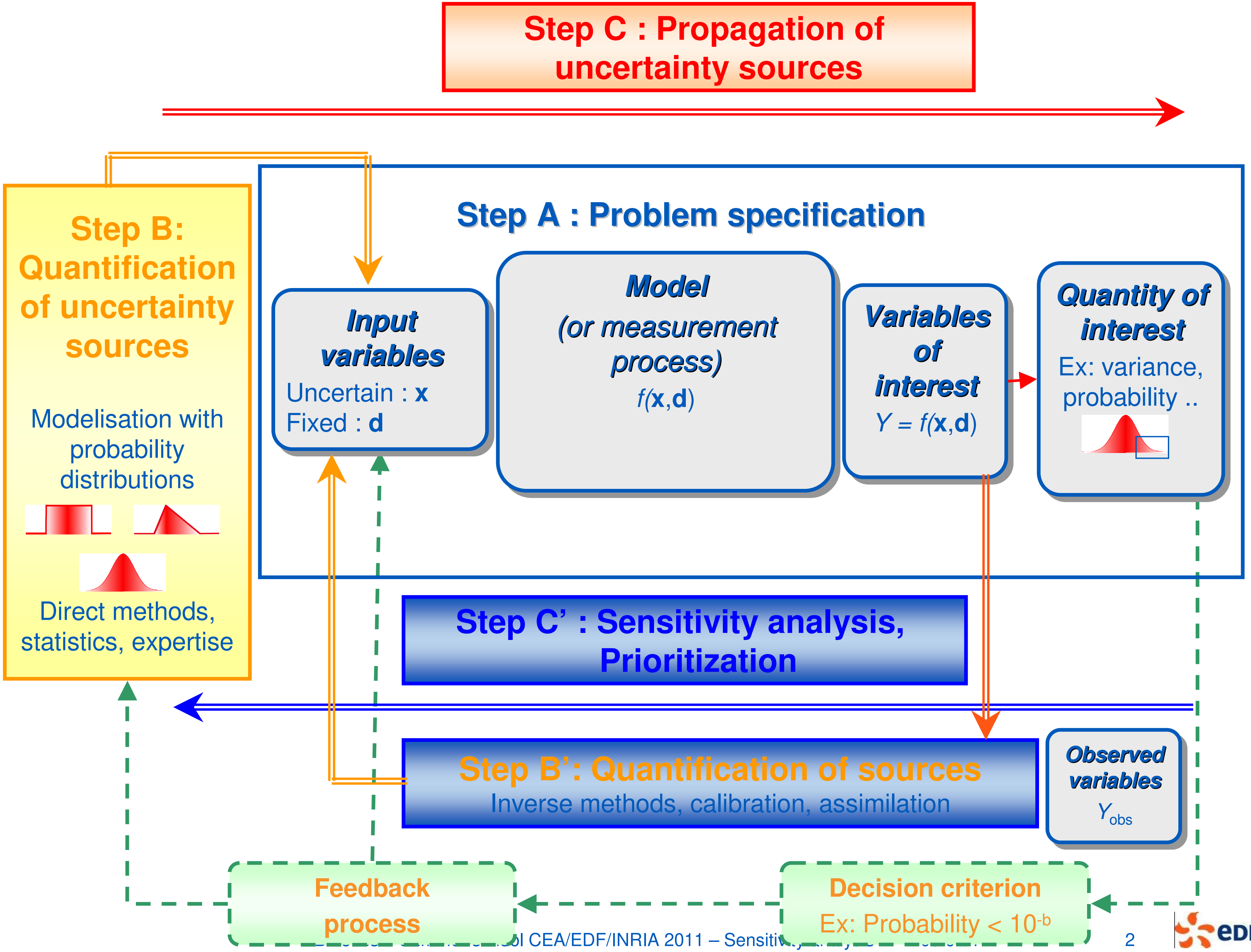
Basics of global sensitivity analysis

Summer school CEA-EDF-INRIA 2011

July, 5th, 2011

CHANGER L'ÉNERGIE ENSEMBLE


Step C : Propagation of uncertainty sources
Step B: Quantification of uncertainty sources
Modelisation with probability distributions
Direct methods, statistics, expertise
Step A : Problem specification
Input variables
Uncertain : x
Fixed : d
Model
(or measurement process)
f(x,d)
Variables of interest
Y = f(x,d)
Quantity of interest
Ex: variance, probability ..
Step C' : Sensitivity analysis, Prioritization
Step B': Quantification of sources
Inverse methods, calibration, assimilation
Observed variables
Y obs
Feedback process
Decision criterion
Ex: Probability < 10-b

Main objectives of sensitivity analysis

- **Reduction of the uncertainty of the model outputs by prioritization of the sources**
  - Variables to be fixed in order to obtain the **largest reduction** (or a fixed reduction) **of the output uncertainty**

    A purely mathematical variable ordering
  - Most influent variables in a given output domain

  → if reducibles, then R&D prioritization

  → else, modification of the system

- **Simplification of a model**
  - **determination of the non-influent variables,** that can be fixed without consequences on the output uncertainty
  - building a simplified model, a metamodel

Goal of this talk

In the jungle of sensitivity analysis methods:

Goal of this talk

In the jungle of sensitivity analysis methods:

I wish you to go from this
uncomfortable position

Goal of this talk

… to this comfortable
one =>

Plan

Restrictions in this talk:

- **scalar variable of interest (output)**
- **quantity of interest: variability of the output**

Screening with n < p (supersaturated designs)

Many inputs (p >> 10) and cpu time costly computer code

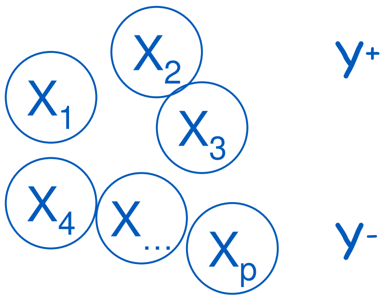
<u>Objective:</u> less computations than number of inputs

<u>Hypotheses:</u>

- Number of influent inputs << total number of inputs
- Monotony of the model, no interaction between inputs
- Knowledge of the direction of the output variation / each input

Example: method of sequential bifurcations

2 calculs



Screening with n < p (supersaturated designs)

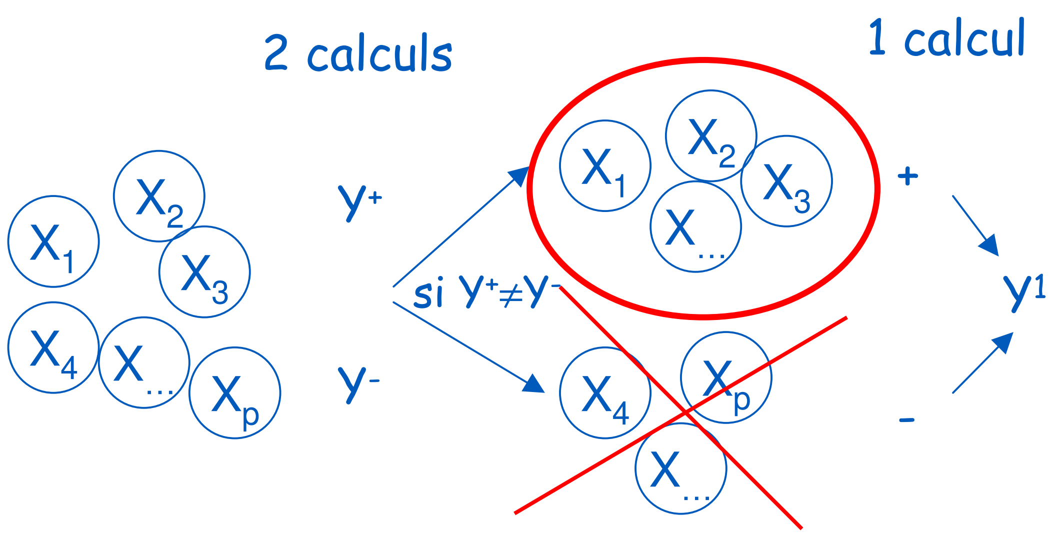
Many inputs (p >> 10) and cpu time costly computer code

Objective: less computations than number of inputs

Hypotheses:

- Number of influent inputs << total number of inputs
- Monotony of the model, no interaction between inputs
- Knowledge of the direction of the output variation / each input

Example: method of sequential bifurcations



Screening with n < p (supersaturated designs)

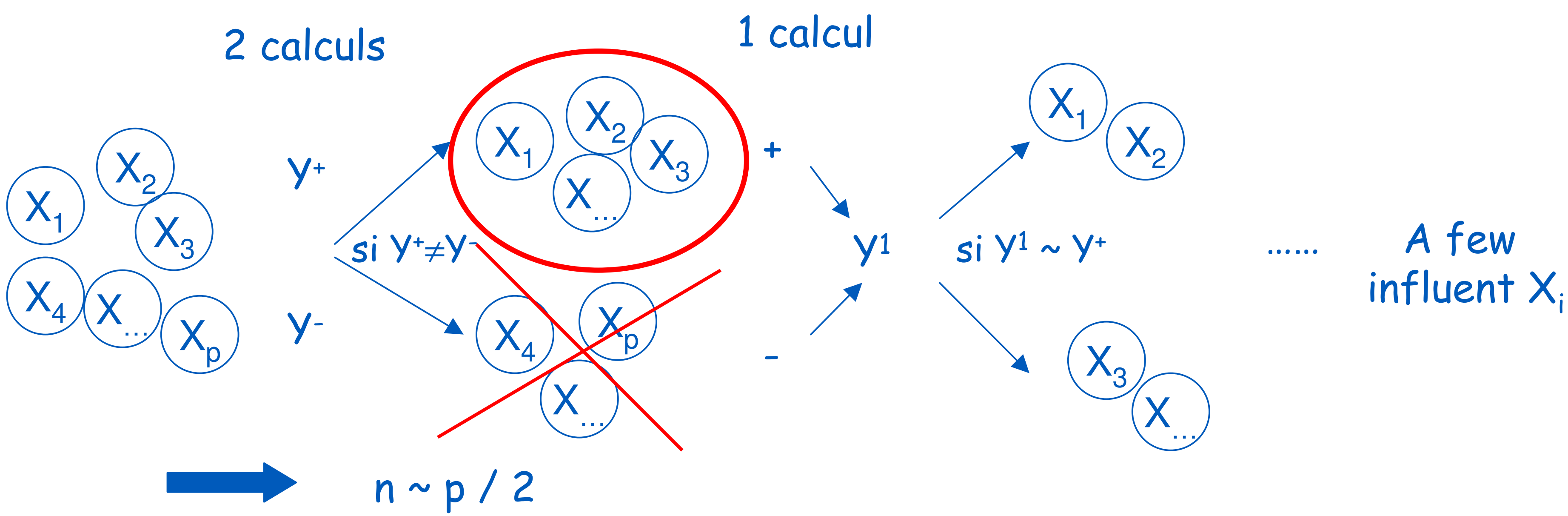
Many inputs (p >> 10) and cpu time costly computer code

Objective: less computations than number of inputs

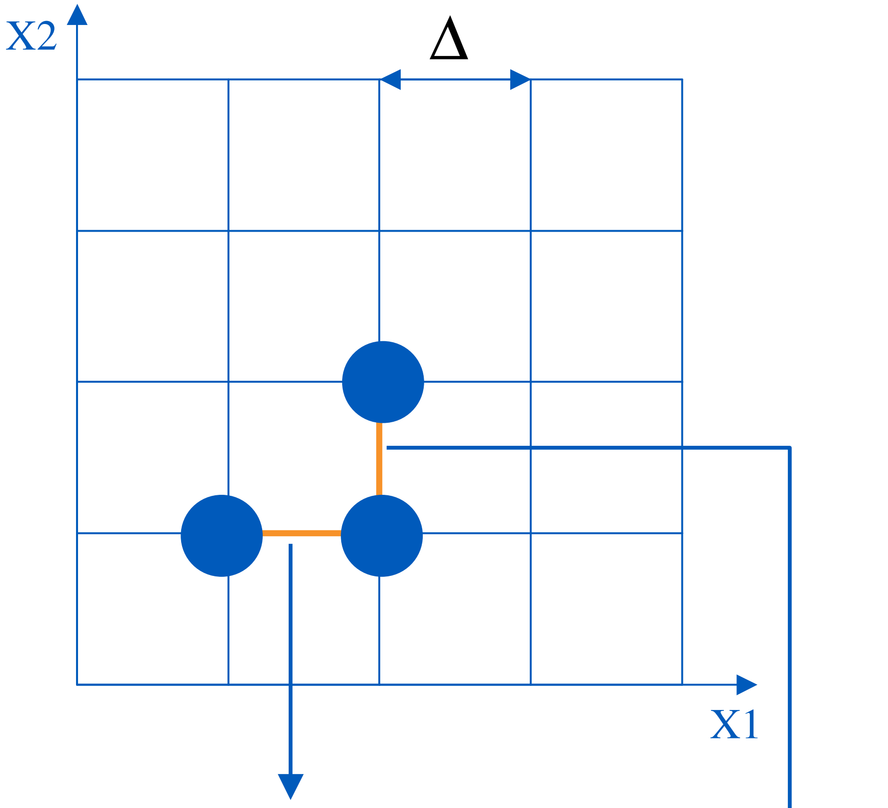
Hypotheses:

- Number of influent inputs << total number of inputs
- Monotony of the model, no interaction between inputs
- Knowledge of the direction of the output variation / each input

Example: method of sequential bifurcations



Screening without hypothesis on function: Morris' method

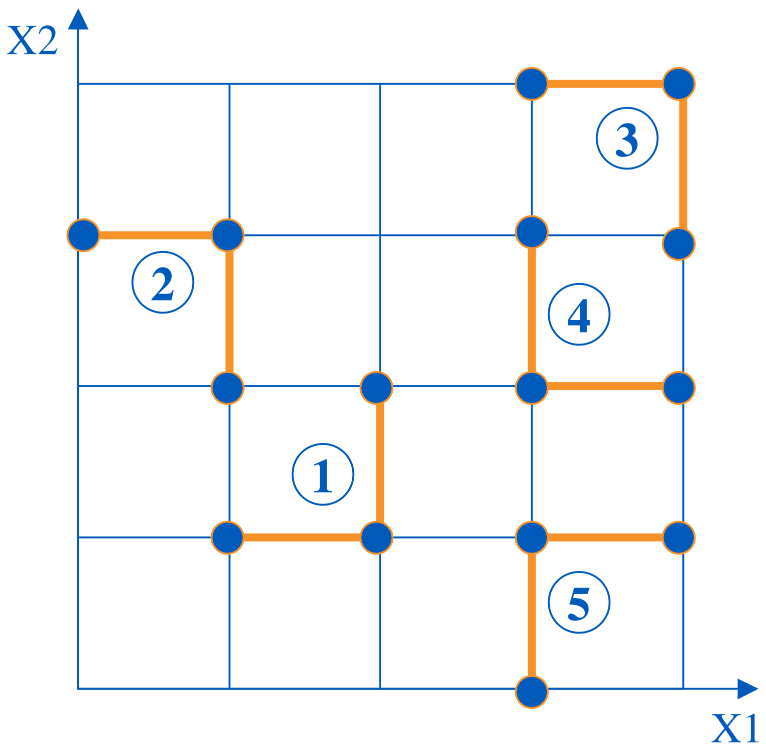


- Discretization of input space
- Needs p+1 experiments
- OAT (One-at-A-Time)
- computation of one elementary effect for each input

$$d_{X_1} = \frac{f(P3) - f(P2)}{\Delta}$$

$$d_{X_2} = \frac{f(P2) - f(P1)}{\Delta}$$

Morris' method



- OAT design is repeated R times (total: n = R*(p+1) experiments)

- It gives an R-sample for each elementary effect

$$\left\{d^i_{X1}\right\}_{i=1\ldots R}$$
$$\left\{d^i_{X2}\right\}_{i=1\ldots R}$$

- Sensitivity measures:

$$\mu_i^* = \mathrm{E}(|d_{X_i}|)$$
$$\sigma_i = \sigma(d_{X_i})$$

Morris: sensitivity measures

- $\mu_i^* = \mathrm{E}(|d_{X_i}|)$ is a measure of the **sensitivity**:

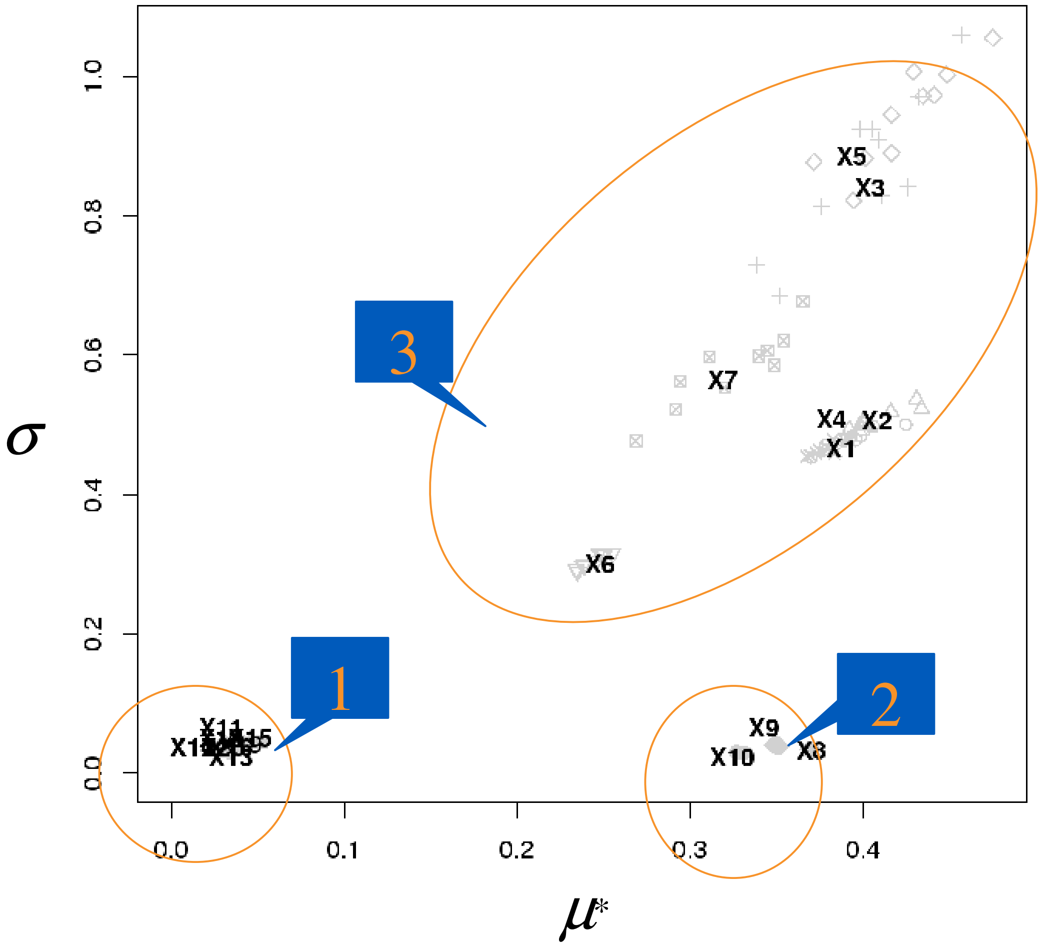
  Important value → important effects (in mean)
  → sensitive model to input variations

- $\sigma_i = \sigma(d_{X_i})$ is a measure of the **interactions** and of the **non linear effects**:

  important value → different effects in the R-sample
  → effects which depend on the value:
  - of the input X_i => non linear effect
  - or of the other inputs => interaction

  (the distinction between the two cases is impossible)

Morris : example



20 factors
210 simulations
→ Graph (mu*, sigma)

Distinction between 3 groups:

1. Negligible effects
2. Linear effects
3. Non linear effects and/or with interactions

Cas test : fontion non monotone de Morris (source Saltelli)

Plan

1. Screening techniques

2. Graphical tools

3. Quantitative methods

4. Conclusions

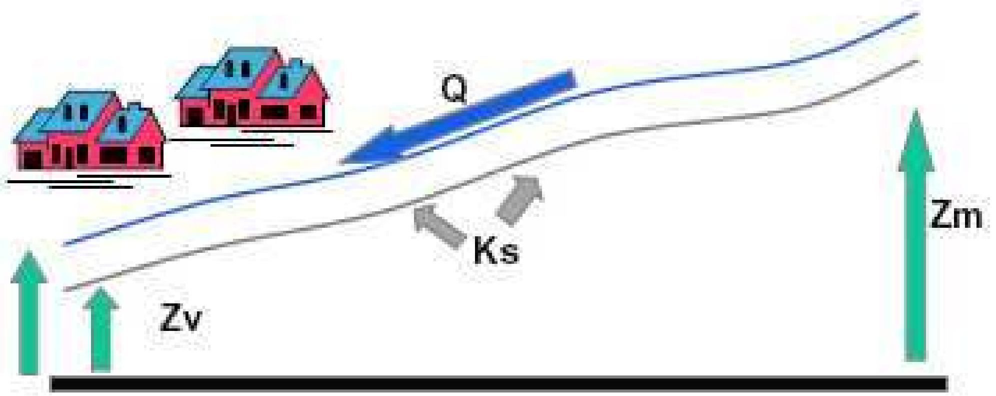
Sensitivity analysis for one scalar output

Sample ($\mathbf{X}$, $Y(\mathbf{X})$) of size n > p , preferably n >> p

Preliminary step: graphical vizualisation (for ex: scatterplots)

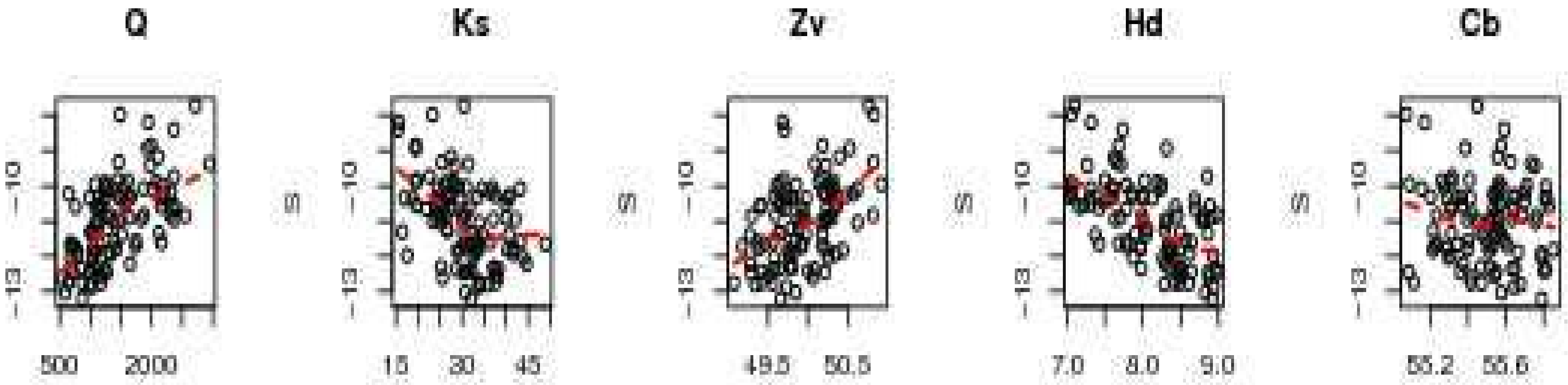
Remark: it can be a Monte Carlo sample, a quasi-Monte Carlo sample or any other designs

Flood model - Scatterplots – Output S



$$S = Z_v + H - H_d - C_b \quad \text{avec} \quad H = \left(\frac{Q}{BK_s \sqrt{\frac{Z_m - Z_v}{L}}} \right)^{0.6}$$

Monte Carlo sample – N = 100

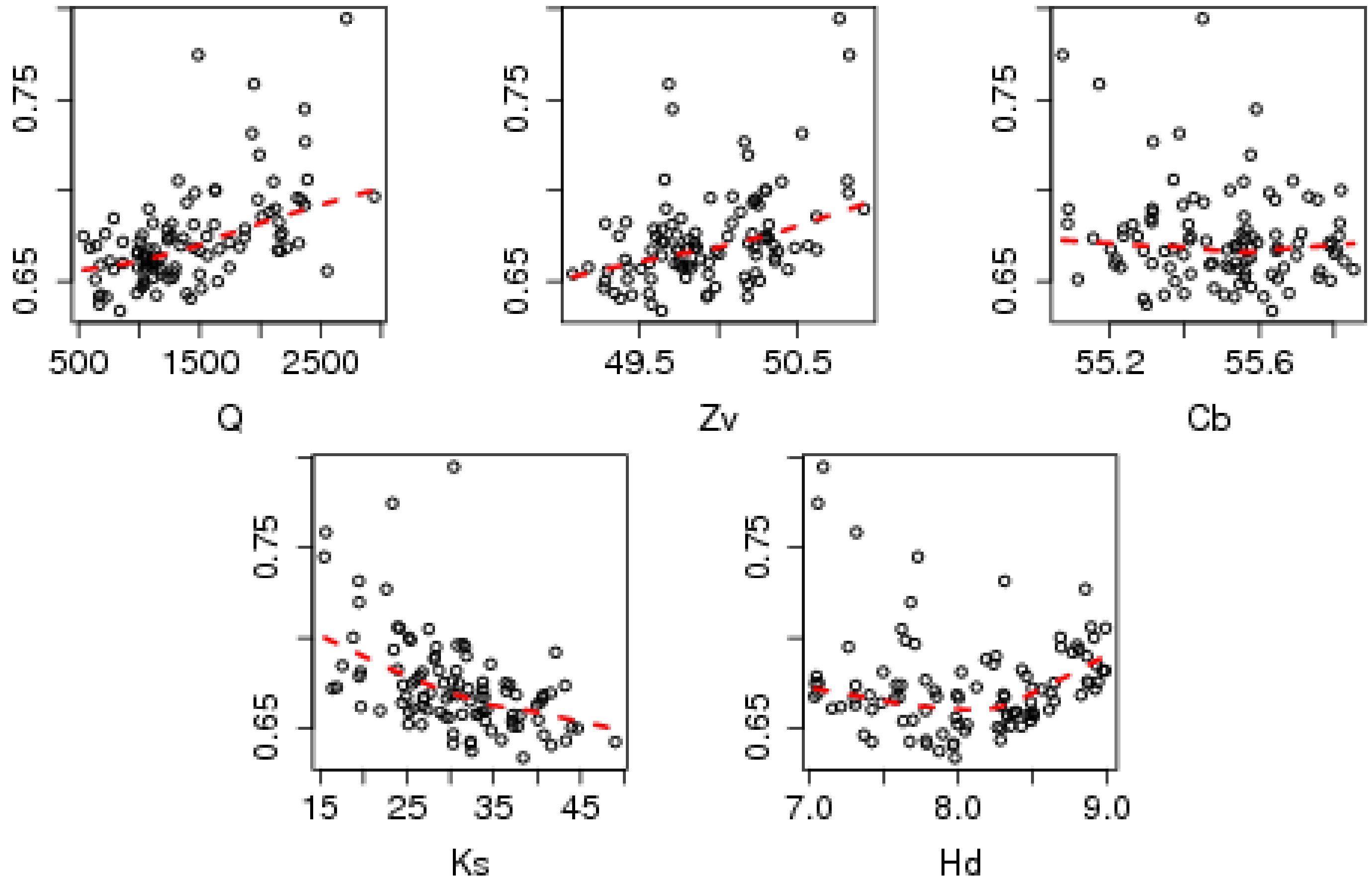


EDF

Flood model - Scatterplots – Output Cp

$$C_p = 1\!\!1_{S>0} + \left\{0.2 + 0.8\left[1 - \exp\left(-\frac{1000}{S^4}\right)\right]\right\} 1\!\!1_{S\leq 0} + \frac{1}{20}\left(H_d 1\!\!1_{H_d>8} + 8 1\!\!1_{8\leq H_d}\right) ,$$

Monte Carlo sample – N = 100



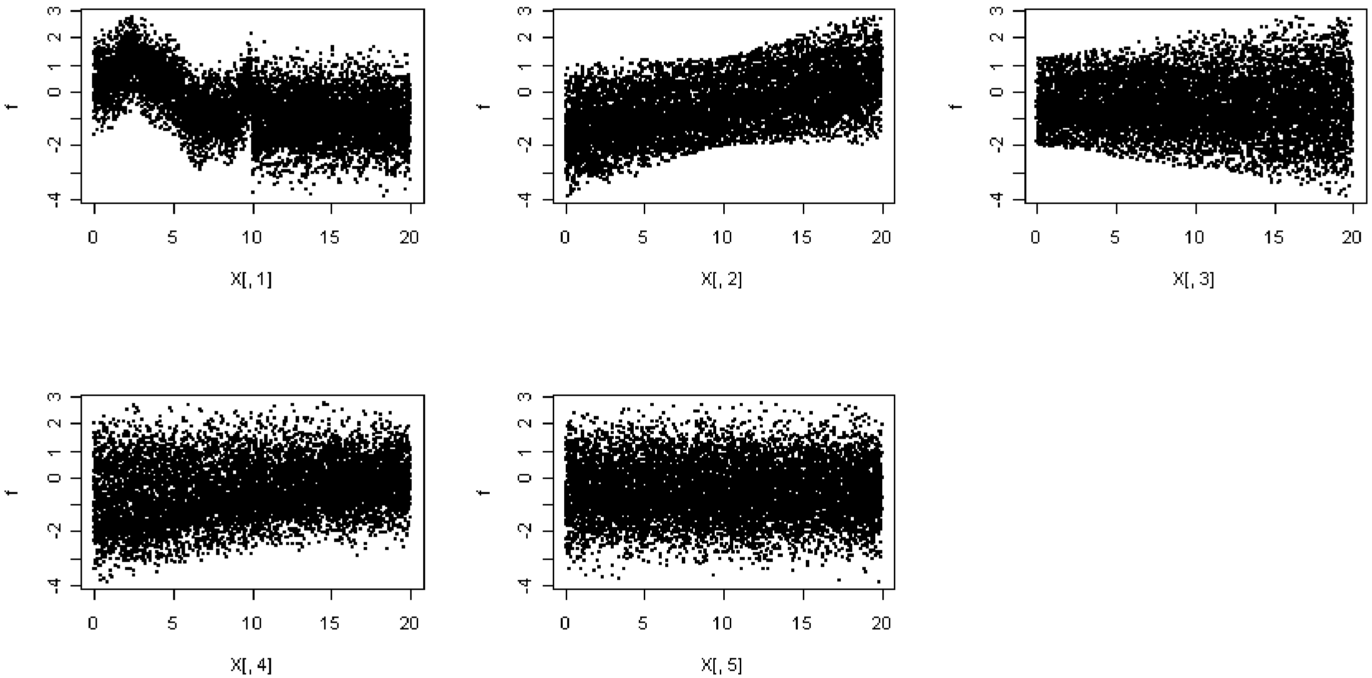
Major drawback : only first order relations between inputs are analyzed and not their interactions

Another example

Analytical function tgp5d

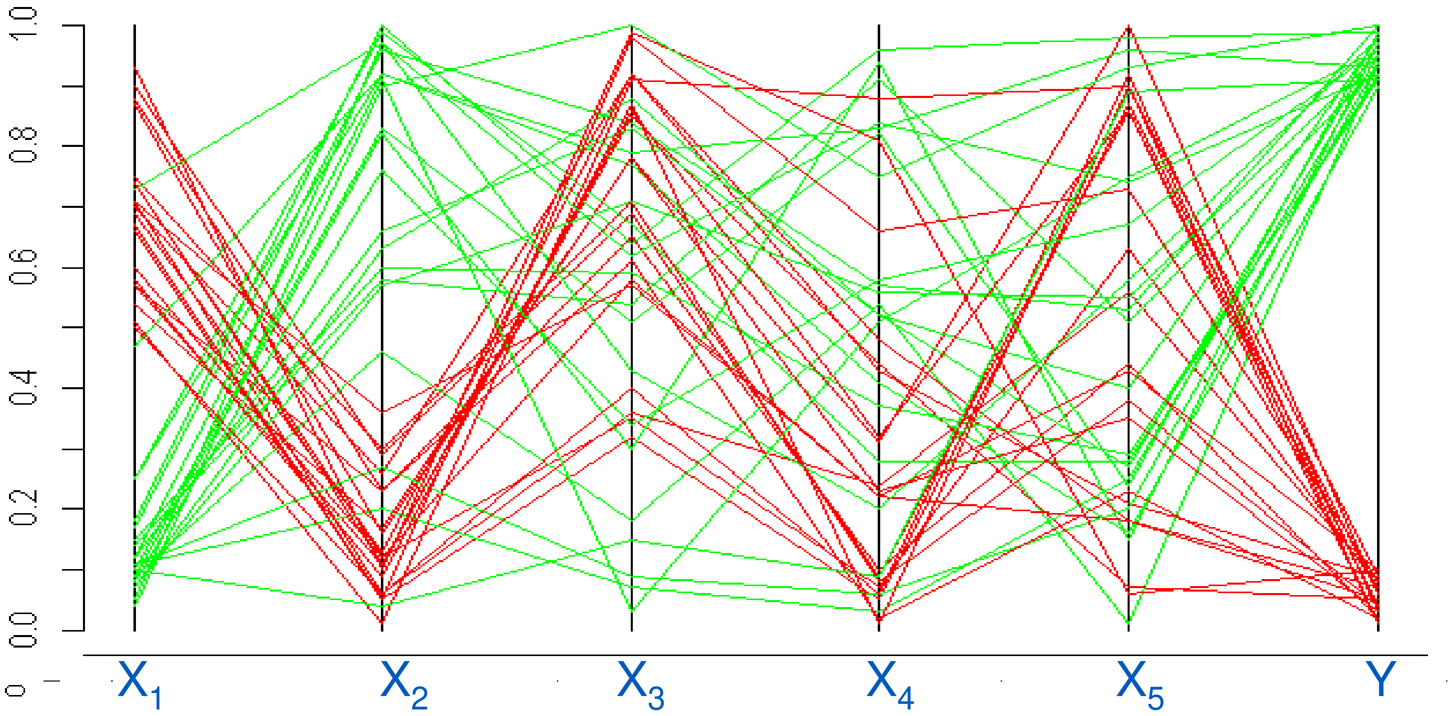
$$f(\mathbf{X}) = \begin{cases} \sin\frac{\pi X_1}{5} + \frac{1}{5}\cos\frac{4\pi X_1}{5} + 0.01(X_2 - 10)X_3 \text{ si } X_1 < 10 \\ 0.01(X_2 - 10)X_3 + 0.1(X_4 - 20) \text{ sinon} \end{cases}$$

$$X_i \sim U[0;20] \text{ pour } i = 1,\ldots,5$$



Another graphical tool : the Cobweb plot

[Source : P. Lemaître]



Detection of interactions

Plan

1. Screening techniques

2. Graphical tools

3. **Quantitative methods**

4. Conclusions

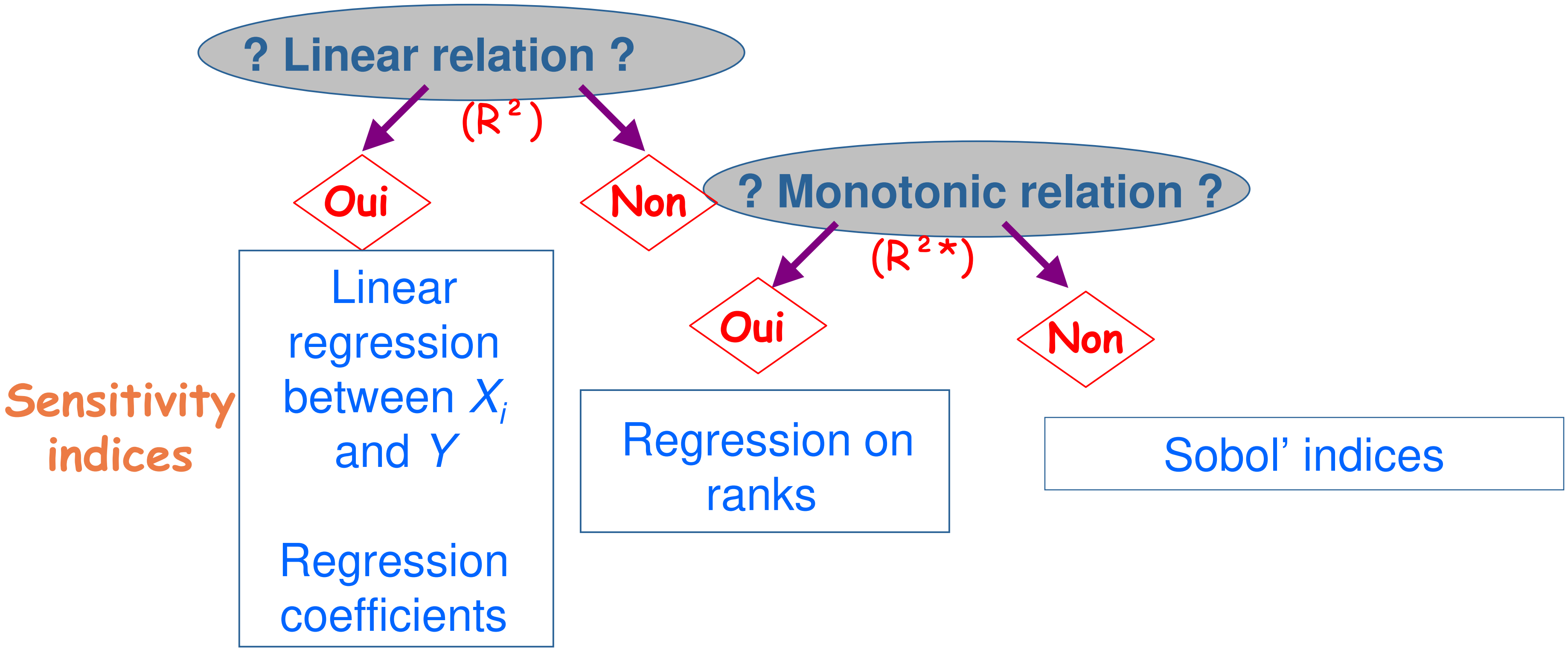
Sensitivity analysis for one scalar output

Sample (**X**, *Y* (**X**)) of size n > *p* , preferably n >> *p*

Preliminary step: graphical vizualisation (for ex: scatterplots)

Quantitative sensitivity analysis methodology

[Saltelli et al. 00, Helton et al. 06]



Indices de sensibilité dans le cas d'une relation entrées/sortie linéaire

Independent inputs $\mathbf{X}=(X_1,\dots,X_p)$

Sample: n realizations of $(\mathbf{X},Y)$

$$Y=\beta_0+\sum_{i=1}^{p}\beta_i X_i$$

- Standard Regression Coefficient (SRC) : $SRC(X_i):=\beta_i\sqrt{\frac{\mathrm{Var}(X_i)}{\mathrm{Var}(Y)}}$

Sign of β_i gives the direction of variation of Y / X_i

- Similar to the linear correlation coefficient (Pearson)

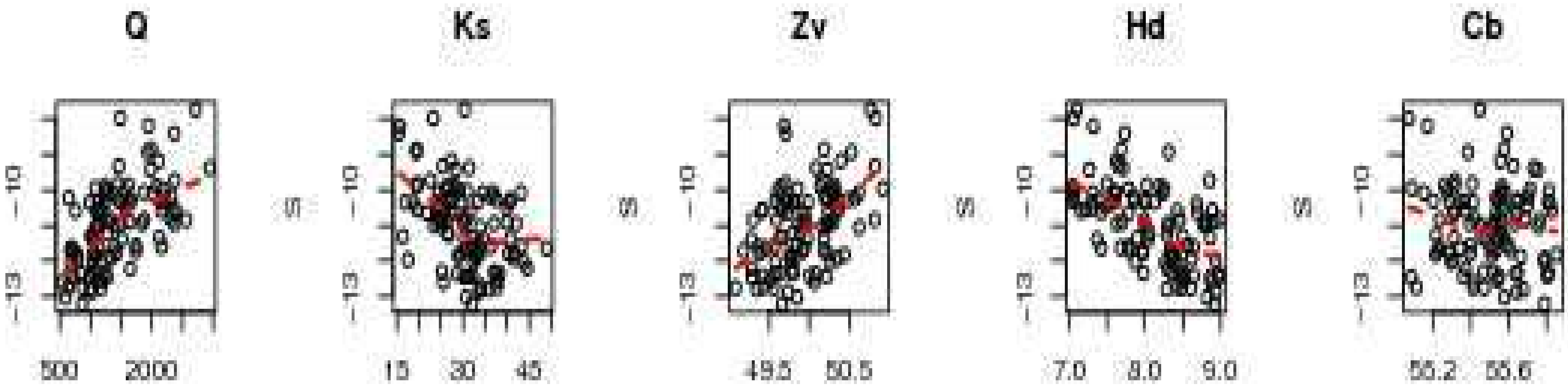
- Validity of the linear model du modèle via regression diagnostics and R^2 :

$$R^2=1-\frac{\sum_{i=1}^{n}\left(\hat{Y}_i-Y_i\right)^2}{\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2}$$

- Theoretically, we have $R^2=\sum_{i=1}^{p}\mathrm{SRC}^2(X_i)$, useful to interpret SRC

Flood model - Output S

Monte Carlo sample – N = 100



31% 15% 18% 29% 6%

Sensitivity indices (SRC^2)

The model is linear (R^2=0.99)

SRC coefficients are sufficient

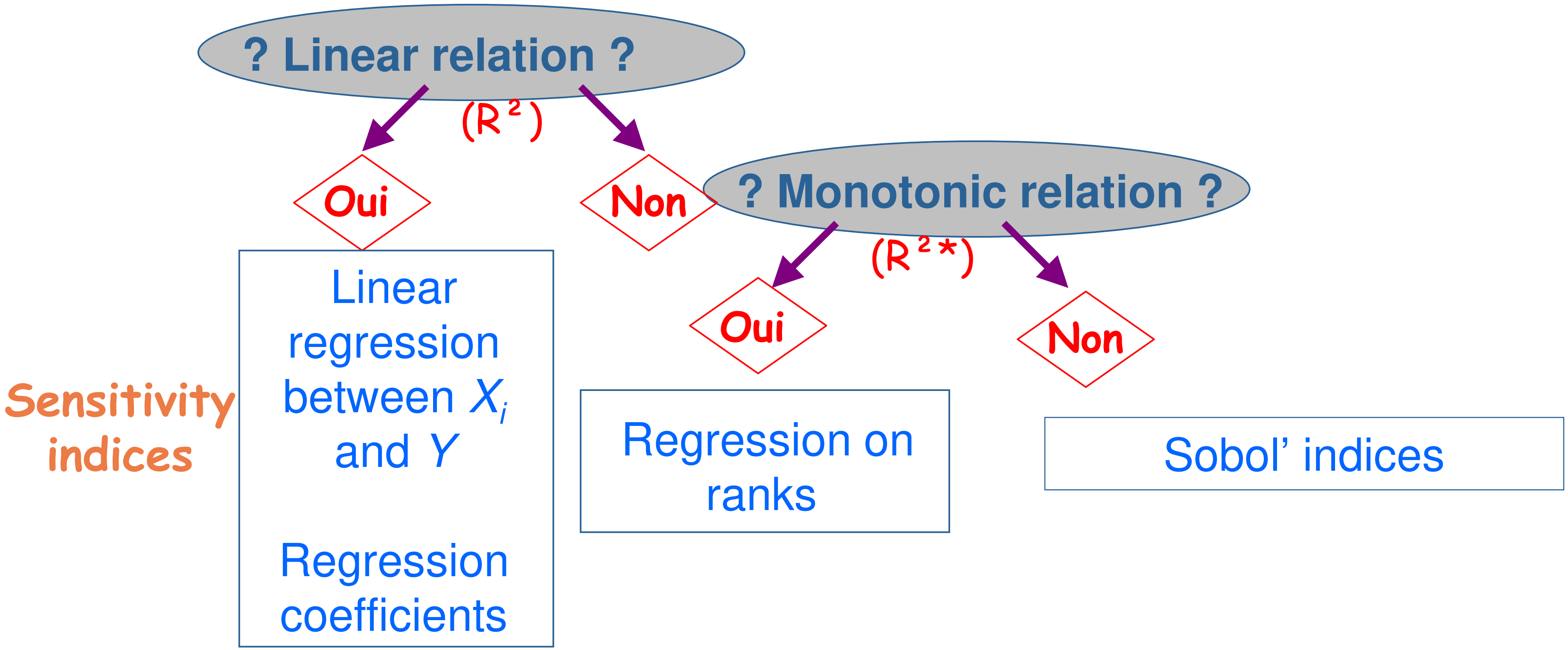
Sensitivity analysis for one scalar output

Sample (**X**, *Y* (**X**)) of size n > *p* , preferably n >> *p*

Preliminary step : graphical vizualisation (for ex: scatterplots)

Quantitative sensitivity analysis methodology

[Saltelli et al. 00, Helton et al. 06]



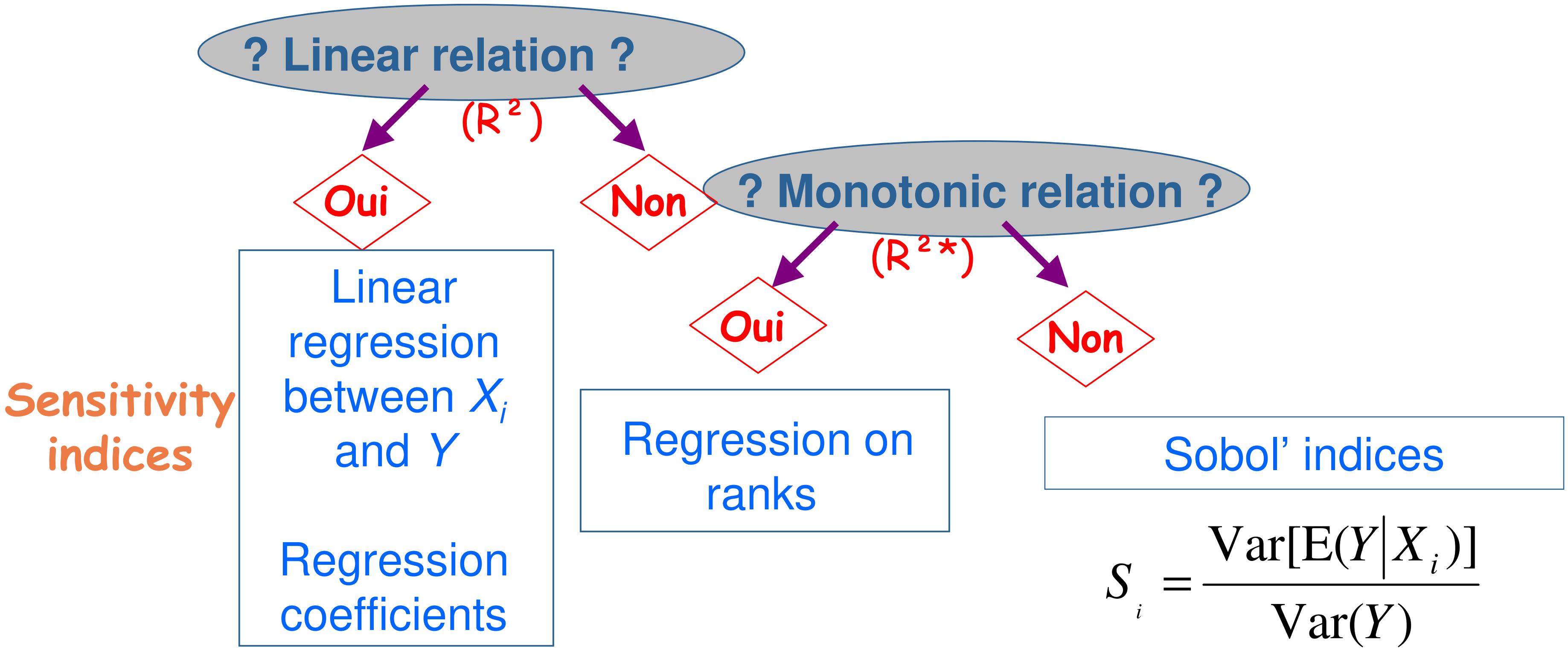
Sensitivity analysis for one scalar output

Sample (**X**, Y(**X**)) of size n > p , preferably n >> p

Preliminary step : graphical vizualisation (for ex: scatterplots)

Quantitative sensitivity analysis methodology

[Saltelli et al. 00, Helton et al. 06]



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

Functional decomposition

$$y = f(\mathbf{x}) = f_0 + \sum_{i=1}^{p} f_i(x_i) + \sum_i \sum_{j>i} f_{ij}(x_i, x_j) + \ldots + f_{1,2,\ldots,p}(x_1, x_2, \ldots, x_p)$$

with $f(\mathbf{x}) \in L^2(\mathbf{x}) \quad \mathbf{x} \in [0;1]^p$

Infinity of possible decompositions

BUT, unicity of decomposition if: $\int f_{i_1 \ldots i_s}(x_{i_1}, \ldots, x_{i_s}) dx_j = 0 \ \forall j = i_1, \ldots, i_s$

Properties ($x_i \sim U[0,1]$ for $i=1,\ldots,p$, the x_i s are independent)

$$f_0 = \int f(\mathbf{x})\, d\mathbf{x} = \mathrm{E}(y)$$

$$f_i(x_i) = \int f(\mathbf{x}) dx_{-i} - f_0 = \mathrm{E}(y \mid x_i) - f_0$$

$$f_{ij}(x_i, x_j) = \mathrm{E}(y \mid x_i, x_j) - \mathrm{E}(y \mid x_i) - \mathrm{E}(y \mid x_j) + f_0$$

Example : $f(x_1, x_2) = x_1 + x_2 \ ; \ x_1 \sim U[0;1] \ ; \ x_2 \sim U[0;1]$

$$f_0 = 1 \ ; \ f_1(x_1) = x_1 - \frac{1}{2}; \ f_2(x_2) = x_2 - \frac{1}{2} \ ; \ f_{12}(x_1, x_2) = 0$$

Indices de sensibilité sans hypothèse sur le modèle

Functional ANOVA *[Efron & Stein 81]* (hyp. of independent X_i s) :

$$\mathrm{Var}(Y) = \sum_{i=1}^{p} V_i(Y) + \sum_{i<j}^{p} V_{ij}(Y) + \cdots + V_{12\ldots p}(Y)$$

where $V_i(Y) = \mathrm{Var}[\mathrm{E}(Y|X_i)]$

$$V_{ij} = \mathrm{Var}\left[\mathrm{E}(Y|X_i X_j)\right] - V_i - V_j, \ldots$$

Sobol indices definition:

- First order sensitivity indices: $S_i = \dfrac{V_i}{\mathrm{Var}(Y)}$
- Second order sensitivity indices: $S_{ij} = \dfrac{V_{ij}}{\mathrm{Var}(Y)}$
- ...

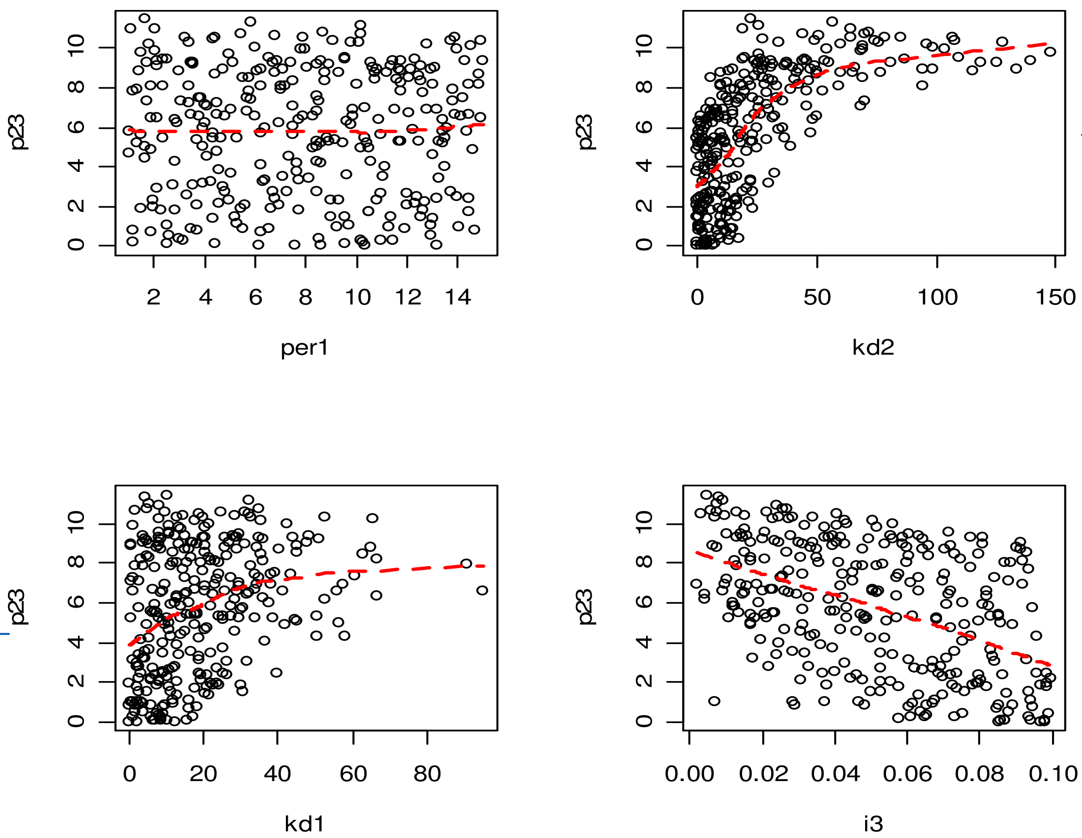
Interprétation graphique

First order Sobol' indices measure the variability of consitional expectation des espérances (mean trend curves in the scatterplots)

Null index

High index

Small index



Propriétés des indices de Sobol

$$1 = \sum_{i=1}^{p} S_i + \sum_i \sum_j S_{ij} + \sum_i \sum_j \sum_k S_{ijk} \ldots + S_{1,2,\ldots,k}$$

$\sum_i S_i \leq 1$ Always

$\sum_i S_i = 1$ Additive model

$1 - \sum_i S_i$ Measure the degree of interactions between variables

Examples : p =4 gives 4 indices S_i, 6 indices S_{ij} , 4 indices S_{ijk}, 1 indice S_{ijkl}

General case : 2^p-1 indices to be estimated

Total sensitivity index: $S_{Ti} = S_i + \sum_j S_{ij} + \sum_{j,k} S_{ijk} + \ldots = 1 - S_{\sim i}$

[Homma & Saltelli 1996]

Direct estimation

- **Monte Carlo (Sobol)**

Taking 2 i.i.d. samples: $\left(X_i^{(j)}\right)_{i=1,..,p;\,j=1,..,n}$ et $\left(X'^{(j)}_i\right)_{i=1,..,p;\,j=1,..,n}$

- Variance (classical estimator): $\hat{V}(Y) = \frac{1}{n}\sum_{k=1}^{n} f\left(\mathbf{X}^{(k)}\right)^2 - \hat{f}_0^2$ avec $\hat{f}_0 = \frac{1}{n}\sum_{k=1}^{n} f\left(\mathbf{X}^{(k)}\right)$

- Conditional variances:

$$V_i(Y) = \mathrm{Var}[\mathrm{E}(Y|X_i)] = \int \mathrm{E}^2(Y|X_i)dX_i - \left(\int \mathrm{E}(Y|X_i)dX_i\right)^2$$

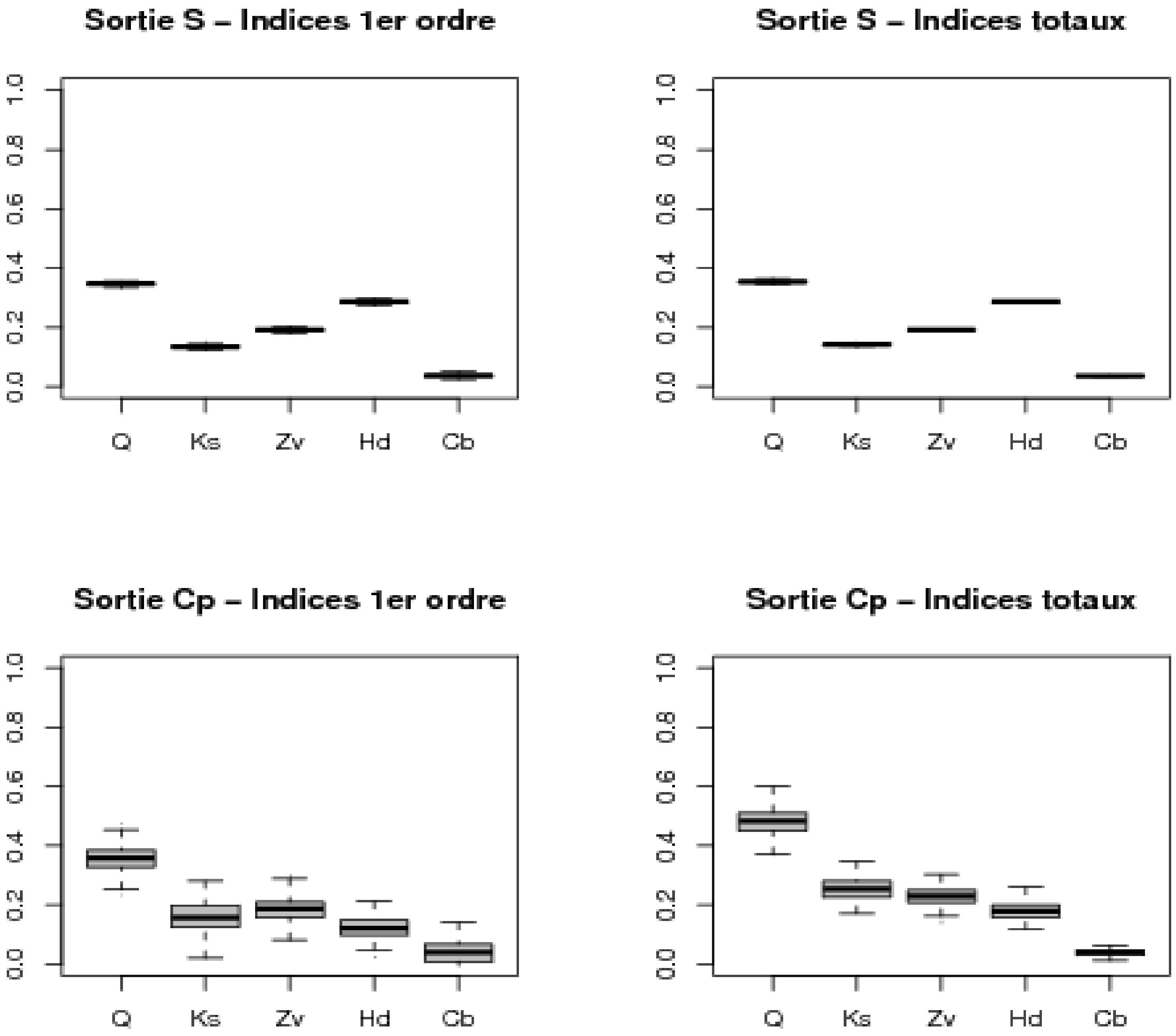
$$\hat{V}_i(Y) = \frac{1}{n}\sum_{k=1}^{n} f\left(X_1^{(k)},\ldots,X_{i-1}^{(k)},X_i^{(k)},X_{i+1}^{(k)},\ldots,X_p^{(k)}\right) f\left(X'^{(k)}_1,\ldots,X'^{(k)}_{i-1},X_i^{(k)},X'^{(k)}_{i+1},\ldots,X'^{(k)}_p\right) - f_0^2$$

$$V_{\sim i}(Y) = \mathrm{Var}[\mathrm{E}(Y|X_{\sim i})]$$

$$\hat{V}_{\sim i}(Y) = \frac{1}{n}\sum_{k=1}^{n} f\left(X_1^{(k)},\ldots,X_{i-1}^{(k)},X_i^{(k)},X_{i+1}^{(k)},\ldots,X_p^{(k)}\right) f\left(X_1^{(k)},\ldots,X_{i-1}^{(k)},X'^{(k)}_i,X_{i+1}^{(k)},\ldots,X_p^{(k)}\right) - f_0^2$$

- Quasi Monte Carlo
- FAST
- Random Balance Design
- …

Flood model


Sortie S – Indices 1er ordre
Sortie S – Indices totaux
Sortie Cp – Indices 1er ordre
Sortie Cp – Indices totaux
0.0
0.2
0.4
0.6
0.8
1.0
Q
Ks
Zv
Hd
Cb

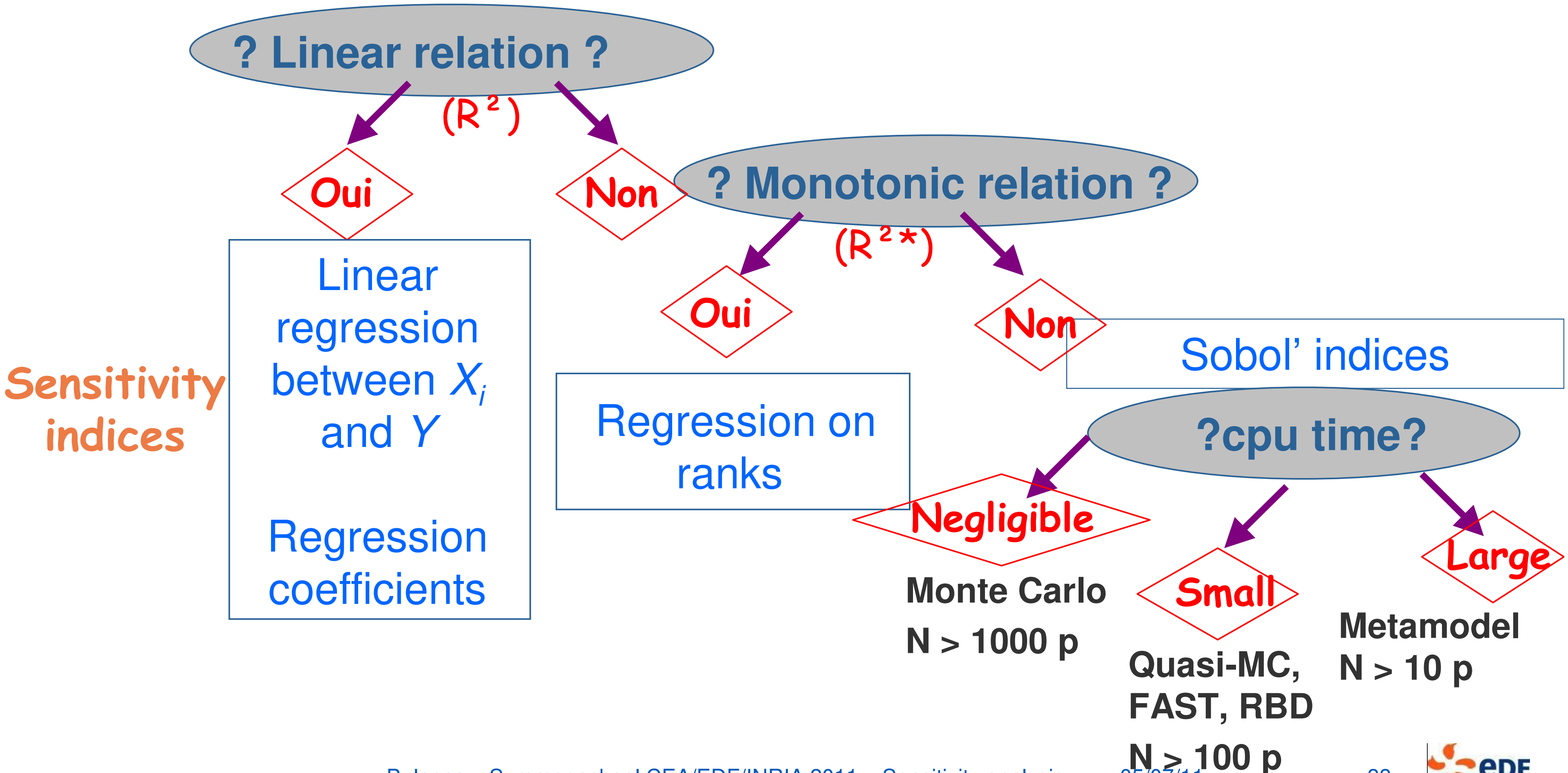
Sensitivity analysis for one scalar output

Sample (**X**, *Y* (**X**)) of size n > *p* , preferably n >> *p*

Preliminary step : graphical vizualisation (for ex: scatterplots)

Quantitative sensitivity analysis methodology

[Saltelli et al. 00, Helton et al. 06]



Flood model – Output Cp

From the 100-size Monte Carlo sample, a **Gaussian process metamodel** is fitted

Predictivity of the Gp metamodel : Q_2 = 99%

Indices (en %)	Q	K_s	Z_v	H_d	C_b
S_i modèle	35.5	15.9	18.3	12.5	3.8
S_i métamodèle	38.9	16.8	18.8	13.9	3.7
S_{T_i} modèle	48.2	25.3	22.9	18.1	3.8
S_{T_i} métamodèle	45.5	21.0	21.3	16.8	4.3

N=1e5

100 replicates

N x (p+2) x 100 = 7e7 evaluations

Plan

1. Screening techniques

2. Graphical tools

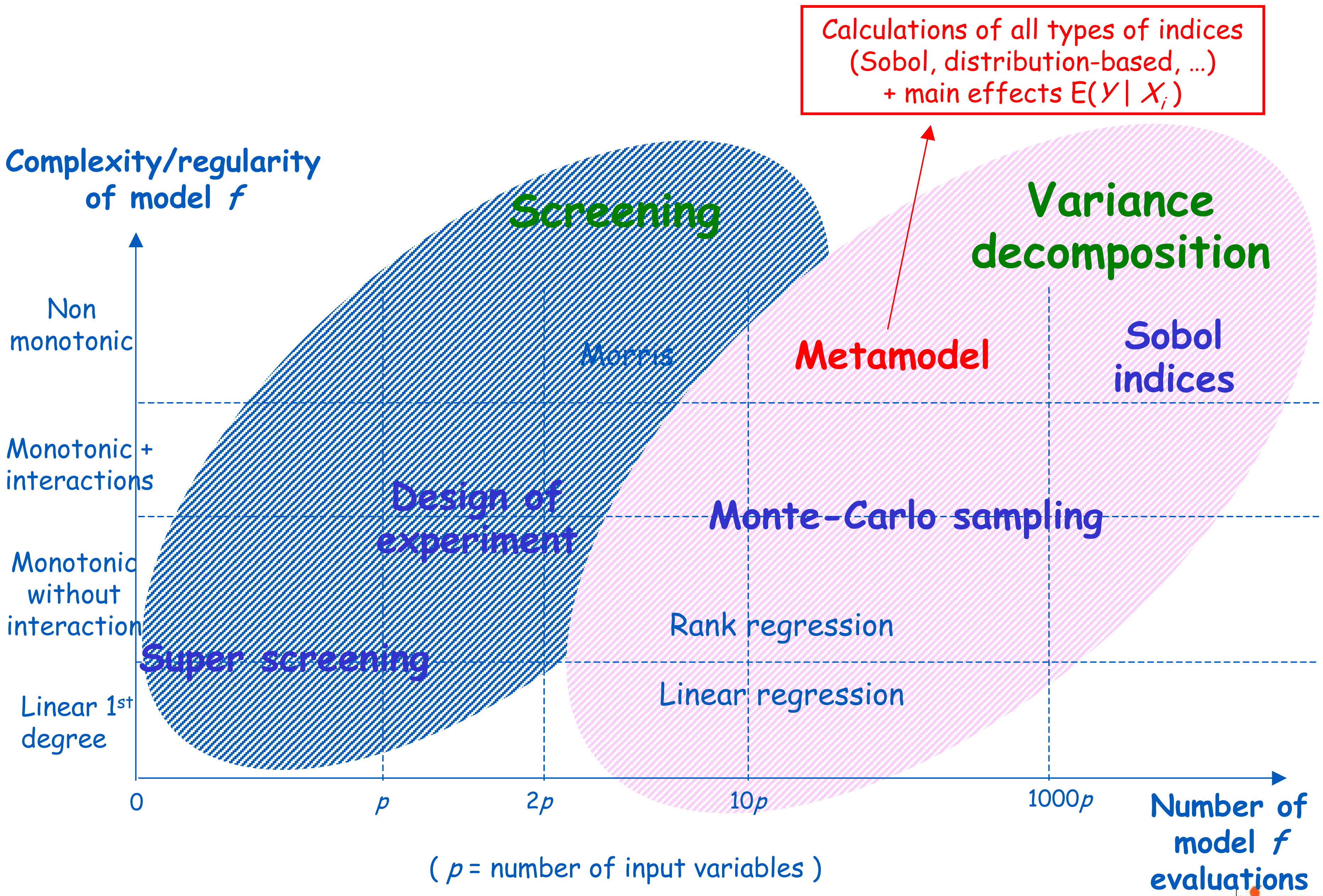
3. Quantitative methods

4. **Conclusions**

Choosing the right method?

- Requested information (qualitative/quantitative)
- Number of inputs
- Regularity of the model (linearity/monotony/continuity)
- Cpu cost of one model evaluation
- Number of outputs
- Additional constraints, for example :
  - Uncertainty/sensitivity joint analysis,
  - Dependency between inputs, …

Classification of sensitivity analysis methods



More complex situations

- Efficient ways to compute Sobol' indices
  => some recent papers (Sudret, Crestaux, Marrel), also discuss with Jean-Marc

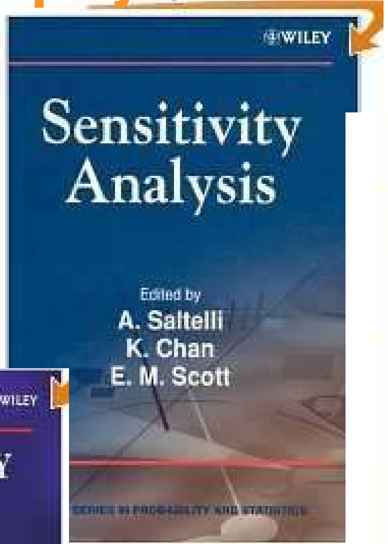
- Non scalar outputs : curve or spatial responses of the model
  => some recent papers (Marrel, Lamboni, Auder), also discuss with Amandine

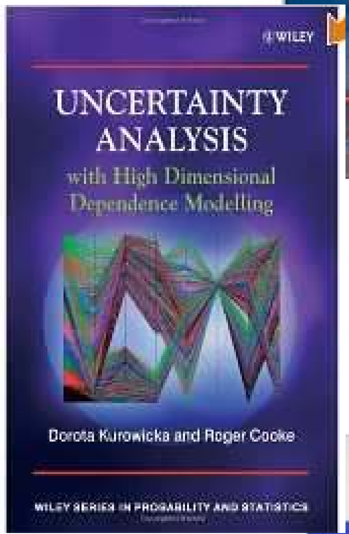
- Quantity of interest: probability of failure instead of variance
  => FORM/SORM method, also discuss with Paul and his « ghost » poster

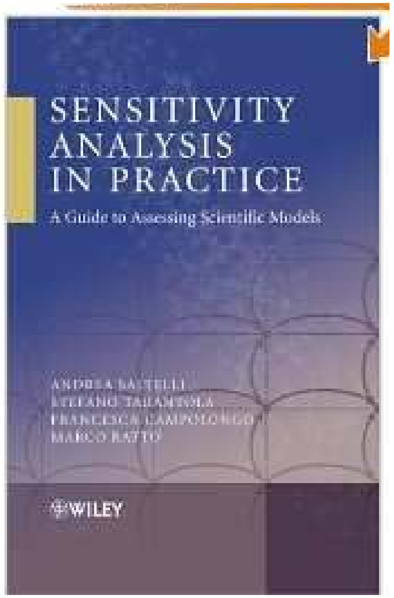
- Non independent inputs
  => some recent papers (Rabitz, Xu & Gertner), also discuss with Gaëlle

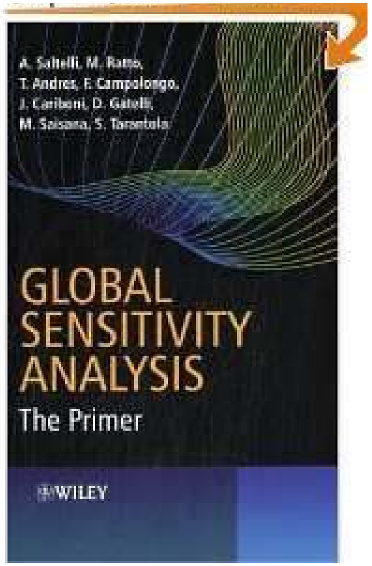
- High dimensional inputs
  => coupling adjoint and stochastic methods, …

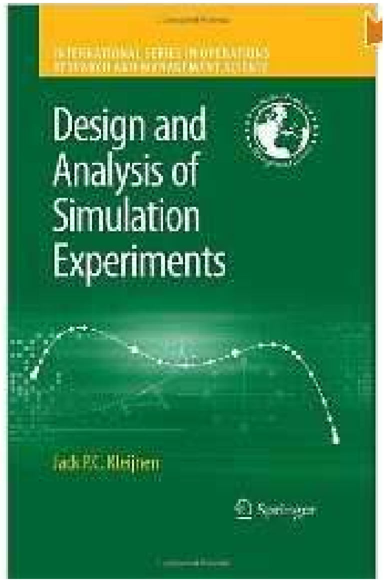
Bibliography

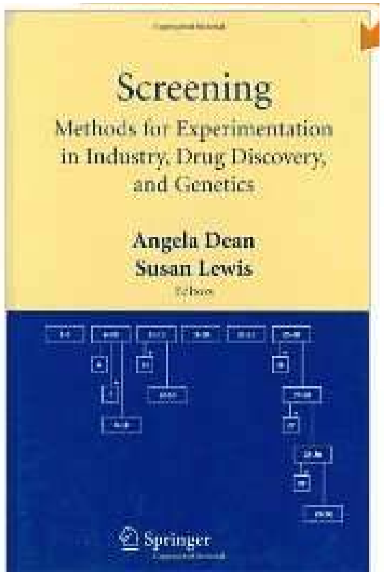


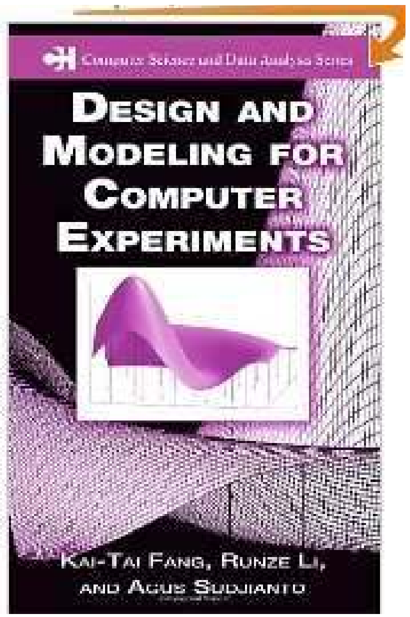


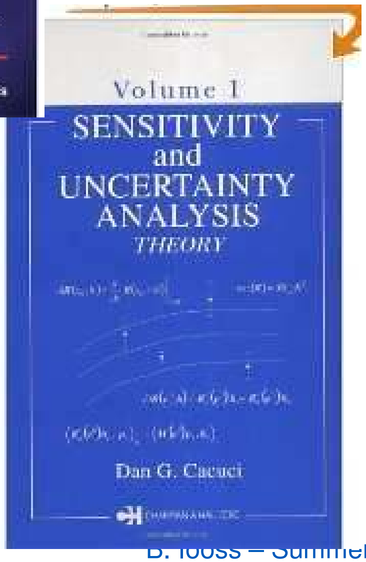


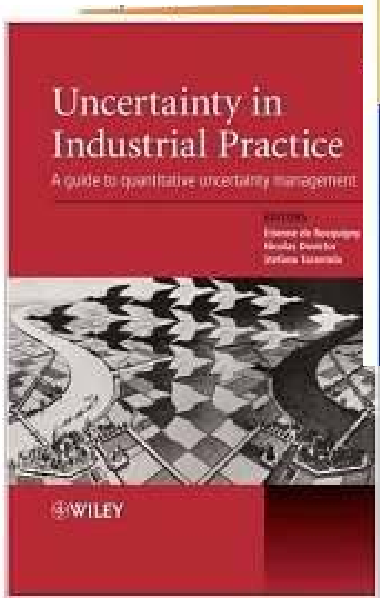












Free softwares/packages on the web

- SIMLAB : under Windows, graphical interface

- Open Turns : uncertainty software of EDF/EADS/Phimeca => see tomorrow

- Matlab packages : SAinterface, DACE, FERUM, …

- R packages :
  - Sensitivity, multisensi
  - SMURFER
  - CompModSA (Storlie's web page)
  - DiceKriging
  - planor (soon)

Thank you