

Analyses statistiques des expériences numériques

Cours 5 : Krigeage et métamodèle processus gaussien

Bertrand Iooss

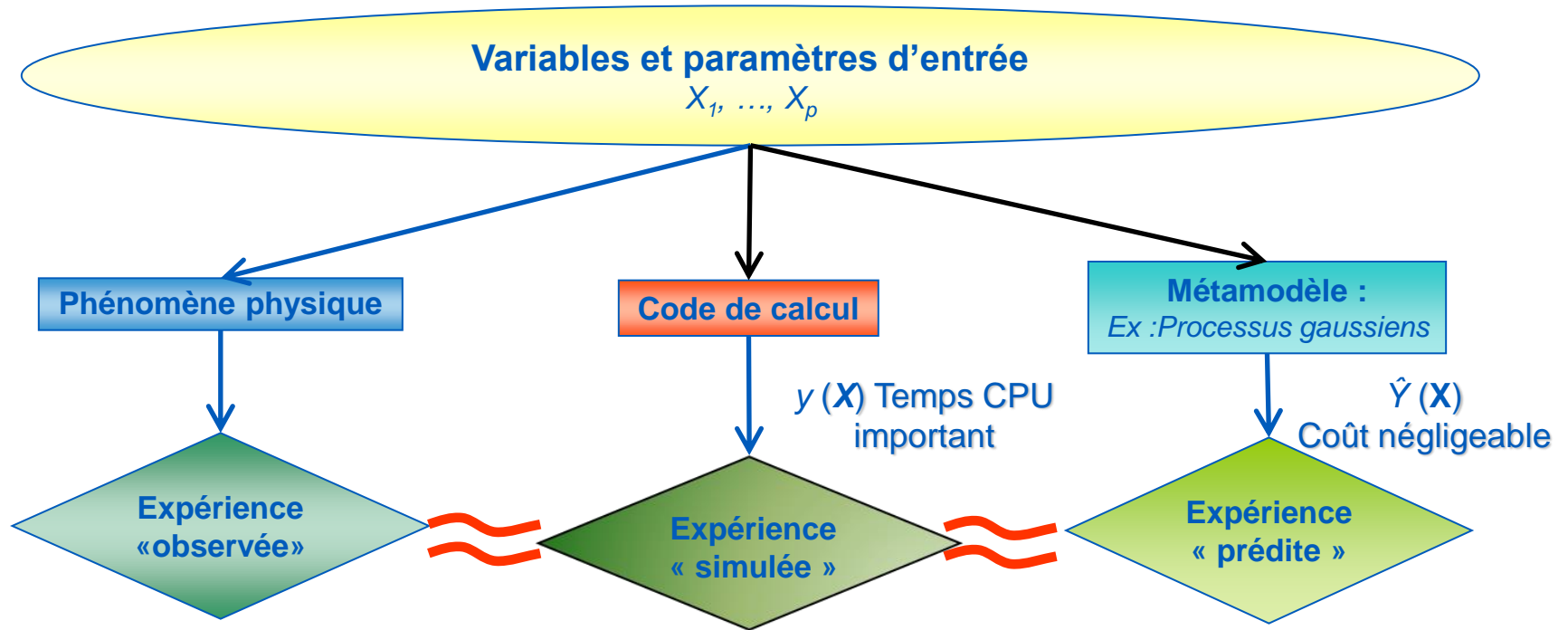
Polytech Nice Sophia

Janvier 2026



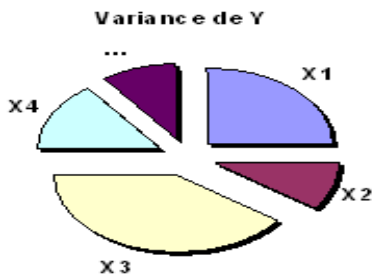
CHANGER L'ÉNERGIE ENSEMBLE

Pourquoi un métamodèle ?

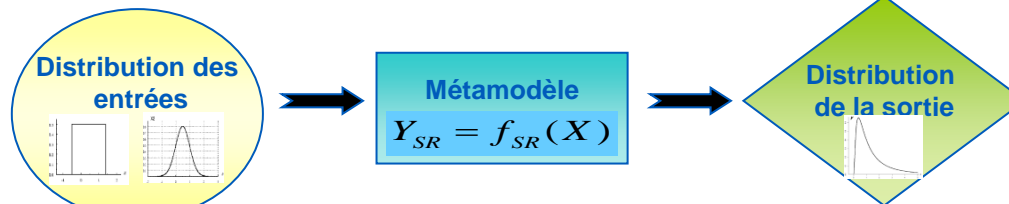


➤ Exploitation du métamodèle (grand nombre de calculs nécessaires)

▪ Analyse de sensibilité



▪ Propagation d'incertitudes



▪ Calibration

Détermination des paramètres
↓
Adéquation expériences simulées et observées

Synonymes : surface de réponse , émulateur, surrogate model, modèle simplifié,

Etapes de la métamodélisation

Planification d'expériences :

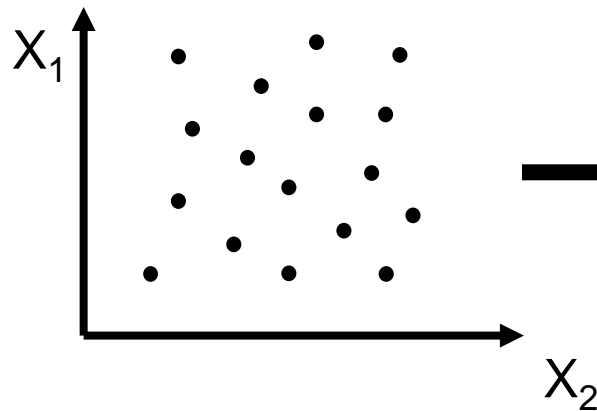
points où l'on va effectuer les simulations (Space Filling Design)

Simulation :

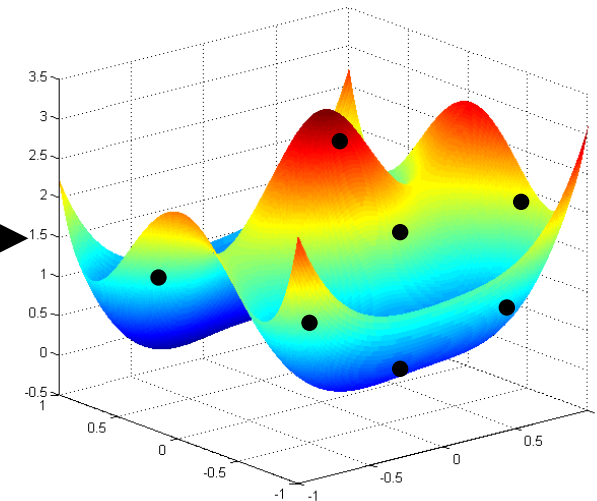
On effectue les simulations

Méta-modélisation :

Approximation de la surface de réponse du code



Code de calcul
Expérience



Méthode : On va choisir parmi une **famille de fonctions**, la fonction qui est la plus proche de la fonction objectif

Différents types de métamodèles

➤ Modèle linéaire, polynômes



➤ Splines

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



➤ Modèles additifs, GAM

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

➤ Arbres de régression, forêts aléatoires

➤ Réseaux de neurones

➤ **Krigeage ou processus gaussiens conditionnels**

Kriging metamodel

Kriging [Matheron 63] for computer codes relies on the idea to interpolate the code outputs in dimension p [Sacks et al. 89] as a spatial cartography

Kriging (or Gaussian process) is interesting because:

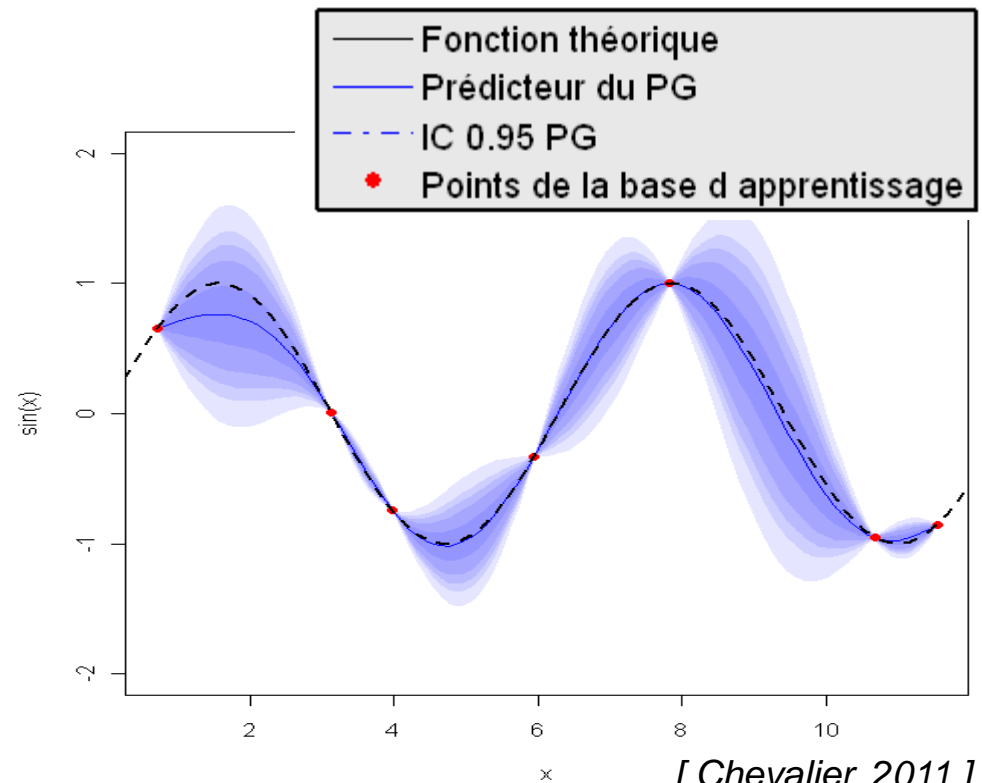
- it interpolates the outputs,
- it gives predictor associated with confidence bands

Example in 1D :

Theoretical function ($p=1$) :

$$Y = G(X) = \sin(X)$$

Simulation of $n = 7$ computation points



[Chevalier, 2011]



Krigeage en stats spatiales

La Géostatistique

Objectifs :

Traitement de **données numériques** à **support spatial** (et/ou temporel) et **quantification des incertitudes**

Contraintes :

Données pouvant **être mesurées en tout point d'un domaine continu**

Aspects principaux :

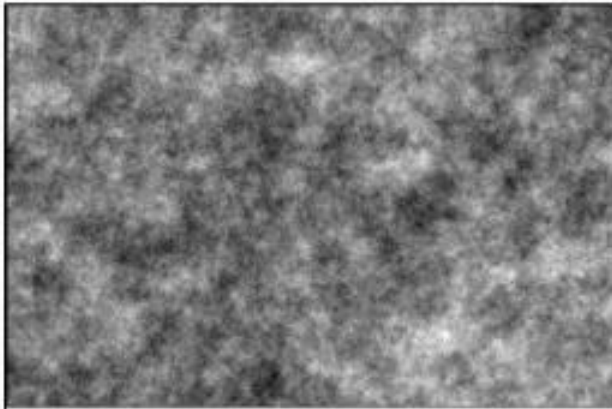
- Prise en compte de la **structure spatiale des données**,
- Échantillonnage peut être irrégulier et incomplet (données fragmentaires),

2 familles de méthodes :

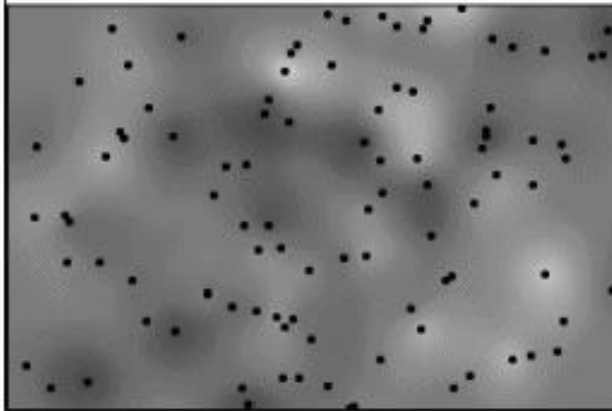
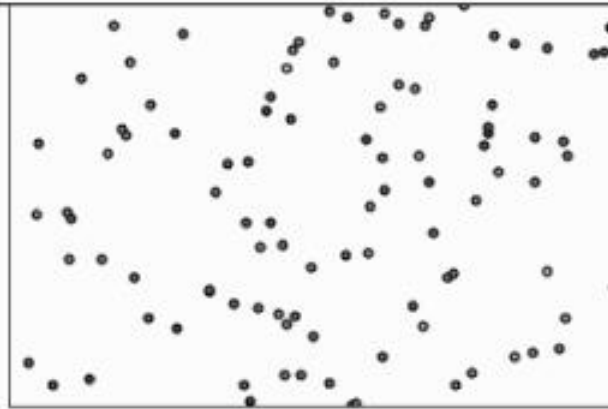
- **Estimation** (prédiction, ...) en un point avec sa précision
- **Simulations** reproduisant la variabilité des échantillons

Exemple : porosité d'un milieu géologique

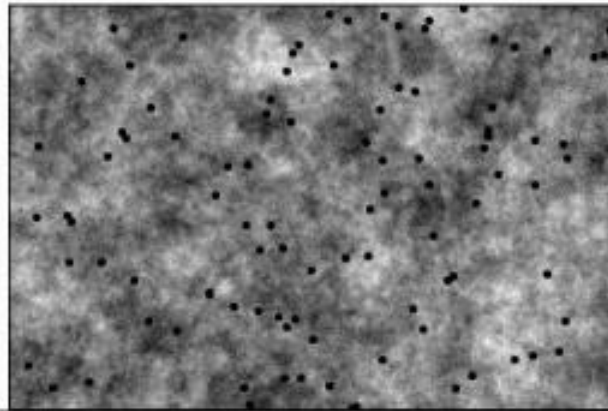
Réalité



Échantillonnage



Estimation optimale
(unique)



Simulation conditionnelle
(infinité de réalisations possibles)

[Chilès]

Utiles pour les calculs qui prennent en entrée ces cartes

N cartes M_i équiprobables $\Rightarrow N$ calculs Monte-Carlo $Y_i = f(M_i)$

Un peu d'histoire

- **1955 : Daniel Krige** (ingénieur de 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** (Ecole des Mines de Paris)

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**

Applications : mines, exploration et production d'hydrocarbures, cartographie géographique, science des matériaux (caractérisation des microstructures)

- **Géostatistique**

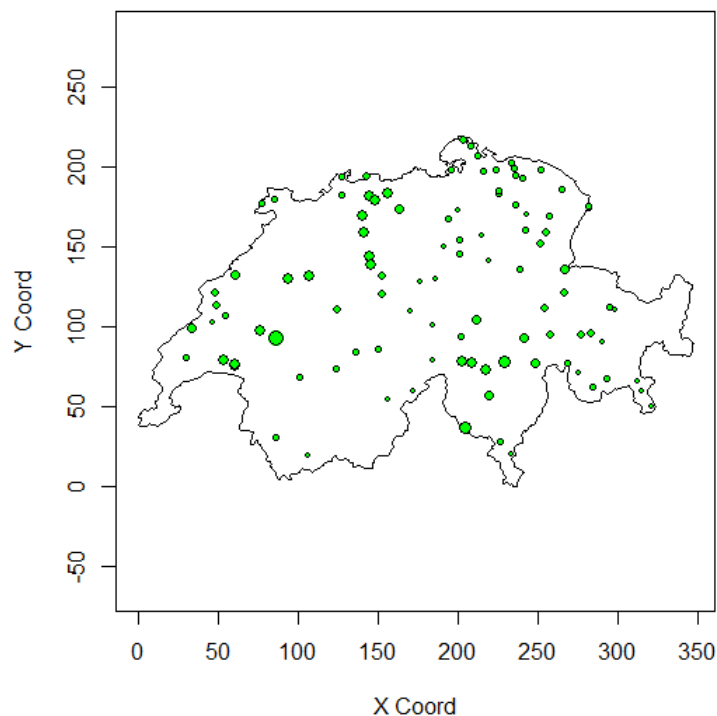
Application de la théorie des fonctions aléatoires à des données localisées dans un espace géographique

=> Méthodes probabilistes pour l'étude de phénomènes corrélés dans l'espace

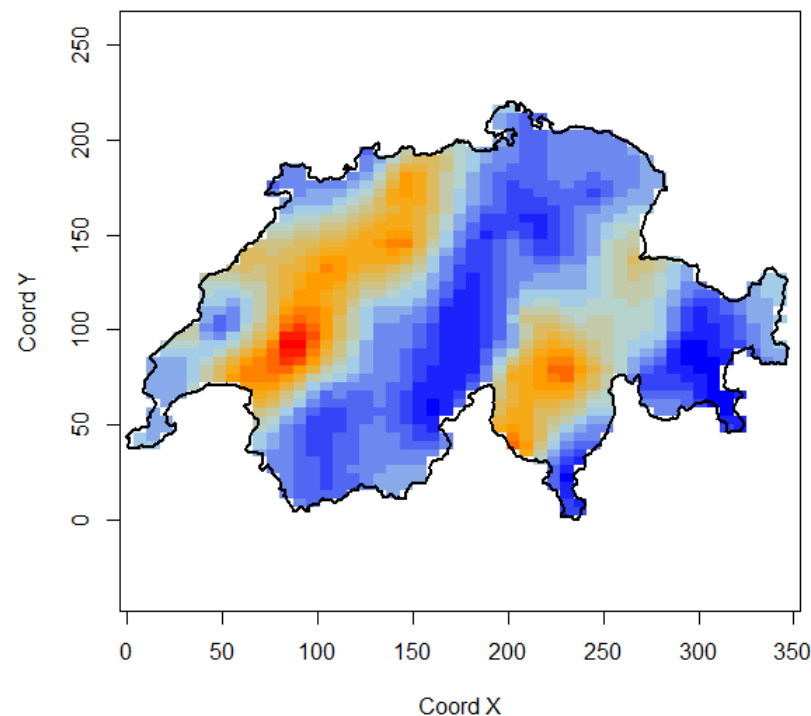
Applications : environnement, pêches industrielles, agriculture, santé, climatologie, ...

Exemple : prévisions pluviométriques

Mesures de précipitations (en mm sur une journée) en Suisse en 100 stations météorologiques

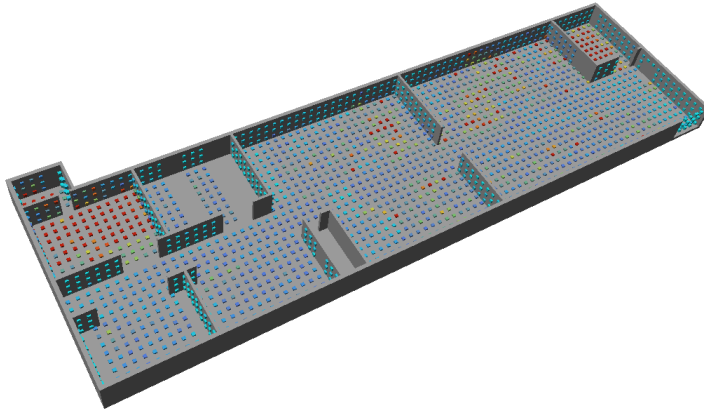


Estimations en tout point du territoire suisse



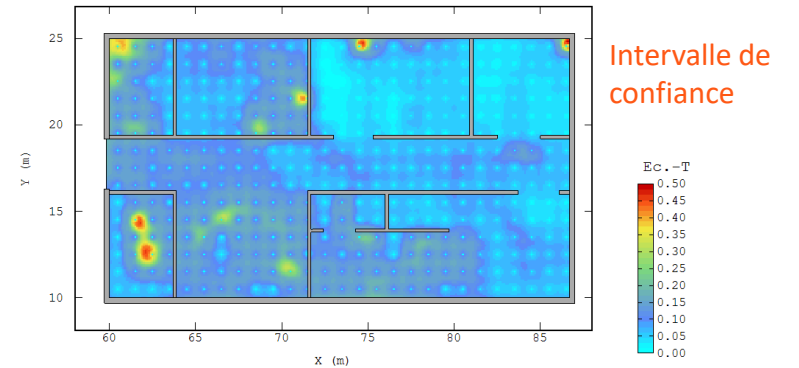
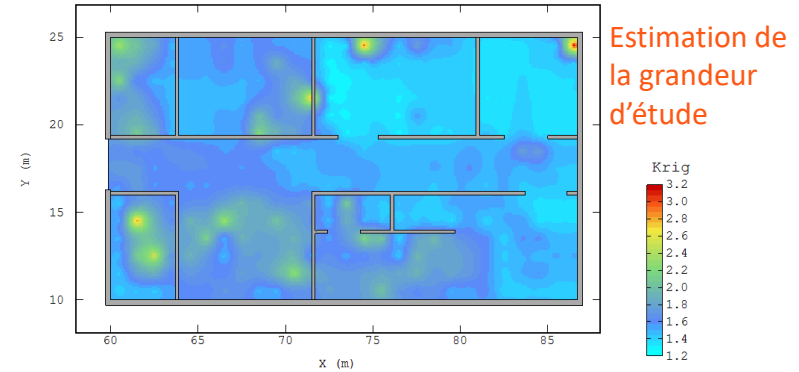
Exemple : pollutions radiologiques

[Desnoyers, 2010, CEA/ENSMP/Géovariances]



Estimations :

- ◆ Cartographie de la grandeur étudiée
- ◆ Estimation de la valeur la plus probable et de l'incertitude associée
- ◆ Probabilité de dépassement de seuil
- ◆ Analyse de risque globale : surface, volumes



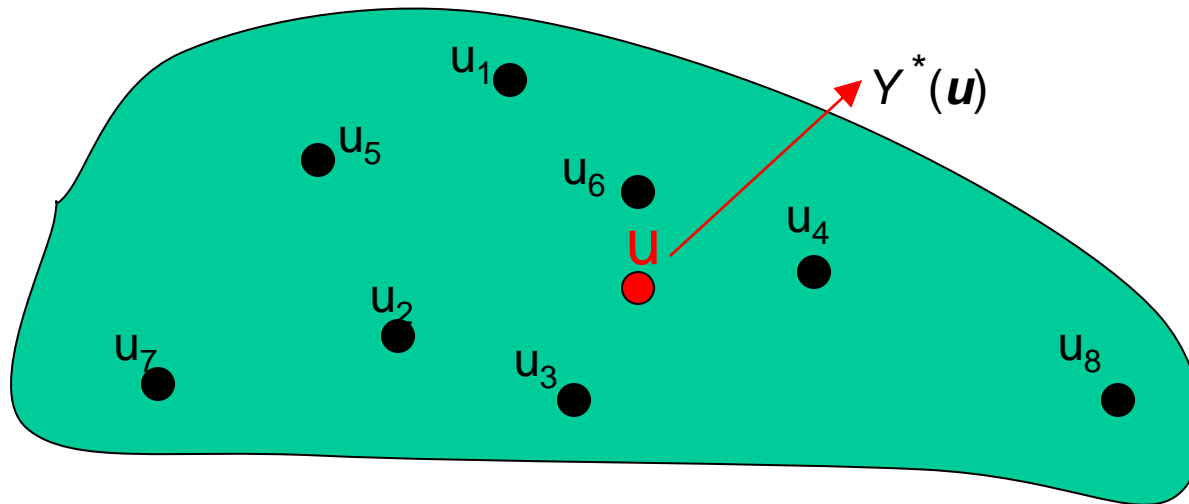
Probabilité de dépassement de seuil

Interpolation par krigage

Linear combination of N data:

$$Y^*(\mathbf{u}) = \sum_{i=1}^N \lambda_i Y(\mathbf{u}_i)$$

Kriging can take into account the data configuration, the distance between data and target, the spatial correlations and potential external information



Probabilistic framework

Estimation without bias: $E [Y^*(\mathbf{u}) - Y(\mathbf{u})] = 0$
the mean of the errors is zero

Estimation $Y^*(\mathbf{u})$ is optimal: $\text{Var} [Y^*(\mathbf{u}) - Y(\mathbf{u})]$ is minimal
the dispersion of the errors is reduced

VARIABLE RÉGIONALISÉE

Modèle constitutif de la géostatistique :

- Le phénomène étudié prend des valeurs dans un domaine \mathcal{D} spatial et/ou temporel :

fonction $y(x)$ avec x position dans \mathbb{R}^d ($d = 1, 2, 3$ ou 4)

En général, on connaît $y(x)$ en quelques points x_α

- La fonction $y(x)$ est une variable régionalisée
= réalisation d'une fonction aléatoire $Y(x)$ implantée sur \mathcal{D}

Les données sont considérées comme ayant été générées par un processus aléatoire :

$$Y(x_\alpha) = y(x_\alpha)$$

INFÉRENCE STATISTIQUE

Connaître les propriétés statistiques de Y à partir d'une réalisation unique $y(x_\alpha)$ est impossible dans le cas général

Chaque $Y(x_i)$ est une variable aléatoire qui a sa loi de proba \Rightarrow **impossible de définir la loi de Y**

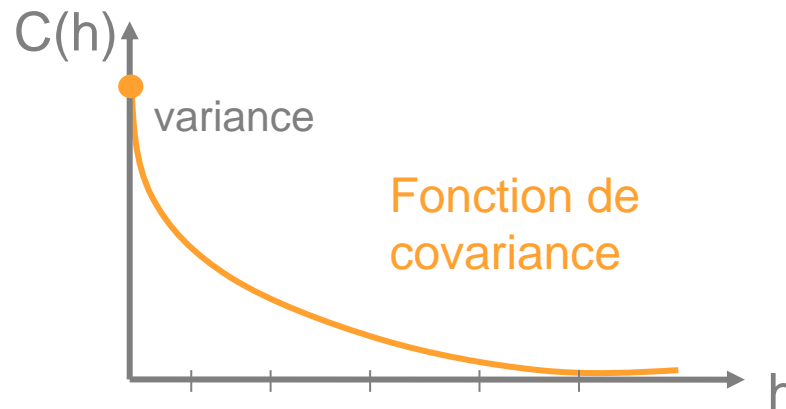
\Rightarrow **Hypothèse de stationnarité**

Z est stationnaire (loi invariante par translation) d'ordre deux :

Moyenne constante : $E[Y(x)] = m$

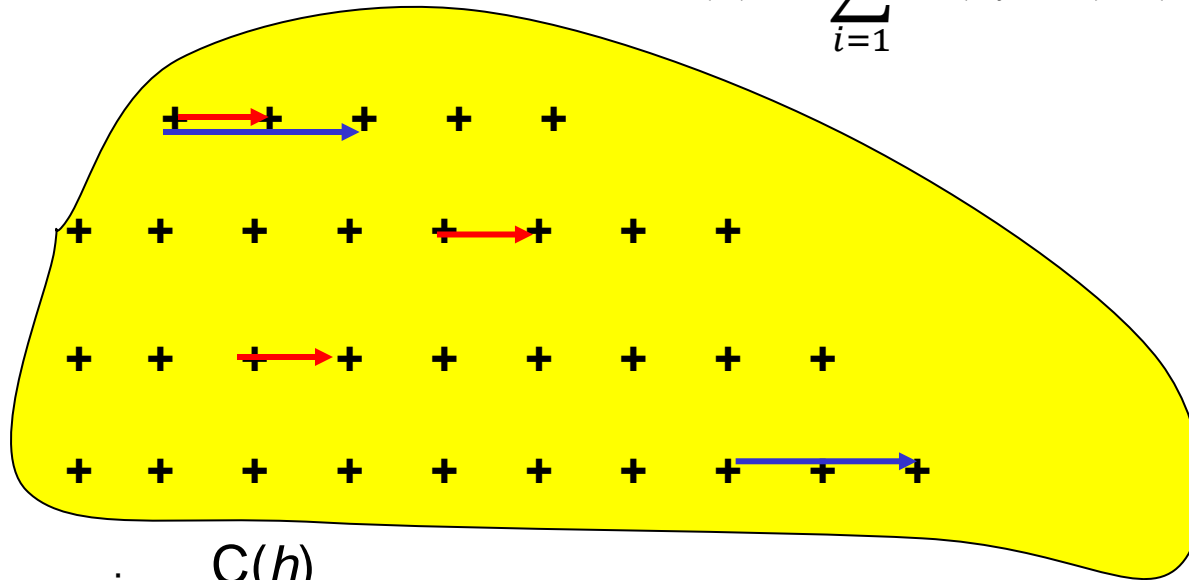
Variance constante : $\text{Var}[Y(x)] = \sigma^2$

Covariance ne dépend pas de x : $\text{Cov}[Y(x_i), Y(x_j)] = C(x_i - x_j)$



En pratique

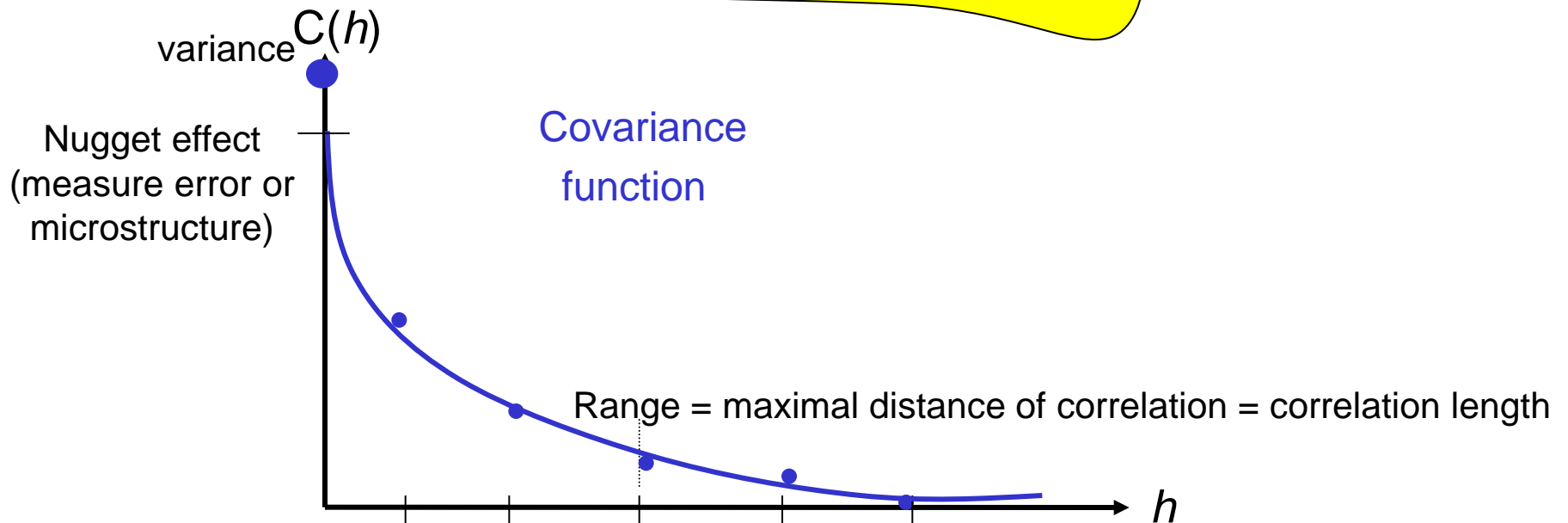
$$C(h) = \sum_{i=1}^{N(h)} [Y(x_i + h)Y(x_i)] - m^2$$



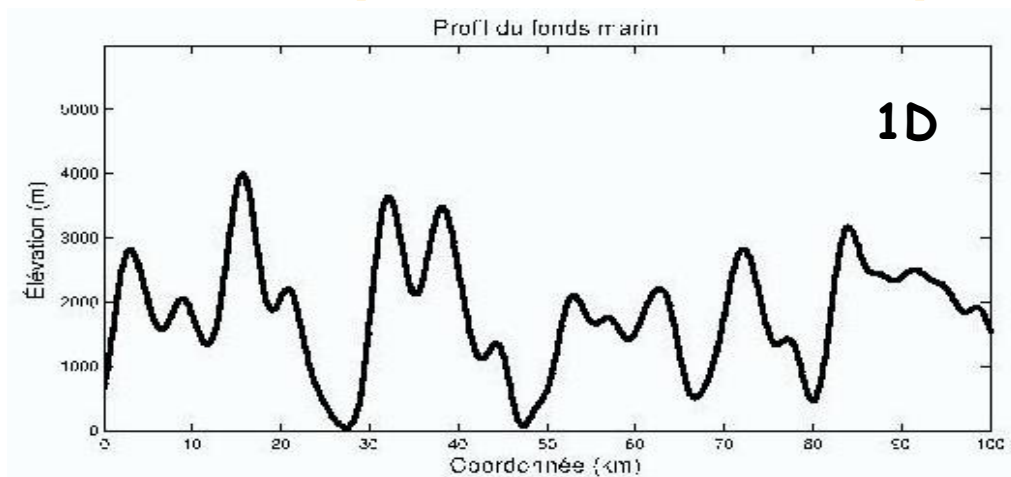
$$\text{Var}[Y(\mathbf{x})] = C(0)$$

$$Y(\mathbf{x}) \longrightarrow Y(\mathbf{x}+h)$$

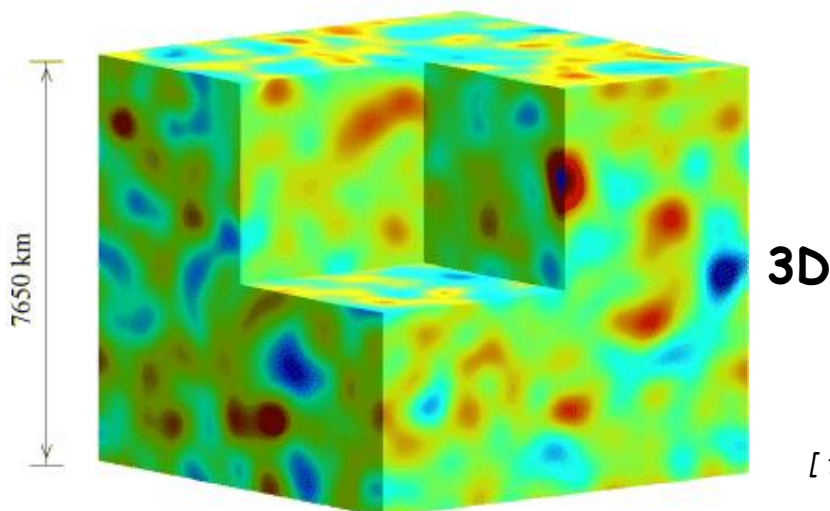
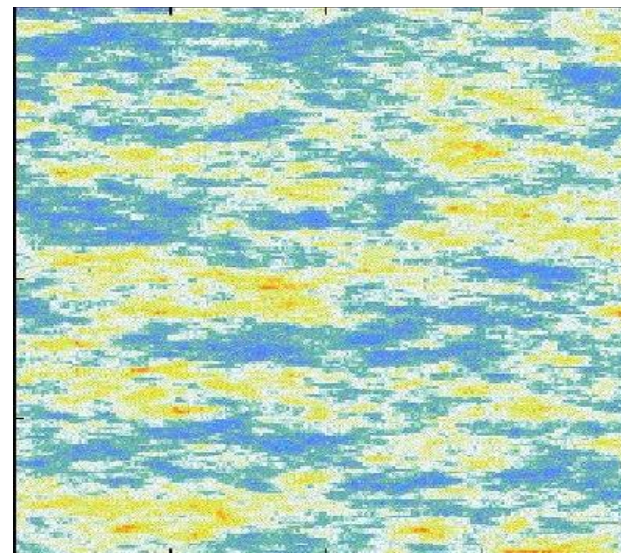
$$Y(\mathbf{x}) \longrightarrow Y(\mathbf{x}+2h)$$



Examples of stochastic processes (Gaussian)



[from: Marcotte]



[from: Baig, 2003]

Model of covariance/correlation functions

1D parametric models:

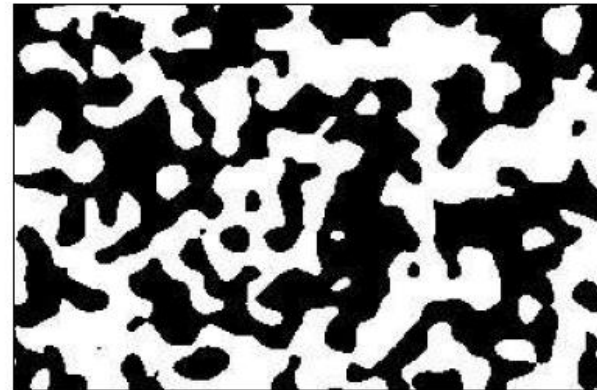
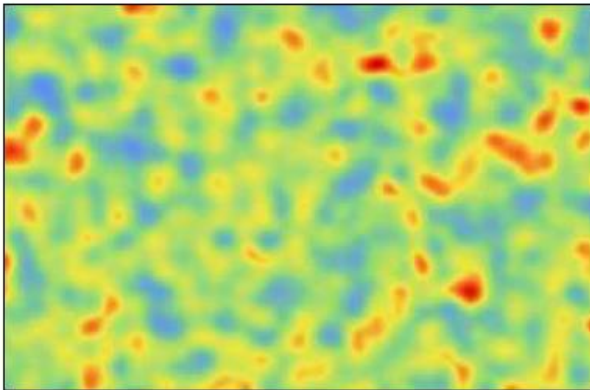
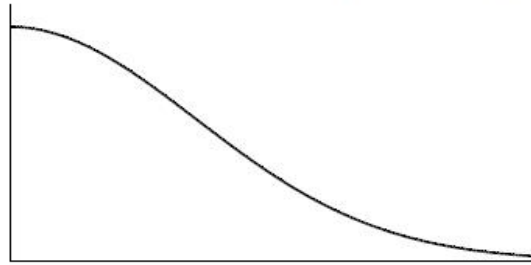
- Exponential: $c(h) = \sigma^2 \exp\left(-\frac{h}{\theta}\right)$
- Gaussian: $c(h) = \sigma^2 \exp\left(-\left(\frac{h}{\theta}\right)^2\right)$
- Generalized exponential: $c(h) = \sigma^2 \exp\left(-\left(\frac{h}{\theta}\right)^p\right)$ avec $p > 0$
- Matern':
$$c(h) = \sigma^2 \frac{1}{2^{\alpha-1} \Gamma(\alpha)} \left(\frac{h}{\theta}\right)^\alpha K_\alpha\left(\frac{h}{\theta}\right) \text{ avec } \alpha > 0$$

K_α is the Bessel function
- Particular cases:
 $p = 0 \Rightarrow$ Exponential
$$p = 1 \Rightarrow c(h) = \sigma^2 \left(1 + \frac{\sqrt{3}h}{\theta}\right) \exp\left(-\frac{\sqrt{3}h}{\theta}\right) \text{ Matern3/2}$$
$$p = 2 \Rightarrow c(h) = \sigma^2 \left(1 + \frac{\sqrt{5}h}{\theta} + \frac{5h^2}{3\theta^2}\right) \exp\left(-\frac{\sqrt{5}h}{\theta}\right) \text{ Matern5/2}$$

Remark: p is related to the regularity of the process (differentiability)

Examples with a Gaussian covariance

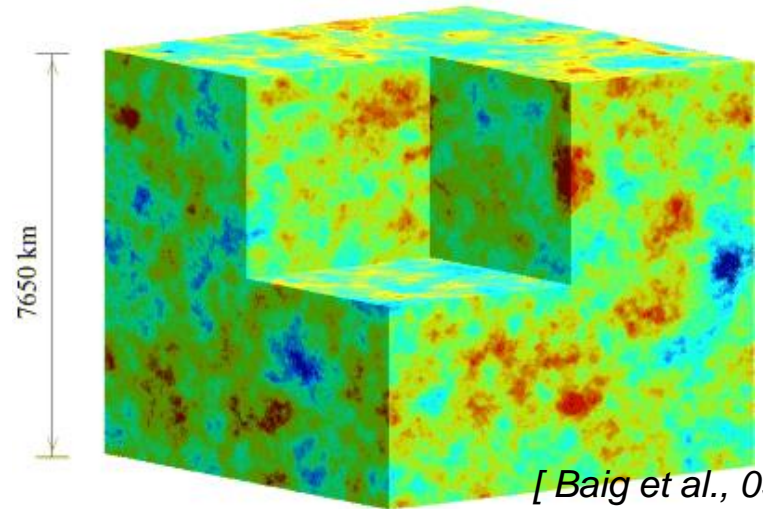
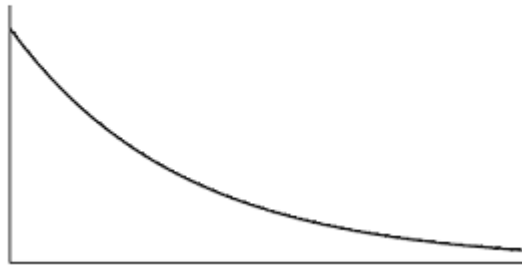
$$C(h) = \exp \left\{ -\frac{|h|^2}{a^2} \right\}$$



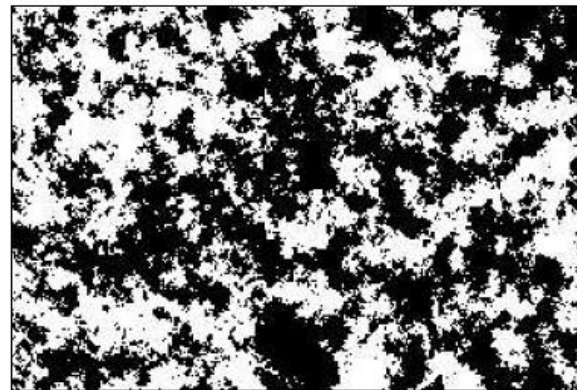
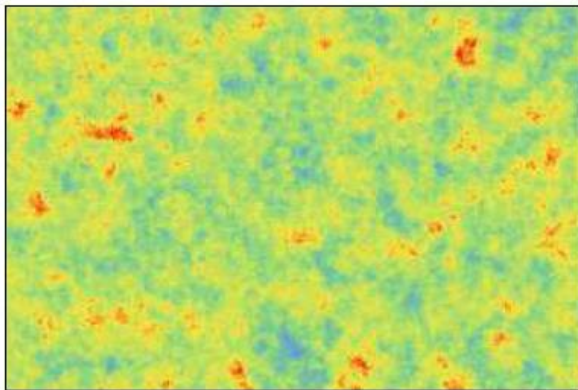
Simulation of 2D isotropic fields

Examples with an exponential covariance

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



Example of a sedimentary medium



Simulation of 2D isotropic fields

Isotropic and anisotropic covariances

- Isotropic covariance $c(\mathbf{h}) = c(\|\mathbf{h}\|)$
- Anisotropic covariance $c(\mathbf{h}) = \prod_{i=1}^d c_i(h_i)$

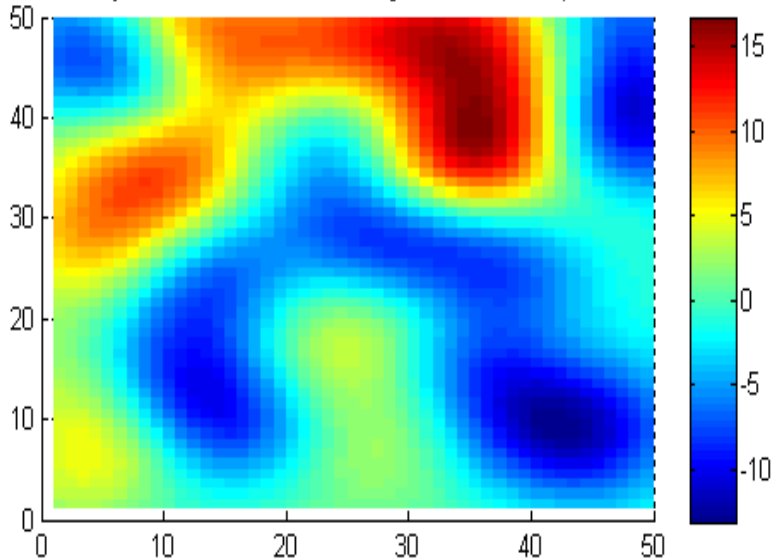
Exponential

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

Gaussian

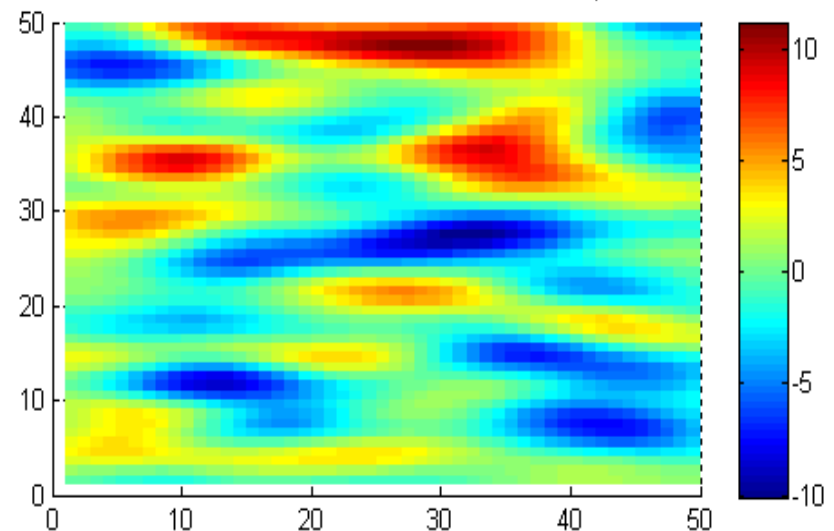
$$c(\mathbf{h}) = \sigma^2 \exp\left(-\sum_{i=1}^d \left(\frac{h_i}{\theta_i}\right)^2\right)$$

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



Simulation of a 2D isotropic field
(Gaussian covariance)

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

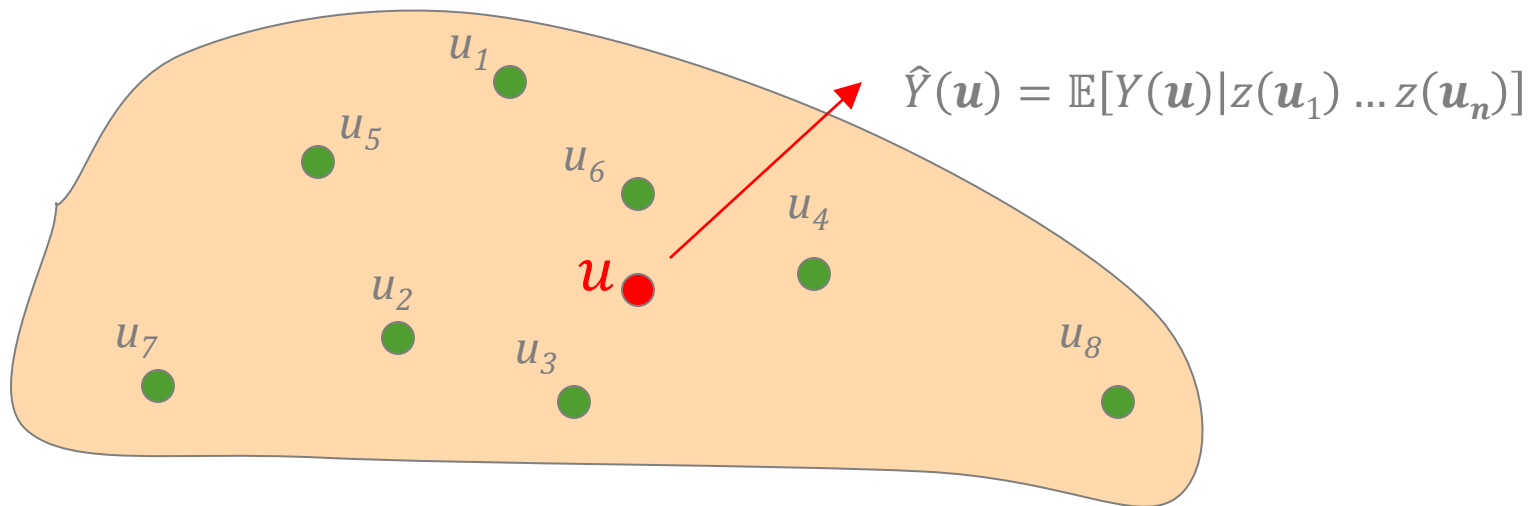


Simulation of a 2D anisotropic field
(Gaussian covariance)

LE KRIGEAGE

Combinaison linéaire des n données : $\hat{Y}(\mathbf{u}) = \sum_{i=1}^n \lambda_i Y(\mathbf{u}_i)$

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



Estimateur sans biais : $\mathbb{E}[\hat{Y}(\mathbf{u}) - Y(\mathbf{u})] = 0$

la moyenne des erreurs est nulle

$\hat{Y}(\mathbf{u})$ est optimal : $\text{Var}[\hat{Y}(\mathbf{u}) - Y(\mathbf{u})]$ est minimale

la dispersion des erreurs est réduite

KRIGEAGE : LES HYPOTHÈSES

- On cherche la valeur d'une fonction G (inconnue) en un point x_0
- On fait l'hypothèse que G est la réalisation d'un champ aléatoire Z dont on va estimer les paramètres
- On fixe trois contraintes :
 - Linéarité : on cherche un estimateur linéaire en fonction des données Z_i , c'est-à-dire de la forme $\hat{Z}(x_0) = \sum_i \lambda_i Z_i$
 - Non-biais : on cherche un estimateur non biaisé, c'est-à-dire tel que $E[\hat{Z} - Z] = 0$
 - Optimalité : on cherche un estimateur de variance minimale

$$\text{Var}[\hat{Z} - Z] = \min_{\lambda \in \mathbb{R}^n} \text{Var}[\sum_i \lambda_i Z_i - Z]$$

KRIGEAGE : LES HYPOTHÈSES

- On cherche la valeur de la fonction G en un point x_0
- On fait l'hypothèse que G est la réalisation d'un champ aléatoire Z dont on va estimer les paramètres
- On fixe trois contraintes :
 - Linéarité : on cherche un estimateur linéaire en fonction des données Z_i , c'est-à-dire de la forme $\hat{Z}(x_0) = \sum_i \lambda_i Z_i$
 - Non-biais : on cherche un estimateur non biaisé, c'est-à-dire tel que $E[\hat{Z} - Z] = 0$
 - Optimalité : on cherche un estimateur de variance minimale

$$\text{Var}[\hat{Z} - Z] = \min_{\lambda \in \mathbb{R}^n} \text{Var}[\sum_i \lambda_i Z_i - Z]$$

Krigeage = BLUP

Best Linear Unbiased Predictor

KRIGEAGE SIMPLE

- On suppose donc que G est la réalisation d'un processus Z :
 - Stationnaire du second ordre

KRIGEAGE SIMPLE

- On suppose donc que f est la réalisation d'un processus Z :
 - Stationnaire du second ordre

Un processus $Z(x)$ est dit stationnaire a l'ordre 1 si son espérance est la même en tout point de l'espace, soit formellement :

$$\exists m \in \mathbb{R}, \forall x \in \mathcal{D}_Z, \mathbb{E}[Z(x)] = m$$

Un processus $Z(x)$ dit stationnaire d'ordre 2 lorsque :

$$\exists m \in \mathbb{R}, \forall x_1, x_2 \in \mathcal{D}_Z, \mathbb{E}[Z(x_1)] = \mathbb{E}[Z(x_2)] = m$$

$$\exists \sigma \in \mathbb{R}, \forall x_1, x_2 \in \mathcal{D}_Z, \text{Var}[Z(x_1)] = \text{Var}[Z(x_2)] = \sigma^2$$

$$\forall x_1, x_2 \in \mathcal{D}_Z, \text{Cov}[Z(x_1), Z(x_2)] = C(x_2 - x_1) \text{ pour une certaine fonction } C$$

KRIGEAGE SIMPLE

- On suppose donc que f est la réalisation d'un processus Z :
 - Stationnaire du second ordre

Un processus $Z(x)$ est dit stationnaire a l'ordre 1 si son espérance est la même en tout point de l'espace, soit formellement :

$$\exists m \in \mathbb{R}, \forall x \in \mathcal{D}_Z, \mathbb{E}[Z(x)] = m$$

Un processus $Z(x)$ dit stationnaire d'ordre 2 lorsque :

$$\exists m \in \mathbb{R}, \forall x_1, x_2 \in \mathcal{D}_Z, \mathbb{E}[Z(x_1)] = \mathbb{E}[Z(x_2)] = m$$

$$\exists \sigma \in \mathbb{R}, \forall x_1, x_2 \in \mathcal{D}_Z, \text{Var}[Z(x_1)] = \text{Var}[Z(x_2)] = \sigma^2$$

$$\forall x_1, x_2 \in \mathcal{D}_Z, \text{Cov}[Z(x_1), Z(x_2)] = C(x_2 - x_1) \text{ pour une certaine fonction } C$$

On peut considérer que la stationnarité signifie que
la loi du processus est « invariante par translation »

KRIGEAGE SIMPLE

- On suppose donc que G est la réalisation d'un processus Z :
 - Stationnaire du second ordre
 - De moyenne nulle $E[Z] = 0$

KRIGEAGE SIMPLE

- On suppose donc que G est la réalisation d'un processus Z :
 - Stationnaire du second ordre
 - De moyenne nulle $E[Z] = 0$
 - De fonction de covariance connue
- La variance de Z est notée σ^2 et la fonction de corrélation est notée k , on a donc :
$$\text{Cov}(Z(x_1), Z(x_2)) = \sigma^2 k(x_1 - x_2)$$

KRIGEAGE SIMPLE

- On suppose donc que G est la réalisation d'un processus Z :
 - Stationnaire du second ordre
 - De moyenne nulle $E[Z] = 0$
 - De fonction de covariance connue
- La variance de Z est notée σ^2 et la fonction de corrélation est notée k , on a donc :
$$\text{Cov}(Z(x_1), Z(x_2)) = \sigma^2 k(x_1 - x_2)$$
- Contrainte de non-biais :

$$\begin{aligned} E[\hat{Z} - Z] &= E\left[\sum_i \lambda_i Z_i - Z\right] \\ &= \sum_i \lambda_i E[Z_i] - E[Z] = 0 \end{aligned}$$

Dans le cas du krigeage simple, cette contrainte est toujours vérifiée

KRIGEAGE SIMPLE

- On suppose donc que G est la réalisation d'un processus Z :
 - Stationnaire du second ordre
 - De moyenne nulle $E[Z] = 0$
 - De fonction de covariance connue
- La variance de Z est notée σ^2 et la fonction de corrélation est notée k , on a donc :
$$\text{Cov}(Z(x_1), Z(x_2)) = \sigma^2 k(x_1 - x_2)$$
- Contrainte de minimisation de la variance d'estimation

La variance d'estimation s'écrit :

$$\begin{aligned}\text{Var}[\hat{Z} - Z] &= \text{Var}\left[\sum_i \lambda_i Z_i - Z\right] \\ &= \text{Var}[Z] + \sum_i \sum_j \lambda_i \lambda_j \text{Cov}(Z_i, Z_j) - 2 \sum_i \lambda_i \text{Cov}(Z_i, Z)\end{aligned}$$

KRIGEAGE SIMPLE

- Contrainte de minimisation de la variance d'estimation

La variance d'estimation s'écrit :

$$\begin{aligned}\text{Var}[\hat{Z} - Z] &= \text{Var}\left[\sum_i \lambda_i Z_i - Z\right] \\ &= \text{Var}[Z] + \sum_i \sum_j \lambda_i \lambda_j \text{Cov}(Z_i, Z_j) - 2 \sum_i \lambda_i \text{Cov}(Z_i, Z)\end{aligned}$$

Qui se minimise en dérivant par rapport à chaque λ_i :

$$\begin{aligned}\frac{d \text{Var}[\hat{Z} - Z]}{d\lambda_i} &= \frac{d \text{Var}[Z]}{d\lambda_i} + \frac{d \sum_i \sum_j \lambda_i \lambda_j \text{Cov}(Z_i, Z_j)}{d\lambda_i} - 2 \frac{d \sum_i \lambda_i \text{Cov}(Z_i, Z)}{d\lambda_i} \\ 0 &= 0 + 2 \sum_j \lambda_j \text{Cov}(Z_i, Z_j) - 2 \text{Cov}(Z_i, Z)\end{aligned}$$

KRIGEAGE SIMPLE

- Contrainte de minimisation de la variance d'estimation

La variance d'estimation s'écrit :

$$\begin{aligned}\text{Var}[\hat{Z} - Z] &= \text{Var}\left[\sum_i \lambda_i Z_i - Z\right] \\ &= \text{Var}[Z] + \sum_i \sum_j \lambda_i \lambda_j \text{Cov}(Z_i, Z_j) - 2 \sum_i \lambda_i \text{Cov}(Z_i, Z)\end{aligned}$$

Qui se minimise en dérivant par rapport à chaque λ_i :

$$\begin{aligned}\frac{d \text{Var}[\hat{Z} - Z]}{d\lambda_i} &= \frac{d \text{Var}[Z]}{d\lambda_i} + \frac{d \sum_i \sum_j \lambda_i \lambda_j \text{Cov}(Z_i, Z_j)}{d\lambda_i} - 2 \frac{d \sum_i \lambda_i \text{Cov}(Z_i, Z)}{d\lambda_i} \\ 0 &= 0 + 2 \sum_j \lambda_j \text{Cov}(Z_i, Z_j) - 2 \text{Cov}(Z_i, Z)\end{aligned}$$

La contrainte $\frac{d \text{Var}[\hat{Z} - Z]}{d\lambda_i} = 0$ **nous donne donc** $\sum_j \lambda_j \text{Cov}(Z_i, Z_j) = \text{Cov}(Z_i, Z)$

KRIGEAGE SIMPLE

- Contrainte de minimisation de la variance d'estimation

La contrainte $\frac{d \text{Var}[\hat{Z}-Z]}{d\lambda_i} = 0$ nous donne donc $\sum_j \lambda_j \text{Cov}(Z_i, Z_j) = \text{Cov}(Z_i, Z)$

KRIGEAGE SIMPLE

- Contrainte de minimisation de la variance d'estimation

La contrainte $\frac{d \text{Var}[\hat{Z}-Z]}{d\lambda_i} = 0$ nous donne donc $\sum_j \lambda_j \text{Cov}(Z_i, Z_j) = \text{Cov}(Z_i, Z)$

... ce qui est vrai pour tout $i \in \{1, \dots, n\}$:

$$\sum_{j=1}^n \lambda_j \text{Cov}(Z_1, Z_j) = \text{Cov}(Z_1, Z)$$

\vdots

$$\sum_{j=1}^n \lambda_j \text{Cov}(Z_i, Z_j) = \text{Cov}(Z_i, Z)$$

\vdots

$$\sum_{j=1}^n \lambda_j \text{Cov}(Z_n, Z_j) = \text{Cov}(Z_n, Z)$$

KRIGEAGE SIMPLE

- Contrainte de minimisation de la variance d'estimation

La contrainte $\frac{d \text{Var}[\hat{Z}-Z]}{d\lambda_i} = 0$ nous donne donc $\sum_j \lambda_j \text{Cov}(Z_i, Z_j) = \text{Cov}(Z_i, Z)$

... ce qui est vrai pour tout $i \in \{1, \dots, n\}$:

$$\sum_{j=1}^n \lambda_j \text{Cov}(Z_1, Z_j) = \text{Cov}(Z_1, Z)$$

\vdots

$$\sum_{j=1}^n \lambda_j \text{Cov}(Z_i, Z_j) = \text{Cov}(Z_i, Z)$$

\vdots

$$\sum_{j=1}^n \lambda_j \text{Cov}(Z_n, Z_j) = \text{Cov}(Z_n, Z)$$

► et suggère une écriture matricielle

KRIGEAGE SIMPLE

- Formulation matricielle

- On note :

$$K = \begin{pmatrix} \text{Cov}(Z_1, Z_1) & \cdots & \text{Cov}(Z_1, Z_j) & \cdots & \text{Cov}(Z_1, Z_n) \\ \vdots & & \vdots & & \vdots \\ \text{Cov}(Z_i, Z_1) & \cdots & \text{Cov}(Z_i, Z_j) & \cdots & \text{Cov}(Z_i, Z_n) \\ \vdots & & \vdots & & \vdots \\ \text{Cov}(Z_n, Z_1) & \cdots & \text{Cov}(Z_n, Z_j) & \cdots & \text{Cov}(Z_n, Z_n) \end{pmatrix}$$

et

$$k_0 = \begin{pmatrix} \text{Cov}(Z_1, Z(x_0)) \\ \vdots \\ \text{Cov}(Z_i, Z(x_0)) \\ \vdots \\ \text{Cov}(Z_n, Z(x_0)) \end{pmatrix}$$

et

$$\lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_i \\ \vdots \\ \lambda_n \end{pmatrix}$$

KRIGEAGE SIMPLE

- Formulation matricielle
- Les n contraintes d'optimalité se réécrivent donc comme une unique équation matricielle :

$$K\lambda = k_0$$

La matrice K étant définie-positive, on peut écrire la solution de ce système de manière très simple :

$$\lambda^* = K^{-1}k_0$$

...écrire la solution de l'estimation de Z en x_0 :

$$\begin{aligned}\hat{Z}(x_0) &= \sum_{i=1}^n \lambda_i^* Z_i \\ &= Z^T \lambda^*\end{aligned}$$

$$\text{où } Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_i \\ \vdots \\ Z_n \\ 0 \end{pmatrix} = \begin{pmatrix} Z(x_{(1)}) \\ \vdots \\ Z(x_{(i)}) \\ \vdots \\ Z(x_{(n)}) \\ 0 \end{pmatrix}$$

KRIGEAGE SIMPLE

- Formulation matricielle
- Les n contraintes d'optimalité se réécrivent donc comme une unique équation matricielle :

$$K\lambda = k_0$$

La matrice K étant définie-positive, on peut écrire la solution de ce système de manière très simple :

$$\lambda^* = K^{-1}k_0$$

...et enfin réécrire la variance d'estimation avec ces notations :

$$\begin{aligned}\text{Var}[\hat{Z}(x_0) - Z(x_0)] &= \text{Var}[Z(x_0)] + \sum_i \sum_j \lambda_i^* \lambda_j^* \text{Cov}(Z_i, Z_j) - 2 \sum_i \lambda_i^* \text{Cov}(Z_i, Z(x_0)) \\ &= k_{0,0} + \lambda^{*T} K \lambda^* - 2k_0^T \lambda^* \\ &= k_{0,0} + (K^{-1}k_0)^T K \lambda^* - 2k_0^T \lambda^* \\ &= k_{0,0} + k_0^T (K^{-1})^T K \lambda^* - 2k_0^T \lambda^* \\ &= k_{0,0} - k_0^T \lambda^* \\ &= k_{0,0} - k_0^T K^{-1} k_0\end{aligned}$$

KRIGEAGE SIMPLE

...et enfin réécrire la variance d'estimation avec ces notations :

$$\begin{aligned}\text{Var}[\hat{Z}(x_0) - Z(x_0)] &= \text{Var}[Z(x_0)] + \sum_i \sum_j \lambda_i^* \lambda_j^* \text{Cov}(Z_i, Z_j) - 2 \sum_i \lambda_i^* \text{Cov}(Z_i, Z(x_0)) \\ &= k_{0,0} - k_0^T K^{-1} k_0\end{aligned}$$

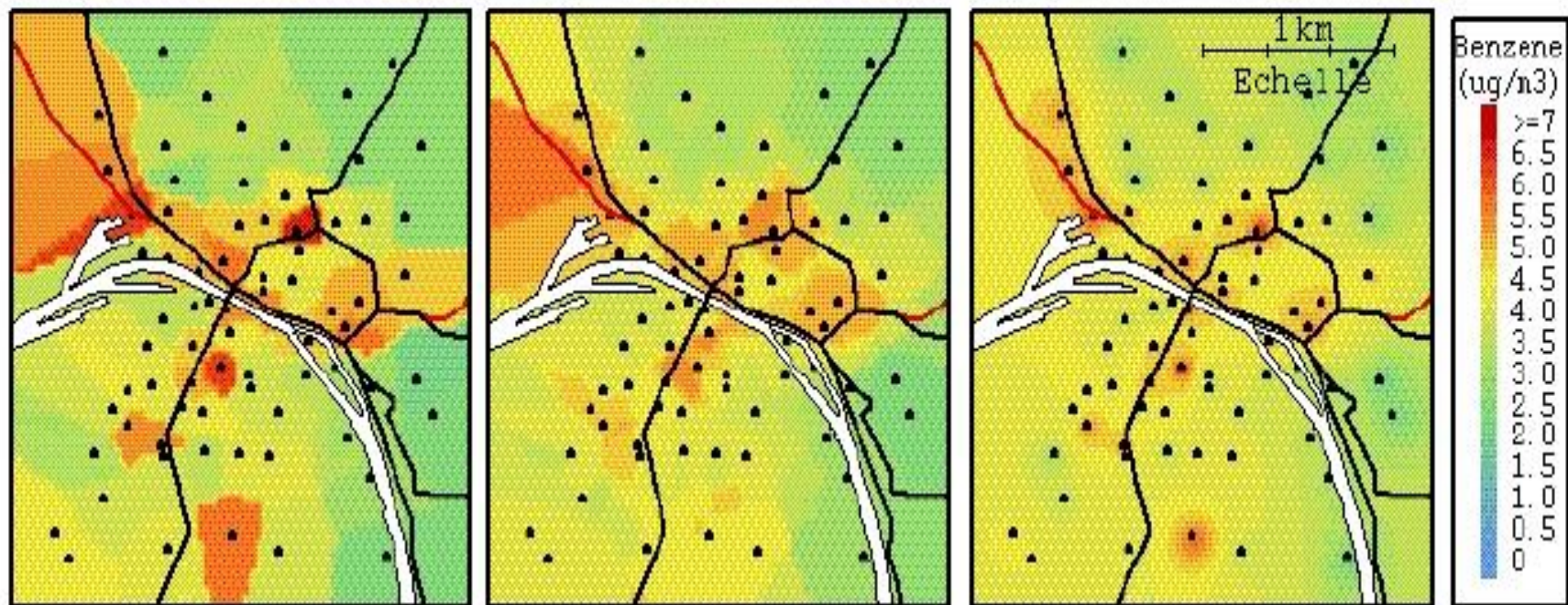
Remarque importante : cette variance au point x_0 où Z n'est pas connue **ne dépend pas de valeurs de la variable Z** mais seulement des localisations de ces valeurs $(x_{(1)}, \dots, x_{(n)}, x_0)$

EXEMPLE : CARTOGRAPHIE DE POLLUANTS ATMOSPHERIQUES

73 données de concentration en benzène sur l'agglomération de Rouen

Méthodes classiques : moyenne pondérée des valeurs aux stations voisines

L'estimation dépend de la localisation des stations / nœud où on estime



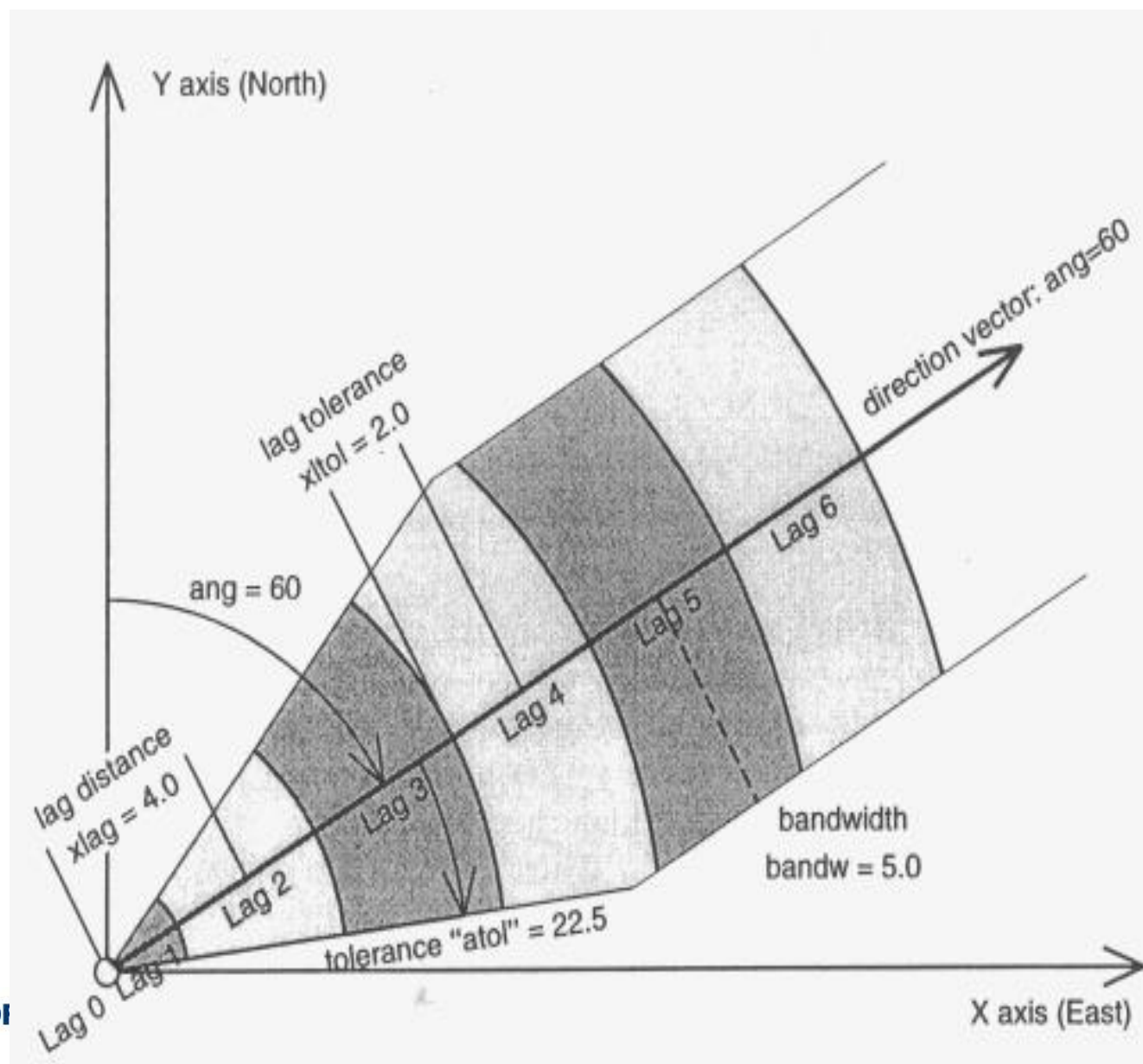
(a) Méthode polygonale

(b) Moyenne mobile (3 pts)

(c) Inverse des (distances)**2



L'ANALYSE VARIOGRAPHIQUE EN PRATIQUE



AJUSTEMENT DU VARIOGRAMME

Krigeage

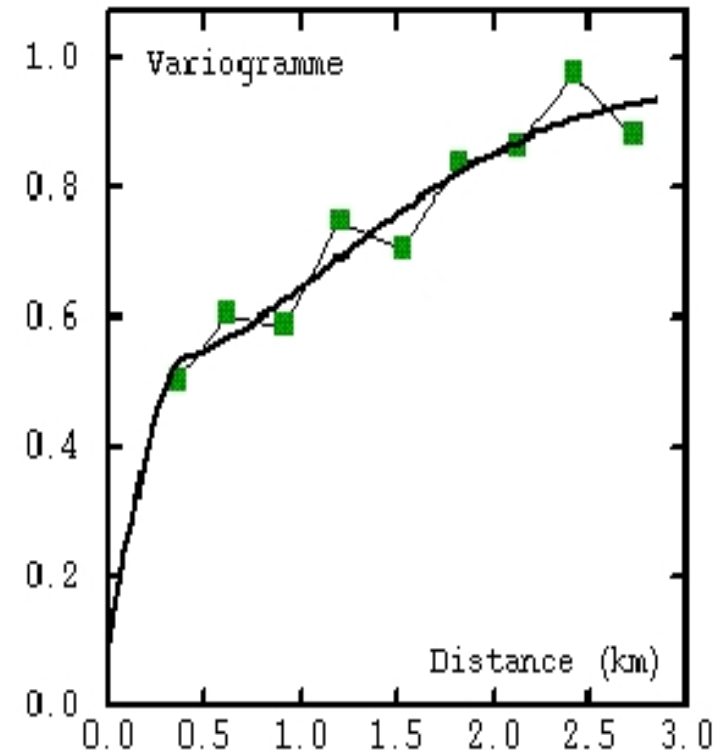
Prise en compte des caractéristiques spatiales du polluant (par l'intermédiaire du variogramme ou de la covariance)

Influence du modèle de variogramme

- Portée
- Présence d'un effet de pépité
(*bruit, microstructure < taille support, microstructure < distance minimale*)

⇒ le krigeage n'est plus interpolant

⇒ carte plus lisse



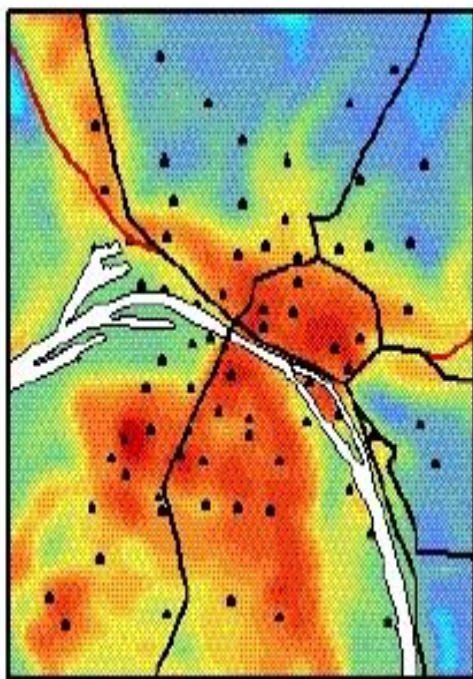
[Bobbia et al., 2000]

EXEMPLES DE KRIGEAGE

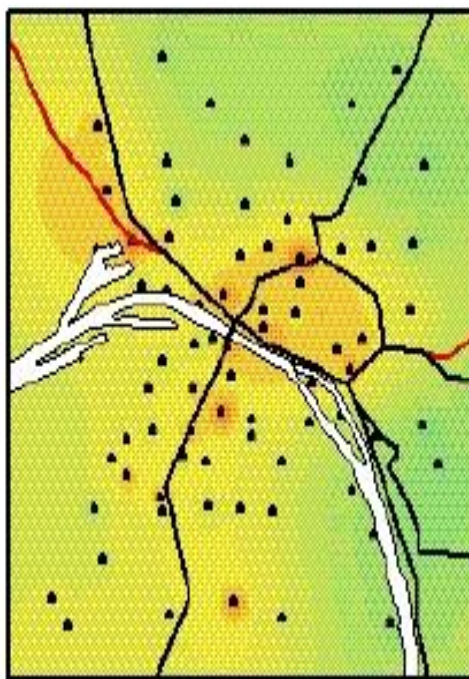
Le krigeage ordinaire se fait à moyenne constante mais inconnue

La dérive externe est une fonction de la densité de population et du relief (liés à la pollution automobile)

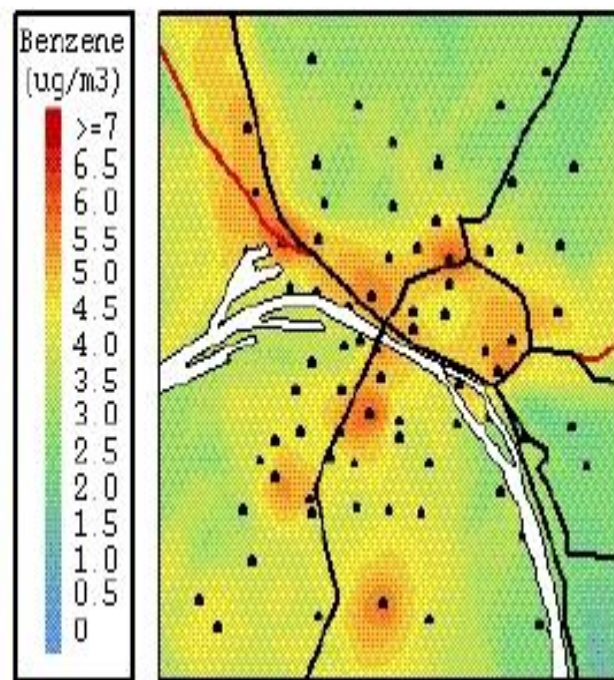
La pollution est supposée en dépendre linéairement



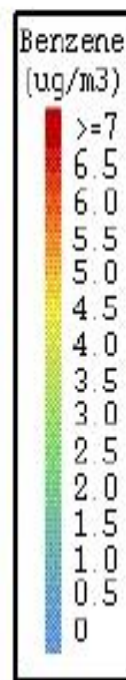
(a) Dérive externe



(b) Krigeage ordinaire



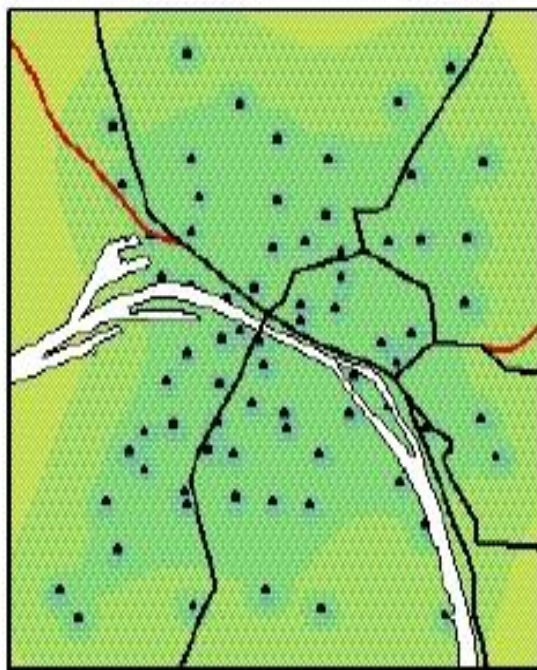
(c) Krigeage avec dérive externe



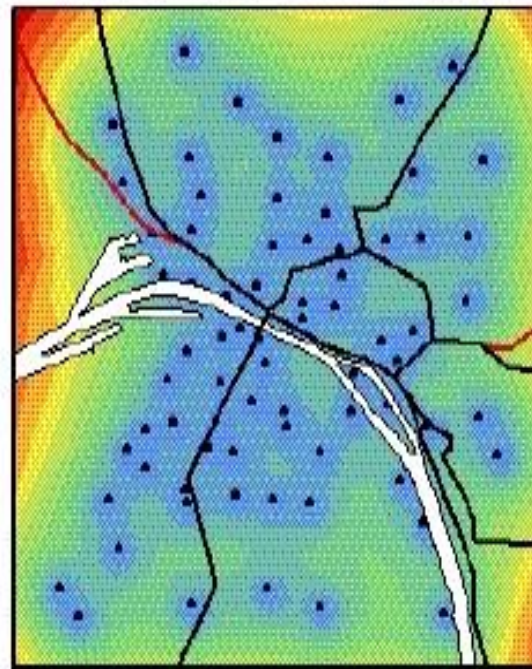
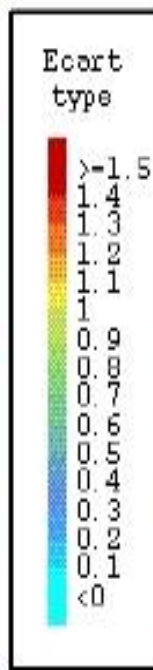
ERREURS DE KRIGEAGE

Deux méthodes quantitatives pour juger la qualité de l'estimation :

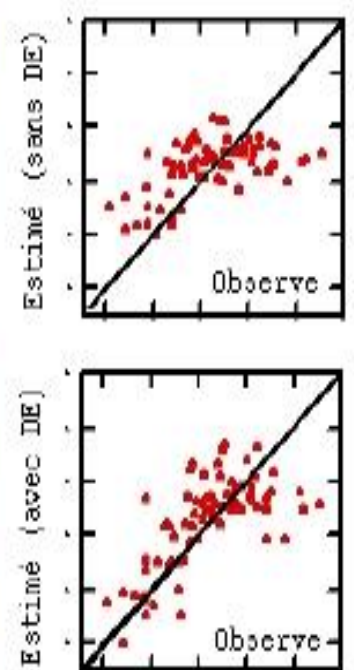
- Écart type du krigeage (→ fortes incertitudes en extrapolation)
- **Validation croisée** (→ la dérive externe améliore les résultats)



(a) Ecart-type du krigeage



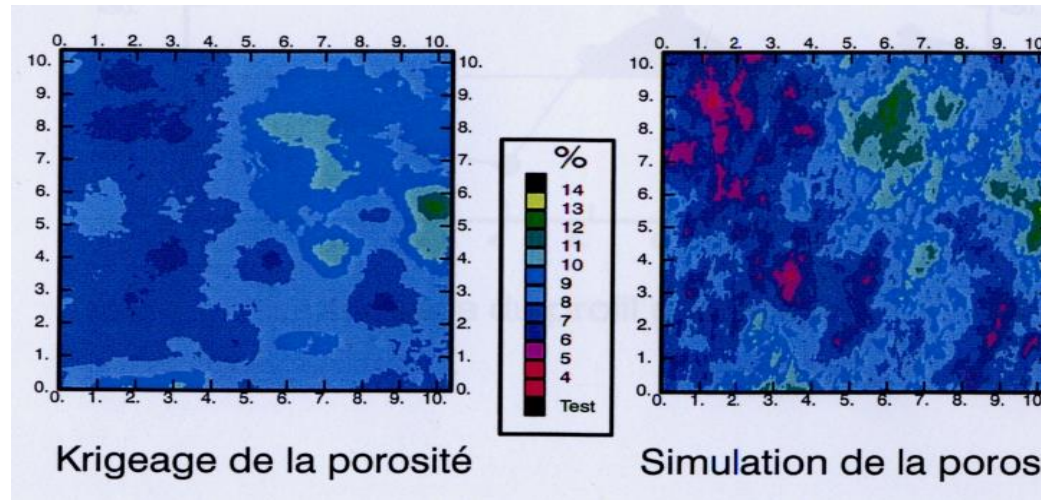
(b) Ecart-type avec dérive externe



(c) Validation croisée

Geostatistical simulations

- **Kriging** give the **optimal estimation** (unbiased, minimal error variance) of the variable at any point, from experimental data
- **A simulation represents a possible realization** of the real phenomenon
It reproduces its true **variability** (distribution, variogram), with respecty to experimental data (**conditional simulation**)



Main goal of simulation : **quantify the uncertainty via sampling** (as Monte Carlo)

Numerous methods of random fields simulation (LU decomposition, turning bands method, spectral method, Karhunen-Loève, etc.)

Gaussian process conditional simulations

- **Conditional law of $Z(x^*) | Z(X_S)$**

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

with
$$\begin{cases} \mu(x^*) = E[Z(x^*) | Z(X_S) = Z_S] = f_0(x^*) + k(x^*)^T K_S^{-1} Z_S \\ \tilde{\sigma}^2(x^*) = \text{Var}[Z(x^*) | Z(X_S) = Z_S] = \sigma^2 - k(x^*)^T K_S^{-1} k(x^*) \end{cases}$$

- **Use of this conditional law:**

- **Build confidence intervals**
- **Realization of conditional simulations**

Non conditional simulation of a Gaussian vector via covariance decomposition

Problem:

Given the $N \times N$ covariance matrix C , simulate multivariate normal (mean 0) random vector X with covariance $E[XX'] = C$

Solution:

- 1) Decompose $C = SS'$ where S is a triangular matrix
- 2) Generate a standard normal vector Z with covariance equal to unit matrix I
- 3) Let $X = SZ$

Proof:

Note that $E[XX'] = E[SZZ'S'] = S E[ZZ'] S' = SS' = C$

The actual method for finding such matrix S is not unique:

LU-decomposition, Cholesky or matrix square root

Simulations conditionnelles

Krigeage obs.

+

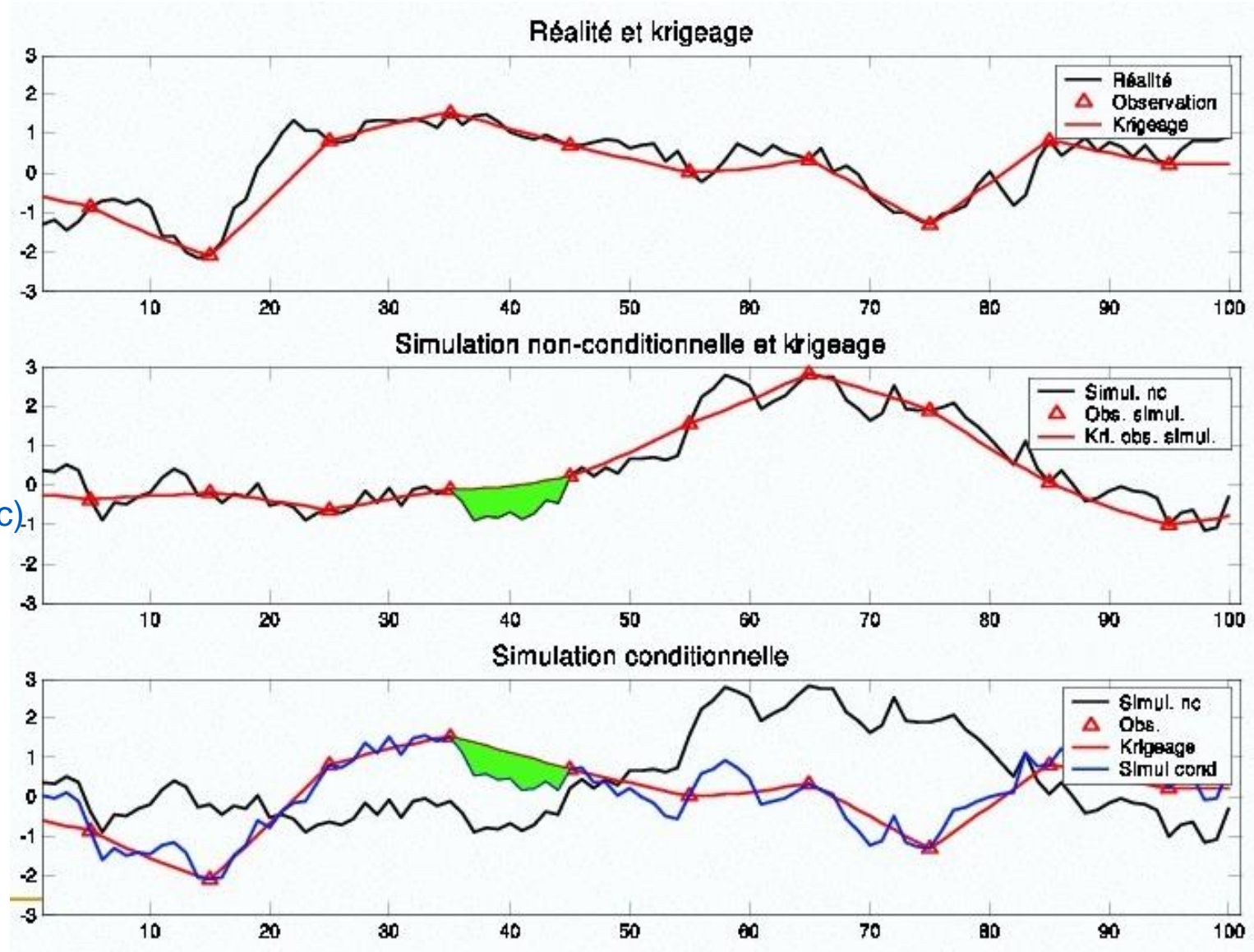
Simulation nc

-

Krigeage(simul. nc)

=

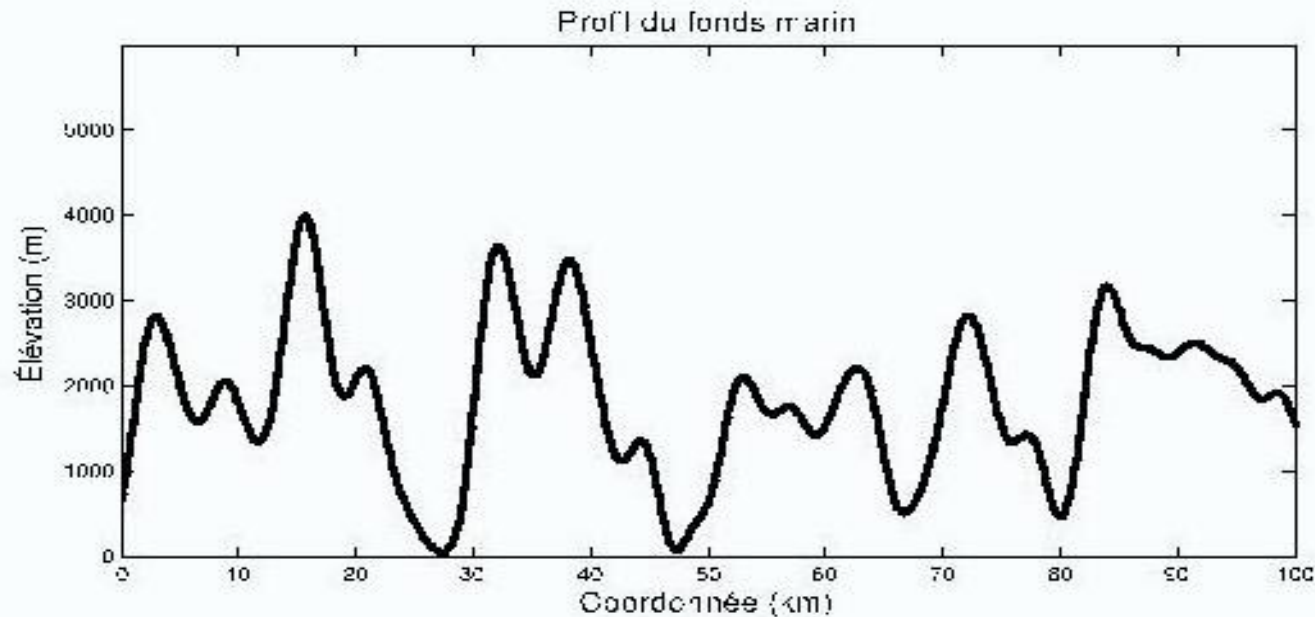
Simulation
conditionnelle



[Marcotte, Cours EPM]

Simulations conditionnelles

➤ Exemple de simulations : profil du fond marin



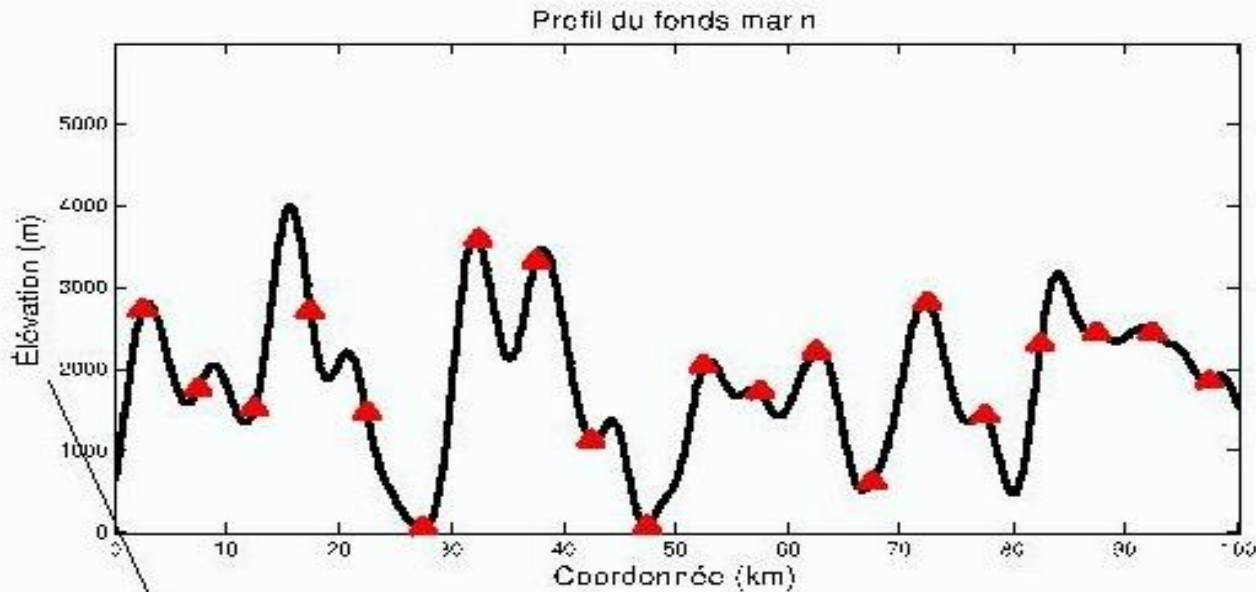
**Vous déposez un câble de fibre optique sur le fond marin.
Quelle longueur devrait avoir ce câble ?**

[Marcotte, Cours EPM]

Simulations conditionnelles

➤ Exemple de simulations : profil du fond marin

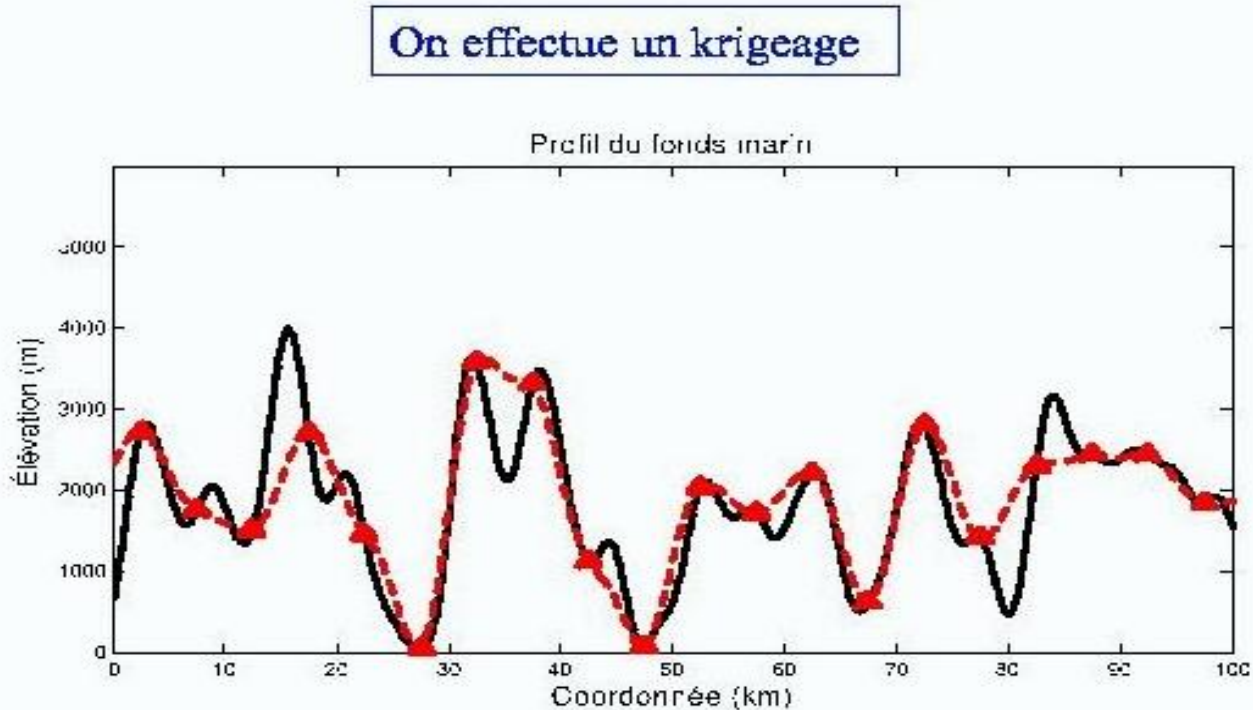
La profondeur exacte est connue uniquement aux points observations



Exagération verticale environ 10

Simulations conditionnelles

➤ Exemple de simulations : profil du fond marin



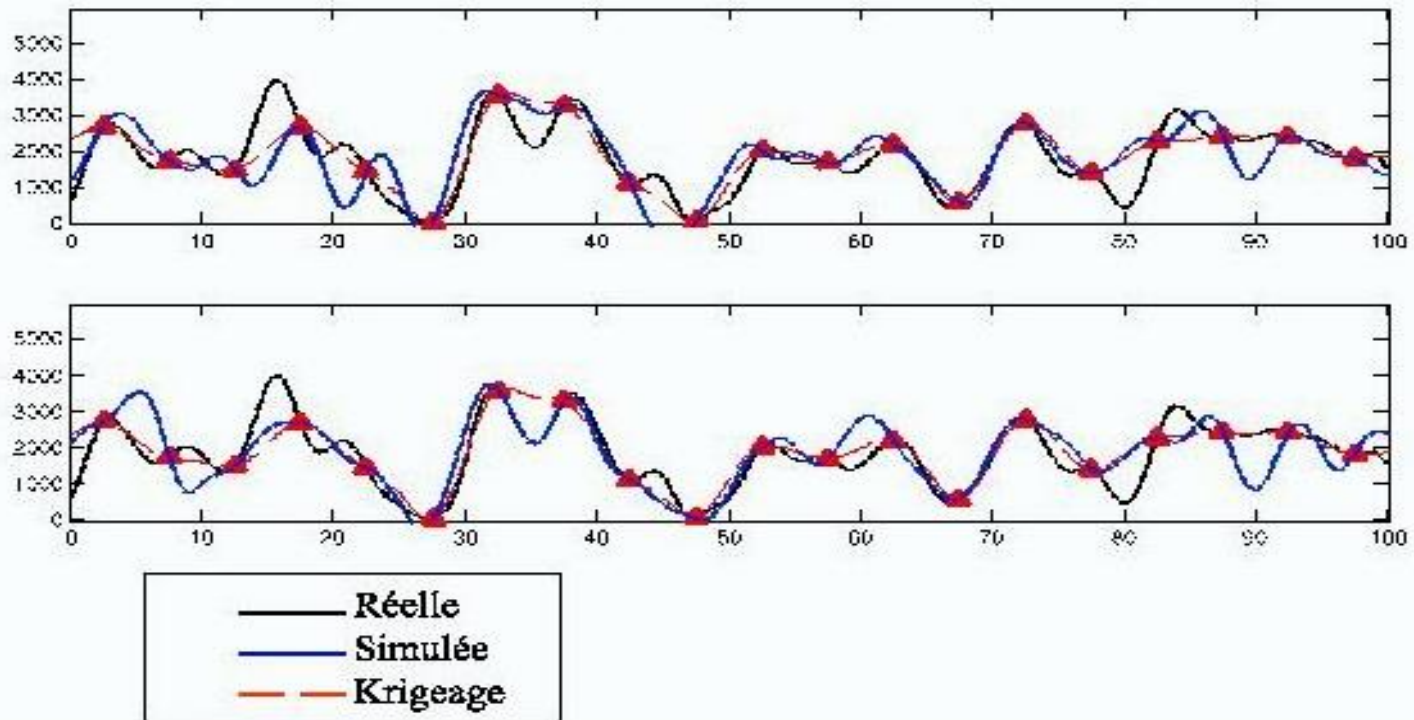
La longueur réelle est 110 km, le krigeage donne 104.6 km

On va manquer de câble !

Simulations conditionnelles

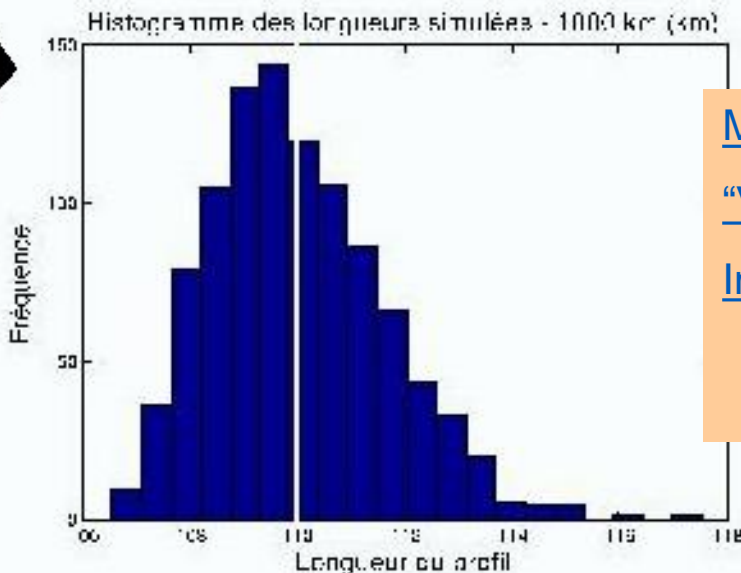
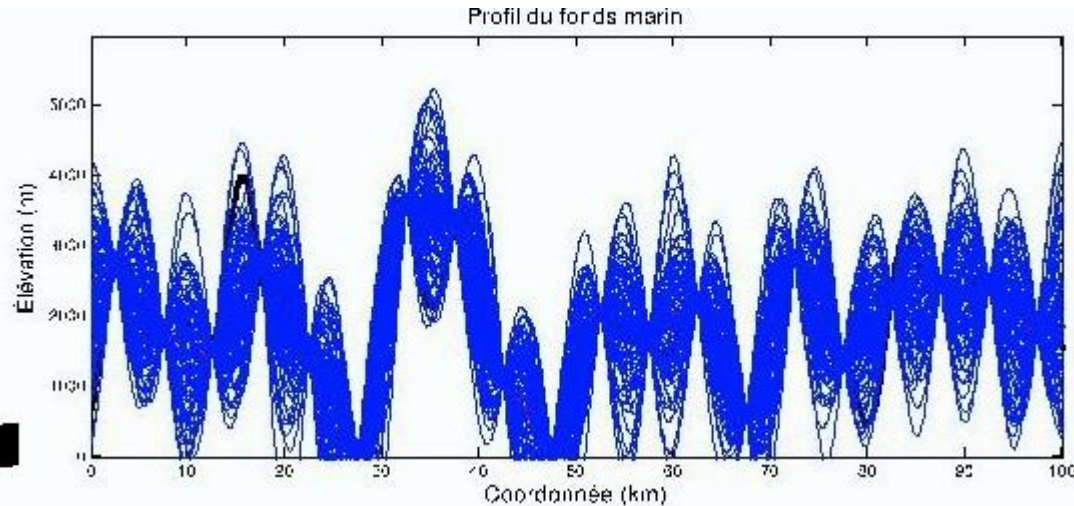
➤ Exemple de simulations : profil du fond marin

Une autre approche : des simulations conditionnelles



Simulations conditionnelles

➤ Exemple de simulations : profil du fond marin



Moyenne des 1000 réalisations : 110

“Vraie” valeur : 110

Intervalle de confiance à 95 % : [108.8 ; 113.5]

=> prédicteur de krigeage à 104.2, non
compris dans l'IC 95 % !!



Métamodèle processus gaussiens (krigeage)

Gaussian process metamodel (1/2)

- Idea: Computer code results are interpolated with the kriging technique
- Necessary hypothesis: Gaussian process

- Definition:

$$Y(x) = bF(x) + Z(x)$$

Regression

stochastic part

Stochastic process Z with :

$$\left\{ \begin{array}{l} E[Z(x)] = 0 \\ \text{Cov}(Z(x), Z(u)) = \sigma^2 R(x, u) \\ \text{where } \sigma^2 \text{ is the variance} \\ \text{and } R \text{ the correlation function} \\ Z \sim \mathcal{N}(0, \sigma^2 R) \end{array} \right.$$

- Parametric choices:

- F : polynomial of degree 1 $\beta F(x) = \beta_0 + \sum_{i=1}^p \beta_i x_i$
- R : stationary \Rightarrow covariance function

Example: Gaussian covariance $R(x, u) = R(x - u) = \exp\left(-\sum_{i=1}^p \theta_i |x_i - u_i|^2\right)$

Anisotropy: θ_i s are not equal (correlation length of each input variable)

Gaussian process metamodel (2/2)

■ Joint distribution :

- Gaussian process (GP) model : $Y(x) = \beta F(x) + Z(x), x \in \mathcal{H}^p$
- Learning sample (LS) of N simulations : (X_{LS}, Y_{LS})

$$X_{LS} = (x^{(1)}, \dots, x^{(N)}), F_{LS} = F(X_{LS}), R_{LS} = (R(x^{(i)}, x^{(k)}))_{i,k}$$
$$Y_{LS} \sim N(\beta F_{LS}, \sigma^2 R_{LS})$$

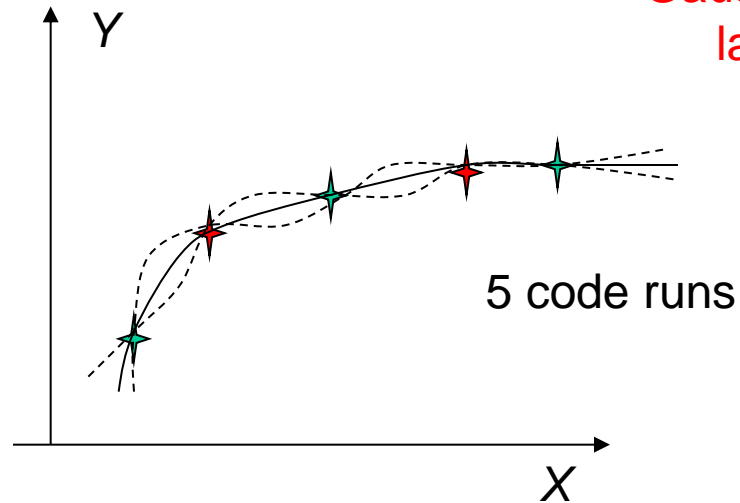
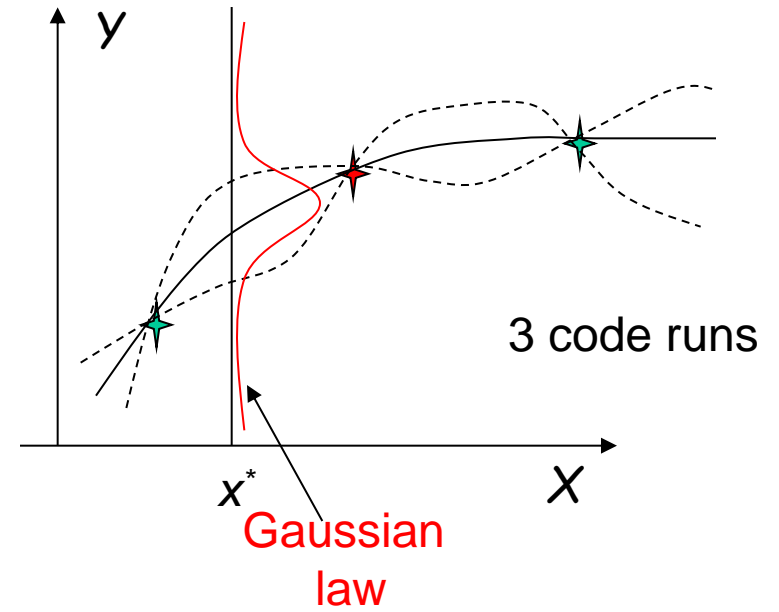
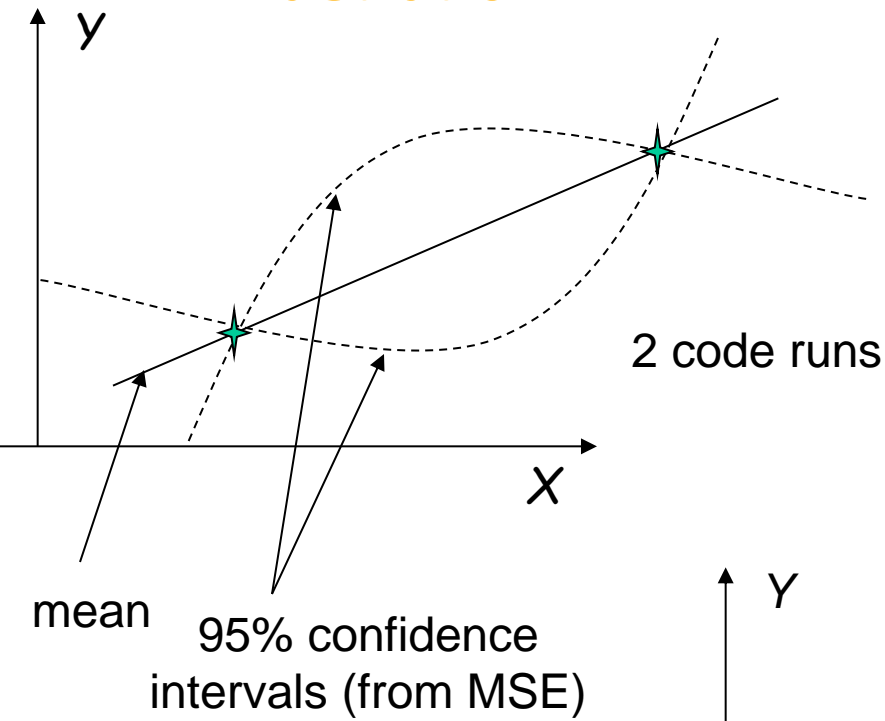
- Conditional GP metamodel :

$$\Rightarrow Y(x)_{|X_{LS}, Y_{LS}} \sim GP$$

$$\left\{ \begin{array}{l} \text{Mean : } \hat{Y}(x) = E[Y(x)_{|X_{LS}, Y_{LS}}] = \beta F(x) + r(x) R_{LS}^{-1} [Y_{LS} - \beta F_{LS}] \\ \text{Covariance : } \text{Cov}(Y(u)_{|X_{LS}, Y_{LS}}, Y(v)_{|X_{LS}, Y_{LS}}) = \sigma^2 (R(u, v) + {}^t r(u) R_{LS}^{-1} r(v)) \\ \Rightarrow \text{Variance} \Rightarrow \text{Mean Square Error (MSE)} \end{array} \right.$$

with $r(x) = [R(x^{(1)}, x), \dots, R(x^{(N)}, x)]$

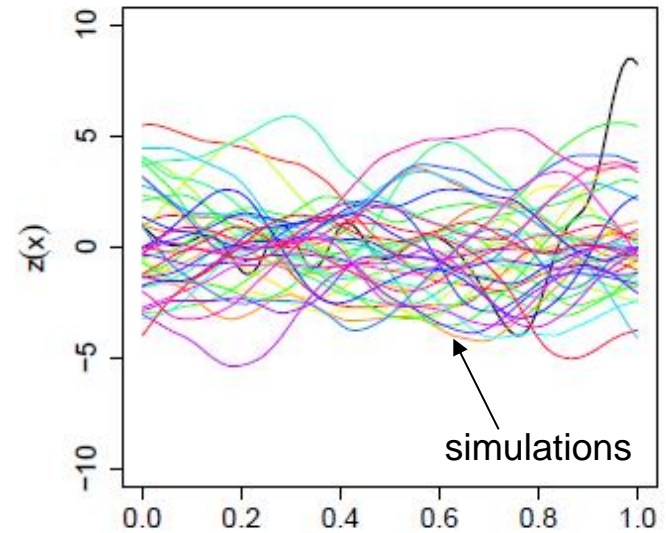
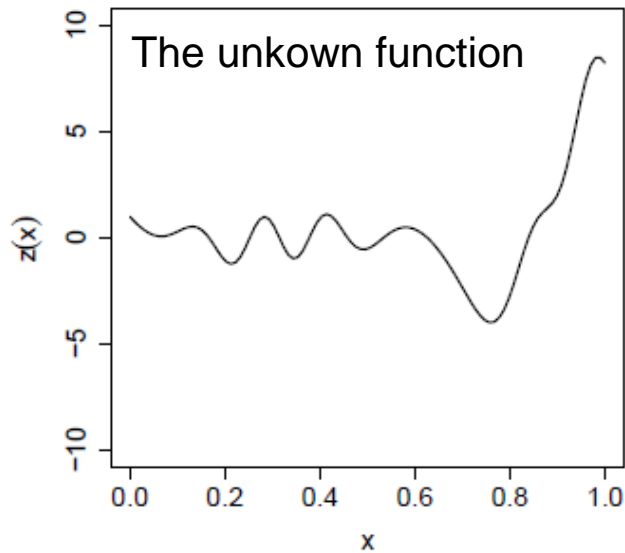
Illustration



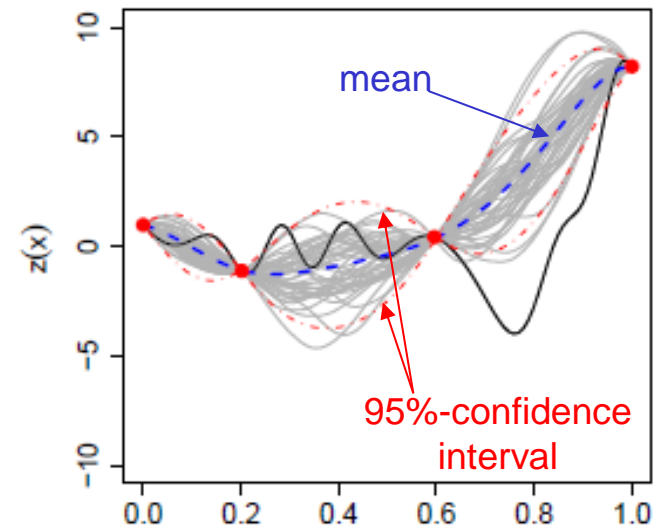
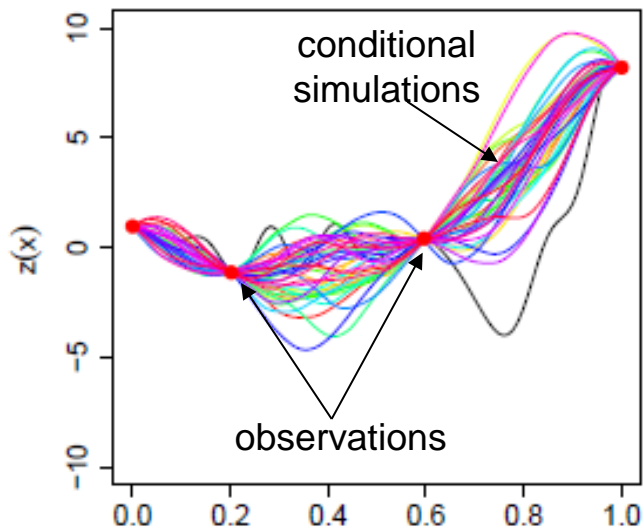
Conclusion: Given a sufficient number of points, we obtain an accurate metamodel

GP metamodel in summary

[Le Gratiot, 2014]



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



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

Its mean gives the GP predictor
Its variance gives the error

Hyperparameters estimation

■ Maximum likelihood method

- Likelihood maximisation on the learning basis (X_s, Y_s) :

$$(\beta^*, \theta^*, \sigma^*) = \underset{(\beta, \theta, \sigma)}{\operatorname{Argmax}} \ln L(Y_{LS}, \beta, \theta, \sigma)$$

- with

$$\ln L(Y_{LS}, \beta, \theta, \sigma) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2} \ln(\det R_{LS}) - \frac{1}{2} \sigma^{-2} [Y_{LS} - \beta F_{LS}]^T R_{LS}^{-1} [Y_{LS} - \beta F_{LS}]$$

- Joint estimation of β and σ :
$$\begin{cases} \beta^* = [{}^t F_{LS} R_{LS}^{-1} F_{LS}]^{-1} {}^t F_{LS} R_{LS}^{-1} Y_{LS} \\ \sigma^{2*} = \frac{1}{N} [{}^t Y_{LS} - \beta^* F_{LS}] R_{LS}^{-1} [Y_{LS} - \beta^* F_{LS}] \end{cases}$$

- Estimation of correlation parameters θ :

$$(\theta^*) = \underset{\theta}{\operatorname{Argmin}} \psi(\theta) \quad \text{with} \quad \psi(\theta) = |R_{LS}|^{-1/N} \sigma^{2*}$$

Estimation and validation

$$R(\mathbf{u}, \mathbf{v}) = R(\mathbf{u} - \mathbf{v}) = \exp\left(-\sum_{i=1}^p \theta_i |u_i - v_i|^2\right)$$

■ Hyperparameters $(\theta_i)_{i=1\dots K}$ estimated by likelihood maximization

Simplex method, **stochastic algorithms**

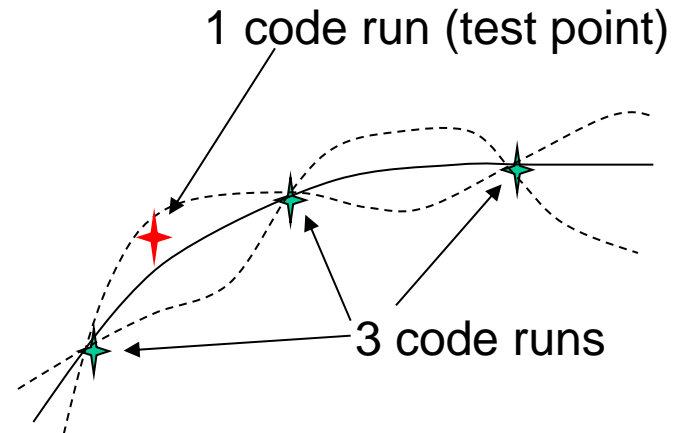
Problems in high dimensional context ($p > 10$), can be solved by sequential fitting algorithms [Marrel et al. 2008]

■ Predictor validation:

Predictivity coefficient:

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

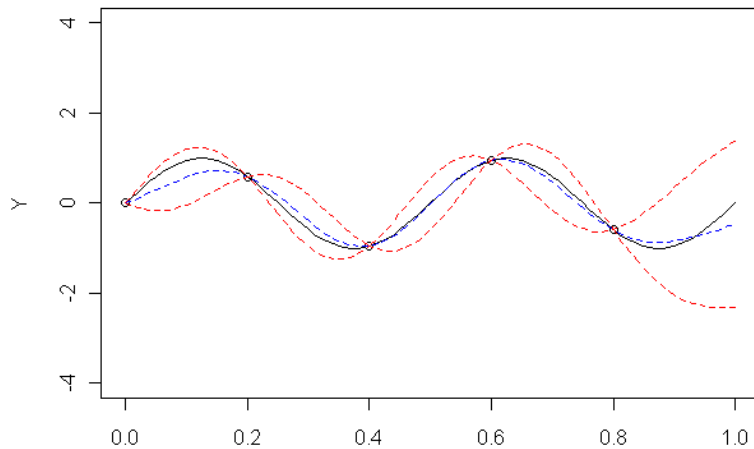
- Test sample
- or leave-one-out
- or k -fold cross validation



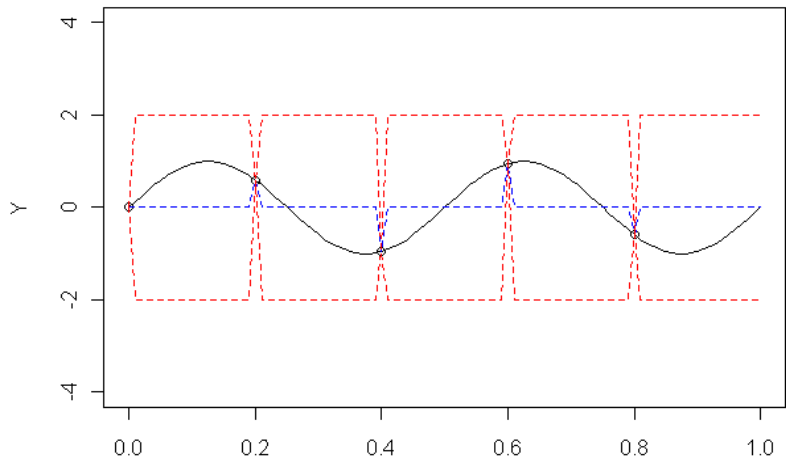
■ MSE validation: Percentage of predicted values inside confidence bounds

Effects of the hyperparameters θ and σ

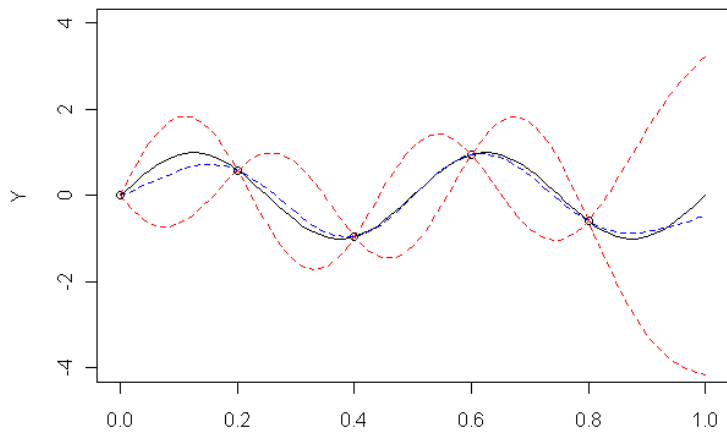
$$f(x) = \sin(4\pi x)$$



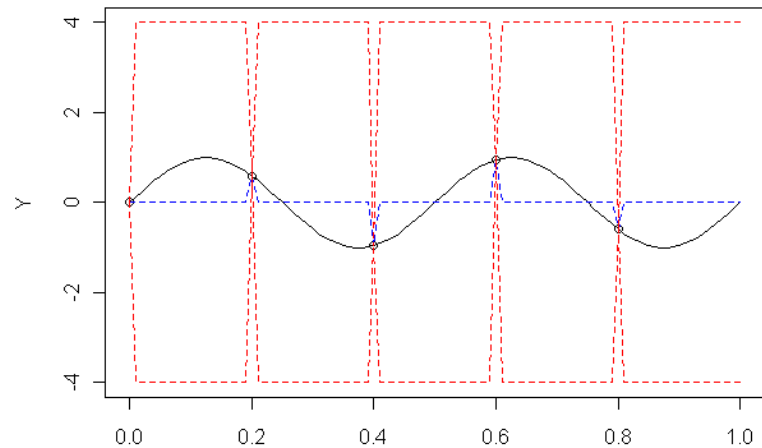
$$\sigma^2 = 1; \theta = 0.2$$



$$\sigma^2 = 1; \theta = 10^{-4}$$



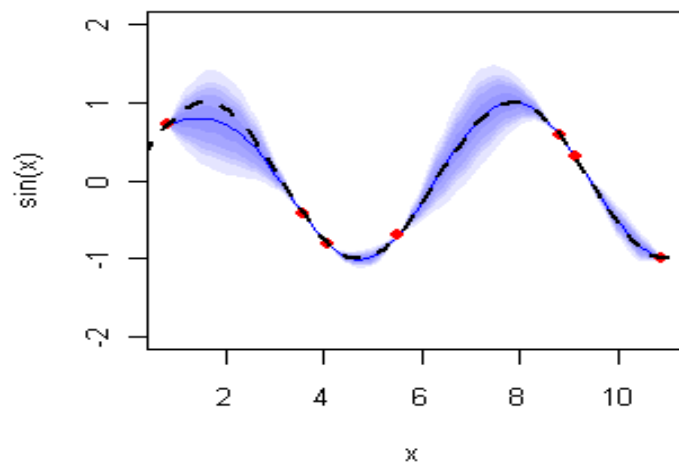
$$\sigma^2 = 4; \theta = 0.2$$



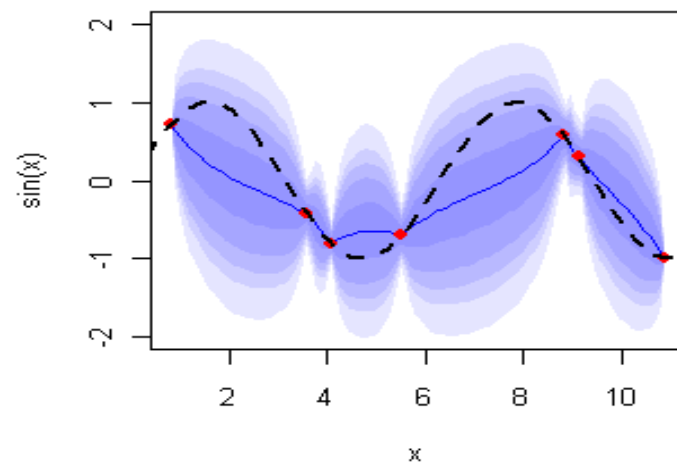
$$\sigma^2 = 4; \theta = 10^{-4} \quad [\text{Le Gratiot, 2011}]$$

Effects of the covariance structure

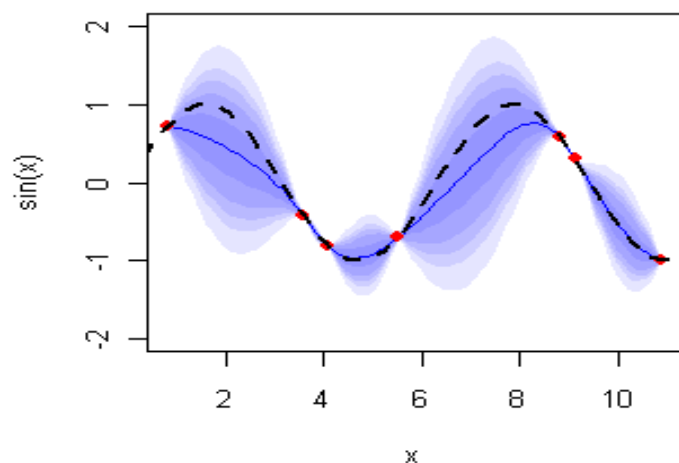
gauss



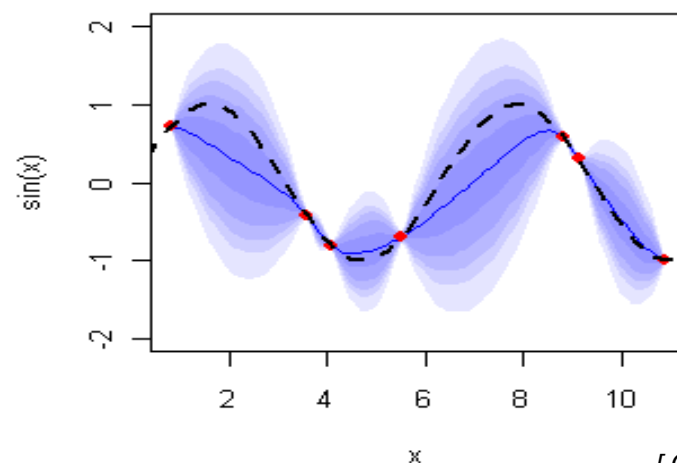
exponential



matern5_2



matern3_2





Planification d'expériences adaptative basée sur le métamodèle processus gaussiens

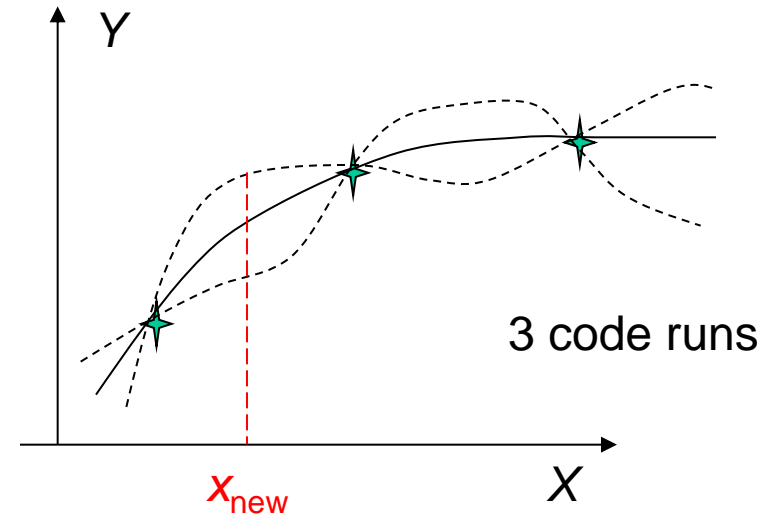
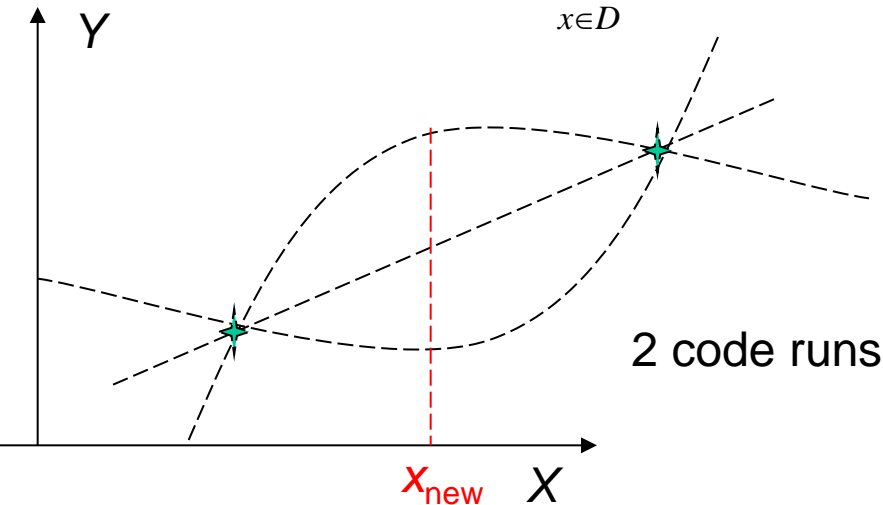
The best way to build GP: Model-based adaptive designs

Example: criterion of the Gaussian process MSE (Mean Square Error)

$$MSE(x) = \sigma^2 + {}^t r(x) R_{LS}^{-1} r(x) + u(x) ({}^t (\beta F_{LS}) R_{LS}^{-1} \beta F_{LS})^t u(x)$$

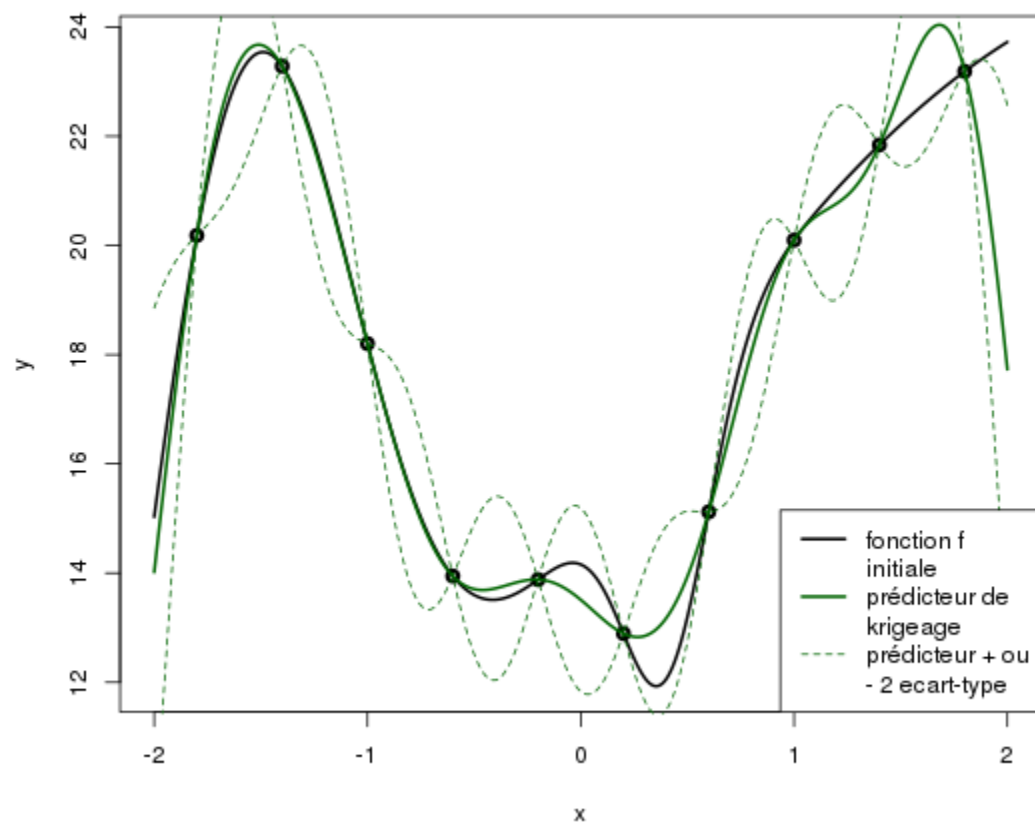
$$u(x) = \beta F(x) - {}^t k(x) R_{LS}^{-1} \beta F_{LS}$$

$$x_{\text{new}} = \arg \max_{x \in D} MSE(x)$$

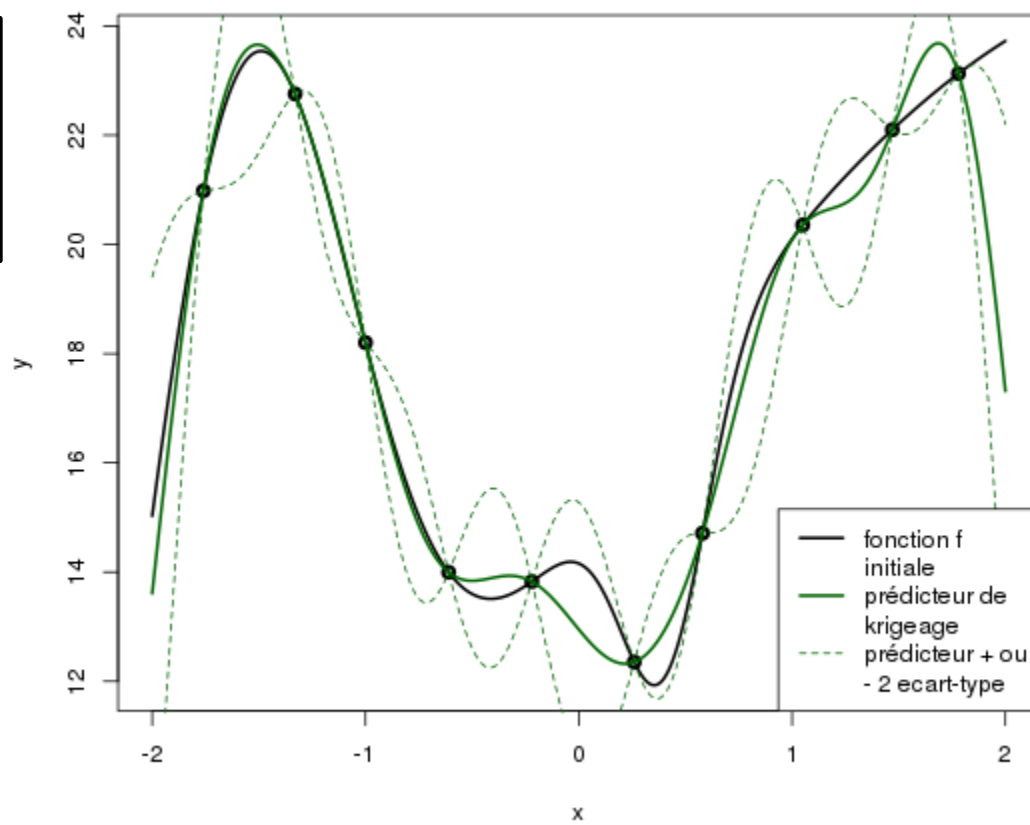


Remark: other criteria are possible (e.g. focusing to active variables)

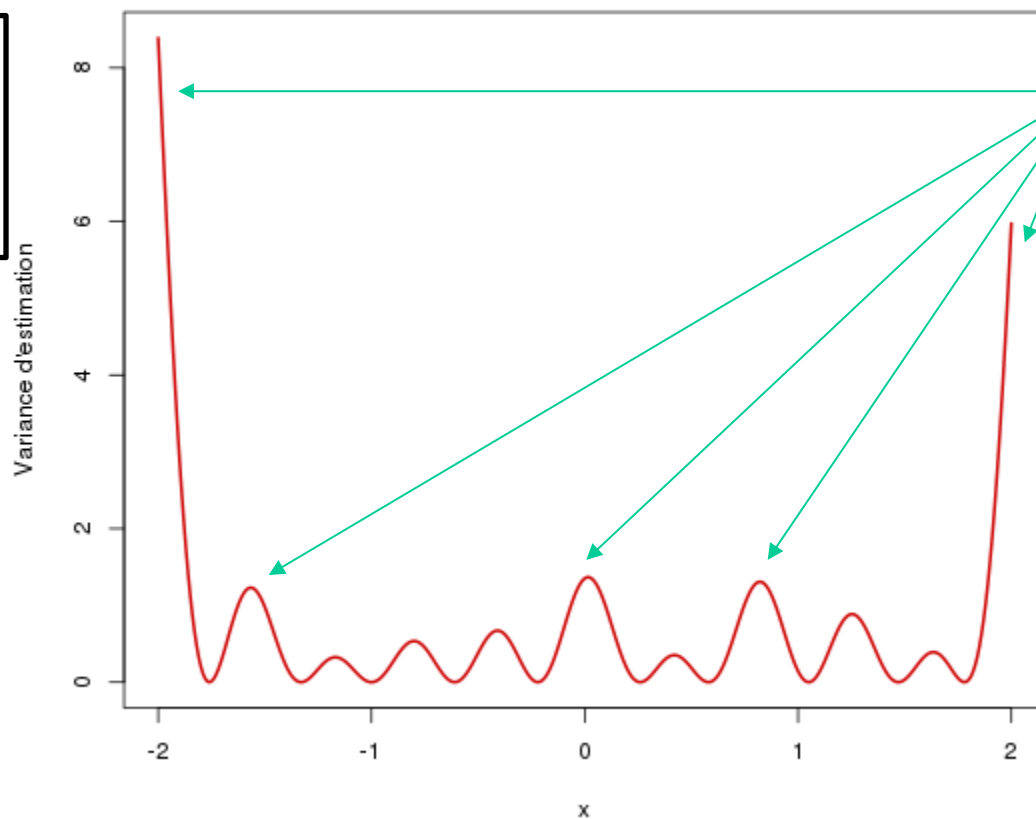
Conclusion: Model-based adaptive designs are the most efficient ones,
but are not always applicable
In practice, we need to initiate the process with a space-filling design



**Initial design
(10 points)**

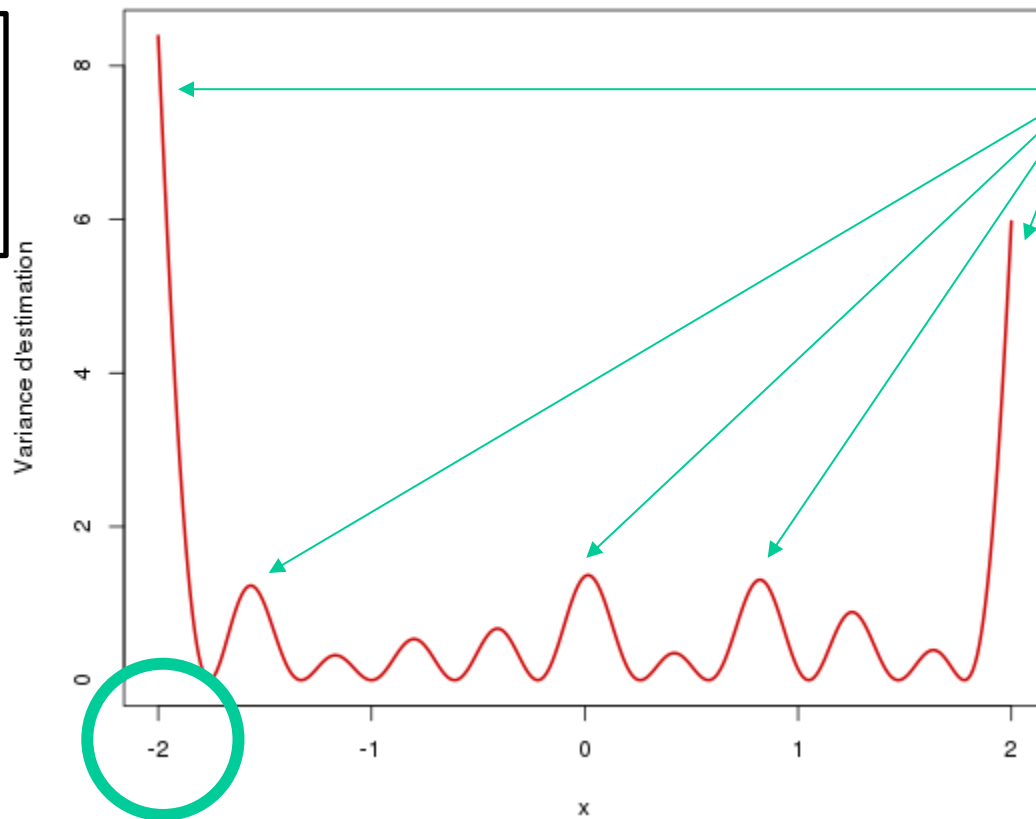


**Initial design
(10 points)**



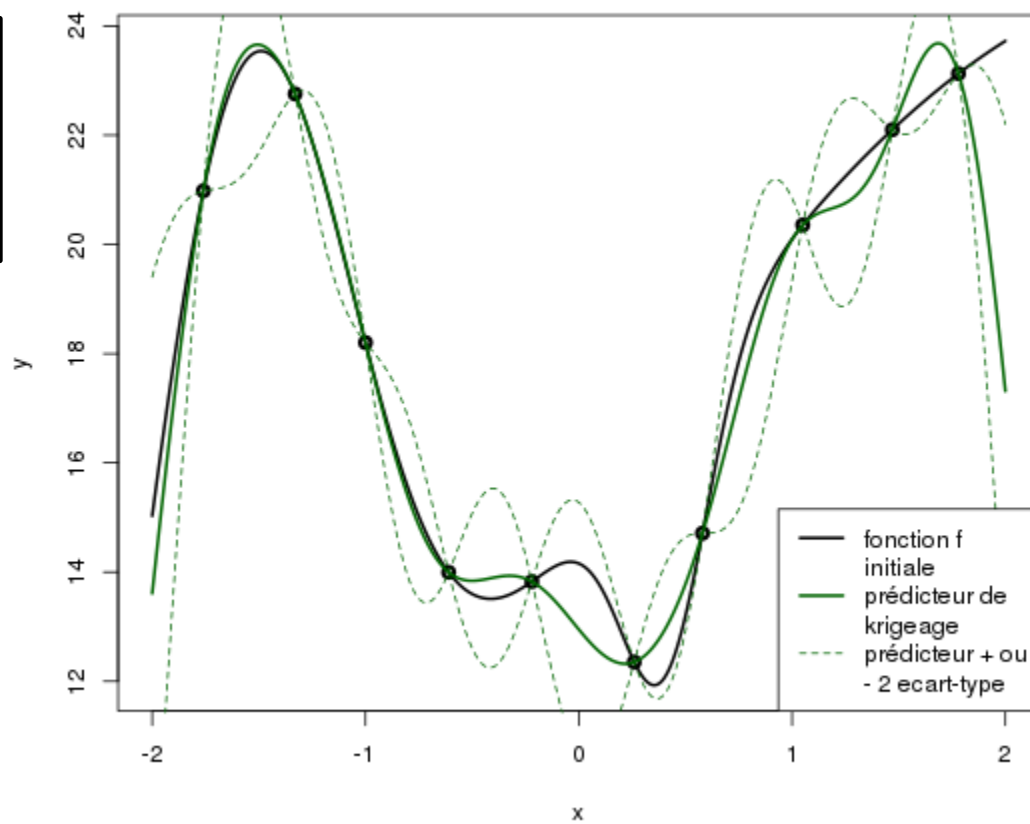
**Idée : acquérir de
l'information là où
le modèle est
« mauvais »**

**Initial design
(10 points)**

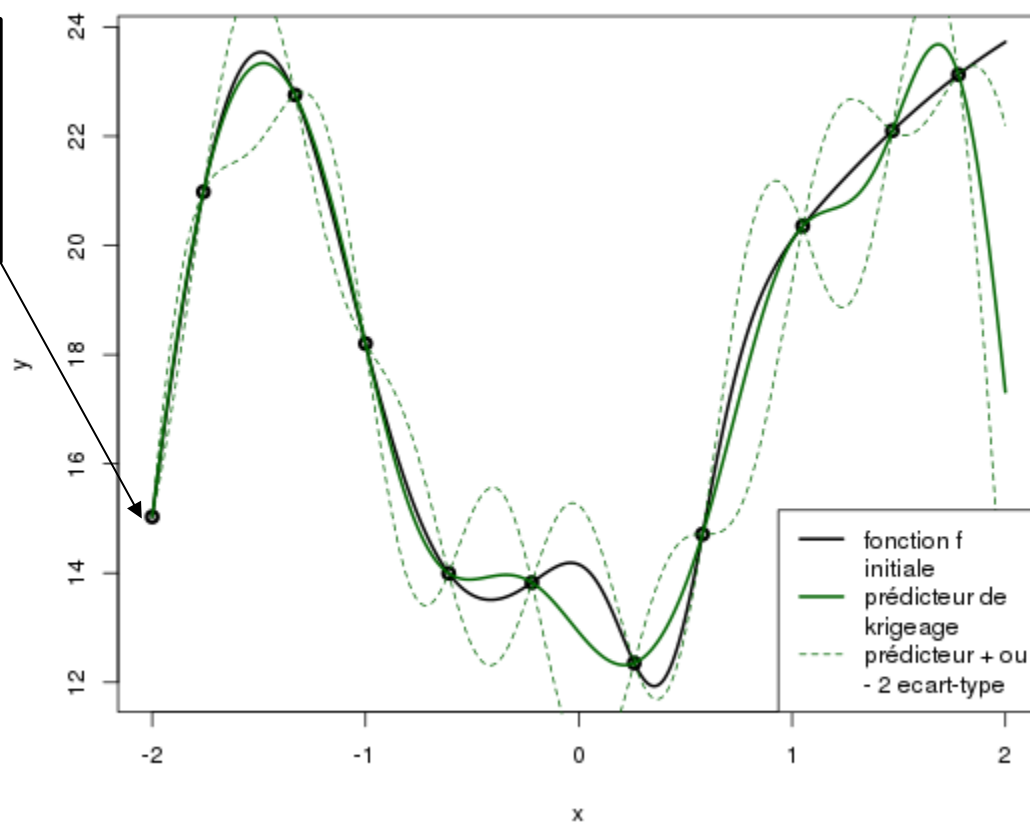


**Idea: bring
information where
the model has
large errors**

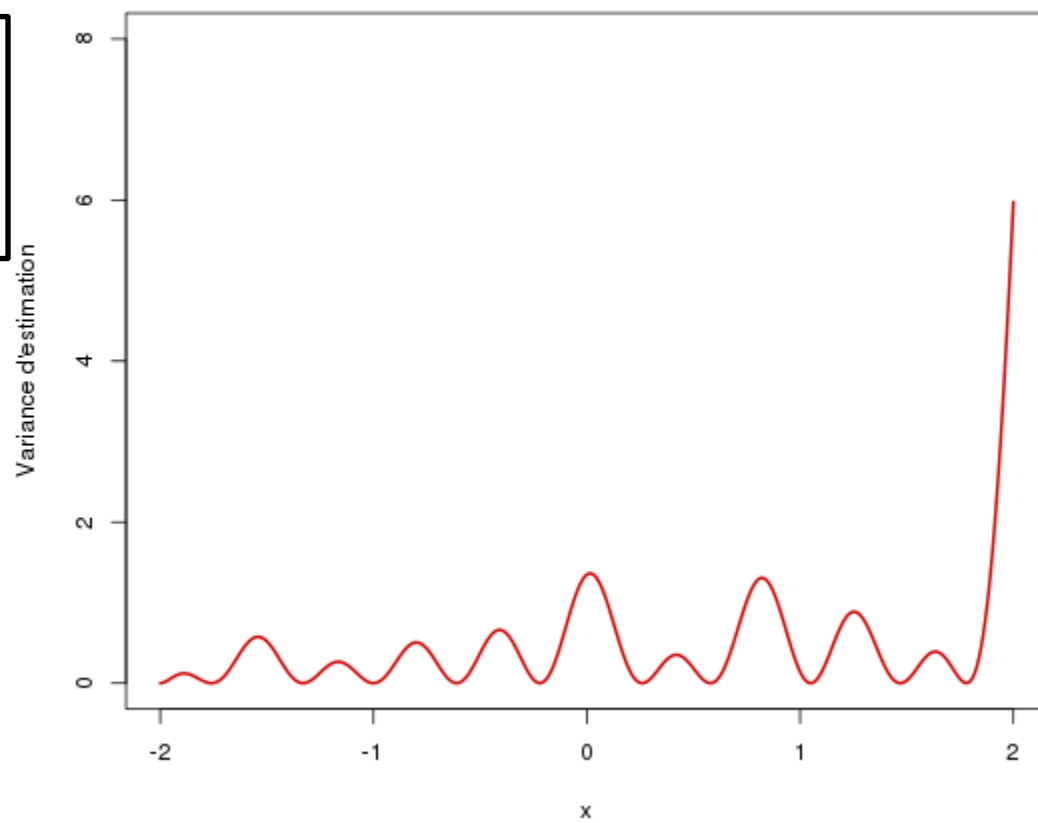
**Initial design
(10 points)**



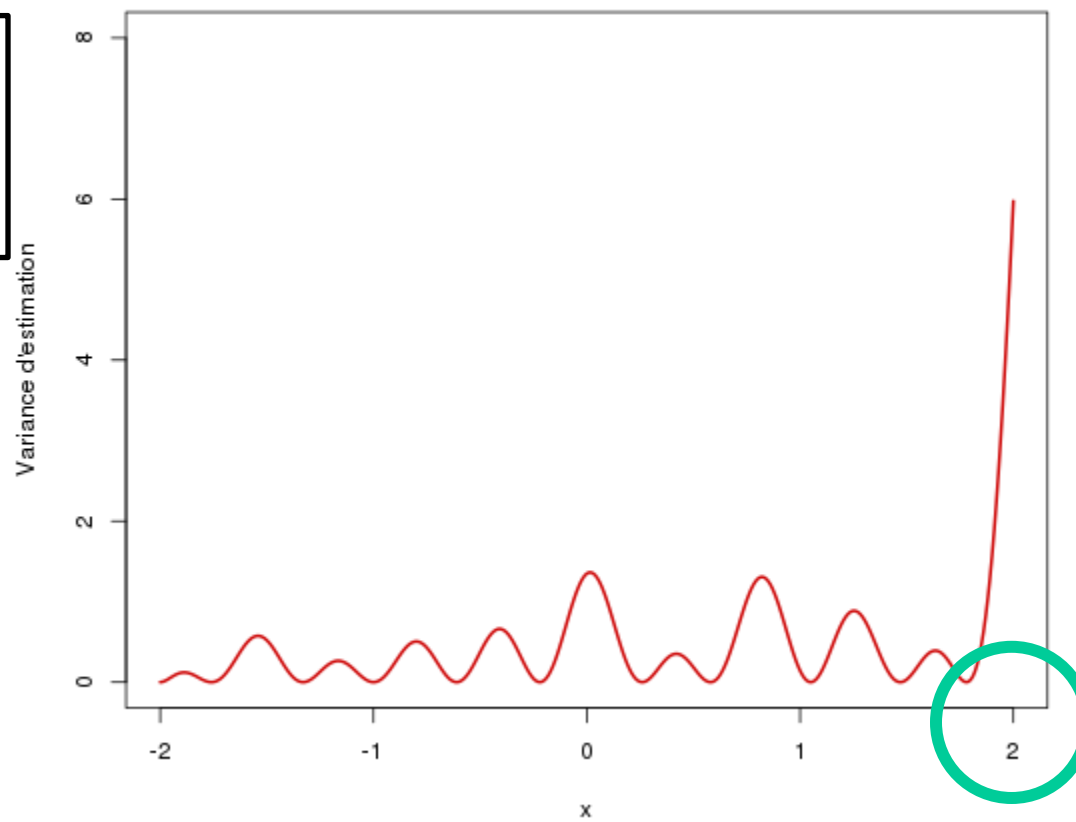
First step



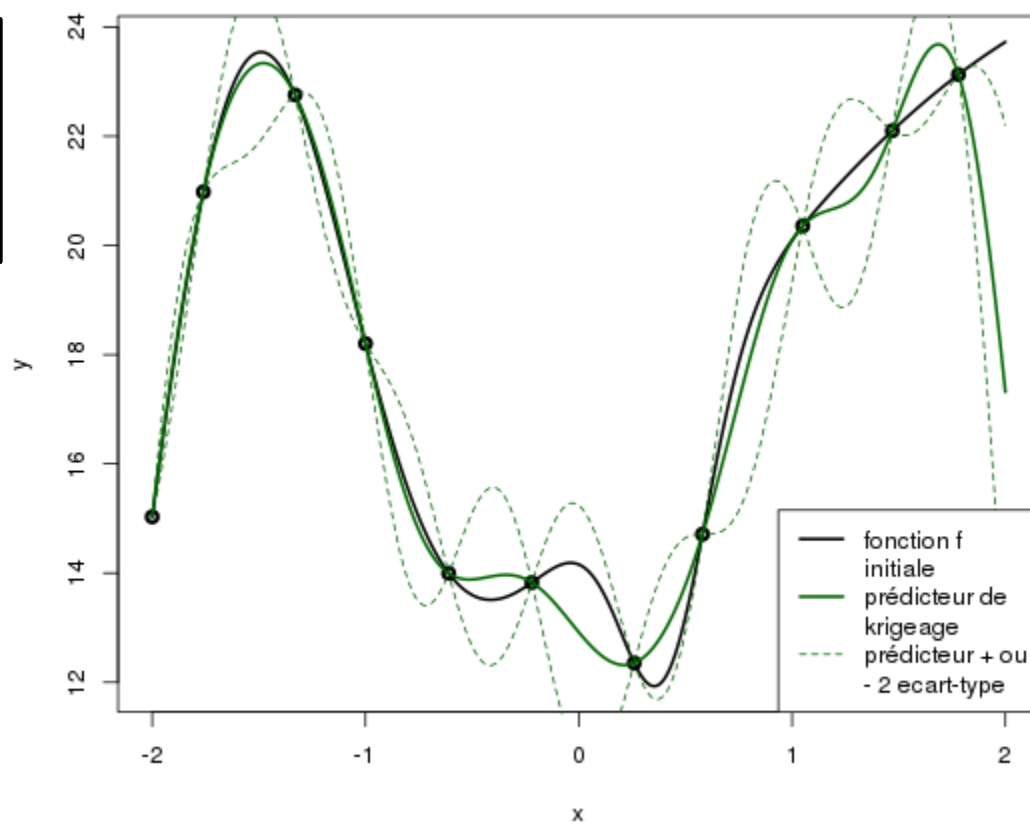
First step



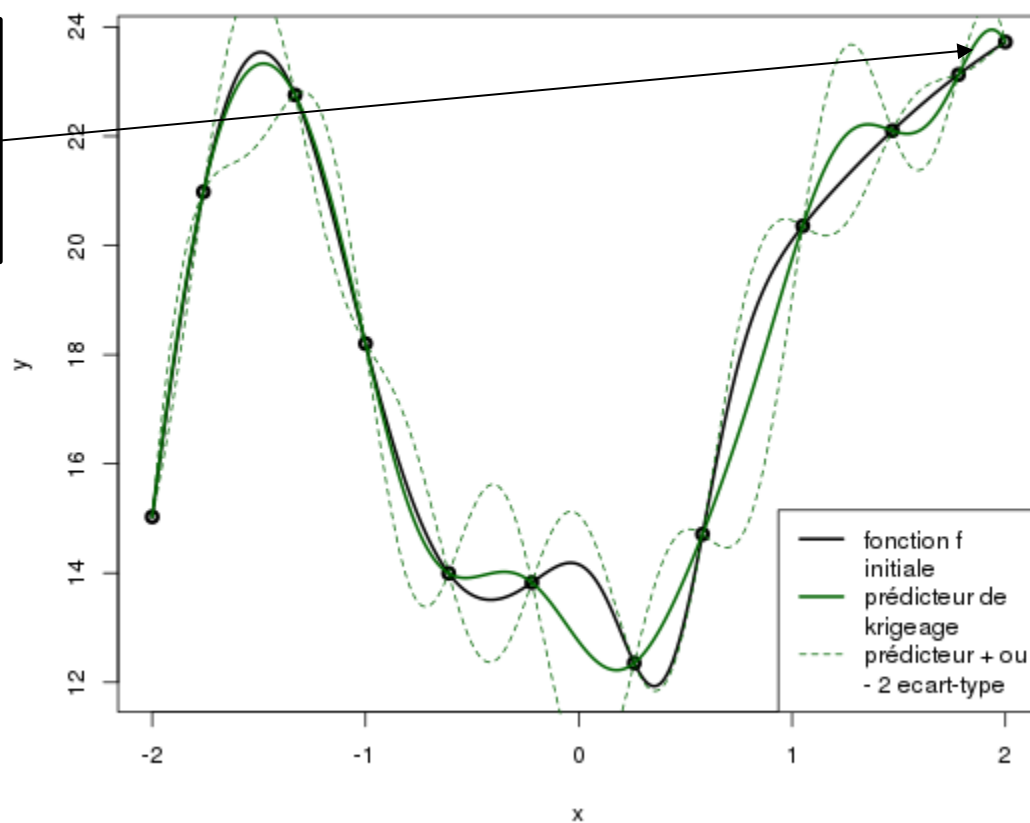
First step



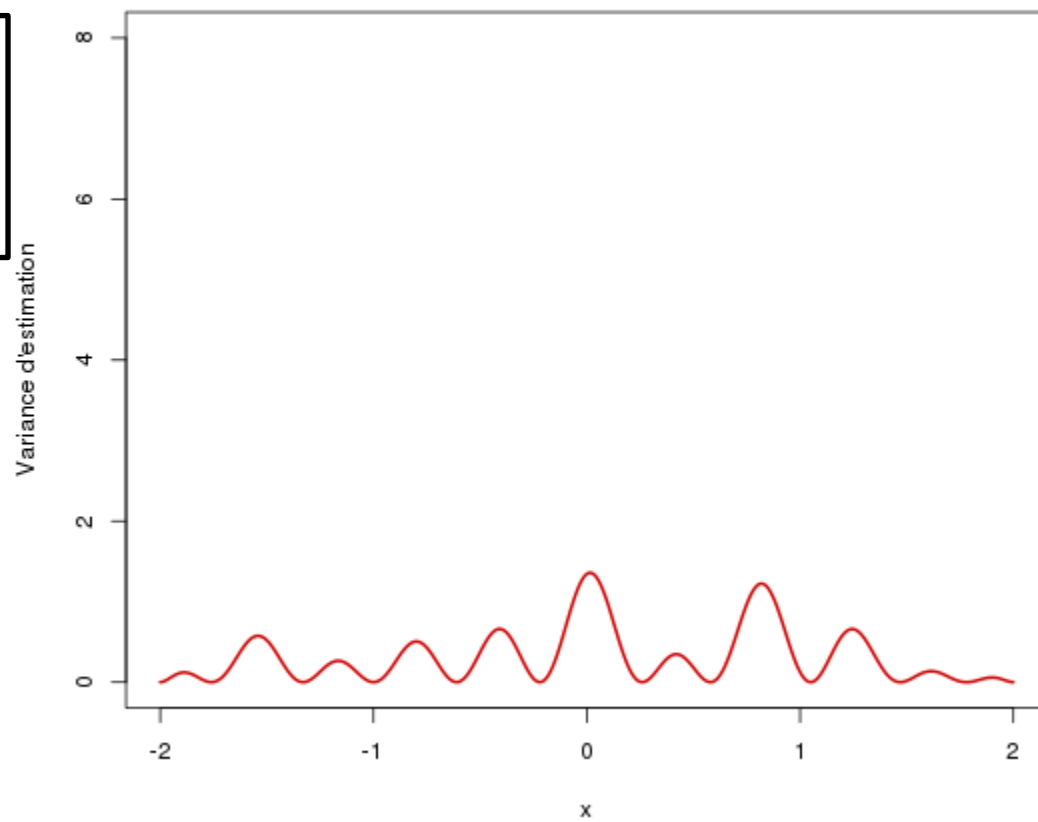
First step



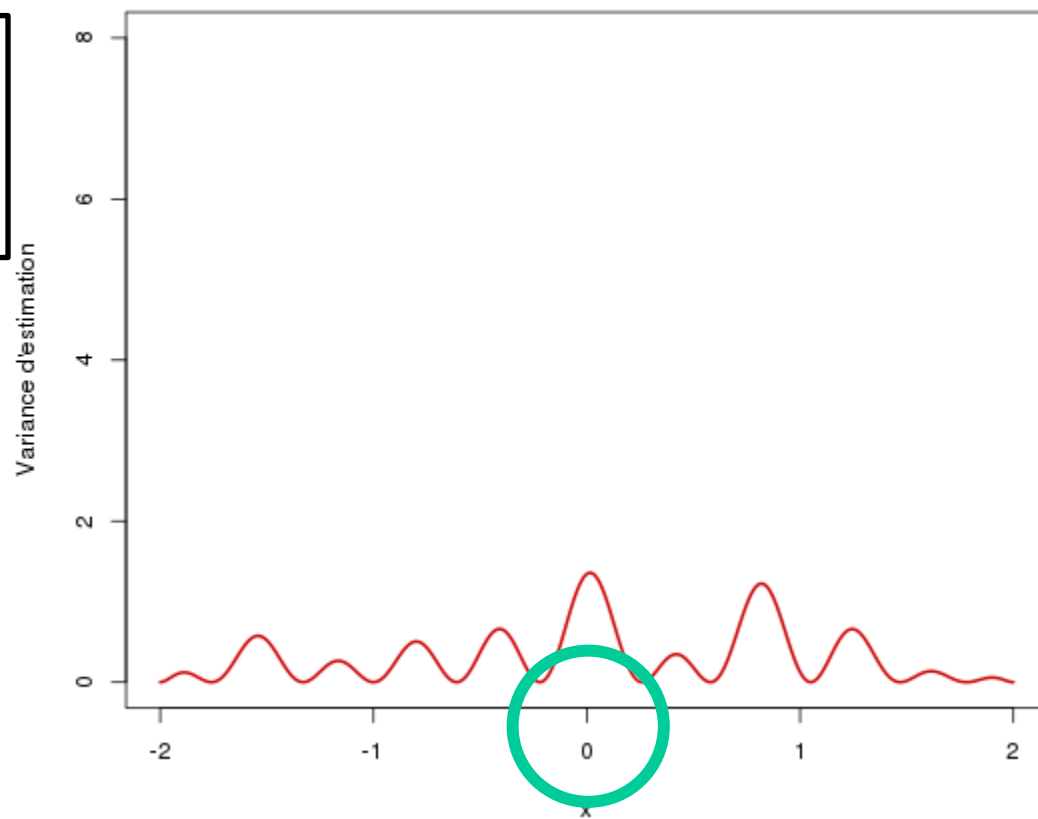
Second step



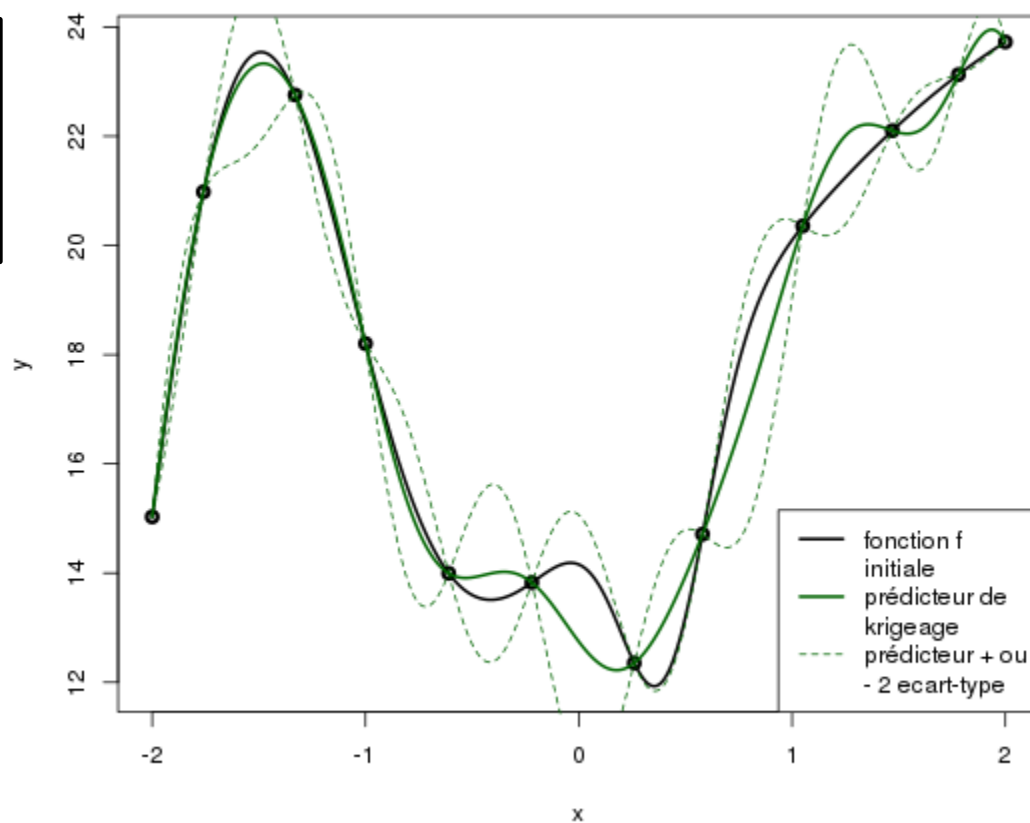
Second step



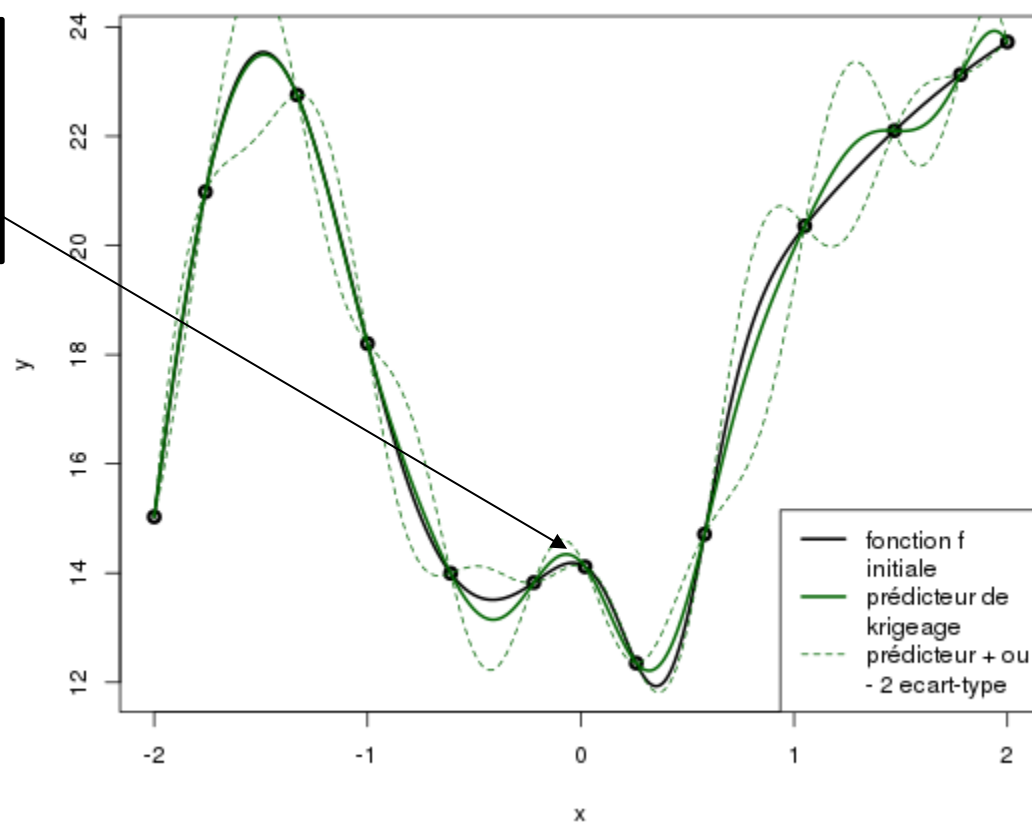
Second step



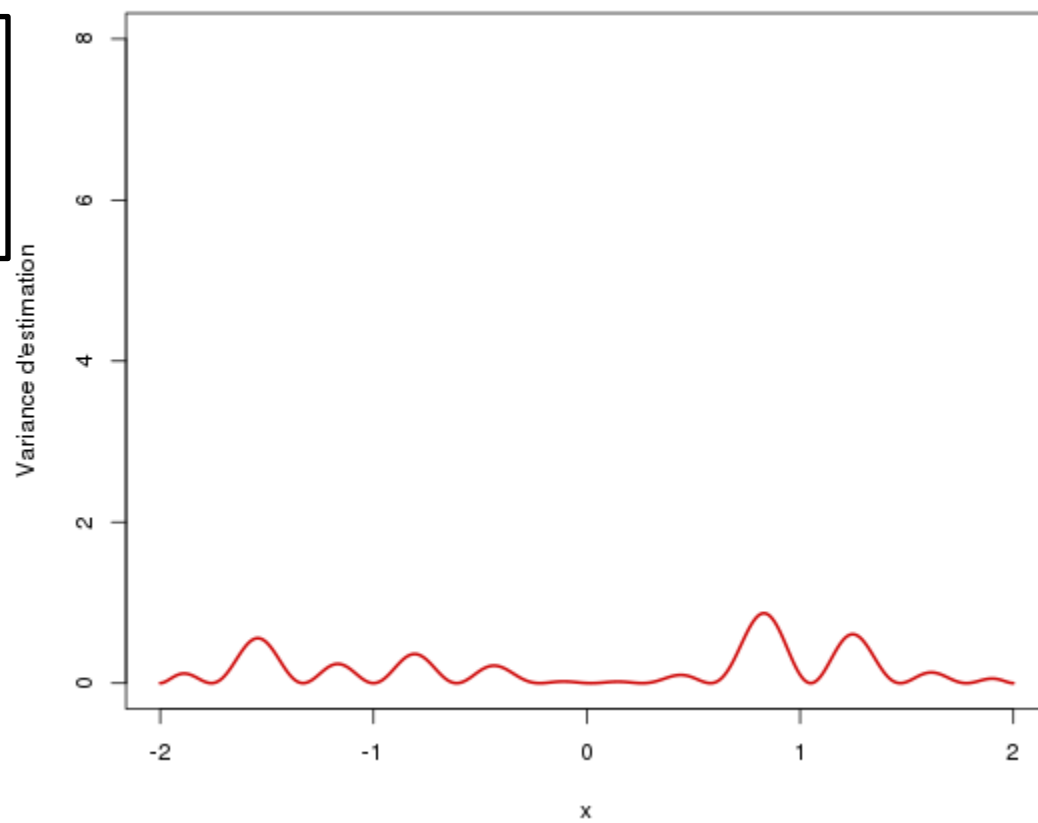
Second step



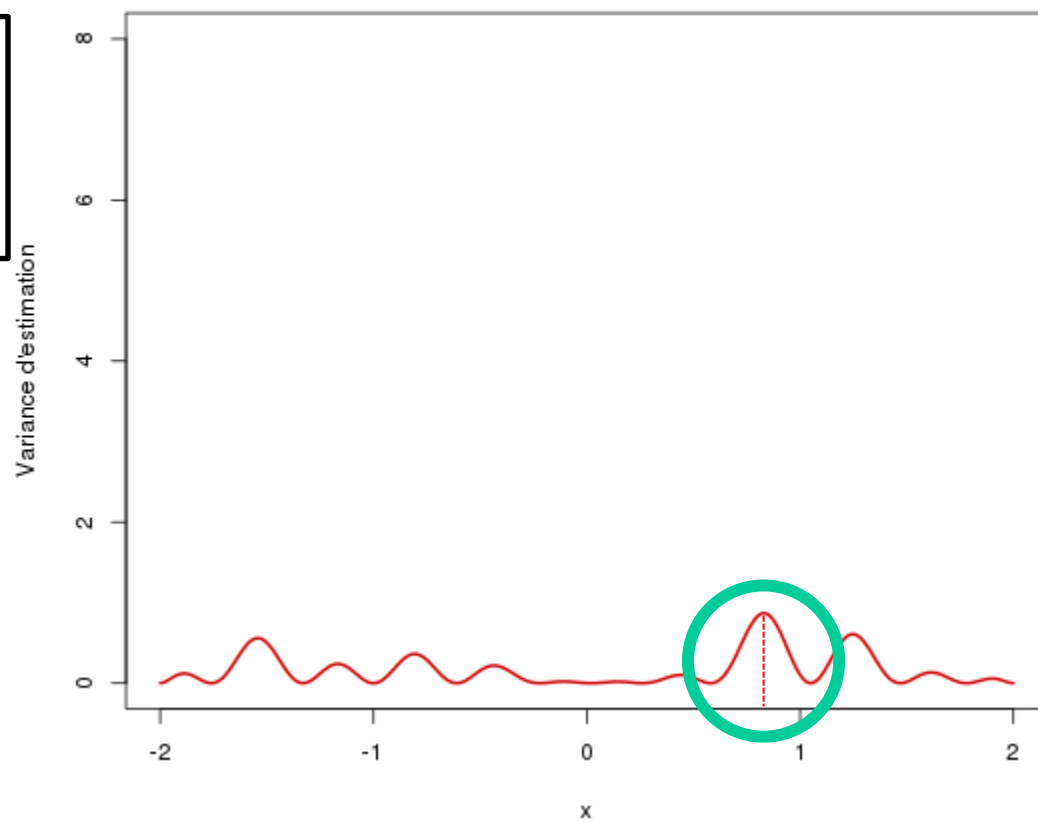
Third step



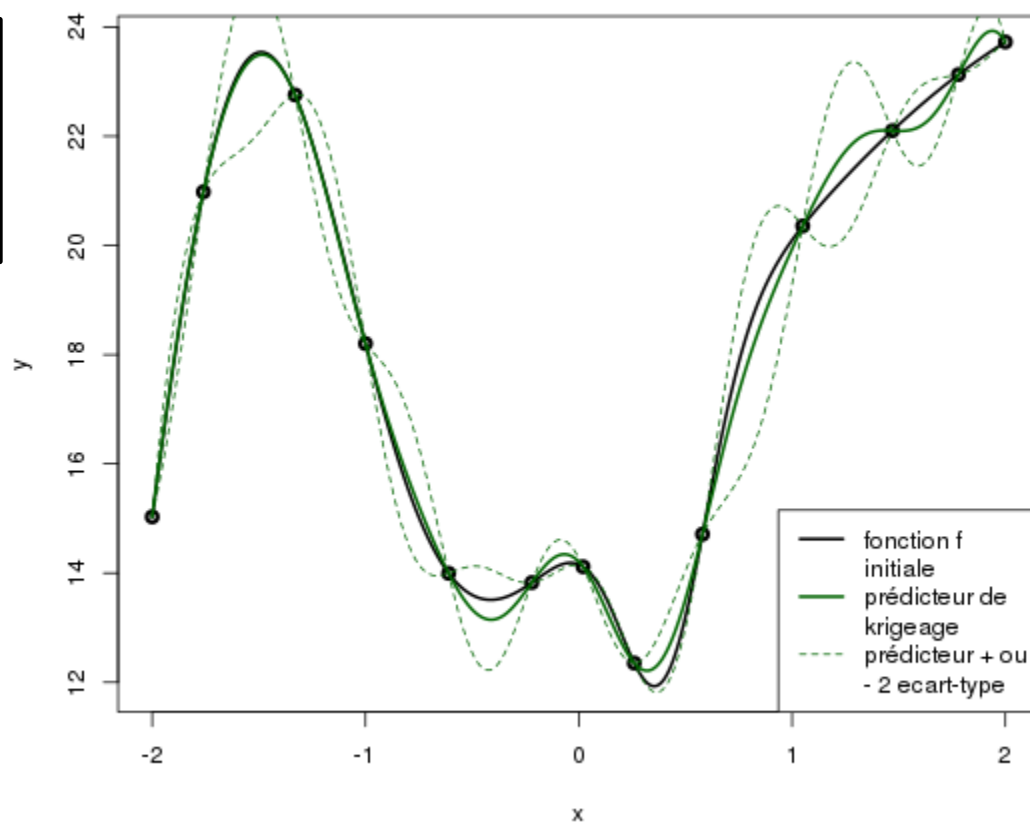
Third step



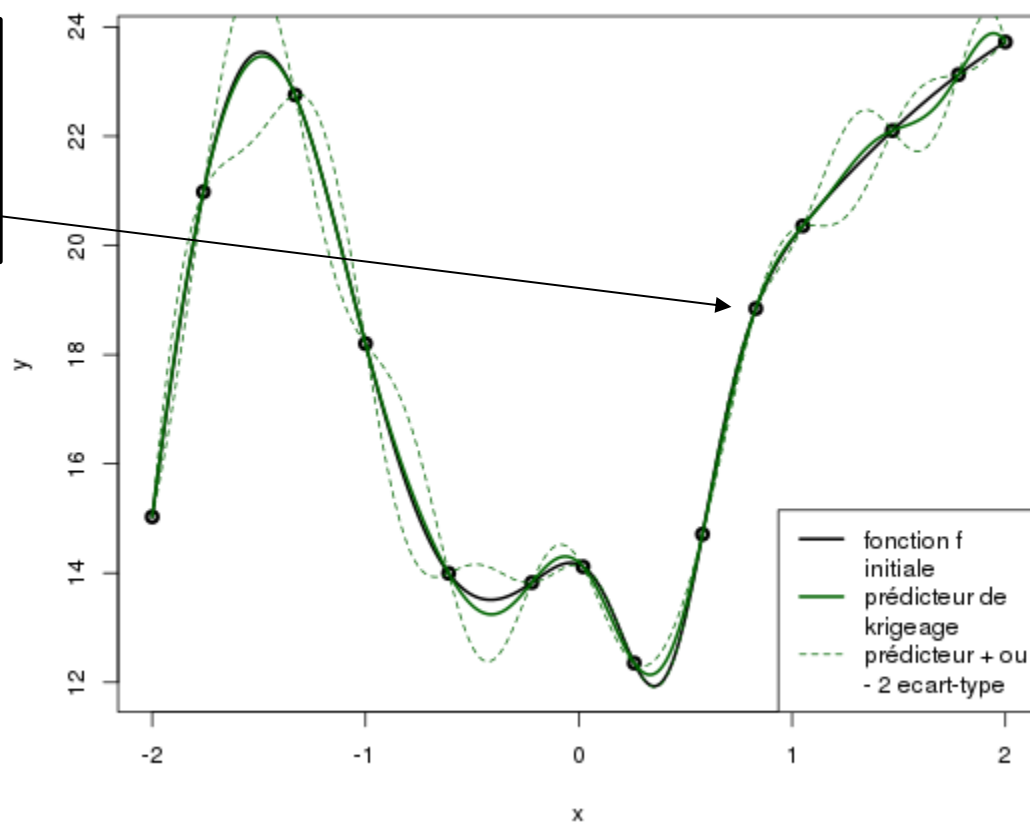
Third step



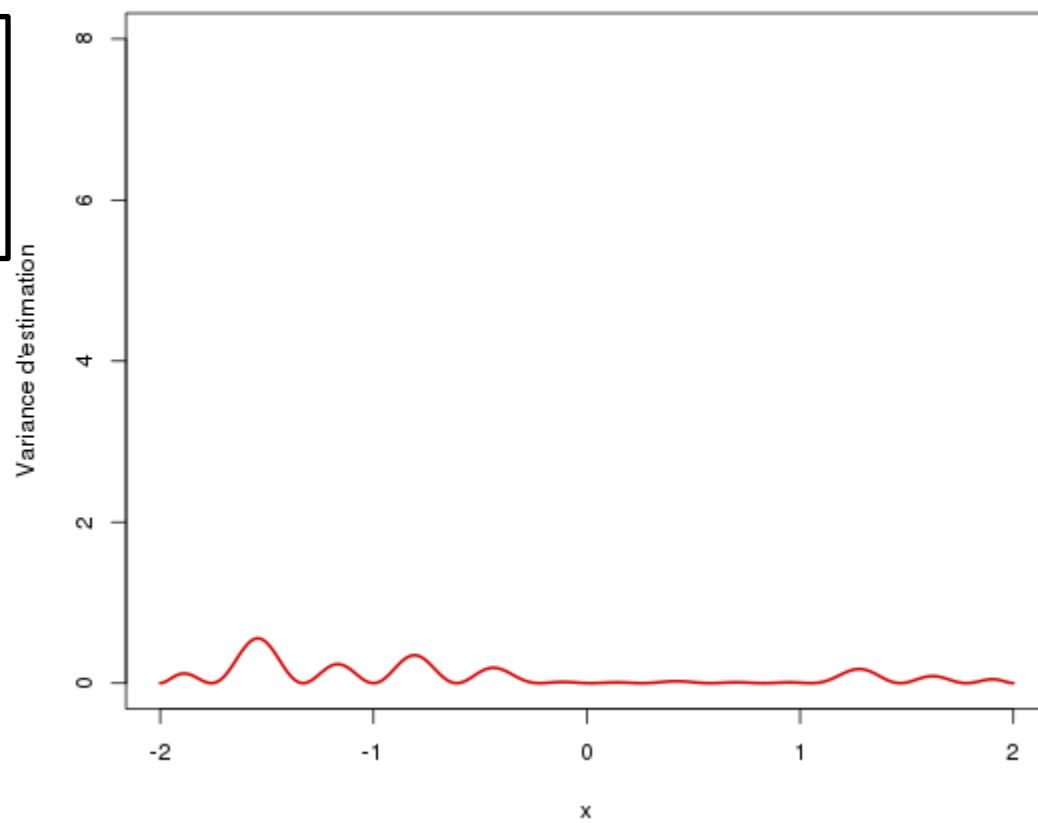
Third step



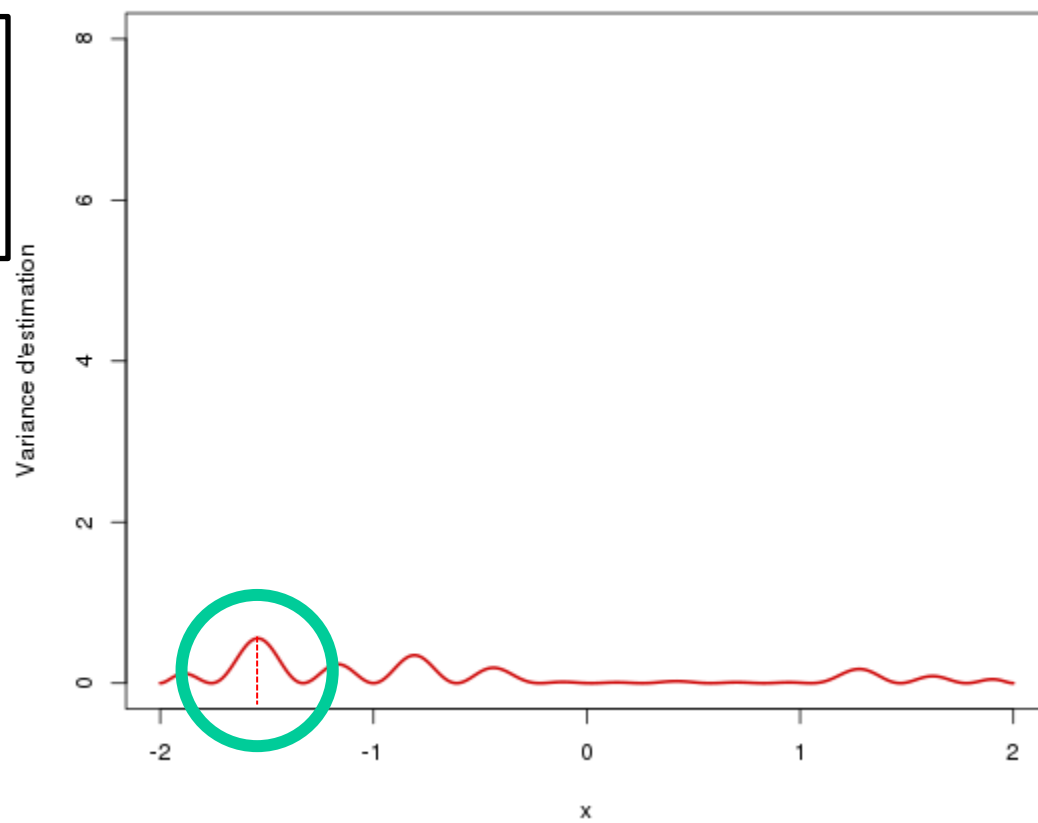
Fourth step



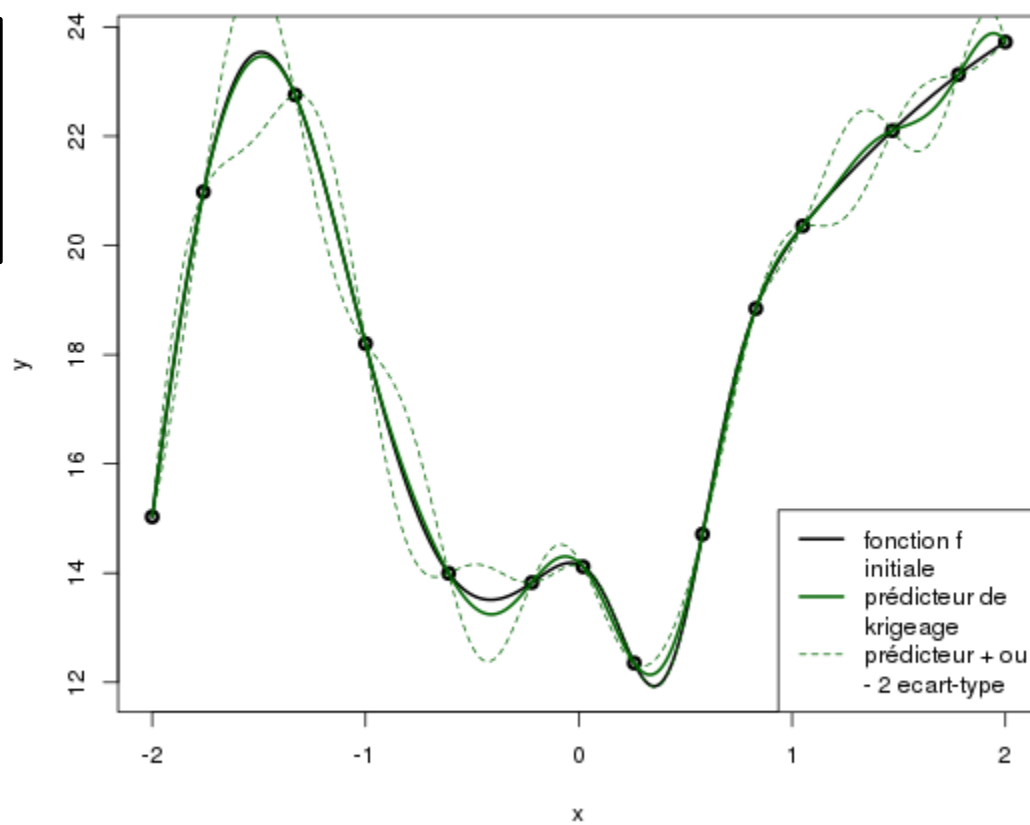
Fourth step



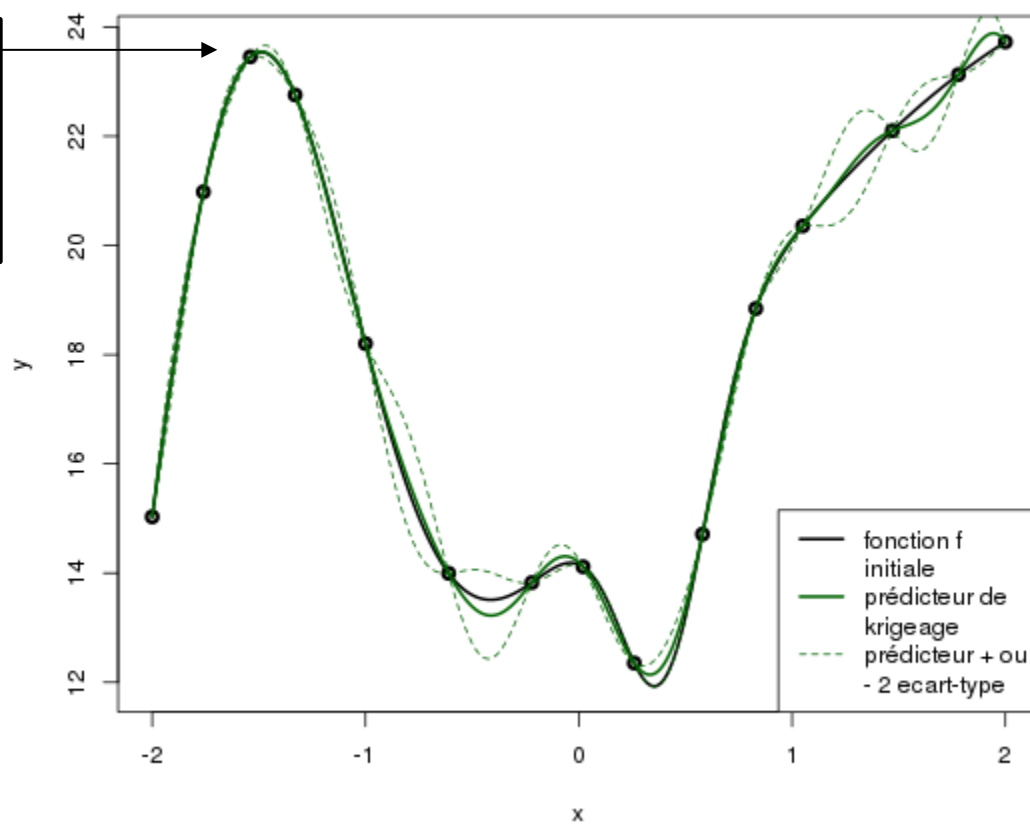
Fourth step



Fourth step



Fifth step



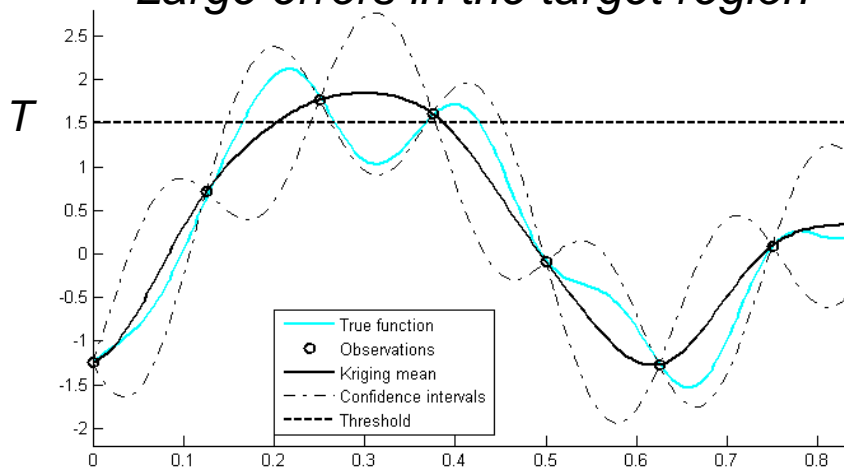
Estimation of rare events probability using GP

[Bect et al. 2012]

Industrial problems: safety analysis with computer code

Problem: find $P_f = \text{Prob} [f(X) > T]$ with $X = \text{random inputs}$; $T = \text{threshold}$

*Reasonable variance everywhere
Large errors in the target region*



[from: Picheny et al. 2010]

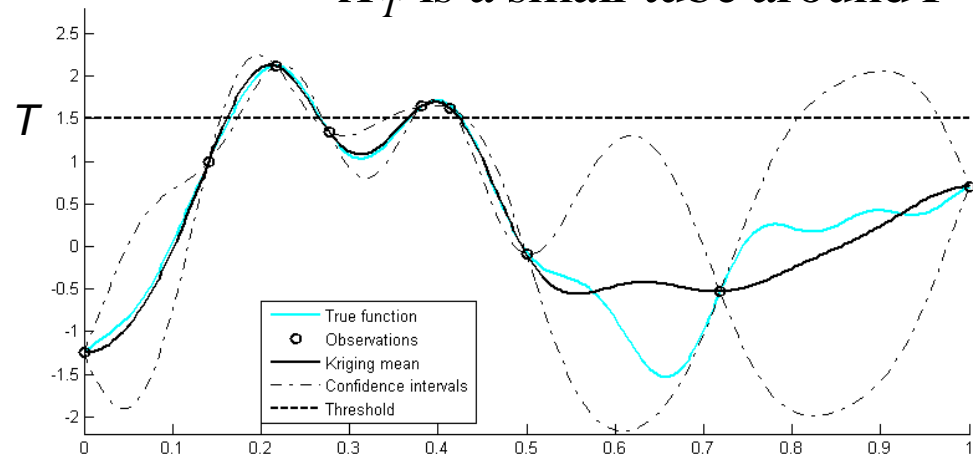
*Large variance in non-target region
Good accuracy in target region*

☺ **New adaptive design**

$$X^* = \arg \min_X (IMSE_T)$$

$$IMSE_T = \int MSE(x) 1_{X_T}(x) dx$$

X_T is a small tube around T



Optimisation of a model output using kriging

Industrial design with costly computer code

Problem: find the values of X which minimize the model output

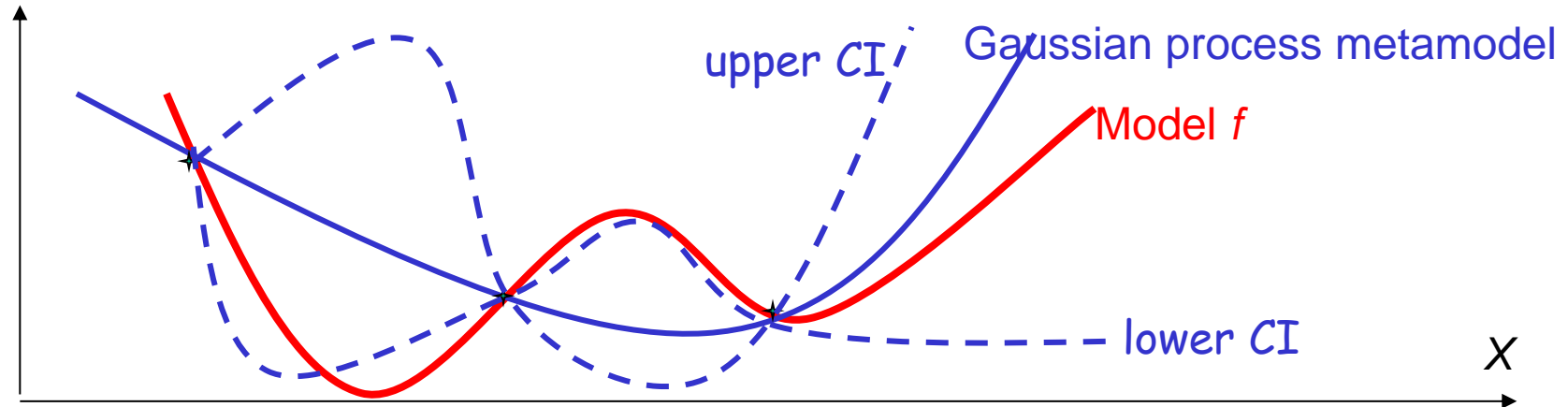
$$X^* = \arg \min_{X \in D} f(X)$$

If f is costly, a natural solution would be to optimize a metamodel of f :

☠ dangerous idea because the metamodel tends to smooth the true model

Gp metamodel allows to take into account the metamodel error,

😊 and to define the expected improvement $EI(X)$ for each $X \in D$



$$EI(x) = E[\max(0 , \text{observed minimum} - f(x))]$$

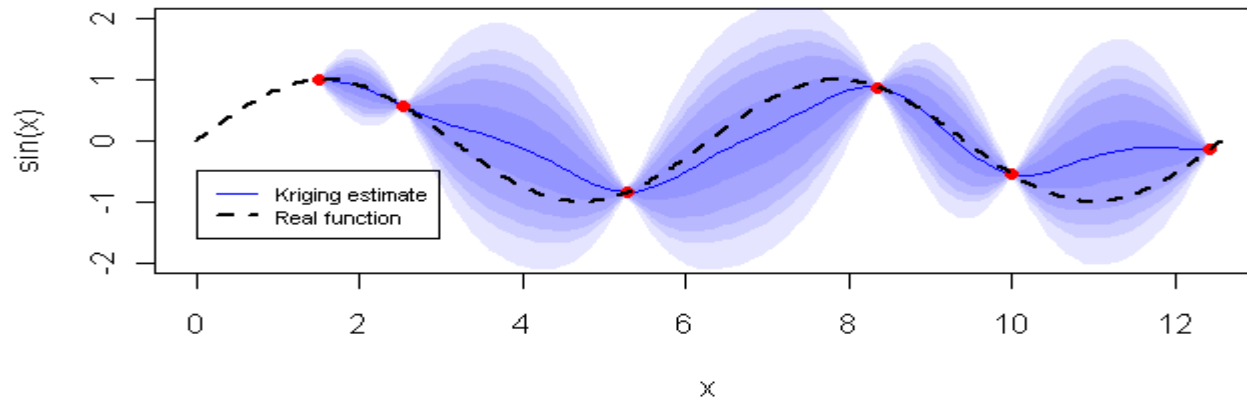
$$EI(x) = \tilde{\sigma}(x)[u(x)\Phi(u(x)) + \phi(u(x))] \text{ with } u(x) = \frac{y_{\min}}{\tilde{\sigma}(x)}$$

Adaptive design for optimization: EGO algorithm

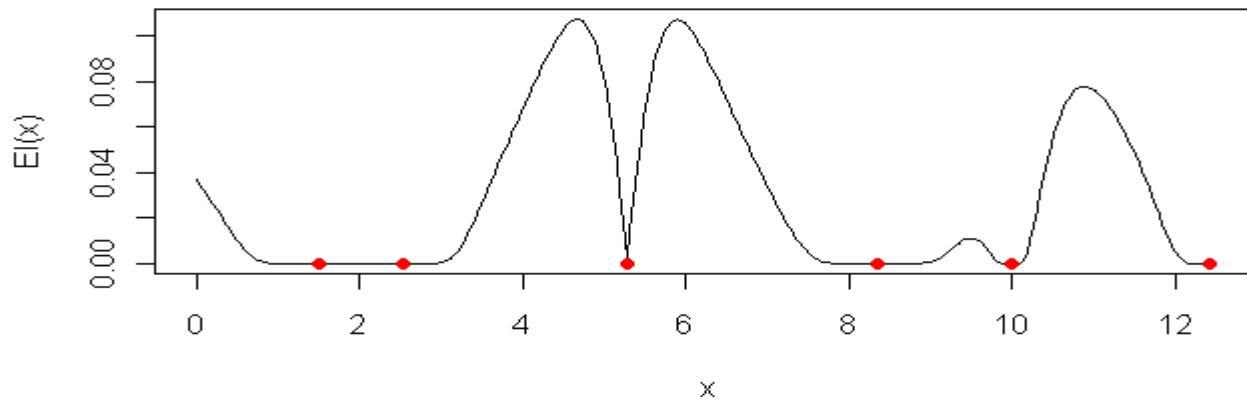
[Jones et al. 1998]

EGO: step 0

kriging the sinus function

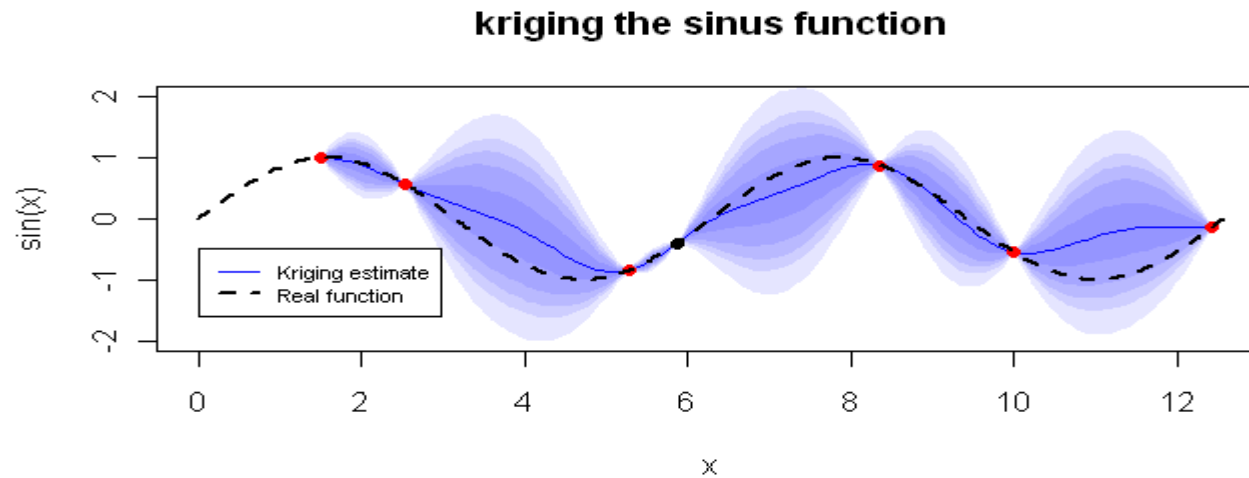


[Chevalier, 2011]

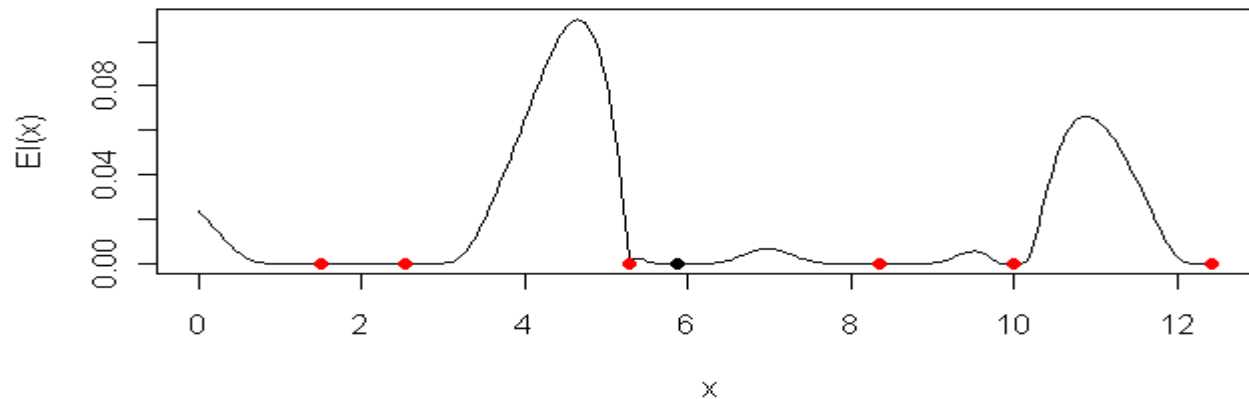


Adaptive design for optimization: EGO algorithm

EGO: step 1



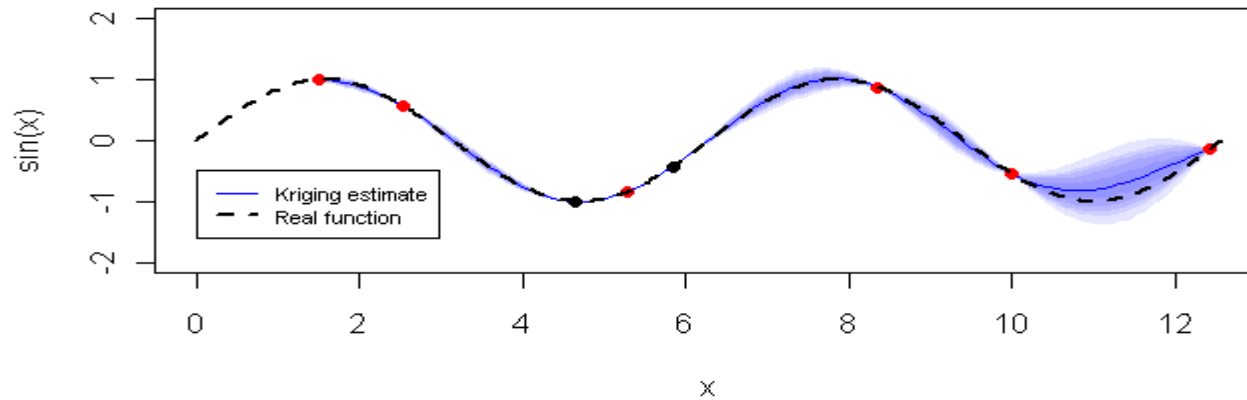
[Chevalier, 2011]



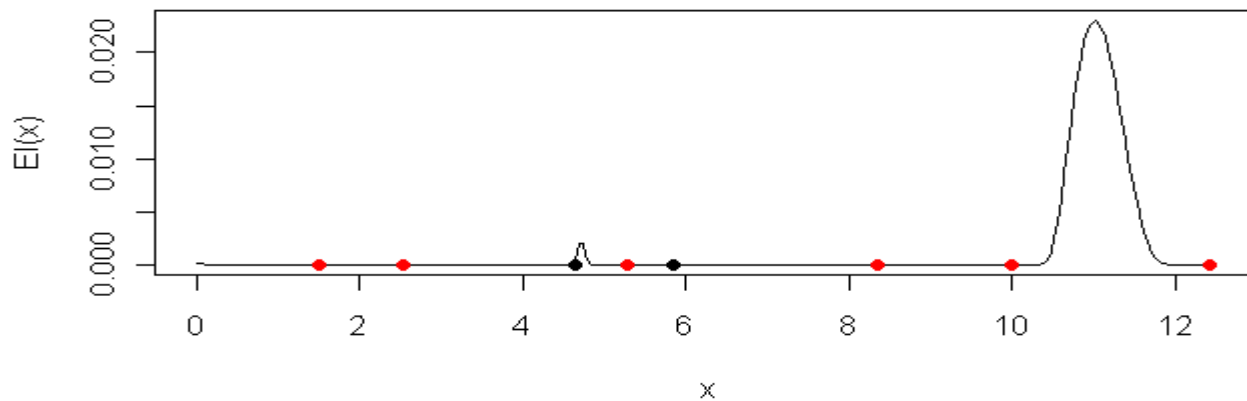
Adaptive design for optimization: EGO algorithm

EGO: step 2

kriging the sinus function



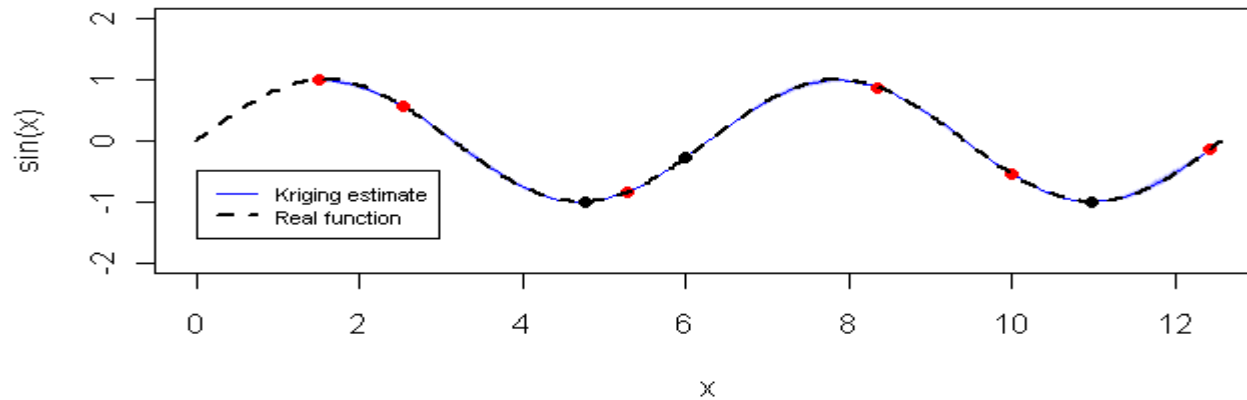
[Chevalier, 2011]



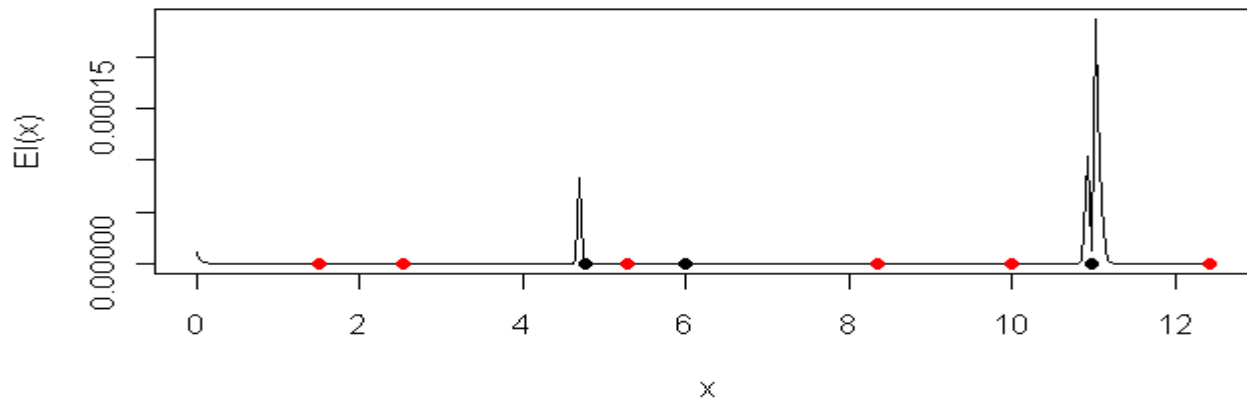
Adaptive design for optimization: EGO algorithm

EGO: step 3

kriging the sinus function



[Chevalier, 2011]





Conclusion

On the Gaussian process metamodel

- Valuable tool when computer code is cpu-time expensive ($N \sim$ hundreds runs)
- GP model construction is possible for moderate dimensional case ($p < 50$)
- Main advantage of GP: probabilistic metamodel which gives confidence bands in addition to a predictor

Full interest in sensitivity analysis

- Fitting quality is dependent of the initial design
GP model is well adapted to sequential and adaptative designs
- Caveats: it can require a large amount of effort during the fitting process and cases with more than 1000 points begin to be difficult (matrix inversion)
- Designs for specific objectives (optimization, quantile, probability, etc.)
- Other hot topics: calibration and validation of computer codes

Bibliography

Cours

- Denis Marcotte (Ecole Polytechnique Montréal)
- Géovariances
- Roman Sueur (EDF R&D)

Books

- Chilès & Delfiner, *Geostatistics*, Wiley, 1999
- Fang, Li & Sudjianto, *Design and modeling for computer experiments*, Chapman, 2006
- Forrester Sobester and Keane (eds), *Engineering Design Via Surrogate Modelling: A Practical Guide*, Wiley, 2008
- Rasmussen, C. E., and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press
- Santner, T. J., Williams, B. J. and Notz, W. I. (2003). *The Design and Analysis of Computer Experiments*. New York: Springer
- Wackernagel, *Multivariate geostatistics*, Springer, 1998

Metamodels

Simpson, Peplinski, Koci & Allen, Metamodels for computer-based engineering design: Survey and recommendations, *Engineering with computers*, 17, 2001

Roustant, Ginsbourger, Deville, DiceKriging, DiceOptim: Two R Packages for the Analysis of Computer Experiments by Kriging-Based Metamodeling and Optimization, *Journal of Stat. Soft.*, 2012

ANNEXES : RAPPELS STATS

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 : $\omega \rightarrow (Z_1(\omega), Z_2(\omega)) \in \mathbb{R}^2$

▪ Fonction de répartition conjointe :

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

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

▪ Densité de probabilité conjointe :

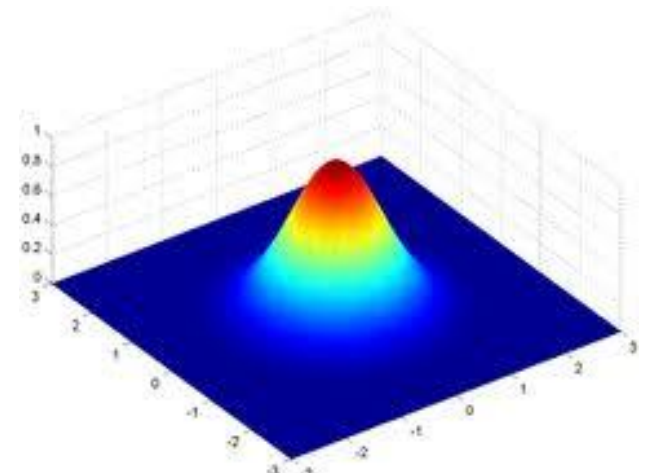
$$F_{Z_1, Z_2}(z_1, z_2) = \int_{-\infty}^{z_2} \int_{-\infty}^{z_1} f_{Z_1, Z_2}(u, v) du dv$$

▪ 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 :

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

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



Rappels statistiques et théorie

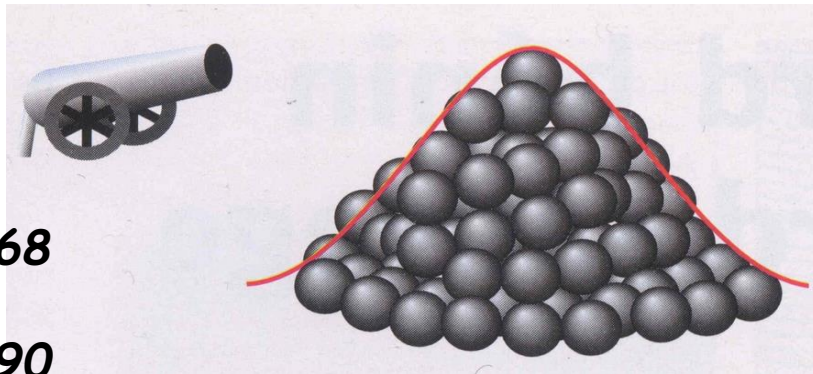
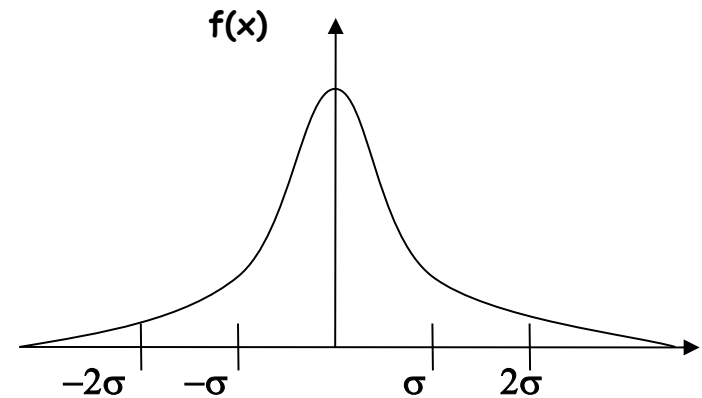
➤ Loi gaussienne ou loi normale $\mathcal{N}(\mu, \sigma^2)$

- **Densité**

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

- **Moments**

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



0.68

0.90

0.95

0.998

$$P(\mu - \sigma < X < \mu + \sigma) =$$

$$P(\mu - 1.64\sigma < X < \mu + 1.64\sigma) =$$

$$P(\mu - 1.96\sigma < X < \mu + 1.96\sigma) =$$

$$P(\mu - 3.09\sigma < X < \mu + 3.09\sigma) =$$

Exemples : impacts des boulets de canon (Touffret 1872)

Rappels statistiques et théorie

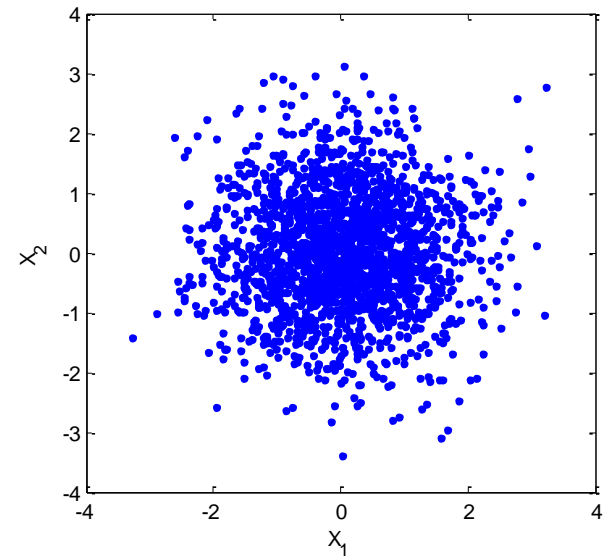
➤ Vecteur gaussien

▪ Loi $\mathcal{N}(0, I_n)$

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



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 :

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

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

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

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



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

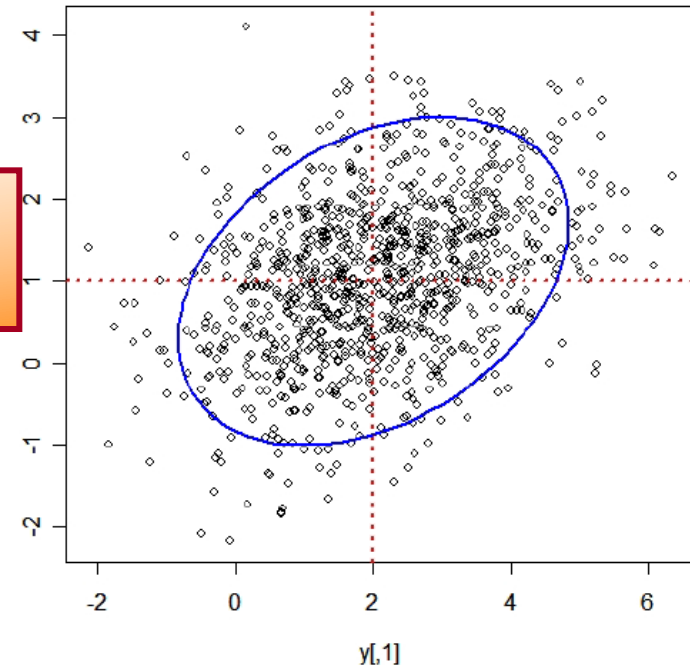
➤ Vecteur gaussien

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

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

Simulation d'un échantillon de 1000 points
avec $\mathbf{m} = (2, 1)^T$; $\sigma_1^2 = 2$; $\sigma_2^2 = 1$ 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 \Leftrightarrow corrélation

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

- Cas non-dégénéré : Σ symétrique définie positive

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



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

Z admet une densité :

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

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 $\mathbf{Z}_2|\mathbf{Z}_1=\mathbf{z}_1$: $f_{Z_2|Z_1=z_1}(z_2) = \frac{f_{Z_2,Z_1}(z_2, z_1)}{f_{Z_1}(z_1)}$

Rappels statistiques et théorie

➤ Vecteur gaussien

▪ Conditionnement :

Extension au cas général $\mathbf{Z}'_1 = (\mathbf{Z}_1, \dots, \mathbf{Z}_p)^\top$ et $\mathbf{Z}'_2 = (\mathbf{Z}_{p+1}, \dots, \mathbf{Z}_n)^\top$

$$(\mathbf{Z}'_1, \mathbf{Z}'_2)^\top \sim N \left(\begin{pmatrix} m'_1 \\ m'_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{1,2} \\ \Sigma_{1,2}^\top & \Sigma_{22} \end{pmatrix} \right)$$

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 $\mathbf{Z}'_2 | \mathbf{Z}'_1 = \mathbf{z}'_1$:

$$\mathbf{Z}'_2 | \mathbf{Z}'_1 = \mathbf{z}'_1 \sim N \left(m'_2 + \Sigma_{1,2}^\top \Sigma_{11}^{-1} (\mathbf{z}'_1 - m'_1), \Sigma_{22} - \Sigma_{1,2}^\top \Sigma_{11}^{-1} \Sigma_{1,2} \right)$$

Rappels statistiques et théorie

➤ Processus Gaussiens (PG)

- **Définition** : un processus gaussien est un processus aléatoire réel $\{Z(x)\}_E$ $D \subset \mathbb{R}^d$ 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

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

$$\text{où } m(x) = E[Z(x)]$$

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

- **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 $\sigma^2 = \text{Var}[Z(x)]$