

Analyse de Sensibilité par Polynômes de Chaos et Planification Active

JM. Martinez, S. Gazut, Th. Crestaux, O. Le Maître

Planification d'expériences - Toulouse - Février 2006

Introduction

- Les différentes analyses d'incertitudes
- L'analyse de sensibilité globale

Méthode de Sobol

- Théorème de Sobol
- Décomposition de Sobol
- Indices de sensibilité
- Calcul des indices

Polynômes de chaos

- Introduction aux PC
- Développements en Polynômes de Chaos
- Indices de sensibilité et coefficients modaux
- Calcul des coefficients modaux
- Exemples

Planification d'expériences numériques

- Apprentissage Actif
- Méthode proposée : Predictive Divergence Resampling

Conclusion

Incertitudes et imprécisions numériques

En simulation numérique, de nombreuses incertitudes/imprécisions apparaissent

- ▶ incertitudes propres à toute modélisation : les modèles ne sont que des outils traduisant plus ou moins fidèlement les phénomènes étudiés (incertitudes épistémiques)
- ▶ incertitudes sur les paramètres physiques, conditions aux limites, la variabilité naturelle de certains phénomènes (incertitudes stochastiques)
- ▶ imprécisions liées à la résolution numérique des modèles : méthodes de résolution, convergence, discrétisation, maillage, ...

Analyse d'incertitude et analyse de sensibilité

- ▶ Fonction déterministe définie sur un domaine de R^p

$$\mathbf{x} \in R^p \rightarrow y = T(\mathbf{x}) \in R$$

- ▶ \mathbf{x} vecteur aléatoire de densité $f_X(\mathbf{x}) \rightarrow f_Y(y)$
- ▶ Deux besoins en modélisation :
 - ▶ **Analyse d'incertitude** : Quelle est l'incertitude $f_Y(y)$ sur y étant donnée l'incertitude $f_X(\mathbf{x})$ sur \mathbf{x} ?
 - ▶ **Analyse de sensibilité** : Quelle est la contribution relative de l'incertitude de la composante x_i sur l'incertitude y ?

Les analyses de sensibilité

- ▶ L'analyse de sensibilité **locale** a produit de bons résultats (grands systèmes, méthodes adjointes, Cacuci 1981). Basée sur une analyse différentielle, elle se limite à l'analyse du système linéarisé

$$\mathbf{y} = T(\mathbf{x}) \Rightarrow \delta y \simeq \frac{\partial T}{\partial \mathbf{x}} \delta \mathbf{x} = \mathbf{S} \delta \mathbf{x} \Rightarrow \text{Var}(\mathbf{y}) = \mathbf{S} \text{Var}(\mathbf{x}) \mathbf{S}^T$$

- ▶ L'analyse de sensibilité **globale** vise une exploration des variations possibles des paramètres sans simplification *a priori* du modèle (Morris, Saltelli, Homma, Iman, Conover, Helton 1995)

Mesures d'analyse de sensibilité globale

L'analyse globale est fondée sur le calcul d'indices représentant la contribution relative due à une variable ou à un groupe de variables sur la variance de la réponse y

- ▶ Indices de sensibilité du premier ordre : part de variance expliquée par une variable

$$S_i = \frac{\text{Var}(E(Y|X_i))}{\text{Var}(Y)}$$

- ▶ Indices de sensibilité d'ordre supérieur : part de variance expliquée par un groupe de variables

$$S_{i_1, \dots, i_k} = \frac{\text{Var}(E(Y|X_{i_1}, \dots, X_{i_k}))}{\text{Var}(Y)}$$

Théorème de Sobol (1993) → analyse de la variance

- ▶ Soit $f()$ définie dans $[0, 1]^p$. Soit x_u les composantes dont les indices $\subset u$, $f()$ se décompose en (A. Antoniadis) :

$$f(x) = \sum_{u \subseteq \{1, 2, \dots, p\}} f_u(x_u)$$

$$\text{avec } \int f_u(x_u) dx_i = 0 \quad \forall i \in u \neq \emptyset \quad \text{et } f_\emptyset(x_\emptyset) = f_0 = \text{constante}$$

⇒ D'où les propriétés :

$$\text{Moyenne } f_0 = \int f(x) dx$$

$$\text{Orthogonalité } \int f_u(x_u) f_v(x_v) dx = 0 \quad \forall u \neq v$$

Calcul de la décomposition de Sobol

- ▶ les f_u s'obtiennent par récurrence, à partir de la moyenne

$$f_0 = \int f(x) dx$$

- ▶ puis en notant \bar{u} le complémentaire de u dans $\{1, \dots, p\}$

$$f_u(x_u) = \int f(x) dx_{\bar{u}} - \sum_{v \subset u} f_v(x_v)$$

Indices de sensibilité

- ▶ Grâce à l'orthogonalité, on décompose la variance

$$\int f^2(x) dx = \sum_{u \subseteq \{1,2,\dots,p\}} \underbrace{\int f_u^2(x_u) dx_u}_{\sigma_u^2}$$

$$\sigma^2(f) = \sum_{u \subseteq \{1,2,\dots,p\}, \neq \emptyset} \sigma_u^2$$

- ▶ puis on obtient les différents indices de sensibilité

$$\frac{\text{Var}(E(Y|X_{i_1}, \dots, X_{i_k}))}{\text{Var}(Y)} = \frac{1}{\sigma^2(f)} \sum_{u \subseteq \{i_1, i_2, \dots, i_k\}, \neq \emptyset} \sigma_u^2$$

Calcul des indices

Le calcul des indices de sensibilité globale se ramène à un calcul d'intégrales. Pour un code $T()$ complexe, vu comme une boîte noire, les indices peuvent être approchés par

- ▶ par méthode de Monte Carlo
- ▶ à partir d'une décomposition des réponses de $T(x)$ sur une base orthogonale de polynômes de chaos :
 - ▶ coefficients obtenus par quadrature de Gauss ou par régression
 - ▶ relations explicites entre les indices et les coefficients

Introduction aux PC

Récemment utilisés pour modéliser les incertitudes (Ghanem, Xiu, Le Maître). Famille de polynômes orthogonaux associée à une mesure

$$\langle f(\xi)g(\xi) \rangle = \int f(\xi)g(\xi)\text{pdf}(\xi)d\xi$$

- ▶ Exemple unidimensionnel pour la pdf normale associée aux polynômes de Hermite

$$\varphi_0 = 1 \quad \varphi_1 = \xi \quad \varphi_2 = \xi^2 - 1 \quad \varphi_3 = \xi^3 - 3\xi \dots$$

- ▶ Généralisation au cas multidimensionnel par tensorisation

$$\text{Nombre de coefficients } p + 1 = \frac{(nx + no)!}{nx!no!}$$

$$\Psi_k(\xi) = \prod_{i=1}^{nx} \varphi_{\alpha_{(k,i)}}(\xi_i) \quad \langle \Psi_k, \Psi_l \rangle = \|\Psi_k\|^2 \delta_{kl}$$

Utilisation des PC

Les polynômes de chaos peuvent être utilisés en mode intrusif (réécriture du code, système d'équations couplées) ou en mode non intrusif :

- approximation de code $T(\cdot)$ par un méta-modèle $y(\cdot)$ (surrogate model) développé sur une base finie de PC

$$T(\mathbf{x}) = \sum_{k=0}^{\infty} \beta_k^* \Psi_k(\mathbf{x}) \Rightarrow y_{\text{opt}}(\mathbf{x}) = \sum_{k=0}^p \beta_k^* \Psi_k(\mathbf{x})$$

$$T(\mathbf{x}) = \underbrace{\tilde{y}(\mathbf{x}) - \sum_{k=0}^p (\beta_k - \beta_k^*) \Psi_k(\mathbf{x})}_{\text{erreur d'estimation}} + \underbrace{\sum_{k=p+1}^{\infty} \beta_k^* \Psi_k(\mathbf{x})}_{\text{erreur d'approximation}}$$

$y_{\text{opt}}(\mathbf{x})$

Indices de sensibilité et coefficients modaux

Une relation simple existe entre les variances partielles σ_u^2 et les coefficients modaux β_k

► Variance

$$y(\mathbf{x}) \simeq \sum_{k=1}^P \beta_k \Psi_k(\mathbf{x}) \Rightarrow \sigma^2(y) \simeq \sum_{k=1}^P \beta_k^2 \|\Psi_k\|^2$$

► Variances partielles

$$\sigma_u^2(y) \simeq \sum_{k \in U(u)} \beta_k^2 \|\Psi_k\|^2 \quad \text{où } U(u) = \{k, \Psi_k(x) = \Psi_k(x_u)\}$$

Méthodes de calcul des coefficients modaux I

- ▶ Par **cubature déterministe** en multidimensionnel

$$\text{Orthogonalité des } \Psi_k \Rightarrow \beta_k = \frac{1}{\|\Psi_k\|^2} \int T(x) \Psi_k(x) \text{pdf}(x) dx$$

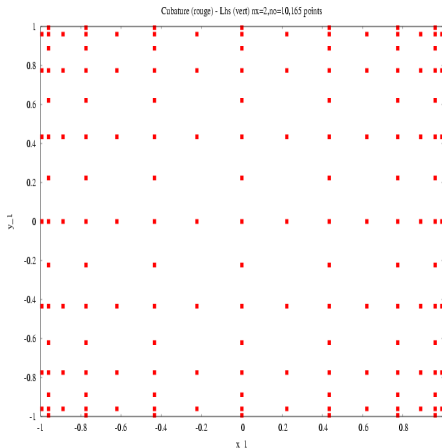
$$\text{Quadrature} \Rightarrow \beta_k \simeq \frac{1}{\|\Psi_k\|^2} \sum_i T(x^{(i)}) \Psi_k(x^{(i)}) \omega_i$$

- ▶ Méthode Smolyak (logiciel Petras) → Plans d'expériences déterministes

Points et poids de la quadrature → $\{x^{(i)}, \omega_i\}_{i=1, \dots, n}$

Plan obtenu par cubature

- Cas $n_x = 2$ $n_o = 10 \Rightarrow 165$ points



Méthodes de calcul des coefficients modaux II

- ▶ Par **régression linéaire** moindres carrés à partir d'un échantillon aléatoire (hypercubes latins)

$$\beta = (Z^T Z)^{-1} ZY$$

où $Z_{ij} = \Psi_j(x^{(i)})$, $Y_i = T(x^{(i)})$

- ▶ Régularisation par régression sur composantes principales (PCR) en ne retenant que les q plus grandes valeurs propres (matrice d'information $Z^T Z$ mieux conditionnée)

par SVD décomposition de $Z = USV^T$

$$\beta = VS^{-1}U^T Y \text{ ou par PCR : } \beta^q = VS_q^{-1}U^T Y$$

Méthodes de calcul des coefficients modaux III

- ▶ Par **régression approchée** (Antoniadis)

$$n \text{ est grand } (Z^T Z) \sim nl \Rightarrow \beta = \frac{1}{n} ZY$$

$$\beta_k = \frac{1}{n} \sum_{i=1}^n y(x_i) \Psi_k(x_i)$$

- ▶ méthode équivalente au Monte Carlo. Amélioration de l'estimation des coefficients par technique de **rétrécissement** :

$$\beta_k = \frac{\beta_k^2}{\beta_k^2 + \text{Var}(\beta_k)} \beta_k$$

Exemple de Homma Saltelli

$$f(x_1, x_2, x_3) = \sin(\pi x_1) + a * \sin^2(\pi x_2) + b * (\pi x_3)^2 \sin(\pi x_1)$$

- ▶ densités des x_i : loi uniforme entre $[-1, +1]$
- ▶ la variance $\sigma^2 = \frac{a^2}{8} + \frac{b\pi^4}{5} + \frac{b^2\pi^8}{18} + \frac{1}{2}$
- ▶ les variances partielles

$$\begin{aligned}\sigma_1^2 &= \frac{b\pi^4}{5} + \frac{b^2\pi^8}{50} + 0.5, \sigma_2^2 = \frac{a^2}{8}, \sigma_{13}^2 = b^2\pi^8\left(\frac{1}{18} - \frac{1}{50}\right) \\ \sigma_3^2 &= \sigma_{12}^2 = \sigma_{23}^2 = \sigma_{123}^2 = 0\end{aligned}$$

- ▶ le cas étudié $a = 7, b = 0.1$

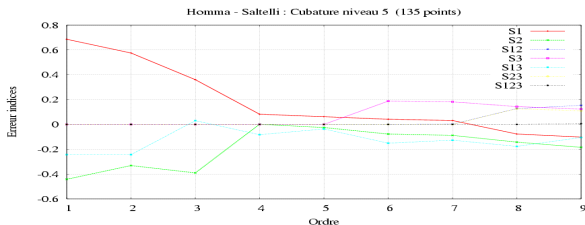
Exemple de Saltelli Sobol

$$f(\mathbf{x}) = \prod_{i=1}^p g_i(x_i) \text{ avec } g_i(x_i) = \frac{|4x_i - 2| + a_i}{1 + a_i} \text{ avec } a_i \geq 0$$

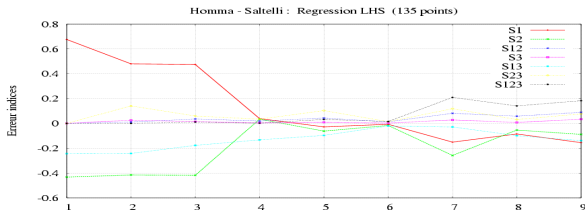
- ▶ les densités $\text{pdf}(x_i) = \begin{cases} 1 & \text{si } 0 \leq x_i \leq 1 \\ 0 & \text{sinon} \end{cases}$
- ▶ la variance $\sigma^2 = -1 + \prod_{i=1}^p (\sigma_i^2 + 1)$ avec $\sigma_i^2 = \frac{1}{3(1+a_i)^2}$
- ▶ les variances partielles $\sigma_{i_1, \dots, i_s}^2 = \prod_{j=1}^s \sigma_{i_j}^2$
- ▶ le cas étudié $p = 4, a_i = i$

Homma Saltelli - 135 évaluations

► Cubature

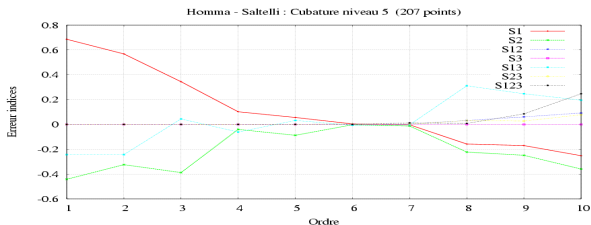


► Régression

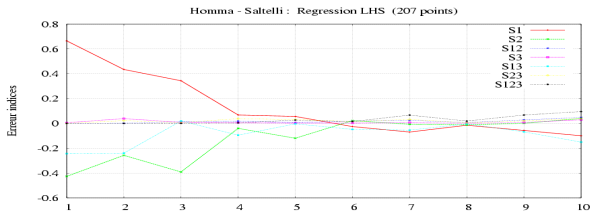


Homma Saltelli - 207 évaluations

► Cubature

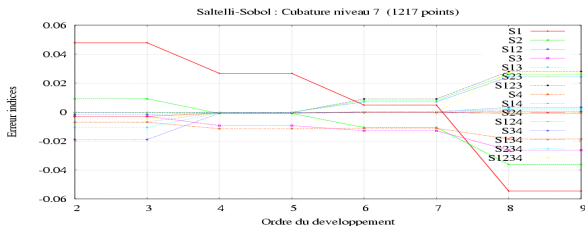


► Régression

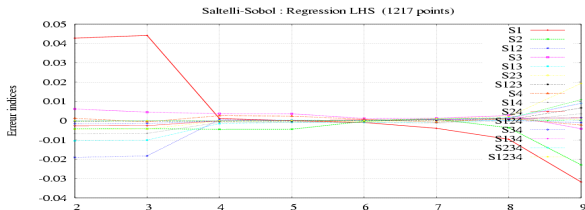


Saltelli Sobol - 1217 évaluations

► Cubature

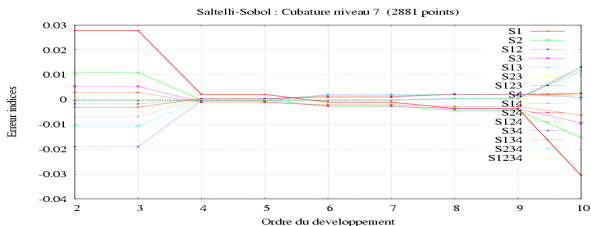


► Régression

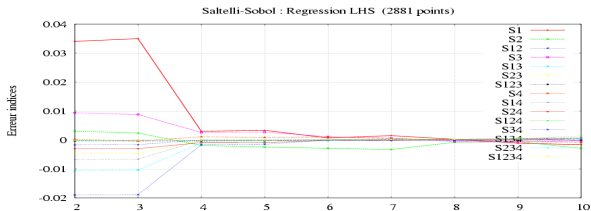


Saltelli Sobol - 2881 évaluations

► Cubature

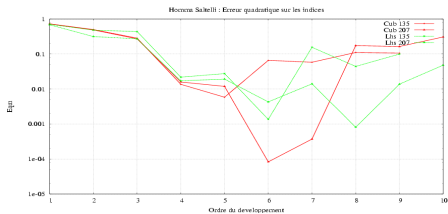


► Régression

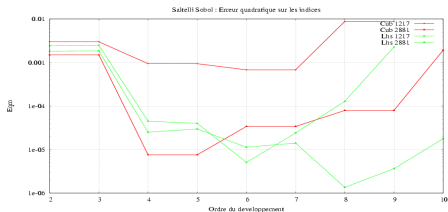


Comparaison des erreurs quadratiques

► Homma Saltelli

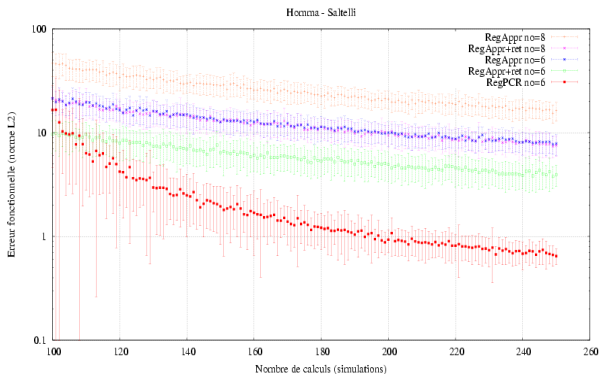


► Saltelli Sobol



Comparaison Régression - Régression approchée

► Homma Saltelli



Comparaison des méthodes

- ▶ Critère erreur fonctionnelle fonction de la taille de l'échantillon

Régression approchée	<	Régression approchée rétrécie
Régression approchée rétrécie	<	Régression Composantes Principales
Cas Homma Saltelli : Cubature	\simeq	Régression Composantes Principales
Cas Saltelli Sobol : Cubature	<	Régression Composantes Principales

- ▶ à envisager les méthodes d'apprentissage fondées sur
 - ▶ des combinaisons de fonctions par PLS (Partial Least Square),
 - ▶ la sélection de fonctions par Boosting (L_2 Regression de Bühlmann)

Planification d'expériences numériques

- ▶ La théorie des plans d'expériences est fondée sur un certain nombre d'hypothèses probabilistes du résidu. Le plan est choisi de façon à réduire l'influence de la variabilité des expériences sur les coefficients du modèle.
- ▶ Analyse spectrale de la matrice de dispersion $(Z^T Z)^{-1}$
 - ▶ D-Optimalité : déterminant
 - ▶ A-optimalité : trace
 - ▶ E-Optimalité : valeur propre maximale
- ▶ En simulation, pour des expériences numériques déterministes, la théorie n'est plus vraiment justifiée, l'état de l'art actuel s'appuie sur les méthodes basées l'apprentissage actif (Cohn, Melville,Chapelle).

Machine Learning Theory - Active Learning

Un site sur Active Learning <http://hunch.net>

- ▶ There is a source of unlabeled data \Rightarrow les simulations numériques potentielles
- ▶ There is an oracle from which labels can be requested for unlabeled data produced by the source \Rightarrow l'oracle est le code de calcul
- ▶ The goal is to perform well with minimal use of the oracle \Rightarrow approximation du code avec le minimum de simulations
- ▶ Choix des requêtes fondées sur l'analyse de la divergence des modèles hypothèses (Melville)

Divergence Prédicative par ré-échantillonnage

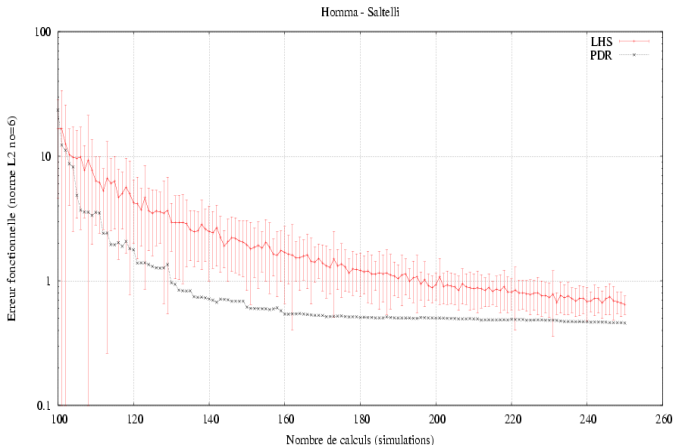
- ▶ Echantillon $\mathcal{L}_n = \{(\mathbf{x}_k, y_k); k = 1, \dots, n\}$
- ▶ Ré-échantillonnage $\mathcal{L}_n^b = \{(\mathbf{x}_k^b, y_k^b); k = 1, \dots, n\}, b = 1, \dots, B$
- ▶ Modèles hypothèses $\tilde{y}_b(\mathbf{x})$ par ré-échantillon
- ▶ Modèle agrégé (Bagging) $y_{\text{bag}}(\mathbf{x}) = \frac{1}{B} \sum_{b=1}^B \tilde{y}_b(\mathbf{x})$
- ▶ la mesure d'utilité de \mathbf{x} est basée sur la divergence des hypothèses représentée (par exemple) par la variance des prédictions $\tilde{y}_b(\mathbf{x})$

$$\mathbf{x}_{n+1} = \text{Arg}_{\mathbf{x} \in U} \text{Max Var}(\tilde{y}_b(\mathbf{x}))$$

$$\mathcal{L}_{n+1} = \mathcal{L}_n \cup \{(\mathbf{x}_{n+1}, y(\mathbf{x}_{n+1}))\}$$

Predictive Divergence Resampling

► PDR-LeaveOneOut et LHS



Conclusions

- ▶ Intérêt des PC pour une analyse de sensibilité globale : lien direct entre les coefficients et les indices de sensibilité
- ▶ Choix entre Monte Carlo, cubature et régression ?
- ▶ Critères de construction/sélection du modèle : ordre, régresseurs, combinaison de régresseurs par PLS, boosting
- ▶ Planification itérative des expériences numériques fondée sur l'apprentissage actif ?
 - ▶ Divergence d'hypothèses par entropie ?
 - ▶ Aggrégation des modèles par le modèle médian ?