



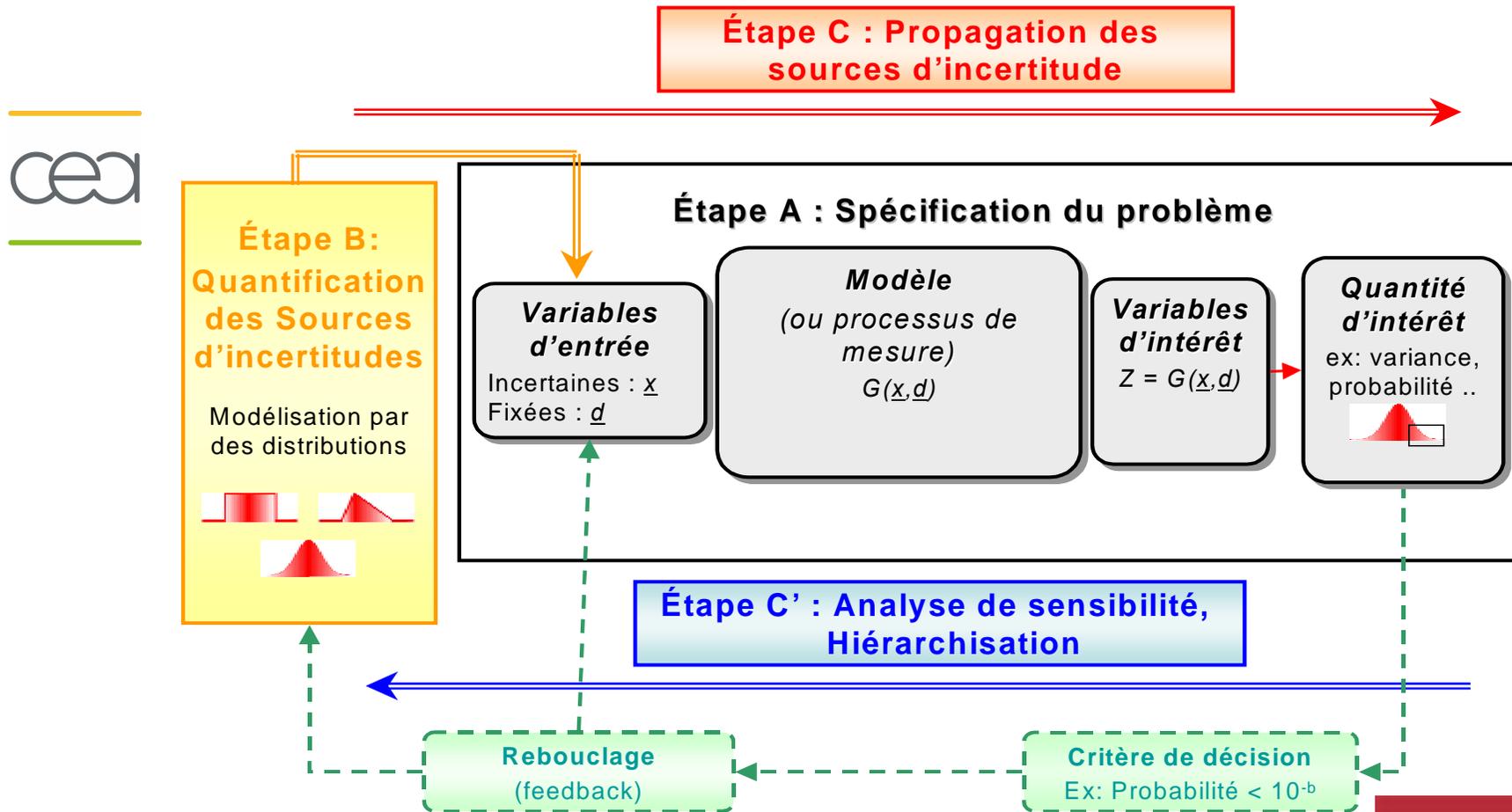
Construction et utilisation de métamodèles

Bertrand looss

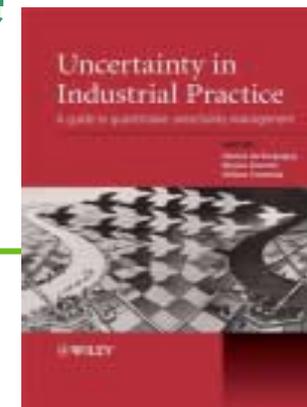
CEA Cadarache
Direction de l'Énergie Nucléaire

2009

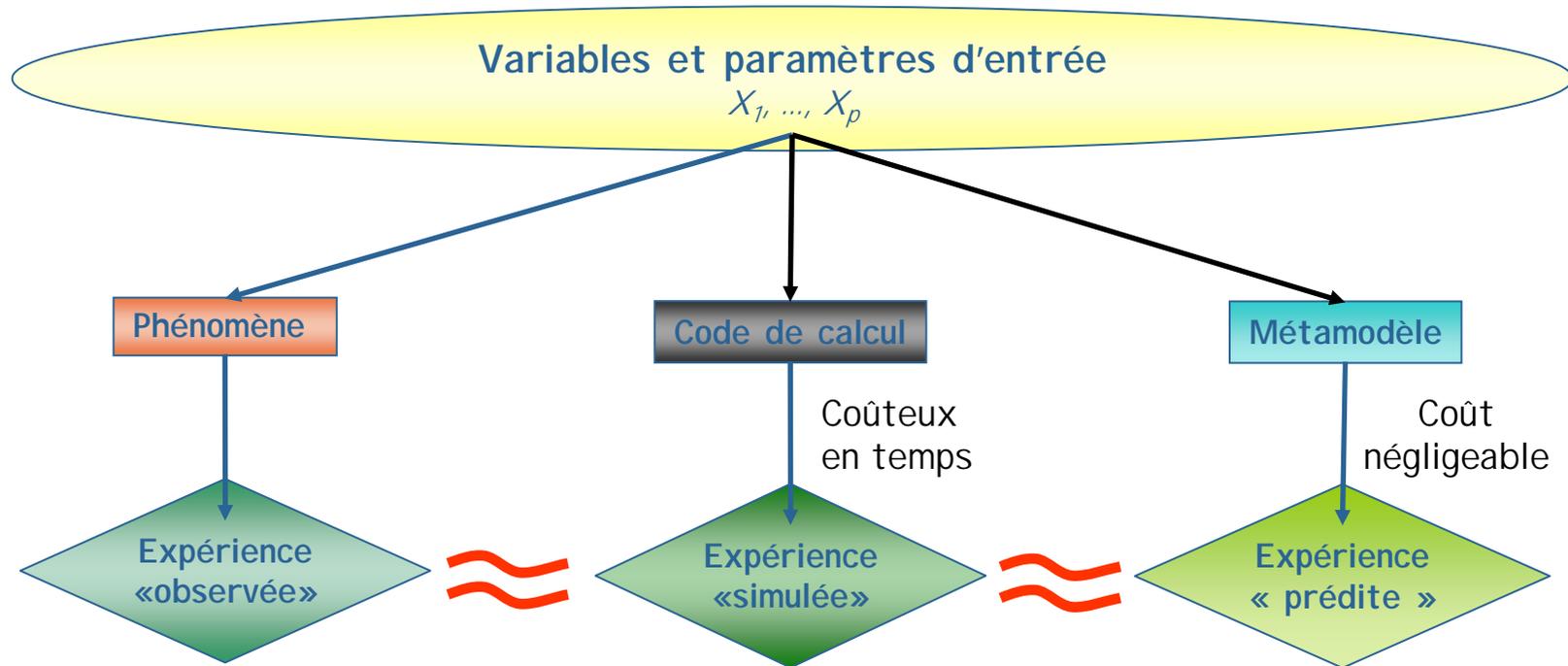
Schéma générique d'une étude d'incertitude sur un code



Dans la suite : $Y = f(\mathbf{X})$, $\mathbf{X} \in \mathbb{R}^p$, $Y \in \mathbb{R}$ [de Rocquigny et al. 08]

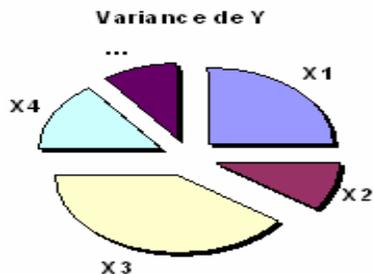


Un outil multifonctionnel : le métamodèle

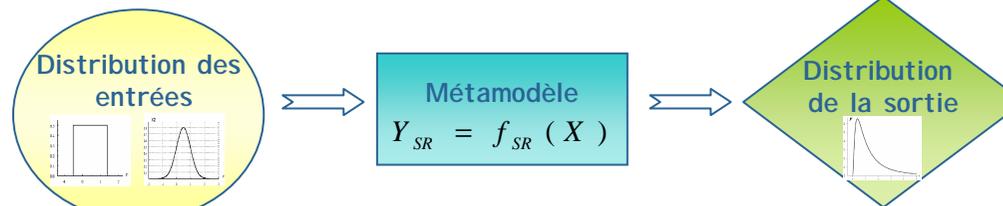


Exploitation du métamodèle :

■ Analyse de sensibilité



■ Propagation d'incertitudes



• Calibration

Détermination des paramètres

Adéquation expériences simulées et observées

The calibration process involves the 'Détermination des paramètres' (determination of parameters), which leads to the 'Adéquation expériences simulées et observées' (adequacy of simulated and observed experiences).

Définitions

[Kleijnen 70's]



Un métamodèle est une fonction mathématique :

- approximant les réponses du modèle étudié,
- de coût négligeable,
- permettant de prédire avec une bonne précision de nouvelles réponses

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

krigeage (processus gaussien - PG)

- Synonymes :
 - Surface de réponse
 - Modèle simplifié
 - Émulateur
 - Modèle proxy
 - Surrogate model

Plan



1. Exemple d'applications : scénario de transfert hydrogéologique de contaminants

2. Le métamodèle Processus Gaussien (krigeage)

3. Plans d'expériences associés

4. Panorama de différents métamodèles

5. Conclusions

Site de stockage temporaire de déchets radioactifs

Collaboration Institut Kurchatov/CEA
[Volkova et al. 08]



- Site (2 ha) en banlieue de Moscou
- De 1943 à 1974 : stockage de déchets radioactifs solides



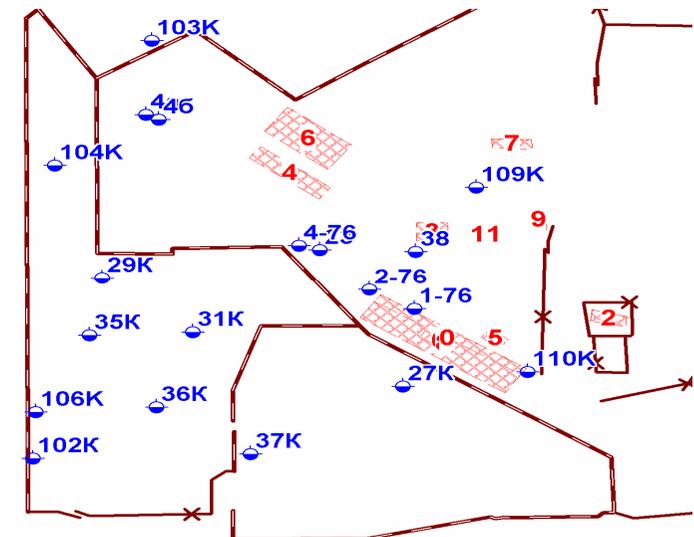
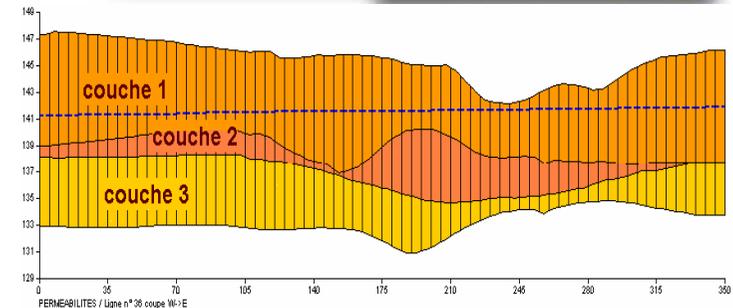
- Reconnaissance du site vers 1990 : réseau de 20 piézomètres
- Contamination nappe supérieure ^{90}Sr

Estimation de l'impact de la contamination sur l'environnement

(degré de contamination de la nappe)

Faut-il réhabiliter le site ?

(excavation des déchets et traitement des sols)



Présentation du modèle du site de Kurchatov (2/2)

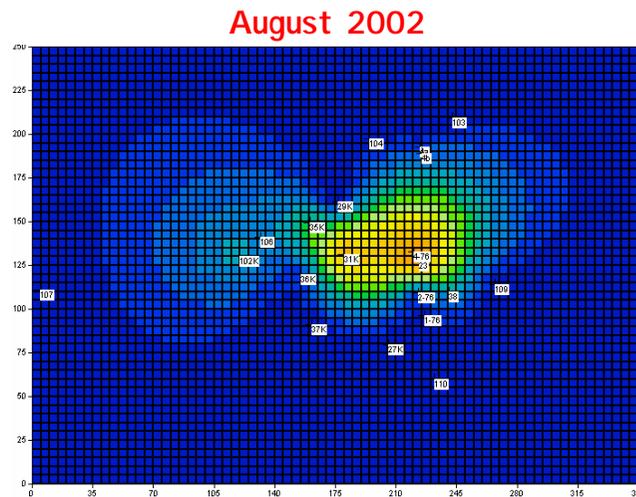
Logiciel MARTHE (BRGM) : Modélisation d'Aquifère par un maillage Rectangulaire en régime Transitoire pour le calcul Hydrodynamique des Ecoulements



250m × 350m, 200 pas de temps

Ecoulement transitoire 3D ; convection-dispersion ; sorption linéaire ; décroissance radioactive ; pas de terme source

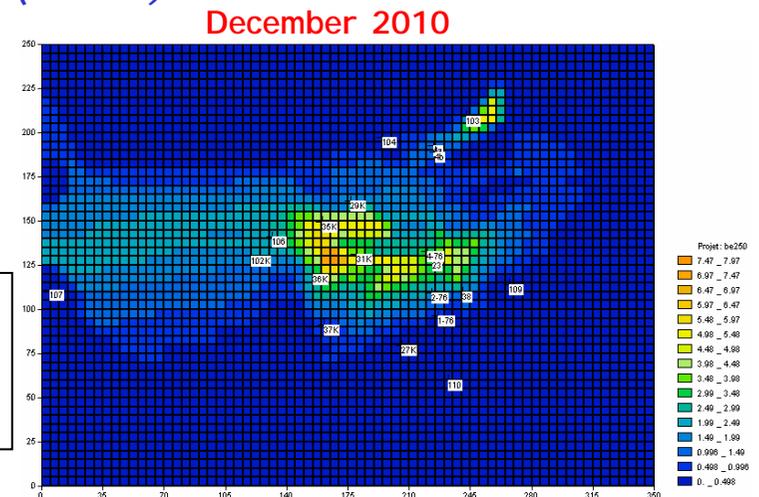
Exemple d'un calcul (20mn)



Cartes de Concentrations



+ Jeu de 20 variables d'entrée



Identification de 20 paramètres d'entrée incertains et de 20 sorties (concentrations aux 20 piézomètres)

1 sortie scalaire – Analyse de sensibilité

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

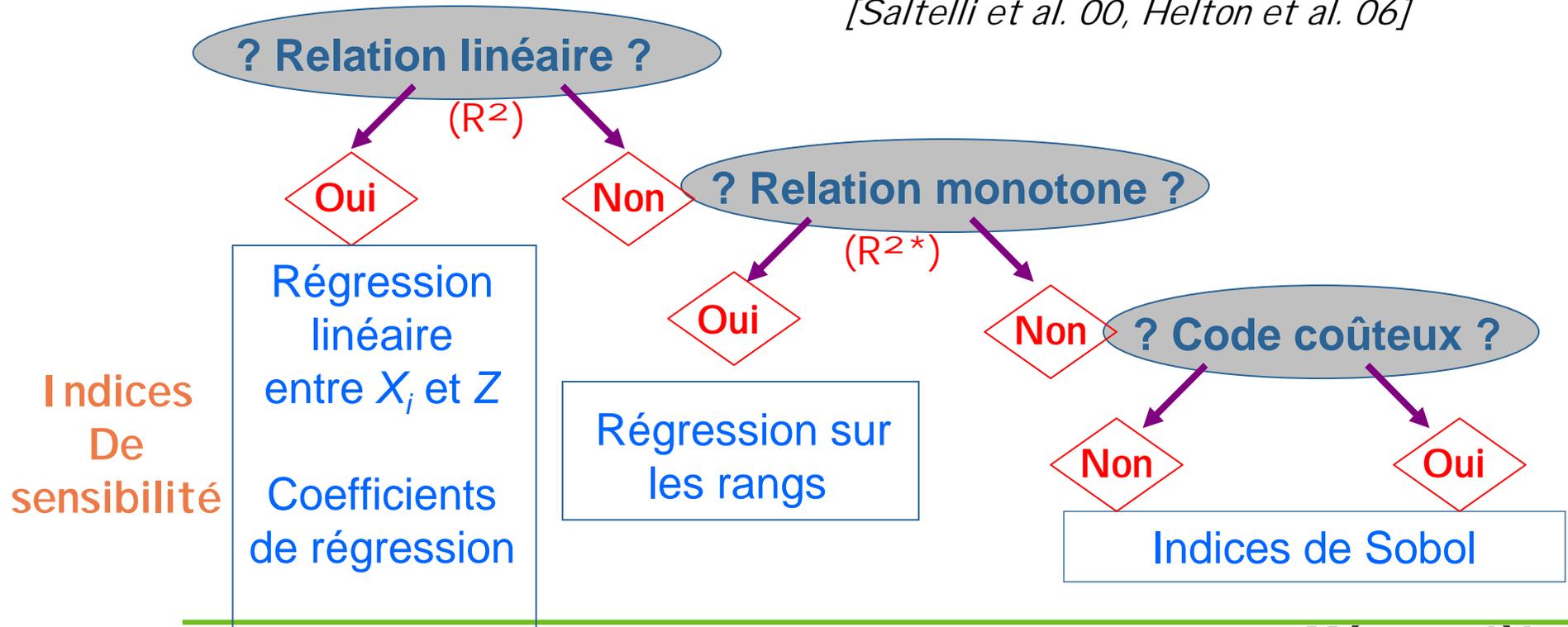
$N = 300$ calculs de type Monte-Carlo (LHS)



Étape préliminaire : visualisation graphique (par ex : scatterplots)

Méthodologie d'analyse de sensibilité quantitative

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



Métamodèle

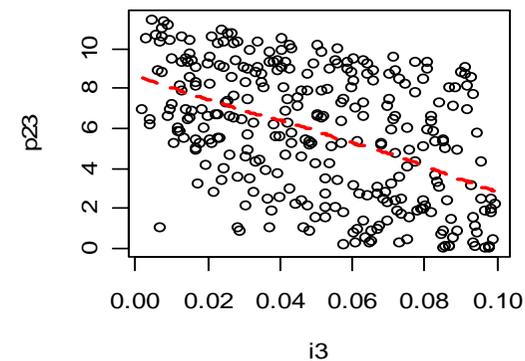
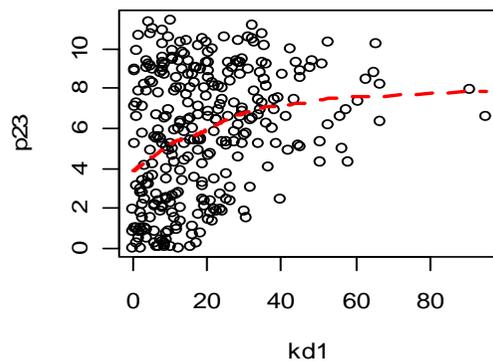
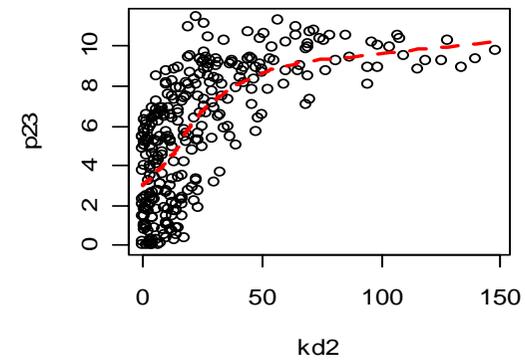
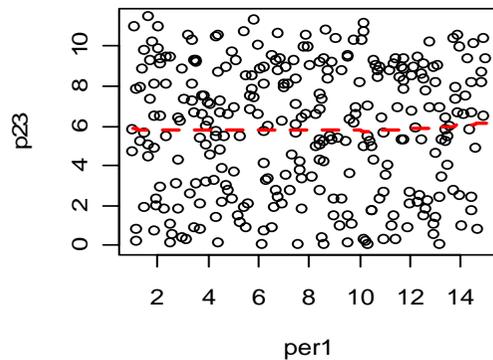
Analyse graphique : scatterplots piézomètre p23

Parfois de simples diagrammes sortie / entrées sont suffisants



1 sortie (p23) - p = 4 entrées

$N = 300$ calculs de type Monte Carlo (toutes les entrées X_i varient en même temps)



Sorties scalaires – Résultats de l'analyse de sensibilité (1/2)

Étude sur 20 variables de sortie



➤ **Relations entrées/sortie approximativement linéaires :**

p23 ($R^2 = 74\%$) ; p4-76 ($R^2 = 67\%$) ; 6 sorties de $R^2 > 50\%$

Remarque : le coef. de détermination R^2 est calculé pour la prédiction

➤ **Relations non monotones** [Marrel et al. 08]

Constructions de métamodèles par krigeage

- $R^2 > 50\%$ pour 15 variables de sortie
- **Calcul des indices de Sobol sur les métamodèles**
- 5 sorties restent à $R^2 < 50\%$ (sorties à très faible variabilités)

Indices de sensibilité pour la sortie « Piézomètre p104 »

Ajustement d'un métamodèle PG : $R^2 = 93\%$ (régression linéaire $R^2 = 68\%$)



Analyse de sensibilité basée sur le PG [Marrel et al. 09]

Estimation des indices de Sobol + construction des intervalles de prédiction

(en %)	SRC_i^2 (régression linéaire)	S_i (PG prédicteur)	$\mu_i = E_\Omega[\tilde{S}_i]$ (PG global)	IC- 90% (\tilde{S}_i) (PG global)
per1	2	8	8	[5 ; 11]
kd1	52	76	69	[56 ; 83]
I3	13	15	13	[10 ; 17]

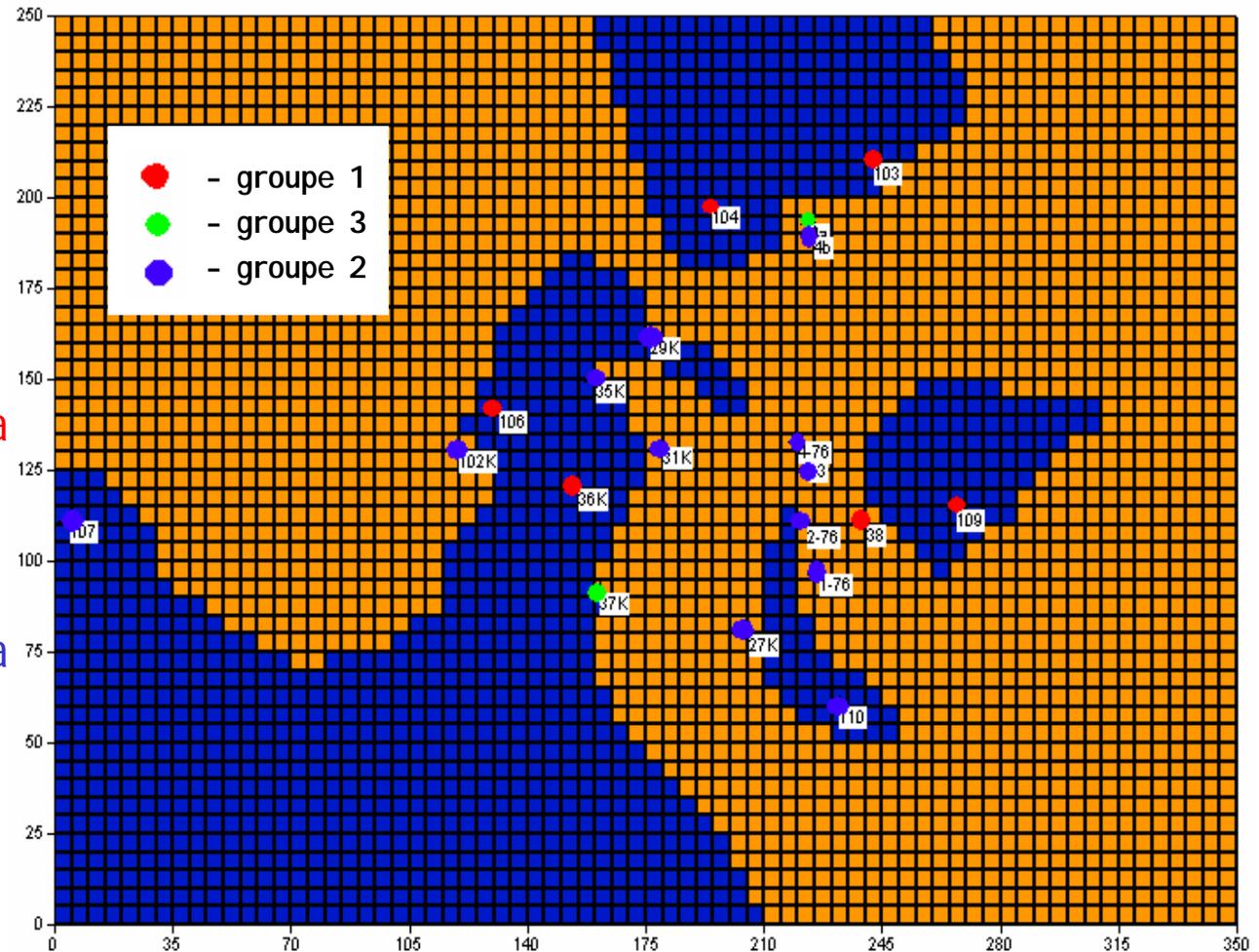
■ Le métamodèle PG permet d'avoir une estimation + fiable des indices de sensibilité qu'en utilisant la régression linéaire (SRC ou SRRC)

■ L'obtention d'IC avec le PG permet d'intégrer l'erreur résiduelle due au remplacement du code par le métamodèle

Sorties scalaires – Résultats de l'analyse de sensibilité (2/2)

 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)



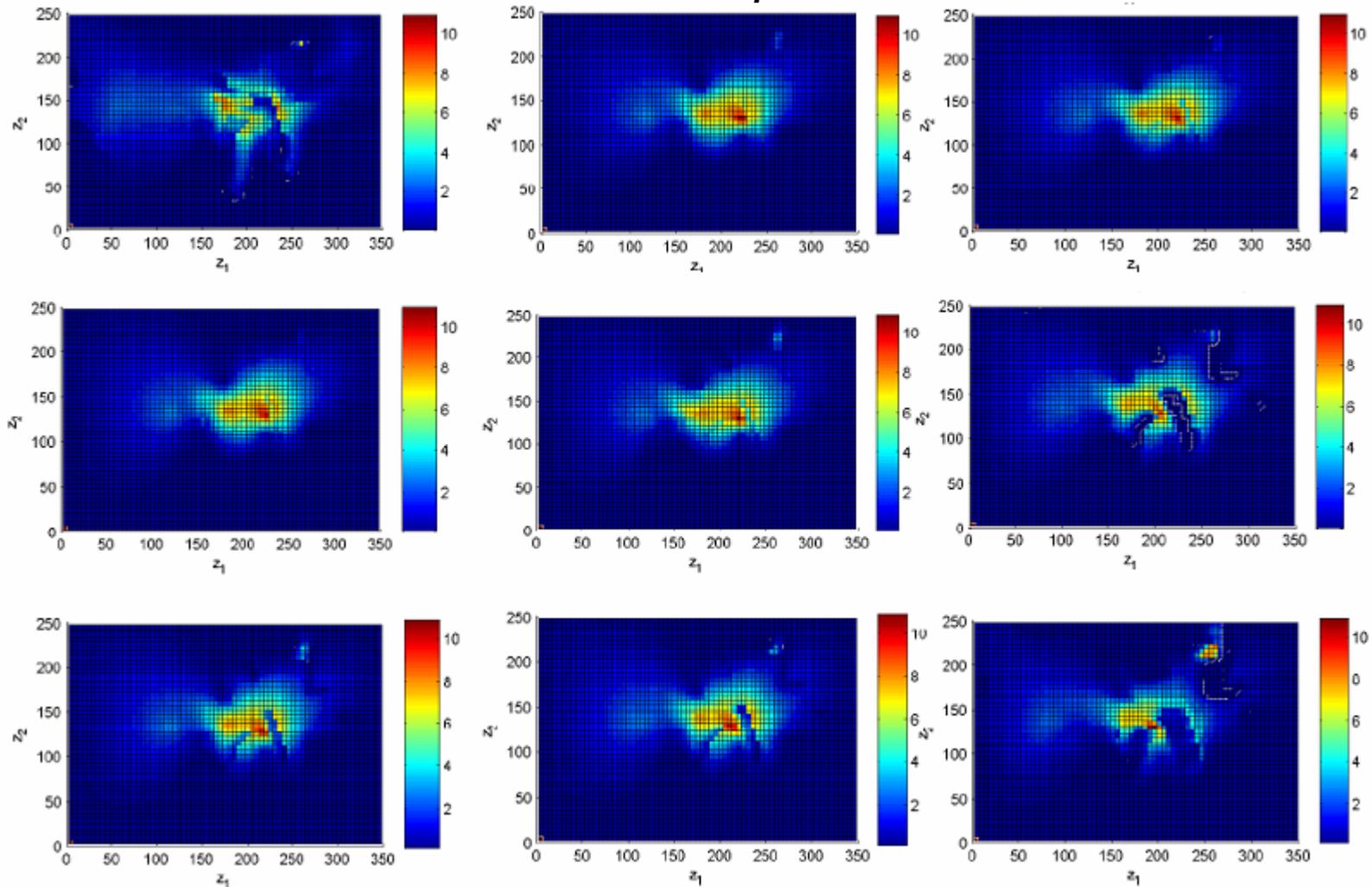
- **Pas d'interaction** entre les variables d'entrée ($S_i \sim ST_i$)

Sortie spatiale

- Quelques exemples de cartes obtenues en sortie du calcul



Concentration en ^{90}Sr prédite en 2010



Méthodologie pour sortie spatiale

[Marrel 08]

■ Etape 1 : Décomposition spatiale des $N = 300$ cartes

- Centrage des cartes
- Décomposition sur une base d'ondelettes
- Tri des coefficients par valeur moyenne en norme L_2

■ Etape 2 : Modélisation des coefficients en fonction de X

- Modélisation des 100 premiers coefficients par métamodèle (krigeage)
- Modélisation des coefficients 101 à 1000 par régression linéaire simple (avec sélection par AIC)

■ Etape 3 : Prédiction pour un nouveau jeu d'entrée x^*

$x^* \Rightarrow$ prediction des coefficients \Rightarrow reconstitution de la carte

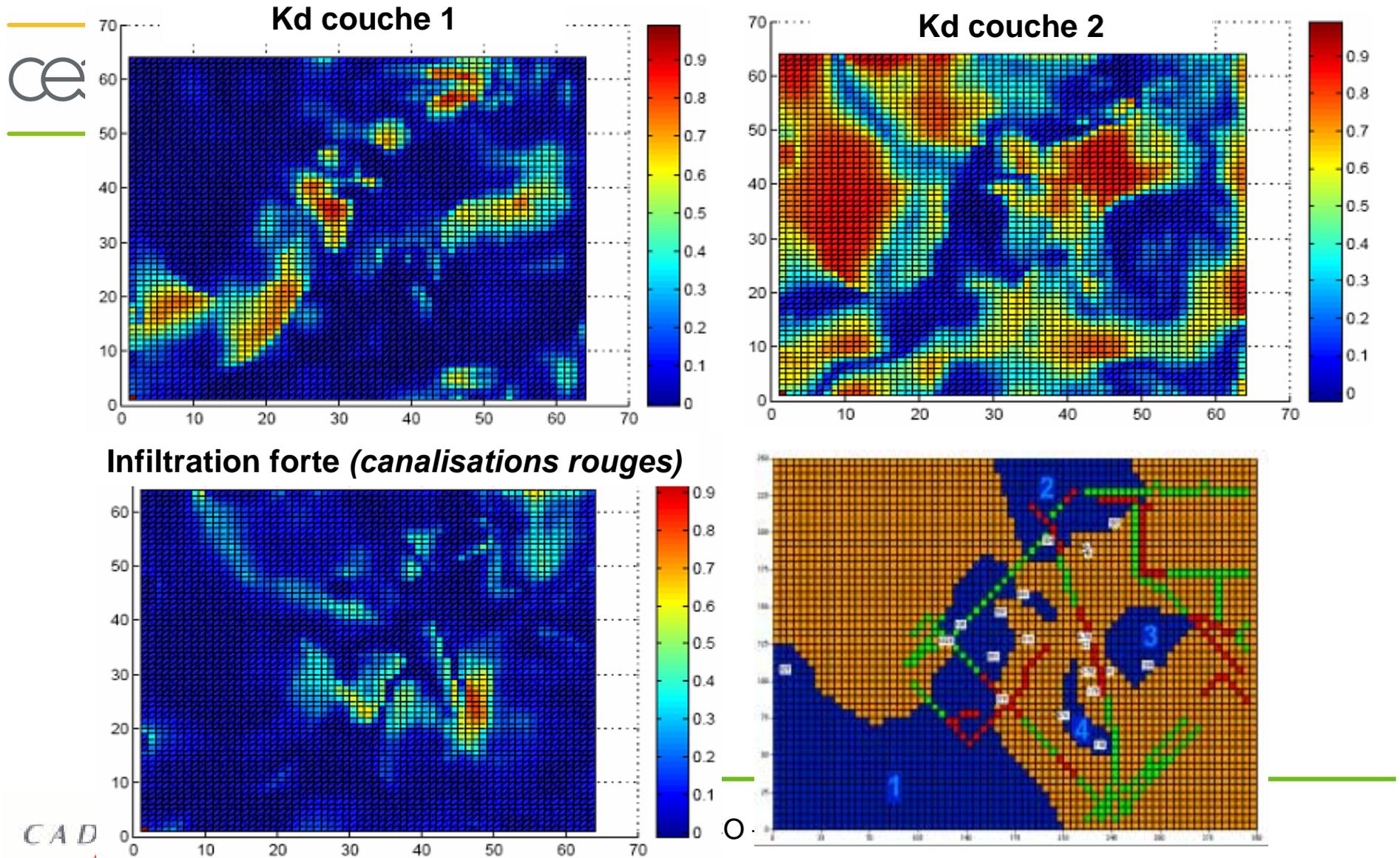


Analyse de sensibilité :

Obtention de cartes spatiales d'indices de Sobol

Sortie spatiale – Résultats de l'analyse de sensibilité

● Cartes d'indices de Sobol pour les entrées kd1, kd2 et i3



Plan



1. Exemple d'applications : scénario de transfert hydrogéologique de contaminants

2. Le métamodèle Processus Gaussien (krigeage)

3. Plans d'expériences associés

4. Panorama de différents métamodèles

5. Conclusions

Métamodèle du krigage

L'idée du krigage [Matheron 63] pour les codes est d'interpoler les réponses du code [Sacks et al. 89] comme pour une cartographie spatiale



Métamodèle intéressant car :

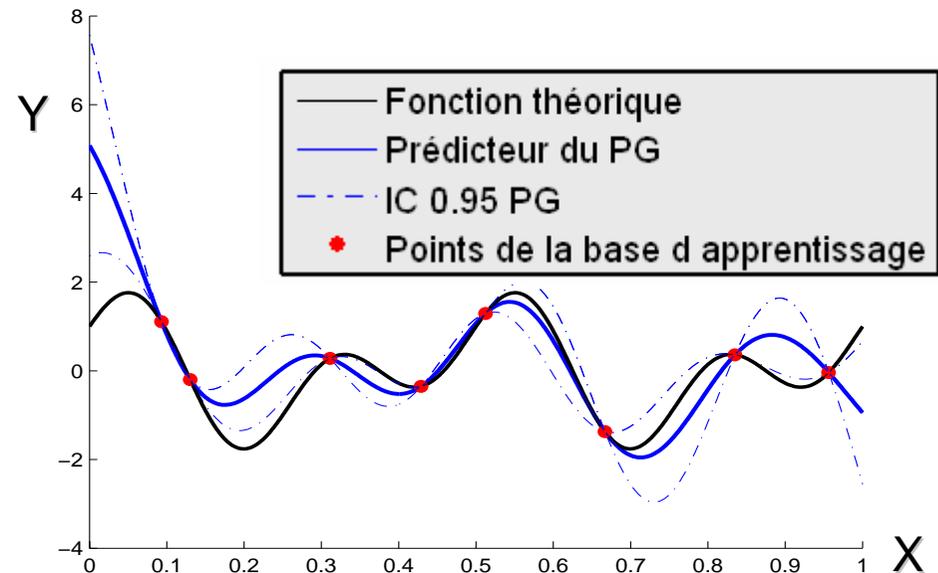
- interpole les réponses (pas d'hypothèse sur le bruit de mesure),
- évalue une nouvelle réponse très rapidement,
- fournit en plus d'une prédiction, une estimation de son erreur

Exemple en 1D :

Fonction théorique ($\rho=1$) :

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

Simulation de $N = 8$ calculs

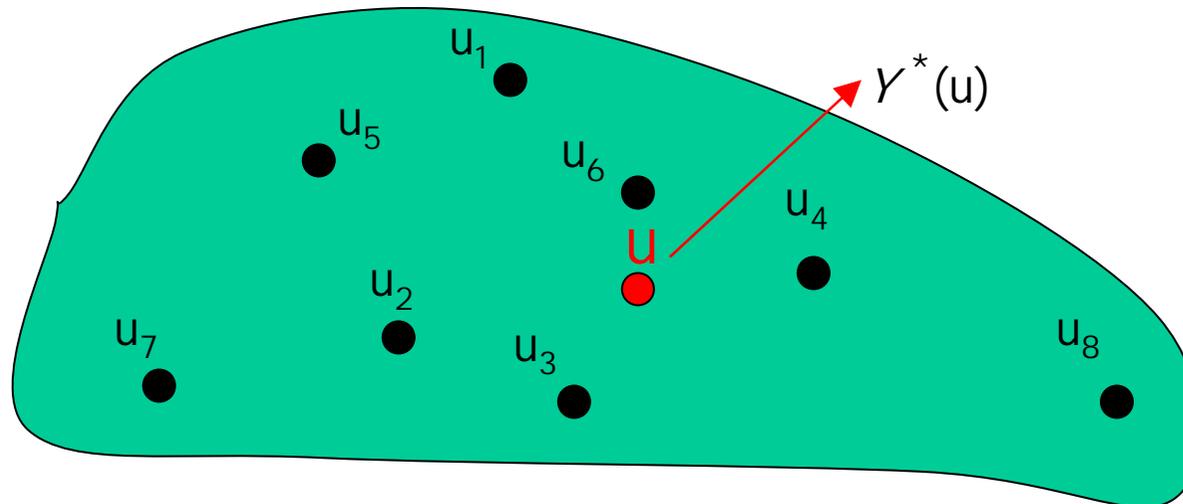


Le krigeage spatial

Combinaison linéaire des N données : $Y^*(u) = \sum_{i=1}^N \lambda_i Y(u_i)$



Prise en compte de la configuration des données, de la distance entre données et cible, des corrélations spatiales, d'informations externes, ...



Estimateur sans biais : $E [Y^*(u) - Y(u)] = 0$
la moyenne des erreurs est nulle.

$Y^*(u)$ est optimal : $\text{Var} [Y^*(u) - Y(u)]$ est minimale
la dispersion des erreurs est réduite.

Modélisation stochastique de $Y(x)$

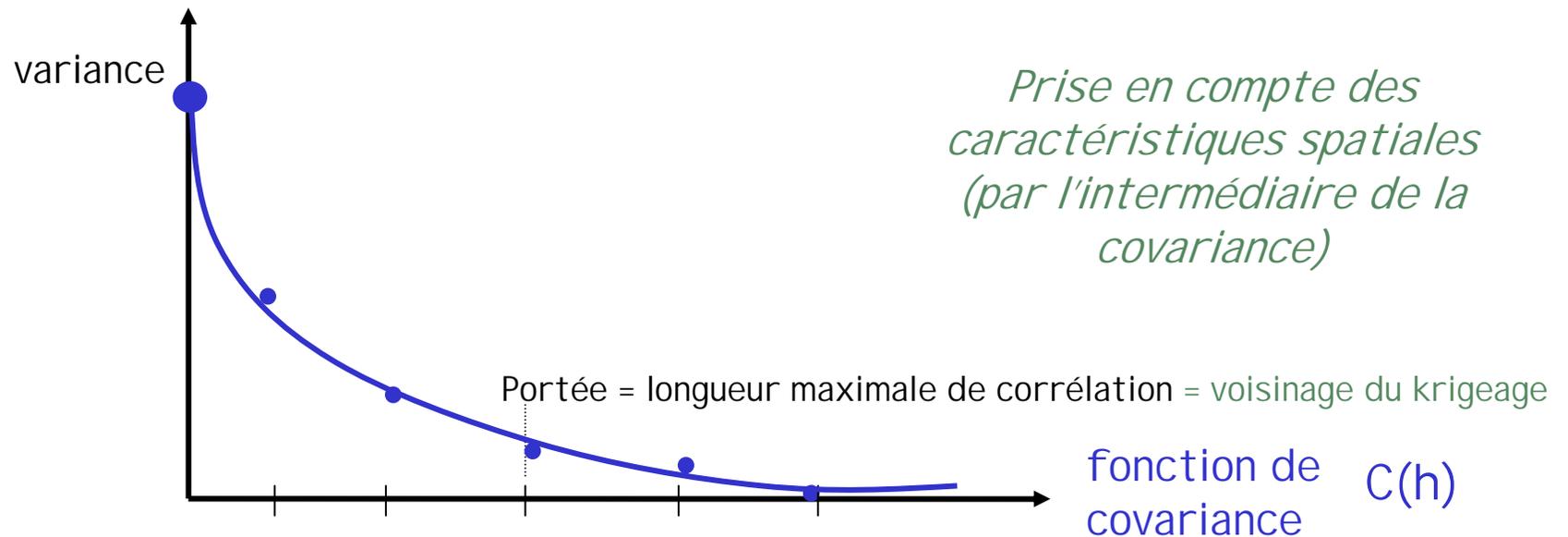
Le champ aléatoire $Y(x)$ est caractérisé par sa moyenne et sa covariance.



$Y(x)$ est stationnaire d'ordre 2 : $E[Y(x)] = m$ ne dépend pas de x

Covariance : corrélation normalisée

$$\text{Cov}[Y(x), Y(x+h)] = E[Y(x+h)Y(x)] - E[Y(x+h)]E[Y(x)] = C(h) \text{ ne dépend pas de } x$$



Le krigeage simple (moyenne connue)

$$Y^*(u) = \sum_{i=1}^N \lambda_i(u) [Y(u_i) - m] + m$$

(m constante et connue)



Min { E [Y*(u) - Y(u)]² }

λ_i  régression linéaire multiple (à résidus corrélés) par moindres carrés.

Les poids de krigeage $\lambda_i(u)$ associés à $Y(u_i)$ sont déterminés par :

$$\left\{ \sum_{j=1}^N \lambda_j(u) C(u_i - u_j) = C(u_i - u) \quad \forall i = 1 \dots N \right.$$

Système de N équations linéaires à N inconnues qui possède une unique solution (si la matrice est non singulière)

Variance du krigeage (erreur d'estimation) :

$$\sigma_K^2(u) = C(0) - \sum_{i=1}^N \lambda_i(u) C(u_i - u)$$

On peut donc visualiser les régions où l'estimation est imprécise et où il conviendrait de placer des nouveaux points de mesure

Métamodèle Processus Gaussien (PG)

Idee : le code est la réalisation d'un champ aléatoire gaussien



Definition :

Processus Gaussien défini sur $R^p \times \Omega$

$$\text{➤ } Y(\mathbf{x}, \omega) = G(\mathbf{x}) + Z(\mathbf{x}, \omega)$$

Regression **Partie stochastique**

Z processus stochastique avec :

$$E_{\Omega}[Z(x)] = 0$$

$$\text{Cov}_{\Omega}(Z(x), Z(u)) = \sigma^2 R(x, u)$$

où σ^2 est la variance

et R la fonction de corrélation

$$Z \sim N(0, \sigma^2 R)$$

Exemples de choix paramétriques :

- G : polynôme de degré 1 $G(\mathbf{x}) = \beta_0 + \sum_{i=1}^p \beta_i x_i$
- Z : processus stationnaire avec covariance exponentielle généralisée

$$R(\mathbf{x}, \mathbf{u}) = R(\mathbf{x} - \mathbf{u}) = \exp\left(-\sum_{i=1}^p \theta_i |x_i - u_i|^{q_i}\right)$$

Loi jointe et loi conditionnelle

• Loi jointe :

- Modèle PG : $Y(X) = \beta F(X) + Z(X)$

- Base d'apprentissage (BA) $X_S = [x_1, \dots, x_N]$, Y_S , $F = F(X_S)$, $R = (R(x_i, x_k))_{i,k}$

- Loi jointe sur la BA : $Y_S \sim \mathcal{N}(\beta F, \sigma^2 R)$

- Loi conditionnelle pour un nouveau point x^* :

→ $Y(x^*)_{|X_S, Y_S} \sim \mathcal{N}(\mu(x^*), \tilde{\sigma}^2(x^*))$

$\mu(x^*) = \beta F(x^*) + r(x^*) R^{-1} [Y - \beta F]$ avec $r(x^*) = [R(x_1, x^*), \dots, R(x_N, x^*)]$

$\tilde{\sigma}^2(x^*) = \sigma^2 (1 + {}^t u(x^*) ({}^t F R^{-1} F)^{-1} u(x^*) - {}^t r(x^*) R^{-1} r(x^*))$

avec $u(x^*) = {}^t F R^{-1} r(x^*) - f(x^*)$

• Prédicteur et erreur associée :

- Meilleur prédicteur linéaire sans biais (BLUP) : $\hat{Y}(x^*) = \mu(x^*)$

- Interpolateur exact des points de la BA

- Formulation de l'erreur : $MSE[\hat{Y}(x^*)] = \tilde{\sigma}^2(x^*)$

Construction du métamodèle PG

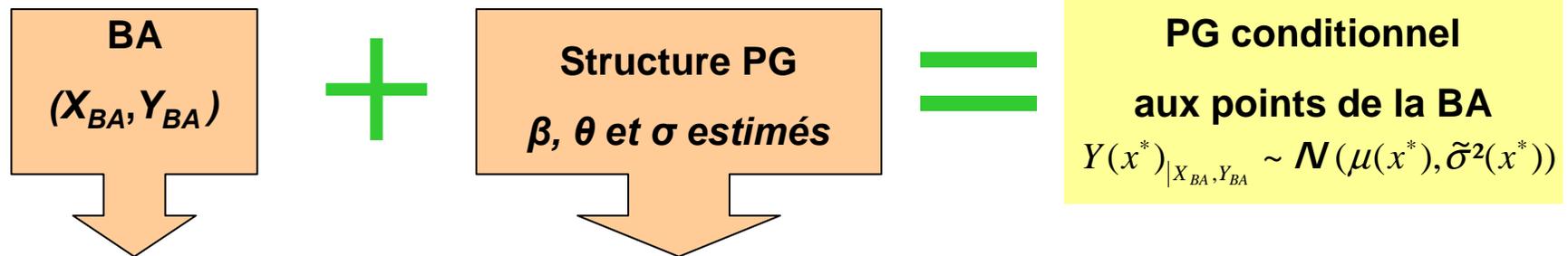
Estimation des paramètres de régression et de corrélation β, θ, σ



- Estimation sur la BA
- Maximum de vraisemblance + algorithme d'optimisation (par ex. stoch)
- Maximisation du R^2 sur une base de test pour θ

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

Formulation et compréhension du modèle :



- Résultat des simulations du code
- Points « d'ancrage »
- Points Interpolés exactement
- Sélection optimale de ces points
- Structure du réseau, de la « toile » 
- Estimation des paramètres par maximum de vraisemblance sur les points de la BA
- Sélection des paramètres β et θ significatifs

Bilan sur le métamodèle PG

- Difficultés : *Estimation en grande dimension, inversion de R*
On y arrive généralement quand $N \sim 10 p$ et $N < 1000$



- Métamodèle souple, performant, interpolateur exact
[ouvrages : Santner et al 03, Rasmussen & Williams 06, Fang et al. 06, Kleijnen 08]

- **Cadre statistique**, expression analytique

➔ Calcul analytique (distribution de la sortie, indices de Sobol)

- Ce métamodèle probabiliste permet d'accéder à d'autres quantités que celles utilisant uniquement le prédicteur :

➔ Estimation de quantiles et de probabilités de dépassement de seuil

➔ Optimisation globale

- La question du plan d'expériences est essentielle

- *Librairies R efficaces pour le PG : mlegp, DiceKriging (à venir)*

Plan



1. 1^{er} exemple d'applications : scénario de transfert hydrogéologique de contaminants

2. Le métamodèle Processus Gaussien (krigeage)

3. Plans d'expériences associés

4. Panorama de différents métamodèles

5. Conclusions

Spécificités des plans d'expériences numériques

(par rapport aux plans d'expériences physiques)



- expériences déterministes (pas d'erreur),
- grand nombre de variable d'entrées,
- larges domaines de variation des entrées,
- variables d'intérêt multiples,
- interactions fortes entre les entrées,
- modèles fortement non linéaires, ...



développement de techniques spécifiques

Space filling designs (répartition uniforme dans l'espace des entrées).

Plan d'expériences maximisant la prédictivité du métamodèle

En grande dimension ($p > 10$), l'espace est creux (très mal rempli)

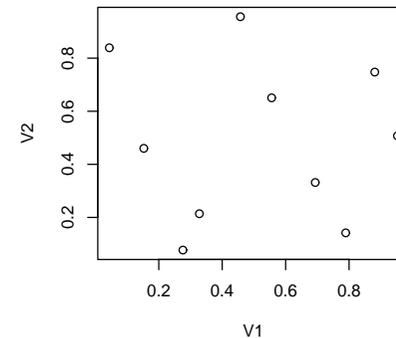
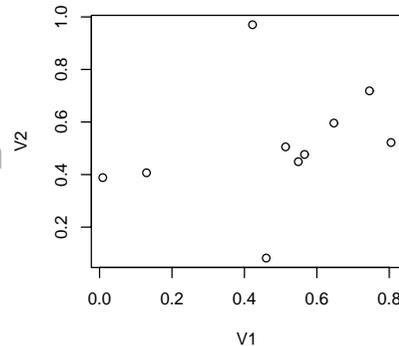


1. Plans « space filling » sont de bons candidats pour bien remplir l'espace

Mesure d'uniformité : la discrédance → suites à faible discrédance

Ex: $p = 2, N = 10$

Simple
Random
Sample

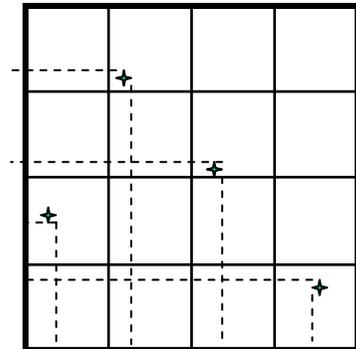


Space
Filling
Design

2. Propriété de *projections uniformes sur les marges*

obtenue via un plan **Hypercube Latin (LHS)** [McKay 79]:
chaque entrée est bien échantillonnée.

Ex: $p = 2, N = 4$



3. Optimisation de LHS [Fang et al. 06] → critères de distance entre les points (minimax/maximin), critères d'uniformité (faible discrédance)

Tests numériques avec des plans à discrédance faible

Discrédance : mesure de la différence entre la configuration des points et une configuration uniforme [Hickernell 98]

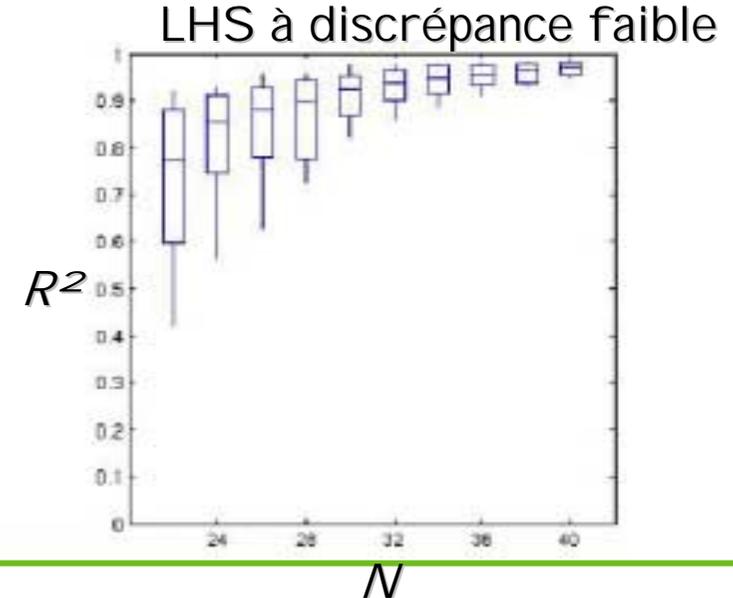
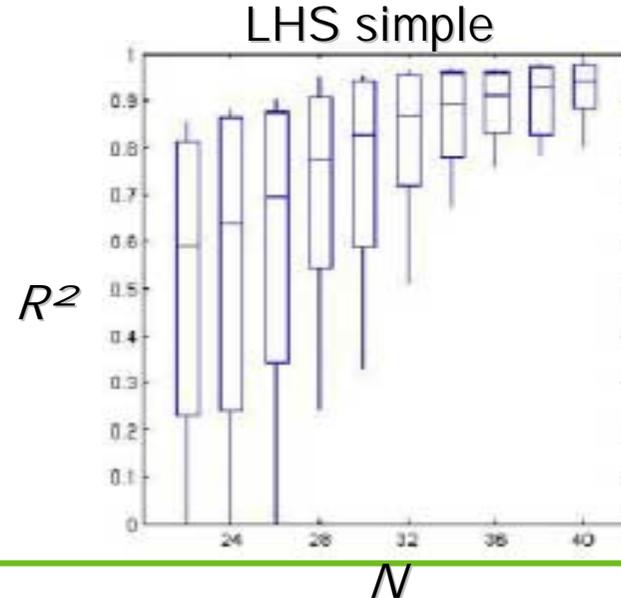


Fonction non monotone ($p = 5$) : g-fonction de Sobol

$$g(X_1, \dots, X_5) = \prod_{i=1}^5 \frac{|4X_i - 2| + a_i}{1 + a_i} \quad \text{avec } X_i \sim U_{[0,1]} \text{ pour } i = 1 \dots 5$$

Ajustement d'un métamodèle PG

Échantillon
d'apprentissage
de taille N



Bilan sur les « space filling designs »

- Les plans LHS de type maximin sont intéressants, mais ne sont pas forcément les meilleurs pour construire un métamodèle

[Marrel 08 , Franco 08]



Axe de recherche encore ouvert

- La **planification adaptative (séquentielle)** est la voie royale pour la construction d'un métamodèle

Pour le métamodèle PG, la disponibilité de la variance permet d'ajouter facilement de nouvelles simulations là où la variance est maximale ; mais il y a bien d'autres stratégies *[Scheidt 06]*

- *Librairies R efficaces pour les Space Filling Designs : lhs, DoCE (à venir)*

Plan



1. Exemple d'applications : scénario de transfert hydrogéologique de contaminants

2. Le métamodèle Processus Gaussien (krigeage)

3. Plans d'expériences associés

4. Panorama de différents métamodèles

5. Conclusions

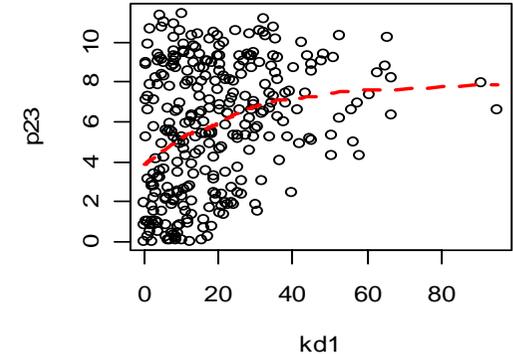
Quelques méthodes de lissage

[Hastie & Tibshirani 90 ; Faraway 06 ; Storlie & Helton 07]



- Polynômes locaux : $\hat{f}(\mathbf{x}) = \hat{\alpha}(\mathbf{x}) + \mathbf{x}\hat{\beta}(\mathbf{x})$
Librairie R loess

- Splines cubiques : $\hat{f}(\mathbf{x}) = \sum_{k=1}^K \hat{\beta}_k B_k(\mathbf{x})$
 K nombre de nœuds,
 $\{B_k\}_{1 \leq k \leq K}$ base des polynômes tronqués (degré 3),
dérivées 2ndes continues sur les nœuds



Problème : fléau de la dimension

- Modèles additifs, GAM : $\hat{f}(\mathbf{x}) = \sum_{i=1}^p s_i(x_i) + \sum_{i<j}^p s_{ij}(x_i, x_j) + \dots$
Par exemple :
 - $s(\cdot)$ = splines de régression pénalisée (*librairie R mgcv*)
 - $s(\cdot)$ = polynômes locaux (*librairie R gam*)

Quelques méthodes d'apprentissage statistique

[Hastie et al. 02 ; Faraway 06 ; Storlie et al. 09]

- Arbres de régression : prédictions constantes par morceaux

$$\hat{f}(\mathbf{x}) = \sum_{k=1}^K \hat{\beta}_k I_k(\mathbf{x})$$



- MARS : prédictions linéaires par morceaux $\hat{f}(\mathbf{x}) = \beta_0 + \sum_{k=1}^K \hat{\beta}_k B_k(\mathbf{x})$

$\{B_k\}_{1 \leq k \leq K}$ produit de termes linéaires par morceaux

- Boosting : méthode adaptative agrégeant des modèles de plus en plus performants (par exemple des arbres de régression)

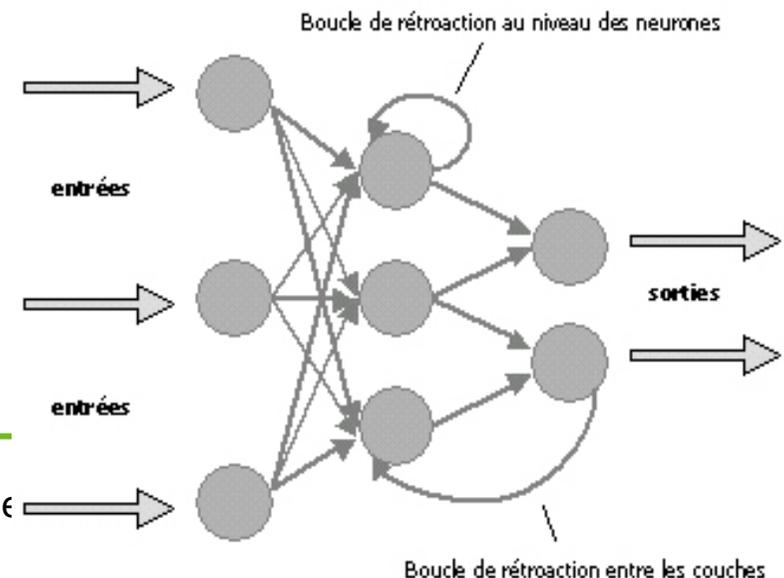
- SVM : méthode d'approximation par noyaux, formellement équivalente au krigeage

- Réseaux de neurones

$$z_m = \sigma(\alpha_{0m} + \mathbf{a}_m^t \mathbf{x}), m = 1, \dots, M$$

$$\hat{f}(\mathbf{x}) = \beta_{0m} + \mathbf{\beta}_m^t \mathbf{z}$$

$\sigma(\cdot)$ est une fonction de transition qui peut être une fct logistique, sigmoïde, radiale, etc.



Méthode spectrale : le chaos polynomial

[Ghanem & Spanos 91 ; Sudret 08]

Base des polynômes du chaos = famille de polynômes orthogonaux $\psi_k(\mathbf{x})$
associés à une mesure $\mu_{\mathbf{x}}$



Ex. : densité gaussienne \Rightarrow pol. d'Hermite ; uniforme \Rightarrow pol. de Legendre

Le développement de chaque v.a. d'entrée sur sa base permet de représenter $Y = f(\mathbf{X})$ par tensorisation (chaos polynomial généralisé) :

$$Y(\mathbf{X}) = \sum_{k \in \mathbb{N}} \alpha_k \psi_k(\mathbf{X})$$

Le jeu consiste à se limiter à un ordre K de développement convenable (précision / nombre de termes).

Par exemple : $p = 2$ et $K = 3 \Rightarrow 10$ polynômes du chaos

Plus : approche déterministe, taux de convergence, atteint un niveau arbitraire d'incertitudes, les coef. du chaos contiennent toute l'info. proba.

Minus : fléau de la dimension, adaptation des outils (en mode intrusif), problème de robustesse (non linéarités, discontinuités)

Plan



1. Exemple d'applications : scénario de transfert hydrogéologique de contaminants

2. Le métamodèle Processus Gaussien (krigeage)

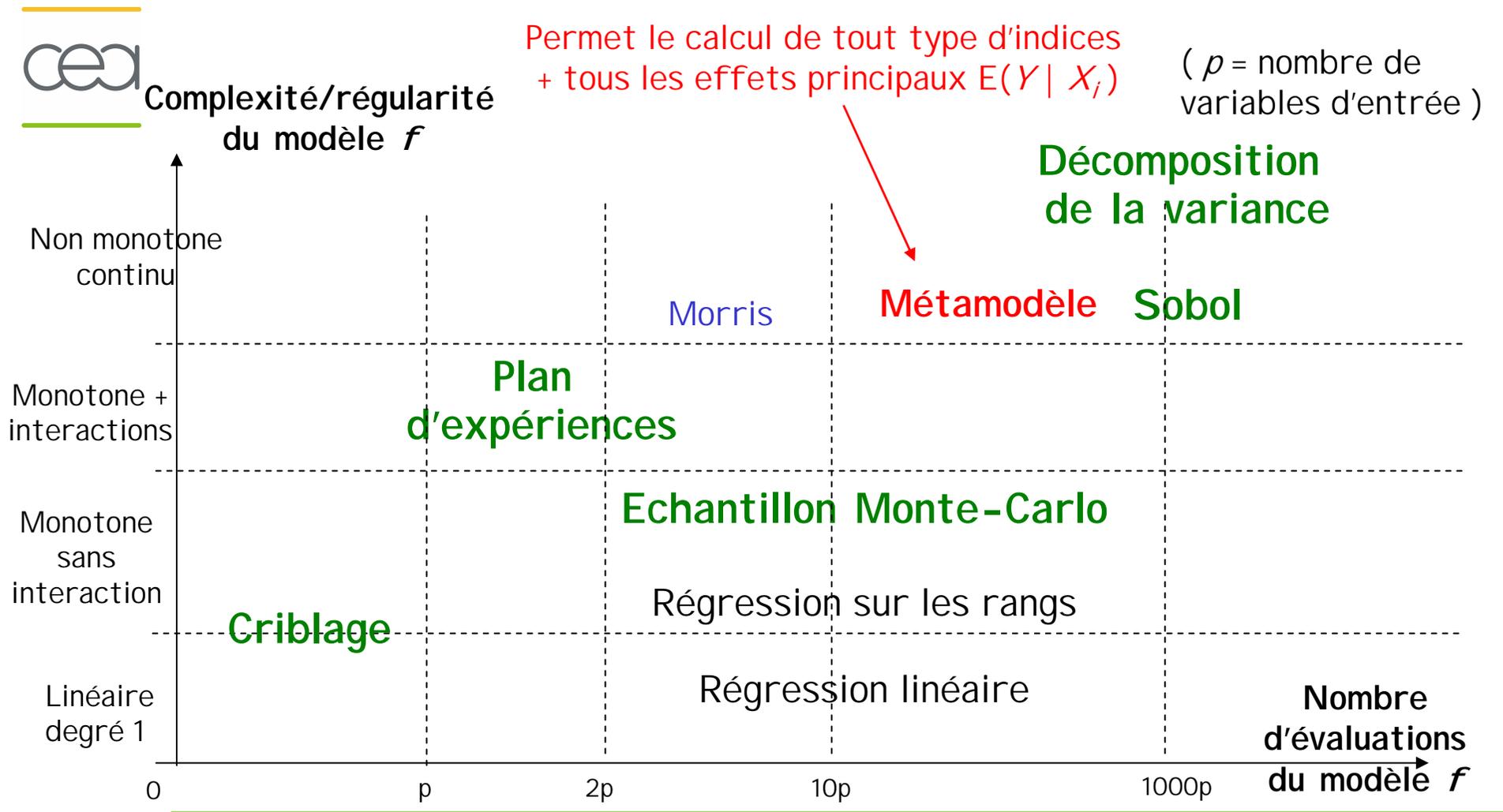
3. Plans d'expériences associés

4. Panorama de différents métamodèles

5. Conclusions

Classification des méthodes d'analyses de sensibilité

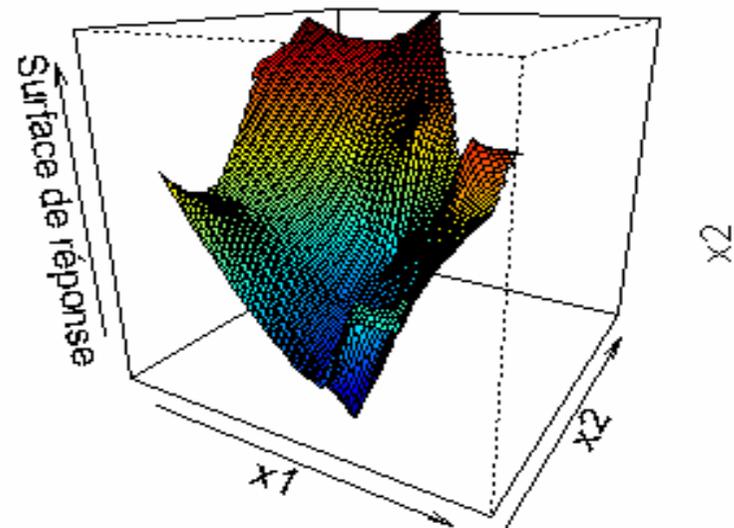
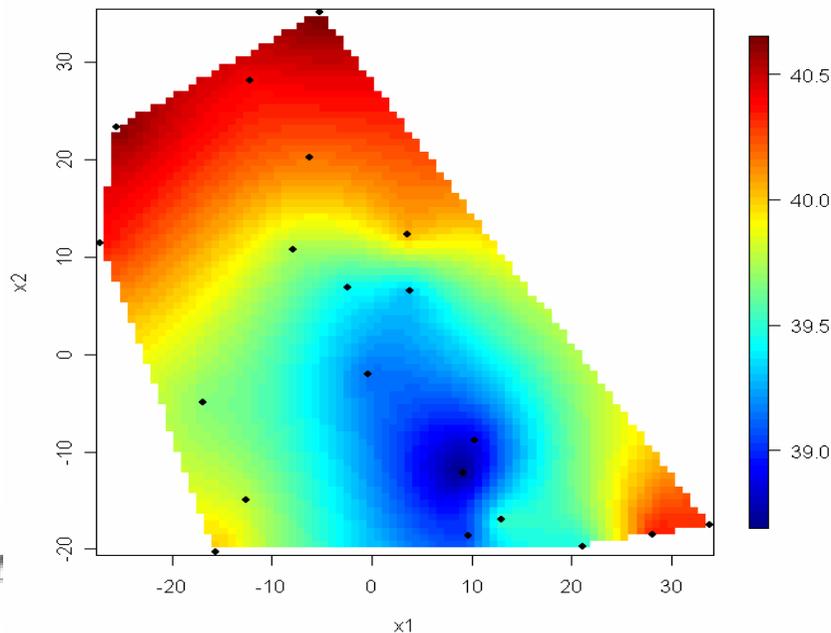
[Iooss 09]



Métamodèle : méthodologie générale

1. Détermination du domaine de variation des paramètres influents
2. Choix d'un plan d'expériences numériques (LHS, SFD, ...)
3. Évaluation du code pour le plan d'expérience choisi à l'étape 2
4. Construction du métamodèle à partir des expériences simulées à l'étape 3 (estimation des paramètres, apprentissage...)
5. Validation du métamodèle (critères stats, validation croisée...)
6. Exploitation du métamodèle (prédiction, A.S., incertitude, calibration...)

Intérêt important du métamodèle : possibilité d'étudier l'impact du choix de la distribution des entrées



Bibliographie

- E. De Rocquigny, N. Devictor & S. Tarantola, *Uncertainty in industrial practice*, Wiley, 2008
- K-T. Fang, R. Li & A. Sudjianto, *Design and modeling for computer experiments*, Chapman, 2006
- J. Faraway, *Extending the linear model with R*, Chapman & Hall/CRC, 2006
- J. Franco, *Planification d'expériences numériques*, Thèse de l'ENSM Saint Etienne, 2008
- R. Ghanem & P. Spanos, *Stochastic finite elements: A spectral approach*, Springer, 1991
- T. Hastie & R. Tibshirani, *Generalized additive models*, Chapman & Hall, 1990
- T. Hastie, R. Tibshirani & J. Friedman, *The elements of statistical learning theory*, Springer, 2002
- J. Helton, J. Johnson, C. Salaberry & C. Storlie, *RESS*, 91, 2006
- F. Hickernell, *Math. Comp.*, 67, 1998
- B. Iooss, *Contributions au traitement des incertitudes en modélisation numérique*, HDR de l'Université Toulouse III, 2009
- B. Iooss & M. Ribatet, Global sensitivity analysis of functional inputs, *RESS*, 94, 2009
- J. Kleijnen, *Design and analysis of simulation experiments*, Springer, 2008
- A. Marrel, Thèse de l'INSA Toulouse, 2008
- A. Marrel, B. Iooss, F. Van Dorpe & E. Volkova, *Comput. Stat. Data Anal.*, 52, 2008
- A. Marrel, B. Iooss, B. Laurent & O. Roustant, *RESS*, 94, 2009
- G. Matheron, *Economic Geology*, 58, 1963
- M. McKay, R. Beckman & W. Conover, *Technometrics*, 21, 1979
- J. Sacks, W. Welch, T. Mitchell, H. Wynn, *Statistical Science*, 4, 1989
- A. Saltelli, K. Chan & E.M. Scott, *Sensitivity analysis*, Wiley, 2000
- C. Scheidt, *Analyse statistique d'expériences simulées*, Thèse de l'Univ. Strasbourg I, 2006
- C. Storlie & J. Helton, *RESS*, 93, 2007
- C. Storlie, L. Swiler, J. Helton & C. Sallaberry, *RESS*, submitted, 2009
- B. Sudret, *Uncertainty propagation and sensitivity analysis in mechanical models*, HDR de l'Université Clermont II, 2008
- E. Volkova, B. Iooss & F. Van Dorpe, *Stoch. Environ. Res. Risk Assess.*, 22, 2008

Voir aussi : <http://www.gdr-mascotnum.fr>