Maximizing Regional Sensitivity Analysis indices to find sensitive model behaviors

Sébastien Roux, Patrice Loisel, Samuel Buis

Maximizing Regional Sensitivity Analysis indices to find sensitive model behaviors

Sébastien Roux [1a], Patrice Loisel[a], Samuel Buis[b]

[a] *MISTEA, Univ Montpellier, INRAE, Institut Agro, Montpellier, France*
[b] *EMMAH, INRAE, Avignon Université, Avignon, France*

Abstract

We address the question of sensitivity analysis for model outputs of any dimension using Regional Sensitivity Analysis (RSA). Classical RSA computes sensitivity indices related to the impact of model inputs variations on the occurrence of a target region of the model output space. In this work, we put this perspective one step further by proposing to find, for a given model input, the region whose occurrence is best explained by the variations of this input. When it exists, this region can be seen as a model behavior whose occurrence is particularly sensitive to the variations of the model input under study. We name this method mRSA (for maximized RSA).
mRSA is formalized as an optimization problem using region-based sensitivity indices. Two formulations are studied, one theoretically and one numerically using a dedicated algorithm. Using a 2D test model and an environmental model producing time series, we show that mRSA, as a new model exploration tool, can provide interpretable insights on the sensitivity of model outputs of various dimensions.

Keywords: Multivariate sensitivity analysis, Cluster analysis, Factor mapping setting, Sobol' indices

1. Introduction

The analysis of models with multidimensional outputs (temporal, spatial, heterogeneous) is one of the current challenges of sensitivity analysis [1, 2].

The most natural way to handle such outputs is to apply classical sensitivity analysis on each of their components in order to generate multidimensional sensitivity indices (e.g. spatial maps or temporal evolution of sensitivity indices, as in [3, 4]). This approach however produces only local information while many other characterizations of the outputs distributions can be of interest in case of multi-dimensional outputs, for instance the influence of model inputs on the change in shape of model outputs. Two main approaches tackle multidimensional outputs in a non purely local way The first one aims at computing a set

[1] corresponding author email: sebastien.roux@inrae.fr

of indices associated to coordinates along a basis of functions on which the model outputs are projected ([5, 6, 7]). Its flexibility and interpretability are limited by the choice of the basis. The second approach summarizes the impact of the model inputs on the variability of all the model outputs using a single index [8, 9, 10, 11]. It provides useful aggregated indices but does not allow a fine understanding of model outputs sensitivity.

Spear and Hornberger introduced in [12] the concept of model behavior in sensitivity analysis through the Regional Sensitivity Analysis (RSA) approach.

The principle is to start from the definition of a target region of the output space (denoted as "behavioral") and to analyze the impact of the variations of model inputs on its occurrence. Using model behaviors expressed as regions of the output space appears to be an efficient method to get interpretable characterizations of model properties [13]. It has also the property to scale to any dimension of the output space. Among the last developments on RSA, two are of particular interest in the present study: i) the application of RSA in the context of reliability engineering to characterize parameter sensitivity in relation to a critical domain of the output space (e.g. the failure domain of a system) using sensitivity measures compatible with rare events and taking into account interactions (Target SA, [14, 15]), ii) its application in combination with a clustering procedure in order to characterize parameter sensitivity with respect to the dominant model behaviors detected in the model output space (Distance-Based Generalized SA [16], Cluster-based GSA, [13]).

These approaches rely on an a priori characterization of the behaviors (regions of the output space) to be analyzed. Behaviors are identified either by experts or by automatic clustering of the simulations. Pannier and Graf [17] for instance introduced sectional sensitivity analysis to study the functional interrelations between model inputs and outputs based on a priori defined subdivisions of the model inputs and outputs spaces. In this work, we propose a new perspective on the link between behaviors and sensitivity analysis based on an extension of the recent developments on RSA. Instead of trying to a priori characterize target regions of the output space, we propose to use an optimization procedure in order to reveal the region of the output space whose occurrence is best explained by the variations of this input. We name this approach mRSA (for maximized Regional Sensitivity Analysis). The formalization of mRSA in terms of principles and numerical algorithm is presented in Section 2. The results of the method application are presented in Section 3 on two examples: a model with 2D outputs and an environmental model producing time series. Software availability is described in Section 5.

2. mRSA methods and algorithm

2.1. Notation

We consider a model whose inputs are noted X_i, for $i = 1, .., n$. We assume that $X = (X_1, ..., X_n)$ is a set of independent random variables. We suppose that they follow an uniform distribution in $[0, 1]$. The model output is noted $Y = f(X_1, .., X_n)$ and is typically a multivariate output in $\mathbb{R}^d$. Model behaviors in the context of RSA are defined as regions of the model output space $Im(f) = \{f(X), X \in [0, 1]^n\}$.

2.2. Principle

The objective of this work is to find, for a given input X_i, the region C^* of the output space whose occurrence is most influenced by the variations of X_i. Note that it is equivalent to look for the partition $(C^*, \overline{C^*})$ (where $\overline{C^*}$ is the complement of C^*) such that the transition of model outputs from C^* to $\overline{C^*}$ (or equivalently from $\overline{C^*}$ to C^*) is best explained by the variations of X_i.

We formalize this question as an optimization problem using the region-based sensitivity indices proposed in [13]. These indices are based on membership functions characterizing the level of membership of any element of the output space to a given region C of the output space. They are defined by applying the variance-based Sobol' indices [18, 19] to the membership functions and quantify the impact of a given input factor on the occurrence of C. In the context of this article, we consider first order indices and binary membership functions that are written using indicator functions (i.e. the function returning 1 if the element of the output space is in the considered region, 0 otherwise). For an input X_i, the region-based index under study, noted S_i^C and associated to a region C, is then defined as:

$$S_i^C = \frac{\mathbb{V}\left[\mathbb{E}\left[\mathbb{1}_C(Y)|X_i\right]\right]}{\mathbb{V}\left[\mathbb{1}_C(Y)\right]} \tag{1}$$

where $\mathbb{V}$ is the variance, $\mathbb{E}$ the expectation, and $\mathbb{1}_C(.)$ the indicator function of set C.

In this case, the mRSA approach aims at finding for each input X_i a partition $(C_i^*, \overline{C_i^*})$ so that C_i^* (or equivalently $\overline{C}_i^*$) maximizes S_i^C. Note that this maximization does not necessarily gives $S_i^{C_i^*} > 0.5$, in other words, X_i is not necessarily the most influential input to explain the occurrence of the optimal region C_i^*.

2.3. First Formulation: unconstrained mRSA

A natural way of formalizing the sensitivity indices maximization problem is to consider the following unconstrained formulation (P1):

$$(P1) : \max_{C \subset Im(f)} S_i^C$$

The maximum value that can take S_i^C is 1. While this maximum might not often occur for real world problems, studying conditions such that solutions of (P1) reach this maximum gives insights on the mRSA approach. This study is presented in section 2.3.1. We then show in Section 2.3.2 some limitations in (P1) that will lead to another formulation (P2) on which the rest of the article focuses.

2.3.1. Characterization of partitions such that $S_i^{C^*} = 1$

We note as E_i^* the set of optimal solutions of (P1):

$$E_i^* = \left\{C \subset Im(f) \mid S_i^C = 1\right\}$$

We introduce a multi-valued function φ_i defined on $Im(f)$ by:

$$\varphi_i(y) = \left\{x_i \in [0,1] \mid \exists x_{-i} \in [0,1]^{n-1} \text{ such that } y = f(x_i, x_{-i})\right\}$$

The set $\varphi_i(y)$ contains all values x_i that can lead to y and contains at least one value (as $y \in Im(f)$). But $\varphi_i(y)$ can also contain several values or even an infinity depending on the model properties.

Let us now give a characterization of the set E_i^* of optimal solutions (see Proof in Appendix A).

Proposition 1. $(C \in E_i^*) \Leftrightarrow$ *There exists a function* F *defined on* $[0,1]$ *and valued in* $\{0,1\}$ *such that:*

$$\begin{cases} C = \{y \in Im(f) |\ \forall z \in \varphi_i(y), F(z) = 1\} \\ C \neq \emptyset \text{ and } C \neq Im(f) \end{cases}$$

Proposition 1 states that an optimal partition of $Im(f)$ verifying $S_i^{C^*} = 1$ is associated to a partition (defined by function F) of the domain of X_i. Note that on the other hand, a partition F of the domain of X_i does not necessarily defines a partition of $Im(f)$.

Remark 1 (Example with $E_i^* = \emptyset$). Let us consider a purely insensitive input factor X_i, a function F valued in $\{0,1\}$ and the set $C = \{y |\ \forall z \in \varphi_i(y), F(z) = 1\}$. As $\forall (x_i, x_i', x_{-i})$, $f(x_i, x_{-i}) = f(x_i', x_{-i})$, we deduce that for any $y \in Im(f)$, $\varphi_i(y)$ is equal to $[0,1]$. F is therefore necessarily constant and equal to one. But this implies that $C = Im(f)$, which is excluded in Proposition 1. Proposition 1 thus shows that there exists no set C^* verifying $S_i^{C^*} = 1$ in this case. Actually, in this example, for every $C \subset Im(f)$, we have $S_i^C = 0$.

Remark 2 (Example with an infinite sets of optimal solutions). Let us now consider a model verifying that each point y of the output space can be uniquely reached from a single possible value x_i. In this case, $\varphi_i(.)$ is a single valued function. Then let us consider any non-degenerated partition of $[0,1]$. Any such partition defines a non-degenerated partition of $Im(f)$ by applying model f. The resulting partition is characterized by the fact that every output point reached using a same value x_i is in the same set of the partition. Following Proposition 1, any such partition is optimal. In this case, as there is an infinite number of non-degenerated partitions of $[0,1]$, there is also an infinite set of solutions in E_i^*.

Remark 3 (Example with $S_i = 0$ and $S_i^{C^*} = 1$). Let us consider the model $Y = Sign(X_1) \cdot |X_2|$, with X_1 and X_2 following uniform distributions on $[-1,1]$ with a focus on X_2. The first order Sobol' index of X_2, noted S_2, is equal to zero. We now apply Proposition 1: as for any $y \in [-1,1]$, $\varphi_2(y) = \{-y, y\}$, all not-degenerated partitions of the output space $[-1,1]$ gathering y and $-y$ in the same set of the partition are optimal. Among others, the set $C^* = \{y \in [-1,1] \mid\ |y| < \alpha, \text{ with } \alpha \in]0,1[\}$ defines a partition $(C^*, \overline{C^*})$ verifying $S_2^{C^*} = 1$. X_2 has therefore the effect of leading the output from the center of the domain toward its boundaries (or equivalently from the boundaries to the center). This example shows the potential of mRSA in revealing the strong influence of a model input on the occurrence of specific patterns of model outputs ($S_i^{C^*} = 1$) while the standard application of sensitivity analysis methods may have led to consider it as non-important ($S_i = 0$).

2.3.2. *Limits of the unconstrained formulation*

Let us consider the function $f(x_1, x_2) = \dfrac{1}{x_1 + \lfloor 2 \cdot x_2^{-0.5} \rfloor}$ where $\lfloor x \rfloor$ is the integer part of x, and let us study $Y = f(X_1, X_2)$, with X_1 and X_2 following independent uniform distributions on $[0, 1]$. We consider problem (P1) on this model and input X_1. A sample of model responses is presented in Figure 1.

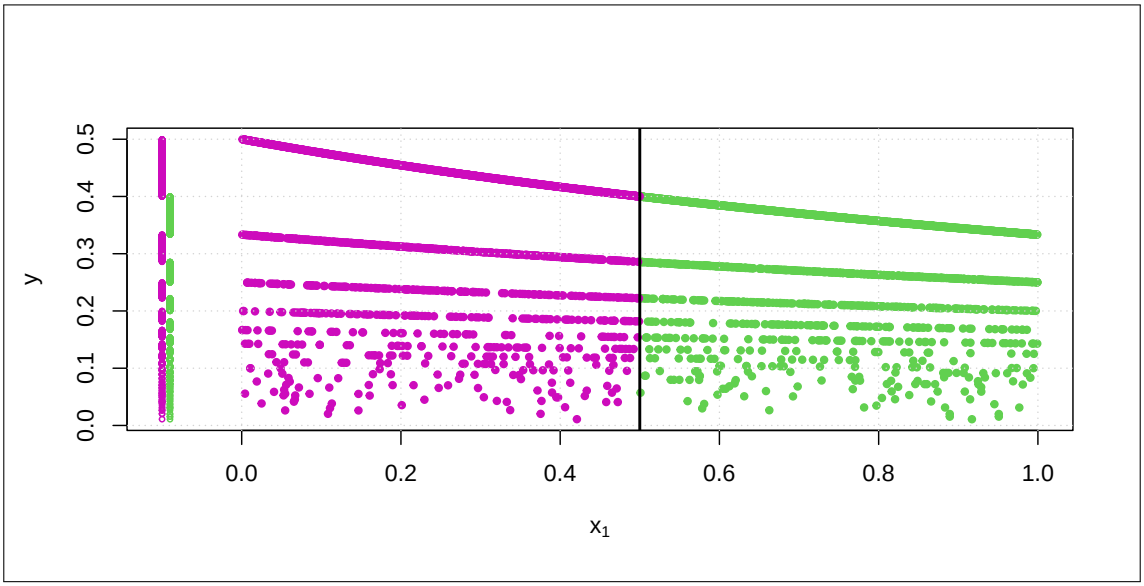

Figure 1: An optimal solution of problem (P1) on model $f(X_1, X_2) = \frac{1}{X_1 + \lfloor 2 \cdot X_2^{-0.5} \rfloor}$. Samples of Y as a function of X_1 are represented. A particular partition of the output space is showed vertically on the left, defined from a partitioning of the x_1 domain above (green values) and below 0.5 (pink values). The partition is optimal from Proposition 1, but difficult to interpret as it is made of an infinite set of sub-intervals when y tends towards zero.

This model has several interesting properties for a mRSA exploration with the unconstrained formulation (P1):

- First, neither X_1 nor X_2 is purely insensitive. Indeed, we find numerically that their first Sobol' indices are respectively about $S_1 = 0.086$ and $S_2 = 0.886$.
- This model has, for X_1, the property mentioned in *Remark* 2: any point of the output space $Im(f) = [0, \frac{1}{2}]$ is reached from a single possible value x_1. Thus the multi-valued function $\varphi_1(.)$ produces in this case only single values and can be noted as a function $f_1^{-1}(.)$. Let us consider the partitioning $(C, \overline{C})$ of $[0, \frac{1}{2}]$ defined by:

$$C = \{y \in Im(f) |\ F(f_1^{-1}(y)) = 1\} \text{ with } F(x_1) = \begin{cases} 0 \text{ if } x_1 > 0.5 \\ 1 \text{ if } x_1 \leq 0.5 \end{cases}$$

This partition is plotted vertically on Figure 1. Following Proposition 1, we have $S_1^C = 1$, which means that X_1 perfectly explains the transition of the model output between sets C and $\overline{C}$.

Nevertheless, this solution lacks interpretability, in particular because it is made of an infinite set of connected components when y tends towards 0 (see Figure 1). Moreover, when y tends toward 0, there is a full mixing of sets C and $\overline{C}$ in the sense that $\forall \epsilon > 0$, we can find $0 < y_1 < y_2 < y_3 < \epsilon$ such that $y_1 \in C, y_2 \in \overline{C}, y_3 \in C$. The region whose occurrence is best explained by the variation of X_1 is thus difficult to characterize in this case. This example shows that additional constraints must be added to the formulation of Problem (P1) in order to increase the interpretability of the solutions, in particular regarding the risk of near-overlapping between C and $\overline{C}$.

2.4. Improved Formulation: constrained mRSA

To overcome the limitations described in the previous section, we now add constraints to Problem (P1) by searching only solutions belonging to a set denoted as $\mathcal{I}$ (Problem (P2)):

$$(P2) : \max_{C \subset Im(f),\ C \in \mathcal{I}} S_i^C$$

This set $\mathcal{I}$ is introduced to increase the chances that a solution is interpretable by an expert of the model. There is no general definition of this interpretability, but we propose a trick to improve the method in this direction. We first suppose that a set of K_Y non-overlapping clusters $(C_1, .., C_{K_Y})$ covering $Im(f)$ has been defined. We denote as *elementary clusters* these K_Y clusters. A *meta-cluster* is defined as a union of elementary clusters. $\mathcal{I}$ is then defined as the set of all possible meta-clusters. This choice: i) allows to avoid the mixing effect illustrated in the 1D example presented in the previous section by choosing an appropriate value for K_Y, ii) adds an indirect constraint on the number of connected components of the solution (depending on the number of elementary clusters), iii) adds the possibility to define elementary clusters either from expert knowledge or using an automatic algorithm. For these reasons, we found that Problem (P2) is more relevant in practice than Problem (P1). However, Problem (P2) must be solved numerically using a dedicated algorithm, since the cost of testing all possible meta-clusters ($2^{K_Y - 1} - 1$) is prohibitive.

2.5. Algorithm for the constrained formulation (P2)

2.5.1. Required simulated data

The numerical resolution of Problem (P2) requires a simulation sample noted $(\mathbf{X}, \mathbf{Y})$, where $\mathbf{X}$ is a matrix of parameters values, and $\mathbf{Y}$ the matrix of multivariate simulated values obtained by applying the model under study on $\mathbf{X}$. For the sake of simplicity, we suppose here that $\mathbf{X}$ is made of N i.i.d parameters values, but note that the algorithm presented in Section 2.5.3 can also be applied on not i.i.d samples with the only consequence that this can lead to sub-optimal results.

2.5.2. Aggregation of elementary clusters $(C_1, .., C_{K_Y})$

We designed a hierarchical clustering procedure using an aggregating property in order to merge iteratively elementary clusters $(C_1, .., C_{K_Y})$.

Let us introduce a partition of $[0, 1]$ into n_X bins $(B_1, .., B_{n_X})$, typically with $B_j =]\frac{j-1}{n_X}, \frac{j}{n_X}]$, for $j = 1, .., n_X$. For a subset C of the simulated output values and a bin B_j, let us define h_j^C such that:

$$h_j^C = Card\{y \in C | \ \exists \ x_i \in B_j \text{ such that } y = f(x_i, x_{-i})\}$$

The vector $h^C = (h_1^C, .., h_{n_X}^C)$ is an approximate representation of the distribution $\mathcal{L}(X_i|Y \in C)$ and is a (non normalized) histogram associated to cluster C. Let us note $\widetilde{S_i}^C$ the estimate of ${S_i}^C$ obtained by using histogram h^C (its expression can be found in Appendix B).

Proposition 2. *Let us consider two elementary clusters C and C' with histograms h and h' satisfying $h' = \theta h$, with $\theta > 0$. Let us denote $(C^*, \overline{C^*})$ the partition of the output space maximizing $\widetilde{S_i}^C$ over the set $\mathcal{I}$. Then clusters C and C' belong both to C^* or to $\overline{C^*}$.*

Proposition 2 (see Proof in Appendix C) states that elementary clusters having proportional histograms cannot be in different sets of the optimal partition $(C^*, \overline{C^*})$ for criterion $\widetilde{S_i}^C$. It therefore suggests that meta-clusters should be defined by aggregating elementary clusters whose histograms are the closest to be proportional. We proposed to do that by using a clustering method based on the histogram correlation distance $d(h, h') = 1 - Cor(h, h')$. This procedure will be used in Algorithm mRSA presented in the coming section. Note that while the estimator $\widetilde{S_i}^C$ was useful to establish Proposition 2, it is however not as precise as classical Sobol' indices estimators. It is therefore not used in practice to make any computation.

2.5.3. Algorithm mRSA

Algorithm mRSA is based on: i) the restriction of the search space to meta-clusters defined from $(C_1, .., C_{K_Y})$ to handle the constraint $C \in \mathcal{I}$, ii) the use of the aggregation method described in the previous section.

The first step of Algorithm mRSA consists in clustering the simulated values using an automatic or user-defined clustering procedure $ClustFun^Y$ to get K_Y elementary clusters. Then a clustering of these K_Y elementary clusters into K_H meta-clusters is done using the method noted $ClustFun^H$. $ClustFun^H$ is a classical hierarchical clustering method using the aggregation procedure described in the previous section. As the last step of the algorithm, an exhaustive search for the best partition over all the $(2^{K_H-1} - 1)$ possible merging of the K_H meta-clusters is performed by estimating $S_i^{P_k}$ for each candidate partition $(P_k, \overline{P_k})$. This algorithm has the following parameters: the index i of the input under study, the pre-clustering method $ClustFun^Y$, the number K_Y of clusters for the pre-clustering step, the number n_X of bins for the histograms computations, the number of meta-clusters K_H and a size parameter γ which is introduced to exclude meta-clusters having to few elements.

Algorithm mRSA

Input data: $(\mathbf{X}, \mathbf{Y})$
Parameters: $(i, ClustFun^Y, K_Y, n_X, K_H, \gamma)$
Output : $(P_{k^*}, \overline{P_{k^*}}), \widehat{S}_i^{P_{k^*}}$

Apply $ClustFun^Y$ on $\mathbf{Y}$ to get K_Y elementary clusters $(C_1, .., C_{K_Y})$
Compute all histograms $(h^{C_k})_{k=1..K_Y}$ for input X_i discretized with n_X bins
Apply $ClustFun^H$ on $(h^{C_1}, .., h^{C_{K_Y}})$ to get K_H meta-clusters $(\widehat{C}_1, .., \widehat{C}_{K_H})$
for all 2-partitions $(P_k, \overline{P_k})$ of $(\widehat{C}_1, .., \widehat{C}_{K_H})$ **do**
  Compute an estimate $\widehat{S}_i^{P_k}$ of $S_i^{P_k}$ from $(\mathbf{X}, \mathbf{Y})$
  Compute $\gamma_k = \frac{1}{N} \min\left(Card(P_k), Card(\overline{P_k})\right)$, where N=number of rows in $\mathbf{X}$
end for
$k^* = \underset{k, \gamma_k \geq \gamma}{\arg\max}\, \widehat{S}_i^{P_k}$

2.5.4. Parameter tuning guidelines

We propose in Table 1 guidelines to set the parameters of Algorithm mRSA based on parameters definitions and numerical experiments.
The sample size N (number of rows in $\mathbf{X}$ and $\mathbf{Y}$) has to be i) large enough for precise estimation of first order indices using $(\mathbf{X}, \mathbf{Y})$, but also ii) large enough for a good covering of the output space in order to obtain partitions as close as possible to an optimal solution. There is no general rule for this tuning. For the two case studies, samples of size 4000 and 5000 were enough to reveal interpretable partitions of the output space associated with high first order sensitivity indices.

Parameter K_H which drives the exhaustive search should be set to the highest value allowed by the computational budget (typically lower than 20). Parameter γ was typically set to 0.1 (10% of the number of simulated points). Parameter n_X should be set in order to avoid histograms with too few values per bin, which was satisfied with n_X around 10-20 in our tests. It also turned out to be quite insensitive in the numerical experiments we carried out. On the other hand, the pre-clustering is a critical step. As explained previously, the elementary clusters can be defined by the user to constrain the shape of the solution according to a prior knowledge. An automatic clustering can also be used, such as the hierarchical clustering using euclidean distances used in the numerical examples presented in Section 3. Note that depending on the output space and user knowledge, other procedures may lead to better results. The number of elementary clusters K_Y for this pre-clustering step was found very sensitive in the numerical experiments (see Section 3). Its optimal setting appeared to be a trade-off between the interpretability of the partition (better for low K_Y) and the sensitivity score of the partition (better for large K_Y). A value of $K_Y = 500$ was found to be satisfactory for the different case studies.

Name	Function	Guideline	Value	
			models $\mathcal{M}_{2D}$ (§ 3.1, 3.2)	model CANTIS (§ 3.3)
N	initial sample size	large enough for Sobol' indices estimation and for good covering of output space	4000	5000
K_Y	number of elementary clusters	trade-off between interpretability (small K_Y) and optimality (large K_Y); $K_Y \leq N$	2000, 500, 30	500
$ClustFun^Y$	method for pre-clustering the output space	to be adapted to the output space/ user expertise	HCA^{euc}	HCA^{euc}
K_H	number of meta-clusters	maximum value allowed by computing time budget	12	12
n_X	number of histogram bins	10-20	18	18
γ	relative minimum size of meta-clusters	user requirement	0.1	0.1

Table 1: Guidelines for setting the parameters of the mRSA algorithm and values used for the different tests. HCA^{euc} is a classical hierarchical clustering algorithm used with an euclidean distance.

3. Numerical applications

3.1. Noise-free 2D test model

3.1.1. Model definition and properties

We first consider the following test model $(Y_1, Y_2) = \mathcal{M}_{2D}(X_1, X_2, X_3)$, where:

$$\begin{aligned} Y_1 &= c_1 + 0.25 \cdot \sqrt{X_3} \cdot \cos(2\pi X_2) \\ Y_2 &= c_2 + 0.25 \cdot \sqrt{X_3} \cdot \sin(2\pi X_2) \end{aligned} \quad \text{with } (c_1, c_2) = \begin{cases} (0.5, 0.316), & \text{if } X_1 < 0.33 \\ (0.25, 0.75), & \text{if } 0.33 \leq X_1 < 0.66 \\ (0.75, 0.75), & \text{if } X_1 \geq 0.66 \end{cases}$$

Model $\mathcal{M}_{2D}$ is studied for X_1, X_2, X_3 following independent uniform distributions on $[0, 1]$. Basically, the model draws a point in the plane at a distance from a reference point related

to the value of X_3 and with an angle to the horizontal line related to the value of X_2. The coordinates of the three possible reference points are defined from the value of X_1. The set of possible positions for (Y_1, Y_2) is the union of three disks $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$ centered on the reference points (see Figure 2). We choose the coordinates of the reference points so that any point of the output space is obtained from exactly one set of values of the parameters.

By applying Proposition 1, we know what the solutions to Problem (P1) are for model $\mathcal{M}_{2D}$:

- X_1: $\varphi_1(y)$ is one of the three intervals $I_1 = [0, \frac{1}{3}[, I_2 = [\frac{1}{3}, \frac{2}{3}[, I_3 = [\frac{2}{3}, 1]$, each associated to one of the three disks $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$. There are therefore three non-degenerated solutions of (P1) for model $\mathcal{M}_{2D}$: $(\mathcal{C}_1, \overline{\mathcal{C}_1}), (\mathcal{C}_2, \overline{\mathcal{C}_2}), (\mathcal{C}_3, \overline{\mathcal{C}_3})$.

- X_2, X_3: for each of these inputs and for each $y \in Im(\mathcal{M}_{2D})$, $\varphi_i(y)$ has a single value. Thus, from Remark 2 following Proposition 1, any non-degenerated partition of the X_2 (resp. X_3) domain defines an optimal partition of the output space. For both X_2 and X_3, the optimal solution is made of a sub-partition replicated on all disks $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$. For X_2, optimal partitions have an angular pattern (i.e. all outputs y having the same angle to the closest center are in the same set of the partition) resulting in partitions made of angular sectors. For X_3, optimal partitions have a radial pattern (i.e. all y at the same distance from the closest center are in the same set of the partition) resulting in partitions made of rings.

3.1.2. Results

We applied Algorithm mRSA to the model $\mathcal{M}_{2D}$ for problem (P2). A set of 2000 elementary clusters is used to define the set $\mathcal{I}$. Elementary clusters are obtained using a hierarchical clustering algorithm with an euclidean distance on $\mathbb{R}^2$. The values of the parameters used for the application of Algorithm mRSA are given in Table 1. The sensitivity indices are computed using the estimation method based on ranks [20] implemented in sobolrank function of the R package sensitivity [21], applied on a Monte Carlo sample of size $N = 4000$.

The results are displayed as a 35×35 image with the following convention: red pixels contain only points belonging to C^*, blue pixels contain only points belonging to $\overline{C^*}$, yellow pixels include points from C^* and $\overline{C^*}$. This display allows to quickly assess regions where many points of the two clusters are close together (which leads to poor interpretability).

The results are presented in Figure 2. As we can see, the optimized values of $\widehat{S}^{C^*}$ are equal to one for each input, showing that the algorithm is able to find optimal solutions for $i = 1, 2, 3$. The partitions found are consistent with the set of optimal solutions of (P1), meaning that in this simple example and with the quite large number of elementary clusters used, (P1) and (P2) are equivalent, and the solutions obtained are interpretable. This example shows the ability to find optimal numerical solutions in a favorable case.

3.2. 2D test model with noise

3.2.1. Model definition and properties

In order to test the algorithm properties in a more complicated case, we add a noise in the formulation of model $\mathcal{M}_{2D}$ to define a new model $(Y_1^{(r)}, Y_2^{(r)}) = \mathcal{M}_{2D}^{(r)}(X_1, X_2, X_3, X_4)$.

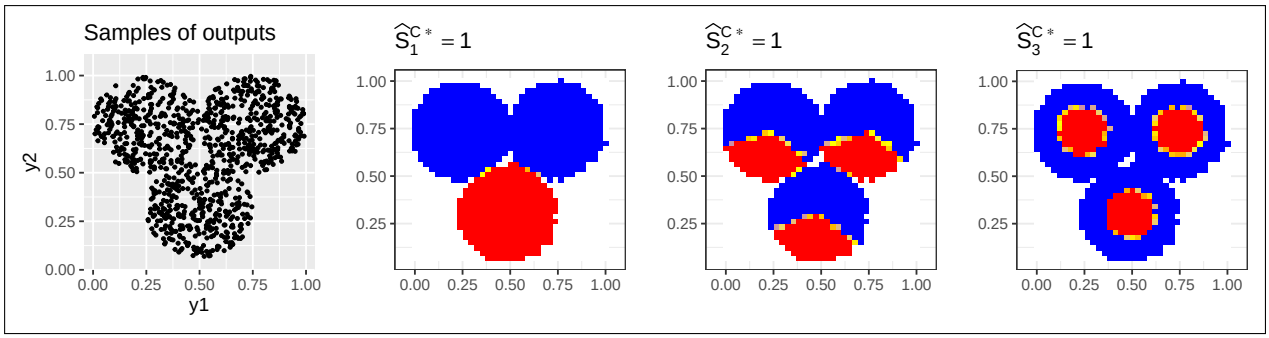

Figure 2: Results of the mRSA algorithm on model $\mathcal{M}_{2D}$ (model with 2D outputs without noise). Left: sample of model outputs, Right: optimized partitions found for each model input (red: pixels belonging to C^*, blue: pixels belonging to $\overline{C^*}$, yellow: pixels including points from C^* and points from $\overline{C^*}$). For each X_i, the values of $\widehat{S}_i^{C^*}$ are equal to 1.

The idea is to have a fourth parameter X_4 (following a uniform distribution in $[0, 1]$) that adds a small displacement to the output of $\mathcal{M}_{2D}$.

In order to avoid ending up with a stochastic model (which would raise additional issues for the definition and estimation of the Sobol' indices [22], outside the scope of the present study), the perturbation is defined in a deterministic way by taking the row of index $\lfloor 100 \cdot X_4 \rfloor$ of a pre-defined matrix Υ. Matrix Υ is of size $(100, 2)$ and made of small 2D displacements sampled from random distributions. $(Y_1^{(r)}, Y_2^{(r)}) = \mathcal{M}_{2D}^{(r)}(X_1, X_2, X_3, X_4)$ is therefore defined as:

$$Y_1^{(r)} = Y_1 + \Upsilon_{\lfloor 100 \cdot X_4 \rfloor, 1}$$
$$Y_2^{(r)} = Y_2 + \Upsilon_{\lfloor 100 \cdot X_4 \rfloor, 2}$$

Due to the perturbations, Problem (P1) no longer has simple solutions when considering model $\mathcal{M}_{2D}^{(r)}$. Indeed, for many points y of the output space, $\varphi_i(y)$ is multi-valued: there are many different settings leading to a same value y (actually all (x_1, x_2, x_3) such as $\mathcal{M}_{2D}(x_1, x_2, x_3) = y - \Upsilon_{i,.}$, where $\Upsilon_{i,.}$ is one row of Matrix Υ). We therefore do not know a priori the solutions of (P1) or (P2) in this case. Nevertheless, as the amplitude of the noise decreases, we expect to retrieve the solutions found for the model $\mathcal{M}_{2D}$.

3.2.2. Results

We applied Algorithm mRSA on problem (P2) for model $\mathcal{M}_{2D}^{(r)}$ for $K_Y = \{2000, 500, 30\}$ elementary clusters in the pre-clustering phase of the algorithm. The results are presented in Figure 3 for all the inputs (including the one driving the perturbations).

As expected, the partitions found are related to the ones obtained with model $\mathcal{M}_{2D}$, but are blurred. We can also notice that the level of noise depends on the setting of the parameter K_Y: there is more noise for high values of K_Y, which makes the interpretation of the results more difficult. On the other hand, lowering K_Y too much leads to more clearly defined partitions but at the expense of their optimality. For example, when looking at the

sensitivity scores $\widehat{S}_2^{C*}$ as a function of K_Y, we have $\widehat{S}_3^{C*}(2000) = 0.435$, $\widehat{S}_3^{C*}(500) = 0.315$, and $\widehat{S}_3^{C*}(30) = 0.234$. As mentioned in the guidelines given in Table 1, there is a compromise to be found in the value of K_Y to obtain results that are both precise and interpretable. In this example, $K_Y = 500$ seems to be a relevant choice.

Another notable result is that the patterns found are not the same depending on the value of K_Y. However, we can note that they all correspond to an optimal solution of the previous problem. For example there are different angular patterns for the solutions found for X_2. Lastly, it is interesting to look at the patterns and associated sensitivity scores obtained for the perturbation parameter X_4. As expected, the scores are low and the patterns made of a lot of connected components even for $K_Y = 500$. The algorithm therefore does not create a falsely interpretable structure associated with this perturbation parameter.

3.3. Environmental model simulating temporal outputs

3.3.1. Model and input distributions

We applied the mRSA algorithm on the environmental model CANTIS [23], which simulates the decomposition of organic biomass in crop residues over time based on a set of differential equations. We use the same model setting as in [13]. The model setting consists in ten uncertain model inputs and the analysis of one time-varying output: the zymogenous microbial biomass, here-after referred to as ZYB. The reader is invited to refer to [13] for more details. The sensitivity indices are computed using the estimation method based on ranks [20] implemented in sobolrank function of the R package sensitivity [21], applied on a Monte Carlo sample of size $N = 5000$ producing 5000 output curves of ZYB. The other parameters values used to apply Algorithm mRSA are given in Table 1.

3.3.2. Results

We present in Table 2 the sensitivity indices $\widehat{S}_i^{C^*}$ obtained when applying the Algorithm mRSA on each input of the CANTIS model in the considered uncertainty setting. Only X_8 and X_9 exhibits large values that deserve further analysis. For these two inputs, we plotted in Figure 4 the optimal clusters found.

For input X_8, we obtained $\widehat{S}_8^{C^*} = 0.851$. Looking at Figure 4, we can see that the two clusters identified differ the most at the beginning of the simulation time period. The two clusters then overlap rapidly. The application of the method indicates that X_8 drives the occurrence of biomass dynamics that start by increasing (or equivalently that start by decreasing).

For input X_9, a very high score ($\widehat{S}_9^{C^*} = 0.874$) was also obtained. For this input, when examining Figure 4, it appears that what best distinguishes the two clusters is the level of ZYB values over the entire simulated time period, and particularly the final values. We conclude that X_9 drives the occurrence of dynamics with the highest (or the lowest) biomass levels all along the considered time scale and particularly for the final value.

4. Conclusion

In this work, a new sensitivity analysis approach extending Regional Sensitivity Analysis and named mRSA was introduced with the aim of automatically identifying the model

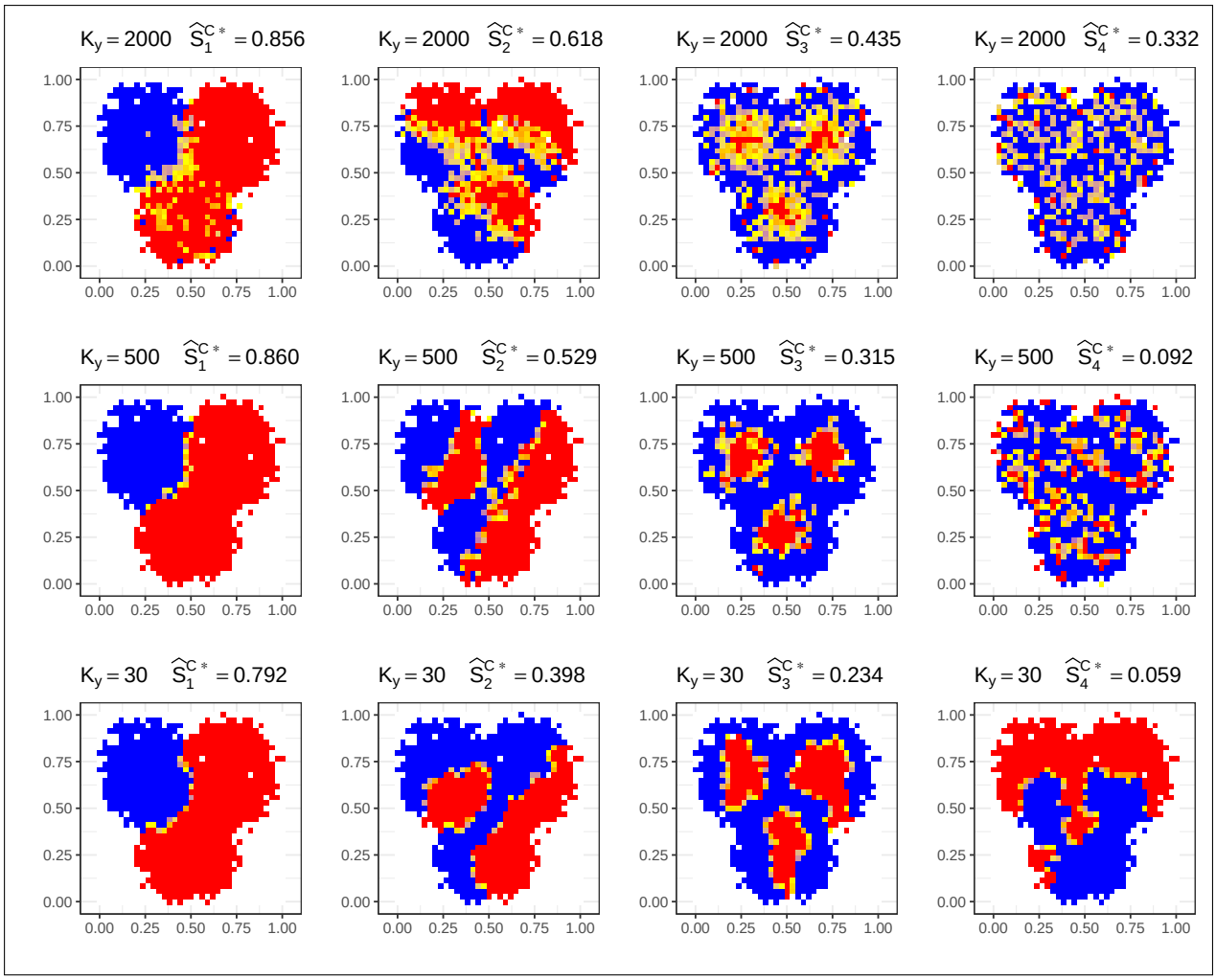

Figure 3: Results of the mRSA algorithm on the 2D model with noise $\mathcal{M}_{2D}^{(r)}$ with $K_Y = \{2000, 500, 30\}$ for the four model inputs. Red: pixels belonging to C^*, blue: pixels belonging to $\overline{C^*}$, yellow: pixels including points from C^* and points from $\overline{C^*}$.

behaviors best explained by the variations of the different model inputs. A first formalization of mRSA as an unconstrained maximization problem using a sensitivity based criterion was theoretically studied. An improved formulation (constrained mRSA), more interesting in practice, was introduced and solved numerically with a dedicated algorithm. The application of this algorithm on an environmental model producing biomass dynamics showed the ability of mRSA to reveal, both in a quantitative and graphical way, some impact of the model inputs variations that would not have been easy to detect without strong prior knowledge. This approach opens new perspectives, particularly for the study of complex models with multivariate outputs.

In this study, only first order indices were considered in the optimization procedure.

X_i	X_1	X_2	X_3	X_4	X_5	X_6	X_7	$\boldsymbol{X_8}$	$\boldsymbol{X_9}$	X_{10}
$\widehat{S}_i^{C*}$	0.069	0.066	0.083	0.069	0.078	0.059	0.069	**0.851**	**0.874**	0.062

Table 2: Optimized region-based sensitivity indices obtained for each input X_i for the environmental model CANTIS.

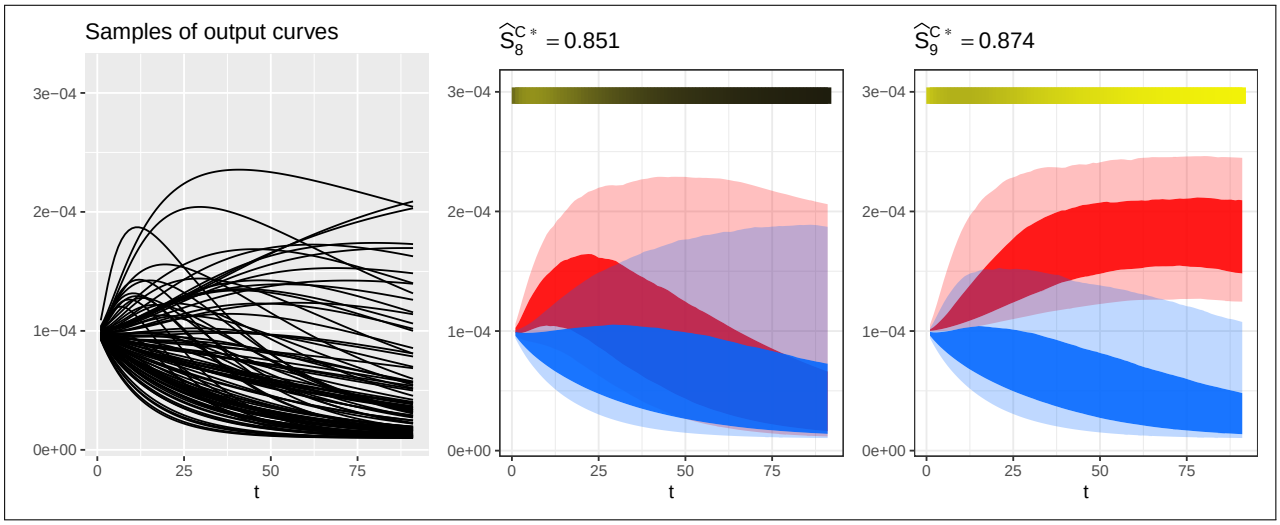

Figure 4: Clustering obtained with Algorithm mRSA on the CANTIS model. Left: samples of output curves. Right: results of the mRSA application for input X_8 and X_9. Sensitivity scores and optimal partitions are represented. The optimal partitions are plotted using quantiles (0.05,0.25,0.75,0.95) as a function of time. The horizontal colorbar indicates the significance of the difference between the two partitions at a given time: yellow (resp. black) corresponds to a high (resp. low) difference as given by Kolmogorov–Smirnov statistic.

Other indices could have been considered, such as indices of groups of parameters, or indices quantifying some interaction effects of interest. Looking for partitions optimizing such indices would enlarge the possible model properties explored. The question of having efficient algorithm to find solutions in such cases is however challenging and is a direction for future work.

5. Software availability

R codes for applying Algorithm mRSA on the toy model and on the CANTIS model are available from a Gitlab repository located at `https://forgemia.inra.fr/sebastien.roux/mrsa_paper/`.

References

[1] A. Marrel, N. Saint-Geours, M. De Lozzo, Sensitivity analysis of spatial and/or temporal phenomena, Handbook of Uncertainty Quantification (2017) 1327–1357.

[2] A. Saltelli, A. Jakeman, S. Razavi, Q. Wu, Sensitivity analysis: A discipline coming of age, Environmental Modelling & Software 146 (2021) 105226.

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

[4] M. De Lozzo, A. Marrel, Sensitivity analysis with dependence and variance-based measures for spatio-temporal numerical simulators, Stochastic environmental research and risk assessment 31 (2017) 1437–1453.

[5] K. Campbell, M. D. McKay, B. J. Williams, Sensitivity analysis when model outputs are functions, Reliability Engineering & System Safety 91 (10-11) (2006) 1468–1472.

[6] M. Lamboni, D. Makowski, S. Lehuger, B. Gabrielle, H. Monod, Multivariate global sensitivity analysis for dynamic crop models, Field Crops Research 113 (3) (2009) 312–320.

[7] S. Xiao, Z. Lu, P. Wang, Multivariate global sensitivity analysis for dynamic models based on wavelet analysis, Reliability Engineering & System Safety 170 (2018) 20–30.

[8] M. Lamboni, H. Monod, D. Makowski, Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models, Reliability Engineering & System Safety 96 (4) (2011) 450–459.

[9] F. Gamboa, A. Janon, T. Klein, A. Lagnoux, Sensitivity analysis for multidimensional and functional outputs, Electron. J. Statist. 8 (1) (2014) 575–603.

[10] M. Lamboni, Multivariate sensitivity analysis: Minimum variance unbiased estimators of the first-order and total-effect covariance matrices, Reliability Engineering & System Safety 187 (2019) 67–92.

[11] L. Xu, Z. Lu, S. Xiao, Generalized sensitivity indices based on vector projection for multivariate output, Applied Mathematical Modelling 66 (2019) 592–610.

[12] R. Spear, G. Hornberger, Eutrophication in peel inlet—ii. identification of critical uncertainties via generalized sensitivity analysis, Water Research 14 (1) (1980) 43–49.

[13] S. Roux, S. Buis, F. Lafolie, M. Lamboni, Cluster-based gsa: Global sensitivity analysis of models with temporal or spatial outputs using clustering, Environmental Modelling & Software 140 (2021) 105046.

[14] A. Marrel, V. Chabridon, Statistical developments for target and conditional sensitivity analysis: application on safety studies for nuclear reactor, Reliability Engineering & System Safety (2021) 107711.

[15] M. Il Idrissi, V. Chabridon, B. Iooss, Developments and applications of shapley effects to reliability-oriented sensitivity analysis with correlated inputs, Environmental Modelling & Software 143 (2021) 105115.

[16] D. Fenwick, C. Scheidt, J. Caers, Quantifying asymmetric parameter interactions in sensitivity analysis: application to reservoir modeling, Mathematical Geosciences 46 (4) (2014) 493–511.

[17] S. Pannier, W. Graf, Sectional global sensitivity measures, Reliability Engineering & System Safety 134 (2015) 110–117.

[18] I. M. Sobol, Sensitivity estimates for nonlinear mathematical models, Mathematical Modelling and Computational Experiments 1 (4) (1993) 407–414.

[19] A. Saltelli, K. Chan, E. M. Scott, et al., Sensitivity analysis, Wiley New York, 2000.

[20] F. Gamboa, P. Gremaud, T. Klein, A. Lagnoux, Global sensitivity analysis: A novel generation of mighty estimators based on rank statistics, Bernoulli 28 (4) (2022) 2345–2374.

[21] B. Iooss, S. Da Veiga, A. Janon, G. Pujol, Package 'sensitivity', R package https://cran.r-project.or/package=sensitivity (2024).

[22] G. Mazo, A trade-off between explorations and repetitions for estimators of two global sensitivity indices in stochastic models induced by probability measures, SIAM/ASA Journal on Uncertainty Quantification 9 (4) (2021) 1673–1713.

[23] P. Garnier, C. Néel, B. Mary, F. Lafolie, Evaluation of a nitrogen transport and transformation model in a bare soil, European Journal of Soil Science 52 (2) (2001) 253–268.

Appendix A: Proof of Proposition 1

We first recall the Sobol' Hoeffding decomposition over $[0,1]^n$, which states that any function g square-integrable over $[0,1]^n$ can be uniquely decomposed in the following form:

$$g(X) = g_0 + \sum_{i_1=1}^{n} g_{i_1}(X_{i_1}) + \sum_{i_2>i_1} g_{i_1 i_2}(X_{i_1}, X_{i_2}) + ... + g_{i_1..i_n}(X_{i_1}, .., X_{i_n})$$

where

$$\mathbb{E}[g_{i_1..i_s}(X_{i_1}, .., X_{i_s})] = 0, \forall \{i_1...i_s\} \subseteq \{1, 2, ..., n\}$$
$$\mathbb{E}[g_{i_1..i_s}(X_{i_1}, .., X_{i_s}) g_{j_1..i_t}(X_{j_1}, .., X_{j_t})] = 0, \forall \{i_1, ...i_s\}) \neq \{j_1, ..., j_t\}$$

We also recall that $S_i = 1$ (where S_i is the first Sobol' index associated to X_i) implies that for all multi-index $j \neq \{i\}$, the total Sobol' index $S_j^T = 0$, which implies that all terms in decomposition except g_i have null variances. Therefore, $S_i = 1$ implies that:

$$g(X) = g_0 + g_i(X_i) \tag{2}$$

Necessary Condition We suppose $C \in E_i^*$ (thus $C \neq \emptyset$ and $C \neq Im(f)$). As $S_i^C = 1$, when applying Equation (2) to $\mathbb{1}_C(.)$, we find that there exists a function $F_i(.)$ valued in {0,1}, such that $\forall (X_1, .., X_n)$ in their variation domain:

$$\mathbb{1}_C(f(X_1, .., X_n)) = F_i(X_i)$$

If $y \in C$ (resp. $\overline{C}$), any (x_i, x_{-i}) such that $y = f(x_i, x_{-i})$ verifies $F_i(x_i) = 1$ (resp. 0). Therefore $y \in C$ (resp. $\overline{C}$) implies that $\forall z \in \varphi_i(y), F_i(z) = 1$ (resp. 0). Therefore F_i satisfies the conditions of Proposition 1.

Sufficient Condition We suppose the existence of a set C satisfying the right hand side of Proposition 1 and calculate S_i^C. There exists a function F such that $\forall (X_1, .., X_n)$ in their variation domain:

$$\mathbb{1}_C(f(X_1, .., X_n)) = F(X_i)$$

This implies that:

$$\mathbb{V}[\mathbb{E}[\mathbb{1}_C(f(X_1, .., X_n))|X_i]] = \mathbb{V}[F]$$

Moreover, as $C \neq \emptyset$ and $C \neq Im(f)$, we have $\mathbb{V}[\mathbb{1}_C(f(X_1, .., X_n))] = \mathbb{V}[F] \neq 0$. Therefore $S_i^C = 1$.

Appendix B: Expression of $\widetilde{S_i^C}$, an approximation of S_i^C using h^C

Equation (1) can be written as:

$${S_i}^C = \frac{1}{\mathbb{V}[\mathbb{1}_C(Y)]} \mathbb{E}_{X_i} \left[\mathbb{E}[\mathbb{1}_C(Y)|X_i] - \mathbb{E}[\mathbb{E}[\mathbb{1}_C(Y)|X_i]] \right]^2$$

We note that, with the discretization of X_i into n_X bins:

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[\mathbb{1}_C(Y)|X_i]] &= \frac{1}{n_X}\sum_{j=1}^{n_X}\mathbb{E}[\mathbb{1}_C(Y)|X_i \in B_j] \\
\mathbb{V}[\mathbb{1}_C(Y)] &= \mathbb{E}[(\mathbb{E}[\mathbb{1}_C(Y)|X_i] - \mathbb{E}[\mathbb{E}[\mathbb{1}_C(Y)|X_i]])^2] \\
&= \frac{1}{n_X}\sum_{j=1}^{n_X}\mathbb{E}[\mathbb{1}_C(Y)|X_i \in B_j] - (\frac{1}{n_X}\sum_{j=1}^{n_X}\mathbb{E}[\mathbb{1}_C(Y)|X_i \in B_j])^2
\end{aligned}$$

S_i^C can be rewritten with the discretization of X_i into n_X bins by:

$$S_i^C = \frac{\frac{1}{n_X}\sum_{j=1}^{n_X}\left(\mathbb{E}[\mathbb{1}_C(Y)|X_i \in B_j] - \frac{1}{n_X}\sum_{l=1}^{n_X}\mathbb{E}[\mathbb{1}_C(Y)|X_i \in B_l]\right)^2}{\frac{1}{n_X}\sum_{j=1}^{n_X}\mathbb{E}[\mathbb{1}_C(Y)|X_i \in B_j].(1 - \frac{1}{n_X}\sum_{j=1}^{n_X}\mathbb{E}[\mathbb{1}_C(Y)|X_i \in B_j])}$$

Moreover, using the simulation sample of size N, we have the approximation:

$$\mathbb{E}[\mathbb{1}_C(Y)|X_i \in B_j] \approx \frac{n_X.h_j^C}{N}$$

Using this approximation, we deduce that the following expression noted $\widetilde{S_i}^C$ is an approximation of S_i^C:

$$\widetilde{S_i^C} = \frac{n_X}{\sum_{j=1}^{n_X} h_j^C.(N - \sum_{j=1}^{n_X} h_j^C)}\sum_{j=1}^{n_X}(h_j^C - \frac{1}{n_X}\sum_{l=1}^{n_X} h_l^C)^2 = \frac{||h^C - H^C||^2}{H^C.(N - n_X H^C)}$$

where $H^C = \frac{1}{n_X}\sum_{j=1}^{n_X} h_j^C$.

Appendix C: Proof of Proposition 2

Let us consider a cluster C_0 (having histogram h^0 with mean H_0) that we either merge with a cluster C (having histogram h and histogram mean H) to form cluster $C_1 = C_0 \cup C$, or with two clusters C and C' having respective histograms h and $h' = \theta h$ to form cluster $C_2 = C_0 \cup C \cup C'$. The region-based sensitivity criteria based on histograms for these three sets are $\widetilde{S_i}^{C_0}$, $\widetilde{S_i}^{C_1}$ and $\widetilde{S_i}^{C_2}$.

We suppose that $\widetilde{S_i}^{C_0} \leq \widetilde{S_i}^{C_1}$, i.e that merging C_0 and C, forming partition $(C_1, \overline{C_1})$ leads to a better partition than $(C_0, \overline{C_0})$. The principle of the proof is to show that there is an improvement in merging C' with C_1, i.e. that $\widetilde{S_i}^{C_1} < \widetilde{S_i}^{C_2}$. Hence we will conclude

that, supposing now that two elementary clusters have proportional histograms, then they are necessarily of the same side of the optimal partition.

We denote: $N_0 = \frac{N}{n_X}$, and $u = h^0 - H_0, v = h - H$.

From $C_0 \cup C \cup C' \subsetneq Im(f)$ we deduce $\sum_j (h_j^0 + (1+\theta)h_j) < N$ and $H_0 + (1+\theta)H < N_0$.

Then, with these notations, we express the sensitivity criteria based on histograms:

$$n_X \widetilde{S}_i^{C_0} = \frac{1}{H_0(N_0 - H_0)} ||u||^2$$
$$n_X \widetilde{S}_i^{C_1} = \frac{1}{(H_0 + H)(N_0 - (H_0 + H))} ||u + v||^2$$
$$n_X \widetilde{S}_i^{C_2} = \frac{1}{(H_0 + (1+\theta)H)(N_0 - (H_0 + (1+\theta)H))} ||u + (1+\theta)v||^2$$

From $\widetilde{S}_i^{C_0} \leq \widetilde{S}_i^{C_1}$ we deduce:

$$||u + v||^2 \geq \frac{(H_0 + H)(N_0 - (H_0 + H))}{H_0(N_0 - H_0)} ||u||^2 \tag{3}$$

As $H_0 + (1+\theta)H < N_0$, we consider $\frac{\widetilde{S}_i^{C_2}}{\widetilde{S}_i^{C_1}}$:

$$\frac{\widetilde{S}_i^{C_2}}{\widetilde{S}_i^{C_1}} = \frac{||u + (1+\theta)v||^2}{||u + v||^2} \frac{(H_0 + H)(N_0 - (H_0 + H))}{(H_0 + (1+\theta)H)(N_0 - (H_0 + (1+\theta)H))}$$
$$= \frac{||u + v||^2 + (\theta^2 + 2\theta)||v||^2 + 2\theta u.v}{||u + v||^2} \frac{(H_0 + H)(N_0 - (H_0 + H))}{(H_0 + (1+\theta)H)(N_0 - (H_0 + (1+\theta)H))}$$

From $(\theta^2 + 2\theta)||v||^2 + 2\theta u.v = \theta(||u+v||^2 - ||u||^2 + (1+\theta)||v||^2)$ and Equation (3) we deduce:

$$(\theta^2 + 2\theta)||v||^2 + 2\theta u.v \geq \theta(1 - \frac{H_0(N_0 - H_0)}{(H_0 + H)(N_0 - (H_0 + H))})||u + v||^2$$
$$||u + v||^2 + (\theta^2 + 2\theta)||v||^2 + 2\theta u.v \geq (1 + \theta - \frac{\theta H_0(N_0 - H_0)}{(H_0 + H)(N_0 - (H_0 + H))})||u + v||^2$$

Hence:

$$\frac{\widetilde{S}_i^{C_2}}{\widetilde{S}_i^{C_1}} \geq \frac{(1+\theta)(H_0 + H)(N_0 - (H_0 + H)) - \theta H_0(N_0 - H_0)}{(H_0 + (1+\theta)H)(N_0 - (H_0 + (1+\theta)H))}$$
$$\frac{\widetilde{S}_i^{C_2}}{\widetilde{S}_i^{C_1}} \geq \frac{(H_0 + (1+\theta)H)(N_0 - (H_0 + H)) - \theta H_0 H}{(H_0 + (1+\theta)H)(N_0 - (H_0 + H)) - \theta H(H_0 + (1+\theta)H)}$$
$$\frac{\widetilde{S}_i^{C_2}}{\widetilde{S}_i^{C_1}} \geq 1 + \frac{\theta(1+\theta)H^2}{(H_0 + (1+\theta)H)(N_0 - (H_0 + (1+\theta)H))} > 1$$

so $\widetilde{S}_i^{C_1} < \widetilde{S}_i^{C_2}$.