Retour sur l'école chercheurs MEXICO « Analyse de sensibilité globale, métamodélisation et optimisation de modèles complexes »

Sébastien Guyader

25 mai 2018

- Réseau méthodologique sur les « Méthodes d'EXploration Informatique de modèles COmplexes »
- Animé par des chercheurs du département MIA de l'INRA
- Autres organismes participants : IFREMER, IRSTEA, CIRAD, Université du Littoral Côte d'Opale...
- Objectifs :
  - ouvrir les biologistes-modélisateurs au traitement stat. des simulations et à l'exploration raisonnée des modèles
  - initier de nouveaux fronts de recherches en statistique
  - contribuer à la réflexion méthodologique en modélisation
  - rendre ces méthodes accessibles au modélisateur

Les actions du réseau sont organisées autour de 4 axes :

- Journées thématiques : journées plus ou moins ouvertes, sur un thème méthodologique spécifique
- Rencontres Mexico : journées destinées aux retours d'expérience des utilisateurs des méthodes d'analyse et d'exploration de modèles
- Écoles-chercheurs : session de formation sur l'exploration et l'analyse de sensibilité de modèles
- Développements informatiques : développement d'une boîte à outils mettant ces méthodes à disposition d'utilisateurs de R ou de gestionnaires de plateformes de modélisation est en cours.

L'école chercheurs 2018 a eu lieu à La Rochelle, du 26 au 30 mars 2018
Au sommaire des réjouissances : exposés, cours, et TP portant sur :

1. La conception de plans d'expérience
2. L'analyse de sensibilité globale et d'incertitude
3. La calibration et l'optimisation de modèles
    1. Modélisation statistique
    2. Optimisation basée sur modèles (métamodélisation)
    3. Optimisation sans métamodèle
    4. Inférence/calibration bayésienne

Analyse d'incertitude

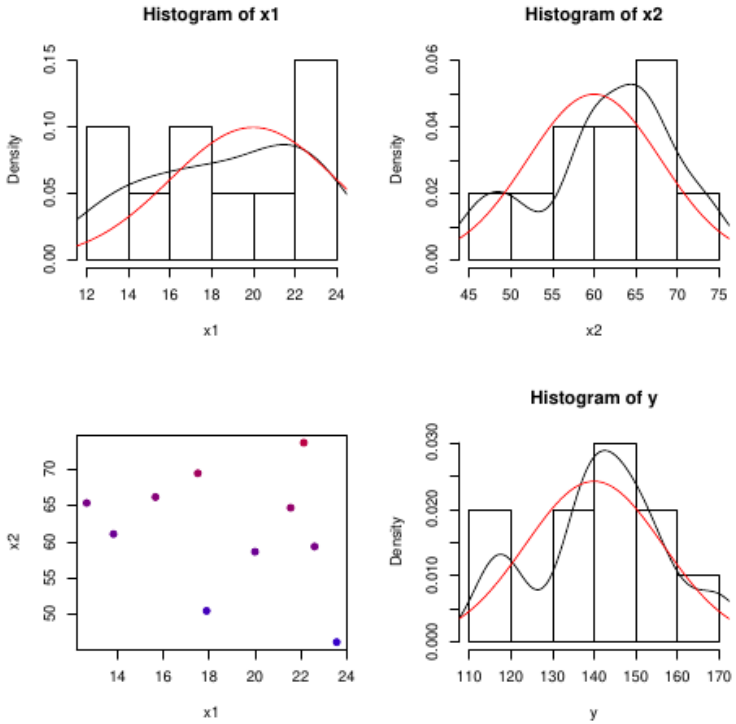
Question

Quelle incertitude sur fonction $\mathcal{G}(x)$, et ses sources ?

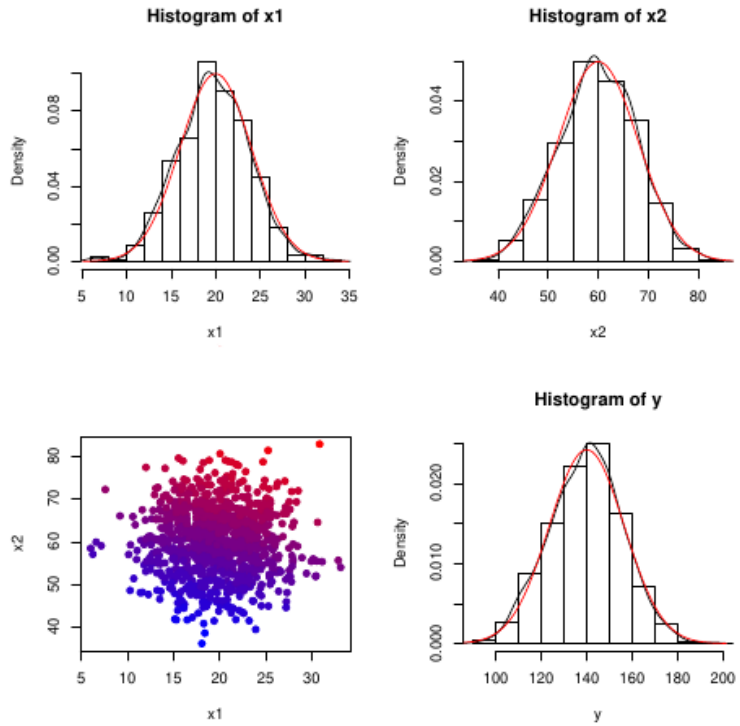
- si on connait la distribution des paramètres $x_1, ..., x_k$, et si modèle simple on peut déduire la distribution de $\mathcal{G}(x)$
- mais si fonction complexe, ou si la distribution des paramètres est inconnue, on échantillonne N combinaisons des paramètres (Monte Carlo, Hypercubes latins, ...)
- on fait appel N fois au modèle et on décrit la variation de $\mathcal{G}(x)$
- la précision dépend uniquement de N, et non de la dimension K du modèle $\mathcal{G}$

Analyse d'incertitude

N=10



N=1000



Analyse de sensibilité

Question

Quelles sont les principales sources de variation qui influencent $\mathcal{G}(x)$?
Quels paramètres sont influents ?

- AS locale : variation de $\mathcal{G}(x)$ autour de la valeur x_0 (basé sur calcul de dérivées)
- AS globale : variation de $\mathcal{G}(x)$ quand x varie dans son domaine d'incertitude entre x_{min} et x_{max} (basée sur calcul d'indices)
- on définit des indices de sensibilité, puis on les calcule en faisant varier les paramètres sur leurs domaines

Analyse de sensibilité

4 étapes :

1. définir les distributions des paramètres (on suppose qu'ils sont indépendants)
2. générer des échantillons à partir de ces distributions
3. calculer $\mathcal{G}(x)$ pour chaque série de paramètres générée
4. calculer les indices de sensibilité

Analyse de sensibilité

2 familles de méthodes basées sur :

1. criblage, discrétisation de l'espace des paramètres :
   - plan factoriel/anova $\rightarrow$ indices de sensibilité = contributions principales et totales des facteurs, $\frac{SCE(effetprincipal+interaction)}{SCT}$
   - méthode de Morris
2. espace des paramètres continu :
   - indices basés sur la régression
   - indices de Sobol' estimés par échantillonnage intensif
   - Pick & Freeze, et extended FAST

Analyse de sensibilité

Criblage (1)

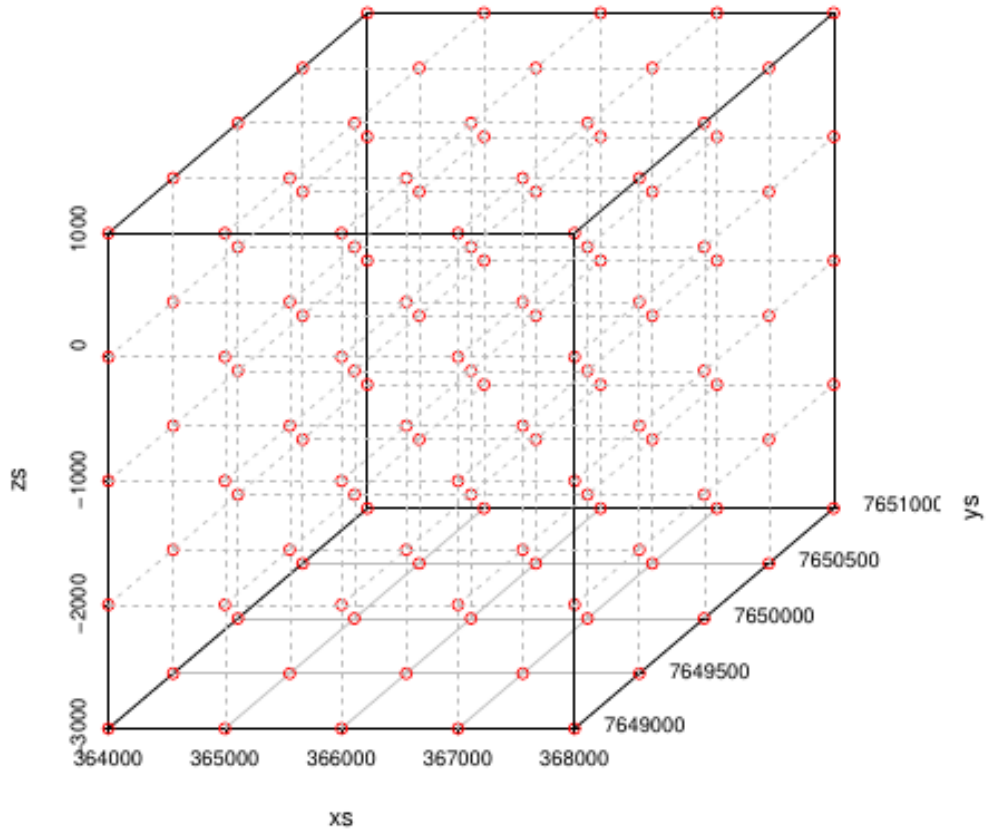
L'ANOVA nécessite un plan factoriel pour décompostion totale de la variance

- un plan factoriel complet est très grand (chaque facteur est décomposé en niveaux, et tous les niveaux sont croisés)
- ex : un modèle à 20 paramètres incertains, avec 3 niveaux il faut $3^{20} = 3486784401$ évaluations du modèle
- si 1 évaluation prend 0.01 seconde, il faut 24.2 jours (24.2 * 200 Watts = 116 226 kWh, coût d'environ 11.6 k€ à 10cts / kWh) = consommation annuelle de 16 foyers français
- les plans fractionnaires sont utiles

Analyse de sensibilité

Criblage (1)

Plan factoriel : 3 facteurs (paramètres), 5 niveaux



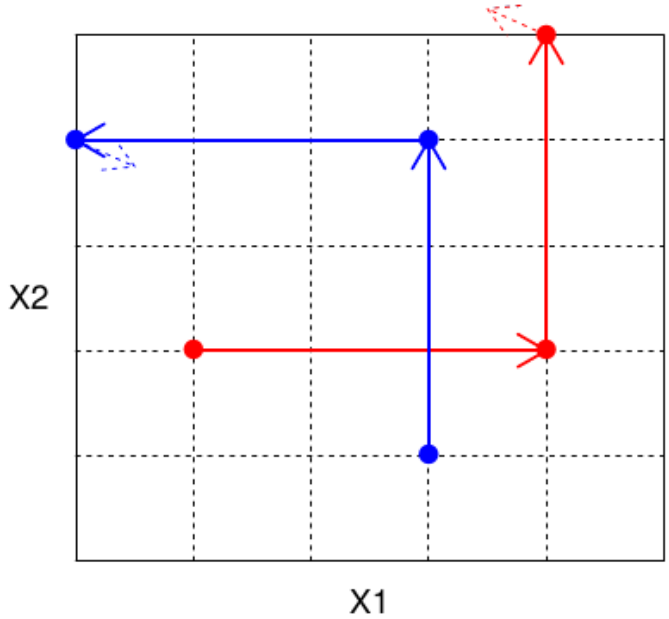
Analyse de sensibilité

Criblage (2)

- Morris : exploration "astucienne" de l'espace des paramètres
  1. discrétiser l'espace → point de départ aléatoire → déplacement sur 1 seul facteur à la fois (OAT) ⇒ 1 trajectoire K déplacements ($K+1$ points)
  2. mesurer l'importance du facteur x_i sur chaque trajectoire n évaluant le modèle en fonction de la taille du saut
     - effet élémentaire (effet d'un saut rapporté à la taille du saut)
     - effet moyen (somme effets des trajectoires divisé par le nb de trajectoires)
     - effet absolu moyen (prend la valeur absolue des effets élémentaires)
     - écart-type des effets élémentaires

Analyse de sensibilité

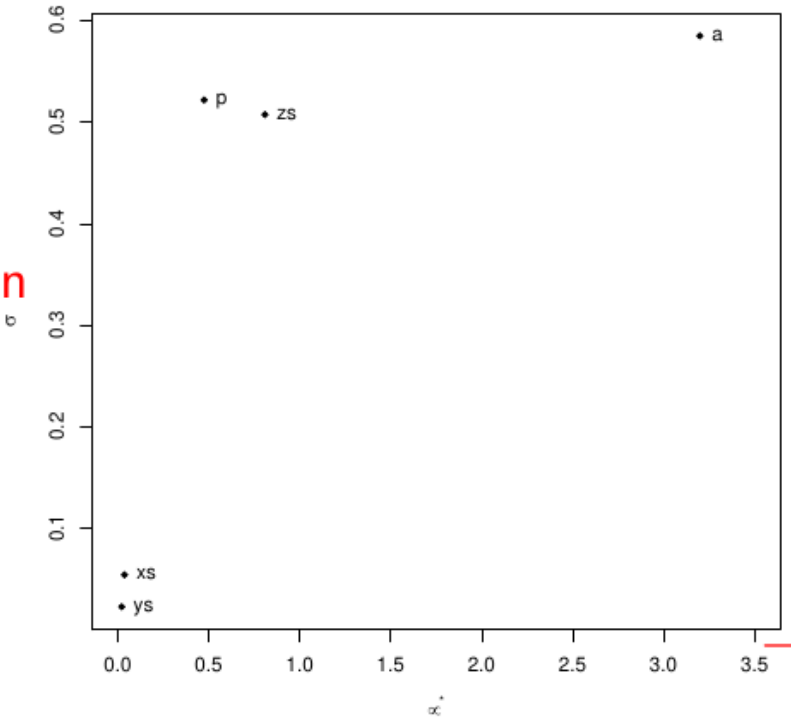
Criblage (2)



Analyse de sensibilité

Criblage (2)

Effet non linéaire et/ou en interaction



Effet principal linéaire

Analyse de sensibilité

Espace continu des paramètres (1)

- tirer les entrées X sur leur intervalle (hypercube latin, Monte Carlo, ou suite à faible discrépance)
- 2 méthodes :
  1. indices basés sur la régression : coefficient de corrélation linéaire ("src") ou partielle ("pcc", plus approprié s'il y a corrélation entre les facteurs)
  2. décomposition d'Hoeffding - Sobol' : décomposition analogue à celle de la variance, indices de sensibilité = effets principal des paramètres et de leur interaction
- problème : ces méthodes nécessitent de nombreux appels au code
- il y a des méthodes alternatives pour estimer les indices :
  - méthode Sobol' - Saltelli ou Pick & Freeze
  - méthode FAST (Fourier Amplitude Sensitivity Test)
  - passage par un métamodèle

Analyse de sensibilité

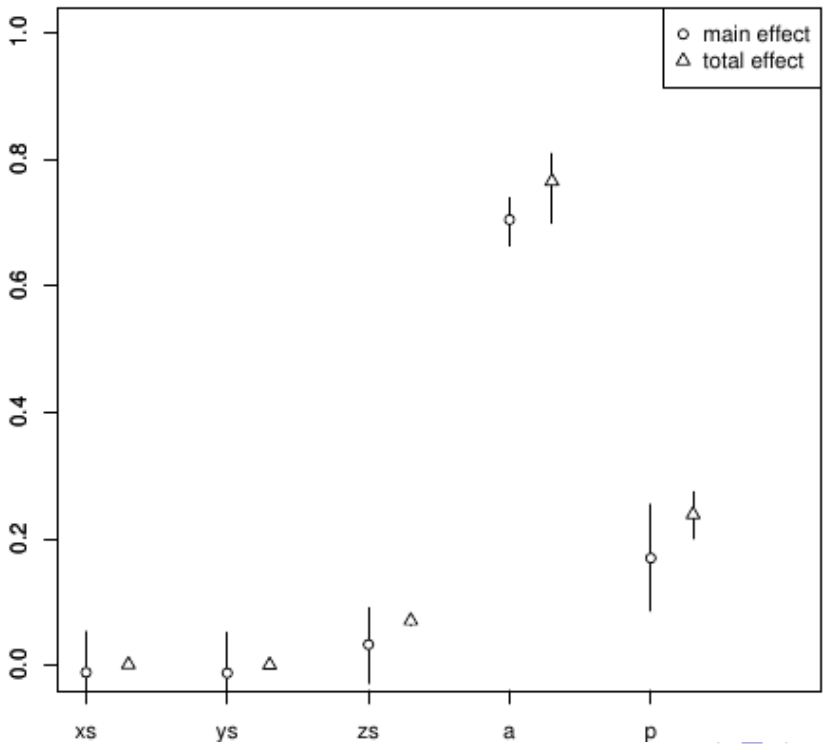
Espace continu des paramètres (2)

- méthode Sobol' ou Pick & Freeze
  - 2 échantillonnages A et B (hypercube latin, Monte Carlo, ou suite à faible discrépance), taille $N * K$
  - évaluations sur $N * (K + 2)$ combinaisons C issues de A et B
  - estimation des indices
  - évaluation de la précision par répétition ou bootstrap
- méthode FAST
  - échantillonnage par trajectoire déterministe remplissant l'espace ("space-filling path")
  - choix raisonné du jeu de fréquences ω_i (harmoniques distinctes entre facteurs jusqu'à l'ordre M (4 à 6)
  - estimation par analyse fréquentielle
  - eFAST (FAST étendu) : 1 trajectoire FAST par paramètre x_i

Analyse de sensibilité

Criblage (2)

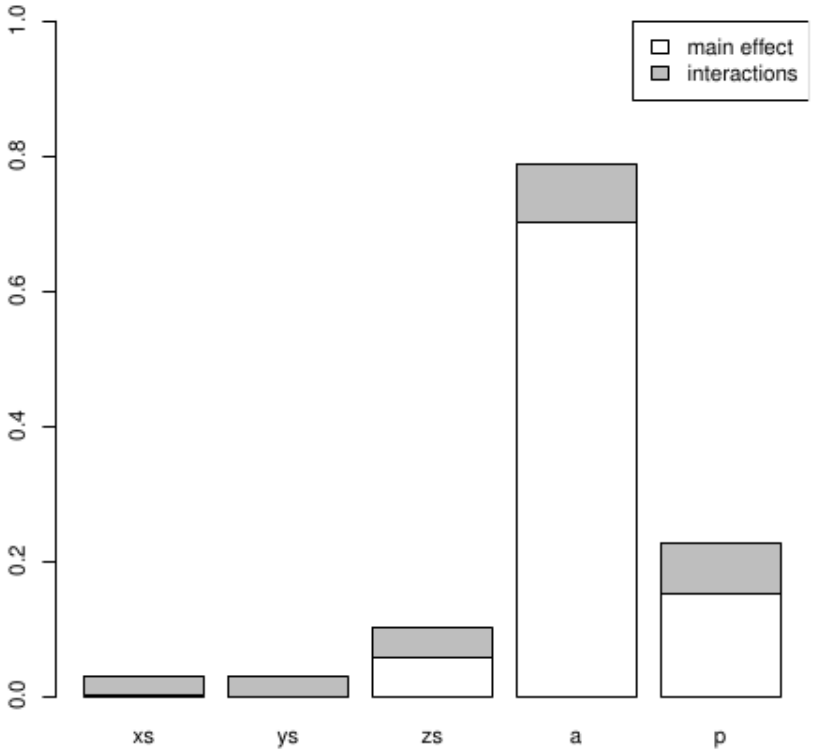
Sobol



Analyse de sensibilité

Criblage (2)

Fast

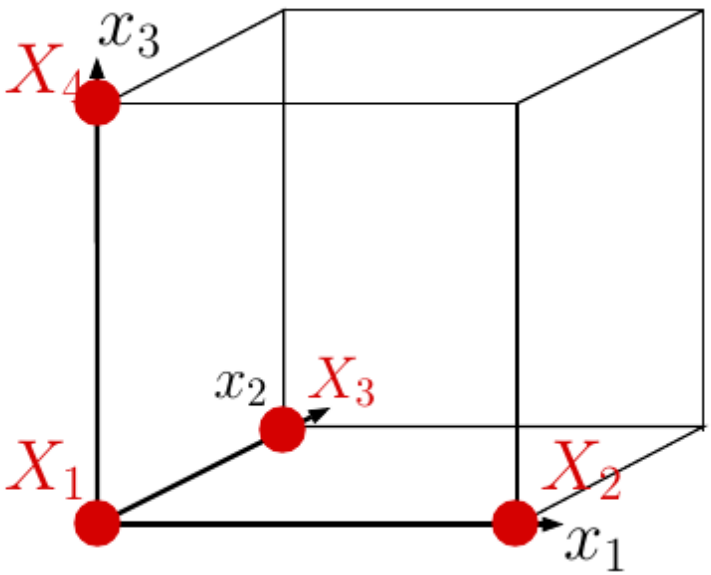


- Souvent, on approche le vrai modèle à l'aide d'une fonction mathématique
- Il faut évaluer la fonction de nombreuses fois, question : pour un budget de n appels à la fonction, quels points de l'espace des paramètres doit-on évaluer ?
- Il faut une stratégie pour choisir les points à évaluer :
  - "One at a Time" (OAT)
  - Pans factoriels
  - Stratégies de remplissage de l'espace

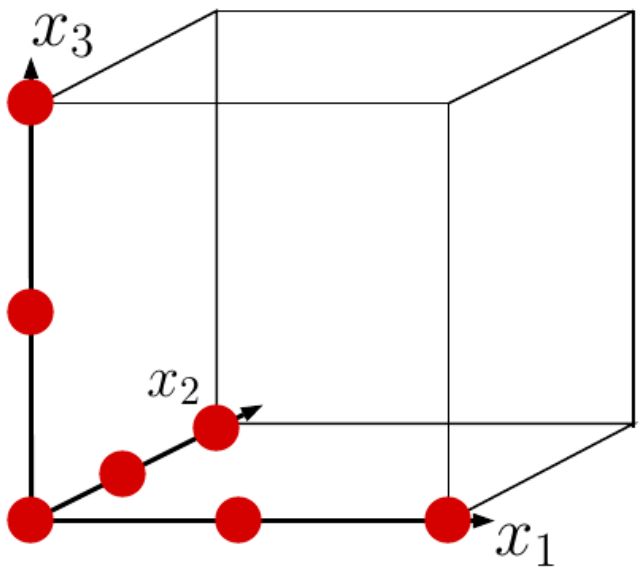
OAT

- Toutes les variables sont fixées, sauf 1 qui varie pour évaluer son influence sur G(x)
- Il faut plus de 2 niveaux pour estimer les effets quadratiques

2 niveaux



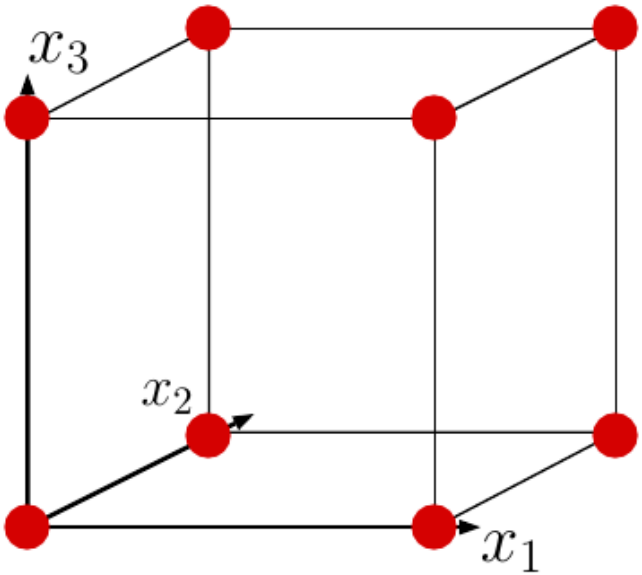
3 niveaux



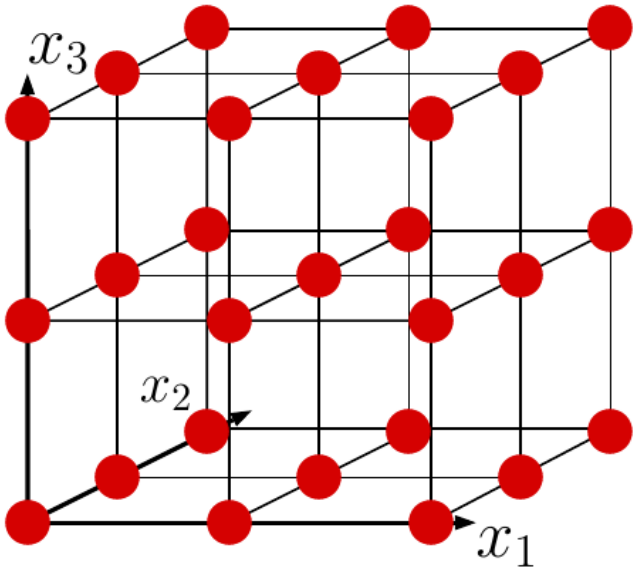
Plans factoriels

- Toutes les combinaisons de paramètres sont testées
- Le plus simple à 2 niveaux (min et max), on peut tester plus de niveaux
- Problème : le temps de calcul explose avec le nombre de paramètres et de niveaux

2 niveaux



3 niveaux

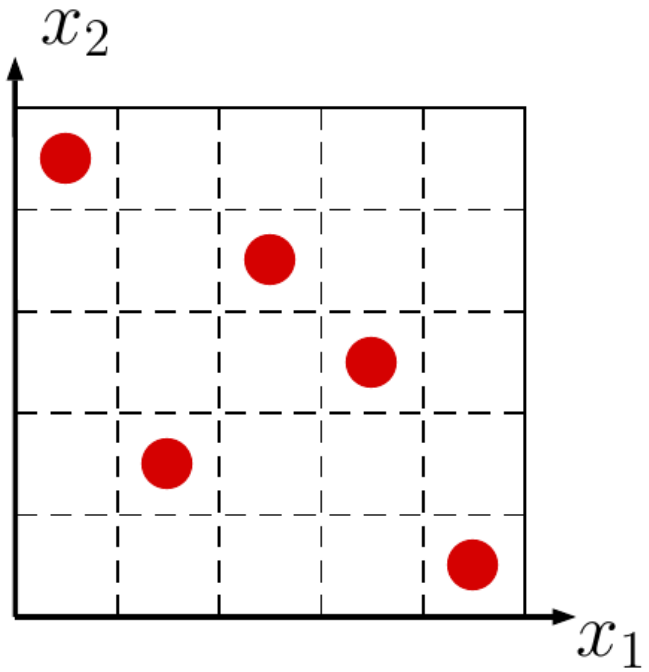


Stratégies de remplissage de l'espace

- Quels critères pour évaluer le remplissage de l'espace :
  - "maximin" (maximiser la distance minimale entre 2 points) et "minimax" (minimiser la distance maximale entre 2 points)
  - "discrépance" : compare le nb de points dans un hyper-rectangle au nb de points attendus par distrib uniforme
- Il y a 3 principales stratégies de remplissage :
  - Hypercubes latins
  - Séquences à faible discrépance
  - Tesselation centroïde de Voronoi

Hypercubes latins (lhs)

- découpage du domaine en n^d blocs avec seulement 1 point par "ligne" et "colonne"
- à combiner avec critère tel que "maximin"



- optimisation d'un LHS par méthode de "Morris & Mitchell"

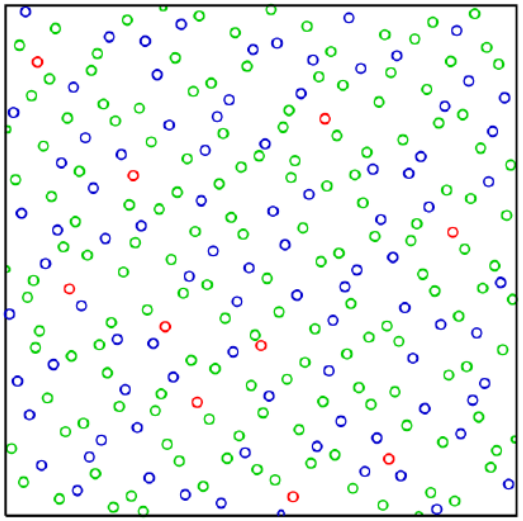
Séries à faible discrépance

- séquences déterministes qui convergent vers une distribution uniforme

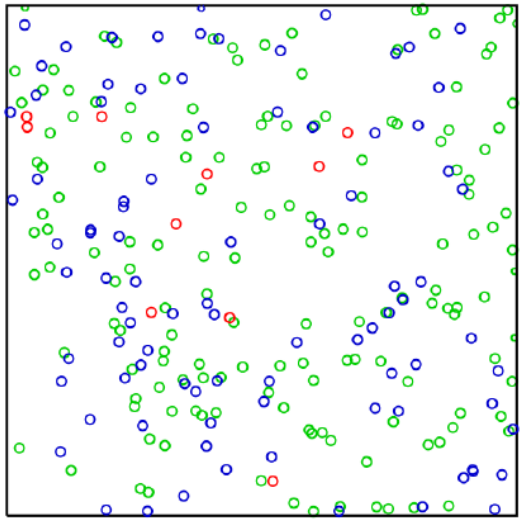
$$x_1 = 1/2,\ 1/4,\ 3/4,\ 1/8,\ 5/8,\ 3/8,\ 7/8,\ 1/16,\ 9/16, \ldots$$
$$x_2 = 1/3,\ 2/3,\ 1/9,\ 4/9,\ 7/9,\ 2/9,\ 5/9,\ 8/9,\ 1/27, \ldots$$

- couvrent l'espace rapidement et de manière homogène

Halton Sequence

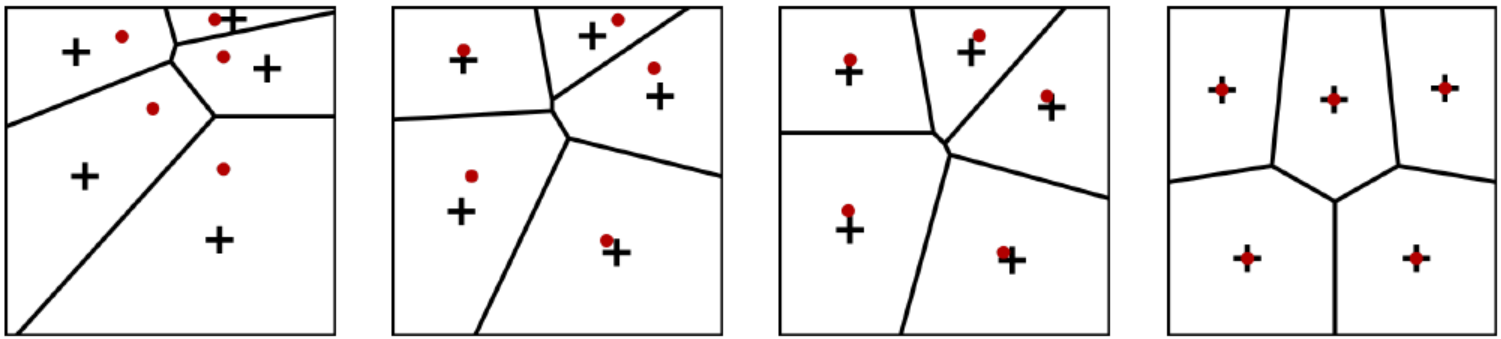


uniform pseudo random



Tesselation centroïde de Voronoi

- partition de l'espace en cellules à partir d'un jeu de points, telle que chaque point d'une cellule est plus proche de son "germe" que de tout autre point de l'espace ("zone d'influence")
- algorithme de Lloyd pour la tesselation centroïde



iteration 1 | iteration 2 | iteration 3 | iteration 15

- autre algorithme : k-means

Calibration

Processus d'ajustement des paramètres d'un modèle en intégrant l'incertitude des paramètres et/ou du modèle, pour obtenir une représentation du système modélisé qui satisfasse un critère prédéfini

Pour des modèles simples, la résolution analytique du problème est possible, mais pas pour des modèles complexes (nombreux paramètres, nombreuses sorties, processus mal connus)

Calibrer pour :

- estimer les paramètres pas ou difficilement mesurables
- comprendre le fonctionnement du système étudié
- crédibiliser/améliorer le modèle pour l'utiliser en décision, prédiction...

Objectif de la calibration :
trouver les valeurs optimales des paramètres qui minimisent la différence entre Y_{obs} et Y_{sim}

C'est donc l'optimisation d'une fonction mathématique (fonction d'objectif), avec 3 cas :

- pas d'hypothèse de distrib. sur Y ni sur $X \rightarrow$ analyse numérique ou statistique inférentielle
- hypothèse de distrib. sur Y, mais pas sur $X \rightarrow$ statistique inférentielle
- hypothèses de distrib sur Y et sur $X \rightarrow$ statistique bayésienne

Commencer par construire une première fonction d'objectif :

- les plus communes : moindres carrés, vraisemblance
- plus spécifiques : ABC (Approximate Bayesian Computation)
- optimisation multi-objectifs : pondérations, fronts de Pareto

Évaluer la qualité de l'optimisation :

- convergence, global/local (est-on dans tombé un creux ?)
- identifiabilité des paramètres (est-ce que plusieurs solutions donnent un même optimum ?)

A chaque itération, calcul d'un ensemble de solutions et des valeurs de la fonction d'objectif associée : trace de l'algorithme dans l'espace Y et l'espace X

Modélisation statistique

Les modèles statistiques sont utilisés pour :

- interpoler ou approximer des fonctions
- calculer des quantités d'intérêt (valeur moyenne, optimum, minimum...)
- mesurer l'erreur

Il y 2 principaux types de modèles statistiques :

- régression linéaire
- régression de processus gaussiens

Processus gaussiens

- Basé sur la distribution Normale de X (jeu de points d'observation)
- Propriété : une combinaison linéaire de variables aléatoires indépendantes suit toujours une loi Normale
- la distribution Gaussienne multivariée peut être généralisée par un processus aléatoire, caractérisé par ses fonctions de moyenne et de covariance (ou noyau)
- on peut générer des vecteurs aléatoires à l'aide fonctions de covariance connues (constante, bruit blanc, exponentiel, Matern 3/2 et 5/2, gaussien...)

Processus gaussiens

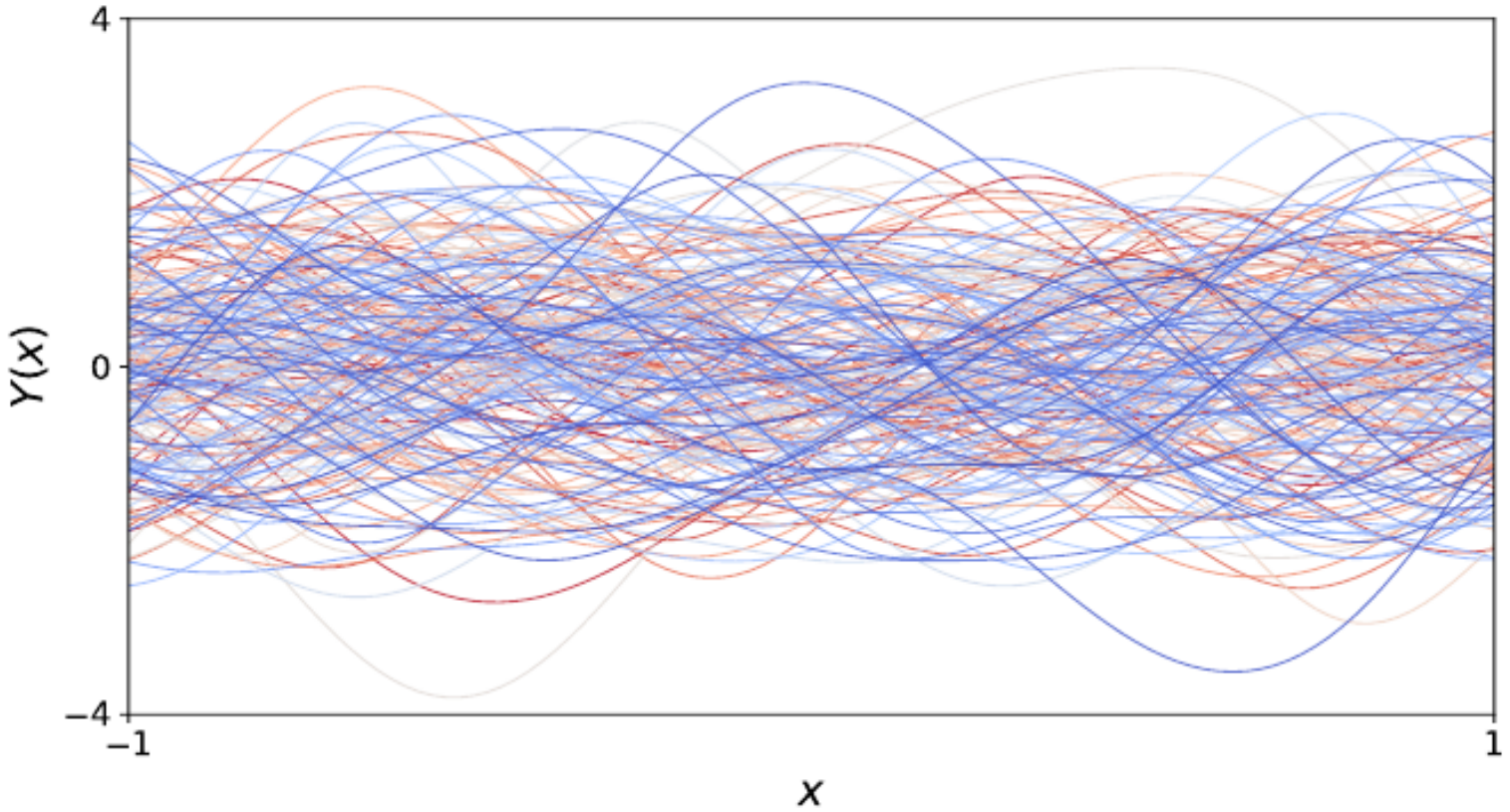
Régression sur processus gaussiens :

- on part de l'observation de la fonction $Y = f(X)$ sur un jeu de points X
- on ne connait pas $f(X)$, mais on suppose qu'elle suit le chemin d'un processus Gaussien $Z \sim \mathcal{N}(0, k)$
- on calcule analytiquement la distribution *a posteriori* $Y(\cdot)|Y(X) = Y$: échantillon de chemins qui passent par les points d'observation

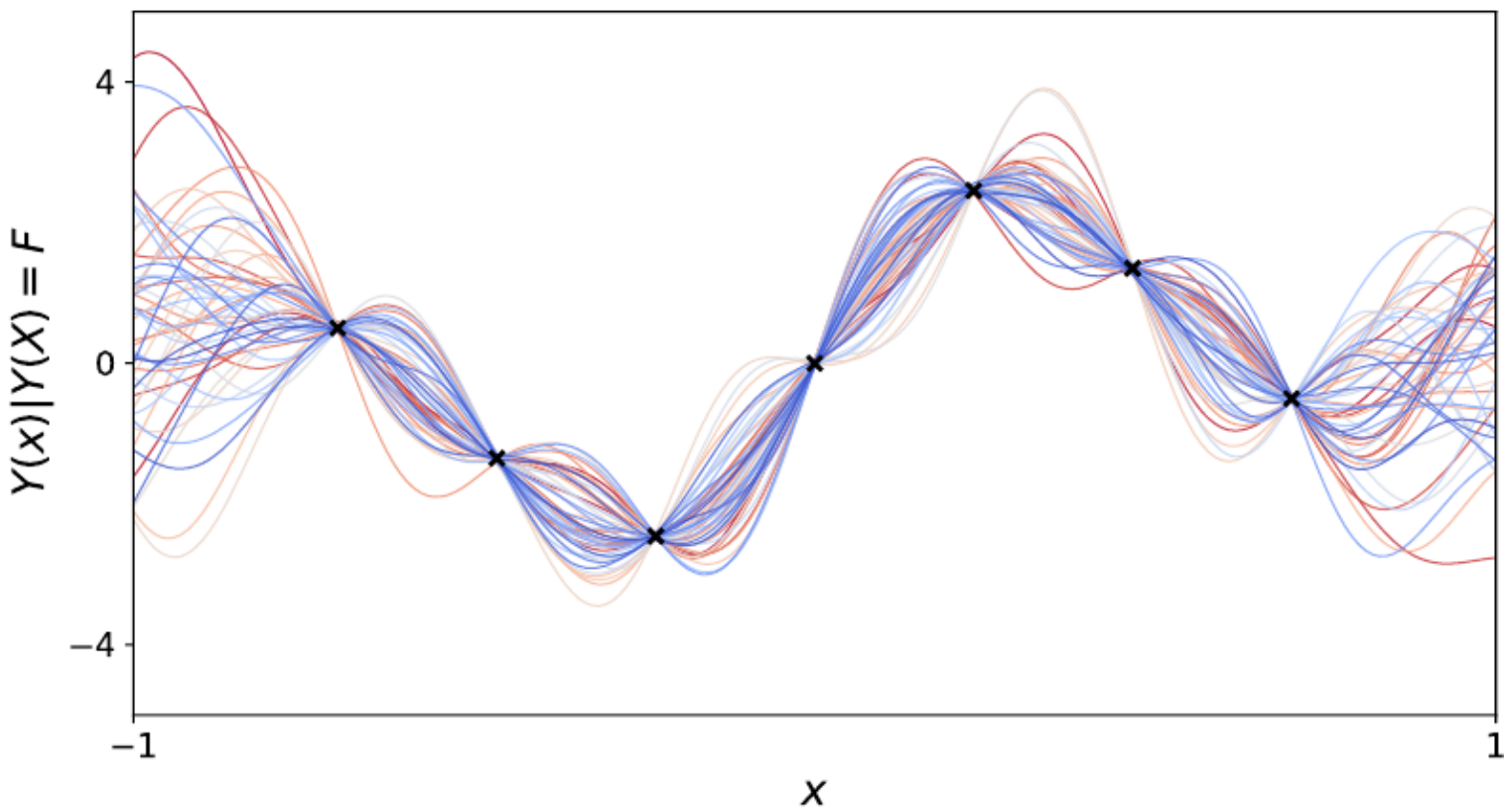
Propriétés :

- permet d'interpoler entre les points
- il faut choisir le noyau en fonction des *a priori* sur la fonction à étudier
- la variance de prédiction dépend uniquement du noyau choisi
- la prédiction de la moyenne ne dépend pas du paramètre de variance
- on peut aussi ajouter du bruit (incertitude) sur les observations

Processus gaussiens



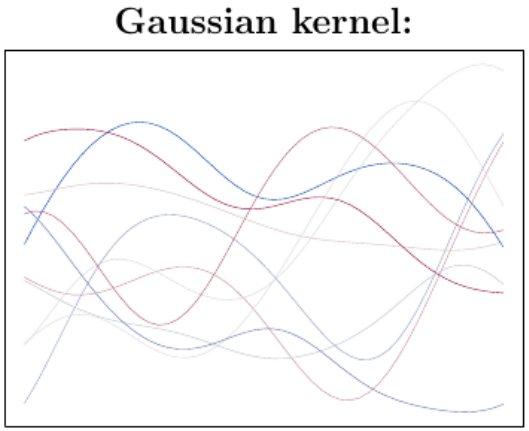
Processus gaussiens

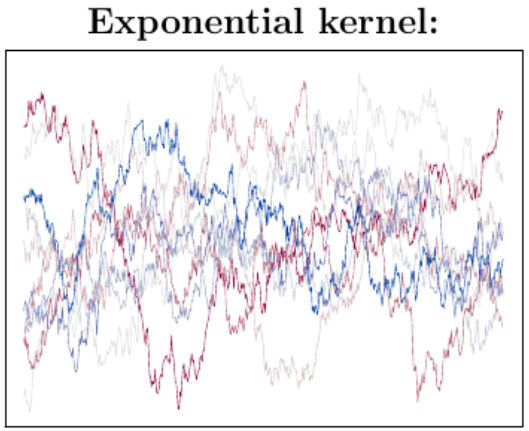


Processus gaussiens

La régression sur processus gaussiens en pratique :

1. Créer un plan d'expérience (quel budget global pour l'évaluation ?)
2. Choisir un noyau





3. Estimer les paramètres (max. vraisemblance, validation croisée, multiples départs)
4. Valider le modèle (jeu de test, Leave-one-out pour moyenne et intervalles de confiance, leave-k-out pour covariances)

Souvent, on fait plusieurs itérations des étapes 2, 3 et 4

Optimisation basée sur métamodèles

L'optimisation fait de nombreux appels au code, il peut être utile d'avoir recours à un métamodèle

On espère qu'à la fin, l'optimum trouvé par le métamodèle soit proche de celui du modèle réel

Compromis à trouver entre exploration/intensification

Optimisation globale :

- une phase d'exploration (recherche partout dans l'espace pour ne pas rater la zone optimale)
- une phase d'intensification une fois la zone identifiée (on recherche le minimum local)

Remarque : il existe une méthode d'optimisation globale sans métamodèle : DIRECT (DIviding RECTangles)

Optimisation par krigeage

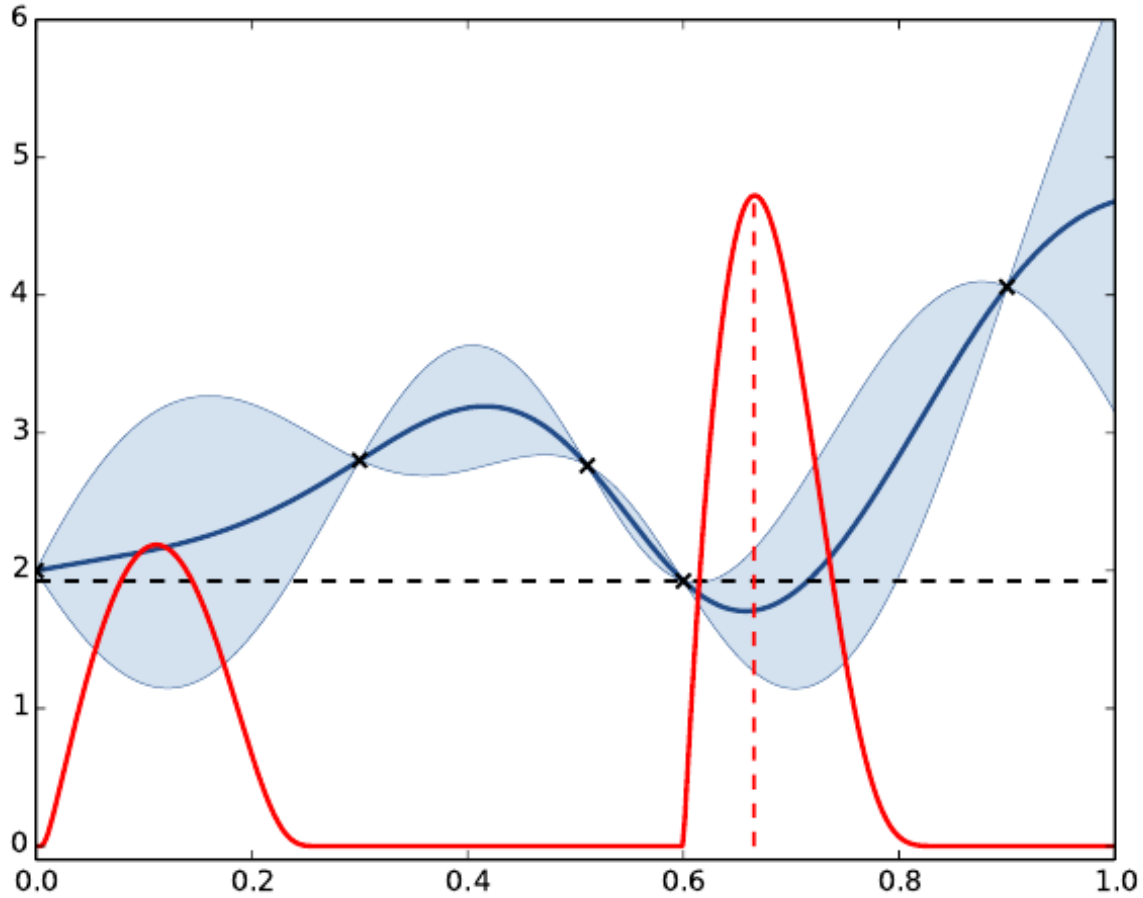
Algorithme "Efficient Global Optimization" (EGO) :

- on cherche un bon compromis entre l'exploitation des bons résultats des itérations précédentes, et l'exploration pour les itérations à venir
- on commence par établir un plan d'expérience
- on calcule la fonction d'objectif F sur ces points
- on élabore un métamodèle, par exemple un modèle de RPG
- en se basant sur les intervalles de confiance et la valeur optimale actuelle de la fonction d'objectif, on calcule la valeur d' "Expected Improvement" (quelles zones sont suceptibles d'héberger l'optimum)
- on ajoute un point là où l'EI est plus élevé
- on réitère
- on arrête l'algorithme quand on a atteint un seuil défini d'itérations
- si on n'a pas trouvé d'optimum, on recommence avec un nouveau modèle RPG

Remarque : il existe une méthode d'optimisation globale sans métamodèle : DIRECT (DIviding RECTangles)

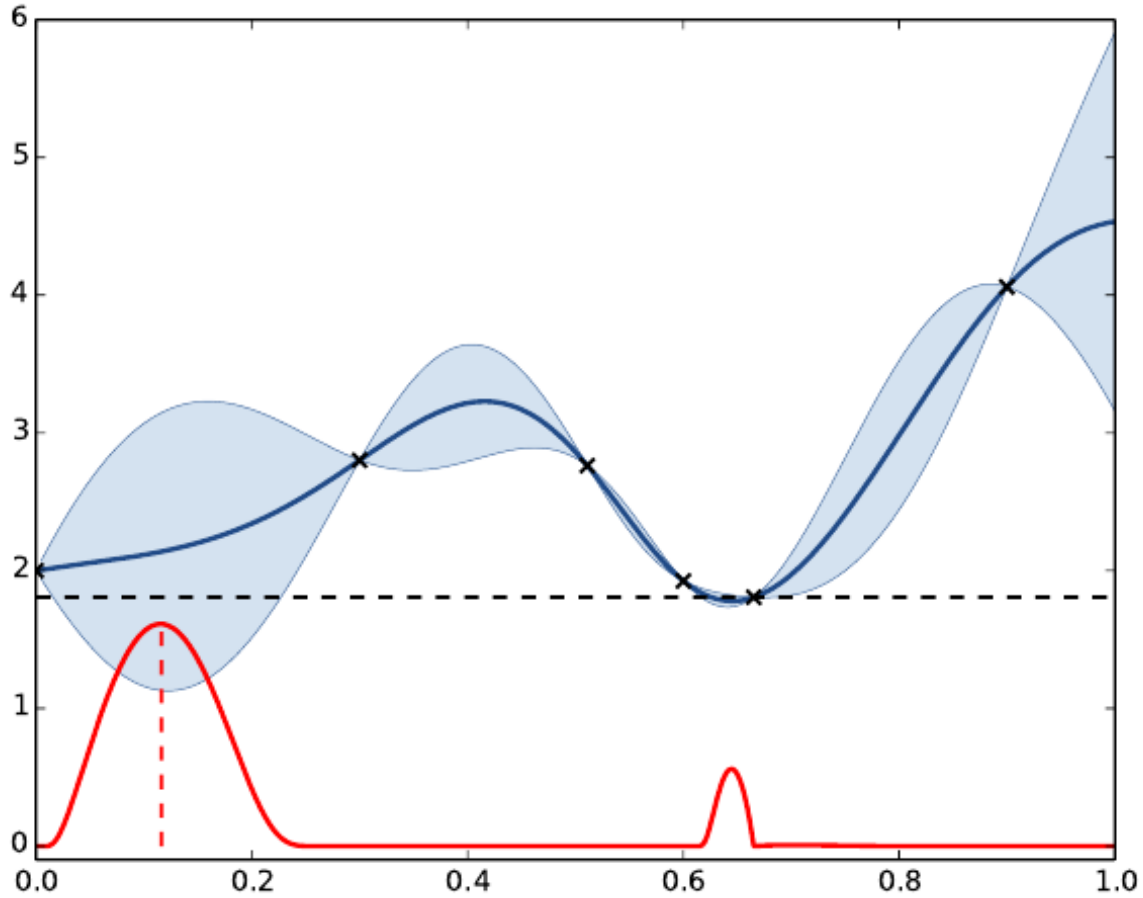
Optimisation par krigeage

Iteration 1



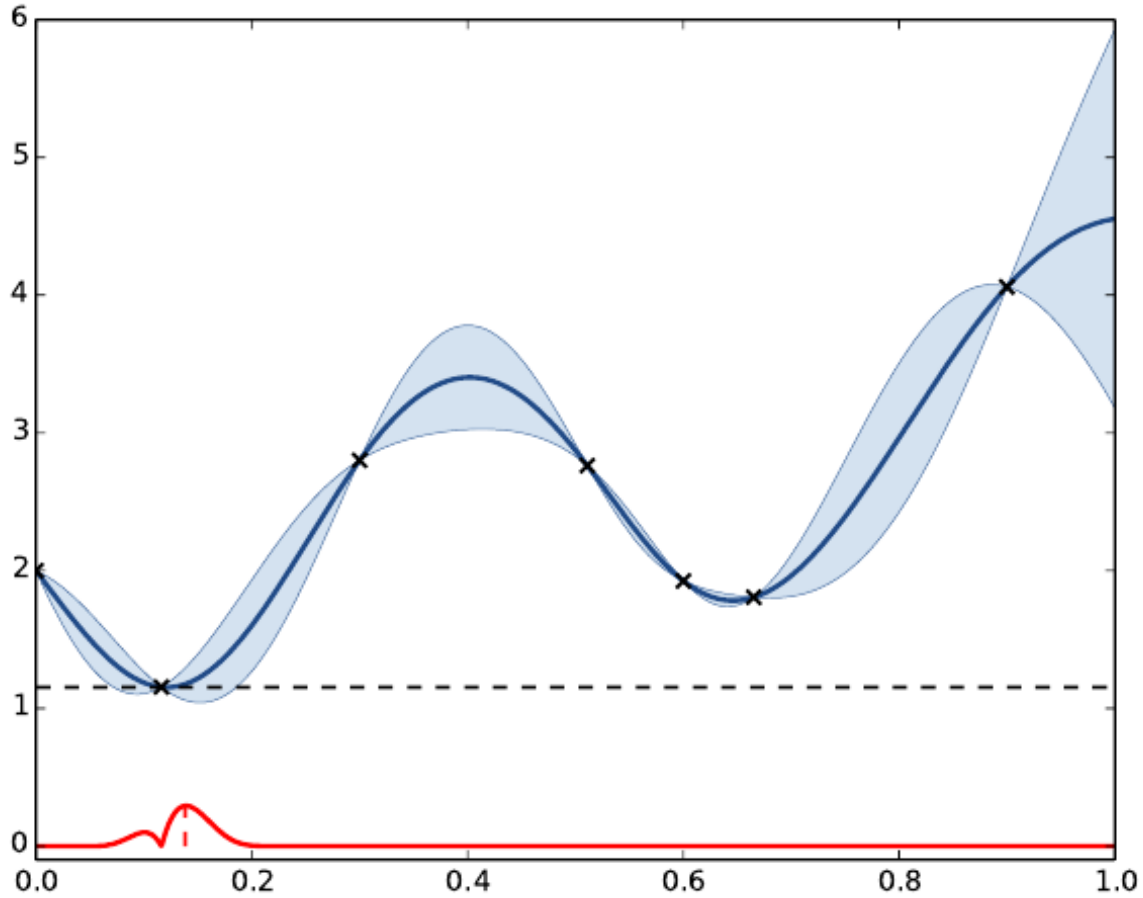
Optimisation par krigeage

Iteration 2



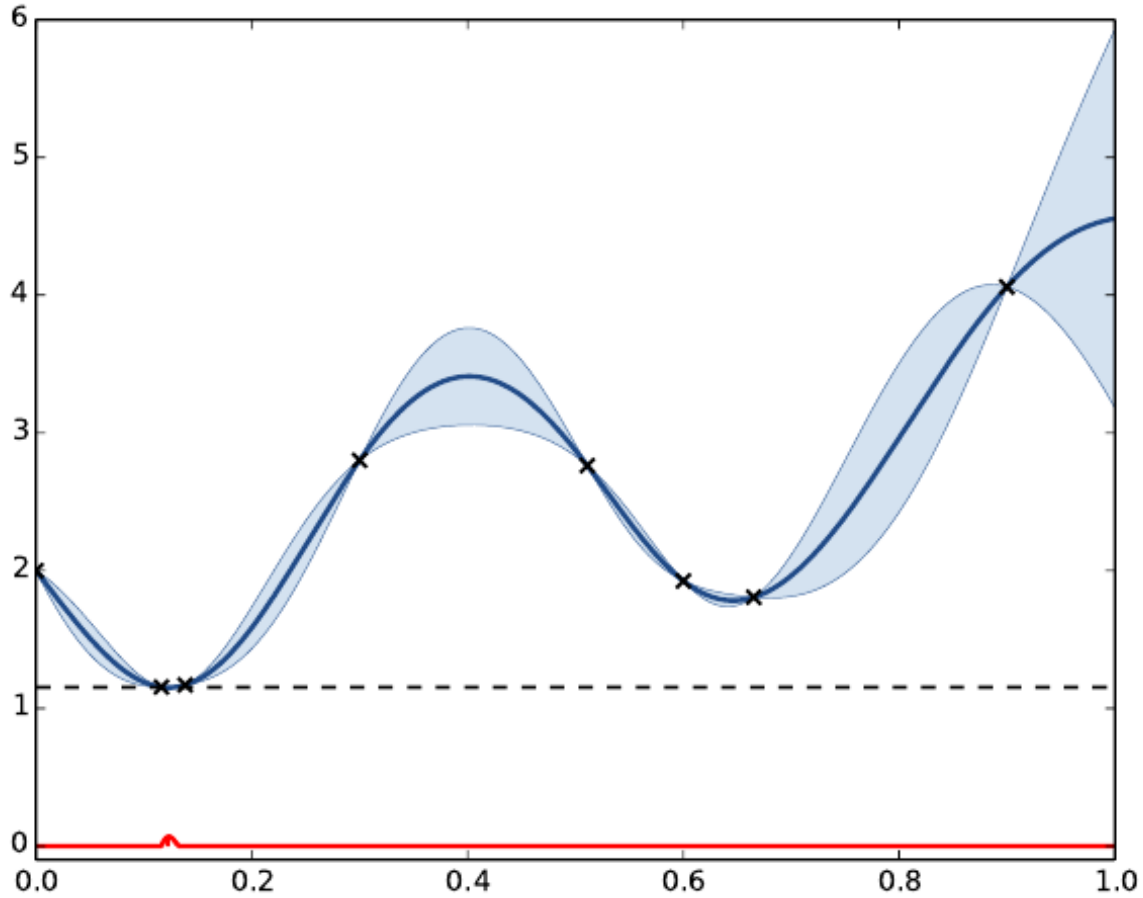
Optimisation par krigeage

Iteration 3



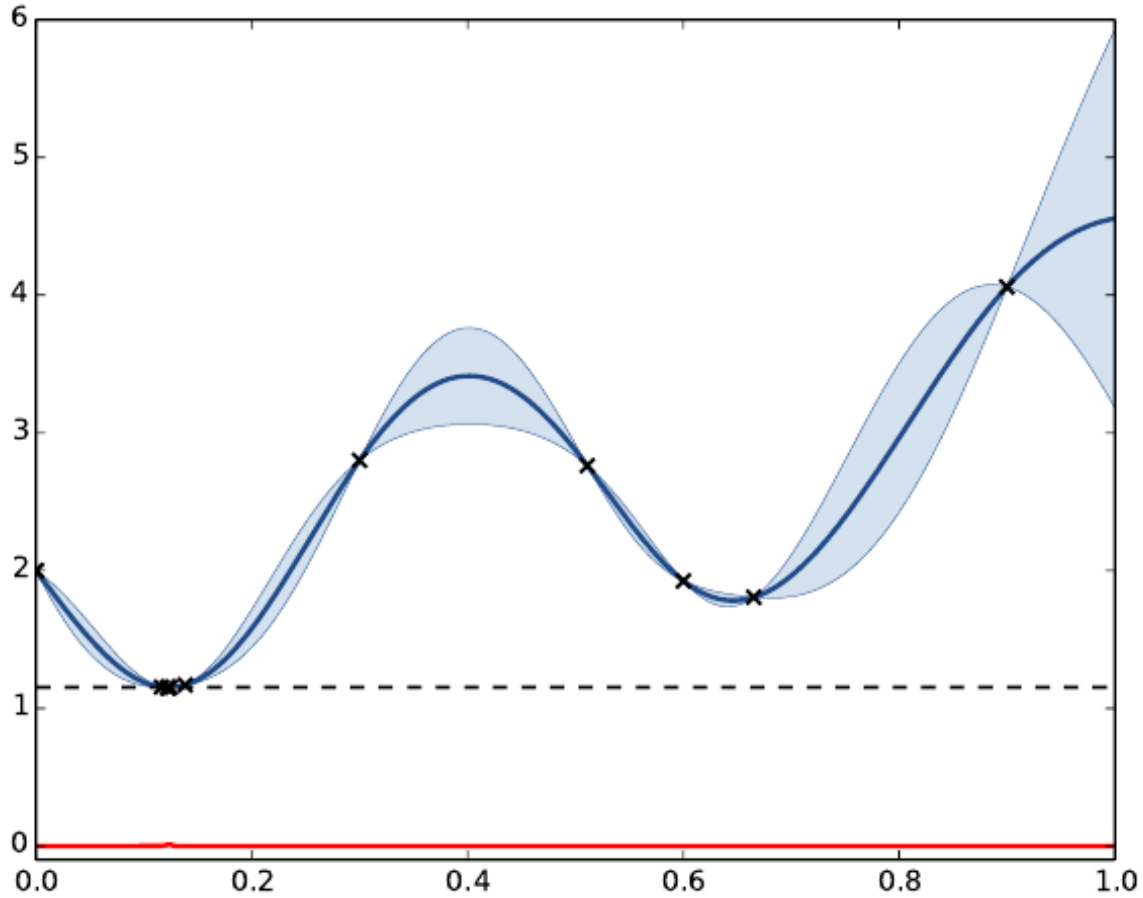
Optimisation par krigeage

Iteration 4



Optimisation par krigeage

Iteration 5



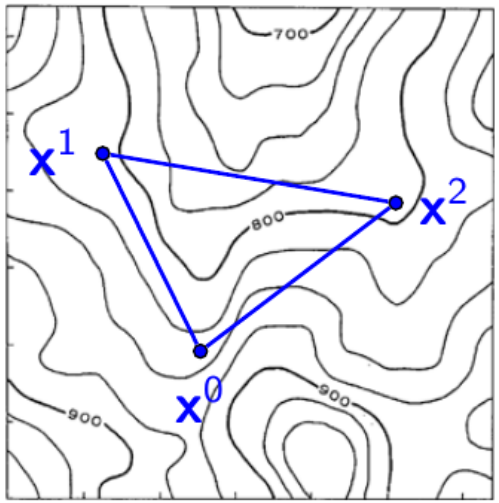
Optimisation sans métamodèle

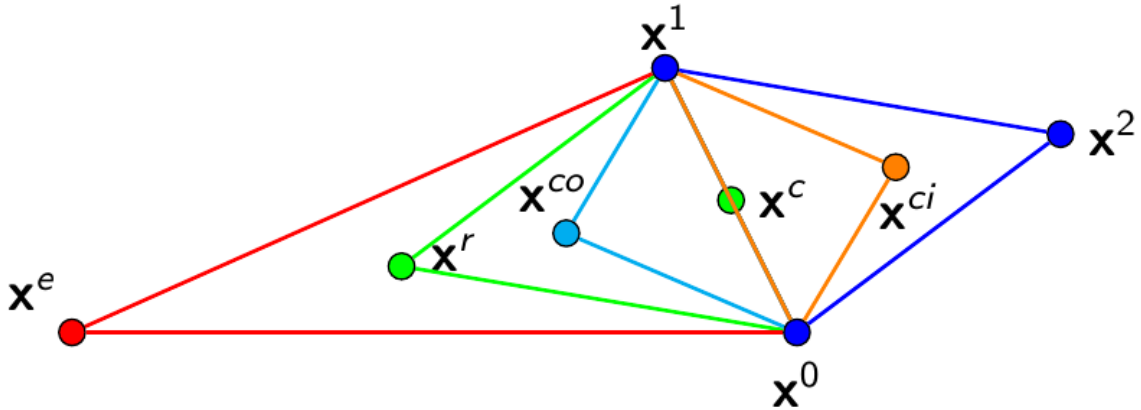
Nelder-Mead :

- on construit des simplex ($d+1$ points), on ajoute 1 point à l'opposé du point le moins bon
- on évalue F à ce point, si pas d'amélioration par rapport au meilleur point, stop (phase de "réflexion")
- si amélioration, on ajoute un point dans la même direction (on augmente la taille du pas, phase "expension")
- en fonction du résultat, stop ou phase de "contraction" ou de "rétractation"

Optimisation sans métamodèle

Nelder-Mead





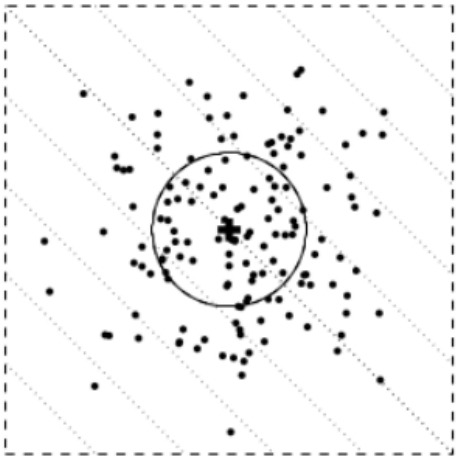
Optimisation sans métamodèle

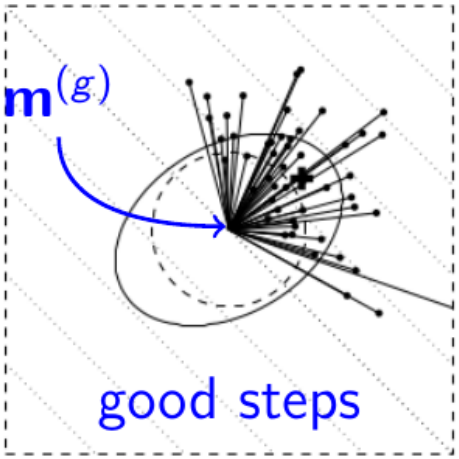
CMA-ES :

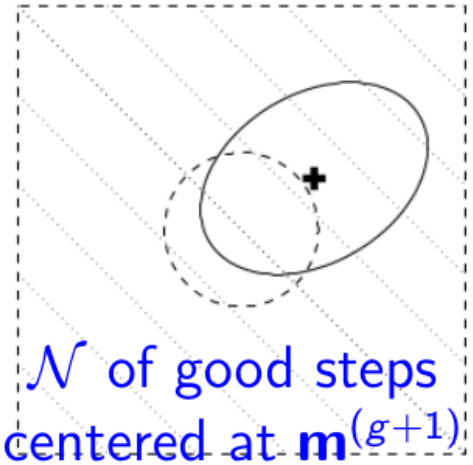
- au lieu de simplex, basé sur l'échantillonnage et l'estimation d'une distribution Normale multivariée
- on évalue la fonction d'objectif sur un premier point
- on échantillonne de nouveaux points selon une distribution centrée sur le premier point
- on évalue la fonction à ces nouveaux points
- on garde les meilleurs points, on réitère en centrant la nouvelle distribution sur la moyenne des meilleurs points
- la direction et la taille du pas (variance) sont adaptés à chaque itération

Optimisation sans métamodèle

CMA-ES







Les algorithmes Nelder-Mead et CMA-ES sont plutôt locaux et peuvent rater l'optimum global Des variantes avec "restart" permettent de mieux explorer l'ensemble de l'espace

Inférence/calibration bayésienne

On considère un paramètre à estimer comme étant une variable aléatoire au même titre que les observations

On cherche à caractériser la distribution *a posteriori* des paramètres du modèle à l'aide de la formule de Bayes : $[\theta|y] = \frac{[\theta][y|\theta]}{[y]}$

Principe général : à partir de la connaissance *a priori* de $[\theta]$, des données $[y|\theta]$ et de la vraisemblance, on met à jour la connaissance de $[\theta|y]$

Modèle non hiérarchique : il suffit de la vraisemblance et de l'*a priori* pour estimer l'*a posteriori*

Modèle hiérarchique : exemple quand un paramètre varie entre individus ou sites $\rightarrow$ hyper-paramètres à estimer en plus

Problème : multiples intégrales à calculer $\rightarrow$ utilisation de méthodes numériques stochastiques pour générer un échantillon issu de la distribution *a posteriori*

Inférence/calibration bayésienne

A quoi sert l'échantillon issu de la loi *a posteriori* :

- connaître intimement la loi *a posteriori*
- estimer la densité *a posteriori*
- estimer les moments *a posteriori* (espérance, variance)
- estimer des intervalles de crédibilité

Méthodes :

- ré-échantillonnage ("sampling importance resampling")
- Monte Carlo par chaîne de Markov (MCMC) : génération d'une chaîne (séquence de réalisations dépendantes de θ)
- échantillonneur de Gibbs (loi à l'itération i repose sur une loi conditionnelle à $i-1$)
- algorithme Metropolis-Hastings (loi arbitraire au lieu de conditionnelle)

Inférence/calibration bayésienne

A la fin du processus, diagnostiquer la convergence des chaînes (stationnarité, variances intra- et interchaînes)

Logiciels souvent utilisés : WinBUGS, OpenBUGS

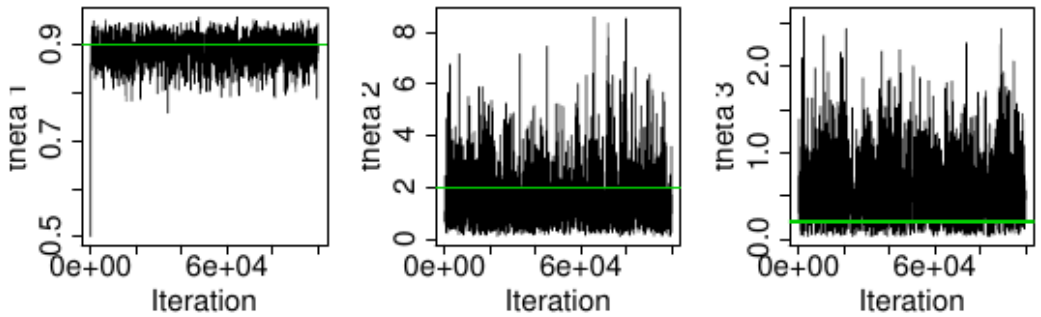
Problème : le calcul analytique de la vraisemblance peut devenir rapidement infaisable pour des modèles complexes

Solution : méthodes Approximate Bayesian Computation (ABC), permettent de ne pas mesurer la vraisemblance

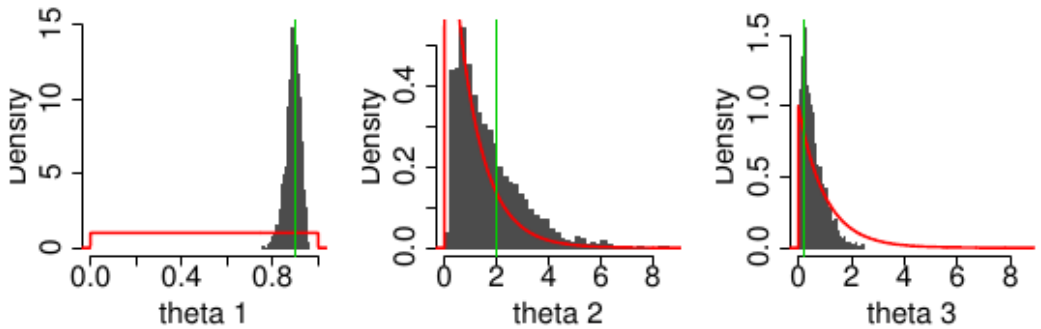
Elles réalisent des simulations, et comparent les résultats avec les données observées pour construire la distribution *a posteriori* à chaque itération

Inférence/calibration bayésienne

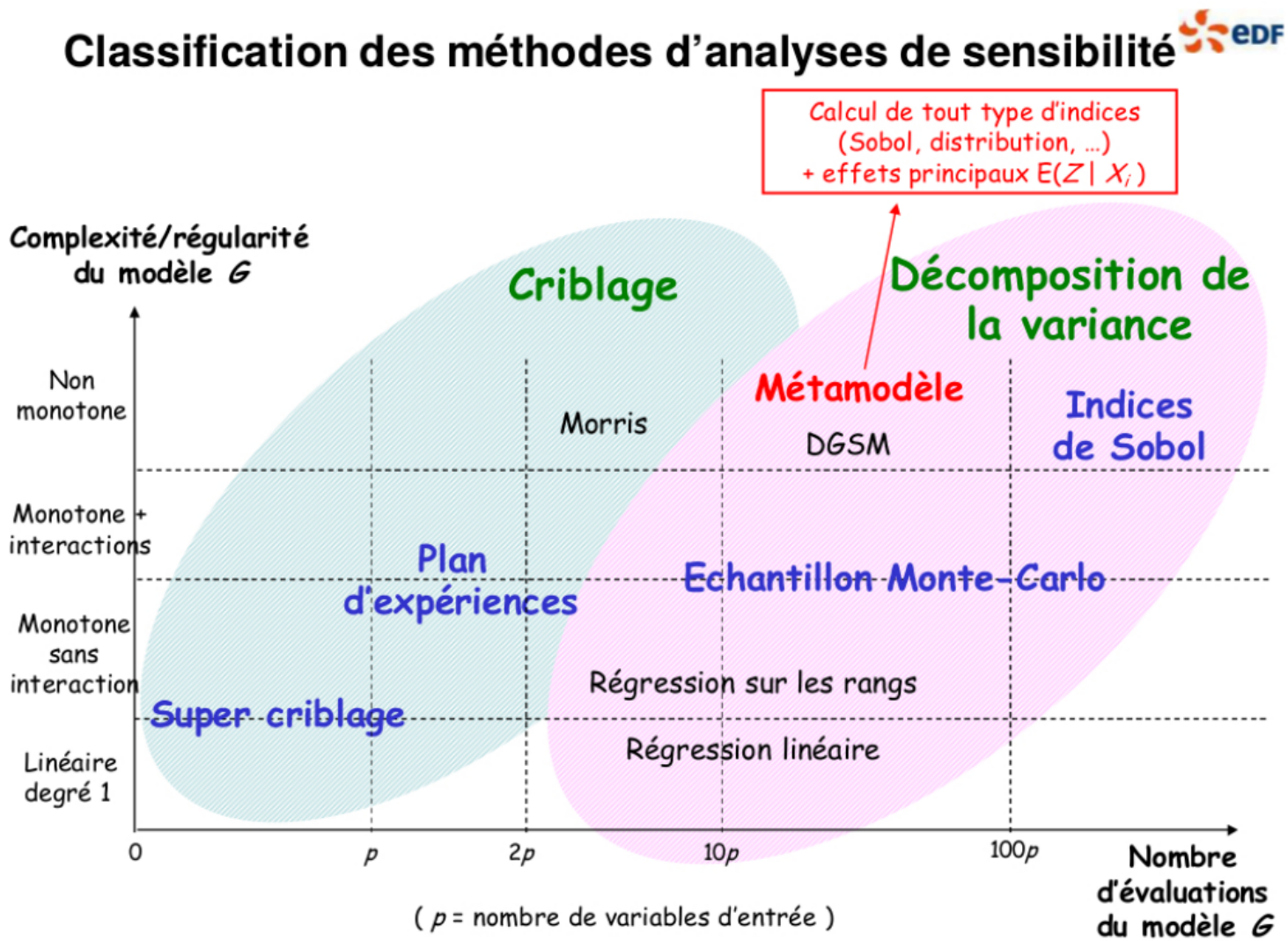
- Chaînes :



- Distributions a posteriori :

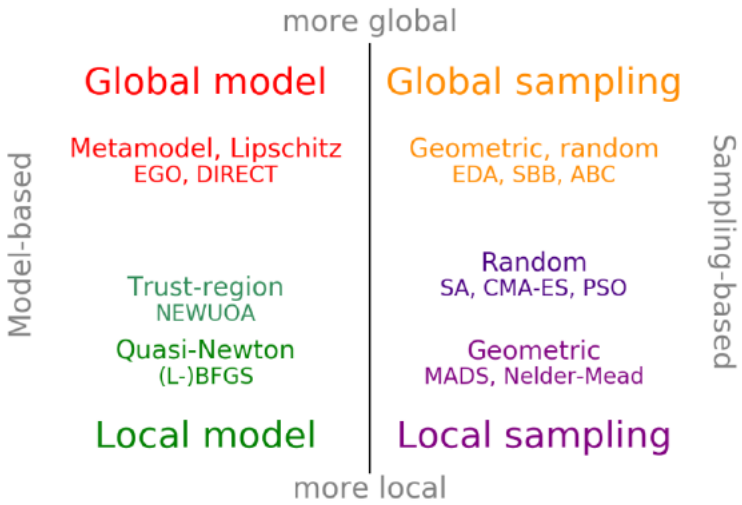


Synthèse : méthodes d'analyse de sensibilité

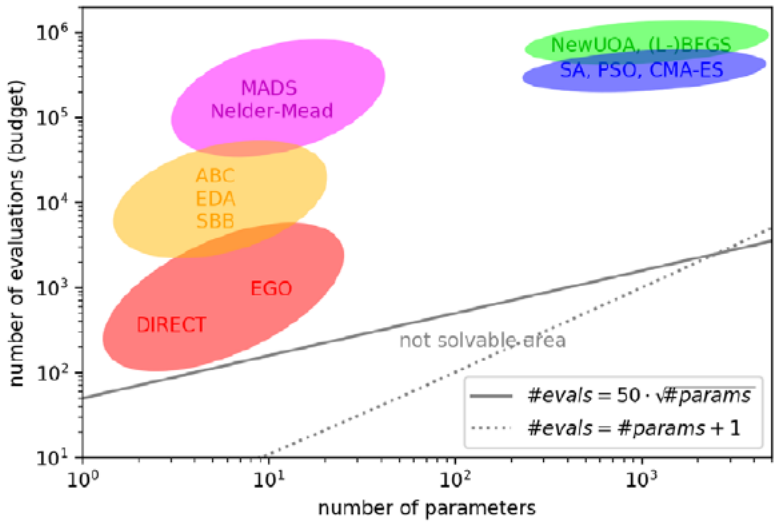


Synthèse : méthodes de calibration/optimisation/inférence

Grille 1 : Espace de projection



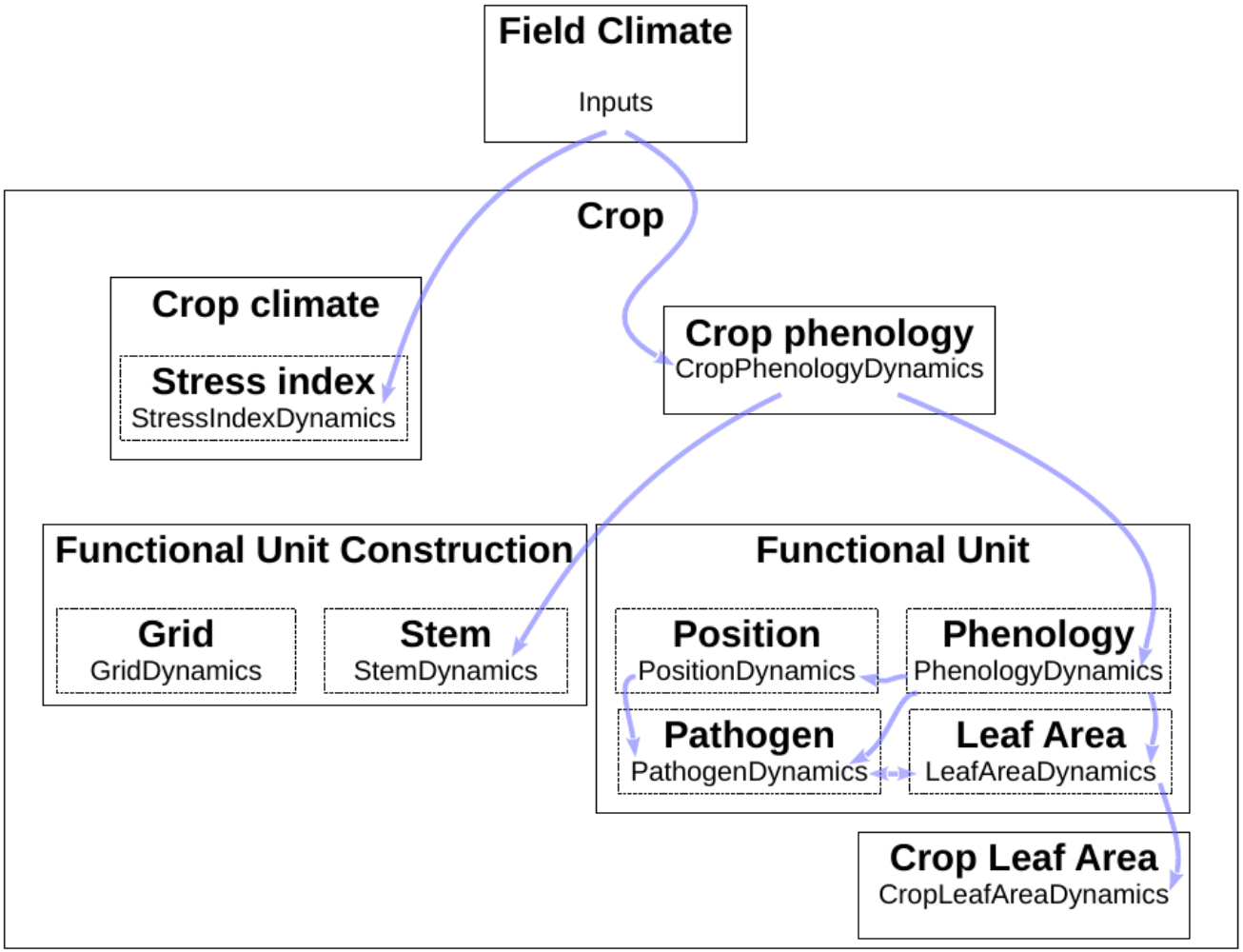
Grille 2: Aide à la sélection



Synthèse : implémentations dans R

- `lhs` (création de plans d'expériences par hypercubes latins)
- `sensitivity` (morris, sobol, src, pcc, fast)
- `DiceDesign` (plans d'expérience, dont hypercubes latins))
- `DiceEval` (construction et évaluation de métamodèles)
- `DiceKriging` (estimation, validation et prediction de modèles de krigeage)
- `DiceView` (construction de graphiques pour les métamodèles de krigeage)
- `DiceOptim` (optimisation par EGO, optimisation avec bruit)
- `cmaes` (optimisation par CMA-ES)
- `optim` (optimisation par Nelder-Mead)
- `GPareto` (optimisation multi-objectif)
- `R2OpenBUGS` (interagit avec OpenBUGS pour l'inférence bayésienne, avec échantillonneur de Gibs)
- `easyABC` (méthode ABC)

Structure du modèle



Dynamique des surfaces foliaires

Dynamique foliaire

$$\begin{cases} Z(t) = \frac{Z_{max}}{1+e^{-k_z \cdot (t-t_z)}} \\ E(t) = \frac{E_{max}}{1+e^{-k_e \cdot (t-t_e)}} \\ S(t) = \frac{E_{max}}{1+e^{-k_s \cdot (t-t_s)}} \end{cases} \qquad A(t) = E(t) - R(t)$$

Résistance ontogénique : fonction de l'âge des feuilles

$$\rho(t_d) = 1 - \frac{1}{1 + e^{-k_r \cdot (t-t_r)}}$$

Porosité : fonction de la densité foliaire

$$\phi_d = \phi_0 \cdot \frac{E_d - R_d}{Z_d}$$

Dynamique du pathogène

Dynamique des surfaces malades

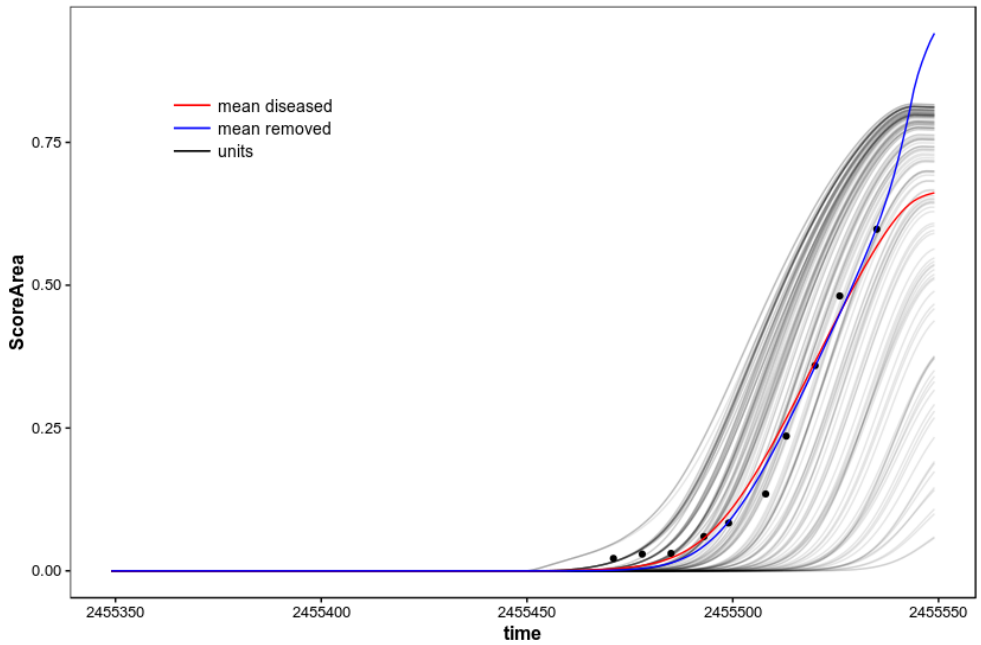
$$\begin{cases} \Delta H = \Delta E - \left(\tau_u \cdot I \cdot \frac{H}{E} \cdot \rho_d \cdot \theta_d\right) - \left(\delta i_d \cdot \frac{H}{E} \cdot \rho_d \cdot \theta_d\right) - \left(\Delta S \cdot \frac{H}{A}\right) \\ \Delta L = \left(\tau_u \cdot I \cdot \frac{H}{E} \cdot \rho_d \cdot \theta_d\right) + \left(\delta i_d \cdot \frac{H}{E} \cdot \rho_d \cdot \theta_d\right) - \left(\frac{1}{lp} \cdot L\right) - \left(\Delta S \cdot \frac{L}{A}\right) \\ \Delta I = \left(\frac{1}{lp} \cdot L\right) - \left(\frac{1}{ip} \cdot I\right) - \left(\Delta S \cdot \frac{I}{A}\right) \\ \Delta R = \left(\frac{1}{ip} \cdot I\right) + \Delta S \end{cases}$$

Flux d'échanges

$$\begin{cases} \delta o_d = \tau_n \cdot I \cdot \phi_d \\ \delta i_d = \sum_n \delta o_{d-1} \end{cases}$$

Paramétrage

Premier calibrage : 1 variable de sortie), 3 paramètres
Algorithme CMA-ES, Fonction d'objectif : minimiser la SCE



⇒ Utiliser les connaissances sur les méthodes pour calibrer le modèle sur l'ensemble des paramètres