

Analyses statistiques des expériences numériques

Cours 4 : Analyse de sensibilité – Décomposition de la variance

Bertrand Iooss

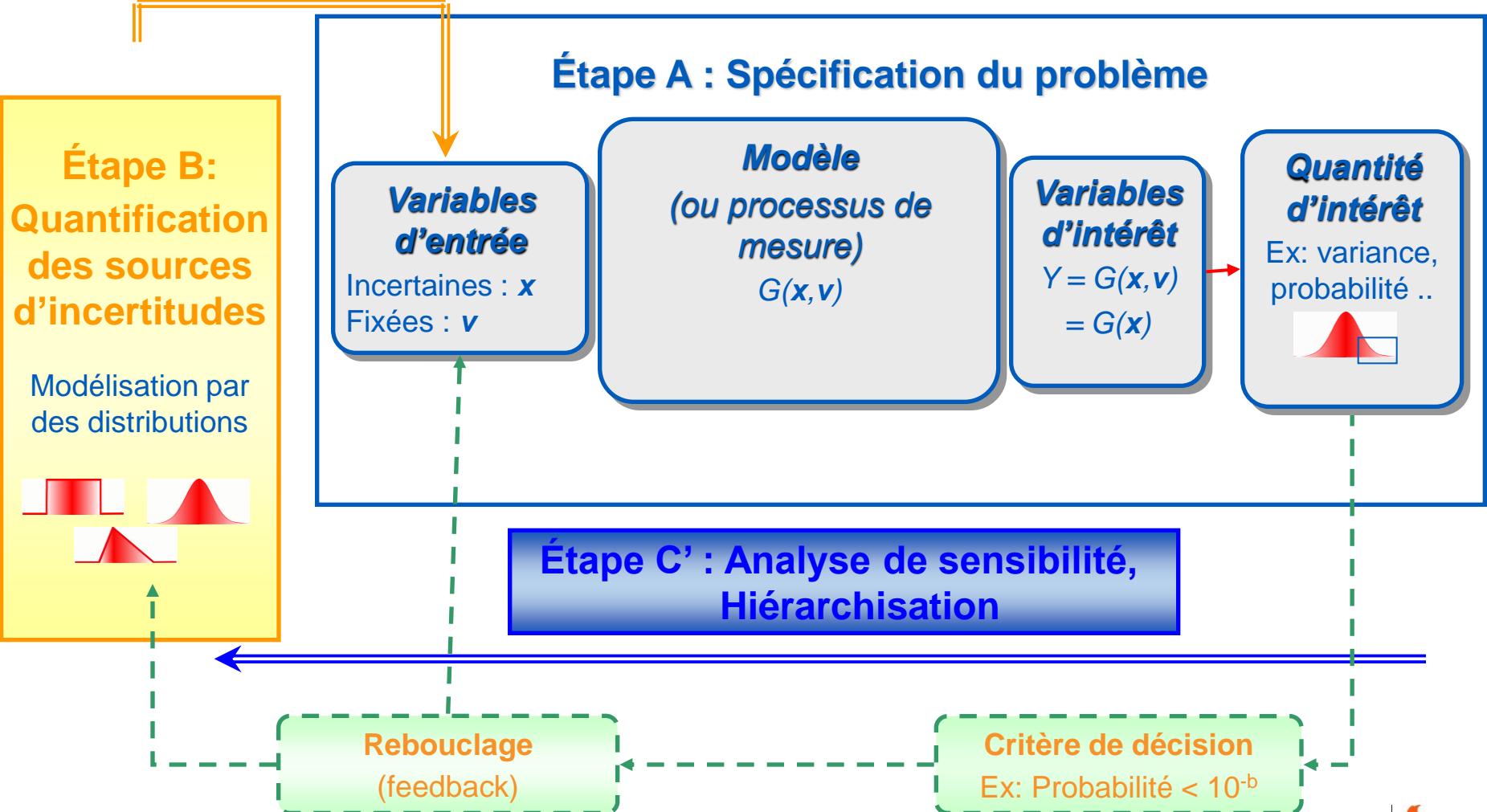
Polytech Nice Sophia

Janvier 2026



Schéma générique

Étape C : Propagation des sources d'incertitude



DEUX NOTIONS

► **Sensibilité**, par exemple $\partial Y / \partial X_i$

Donne une idée de la manière dont peut répondre la réponse en fonction de **variations potentielles** des facteurs

=> Analyse locale

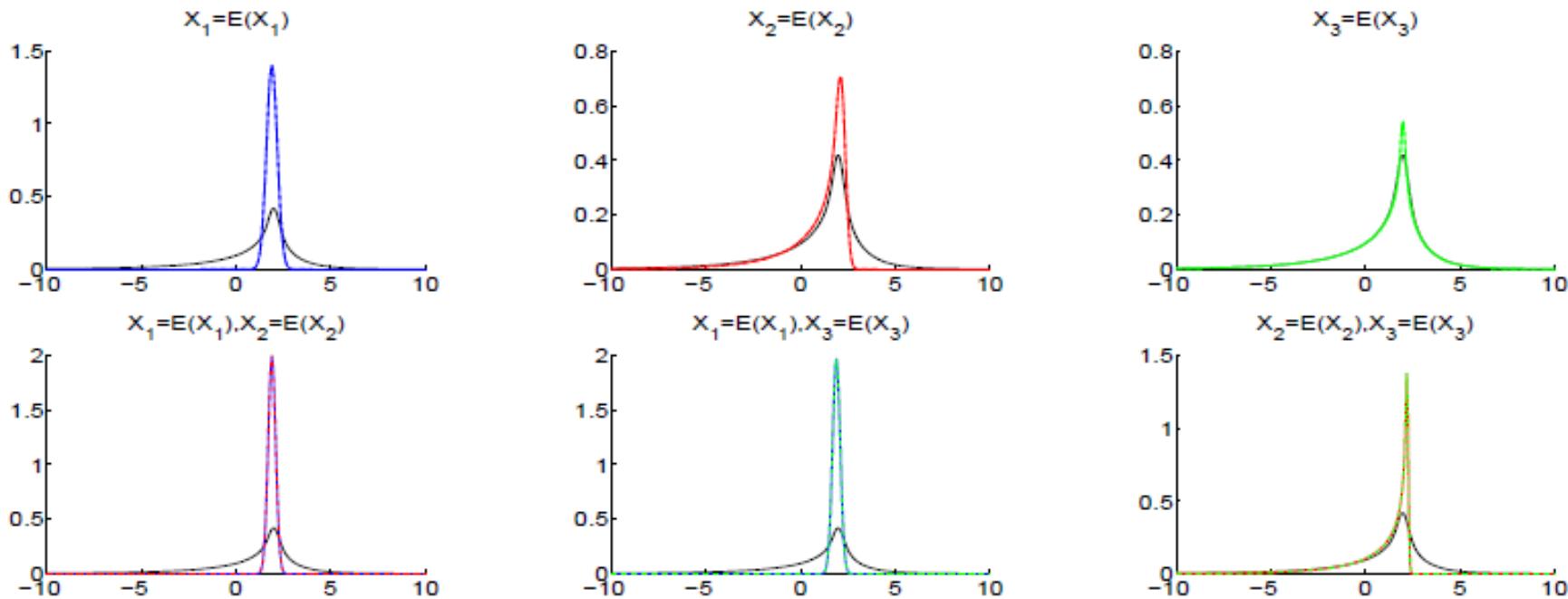
► « **Contribution** » = **Sensibilité x “Importance”**, par exemple

$$\frac{\partial Y}{\partial X_i} \sigma(X_i)$$

=> Analyse globale

Exercice introductif

- 1) Indiquer la/les variable(s) les plus influente(s) sur la variabilité de $Y = G(X_1, X_2, X_3)$
- 2) Indiquer la variable à contrôler en priorité pour garantir $Y > 0$



En noir : densité de probabilité de Y

En couleurs : densités de probabilité conditionnelles (de Y conditionnée aux différentes variables d'entrée)

Mesures d'influence globale

Échantillon $(X, Y(X))$ de taille $N > p$, de préférence de taille $N \gg p$

Étape préliminaire : visualisation graphique (par ex : scatterplots)

→ Étape primordiale pour le choix de la méthode :
gain de temps dans la suite, et dans l'analyse critique des résultats

Mesures d'influence globale

=> Représentation graphique : scatterplots

Mesure le caractère linéaire du nuage de points

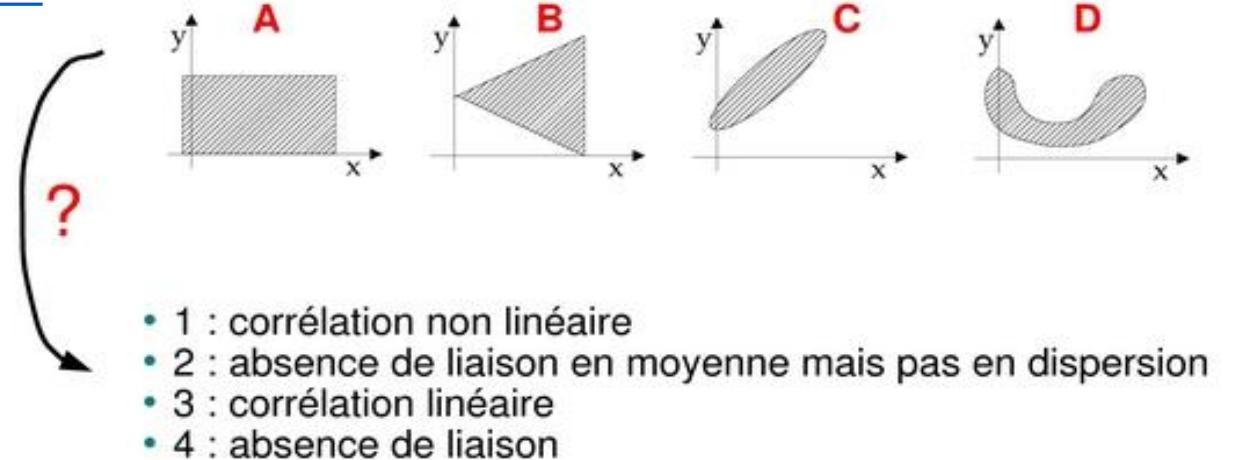
N calculs

Graphe Sortie / chaque entrée

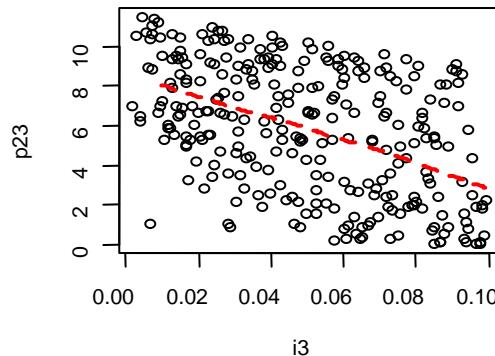
$$\rho = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

$$\hat{\rho} = \frac{1/N \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{S_x S_y}$$

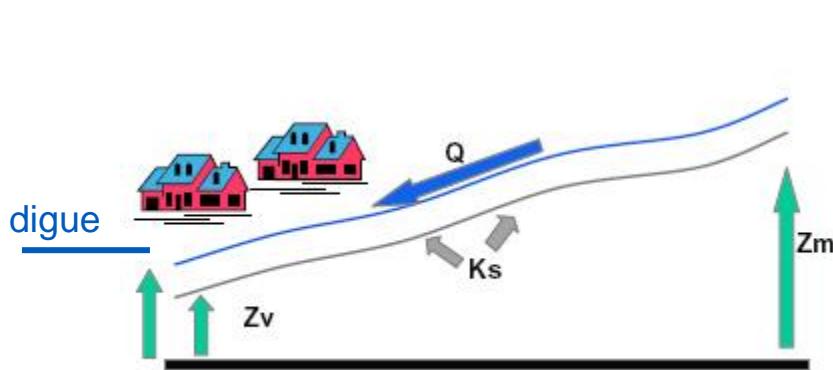
• Nuage de points : exemples



Exemple : échantillon de taille $N = 300$



APPLICATION « JOUET » : MODÈLE DE CRUES



Modèle physique simplifié (hydraulique)

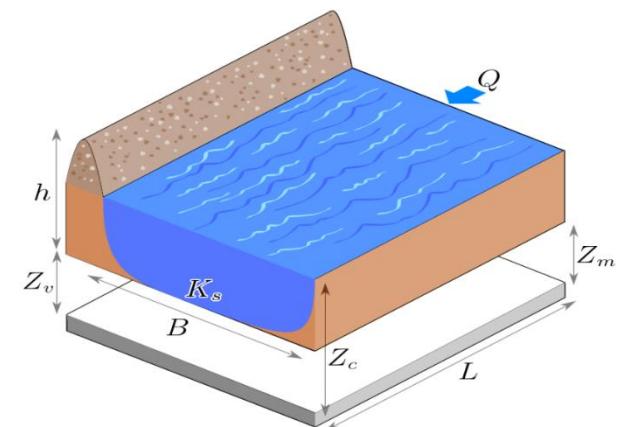
S = surverse

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

$$S_0 = H - (C_b - Z_v) - 8$$

C_p = coût annuel de maintenance de la digue

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



Entrées aléatoires

Q = débit $\sim G(1013,558 | [500,3000])$

K_s = coef. Frottement $\sim \mathcal{N}(30,8)$

Z_v = hauteur du lit aval $\sim \mathcal{T}[49,51]$

Z_m = hauteur lit amont $\sim \mathcal{T}[54,56]$

H_d = hauteur digue $\sim \mathcal{U}[7,9]$

C_b = côte de la berge $\sim \mathcal{T}[55,56]$

L = longueur tronçon $\sim \mathcal{T}[4990,5010]$

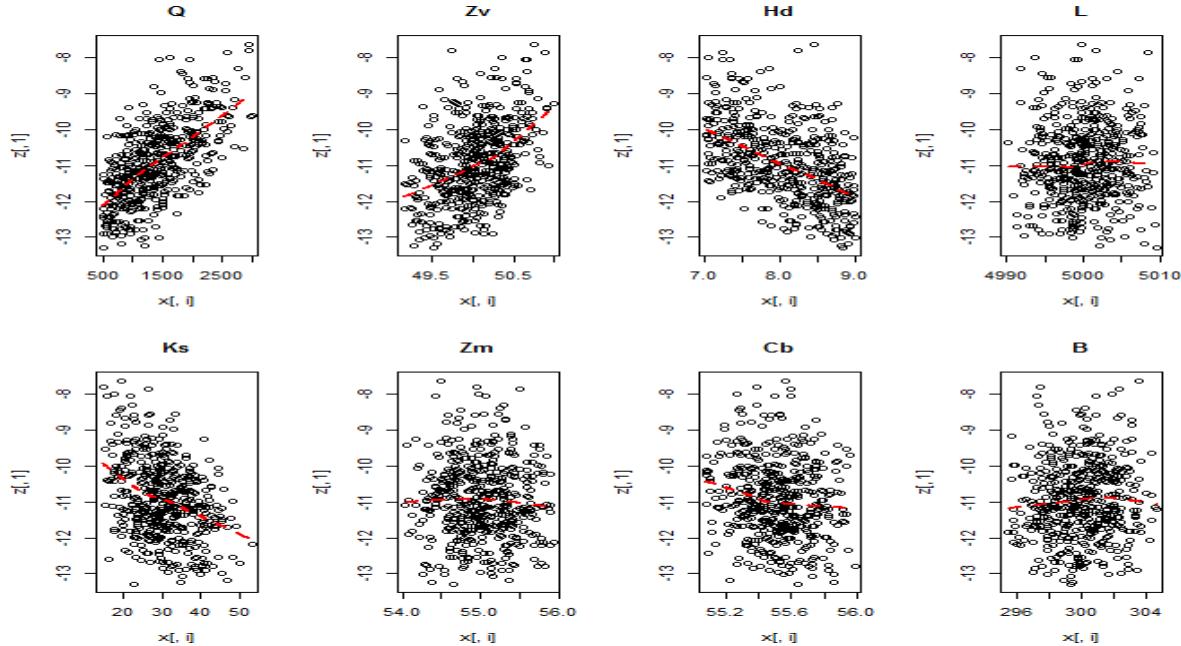
B = largeur rivière $\sim \mathcal{T}[295,305]$

MESURES D'INFLUENCE GLOBALE

► MODÈLE DE CRUES - SCATTERPLOTS – SORTIE S

$$S = H - (C_b - Z_v) - H_d$$

avec $H = \left(\frac{Q}{B K_s \sqrt{\frac{Z_m - Z_v}{L}}} \right)^{0.6}$

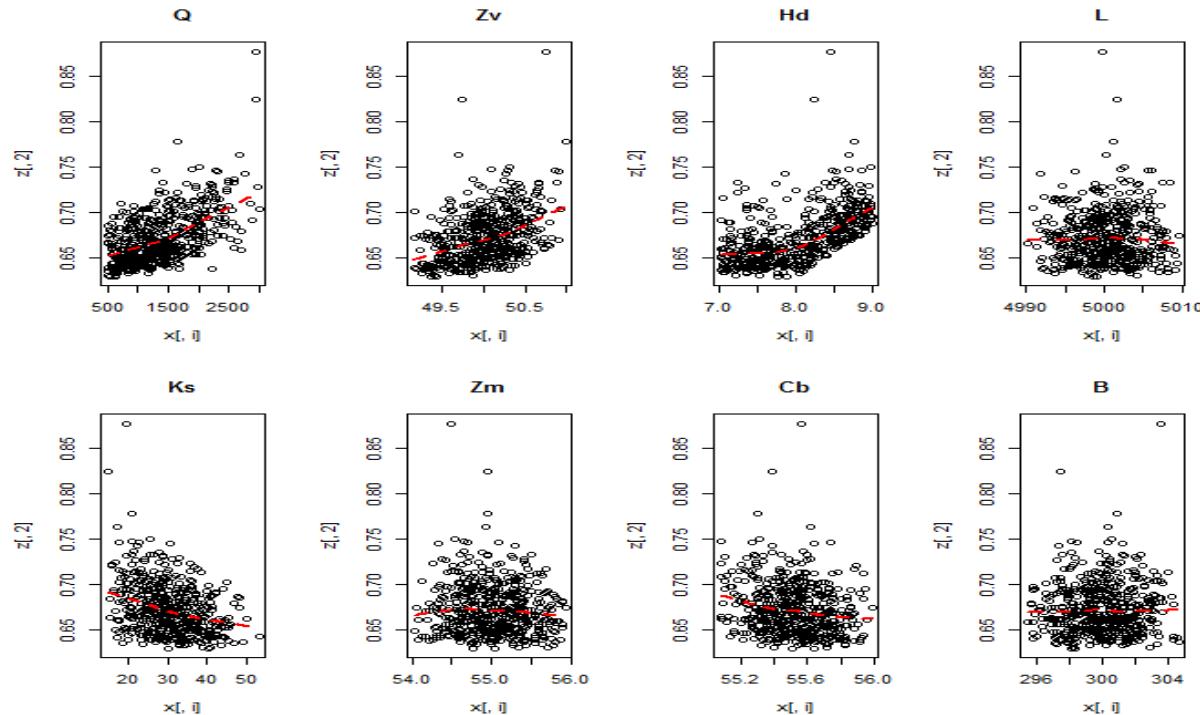


Echantillon Monte
Carlo $N = 500$

MESURES D'INFLUENCE GLOBALE

► MODÈLE DE CRUES - SCATTERPLOTS – SORTIE C_p

$$\begin{aligned} C_p &= \mathbb{1}_{S>0} + \left\{ 0.2 + 0.8 \left[1 - \exp \left(-\frac{1000}{S_0^4} \right) \right] \right\} \mathbb{1}_{S \leq 0} \\ &+ \frac{1}{20} (H_d \mathbb{1}_{H_d > 8} + 8 \mathbb{1}_{8 \leq H_d}) , \end{aligned}$$

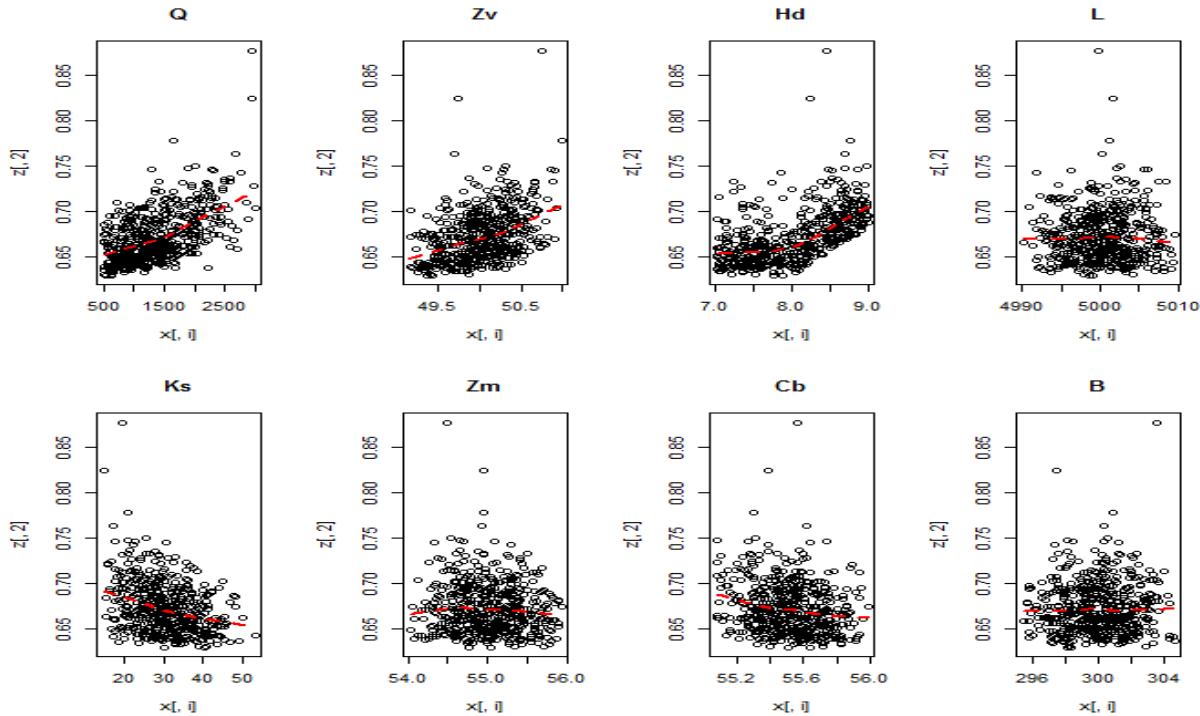


Echantillon Monte
Carlo $N = 500$

MESURES D'INFLUENCE GLOBALE

► MODÈLE DE CRUES - SCATTERPLOTS – SORTIE C_p

$$C_p = \mathbb{1}_{S>0} + \left\{ 0.2 + 0.8 \left[1 - \exp \left(-\frac{1000}{S_0^4} \right) \right] \right\} \mathbb{1}_{S \leq 0} \\ + \frac{1}{20} (H_d \mathbb{1}_{H_d > 8} + 8 \mathbb{1}_{8 \leq H_d}) ,$$



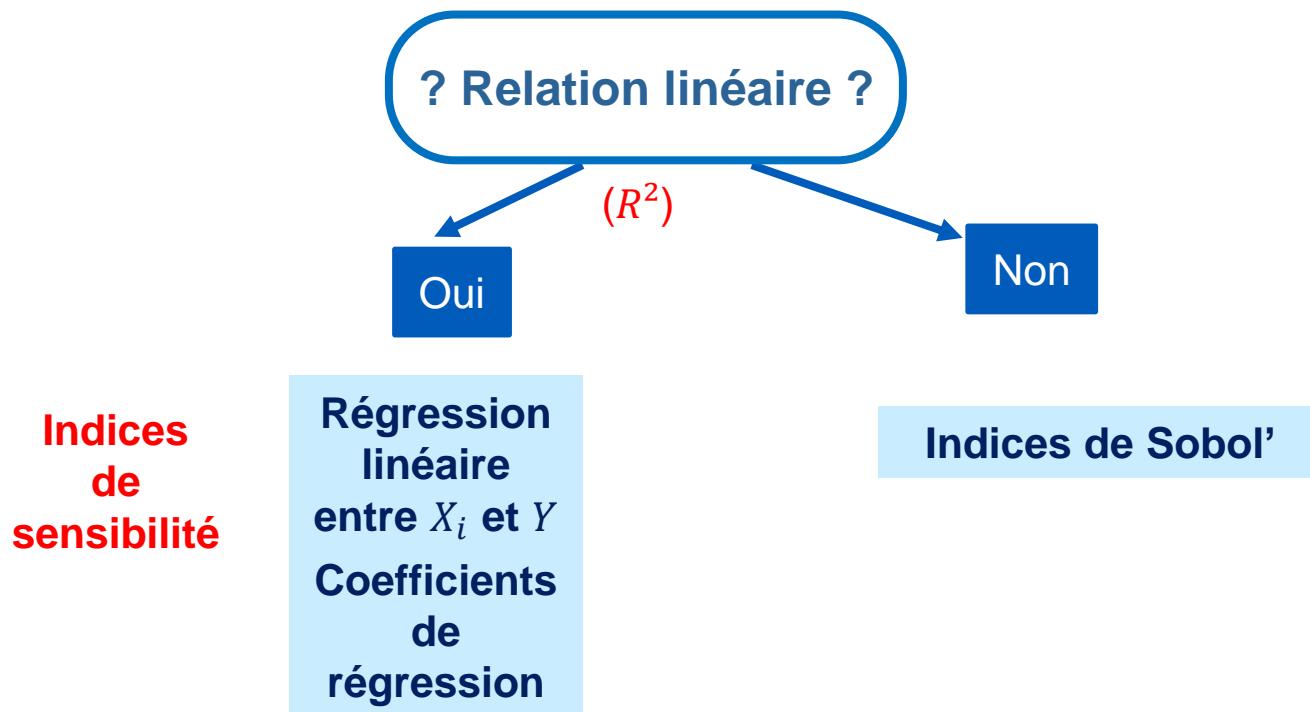
Limite majeure : analyse seulement les relations du premier ordre et pas les effets des interactions entre les entrées

MESURES D'INFLUENCE GLOBALE

Échantillon $(X, Y(X))$ de taille $N > p$, de préférence de taille $N \gg p$

Étape préliminaire : visualisation graphique (par ex : scatterplots)

Méthodologie d'analyse de sensibilité quantitative [Saltelli et al. 00, Helton et al. 06]



INDICES DE SENSIBILITÉ DANS LE CAS D'UNE RELATION ENTRÉES/SORTIE LINÉAIRE

Variables d'entrées $X = (X_1, \dots, X_p)$ indépendantes

Echantillon : N réalisations de (X, Y)

$$Y = \beta_0 + \sum_{i=1}^p \beta_i X_i \quad (\text{cf. rappels sur la régression linéaire en annexe})$$

- ▶ L'indice SRC : $SRC(X_i) \triangleq \beta_i \sqrt{\frac{\text{Var}(X_i)}{\text{Var}(Y)}}$

Le signe de β_i donne le sens de variation de Y par rapport à X_i

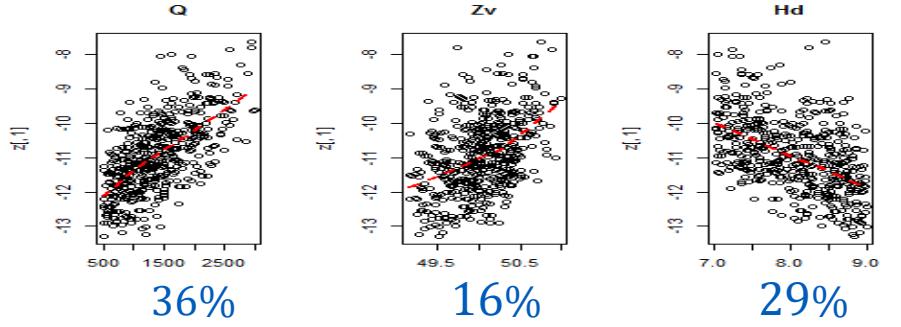
- ▶ Similaire au coefficient de corrélation linéaire (Pearson)
- ▶ Validité du modèle linéaire via les diagnostics de la régression et le R^2 :

$$R^2 = 1 - \frac{\sum_{k=1}^N (\hat{Y}_k - Y_k)^2}{\sum_{k=1}^N (Y_k - \bar{Y})^2}$$

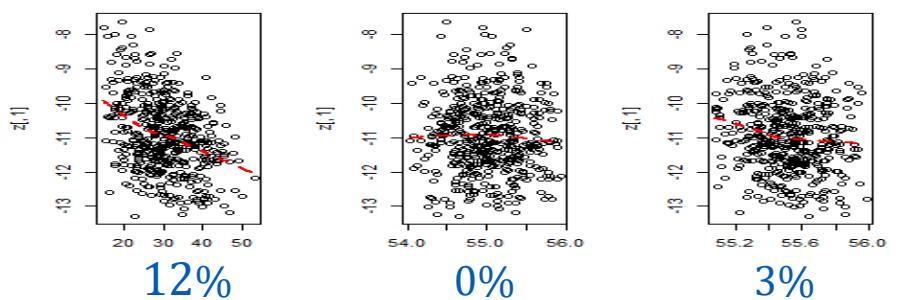
- ▶ On a $R^2 = \sum_{i=1}^p SRC^2(X_i)$ quand R^2 est proche de 1, ce qui permet d'interpréter aisément les SRC

INDICE SRC² - MODÈLE DE CRUES – SORTIE S

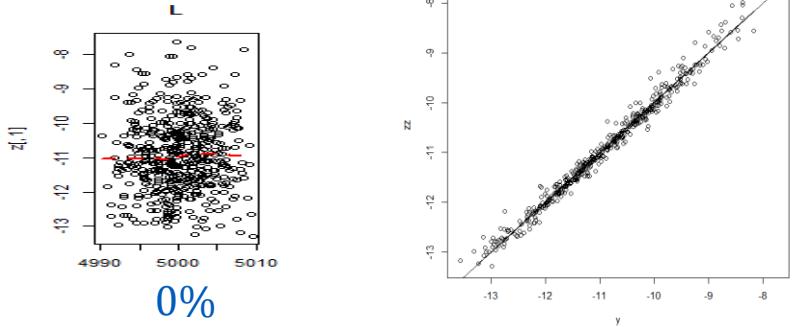
Echantillon Monte Carlo – $N = 500$



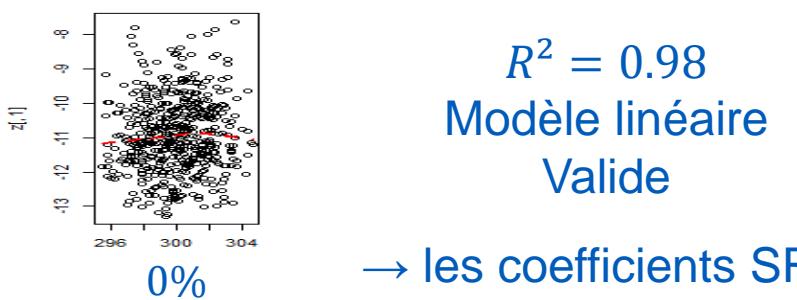
36%



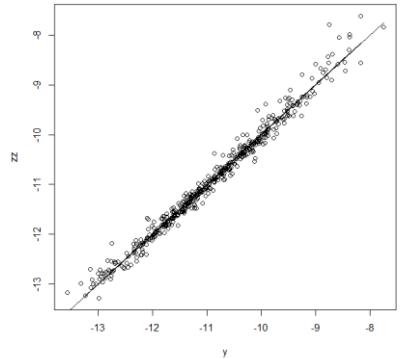
12%



0%



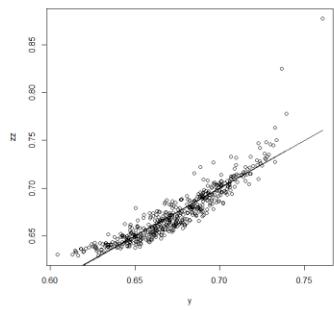
0%



$R^2 = 0.98$
Modèle linéaire
Valide

→ les coefficients SRC
sont des indices de
sensibilité valides

Sortie C_p : $R^2 = 0.83$;



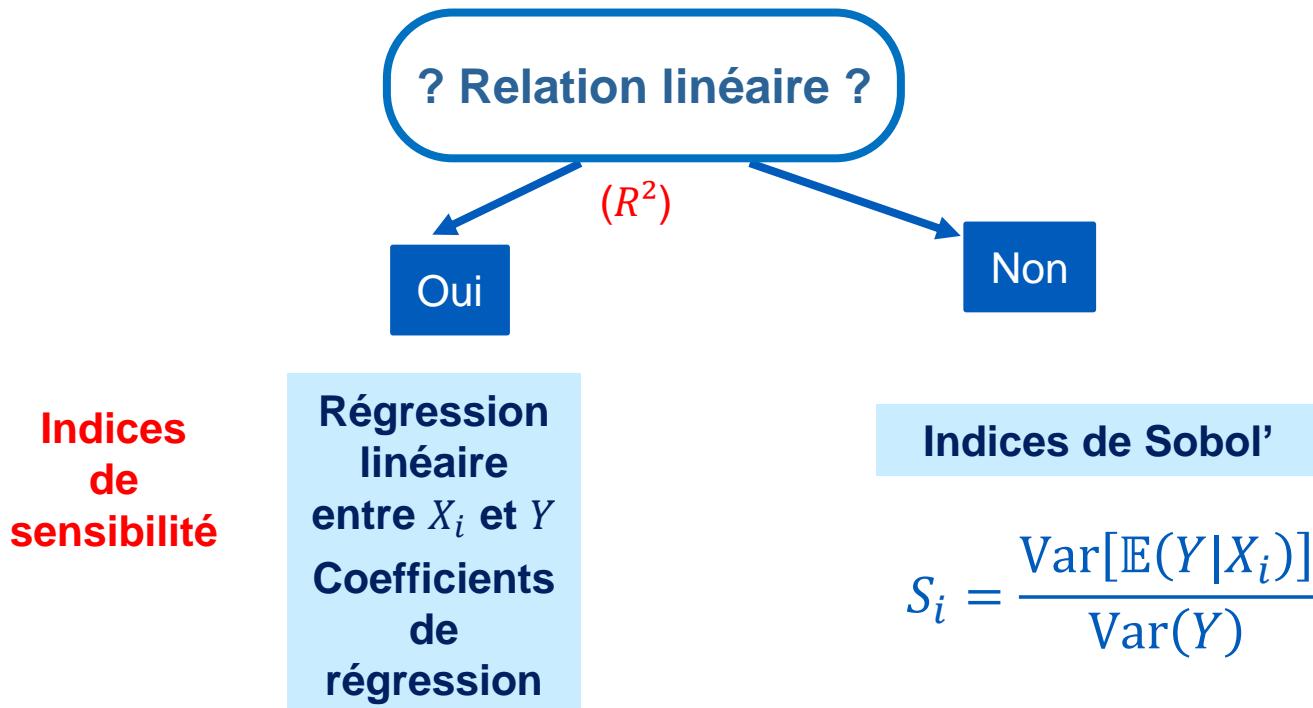
→ modèle linéaire insuffisamment valide

MESURES D'INFLUENCE GLOBALE

Échantillon $(X, Y(X))$ de taille $N > p$, de préférence de taille $N \gg p$

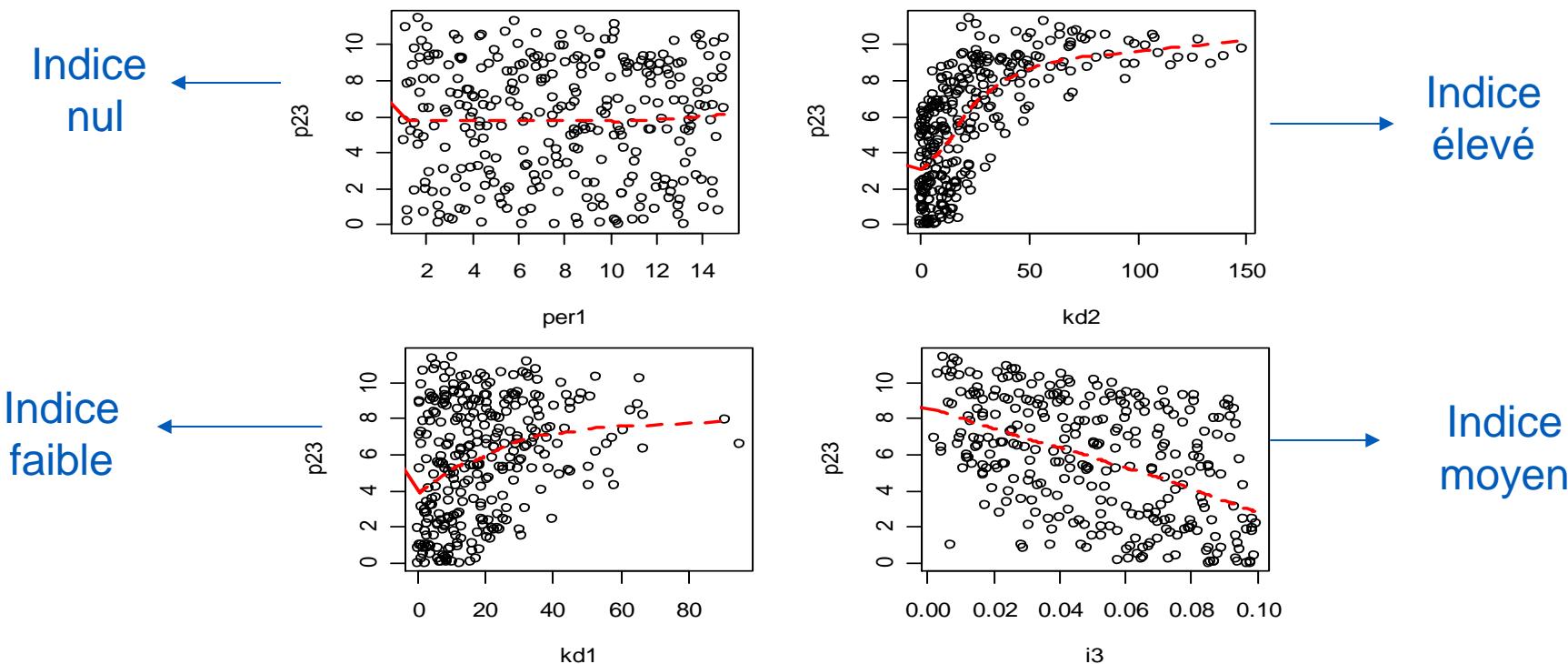
Étape préliminaire : visualisation graphique (par ex : scatterplots)

Méthodologie d'analyse de sensibilité quantitative [Saltelli et al. 00, Helton et al. 06]



Indices de Sobol : interprétation graphique

Les indices de Sobol du premier ordre mesurent la variabilité des espérances conditionnelles (courbes de tendance dans les scatterplots)



La courbe rouge correspond à un lissage, c'est-à-dire à $E(Y|X_i)$

Le lissage est réalisé ici par polynômes locaux : $\hat{f}(\mathbf{x}) = \hat{\alpha}(\mathbf{x}) + \mathbf{x}\hat{\beta}(\mathbf{x})$

Autres méthodes envisageables : moyenne mobile, polynômes, splines, modèles additifs, ...

Mesures d'influence globale : décomposition fonctionnelle

$$y = G(\mathbf{x}) = G_0 + \sum_{i=1}^p G_i(x_i) + \sum_{i=1}^p \sum_{j>i} G_{ij}(x_i, x_j) + \dots + G_{12\dots p}(x_1, \dots, x_p)$$

avec $G(\mathbf{x}) \in \mathcal{L}^2(\mathbf{x}), \mathbf{x} \in [0; 1]^p$ [Hoeffding 1946]

→ Il existe une infinité de décompositions possibles

MAIS, la décomposition est unique si : $\int G_{i_1, \dots, i_s}(x_{i_1}, \dots, x_{i_s}) dx_j = 0, \forall j = i_1, \dots, i_s$

Propriétés ($x_i \sim \mathcal{U}[0; 1]$ pour $i = 1, \dots, p$, les x_i sont indépendants)

$$G_0 = \int G(\mathbf{x}) d\mathbf{x} = \mathbb{E}[Y]$$

$$G_i(x_i) = \int G(\mathbf{x}) d\mathbf{x}_{-i} - G_0 = \mathbb{E}[Y|x_i] - G_0$$

$$G_{ij}(x_i, x_j) = \mathbb{E}[Y|x_i, x_j] - \mathbb{E}[Y|x_i] - \mathbb{E}[Y|x_j] + G_0$$

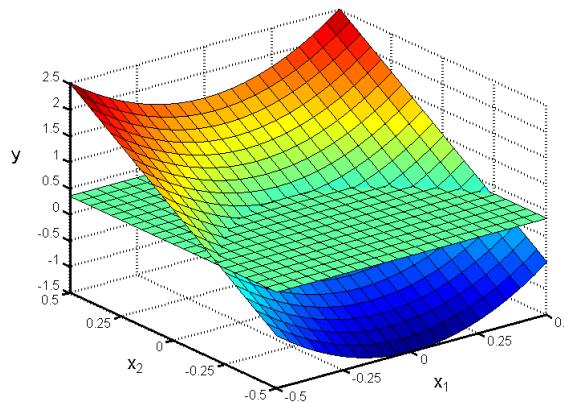
...

Exercice : $G(x_1, x_2) = x_1 + x_2; x_1 \sim U[0; 1]; x_2 \sim U[0; 1]$
 $G_0 = ?; G_1(x_1) = ?; G_2(x_2) = ?; G_{12}(x_1, x_2) = ?$

Un exemple

$$G(x_1, x_2) = 4x_1^2 + 3x_2$$

$$x_1, x_2 \sim \mathcal{U}[-1/2; 1/2]$$



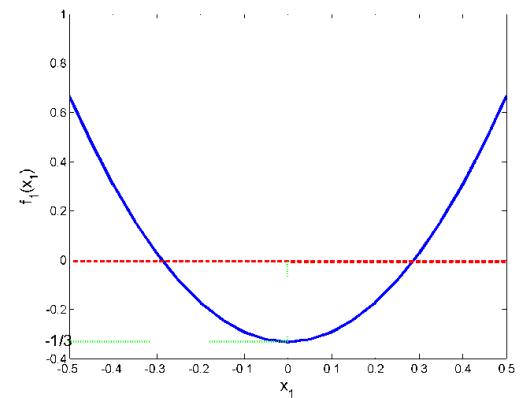
$$\begin{aligned} G_0 &= 0 \\ G_1(x_1) &= 4x_1^2 \\ G_2(x_2) &= 3x_2 \\ G_{12}(x_1, x_2) &= 0 \end{aligned}$$

$$G_0 = \mathbb{E}(y) = \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} (4x_1^2 + 3x_2) dx_1 dx_2 = 1/3$$

$$G_1(x_1) = \mathbb{E}(y|x_1) - G_0 = \int_{-1/2}^{1/2} (4x_1^2 + 3x_2) dx_2 = 4x_1^2 - 1/3$$

$$G_2(x_2) = \mathbb{E}(y|x_2) - G_0 = x_2$$

$$G_{12}(x_1, x_2) = 0$$



Mesures d'influence globale : Indices de Sobol

ANOVA fonctionnelle [Efron & Stein 81] (hyp. de X_i indépendants) :

$$\text{Var}(Y) = \sum_{i=1}^p V_i(Y) + \sum_{i=1}^p \sum_{j>i} V_{ij}(Y) + \dots + V_{12\dots p}(Y)$$

où $V_i(Y) = \text{Var}[\mathbb{E}(Y|X_i)]$, $V_{ij}(Y) = \text{Var}[\mathbb{E}(Y|X_i, X_j)] - V_i - V_j$, ...

Définition des indices de Sobol

- ↳ Indice de sensibilité d'ordre 1 :

$$S_i = \frac{V_i}{\text{Var}(Y)}$$

- ↳ Indice de sensibilité d'ordre 2 :

$$S_{ij} = \frac{V_{ij}}{\text{Var}(Y)}$$

...

Remarque : si le modèle est linéaire : $S_i = \text{SRC}^2(X_i)$

Exercice : $f(X_1, X_2) = X_1^2 + X_2$; $X_1 \sim U[0,1]$; $X_2 \sim U[0,1]$; $X_1 \perp X_2$
 $S_1 = ?$; $S_2 = ?$

Mesures d'influence globale : Indices de Sobol

$$y = G(x_1, x_2) = 4x_1^2 + 3x_2$$

avec $x_1, x_2 \in \mathcal{U}[-1/2; 1/2]$

On a vu :

$$G_0 = \mathbb{E}(Y) = 1/3$$

$$G_1(x_1) = \mathbb{E}[Y|x_1] - G_0 = 4x_1^2 - 1/3$$

$$G_2(x_2) = \mathbb{E}[Y|x_2] - G_0 = 3x_2$$

$$G_{12}(x_1, x_2) = 0$$

Ce qui donne après calcul :

$$S_1 = \frac{\text{Var}[G_1(x_1)]}{V} = \frac{0,08}{0,838} = 0,106$$

$$S_2 = \frac{\text{Var}[G_2(x_2)]}{V} = \frac{0,75}{0,838} = 0,894$$

Indices de Sobol : autre interprétation

- ▶ Question centrale de l'analyse de sensibilité : la sortie Y est-elle plus ou moins variable lorsque l'on fixe une des entrées X_i ?
- ▶ On étudie $\text{Var}(Y|X_i = x_i)$ et on prend sa moyenne sur les valeurs de X_i :
$$\mathbb{E}[\text{Var}(Y|X_i = x_i)]$$
- ▶ Plus $\mathbb{E}[\text{Var}(Y|X_i = x_i)]$ est petit, plus X_i est influente
- ▶ Théorème de la variance totale : $\text{Var}(Y) = \mathbb{E}[\text{Var}(Y|X_i)] + \text{Var}[\mathbb{E}(Y|X_i)]$

- ▶ On en déduit les indices de Sobol : $S_i = \frac{\text{Var}[\mathbb{E}(Y|X_i)]}{\text{Var}(Y)}$

Propriétés des indices de Sobol

$$1 = \sum_{i=1}^p S_i + \sum_{i=1}^p \sum_{j>i} S_{ij} + \sum_i \sum_j \sum_k S_{ijk} + \cdots + S_{12\dots p}$$

$\sum_{i=1}^p S_i \leq 1$ → Toujours vrai

$\sum_{i=1}^p S_i = 1$ → Le modèle est additif

$1 - \sum_{i=1}^p S_i$ → Mesure le degré d'interactions entre variables

Exemple : $p = 4$ donne 4 indices S_i , 6 indices S_{ij} , 4 indices S_{ijk} , 1 indice S_{ijkl}

Cas général : $2^p - 1$ indices à estimer

Indice de sensibilité total :

$$S_{Ti} = S_i + \sum_{\substack{j=1 \\ j \neq i}}^p S_{ij} + \sum_{\substack{j=1 \\ j \neq i}}^p \sum_{\substack{k>j \\ k \neq i}} S_{ijk} + \cdots = 1 - S_{-i}$$

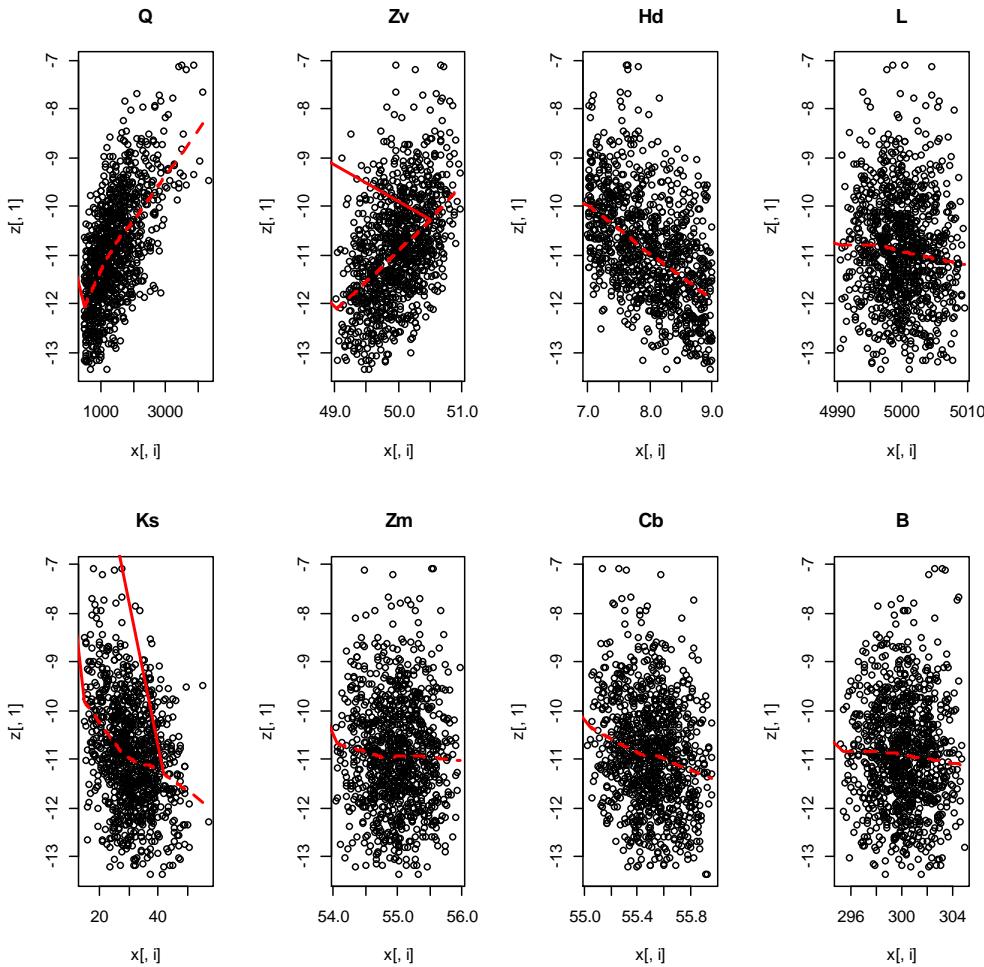
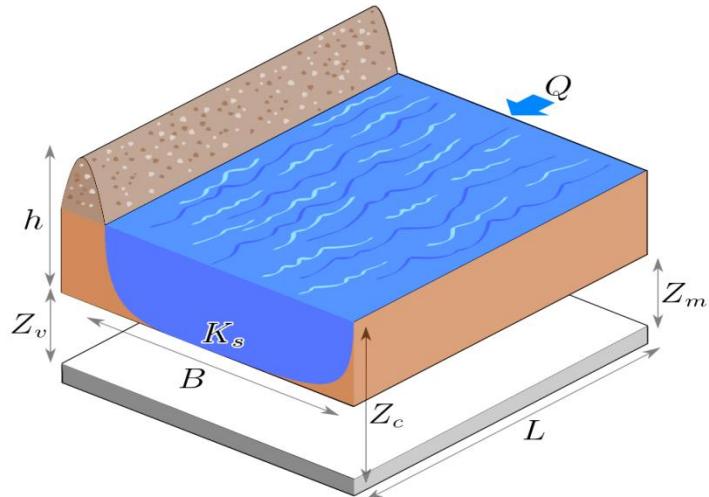
[Homma & Saltelli 1996]

Exemple - Modèle de crues – Sortie « Surverse » S

Echantillon Monte Carlo – $N = 1000$

$$S = Z_v + H - H_d - C_b$$

avec $H = \left(\frac{Q}{BK_s \sqrt{(Z_m - Z_v)/L}} \right)^{0.6}$

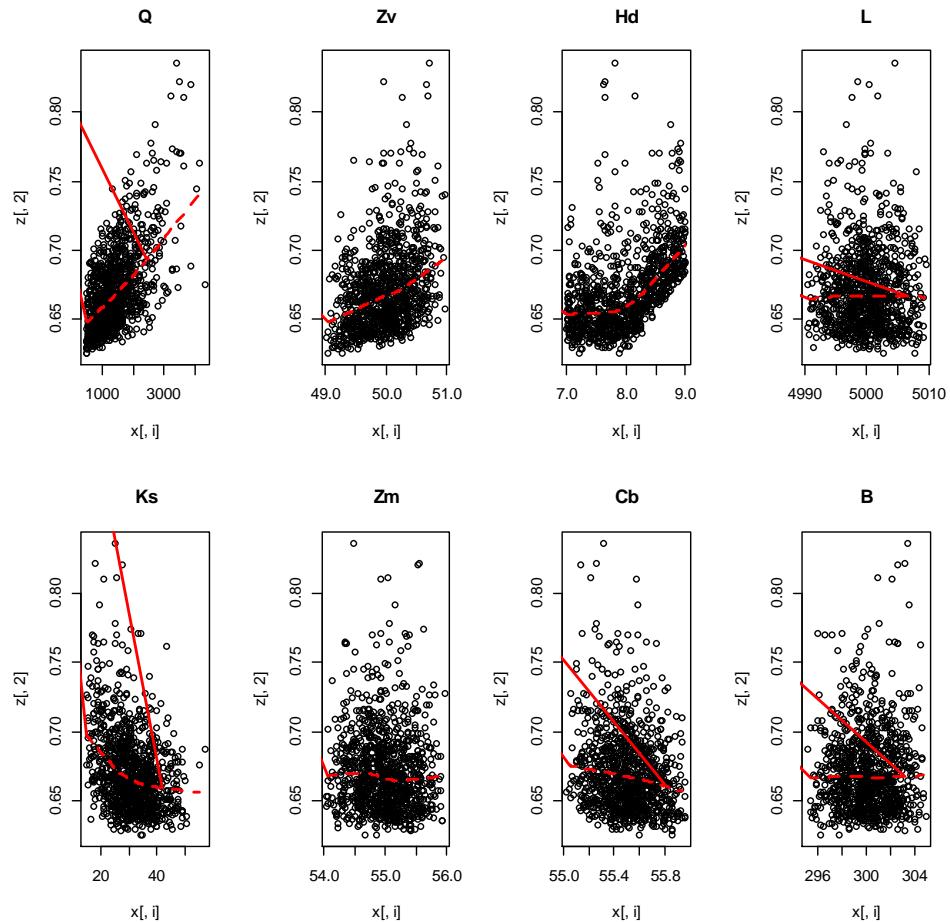


Exemple - Modèle de crues – Sortie « Coût » Cp

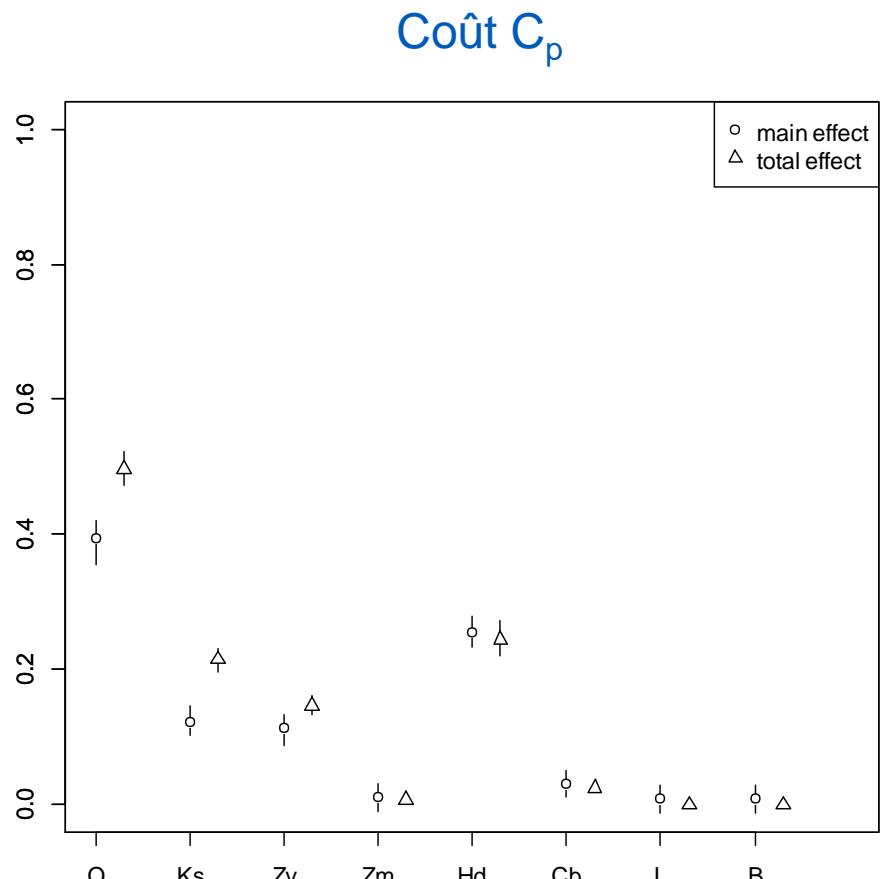
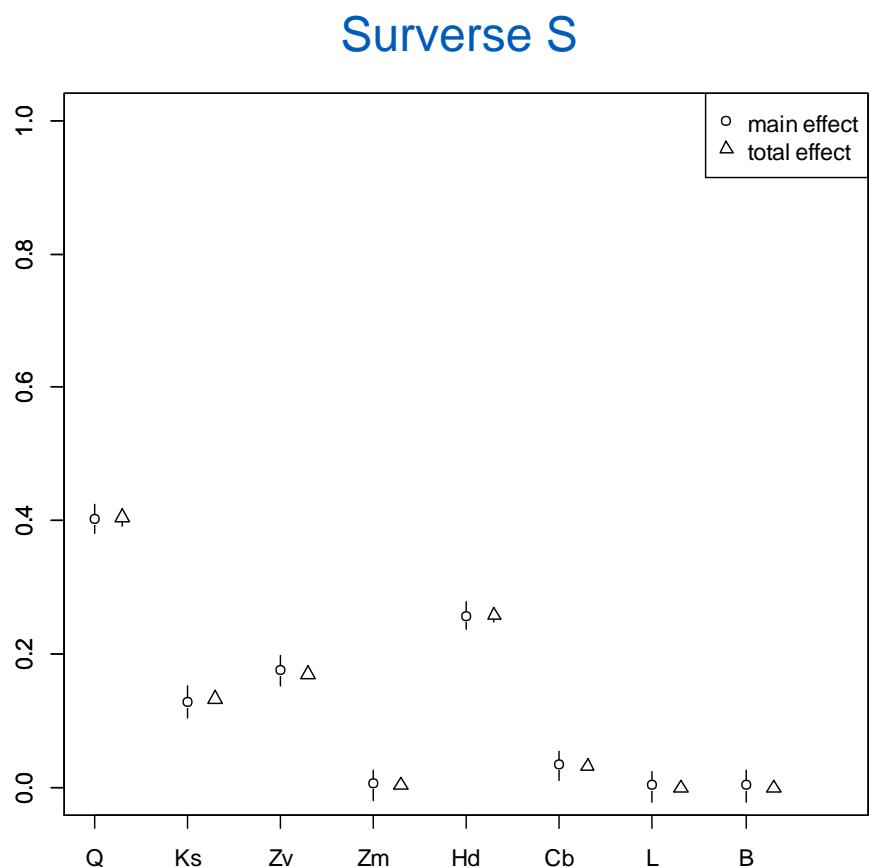
$$C_p = 1_{S>0} + \{0.2 + 0.8[1 - \exp(-1000 / S_0^4)]\} 1_{S \leq 0} + \frac{1}{20} (H_d 1_{H_d > 8} + 8 1_{H_d \leq 8})$$

avec $S = Z_v + H - H_d - C_b$ et $S_0 = Z_v + H - C_b$

Echantillon Monte Carlo
 $N = 1000$



Indices de Sobol : application sur le modèle de crues



Formule pour le calcul des indices de Sobol

► Formulations des indices pour X_i (1^{er} ordre et total) :

$$S_i = \frac{V_i}{\text{Var}(Y)} \text{ et } S_{Ti} = 1 - \frac{V_{-i}}{\text{Var}(Y)}$$

► Formulations des variances conditionnelles :

Soient $\mathbf{X} = (X_i, X_{-i})$ et $\mathbf{X}' = (X'_i, X'_{-i})$ une copie indépendante de \mathbf{X} , on a :

$$\begin{aligned} V_i(Y) &= \text{Var}[\mathbb{E}(Y|X_i)] = \int (\mathbb{E}[Y|x_i])^2 dx_i - \left(\int \mathbb{E}(Y|x_i) dx_i \right)^2 \\ &= \text{cov}[G(X_i, X_{-i}), G(X_i, X'_{-i})] = \text{cov}[G(\mathbf{X}), G(X_i, X'_{-i})] \end{aligned}$$

$$V_{-i}(Y) = \text{Var}[\mathbb{E}(Y|X_{-i})] = \text{cov}[G(X_i, X_{-i}), G(X'_i, X_{-i})]$$

Preuve : on utilise que $G(X_i, X_{-i})$ et $G(X_i, X'_{-i})$ sont :

- de même loi,
- indépendants conditionnellement à X_i

Estimation des indices de Sobol par Monte Carlo

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

► **Variance (estimateur classique)** : $\hat{V}(Y) = \frac{1}{n} \sum_{k=1}^n G(X^{(k)})^2 - \hat{G}_0^2$ avec $\hat{G}_0 = \frac{1}{n} \sum_{k=1}^n G(X^{(k)})$

► **Estimations des variances conditionnelles :**

$$\hat{V}_i(Y) = \frac{1}{n} \sum_{k=1}^n G\left(X_1^{(k)}, \dots, X_{i-1}^{(k)}, X_i^{(k)}, X_{i+1}^{(k)}, \dots, X_p^{(k)}\right) G\left(X'_1^{(k)}, \dots, X'_{i-1}^{(k)}, X_i^{(k)}, X'_{i+1}^{(k)}, \dots, X'_p^{(k)}\right) - G_0^2$$

$$\hat{V}_{-i}(Y) = \frac{1}{n} \sum_{k=1}^n G\left(X_1^{(k)}, \dots, X_{i-1}^{(k)}, X_i^{(k)}, X_{i+1}^{(k)}, \dots, X_p^{(k)}\right) G\left(X_1^{(k)}, \dots, X_{i-1}^{(k)}, X'_i^{(k)}, X_{i+1}^{(k)}, \dots, X_p^{(k)}\right) - G_0^2$$

Indices 1^{er} ordre : coût = $n(p + 1)$

Indices 1^{er} ordre + indices totaux : coût = $n(p + 2)$

► Astuce : on intervertit $\left(X_i^{(j)}\right)_{i=1,\dots,p; j=1,\dots,n}$ et $\left(X'_i^{(j)}\right)_{i=1,\dots,p; j=1,\dots,n}$ dans $\hat{V}_{-i}(Y)$

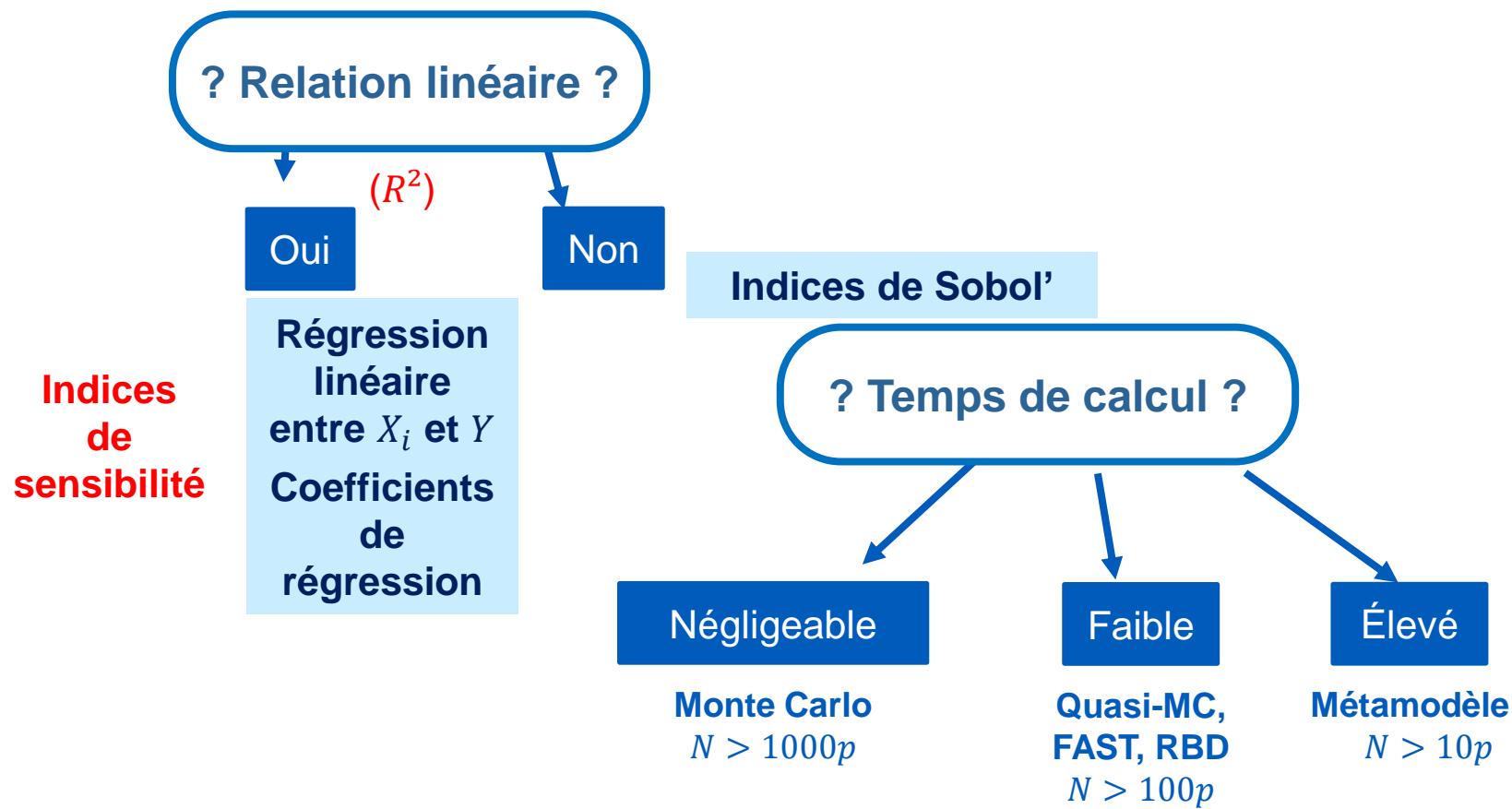
En pratique, $n \sim 10^4 \rightarrow$ problème du coût en nombre d'évaluations nécessaires

Analyse de sensibilité globale

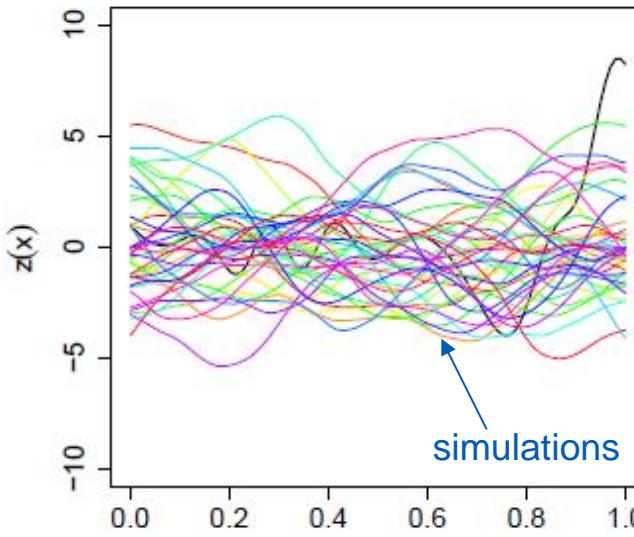
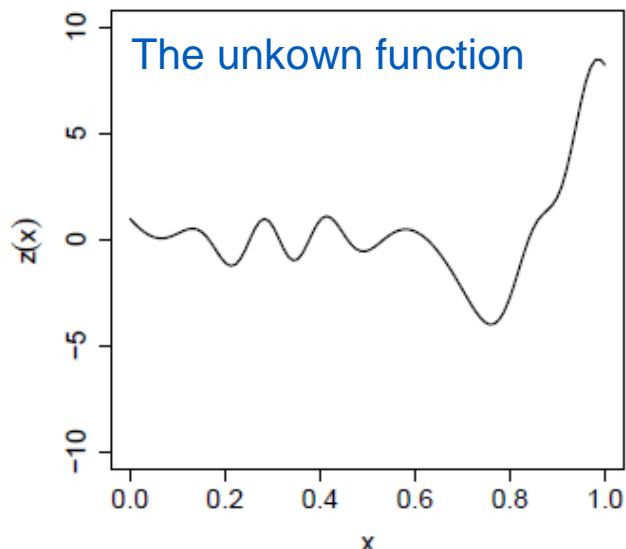
Échantillon $(X, Y(X))$ de taille $N > p$, de préférence de taille $N \gg p$

Étape préliminaire : visualisation graphique (par ex : scatterplots)

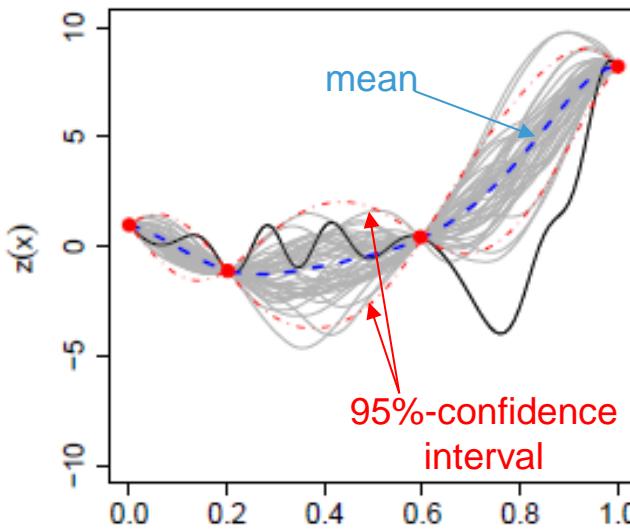
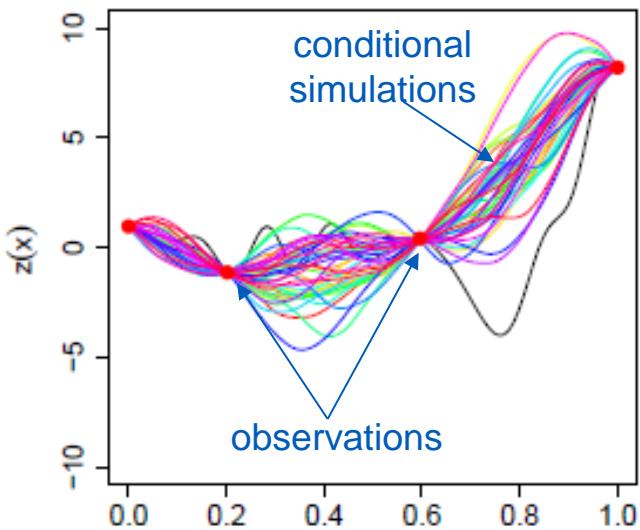
Méthodologie d'analyse de sensibilité quantitative [Saltelli et al. 00, Helton et al. 06]



Exemple de métamodèle : le krigage (processus gaussien)



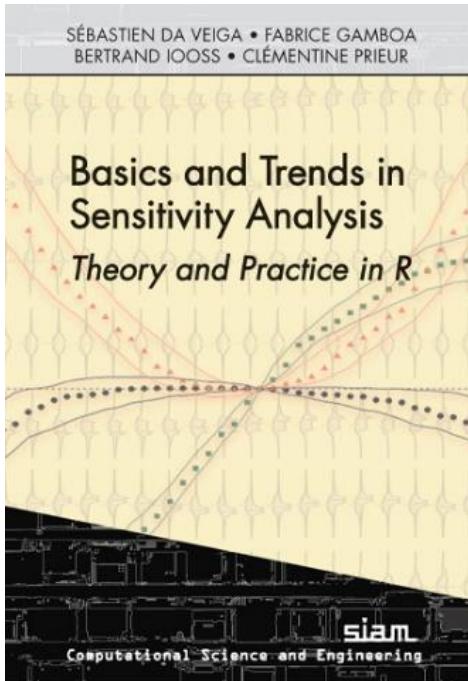
Main hypothesis: $z(x)$ is the realization of a Gaussian process (defined by its mean and cov)



GP metamodel is given by the probability law of $Z(x)$ conditionally to the observations

Its mean gives the GP predictor
Its variance gives the error

Crédits & Bibliographie



- Baudin & Martinez, *Introduction to sensitivity analysis with NISP*, 2013
- L. Clouvel, V. Chabridon, B. Iooss, M. II Idrissi and F. Robin, An overview of variance-based importance measures in the linear regression context: comparative analyses and numerical tests, *Socio-Environmental Systems Modelling*, vol. 7, 18681, 2025, doi:10.18174/sesmo.1868
- Saltelli et al., Global sensitivity analysis – The primer, Wiley, 2008