

MaiMoSiNE



METAMODELISATION PAR PROCESSUS GAUSSIEN

Amandine MARREL

CEA Cadarache

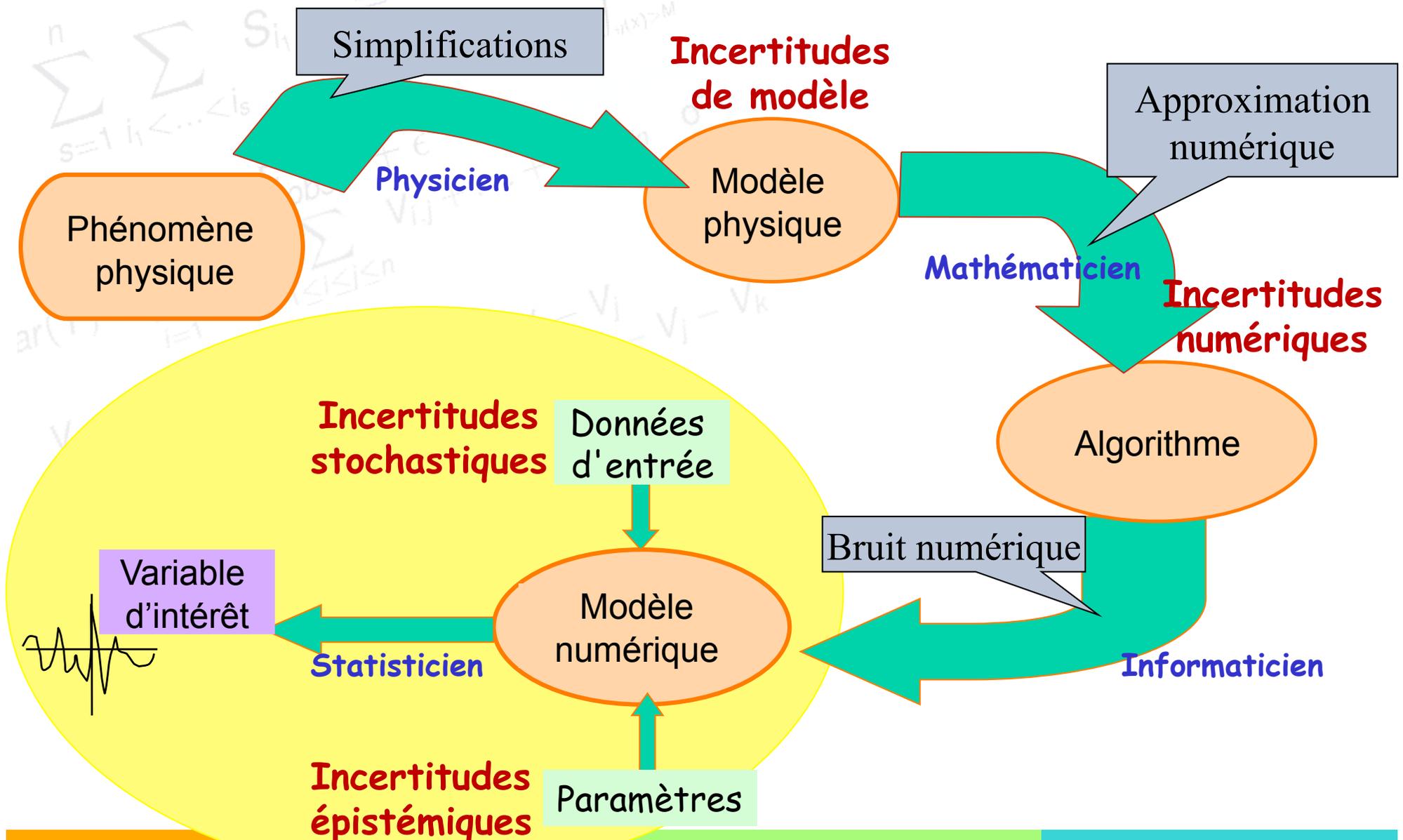
Ecole de Physique des Houches

11 Avril 2013



Pourquoi un métamodèle ?

➤ Du phénomène physique au modèle numérique...



Pourquoi un métamodèle ?

➤ Incertitudes sur les paramètres d'entrée des modèles numériques



➤ Gestion des incertitudes :

- Comment propager les incertitudes sur la sortie du code ?
- Quelles sont les incertitudes les plus préjudiciables ?
- Quelles sont les variables les plus influentes ?
- Incertitude au final sur la sortie ?
- Estimation de la marge de confiance sur la prise de décision ?



Pourquoi un métamodèle ?

➤ Problème :

- Code souvent complexe & « coûteux »
- Grand nombre de variables d'entrée
- Grand nombre de simulations nécessaires pour réaliser les études de sensibilité et de propagation d'incertitude
- Etude souvent « multi-objectif »



Exploitation directe du code difficile



Utilisation d'un métamodèle

Pourquoi un métamodèle ?

➤ Métamodèle :

- Fonction statistique représentative du code de calcul
- Construction à partir de n simulations du code
- Temps d'évaluation négligeable par rapport à celui du code
- Approximant les réponses du code
- Permettant de prédire avec une « bonne précision » de nouvelles réponses dans le domaine de variation des paramètres incertains



Contrôle de la qualité d'approximation et de prédiction

- Outil multifonctionnel : analyse de sensibilité, propagation d'incertitudes, calibration...

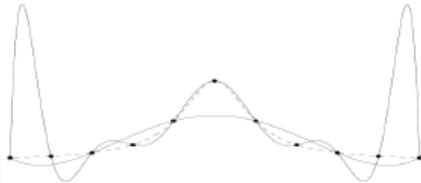
Autres appellations : *surface de réponse* , *émulateur* , *surrogate model* , *modèle simplifié* , ...

Pourquoi un métamodèle ?

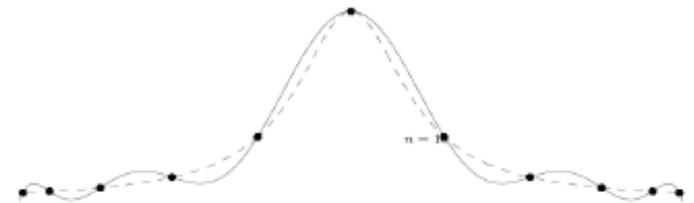
➤ Différents types de métamodèles :

- Modèle linéaire, modèle linéaire généralisé (McCullagh & Nelder [1989])

- Polynômes



- Splines $\hat{G}(\mathbf{x}) = \sum_{k=1}^K \hat{\beta}_k B_k(\mathbf{x})$ avec K le nb de noeuds

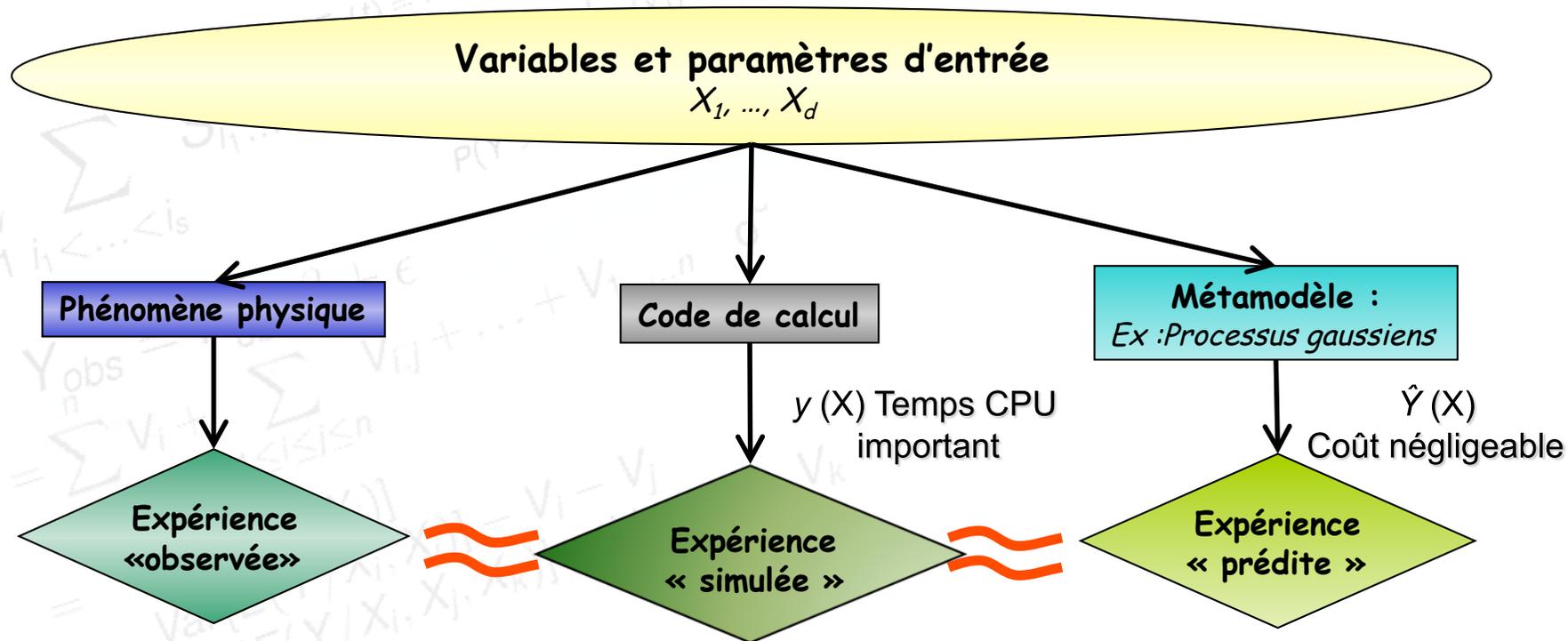


- Modèle additif, modèle additif généralisé (Hastie & Tibshirani, [1990])

$$\hat{G}(\mathbf{x}) = \sum_{i=1}^p s_i(x_i) + \sum_{i<j}^p s_{ij}(x_i, x_j) + \dots$$

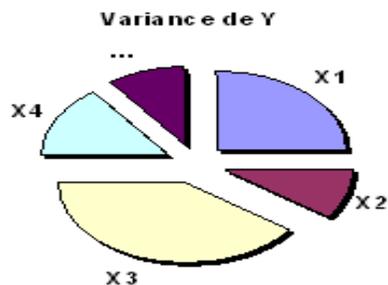
- Arbres de régression (Hastie et al. [2001])
- Réseaux de neurones (Dreyfus et al. [2002])
- Polynômes de chaos (cf. exposé G. Blatman)
- **Krigeage ou processus gaussiens conditionnels**

Pourquoi un métamodèle ?

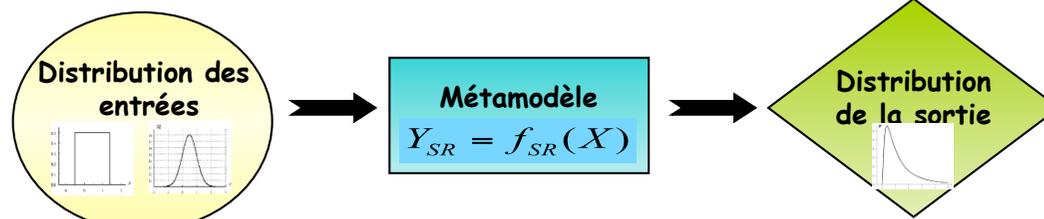


➤ Exploitation du métamodèle

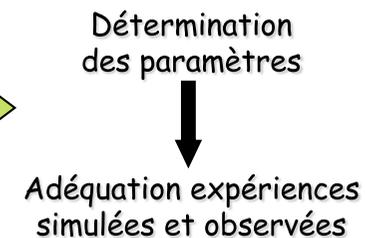
▪ Analyse de sensibilité



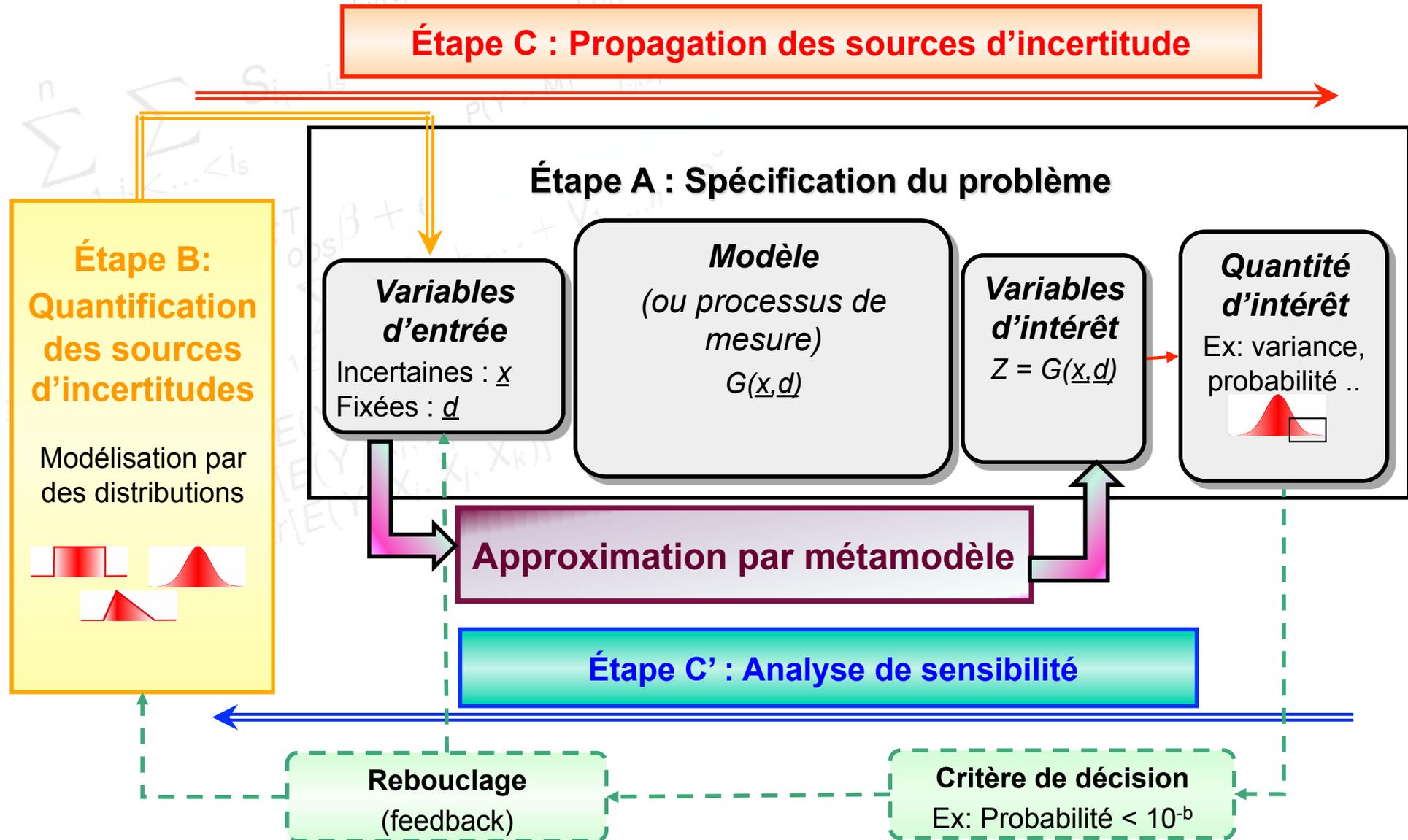
▪ Propagation d'incertitudes



▪ Calibration



Pourquoi un métamodèle ?



De Rocquigny et al. [2008]. Uncertainty in industrial practice. Wiley.

Quelques références sur les métamodèles

➤ Méthode des métamodèles/surfaces de réponse :

▪ Pour les expériences physiques :

Box & Draper [1987]. Empirical model building and responses surfaces. Wiley Series in Probability and Mathematical Statistics.

▪ Généralisation aux codes de calcul avec temps CPU trop important :

Sacks et al. [1989]. Design and analysis of computer experiments. Statistical Science, 4, 409-435.

Fang et al [2006]. Design and modeling for computer experiments. Chapman & Hall/CRC.

➤ Choix du métamodèle :

Hastie et al. [2002]. The elements of statistical learning. Springer.

Chen et al. [2006]. A review on design, modeling and applications of computer experiments. IIE Transactions, 38:273-291.

Fang et al [2006]. Design and modeling for computer experiments. Chapman & Hall/CRC.

Simpson et al [2001]. Metamodel for computer-based engineering designs : survey and recommandations. Engineering with computers, 17:129-150.

Choix délicat, une stratégie possible :

=> **privilégier la simplicité** : métamodèle le + simple en fonction des objectifs de l'étude.

=> Si validation non satisfaisante => choix d'un métamodèle plus complexe.

Cf. package SMURFER, B. Iooss

PLAN

● Rappels statistiques et théorie

- Variable, vecteur et processus gaussien

● Définition du métamodèle processus gaussien (PG)

- Du krigeage à la régression par PG
- Formulation analytique
- Analogie avec le krigeage

● Construction du métamodèle PG

- Choix de la fonction moyenne
- Choix de la fonction de covariance
- Estimation des paramètres

● Utilisation du métamodèle PG

- Validation du métamodèle PG
- Illustration sur des exemples analytiques
- Plans d'expériences et planification adaptative

Rappels statistiques et théorie

➤ De la variable au champ aléatoire :

- **Variable aléatoire réelle** : fonction définie sur l'ensemble des éventualités Ω et qui associe à chaque éventualité ω un élément de \mathfrak{R}

$$X : \omega \rightarrow X(\omega)$$

- **Vecteur aléatoire réel ou variable aléatoire vectorielle** : à chaque ω est associée n éléments de \mathfrak{R}

$$\omega \rightarrow [X_1(\omega), X_2(\omega), \dots, X_n(\omega)]$$

- **Fonction aléatoire réelle** : variable aléatoire vectorielle avec une infinité de composantes

- **Dénombrable** : $Z_1(\omega), Z_2(\omega), \dots, Z_i(\omega), \dots$

- **Non dénombrable** : fonction $Z(t)$, fonction $Z(x)$

- **Champ aléatoire réel** : fonction aléatoire dans un espace à plusieurs dimensions : $Z(x,y)$, $Z(x,y,z)$, $Z(x,t)$...

Rappels statistiques et théorie

➤ Variable aléatoire vectorielle :

▪ Variable aléatoire réelle à 2 composantes : $\omega \rightarrow (Z_1(\omega), Z_2(\omega)) \in \mathfrak{R}^2$

▪ Fonction de répartition conjointe :

$$F_{Z_1, Z_2} : \mathfrak{R}^2 \rightarrow [0, 1]$$

$$(z_1, z_2) \rightarrow F_{Z_1, Z_2}(z_1, z_2) = P(Z_1 \leq z_1, Z_2 \leq z_2)$$

▪ Densité de probabilité conjointe :

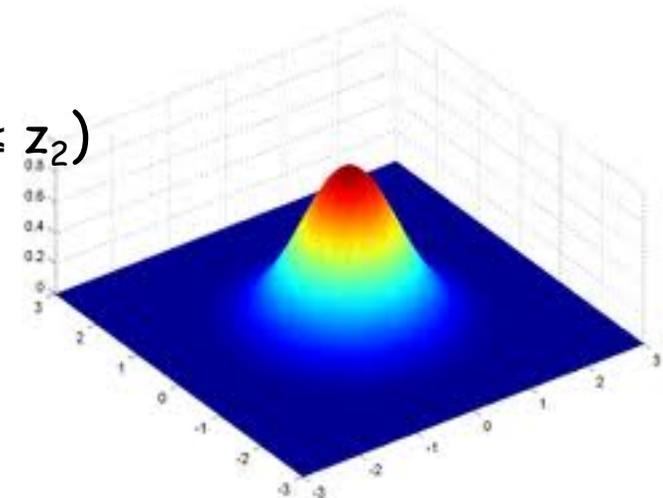
$$F_{Z_1, Z_2}(z_1, z_2) = \int_{-\infty}^{z_2} \int_{-\infty}^{z_1} f_{Z_1, Z_2}(u, v) du dv$$

▪ Indépendance entre Z_1 et Z_2 $\iff f_{Z_1, Z_2}(z_1, z_2) = f_{Z_1}(z_1) f_{Z_2}(z_1)$

▪ Covariance et coefficient de corrélation :

$$\begin{aligned} \text{cov}(Z_1, Z_2) &= E[(Z_1 - E[Z_1])(Z_2 - E[Z_2])] \\ &= E[Z_1 Z_2] - E[Z_1]E[Z_2] \end{aligned}$$

$$\begin{aligned} \rho_{1,2} &= \frac{\text{Cov}(Z_1, Z_2)}{\sqrt{\text{Var}(Z_1)\text{Var}(Z_2)}} = \frac{\text{Cov}(Z_1, Z_2)}{\sigma_1 \sigma_2} \\ &-1 \leq \rho_{1,2} \leq 1 \end{aligned}$$



Rappels statistiques et théorie

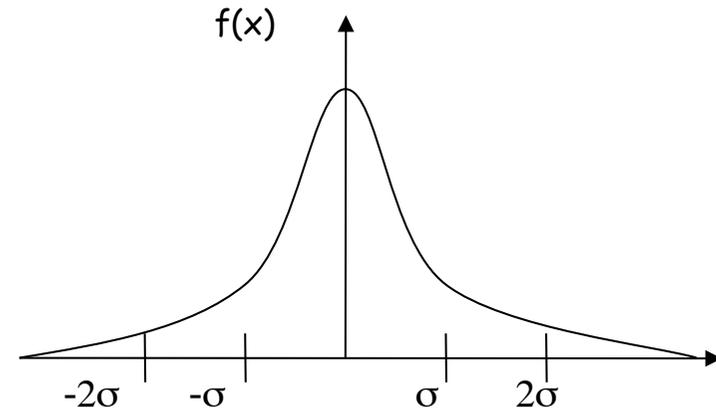
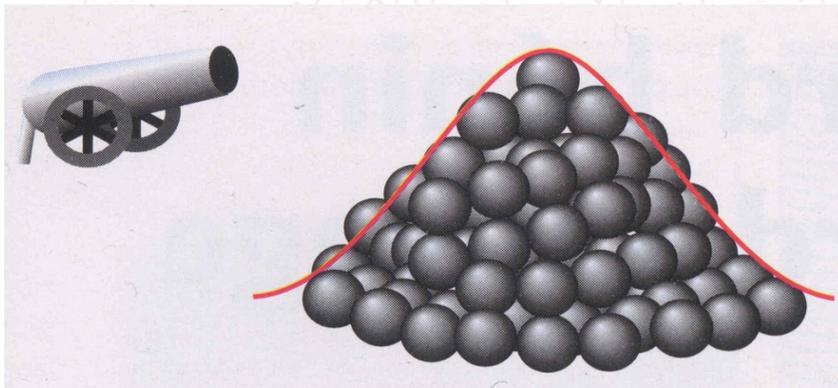
➤ Loi gaussienne ou loi normale $\mathcal{N}(\mu, \sigma^2)$

▪ Densité

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

▪ Moments

$$E(X) = \mu ; \text{var}(X) = \sigma^2$$



$$\begin{aligned} P(\mu - \sigma < X < \mu + \sigma) &= 0.68 \\ P(\mu - 1.64\sigma < X < \mu + 1.64\sigma) &= 0.90 \\ P(\mu - 1.96\sigma < X < \mu + 1.96\sigma) &= 0.95 \\ P(\mu - 3.09\sigma < X < \mu + 3.09\sigma) &= 0.998 \end{aligned}$$

Exemples : *impacts des boulets de canon (Jouffret, 1872)*

Rappels statistiques et théorie

➤ Vecteur gaussien

- Loi $\mathcal{N}(0, I_n)$

$$\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)^T \sim \mathcal{N}(0, I_n)$$



tous les Z_i ($i=1, \dots, n$) suivent indépendamment des lois gaussiennes centrées réduites $\sim \mathcal{N}(0, 1)$

- **Vecteur gaussien d-dimensionnel \mathbf{Z}**

- Si toute combinaison linéaire de ses composantes suit une loi gaussienne
- Si sa fonction caractéristique est de la forme :

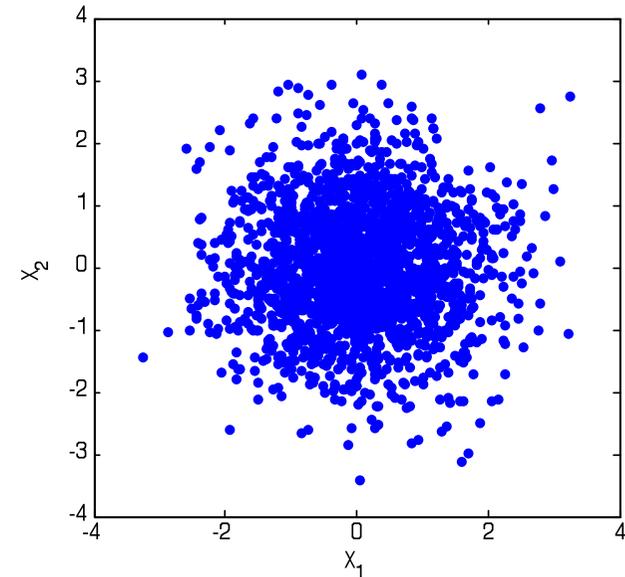
$$\forall u \in \mathbb{R}^n, \phi_{\mathbf{Z}}(u) = \mathbb{E}\left[e^{i\langle u, \mathbf{Z} \rangle}\right] = e^{-i\langle u, m \rangle - \frac{1}{2}\langle u, Au \rangle}$$

avec $m \in \mathbb{R}^n$ et $A \in M^n(\mathbb{R})$ une matrice symétrique positive

- Si \mathbf{Z} s'écrit comme une transformation affine d'une $\mathcal{N}(0, I_n)$:

$$\mathbf{Z} = m + A\mathbf{N} \quad \text{où } \mathbf{N} \sim \mathcal{N}(0, I_n)$$

avec $m \in \mathbb{R}^n$ et $A \in \mathcal{M}_n(\mathbb{R})$



Rappels statistiques et théorie

➤ Vecteur gaussien

▪ Propriétés :

Transformation affine

- Si $A \in \mathcal{M}_{n,k}(\mathbb{R})$, $b \in \mathbb{R}^n$ et $Z \sim \mathcal{N}_n(m, \Sigma)$, alors

$$AZ + b \sim \mathcal{N}(Am + b, A \Sigma A^t)$$

Utilisation pour simuler un vecteur gaussien

$Z \sim N(0, I_n)$ et Σ matrice symétrique définie positive

alors $m + \Sigma^{\frac{1}{2}} Z \sim N(m, \Sigma)$



Simulation à partir d'une $\mathcal{N}(0, I_n)$

Rappels statistiques et théorie

➤ Vecteur gaussien $Z = (Z_1, Z_2, \dots, Z_n)^T \sim \mathcal{N}(m, \Sigma)$

- Vecteur de la moyenne

$$m = E(Z) = [\mu_1, \dots, \mu_n]^T = [E(Z_1), \dots, E(Z_n)]^T$$

- Matrice de covariance (hypo : $\text{Var}[Z_i] = \sigma_i^2 > 0$)

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \text{Cov}(Z_1, Z_2) & \dots & \text{Cov}(Z_1, Z_n) \\ \text{Cov}(Z_1, Z_2) & \sigma_2^2 & \dots & \text{Cov}(Z_2, Z_n) \\ \dots & \dots & \dots & \dots \\ \text{Cov}(Z_1, Z_n) & \text{Cov}(Z_2, Z_n) & \dots & \sigma_n^2 \end{pmatrix}$$

$$\text{avec } \begin{cases} \text{Cov}(Z_i, Z_j) = E[(Z_i - E[Z_i])(Z_j - E[Z_j])] \\ \sigma_i^2 = \text{Cov}(Z_i, Z_i) = \text{Var}[Z_i] \end{cases}$$

Rappels statistiques et théorie

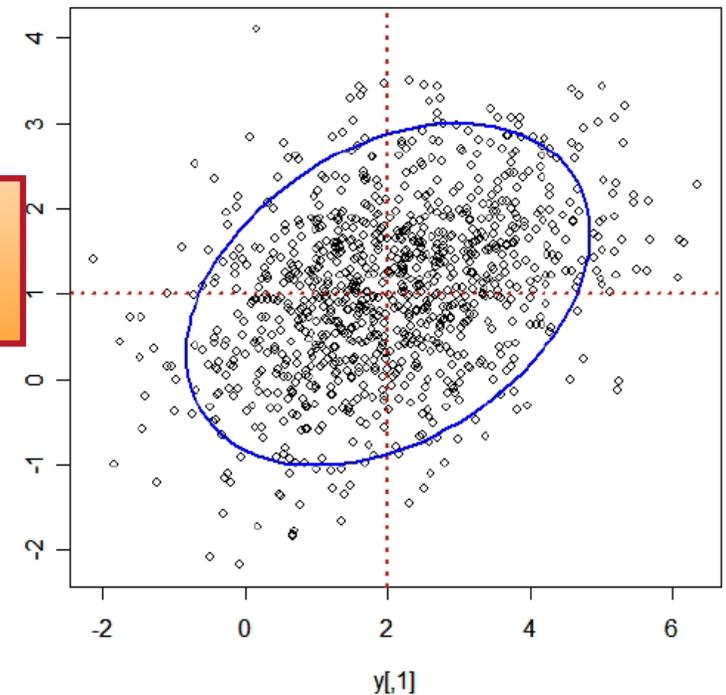
➤ Vecteur gaussien

- Loi bi-gaussienne : $Z = (Z_1, Z_2)^T \sim \mathcal{N}(m, \Sigma)$

$$\begin{cases} m = [\mu_1, \mu_2]^T \\ \Sigma = \begin{pmatrix} \sigma_1^2 & \text{Cov}(Z_1, Z_2) \\ \text{Cov}(Z_1, Z_2) & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \end{cases}$$

Simulation d'un échantillon de 1000 points avec $m = (2, 1)^T$; $\sigma_1^2 = 2$; $\sigma_2^2 = 1$ et $\rho = 0.35$

Ellipse de confiance à 95%



Rappels statistiques et théorie

➤ Vecteur gaussien

▪ Propriétés :

- Les composantes d'un vecteur gaussien sont gaussiennes (Attention : la réciproque n'est pas vraie)
- La somme de 2 vecteurs indépendants est un vecteur gaussien

▪ **Dépendance** ⇔ **corrélation**

▪ Représentation de loi multi-gaussienne : $Z \sim \mathcal{N}(m, \Sigma)$

- Cas non-dégénéré : Σ symétrique définie positive

$$\Rightarrow Y = \Sigma^{-\frac{1}{2}}(Z - m) \sim N(0, I_n)$$

$$\Rightarrow Z = m + \Sigma^{\frac{1}{2}}Y$$



Simulation à partir
d'une $\mathcal{N}(0, I_n)$

Z admet une densité :

$$\forall Z \in \mathbb{R}^n, f_Z(z) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} e^{-\frac{1}{2} \left[\Sigma^{-\frac{1}{2}}(z-m) \right]^T \left[\Sigma^{-\frac{1}{2}}(z-m) \right]}$$

Rappels statistiques et théorie

➤ Vecteur gaussien

- **Propriétés de projection** : Théorème de Cochran
- **Statistiques des échantillons gaussiens**: moyenne et variance empiriques indépendantes et resp. de loi normale et khi2
- **Conditionnement** : $\mathbf{Z} = (Z_1, Z_2)^T$

Loi jointe : $\mathbf{Z} \sim \mathcal{N}(\mathbf{m}, \Sigma)$ avec $\mathbf{m} = (m_1, m_2)^T$ et $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix}$

Loi conditionnelle Z_2 sachant $Z_1 = z_1$:

- $Z_2 | Z_1 = z_1$ suit une loi gaussienne moyenne et variance :

$$\begin{cases} E[Z_2 | Z_1 = z_1] = m_2 + \sigma_{1,2} \sigma_1^{-2} (z_1 - m_1) \\ \text{Var}[Z_2 | Z_1 = z_1] = \sigma_2^2 - \sigma_{1,2} \sigma_1^{-2} \sigma_{1,2} \end{cases}$$

Rappel conditionnement : $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.

Pour des variables continues, densité de $Z_2 | Z_1 = z_1$: $f_{Z_2 | Z_1 = z_1}(z_2) = \frac{f_{Z_2, Z_1}(z_2, z_1)}{f_{Z_1}(z_1)}$

Rappels statistiques et théorie

➤ Vecteur gaussien

▪ Conditionnement :

Extension au cas général $Z'_1 = (Z_1, \dots, Z_p)^T$ et $Z'_2 = (Z_{p+1}, \dots, Z_n)^T$

$$(Z'_1, Z'_2)^T \sim N \left(\begin{matrix} m'_1 \\ m'_2 \end{matrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{1,2} \\ \Sigma_{1,2}^T & \Sigma_{22} \end{pmatrix} \right)$$

avec m'_1 vecteur $p \times 1$, m'_2 vecteur $(n-p) \times 1$,

Σ_{11} matrice $p \times p$, Σ_{22} matrice $(n-p) \times (n-p)$

et $\Sigma_{1,2}$ matrice $(p) \times (n-p)$

⇒ loi conditionnelle de $Z'_2 | Z'_1 = z'_1$:

$$Z'_2 | Z'_1 = z'_1 \sim N \left(m'_2 + \Sigma_{1,2}^T \Sigma_{11}^{-1} (z'_1 - m'_1), \Sigma_{22} - \Sigma_{1,2}^T \Sigma_{11}^{-1} \Sigma_{1,2} \right)$$

Rappels statistiques et théorie

➤ Processus Gaussiens (PG)

- **Définition** : un processus gaussien est un processus aléatoire réel $\{Z(x)\}_{x \in \mathbb{D} \subset \mathbb{R}^d}$ dont toutes les lois finies-dimensionnelles $(Z(x_1), \dots, Z(x_n))$ sont gaussiennes
- **Propriétés** : entièrement caractérisé par sa moyenne et sa fonction de covariance

$$Z(x) \sim PG(m(x), C(x, x'))$$

$$\text{où } m(x) = E[Z(x)]$$

$$\text{et } C(x, x') = \text{Cov}(Z(x), Z(x')) = E[(Z(x) - m(x)) (Z(x') - m(x'))]$$

- **Processus centré** : $m(x) = 0$
- **Covariance stationnaire** : $C(x, x') = C(x - x')$
- **Covariance isotrope** : $C(x, x') = C(\|x - x'\|)$
- **Fonction de corrélation R** : $C(x, x') = \sigma^2 R(x - x')$ avec $\sigma^2 = \text{Var}[Z(x)]$

PLAN

● Rappels statistiques et théorie

- Variable, vecteur et processus gaussien

● Définition du métamodèle processus gaussien (PG)

- Du krigeage à la régression par PG
- Formulation analytique
- Analogie avec le krigeage

● Construction du métamodèle PG

- Choix de la fonction moyenne
- Choix de la fonction de covariance
- Estimation des paramètres

● Utilisation du métamodèle PG

- Validation du métamodèle PG
- Illustration sur des exemples analytiques
- Plans d'expériences et planification adaptative

Définition du métamodèle processus gaussien

➤ Du krigeage à la régression par processus gaussiens

- **Géostatistique/statistique spatiale** : application de la théorie des fonctions aléatoires à des données localisées dans un espace géographique,
=> Application de méthodes probabilistes pour l'étude de phénomènes corrélés dans l'espace
- **1955 : Daniel Krige**, mines d'or sud-africaines
Approche empirique pour corriger les problèmes de biais lors de l'estimation de la teneur d'un bloc de minerai à partir d'un nombre limité d'échantillons pris autour du bloc
- **1960-1970 : Georges Matheron**
Formalisation théorique et introduction d'un outil pour analyser la continuité spatiale des variables d'intérêt : le variogramme
=> Méthode d'estimation basée sur le variogramme : **krigeage**

Définition du métamodèle processus gaussien

➤ Du krigeage à la régression par processus gaussiens

- **1970-actuellement** : Extension des méthodes géostatistiques à d'autres domaines : pétrole, environnement, hydrologie, océanographie, foresterie, **codes de calcul**, ...

L'idée du krigeage pour les codes est d'interpoler les réponses du code en dimension p comme pour une cartographie spatiale

- **Différents points de vue** :
 - **Weight space view** => Krigeage => estimateur BLUP
 - **Function space view** => vecteurs gaussiens, loi jointe et loi conditionnelle => métamodèle processus Gaussien

2 approches différentes mais qui conduisent au même métamodèle



Métamodèle krigeage ou métamodèle processus gaussien

Définition du métamodèle processus gaussien

➤ Formulation analytique du métamodèle PG :

- Entrées du code : $x \in D \subset \mathbb{R}^d$

- Sortie du code : $z(x) \in \mathbb{R}$

- Hypothèse :

la sortie déterministe $z(x)$ est considérée comme la réalisation d'un champ aléatoire $Z(x)$:

$$Z(x) = f_o(x) + W(x)$$

avec :

- f_o la fonction moyenne

- $W(x)$ un processus gaussien centré stationnaire de variance σ^2 et de fonction de corrélation R :

$$\text{Cov}(W(x), W(x')) = c(x-x') = \sigma^2 R(x-x')$$

Définition du métamodèle processus gaussien

➤ Formulation analytique du métamodèle PG :

▪ Base d'apprentissage (BA) :

- N points de simulations du code de calcul $\{x_1, \dots, x_n\}$

$$X_s = [x_1^T, \dots, x_n^T]^T$$

- N réalisations/observations du code associées :

$$Z_s = Z(X_s) = [Z(x_1), \dots, Z(x_n)]^T$$

▪ Loi jointe de la BA: $Z_s \sim \mathcal{N}(F_s, \Sigma_s)$

avec $F_s = f_o(X_s) = [f_o(x_1), \dots, f_o(x_n)]^T$ et Σ_s la matrice de covariance

$$(\Sigma_s)_{i,j} = c(x_i - x_j) = \sigma^2 R(x_i - x_j)$$

▪ Point de prédiction : nouveau point non échantillonné x^*

Définition du métamodèle processus gaussien

➤ Formulation analytique du métamodèle PG :

- **Loi jointe {BA+x*} :**

$$[Z(X_S), Z(x^*)] \sim N\left(\begin{pmatrix} F_S \\ f_0(x^*) \end{pmatrix}, \begin{pmatrix} \Sigma_S & k(x^*) \\ k(x^*) & \sigma^2 \end{pmatrix}\right)$$

- **Loi conditionnelle : $Z(x^*) | Z(X_S)$**

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

$$\text{avec} \begin{cases} \mu(x^*) = E[Z(x^*) | Z(X_S) = Z_S] = f_0(x^*) + k(x^*)^T \Sigma_S (Z_S - F_S) \\ \tilde{\sigma}^2(x^*) = \text{Var}[Z(x^*) | Z(X_S) = Z_S] = \sigma^2 - k(x^*)^T \Sigma_S k(x^*) \end{cases}$$

Définition du métamodèle processus gaussien

➤ Formulation analytique du métamodèle PG :

- **Prédicteur du métamodèle PG :**

$$\hat{Z}(x^*) = \mu(x^*) = f_0(x^*) + k(x^*)^T \Sigma_S (Z_S - F_S)$$

« **Tendance déterministe** »

Terme identique à celui des MC classiques

« **Partie permettant l'interpolation** »

Prise en compte de la configuration des données

- **Erreur du prédicteur :**

$$MSE[\hat{Z}(x^*)] = \tilde{\sigma}^2(x^*) = \sigma^2 - k(x^*)^T \Sigma_S k(x^*)$$

Définition du métamodèle processus gaussien

➤ Propriétés du métamodèle PG :

- Interpolateur exact des points de la base d'apprentissage :

$$\hat{Z}(X_s) = Z_s$$

- Meilleur prédicteur linéaire sans biais (BLUP)

=> prédicteur et variance identiques à ceux du krigeage obtenus avec l'approche « Weight space view »



Krigeage universel

Définition du métamodèle processus gaussien

➤ Analogie avec le krigeage :

▪ Hypothèses :

- Processus spatial Z de covariance c connue
- n sites de mesures $\{s_1, \dots, s_n\} \Rightarrow$ observations de $\{Z(s_1), \dots, Z(s_n)\}$
- Z centré de moyenne nulle

▪ Objectif : Prédire le processus pour un nouveau site non observé s^*

▪ Idée : Construire un estimateur qui soit une combinaison linéaire des observations :

$$\hat{Z}(s^*) = \sum_{i=1}^n \lambda_i Z(s_i) = \lambda' Z_s$$

- Sans biais : $E[\hat{Z}(s^*) - Z(s^*)] = 0$ (1)

- De variance minimale : $Var[\hat{Z}(s^*) - Z(s^*)] = \min_{\lambda \in \mathbb{R}^n} Var\left[\sum_{i=1}^n \lambda_i Z(s_i) - Z(s^*)\right]$ (2)

Estimateur BLUP

Détermination des λ_i

Définition du métamodèle processus gaussien

➤ Analogie avec le krigeage :

- **Détermination des λ_i :**

Condition (2) implique :

$$\text{Var} \left[\sum_i \lambda_i Z(s_i) - Z(s^*) \right] = \text{Var}[Z(s^*)] + \sum_i \sum_j \lambda_i \lambda_j \text{Cov}[Z(s_i), Z(s_j)] - 2 \sum_i \lambda_i \text{Cov}[Z(s_i), Z(s^*)]$$

Dérivation par rapport aux λ_i

=> système linéaire avec n équations à n inconnues

$$\sum_j \lambda_j \text{Cov}[Z(s_i), Z(s_j)] = \text{Cov}[Z(s_i), Z(s^*)] \quad \forall i = 1..n$$

=> Réécriture sous forme matricielle du système linéaire :

$$\underbrace{\begin{bmatrix} \sigma^2 & \text{Cov}(Z_1, Z_2) & \cdots & \text{Cov}(Z_1, Z_n) \\ \text{Cov}(Z_2, Z_1) & \sigma^2 & \cdots & \text{Cov}(Z_2, Z_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(Z_n, Z_1) & \text{Cov}(Z_n, Z_2) & \cdots & \sigma^2 \end{bmatrix}}_{\mathbf{K}_S} \underbrace{\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix}}_{\boldsymbol{\lambda}_S} = \underbrace{\begin{bmatrix} \text{Cov}(Z_1, Z_v) \\ \text{Cov}(Z_2, Z_v) \\ \vdots \\ \text{Cov}(Z_n, Z_v) \end{bmatrix}}_{\mathbf{k}_S}$$

Définition du métamodèle processus gaussien

➤ Analogie avec le krigeage :

- **Détermination des λ_i :** Résolution du système linéaire :

$$K_s \lambda_s = k_s \quad \Rightarrow \quad \lambda_s = K_s^{-1} k_s$$



$$\begin{cases} \hat{Z}(s^*) = {}^t k_s K_s^{-1} Z_s \\ \text{Var}[\hat{Z}(s^*) - Z(s^*)] = \sigma^2 - {}^t k_s K_s^{-1} k_s \end{cases}$$

- **Généralisation aux processus non centré :**

- Moyenne m constante et connue => krigeage simple
- Moyenne m constante et inconnue => krigeage ordinaire
- Moyenne fonction de s , $m(s)$ inconnue (présence d'une « dérive »)
=> krigeage universel ou krigeage avec dérive

PLAN

● Rappels statistiques et théorie

- Variable, vecteur et processus gaussien

● Définition du métamodèle processus gaussien (PG)

- Du krigeage à la régression par PG
- Formulation analytique
- Analogie avec le krigeage

● Construction du métamodèle PG

- Choix de la fonction moyenne
- Choix de la fonction de covariance
- Estimation des paramètres

● Utilisation du métamodèle PG

- Validation du métamodèle PG
- Illustration sur des exemples analytiques
- Plans d'expériences et planification adaptative

Construction du métamodèle PG

▪ **Modèle** : $x \in D \subset \mathbb{R}^d$

$$Z(x) = f_0(x) + W(x)$$

avec :

▪ f_0 la fonction moyenne

▪ $W(x)$: PG centré stationnaire de variance σ^2 et de fx de corrélation R

➤ Choix de la fonction « moyenne » f_0

▪ **Constante** $f_0(x) = c$

▪ **Polynôme de degré 1** $f_0(x) = \beta_0 + \sum_{i=1}^d \beta_i x_i$

▪ **Polynôme d'ordre supérieur**

▪ **Modèle additif généralisé**

$$f_0(x) = \beta_0 + \sum_{i=1}^d \beta_i f_i(x_i) + \sum_{\substack{i,j=1 \\ i \neq j}}^d \beta_{i,j} f_{i,j}(x_i, x_j) + \dots$$

Construction du métamodèle PG

▪ **Modèle** : $x \in D \subset \mathbb{R}^d$

$$Z(x) = f_o(x) + W(x)$$

avec :

▪ f_o la fonction moyenne

▪ $W(x)$: PG centré stationnaire de variance σ^2 et de fx de corrélation R

➤ Choix de la fonction de covariance/corrélation

▪ **Propriétés par définition:**

- Symétrique $c(x, x') = c(x', x)$

- Semi-positif $Var[\sum_{i=1}^n \lambda_i Z(s_i)] \geq 0 \Rightarrow \begin{cases} \sum_{i,j=1}^n \lambda_i \lambda_j c(s_i, s_j) \geq 0 \quad \forall \lambda_j, \lambda_i \text{ réels} \\ c \text{ est une fonction de type positif} \end{cases}$

▪ **Covariance stationnaire :**

$$c(x, x') = c(x - x')$$

▪ **Covariance isotrope :**

$$c(x, x') = c(\|x - x'\|)$$

Rq : différentiabilité de c directement liée à la régularité du processus

Construction du métamodèle PG

➤ Choix de la fonction de covariance/corrélation

▪ Modèles paramétriques 1D:

- Covariance exponentielle $c(h) = \sigma^2 \exp\left(-\frac{h}{\theta}\right)$

- Covariance gaussienne $c(h) = \sigma^2 \exp\left[-\left(\frac{h}{\theta}\right)^2\right]$

- Covariance exponentielle généralisée $c(h) = \sigma^2 \exp\left[-\left(\frac{h}{\theta}\right)^p\right]$ avec $p > 0$

- Covariance de Matern : $c(h) = \sigma^2 \frac{1}{2^{\alpha-1} \Gamma(\alpha)} \left(\frac{h}{\theta}\right)^{\alpha} K_{\alpha}\left(\frac{h}{\theta}\right)$ avec $\alpha > 0$
 K_{α} est la fonction de Bessel

Cas particulier $\alpha = p + 1/2$, $p \in \mathbb{N} \Rightarrow$ exponentiel ($p=0$), Matern3/2 ($p=1$), Matern5/2 ($p=2$)

$$p = 1 \Rightarrow c(h) = \sigma^2 \left(1 + \frac{\sqrt{3}h}{\theta}\right) \exp\left(-\frac{\sqrt{3}h}{\theta}\right)$$

$$p = 2 \Rightarrow c(h) = \sigma^2 \left(1 + \frac{\sqrt{5}h}{\theta} + \frac{5h^2}{3\theta^2}\right) \exp\left(-\frac{\sqrt{5}h}{\theta}\right)$$

Rque : p directement liée à la régularité du processus (différentiabilité de c)

Construction du métamodèle PG

➤ Choix de la fonction de covariance/corrélation

▪ Généralisation en dimension supérieure:

- Covariance isotrope : $c(h) = c(\|h\|)$

- Covariance anisotrope : $c(h) = \prod_{i=1}^d c_i(h_i)$

- Exponentielle

$$c(h) = \sigma^2 \prod_{i=1}^d \exp\left(-\frac{h_i}{\theta_i}\right) = \sigma^2 \exp\left(-\sum_{i=1}^d \frac{h_i}{\theta_i}\right)$$

- Gaussienne

$$c(h) = \sigma^2 \exp\left(-\sum_{i=1}^d \left(\frac{h_i}{\theta_i}\right)^2\right)$$

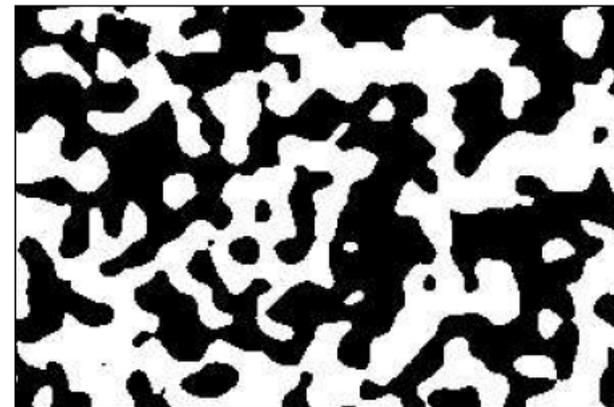
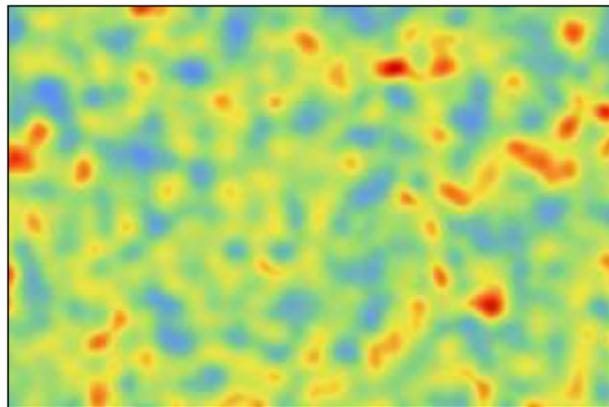
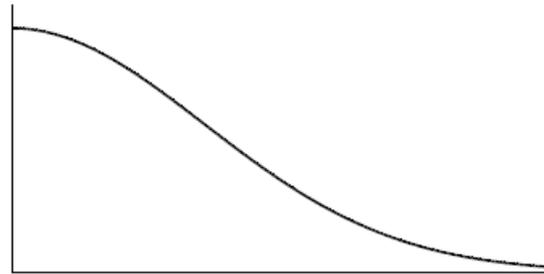
- ...

Construction du métamodèle PG

➤ Choix de la fonction de covariance/corrélation

▪ Covariance gaussienne

$$C(h) = \exp\left\{-\frac{|h|^2}{a^2}\right\}$$



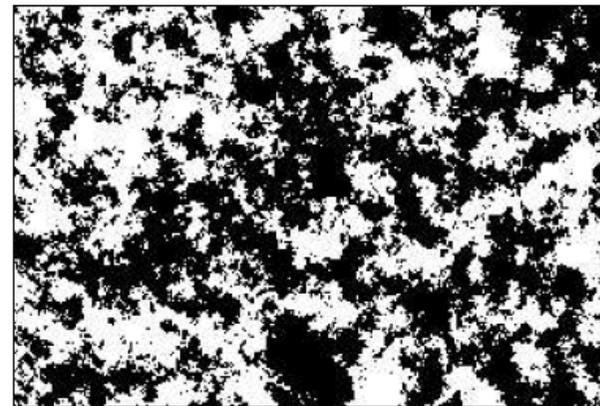
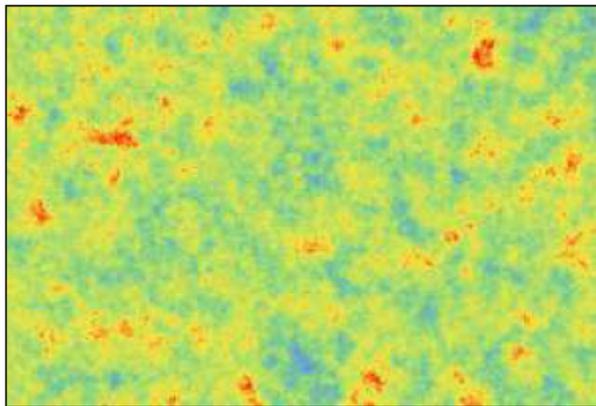
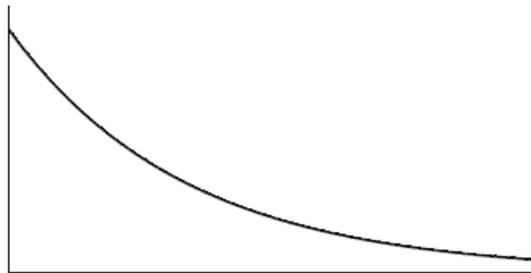
Simulation d'un champ 2D isotrope

Construction du métamodèle PG

➤ Choix de la fonction de covariance/corrélation

▪ Covariance exponentielle

$$C(h) = \exp\left\{-\frac{|h|}{a}\right\}$$



Simulation d'un champ 2D isotrope

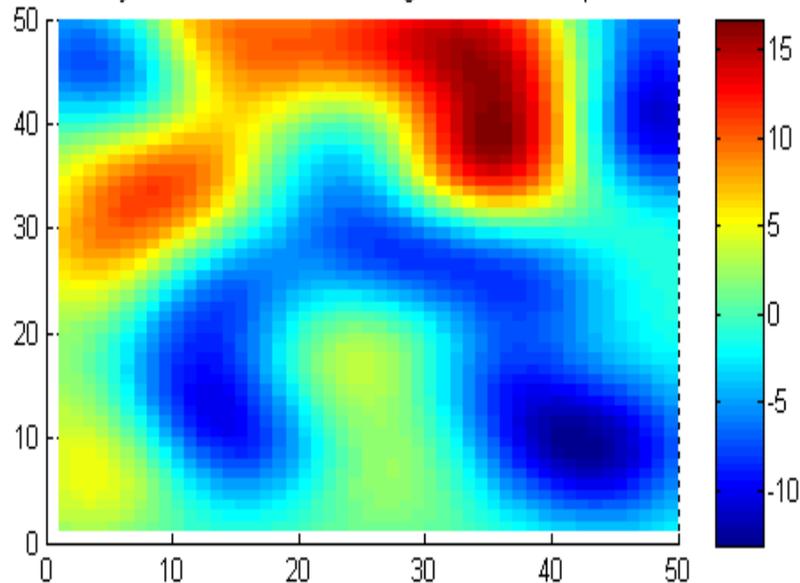
Construction du métamodèle PG

➤ Choix de la fonction de covariance/corrélation

▪ Covariance exponentielle anisotrope

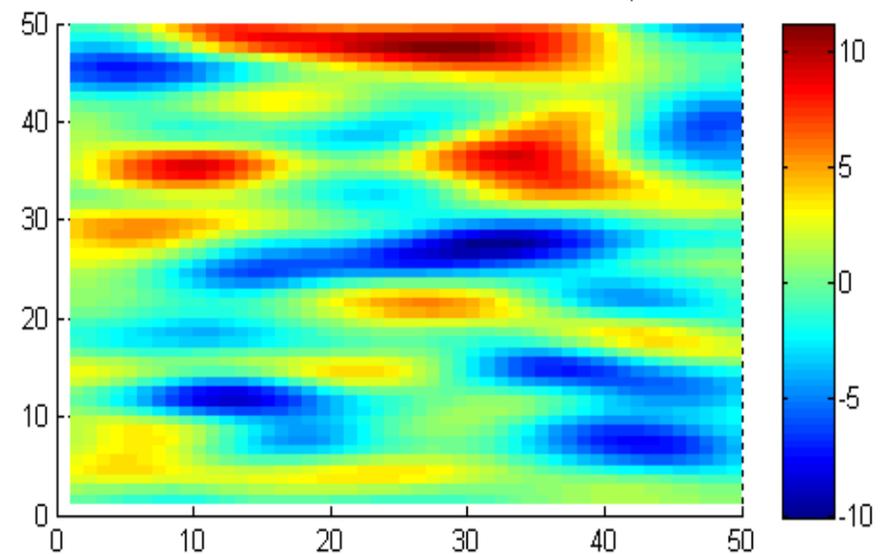
$$c(h) = \sigma^2 \exp\left(-\left(\frac{h_1}{\theta_1}\right)^2 - \left(\frac{h_2}{\theta_2}\right)^2\right) \text{ avec } \theta_1 \neq \theta_2$$

Simulation trajectoires PG de covariance gaussienne isotrope avec $\theta = 0.1$



Simulation d'un champ 2D isotrope
(covariance gaussienne)

Simulation trajectoires PG de covariance gaussienne avec $\theta_1 = 0.1$ et $\theta_2 = 0.03$



Simulation d'un champ 2D anisotrope
(covariance gaussienne)

Construction du métamodèle PG

➤ Choix de la fonction de covariance/corrélation

▪ Effet de pépité

$$\tilde{c}(h) = \tilde{c}(h) + \varepsilon^2 1_{h=0}$$

- Discontinuité de la covariance à l'origine
- Bruit blanc additionnel
- Géostatistique : modélisation d'hétérogénéités locales
- **Pour le métamodèle PG :**
 - Relaxation de la contrainte d'interpolateur exact des données
 - Amélioration du conditionnement de la matrice de covariance (rajout d'un terme sur la diagonale)

Construction du métamodèle PG

➤ Estimation des paramètres du métamodèle PG

- Paramètres de la moyenne β
- Paramètre de variance σ^2
- Paramètres de la covariance θ : appelés « hyperparamètres »

➔ Estimation par maximum de vraisemblance

- Log Vraisemblance sur la BA :

$$\text{Ln}L(Z_S, \beta, \sigma, \theta) = -\frac{N}{2} \text{Ln}(2\pi\sigma^2) - \frac{1}{2} \text{Ln}(\det \Sigma_S) - \frac{1}{2} \sigma^2 (Z_S - F_S)^T \Sigma_S^{-1} (Z_S - F_S)$$

- Maximisation :

$$(\beta^*, \sigma^*, \theta^*) = \underset{(\beta, \sigma, \theta)}{\text{Argmax}} \text{Ln}L(Z_S, \beta, \sigma, \theta)$$

Construction du métamodèle PG

➤ Estimation des paramètres du métamodèle PG

- Si f_0 de forme linéaire

$$f_0(x) = \beta_0 + \sum_{i=1}^k \beta_i x_i \Rightarrow F_S = \tilde{X}_S \beta$$

$$\text{avec } \tilde{X}_S = [1, x_1^T, \dots, x_k^T] \text{ et } \beta = [\beta_0, \beta_1, \dots, \beta_k]$$

=> Solutions analytiques pour β^* et σ^{2*} en fonction de θ

$$\beta^*(\theta) = \left(\tilde{X}_S^T \Sigma_{s,\theta}^{-1} \tilde{X}_S \right)^{-1} \Sigma_{s,\theta}^{-1} \left(\tilde{X}_S^T \Sigma_{s,\theta}^{-1} Z_S \right)$$

$$\sigma^{2*}(\theta) = \frac{1}{N} \left(Z_S - \tilde{X}_S \beta^*(\theta) \right)^T \Sigma_{s,\theta}^{-1} \left(Z_S - \tilde{X}_S \beta^*(\theta) \right)$$

=> Estimation des hyperparamètres θ^* (injection de σ^{2*} dans LnL)

$$\theta^* = \underset{\theta}{\text{Argmax}} \varphi(\theta) \quad \text{avec} \quad \varphi(\theta) = \left| \Sigma_{s,\theta} \right|^{-\frac{1}{N}} \sigma^{2*}(\theta)$$



Optimisation numérique pour l'estimation des θ^*

Construction du métamodèle PG

➤ Etapes de construction du métamodèle PG :

- Choix de la fonction moyenne
- Choix de la fonction de covariance
- Estimation des paramètres :
 - paramètres de la moyenne et hyperparamètres
 - par maximum de vraisemblance sur la base d'apprentissage
 - Par validation croisée (Dubrule [1983], thèse de F. Bachoc CEA Saclay)
- Construction métamodèle PG conditionnel aux points de la BA :
 - Moyenne = prédicteur
 - Variance = MSE du prédicteur

$$\hat{Z}(x^*) = f_0(x^*) + k(x^*)^T \Sigma_S (Z_S - F_S)$$

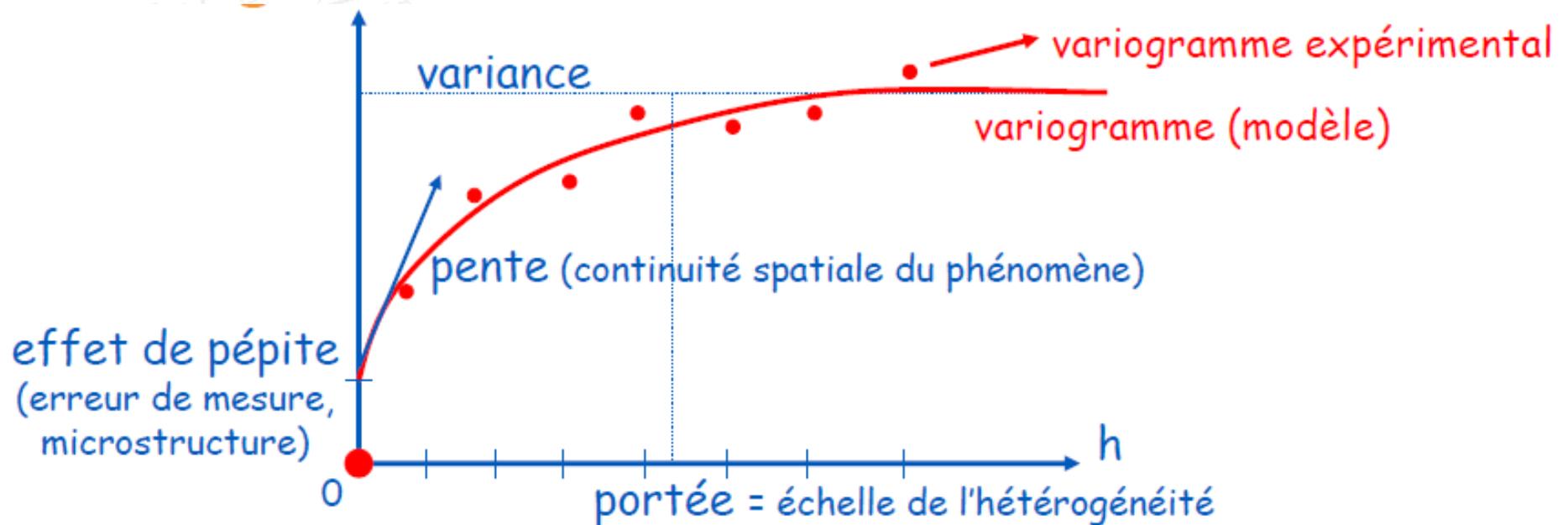
$$MSE[\hat{Z}(x^*)] = \hat{\sigma}^2 - k(x^*)^T \Sigma_S k(x^*)$$

Analogie avec le krigeage

➤ Covariance et variogramme

$$\text{Variogramme : } \gamma(h) = 0.5 \text{ Var}[Z(x+h) - Z(x)]$$

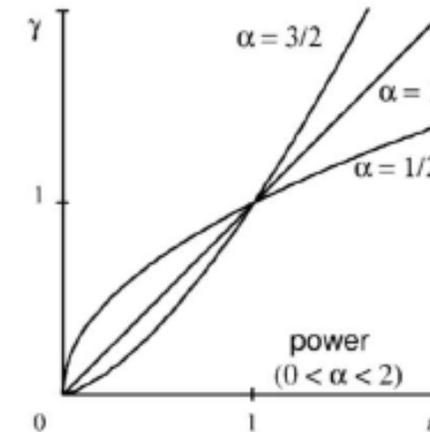
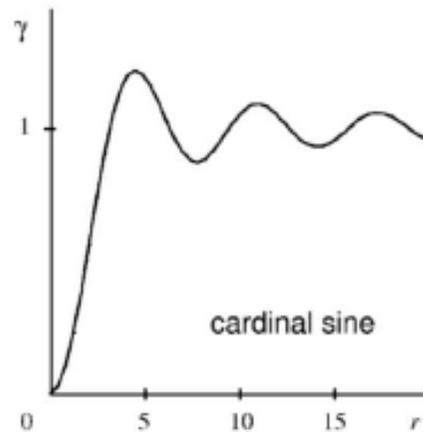
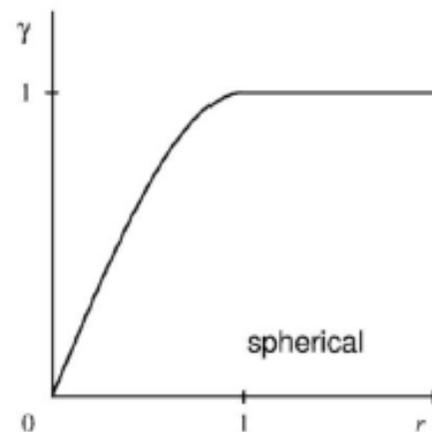
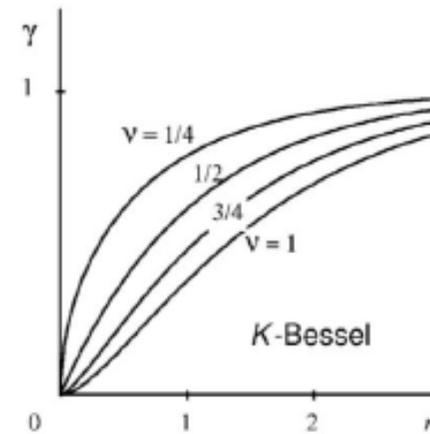
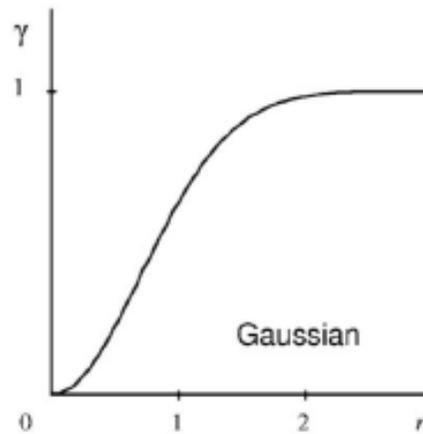
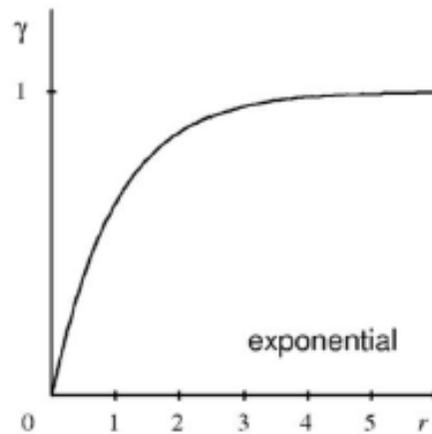
- Variogramme d'un processus stationnaire : $\gamma(h) = C(0) - C(h)$



Analogie avec le krigeage

➤ Covariance et variogramme

Modèles basiques de variogramme



[Chilès]

PLAN

● Rappels statistiques et théorie

- Variable, vecteur et processus gaussien

● Définition du métamodèle processus gaussien (PG)

- Du krigeage à la régression par PG
- Formulation analytique
- Analogie avec le krigeage

● Construction du métamodèle PG

- Choix de la fonction moyenne
- Choix de la fonction de covariance
- Estimation des paramètres

● Utilisation du métamodèle PG

- Validation du métamodèle PG
- Illustration sur des exemples analytiques
- Plans d'expériences et planification adaptative

Utilisation du métamodèle PG

➤ Validation du métamodèle PG

- **Qualité d'ajustement parfaite des données de la base d'apprentissage (BA):** car interpolateur exact des points de la base d'apprentissage
- **Qualité de prédiction pour un nouveau point ?? => prédictivité ?**
 - 2 possibilités :
 - **Utilisation d'une base de test :** points différents de la BA et non utilisés pour la construction du métamodèle
 - **Approche par validation croisée sur la BA :**
 - « **leave one out** » (LOO) (Hastie et al. 1990) : on enlève un point de la BA, on construit le métamodèle avec les points restants, puis on compare la prédiction du métamodèle pour le point enlevé avec sa vraie valeur. Répétition pour les n points de la BA => n prédictions et donc n résidus
 - **Validation croisée par blocs**

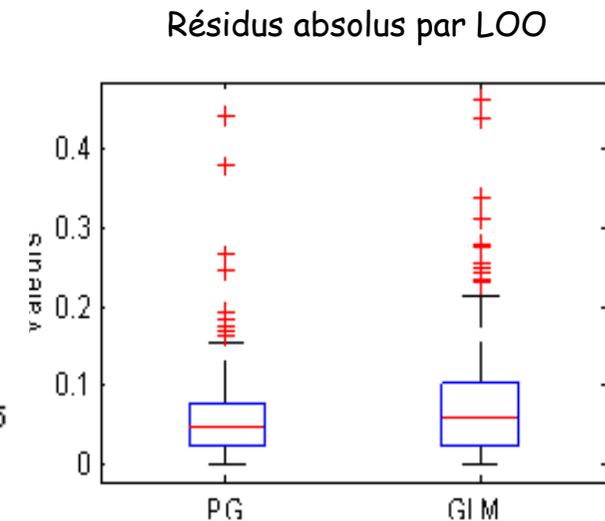
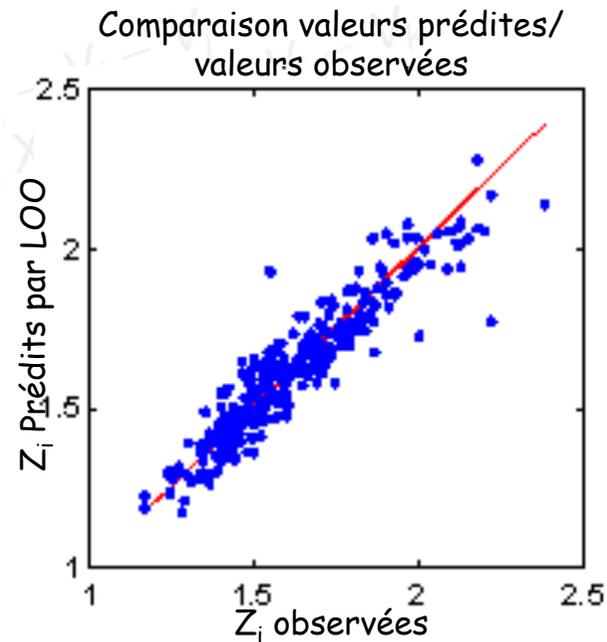
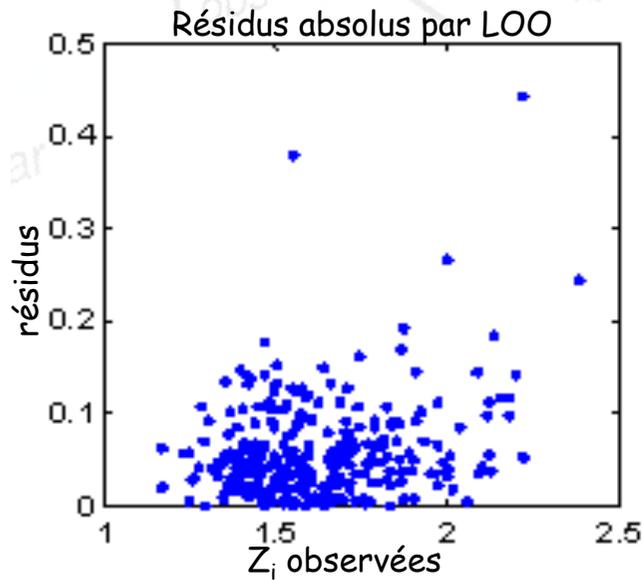


Comparaison des valeurs observées et des valeurs prédites: calcul des résidus

Utilisation du métamodèle PG

➤ Validation du métamodèle PG

- ✓ Estimation des prédictions \hat{Z}_i : sur une base de test ou par **validation croisée**
- ✓ Validation graphique : **Grphe des résidus, comparaison valeurs prédites/valeurs observées**



Utilisation du métamodèle PG

➤ Validation du métamodèle PG

✓ Validation quantitative :

▪ Calcul du coefficient de prédictivité : Q_2

≈ 1 - écart quadratique normalisé entre prédictions et valeurs réelles

$$Q_2(Z, \hat{Z}) = 1 - \frac{\sum_{i=1}^N (Z_i - \hat{Z}_i)^2}{\sum_{i=1}^N (\bar{Z} - Z_i)^2} \quad \text{avec} \quad \begin{cases} Z_i : \text{valeurs observées} \\ \hat{Z}_i : \text{valeurs prédites par le métamodèle} \\ \bar{Z} : \text{moyenne des valeurs observées} \end{cases}$$

▪ Interprétation du Q_2 :

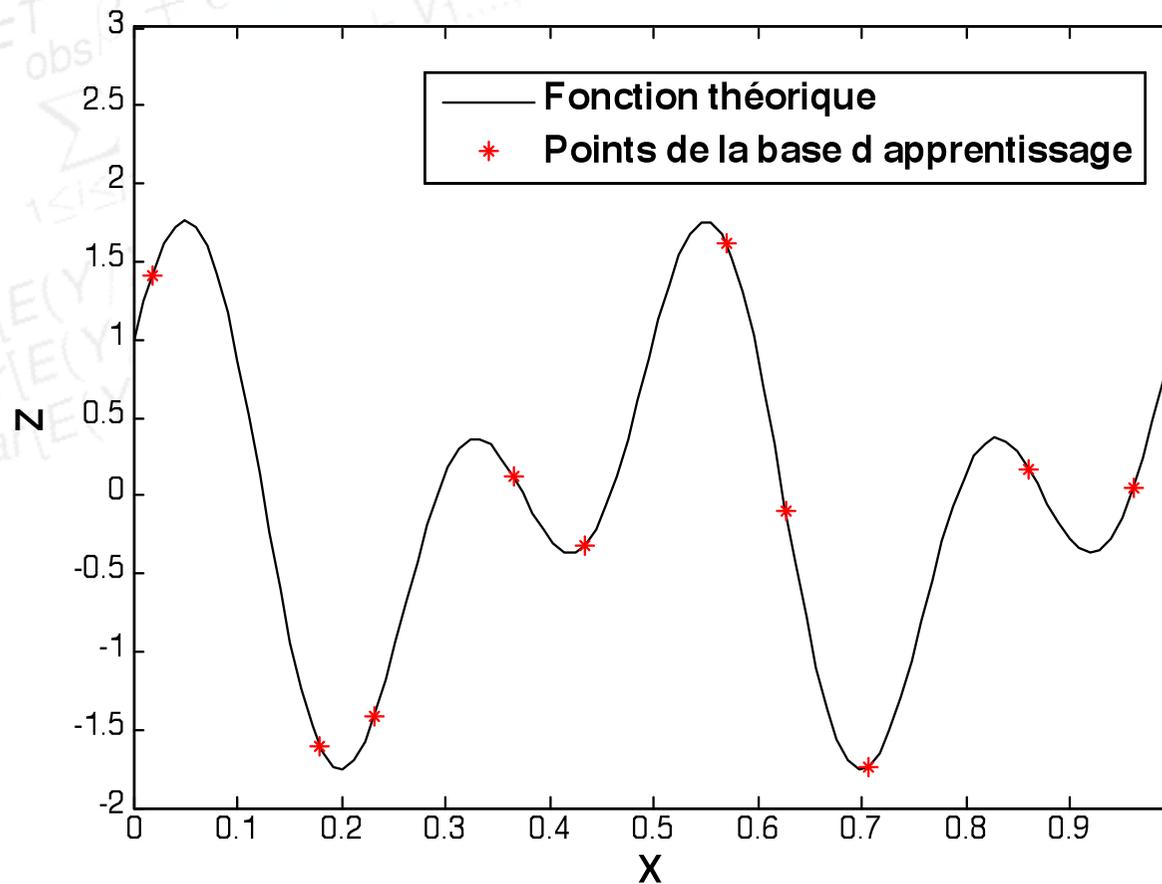
« Plus il est proche de 1, meilleur est l'ajustement du modèle aux observations ».

= Part de la variance de Y expliquée par le métamodèle.

Utilisation du métamodèle PG

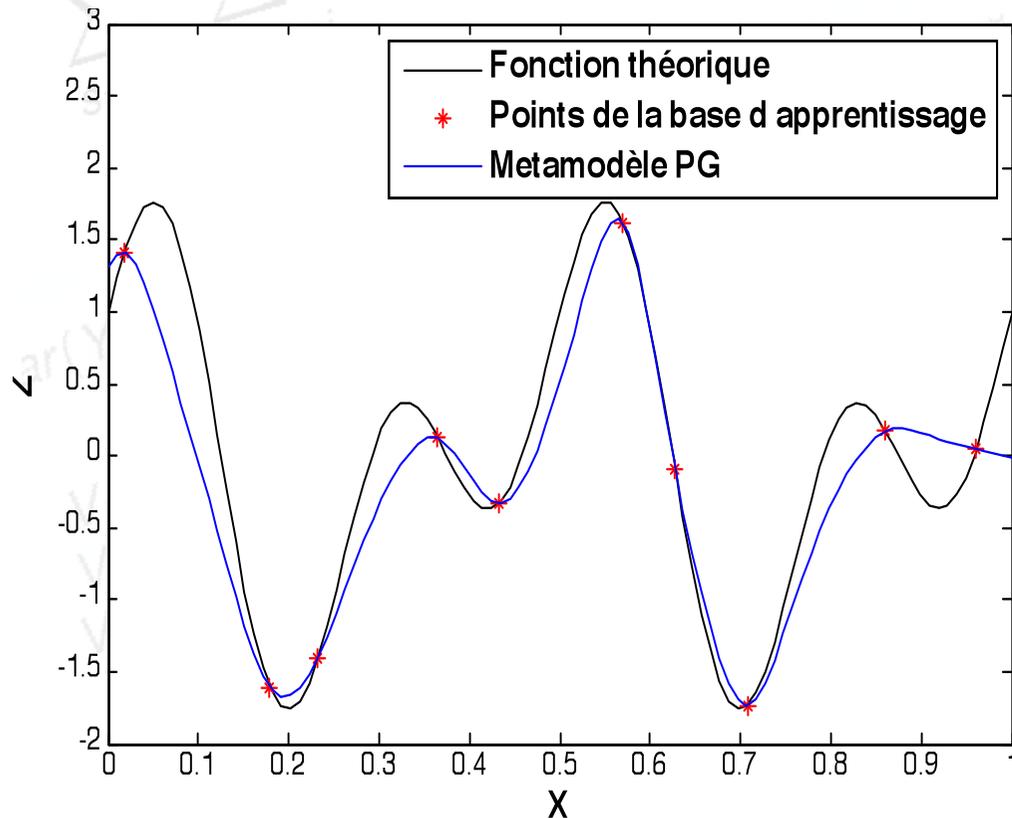
➤ Illustration sur des exemples analytiques

- **Fonction théorique** : $f(x) = \cos(4\pi X) + \sin(8\pi X)$
- **Base d'apprentissage** : tirage aléatoire de 10 points



Utilisation du métamodèle PG

- Illustration sur des exemples analytiques $f(x) = \cos(4\pi X) + \sin(8\pi X)$
 - Ajustement d'un métamodèle PG : avec covariance Matern 5/2



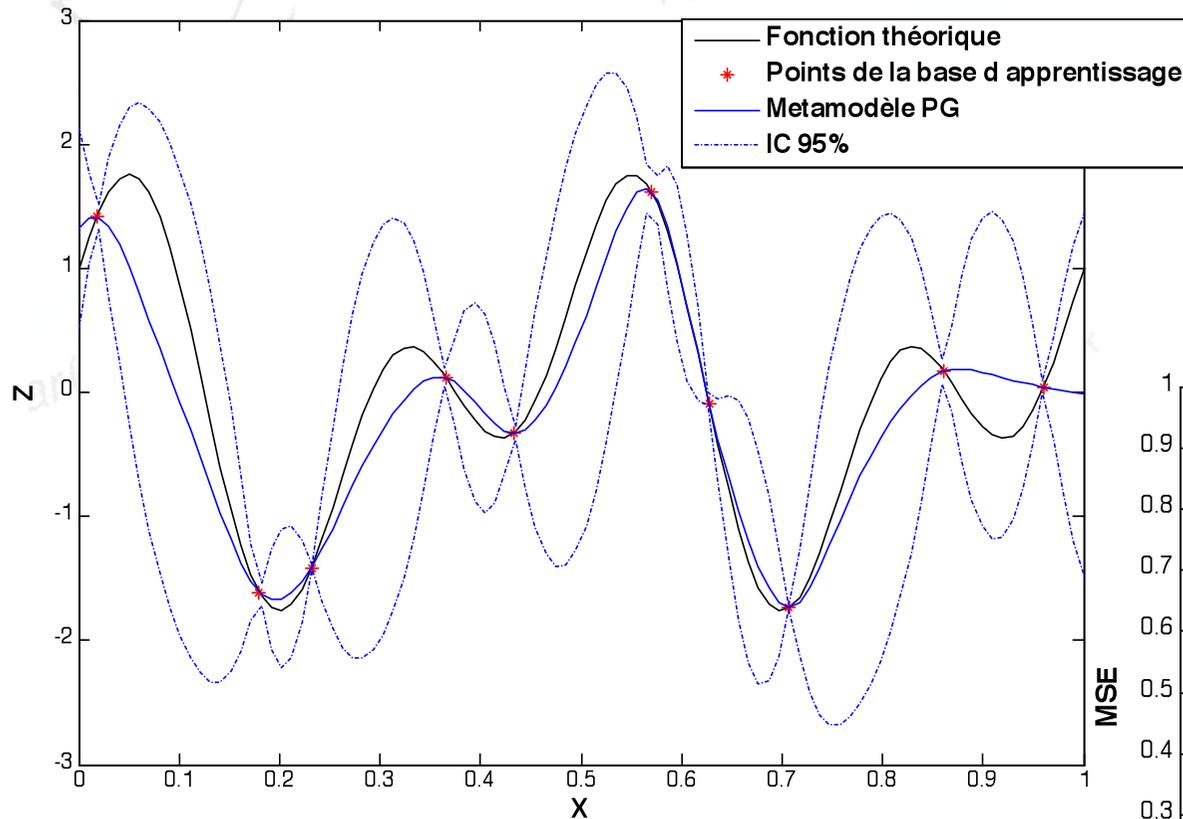
Calcul du Q^2 sur une base de test de 100 points :

$$Q^2 = 0.85$$

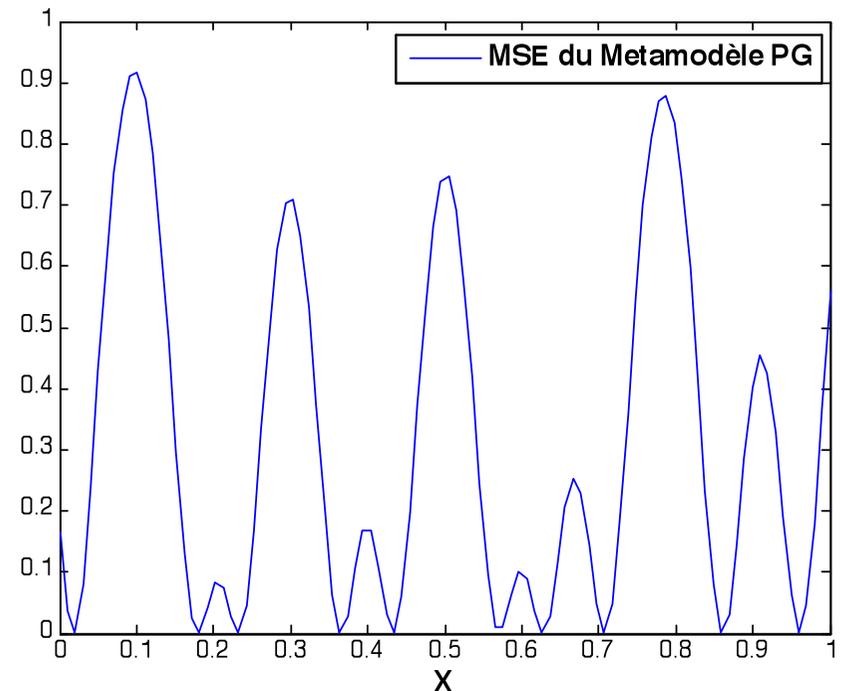
Utilisation du métamodèle PG

➤ Illustration sur des exemples analytiques $f(x) = \cos(4\pi X) + \sin(8\pi X)$

▪ Utilisation du MSE du métamodèle PG



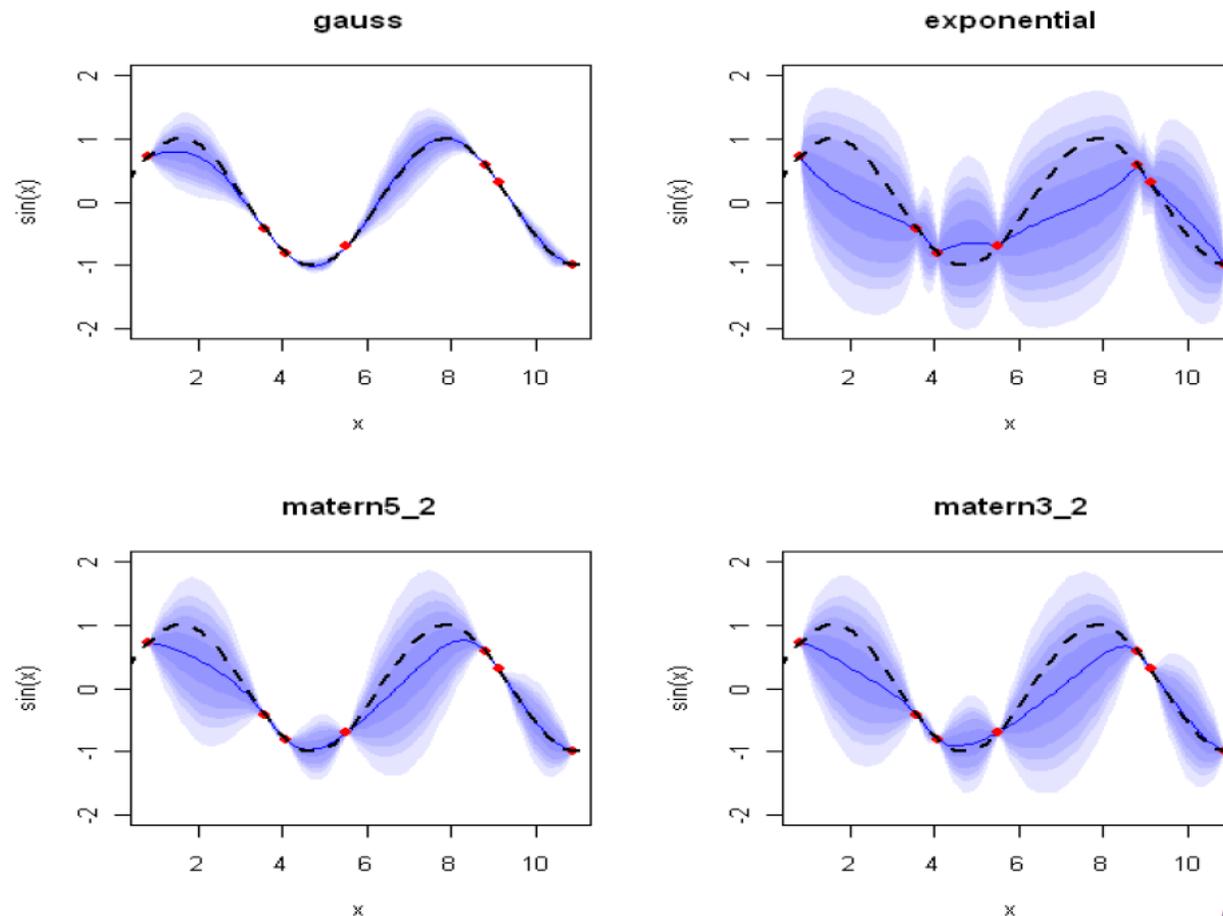
Construction d'intervalles de confiance pour les prédictions



Utilisation du métamodèle PG

➤ Illustration sur des exemples analytiques

- **Fonction théorique** : $Y = \sin(X)$
 - **Base d'apprentissage** : 7 points tirés aléatoirement
- ⇒ **Impact du choix de la fonction de covariance**



[Chevalier, 2011]

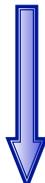
Utilisation du métamodèle PG

➤ Avantages du métamodèle PG

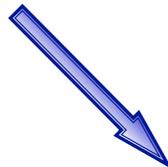
- **Interpolateur exact** des points de la base d'apprentissage (relaxation possible si ajout d'un effet de pépité)
- **Possibilité de « jouer » sur la forme de la fonction moyenne**
=> moyenne du processus devient une fonction de x (krigeage universel)
=> prise en compte d'avis d'expert
- **Possibilité d'estimer les paramètres par maximum de vraisemblance**
=> extension + facile en grande dimension
- **Formulation analytique : exploitation de la loi du prédicteur**
=> Simulation de la loi gaussienne conditionnelle
- **Rapidité d'évaluation du prédicteur**



Analyse de sensibilité



Propagation d'incertitude



Calibration

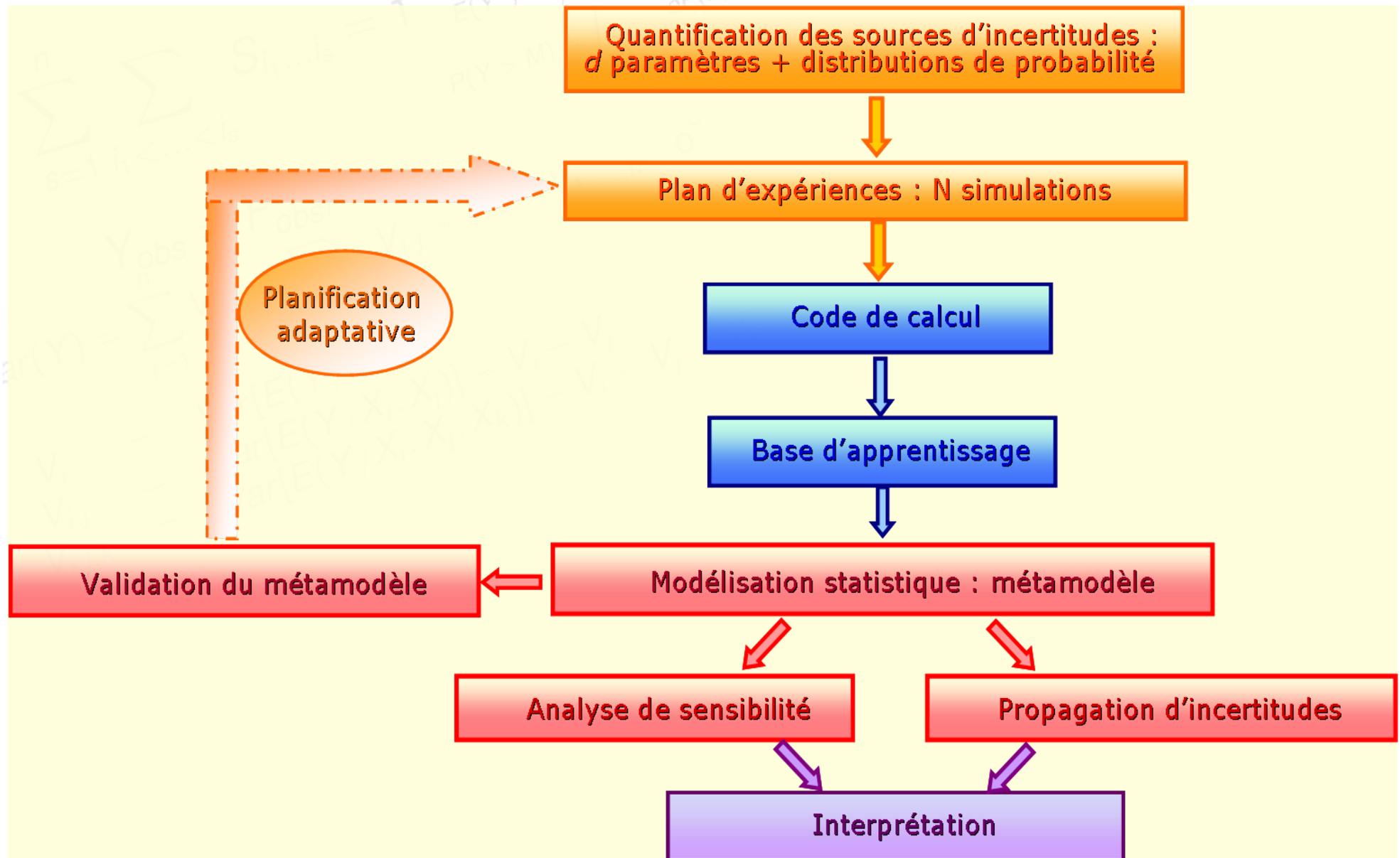
Utilisation du métamodèle PG

➤ Inconvénients du métamodèle PG

- **Choix des fonctions moyenne et de covariance**
- **Estimation des paramètres de régression et de covariance**
=> Estimation en grande dimension
- **Pas adapté quand beaucoup de données**
inversion de la matrice de covariance coûteuse => optimisation de la vraisemblance difficile
=> solution possible : réduction de la base d'apprentissage (cf. présentation de D. Ginsbourger)
- **Prédicteur à tendance à lisser**
=> pas adapté à l'estimation de valeurs extrêmes, de seuil...

Utilisation du métamodèle PG

➤ Utilisation du métamodèle PG dans une démarche « incertitudes »



Utilisation du métamodèle PG

➤ Choix du plan d'expériences pour la base d'apprentissage

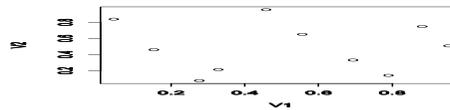
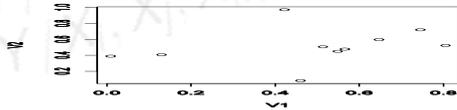
▪ Métamodèle PG adapté :

- au cas des modèles numériques \Rightarrow expériences « déterministes »
- quand quelques dizaines d'entrée et quelques centaines de simulations possibles \Rightarrow points d'expériences dispersés \Rightarrow « **sparsity** »

▪ Plan d'expériences recommandés dans ces cas :

- **Space-filling design** : couverture optimale du domaine de variation des paramètres incertains

Simple
Random
Sample



Space
Filling
Design

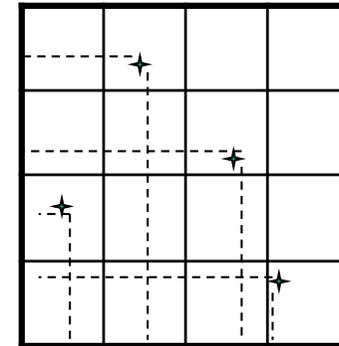
Utilisation du métamodèle PG

➤ Choix du plan d'expériences pour la base d'apprentissage

- **Space-filling design de type hypercubes latins (LHS)** : projections uniformes sur les marginales

⇒ compromis entre "étalement" des points et répartition uniforme

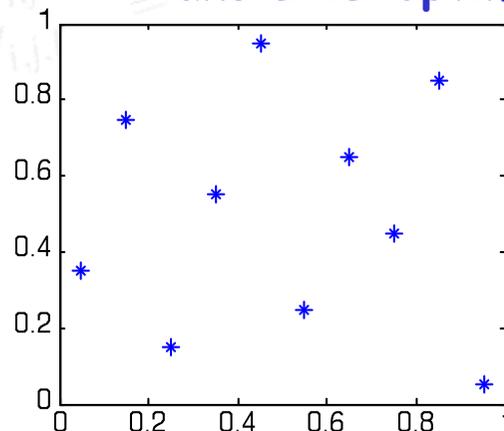
On divise chaque dimension en N intervalles et on prend un point dans chaque intervalle



plan LHS de 4 points en dimension 2

- **Plans LHS optimisés** : optimisation d'un critère d'occupation de l'espace

ex : critère maximin, minimax, discrédance, ...



plan LHS maximin de 10 points en dimension 2

Rque : Plans optimaux pour l'estimation des hyperparamètres ??

Utilisation du métamodèle PG

➤ Planification adaptative avec le métamodèle PG

▪ Principe :

- Plan d'expériences initial
- Construction d'un premier métamodèle
- Ajout de nouveaux points de manière séquentielle en utilisant le métamodèle ⇒ **planification adaptative**

▪ Mise en œuvre avec le métamodèle PG:

▪ Utilisation d'un critère basé sur le MSE du métamodèle PG

- Objectif : choisir comme nouveau point celui qui améliore le plus la qualité du métamodèle
 - Exemple : point où le MSE du métamodèle PG est maximal => point où l'erreur de prédiction est potentiellement la + grande
 - Autres critères utilisant le MSE du PG : point qui minimise l'intégrale du MSE sur le domaine, ajout par blocs de points, ...

▪ Autres critères liés à l'objectif de l'étude : recherche de minima,...

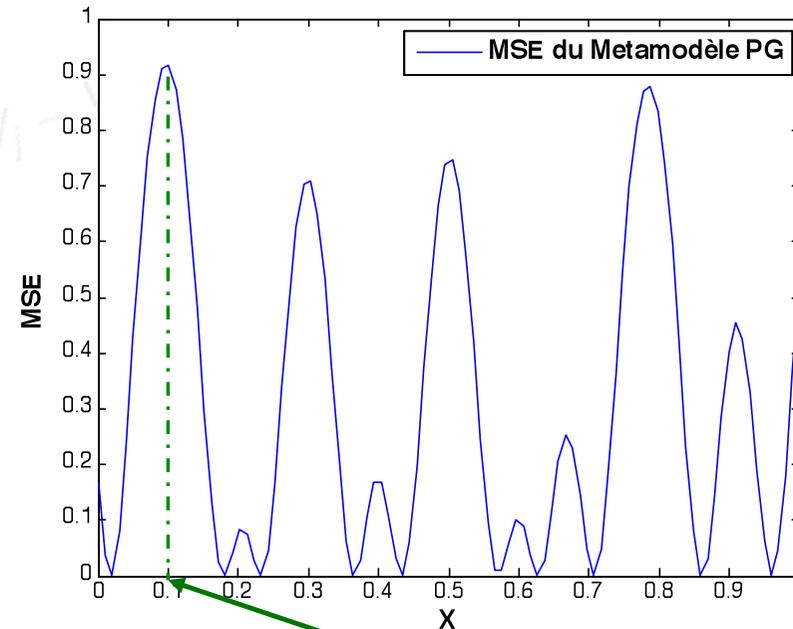
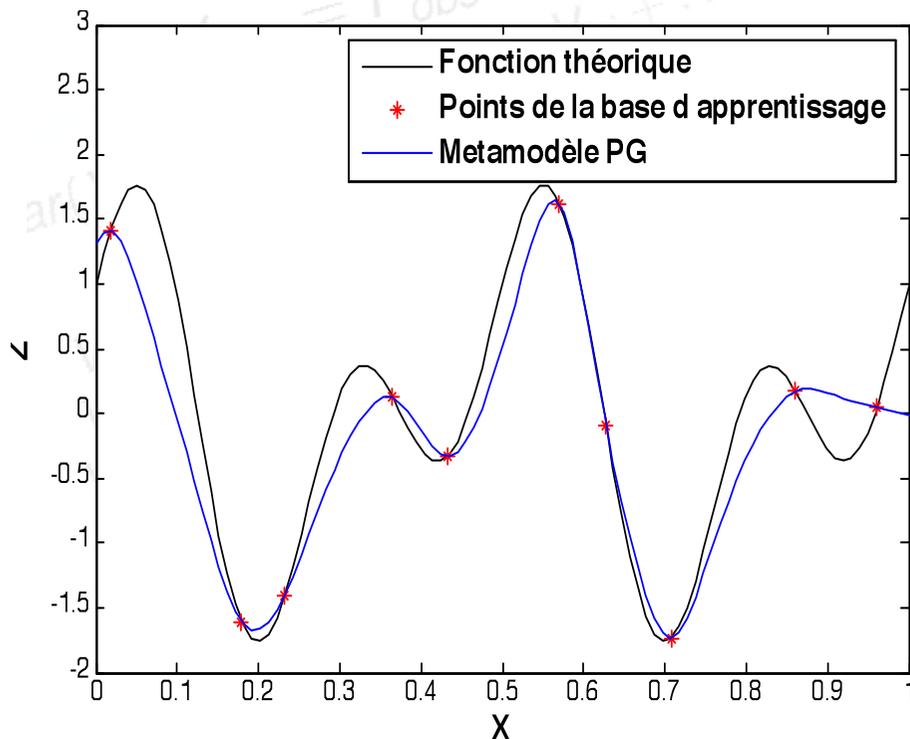
Utilisation du métamodèle PG

➤ Planification adaptative avec le métamodèle PG

■ Illustration : Utilisation d'un critère basé sur le MSE

Nouveau point = celui où le MSE du métamodèle PG est maximal

$$x_{new} = \underset{x}{\operatorname{Argmax}} \operatorname{MSE}[\hat{Z}(x)] = \underset{x}{\operatorname{Argmax}} (\sigma^2 - k(x)^T \Sigma_S k(x))$$



Nouveau point : $x_{new} = 0.0972$

Utilisation du métamodèle PG

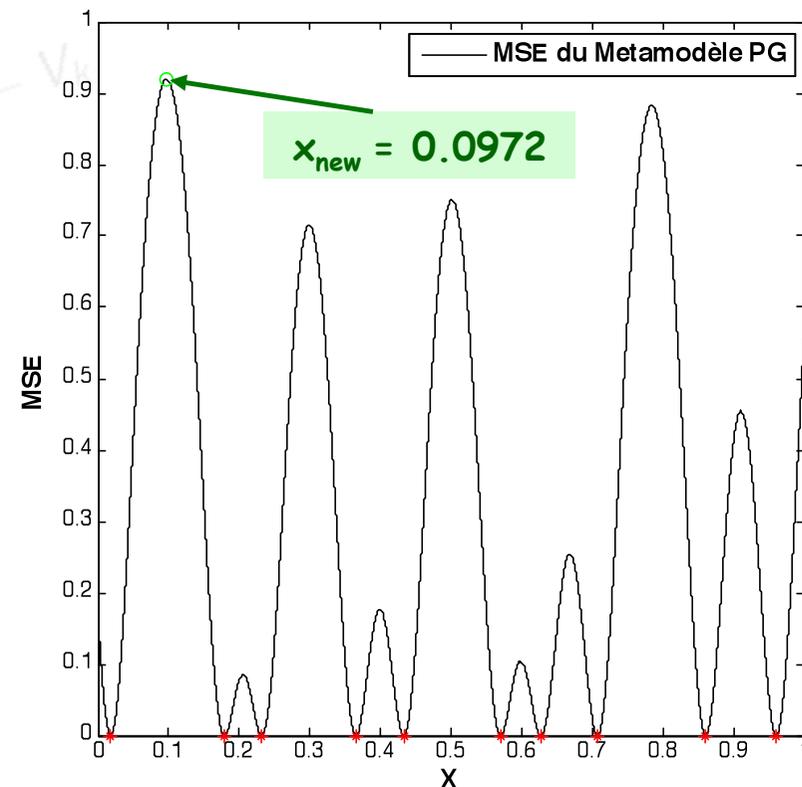
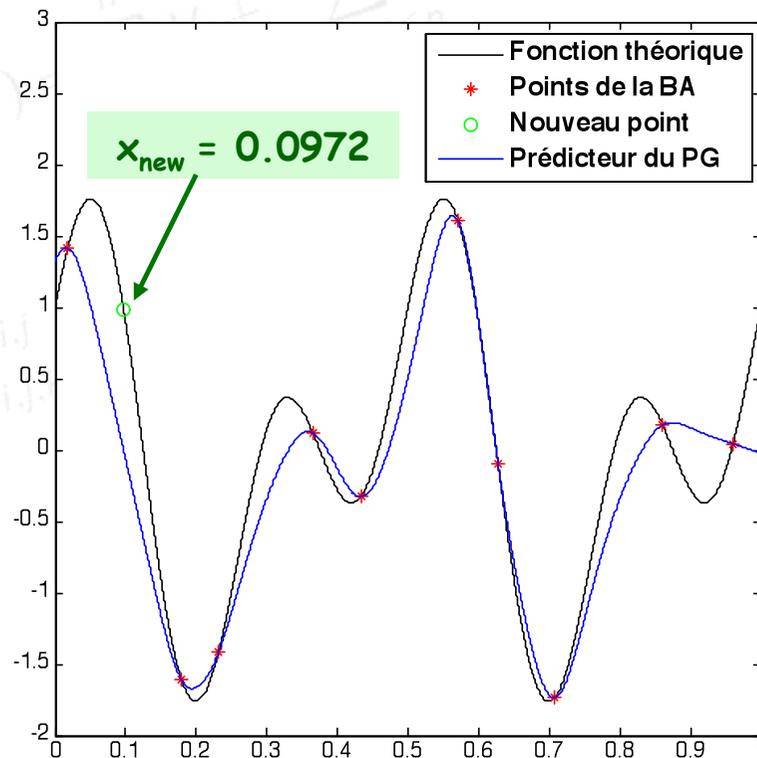
➤ Planification adaptative avec le métamodèle PG

▪ Illustration : Utilisation d'un critère basé sur le MSE

Nouveau point = celui où le MSE du métamodèle PG est maximal

$$x_{new} = \underset{x}{\operatorname{Argmax}} \operatorname{MSE}[\hat{Z}(x)] = \underset{x}{\operatorname{Argmax}} (\sigma^2 - k(x)^T \Sigma_S k(x))$$

Etape 1 : Recherche de x_{new} (maximisation du MSE => pbe d'optimisation)

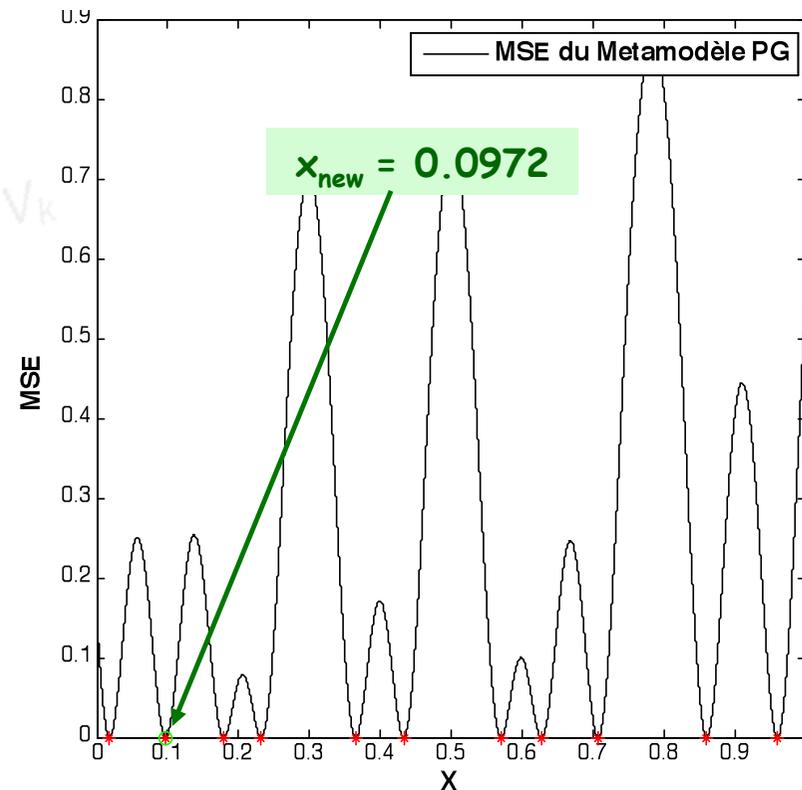
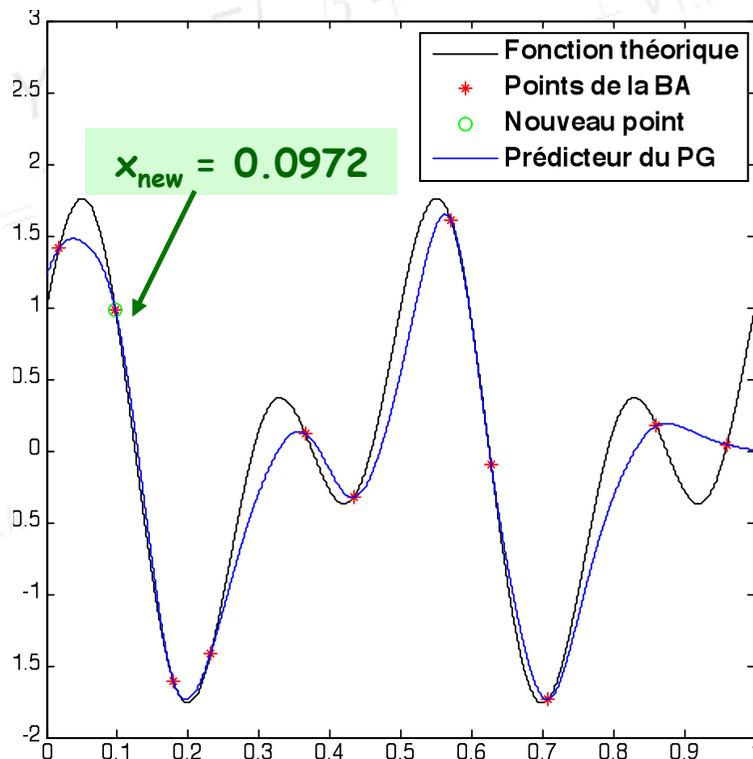


Utilisation du métamodèle PG

➤ Planification adaptative avec le métamodèle PG

Etape 2 : Simulation de $x_{new} \Rightarrow z_{new}$

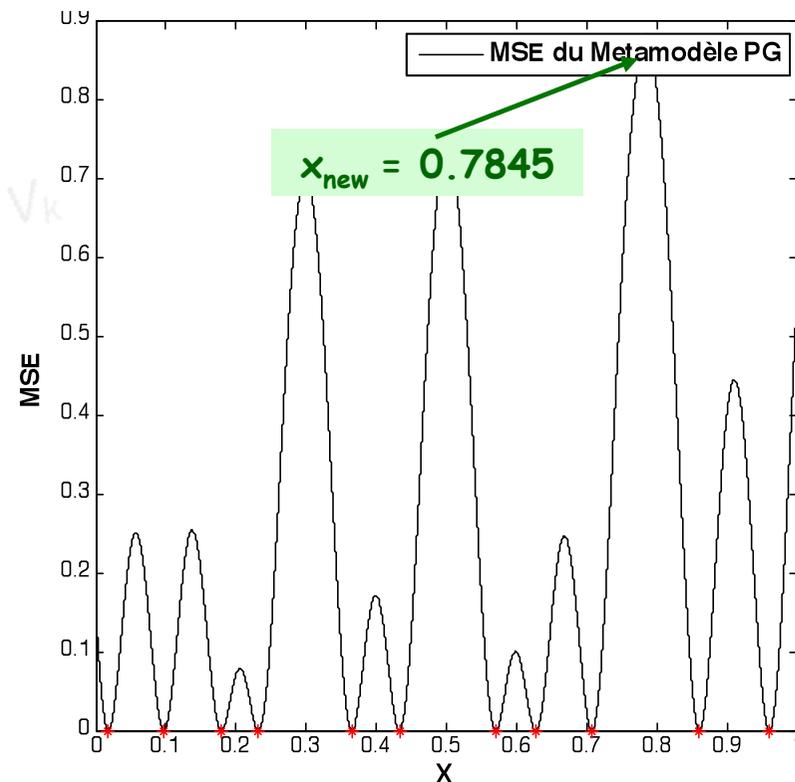
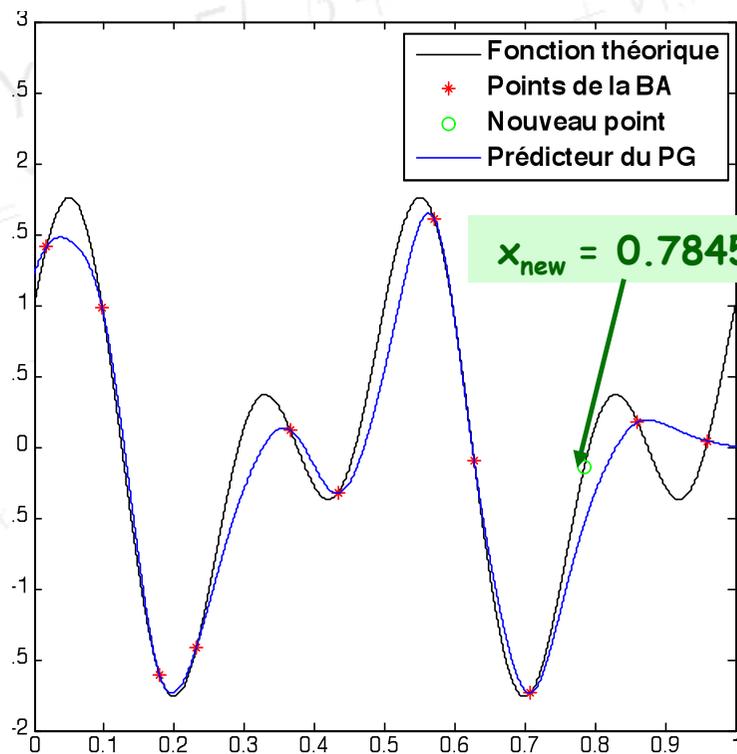
Ajout de (x_{new}, z_{new}) à la BA et mise à jour du métamodèle PG



Utilisation du métamodèle PG

➤ Planification adaptative avec le métamodèle PG

Etape 1 : Recherche de x_{new} (maximisation du MSE => pbe d'optimisation)

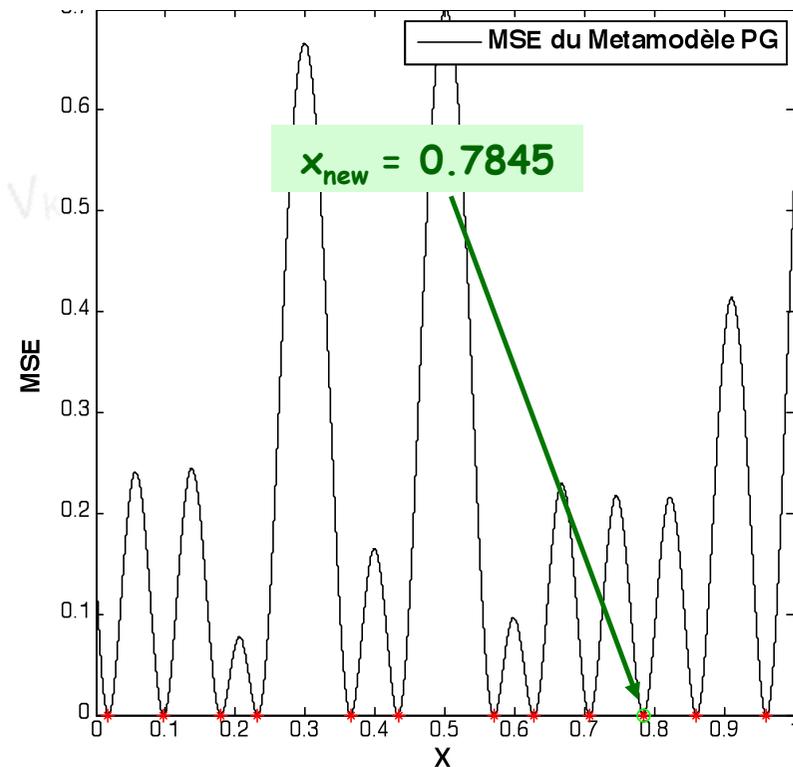
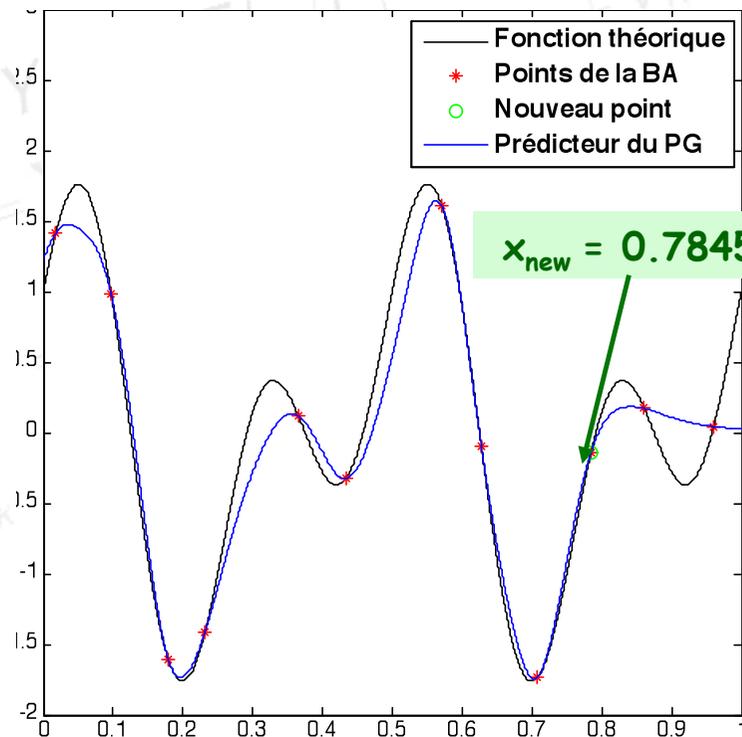


Utilisation du métamodèle PG

➤ Planification adaptative avec le métamodèle PG

Etape 2 : Simulation de $x_{new} \Rightarrow z_{new}$

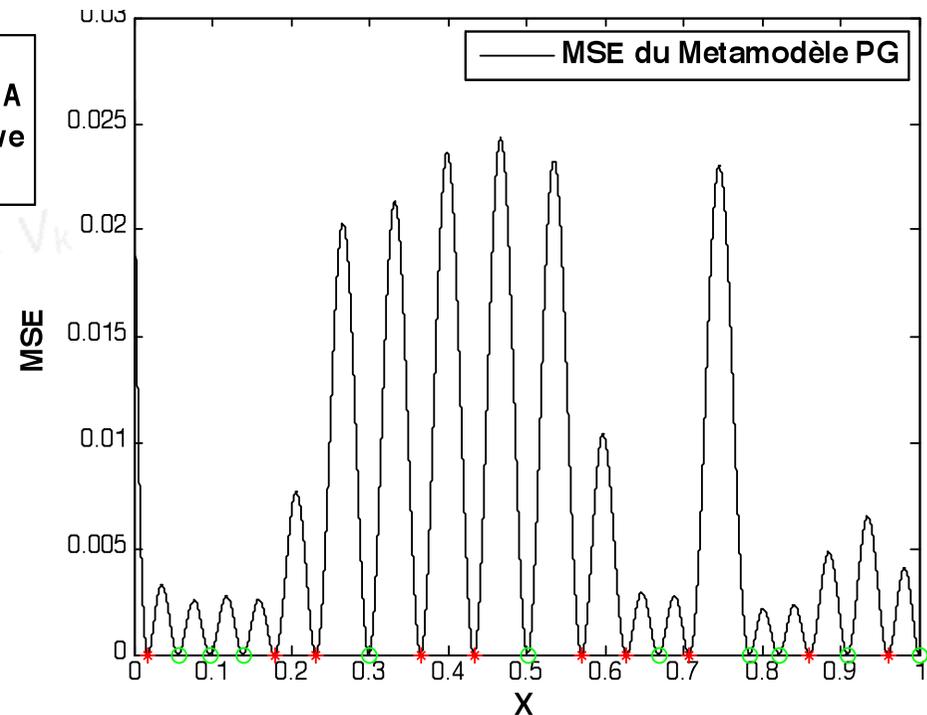
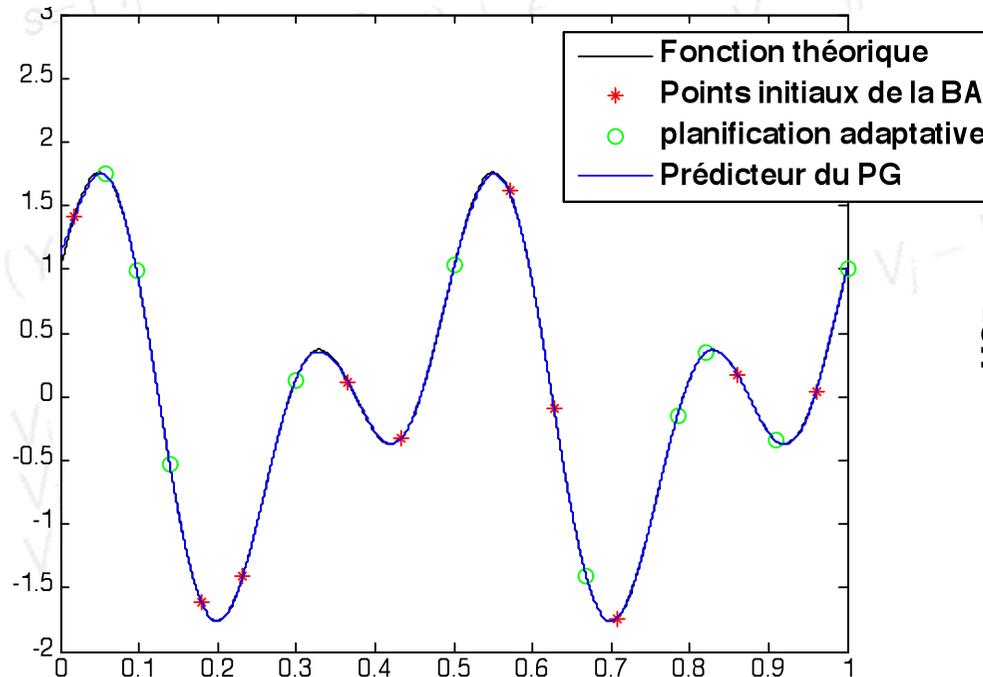
Ajout de (x_{new}, z_{new}) à la BA et mise à jour du métamodèle PG



Utilisation du métamodèle PG

➤ Planification adaptative avec le métamodèle PG

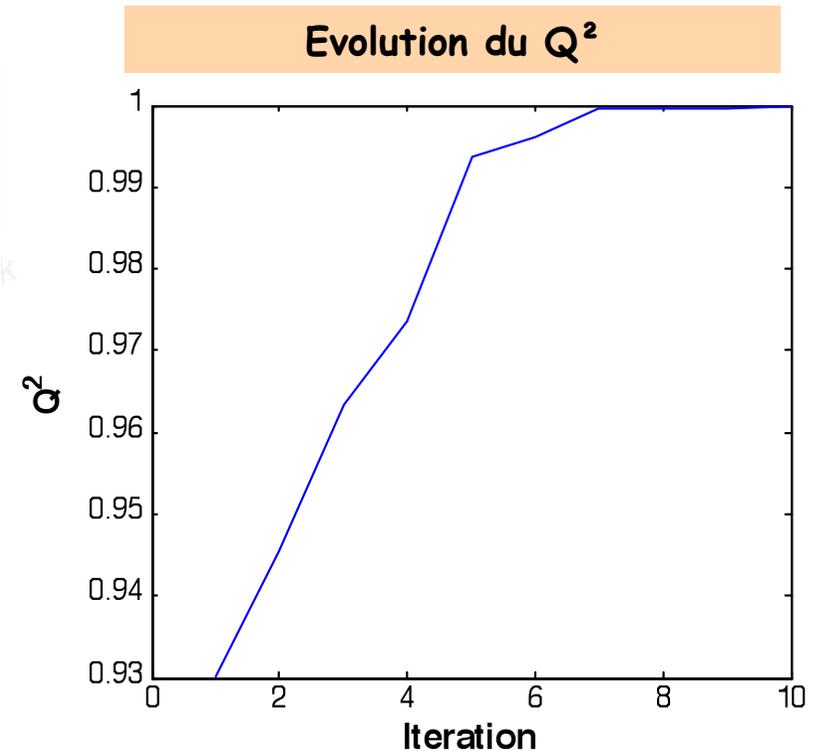
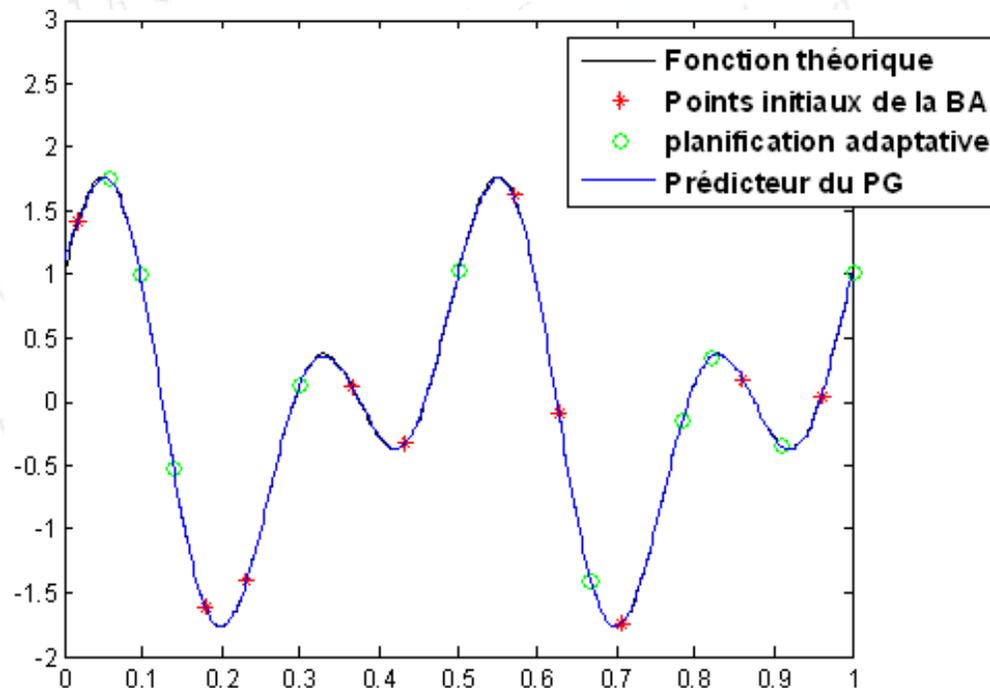
Ajout de 10 points par planification adaptative



Utilisation du métamodèle PG

➤ Planification adaptative avec le métamodèle PG

Ajout de 10 points par planification adaptative



Bibliographie

➤ Krigeage et processus gaussien

▪ Géostatistique :

Krige [1951]. A statistical approach to some basic valuations problems on the witwatersrand. J. Chem. Metall. Min. Soc. 52(6) :119-139.

Matheron [1970]. La théorie des variables régionalisés et ses applications. Les cahiers du centre de Morphologie Mathématique de Fontainebleau, Ecole des Mines de Paris.

Chiles & Delfiner [1999]. Geostatistics : modeling spatial uncertainty. Wiley, New-York.

▪ Extension aux codes de calcul :

Sacks et al. [1989]. Design and analysis of computer experiments. Statistical Science, 4, 409-435.

O'Hagan, A. [2006]. Bayesian analysis of computer code outputs: a tutorial. Reliability Engineering and System Safety 91, 1290-1300.

Rasmussen et al. [2006]. Gaussian Processes for Machine Learning. Cambridge, MA: MIT Press.

Santner et al. [2003]. The Design and Analysis of Computer Experiments. New York: Springer.

Bibliographie

➤ Validation du métamodèle :

Azais & Bardet [2005]. Le modèle linéaire par l'exemple, Dunod.

- **Dans les expériences physiques :**

Droesbecke et al. [1998]. Plans d'expériences (Applications à l'entreprise). Technip, Paris.

- **Dans les expériences numériques :**

Kleijnen, J. & Sargent, R.G. [2000]. A methodology for fitting and validating metamodels in simulation. European Journal of operational Research, 120:14-29.

Meckesheimer et al. [2002]. Computationally inexpensive metamodel assessment strategies. AIA A Journal 40 : 2053-2060.

⇒ **Pratique usuelle** : estimation de critères globaux (erreur quadratique moyenne, erreur en valeur absolue, ...) estimés sur base de test, par validation croisée ou bootstrap.

➤ Analyse de sensibilité avec métamodèle processus gaussien

Oackley & O'Hagan [2004]. Probabilistic sensitivity analysis of computer models : a bayesian approach. Journal of the Royal Statistic Society, Serie B, 66:751-769.

Marrel et al. [2008]. Calculations of the Sobol indices for the Gaussian process metamodel. Reliability Engineering and System Safety, 94:742-751

Packages

➤ Packages métamodèle processus gaussien :

- R : DiceKriging <http://cran.r-project.org/web/packages/DiceKriging/index.html>
- Matlab : DACE www2.imm.dtu.dk/~hbn/dace
- Logiciel GEM <http://www.tonyohagan.co.uk/academic/GEM/index.html>

➤ Package métamodèles

- R : SMURFER <http://www.gdr-mascotnum.fr/iooss1.html>
- R : DiceEval