



Analyse de sensibilité globale de sorties spatiales de modèles numériques

B. Iooss

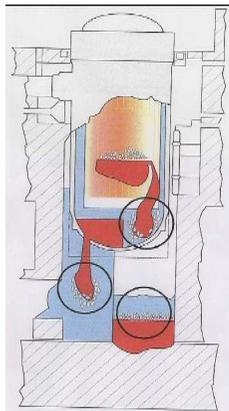
Travail réalisé avec Amandine Marrel (CEA, IFP), Béatrice Laurent (INSA Toulouse), Elena Volkova (Institut Kurchatov) et Michel Jullien (CEA Cad)

CEA Cadarache
Direction de l'Énergie Nucléaire
Département d'Études des Réacteurs
Laboratoire de Conduite et Fiabilité des Réacteurs

2010

Quelques enjeux de la prise en compte des incertitudes

• Modélisation :

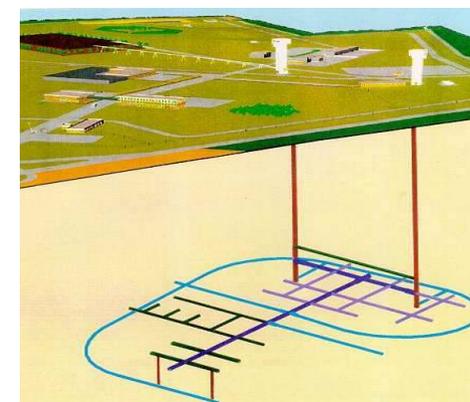


- améliorer le modèle
- explorer « au mieux » différentes combinaisons des entrées
- identifier les entrées prépondérantes (afin de prioriser la R&D)



• Validation :

- réduire l'incertitude de prédiction
- calibrer les paramètres du modèle



• Étude de sûreté : calculer un **risque** de défaillance

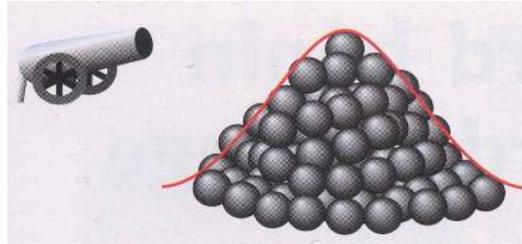
• Étude de conception : optimiser **les performances** d'un système

L'approche statistique - $Y = f(\mathbf{X})$, $\mathbf{X} \in \mathcal{D} \subset \mathbb{R}^p$, $Y \in \mathbb{R}$

Échantillonnage Monte Carlo (MC) / Quasi MC / Plan d'expériences

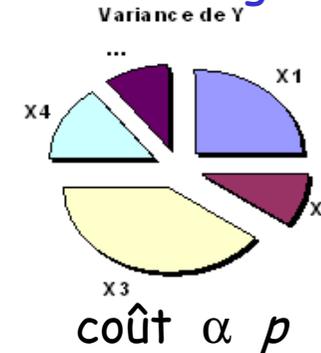
cea

Propagation



coût (nb de calculs) élevé pour probas faibles

Sensibilités globales



Avantages de l'approche statistique :

- Simplicité (Monte Carlo bien connue, méthodologie communicable aux non spécialistes),
- Souplesse (distribution des calculs, changement des lois de proba des entrées, ...),
- Applicable au code de calcul dont les sources sont intouchables, inaccessibles ou incompréhensibles (majorité des codes de l'industrie),
- Mise à disposition des erreurs associées aux estimations, ...

Inconvénients : - perte de la formulation physique,

- coût en nb d'évaluations, malédiction de la dimension

➔ Méthodes proba. et plans d'expériences spécifiques pour résoudre ces pbs

Plan



- *Analyse de sensibilité et métamodèle*
- Métamodèle pour sortie spatiale
- Analyse de sensibilité globale pour sortie spatiale
- Conclusions

MARTHE application

- Computation of the spatio-temporal evolution of ^{90}Sr concentration in an ancient radwaste disposal site

[Volkova et al. 08]



- Goal: Estimate the contamination impact on the environment

Numerical modelisation :

Hydrogeological transport scenario of ^{90}Sr between 2002 and 2010 with the MARTHE software

20 uncertain inputs :

Permeability, dispersivity, Kd, infiltration intensity, ...

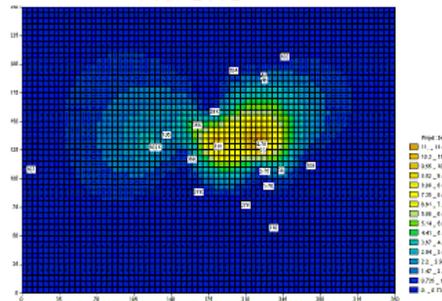
Output of interest :

Concentration values

Quantity of interest :

Distribution, variance

Carte de concentration initiale
Août 2002



Jeu de valeurs
pour les 20 paramètres d'entrée

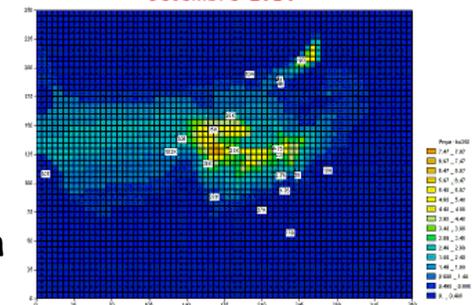
per1	perz3	dt1	kd3
per2	perz4	dt2	poros
per3	d1	dt3	i1
perz1	d2	kd1	i2
perz2	d3	kd2	i3

MARTHE



30 mn / run

Carte de concentration finale
Décembre 2010

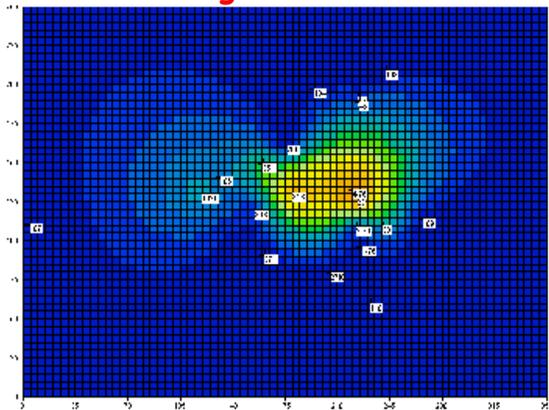


Sorties scalaires : analyse d'incertitudes en 2010



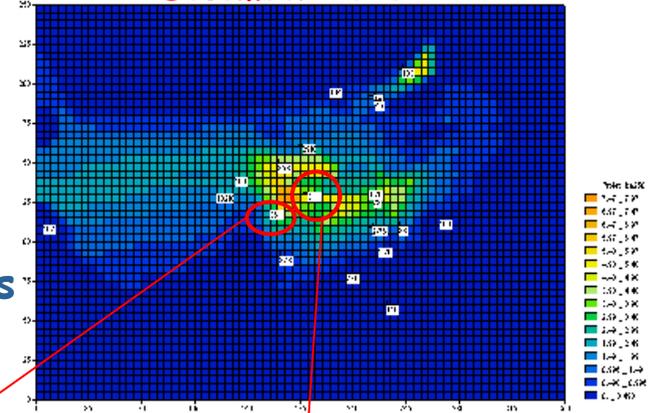
Concentration
initiale

August 2002



Cartes de
concentrations

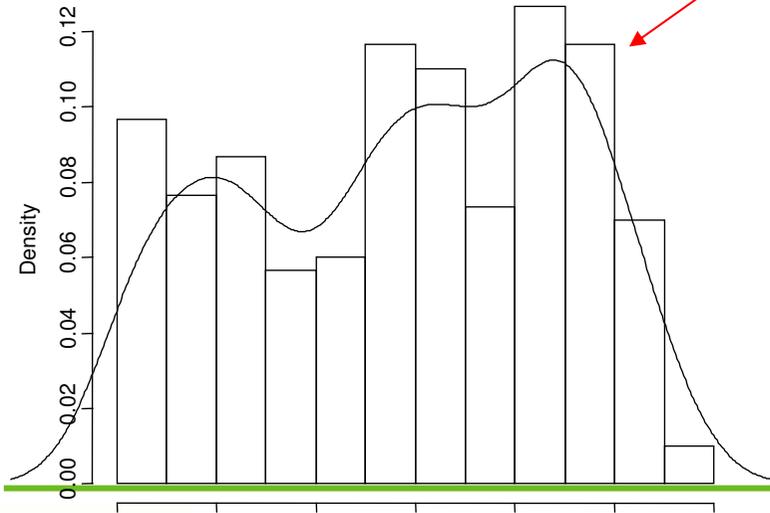
December 2010



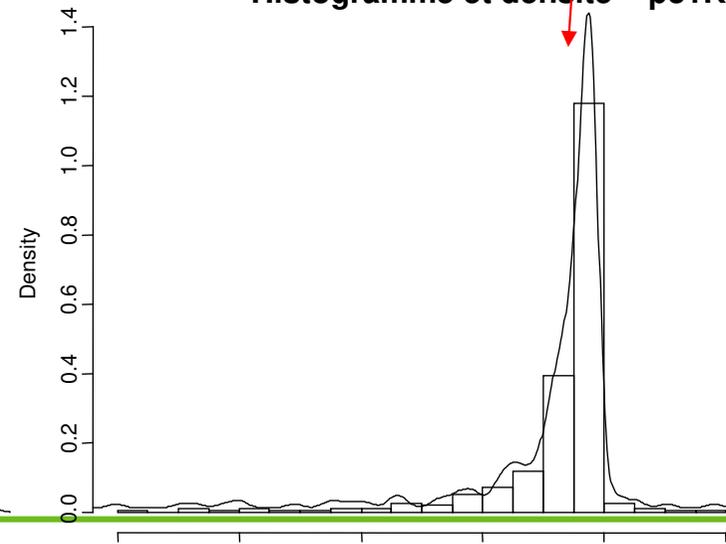
$N = 300$ calculs de type Monte-Carlo (LHS)

distributions des concentrations aux piézomètres (Bq/l)

Histogramme & densité - p23



Histogramme et densité - p31K



1 sortie scalaire – Analyse de sensibilité

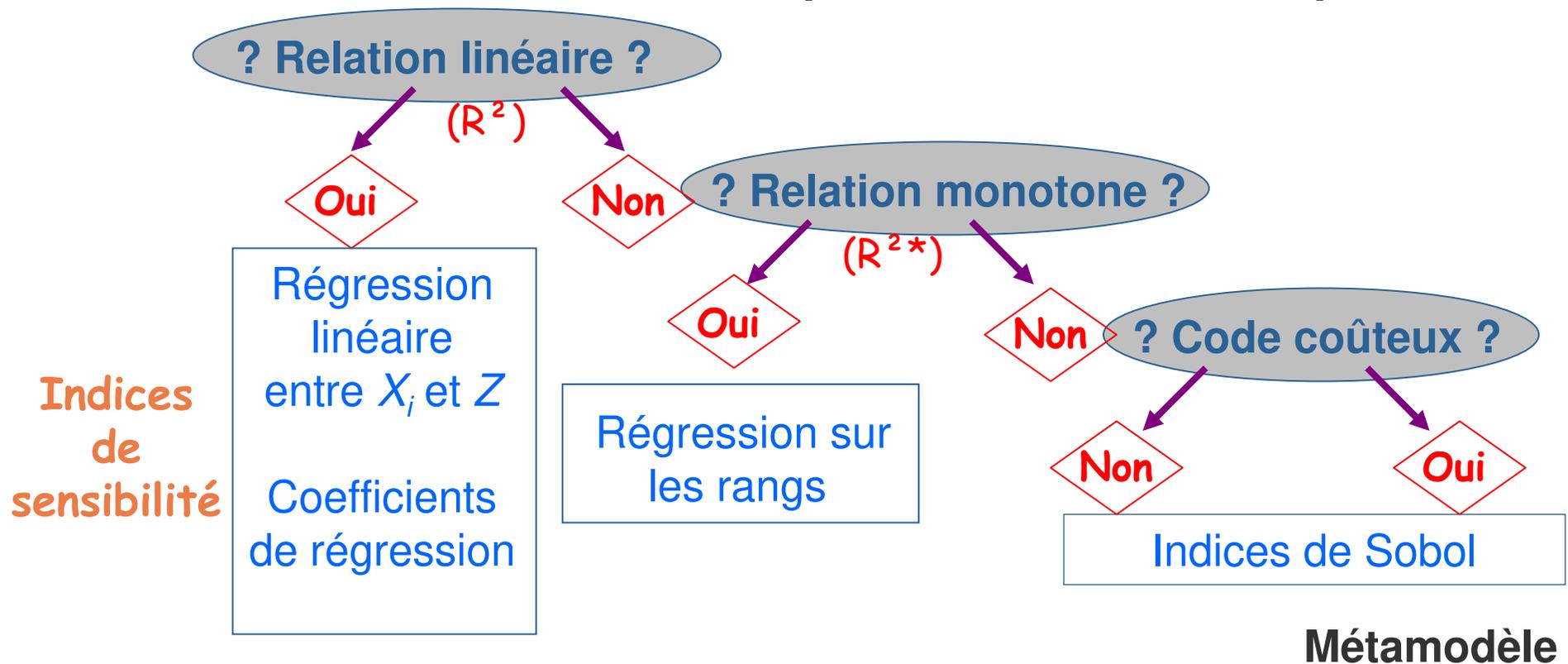
Échantillon $(X, Y(X))$ de taille $N > p$, de préférence de taille $N \gg p$

Étape préliminaire : visualisation graphique (par ex : scatterplots)



Méthodologie d'analyse de sensibilité quantitative

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



Analyse de sensibilité : indices de Sobol

Décomposition : f intégrable sur $[0,1]^p$

$$f(X_1, \dots, X_p) = f_0 + \sum_{i=1}^p f_i(X_i) + \sum_{i < j}^p f_{ij}(X_i, X_j) + \dots + f_{12\dots p}(X_1, \dots, X_p)$$



ANOVA fonctionnelle :

$$\text{Var}(Y) = \sum_{i=1}^p V_i(Y) + \sum_{i < j}^p V_{ij}(Y) + \dots + V_{12\dots p}(Y) \text{ (hyp. de } X_i \text{ indépendants)}$$

$$V_i = \text{Var}_{X_i} [E(Y | X_i)]$$

$$S_i = \frac{V_i}{\text{Var}(Y)} ; S_{ij} ; \dots ; S_{Ti} = S_i + \sum_j S_{ij} + \sum_{j,k} S_{ijk} + \dots$$

1er ordre 2ème ordre **Indice total**

Évaluations de S_i et S_{Ti} par :

- Monte Carlo [*Sobol 93, Saltelli 02*] (coût élevé), quasi Monte Carlo, FAST [*Saltelli et al. 99*] : **coût > 100 p**
- Une fonction mathématique à coût négligeable (métamodèle) : **coût ~ 10 p**
Puis : formulations analytiques ou Monte Carlo

Métamodèle

= modèle mathématique de coût négligeable, approximant les réponses du modèle étudié, en vue de prédire de nouvelles réponses



Utilisations pratiques : analyse de sensibilité, propagation d'incertitudes, optimisation de réponses du modèle, calibration de paramètres

Exemples : polynômes, splines, réseaux de neurones, SVM, ...

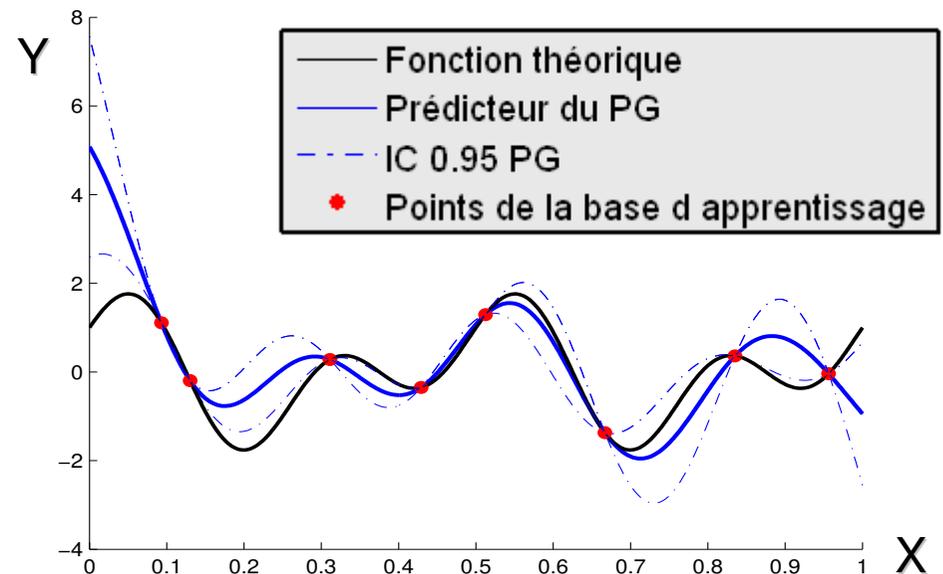
Métamodèle du krigage (processus gaussien) : [Matheron 63, Sacks al. 89]

- interpole les réponses (comme en cartographie spatiale),
- fournit en plus d'une prédiction, une estimation de son erreur

Exemple en 1D :

$$Y(X) = \cos(4\pi X) + \sin(8\pi X)$$

Simulation de $N = 8$ calculs
Prédiction sur $[0,1]$



Métamodèle processus gaussien (PG)

- Hypothèses :

$$Y(\mathbf{x}) = h(\mathbf{x}) + Z(\mathbf{x})$$

ceci **Régression (pol d°1)** **Partie stochastique**
 $\beta F(\mathbf{x})$

$$F(\mathbf{x}) = \beta_0 + \sum_{i=1}^p \beta_i x_i \quad R(\mathbf{u}, \mathbf{v}) = R(\mathbf{u} - \mathbf{v}) = \exp\left(-\sum_{i=1}^p \theta_i |u_i - v_i|^{q_i}\right)$$

(cov. factorisable)

Processus stochastique stationnaire Z

$$E[Z(\mathbf{x})] = 0$$

$$\text{Cov}(Z(\mathbf{u}), Z(\mathbf{v})) = \sigma^2 R(\mathbf{u} - \mathbf{v})$$

où σ^2 est la variance

et $R(\cdot)$ la fonction de corrélation

- Echantillon d'apprentissage de N simulations: $X_{LS} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}), Y_{LS} = f(X_{LS})$

- Métamodèle PG : prédicteur et covariance

$$Y(\mathbf{x})|_{X_{LS}, Y_{LS}} \sim PG$$

$$\hat{Y}(\mathbf{x}) = E^*[Y(\mathbf{x})] = \beta F(\mathbf{x}) + r(\mathbf{x})^t R_{LS}^{-1} [Y_{LS} - \beta F_{LS}]; \quad r(\mathbf{x}) = [R(\mathbf{x}^{(1)}, \mathbf{x}), \dots, R(\mathbf{x}^{(N)}, \mathbf{x})]$$

$$\text{Cov}^*(Y(\mathbf{u})|_{X_{LS}, Y_{LS}}, Y(\mathbf{v})|_{X_{LS}, Y_{LS}}) = \sigma^2 (R(\mathbf{u}, \mathbf{v}) + r(\mathbf{u})^t R_{LS}^{-1} r(\mathbf{v})) \Rightarrow \text{MSE} [\hat{Y}(\mathbf{x})]$$

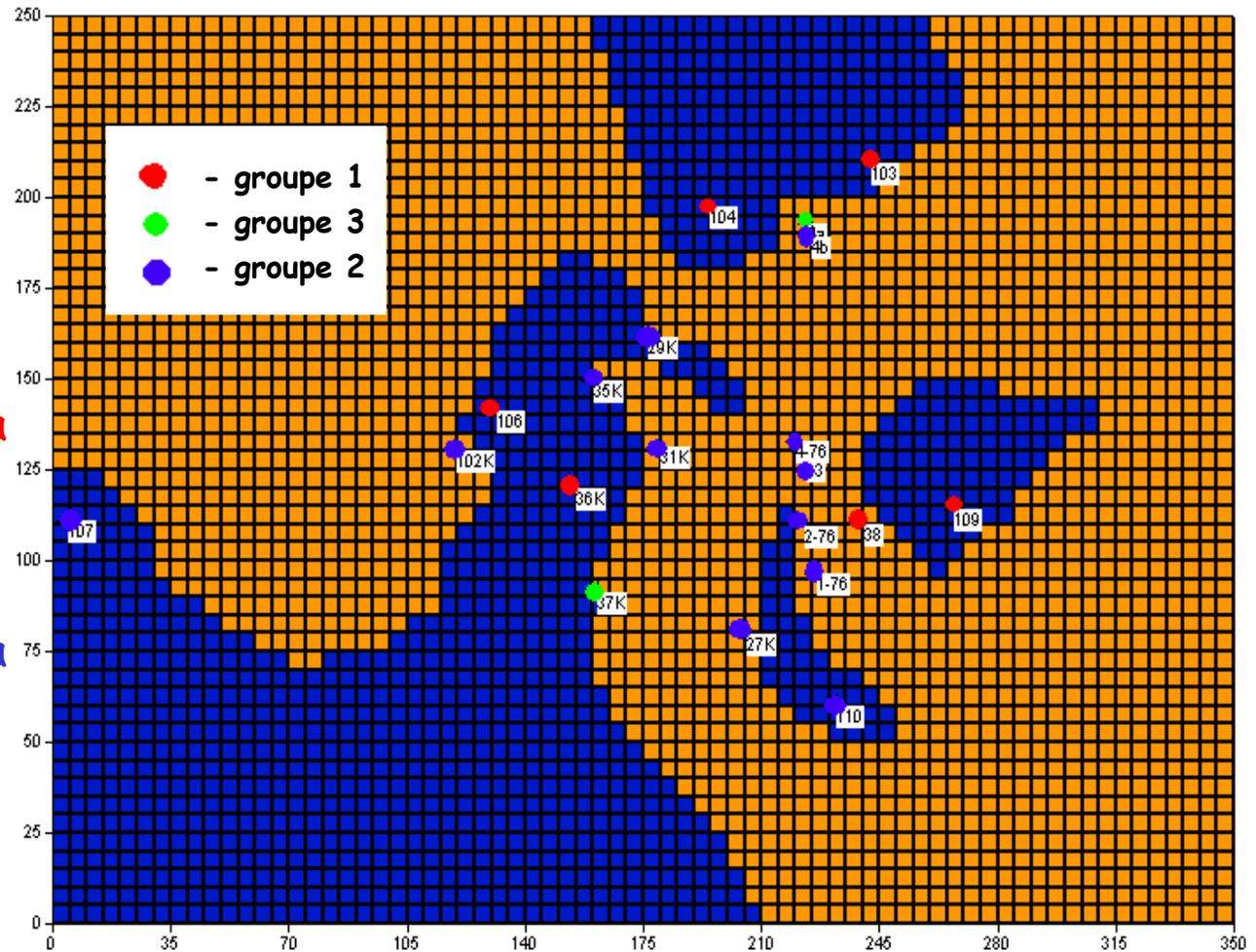
- Construction du PG : estimation des paramètres du PG par max. de vraisemblance
 En grande dimension ($p > 5$), procédure d'optimisation coûteuse (+ algo de sélection)

Sorties scalaires – Résultats de l'analyse de sensibilité



Entrées les + influentes

- **Groupe 1 : kd1**
(coef. de partage de la couche 1)
- **Groupe 2 : kd2**
(coef. de partage de la couche 2)
- **Groupe 3 : i3**
(intensité d'infiltration)



- **Pas d'interaction entre les variables d'entrée ($S_i \sim ST_i$)**

Plan

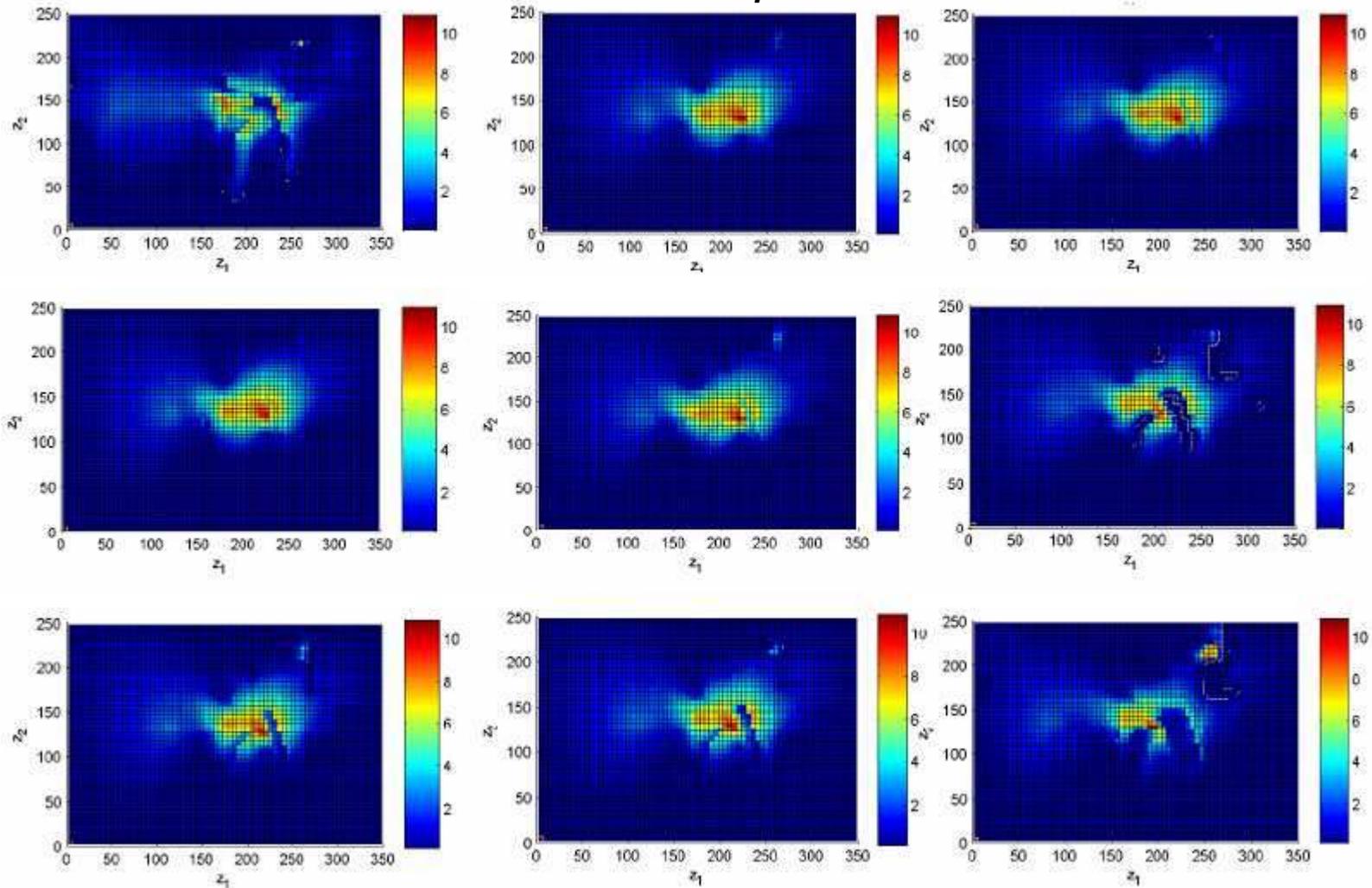


- Analyse de sensibilité et métamodèle
- **Métamodèle pour sortie spatiale**
- Analyse de sensibilité globale pour sortie spatiale
- Conclusions

Sortie spatiale

Quelques exemples de cartes obtenues en sortie du calcul

Concentration en ^{90}Sr prédite en 2010



Méthodologies pour sorties fonctionnelles

Sortie spatiale de 4096 pixels \implies sortie fonctionnelle 2D



Traitement d'une sortie fonctionnelle d'un code de calcul

- Utilisation de la discrétisation complète de la fonction

Construction d'un métamodèle puis analyse de sensibilité en chaque point de discrétisation

\implies Possible mais peut être très coûteux en fonction du métamodèle

\implies Synthèse de l'information ou isolement de l'info principale

- Remplacer la fonction par quelques paramètres d'intérêt (valeur finale, max, moyenne, ...) \implies Exploitation réduite

- Décomposition dans une base fonctionnelle (Fourier, splines, ondelettes,...)

Contexte de code de calcul : Bayarri et al., 2007, Shi et al., 2007, Higdon et al., 2008

Krigeage de courbes : Nerini et al., 2009, ...

Problématique spécifique

[Marrel et al. 09]

- Code de calcul $f(.)$:

Entrée : $\mathbf{X} = (X_1, \dots, X_p)$ vecteur aléatoire

Sortie pour une entrée \mathbf{x}^* : $y = f(\mathbf{x}^*, \mathbf{z})$, $\mathbf{z} \in D_z$ ensemble de \mathbb{R}^2



En pratique, D_z est discrétisé en n_z points (ici $64 \times 64 = 4096$ points)

- **Décomposition dans une base de fonctions orthogonales**
Par ex. base d'ondelettes (adaptée si discontinuités et phénomènes locaux)
- **Modélisation de chaque coefficient de la base par un métamodèle**

- **Prédiction :**



- Analyse de sensibilité  obtention de cartes d'indices de Sobol

Problème : la construction d'un métamodèle est coûteuse (car dim. $p > 5$)

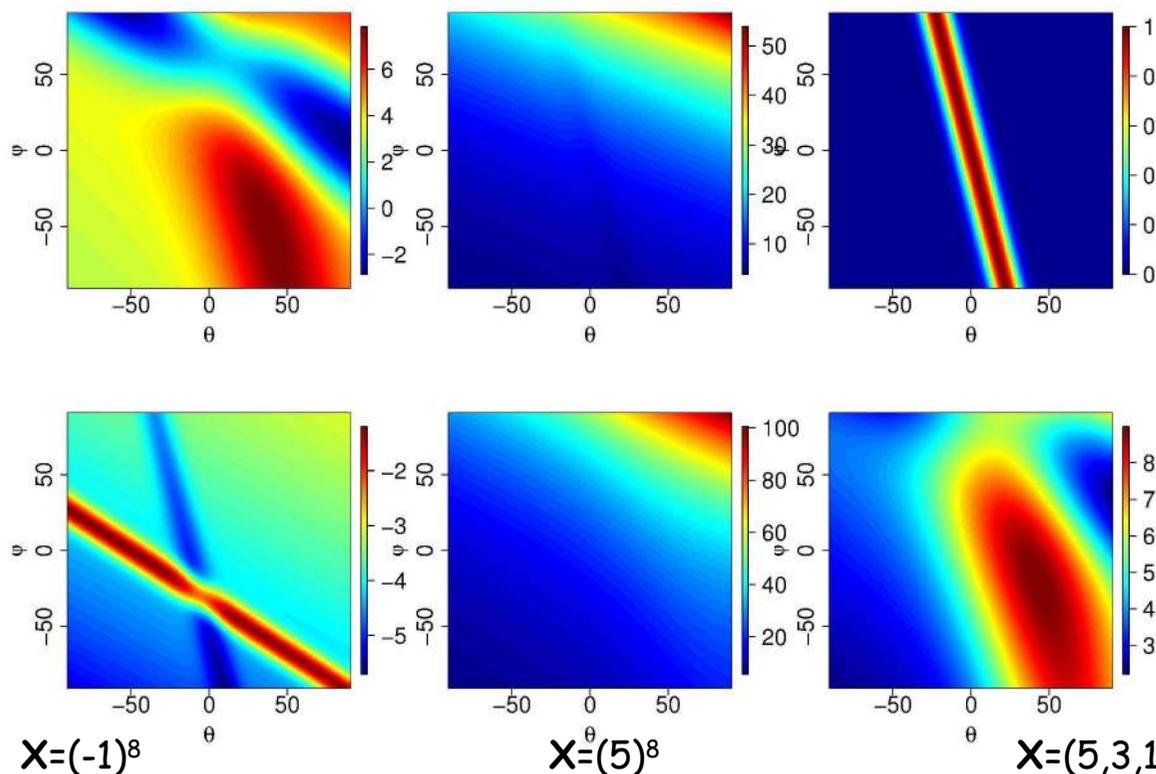
 **Sélection des coefficients à modéliser** (pour limiter leur nombre)

Une fonction « jouet » : fonction Campbell2D



$$g(X, \theta, \phi) = X_1 \exp\left[-\frac{(0.8\theta + 0.2\phi - 10X_2)^2}{60X_1^2}\right] + (X_2 + X_4) \exp\left[\frac{(0.5\theta + 0.5\phi)X_1}{500}\right]$$

$$+ X_5(X_3 - 2) \exp\left[-\frac{(0.4\theta + 0.6\phi - 20X_6)^2}{40X_5^2}\right] + (X_6 + X_8) \exp\left[\frac{(0.3\theta + 0.7\phi)X_7}{250}\right]$$



$$p = 8$$

$X_i \sim U[-1, 5]$
pour $i = 1, \dots, 8$

$$z = (\theta, \phi)$$

$$D_z = [-90, 90]^2$$

$$n_z = 64 \times 64$$

$$= 4096 \text{ points}$$

Etape 1 : décomposition spatiale & tri des coefficients

Décomposition spatiale des N cartes simulées



- Centrage des cartes (moyenne empirique sur N cartes)

$$\mu(\mathbf{z}) = E_{\mathbf{X}} [Y(\mathbf{X}, \mathbf{z})]$$

- Décomposition sur une **base d'ondelettes de Daubechies** ($K= 4096$)

$$Y_K(\mathbf{X}, \mathbf{z}) = \mu(\mathbf{z}) + \sum_{j=1}^K \alpha_j(\mathbf{X}) \phi_j(\mathbf{z})$$

- Tri des coefficients par variabilité décroissante (variance / \mathbf{X})

$$\{ \alpha_1, \dots, \alpha_K \} \longrightarrow \{ \alpha_{(1)}, \dots, \alpha_{(K)} \}$$

Etape 2 : modélisation des coefficients

Modélisation des K coefficients $\alpha_j(\mathbf{X})$ en fonction de \mathbf{X}



3 modèles :

- | | |
|---------------------------------|-----|
| ➤ Constante (moyenne empirique) | Moy |
| ➤ Régression linéaire | RL |
| ➤ Processus Gaussien | PG |

3 méthodes :

1. PG pour les k premiers coef., Moy pour le reste
2. RL pour les k premiers coef., Moy pour le reste
3. PG pour les k premiers coef., RL pour k' coef suivants, Moy pour le reste

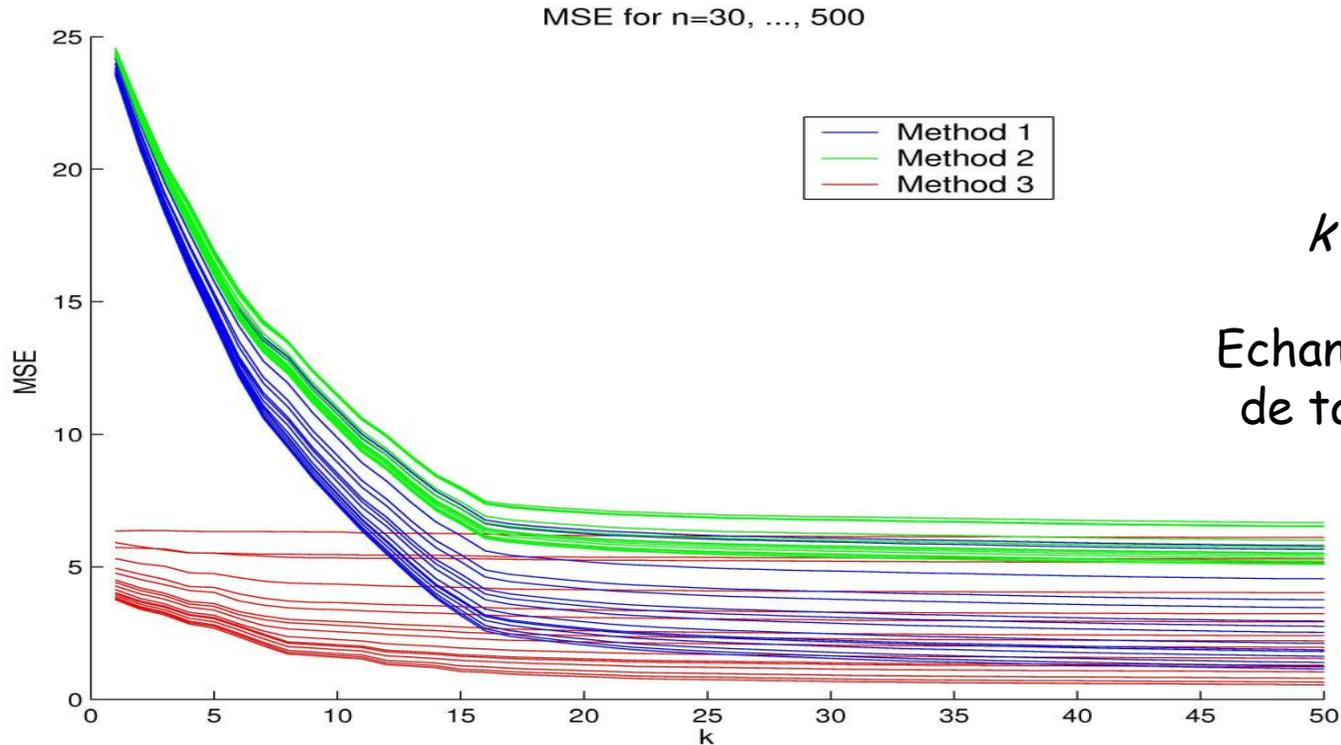
Mesures de qualité :

$$\text{MSE}(\mathbf{X}) = \int_{D_z} [Y(\mathbf{X}, \mathbf{z}) - \hat{Y}_{K,k}(\mathbf{X}, \mathbf{z})]^2 d\mathbf{z} \text{ avec } \hat{Y}_{K,k}(\mathbf{X}, \mathbf{z}) \text{ approximation de } Y_K(\mathbf{X}, \mathbf{z})$$

$$\text{MSE} = E_{\mathbf{X}}[\text{MSE}(\mathbf{X})] \quad ; \quad Q_2 = 1 - \frac{\text{MSE}}{E_z\{\text{Var}_{\mathbf{X}}[Y(\mathbf{X}, \mathbf{z})]\}}$$

Etape 3 : sélection du nombre de coefficients PG

Fonction Campbell2D : Visualisation du MSE en fct de k et de N



➡ **Méthode 3 avec $k = 30$ (stabilisation du MSE avérée)**

$Q_2 = 93\%$ pour $N = 200$; $Q_2 = 97\%$ pour $N = 500$

Plan



1. Analyse de sensibilité et métamodèle

2. Métamodèle pour sortie spatiale

3. Analyse de sensibilité globale pour sortie spatiale

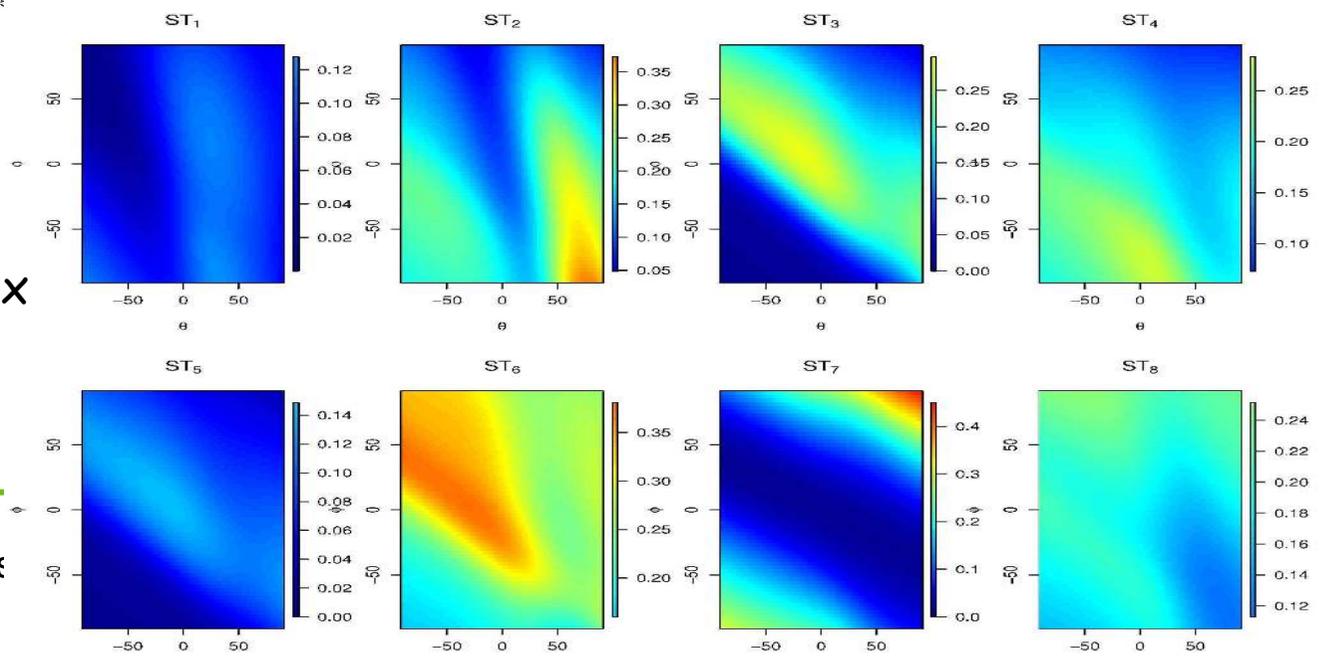
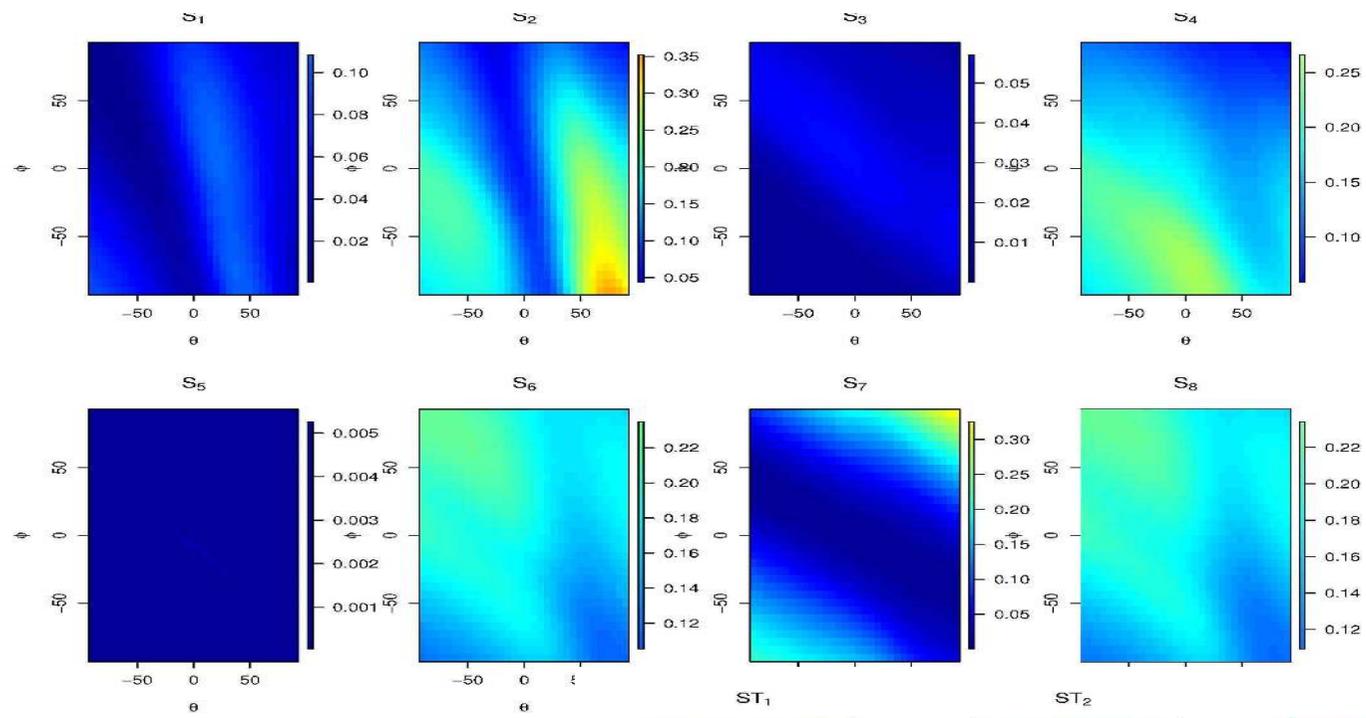
4. Conclusions

Indices de Sobol exacts sur Campbell2D : 1^{er} ordre et total

$$X_i \sim U[-1,5]$$

$$i=1...8$$

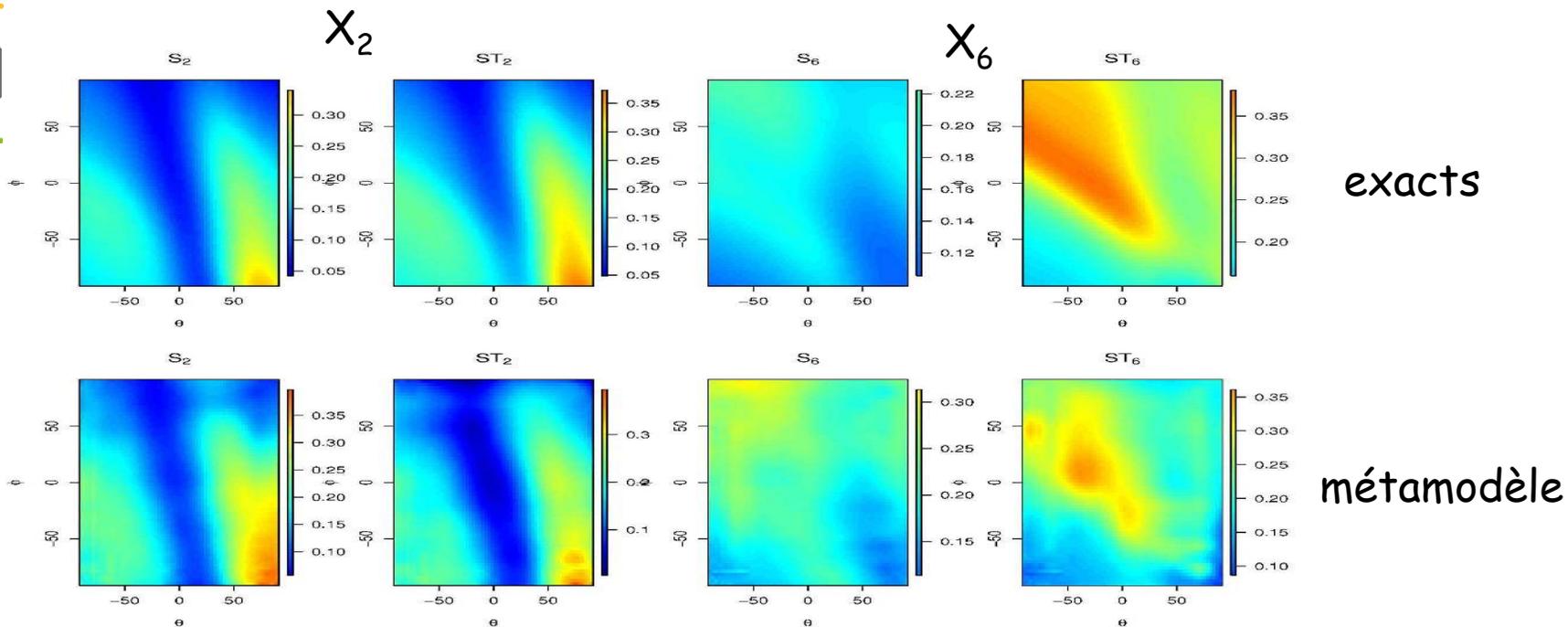
Indices 1^{er} ordre
Formules analytiques



Indices totaux
Monte Carlo

Résultats sur Campbell2D

Estimation des indices de Sobol du 1^{er} ordre et ordre total par Monte Carlo sur le métamodèle fonctionnel



Erreurs absolues moyennes relatives sur les indices du 1^{er} ordre :

X_1	X_2	X_3	X_4	X_6	X_7	X_8
9%	16%	16%	13%	13%	12%	10%

Sortie spatiale – Analyse de sensibilité sur Kurchatov (1/2)

$N = 300$ simulations



$p = 20$ variables d'entrée

$K = 4096$ pixels

$k = 100$ coef. modélisés par PG

$k' = 900$ coef. modélisés par régression linéaire

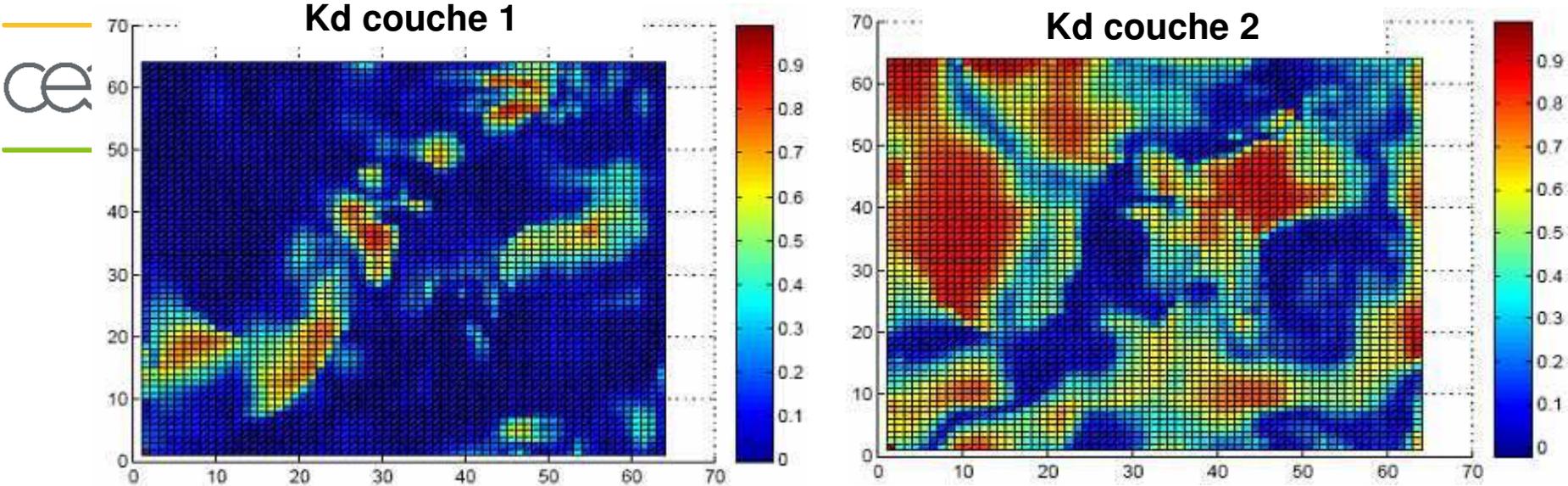
Prédictivité moyenne : $Q_2 = 72\%$

Estimation des indices de Sobol du 1^{er} ordre et ordre total par Monte Carlo (22000 évaluations du métamodèle fonctionnel)

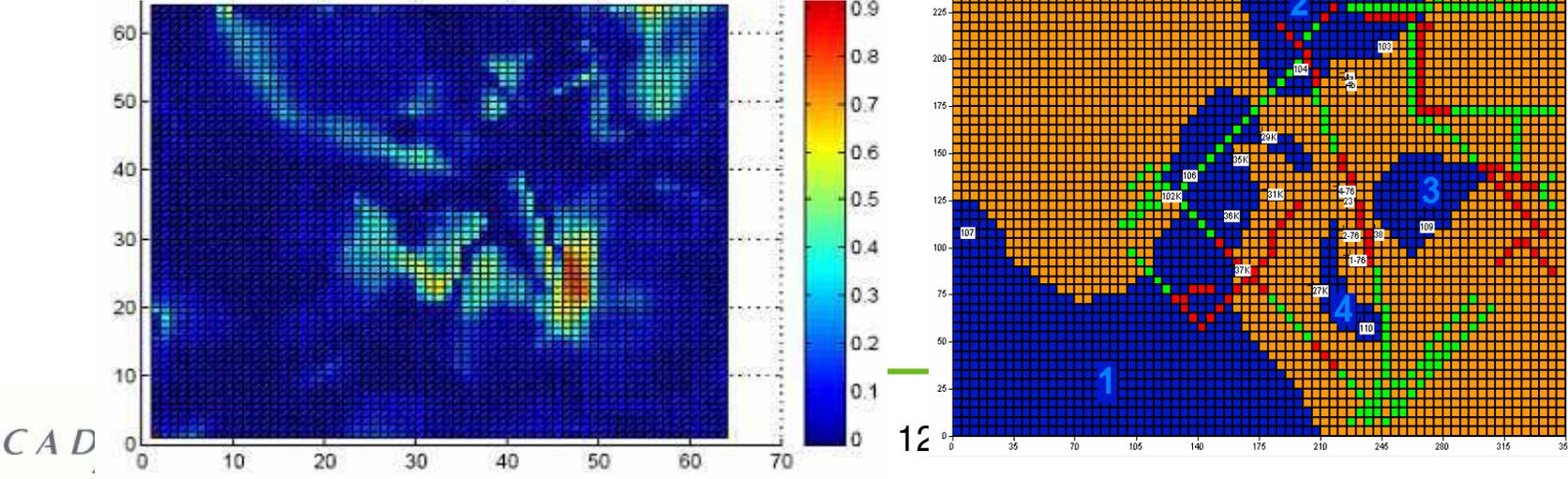
 20 cartes d'indices de sensibilité

Sortie spatiale – Analyse de sensibilité sur Kurchatov (2/2)

Cartes d'indices de Sobol (1^{er} ordre) pour les entrées kd1, kd2, i3



Infiltration forte (canalisations rouges)



Plan



1. Analyse de sensibilité et métamodèle

2. Métamodèle pour sortie spatiale

3. Analyse de sensibilité globale pour sortie spatiale

4. Conclusions

Conclusions principales de l'étude du site de Kurchatov

1. Les prédictions de la contamination au bord du modèle ne dépassent pas le seuil réglementaire (jusqu'en 2010)



2. Les coefficients de partage sont les paramètres les plus influents et interagissent peu avec les autres paramètres

Réduction de leur incertitudes → *réduction de l'incertitude de prédiction*

3. L'influence de la forme de la 2^{ème} couche est forte. On peut l'étudier en faisant varier la forme par simulation géostatistique

→ Analyse de sensibilité avec champs aléatoires en entrée

4. Si on veut réduire l'incertitude de la prédiction au bord du domaine, le paramètre hydrogéologique à re-caractériser en priorité est le Kd de la 2^{ème} couche

Conclusions et perspectives

■ Proposition d'un métamodèle fonctionnel

- Décomposition en ondelettes
- Modélisation par krigeage des coefficients sélectionnés
- Obtention de cartes d'indices de sensibilité
- Interprétation globale et locale, aide à la décision

■ Perspectives

- Extension de l'exploitation du métamodèle fonctionnel
- Sortie 3D voire spatio-temporelle



Méthodologie applicable à d'autres problématiques
Ex : CFD, rejet de polluants atmosphériques, séquestration du CO2

Références

- Volkova, Iooss & Van Dorpe, *Stoch. Env. Res. Risk Asses.*, 2008
- Marrel, *Thèse de l'INSA Toulouse*, 2008
- Marrel, Iooss, Jullien, Laurent & Volkova, *Spatial global sensitivity analysis*, soumis, dispo sur HAL