie Saint-Geours · Christian Lavergne ·
Jean-Stéphane Bailly · Frédéric Grelot**

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 as 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, that is, 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 (for example, 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 Sobol' indices · Model upscaling · Change of support · Regularization
theory · Spatial model

N. Saint-Geours · C. Lavergne
Université Montpellier 3, I3M, Montpellier, France

N. Saint-Geours (✉) · J.-S. Bailly
AgroParisTech, UMR TETIS, Montpellier, France
e-mail: saintge@teledetection.fr

J.-S. Bailly
AgroParisTech, UMR LISAH, Montpellier, France

F. Grelot
Irstea, UMR G-EAU, Montpellier, France

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; 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 (for example, 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). In these works, the computer code under study uses maps derived from field data (for example, 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 (for example, 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 (for example, porosity or evapotranspiration), that is, their large-scale properties derive from small-scale properties by simple averaging (Chilès and Delfiner 1999). 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 (for example, 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). 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, that is, models for which the computation of the model output at some location uses the values of spatial inputs at that same location only (Heuvelink et al. 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). The model output is a two-dimensional 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 Introduc-
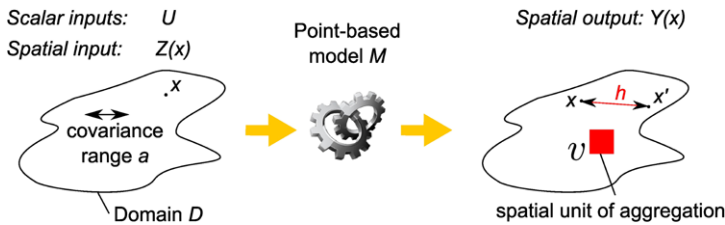
Fig. 1 Spatial model with uncertain inputs $\mathbf{U}$ and $Z(\mathbf{x})$ and spatial output $Y(\mathbf{x})$. The modeler is interested in the block average of $Y(\mathbf{x})$ over some spatial unit v

tion, 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

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

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. (See Fig. 1.)

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

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

$S_i \in [0; 1]$ 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 (with respect to output of interest $Y(\mathbf{x}^*)$) and an index on a larger support (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

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

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 $S_Z(v)$

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

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 nonspatial 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 ψ

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

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 ψ 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 sensitivity indices with Monte Carlo simulation. We chose to use the estimators and the computational procedure described by Lilburne and Rarantole (2009) based on a quasirandom 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.

Table 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)

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

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

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

where $\mathbb{E}_Z Y(\mathbf{x})$ 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 σ_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 σ_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). Using these notations, it follows from Eqs. (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) that the block variance σ_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 σ_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). It follows from Eq. (8) that

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

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 $\mathrm{Var}(Y_v)$ that is explained by the input random field $Z(\mathbf{x})$. For a small ratio (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 two-dimensional random field $Z(\mathbf{x})$ explains a small fraction of the output variance $\mathrm{Var}(Y_v)$. However, for a greater ratio (that is, 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 Lock 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; see Appendix B for a proof)

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

For most of the usual transition covariance functions (spherical, exponential, and Gaussian models), the covariance $C(\mathbf{h})$ is a monotonically 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.
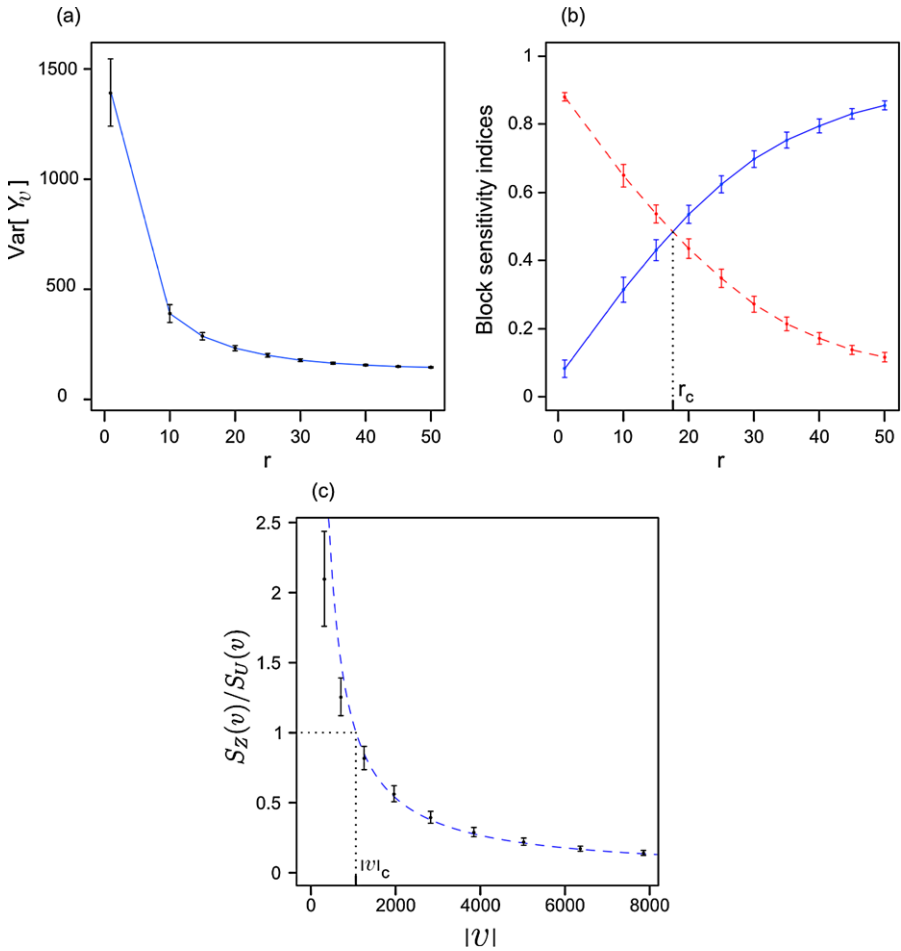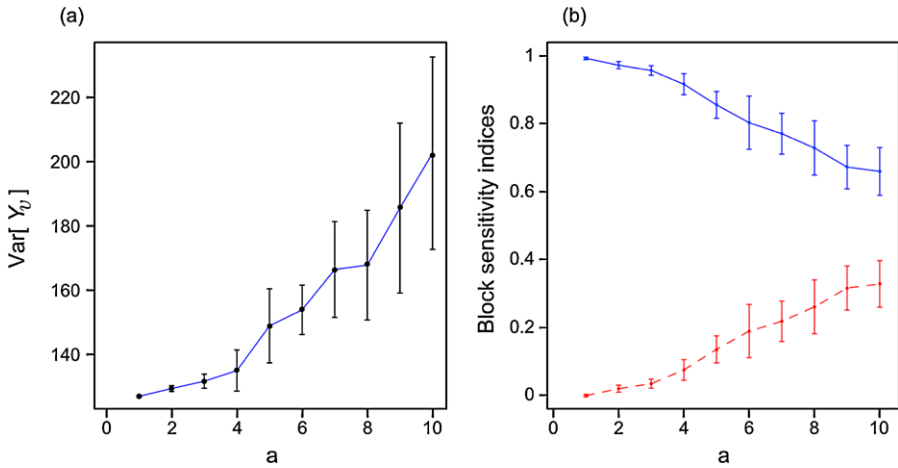
Fig. 2 GSA results depending on the size of disc-shaped support v (with radius r and area $|v| = \pi r^2$), for $a = 5$, $\eta = 0.1$: (**a**) total variance of Y_v, (**b**) block sensitivity indices $S_U(v)$ (solid line) and $S_Z(v)$ (dashed line), (**c**) ratio $S_Z(v)/S_U(v)$ with fitted curve $S_Z(v)/S_U(v) = |v|_c/|v|$ (dashed line). Error bars show 95 % confidence interval computed by bootstrapping (100 replicas)

Nugget parameter's impact on the block sensitivity indices can be interpreted in the same manner. The nugget parameter η controls the relative part of pure noise in the input random field $Z(\mathbf{x})$ (Eq. (1)). The smaller η 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 $\text{Var}(Y_v)$ will be that is explained by $Z(\mathbf{x})$. The critical size $|v|_c$ is thus a decreasing function of nugget parameter η, and the ratio of block sensitivity indices $S_Z(v)$ and $S_U(v)$ satisfies (Eqs. (1), (8), (11))

$$\frac{\partial}{\partial \eta}\left[\frac{S_Z(v)}{S_U(v)}\right] \leq 0. \tag{13}$$

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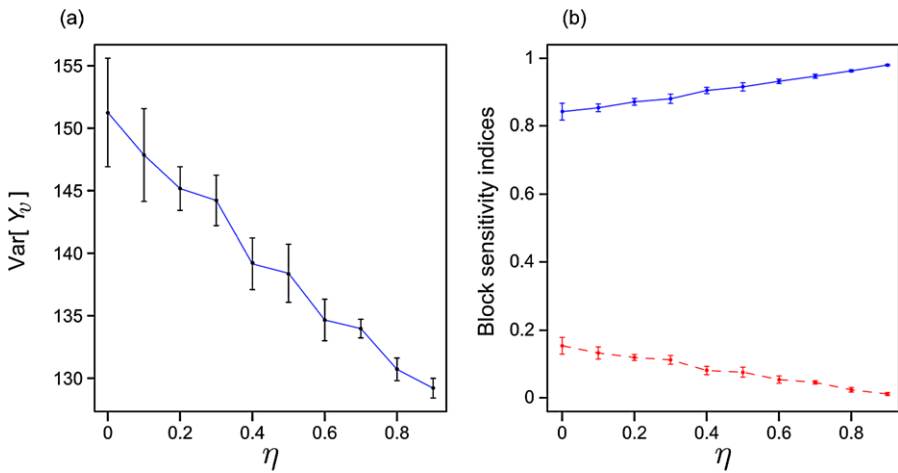
Fig. 4 GSA results depending on covariance nugget parameter η, for $a = 5$ and a disc-shaped support v of radius $r = 50$: (**a**) total variance of Y_v, (**b**) block sensitivity indices $S_U(v)$ (solid line) and $S_Z(v)$ (dashed line). Error bars show 95 % confidence interval computed by bootstrapping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

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 $\text{Var}(Y_v)$, the block sensitivity indices $S_U(v)$, $S_Z(v)$, and the ratio $S_Z(v)/S_U(v)$ 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 Eqs. (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 η (Fig. 4(b)). The opposite trends are observed for sensitivity index $S_U(v)$. 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, Fig. 2(c) shows that the ratio $S_Z(v)/S_U(v)$ is proportional to $1/|v|$ when the support v is large enough. The theoretical curve $S_Z(v)/S_U(v) = |v|_c/|v|$ (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 $|v|_c \simeq 1{,}068$. All calculations and figures were realized in 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 nonspatial inputs is proportional to $|v|_c/|v|$. The critical size $|v|_c$ 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 η. 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 modeler 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 modeler 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 inefficient, 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 η_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 $\sigma_{v,i}^2/\sigma_{v,j}^2$. Second, some environmental models are not point-based and involve spatial interactions (for example, 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 (for example, 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 Conclusions

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 $|v|_c/|v|$, where $|v|$ is the area of the spatial support v and the critical size $|v|_c$ 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 nonpoint-based models and extend our study to outputs of interest other than the average value of model output over support v.

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))

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

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

$$\mathbb{E}[Y_v \mid Z] = \mathbb{E}\left[\left(1/|v| \int_v Y(\mathbf{x})\,d\mathbf{x} \right) \Big| Z(\mathbf{x}) \right] \quad \text{(definition of } Y_v)$$

$$= 1/|v| \int_v \mathbb{E}\big[Y(\mathbf{x}) \mid Z(\mathbf{x}) \big]\,d\mathbf{x} \quad \text{(for a point-based model)}$$

$$= 1/|v| \int_v \mathbb{E}_Z Y(\mathbf{x})\,d\mathbf{x} \quad \big(\text{definition of } \mathbb{E}_Z Y(\mathbf{x})\big).$$

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

$$\mathbb{E}[Y_v \mid \mathbf{U}] = \mathbb{E}\left[\left(1/|v| \int_v Y(\mathbf{x})\,d\mathbf{x} \right) \Big| \mathbf{U} \right] \quad \text{(definition of } Y_v)$$

$$= 1/|v| \int_v \mathbb{E}\big[Y(\mathbf{x}) \mid \mathbf{U} \big]\,d\mathbf{x} \quad \text{(Fubini's theorem)}.$$

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

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

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

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

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

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

where coefficients α_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 $Z(\mathbf{x})$

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

Its covariance then gives (Chilès and Delfiner 1999)

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

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

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 Agroecosyst 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

Development Core Team R (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

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. Math 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 Environ 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