Pratique de l'analyse de sensibilité : comment évaluer l'impact des entrées aléatoires sur la sortie d'un modèle mathématique

Julien JACQUES
http://labomath.univ-lille1.fr/∼jacques/

25 mars 2011

Ce document a pour objetif de guider les praticiens de tous domaines désirant réaliser l'analyse de sensibilité d'un modèle mathématique, et est pour partie extrait de la thèse de l'auteur [4].

Résumé

L'analyse de sensibilité globale (AS) permet d'analyser un modèle mathématique en étudiant l'impact de la variabilité des facteurs d'entrée du modèle sur la variable de sortie. Déterminant les entrées responsables de cette variabilité à l'aide d'indices de sensibilité, l'AS permet de prendre les mesures nécessaires pour diminuer la variance de la sortie si celle-ci est synonyme d'imprécision, ou encore d'alléger le modèle en fixant les entrées dont la variabilité n'influe pas la variable de sortie. Nous présentons dans ce document les principaux indices de sensibilité, basés sur l'hypothèse d'indépendance des variables d'entrée, leurs estimations, puis abordons le cas des modèles à entrées non indépendantes. Deux applications numériques illustrent l'inteprétation des indices de sensibilité dans le cas de modèle à entrées indépendantes et dépendantes.
Mots clès : analyse de sensibilité, décomposition de la variance, indices de Sobol, entrées dépendantes.

1 Introduction : les objectifs de l'analyse de sensibilité

Considérons un modèle mathématique qui, à un ensemble de variables d'entrée aléatoires $\mathbf{X}$, fait correspondre, via une fonction f déterministe, une variable de sortie Y (ou réponse) aléatoire :

$$\begin{array}{rcl} f : \mathbb{R}^p & \rightarrow & \mathbb{R} \\ \mathbf{X} & \mapsto & Y = f(\mathbf{X}) \end{array} \quad (1)$$

La fonction f du modèle peut être très complexe (système d'équations différentielles...), et est en pratique évaluée à l'aide d'un code informatique, plus ou moins onéreux en temps de calcul. L'ensemble des variables d'entrée $\mathbf{X} = (X_1, ..., X_p)$ regroupe toutes les entités considérées comme aléatoires dans le modèle.

L'analyse de sensibilité étudie comment des perturbations sur les variables d'entrée du modèle engendrent des perturbations sur la variable réponse. L'auteur interessé par un ouvrage de référence pourra se référer à [12]. Il est possible de grouper les méthodes d'analyse de sensibilité en trois classes : les méthodes de *screening*, qui consistent en une analyse qualitative de la sensibilité de la variable de sortie aux variables d'entrée, les méthodes d'analyse locale [18], qui évaluent quantitativement l'impact d'une petite variation autour d'une valeur donnée des entrées, et enfin les méthodes d'analyse de sensibilité globale, qui s'intéressent à la variabilité de la sortie du modèle dans l'intégralité de son domaine de variation. L'analyse de sensibilité globale étudie comment la variabilité des entrées se répercute sur celle de la sortie, en déterminant quelle part de variance de la sortie est due à telle entrée ou tel ensemble d'entrées. Si l'analyse de sensibilité locale s'intéresse plus à la valeur de la variable réponse, l'analyse de sensibilité globale s'intéresse quant à elle à sa variabilité. Nous nous intéressons dans ce document à l'analyse de sensibilité globale et omettrons donc par la suite l'adjectif global.

Les enjeux de l'analyse de sensibilité peuvent être multiples : validation d'une méthode ou d'un code de calcul, orientation des efforts de recherche et développement, ou encore justification en terme de sûreté d'un dimensionnement ou d'une modification d'un système. Nous décrivons ci-après les principales questions auxquelles l'analyse de sensibilité permet d'apporter des éléments de réponse.

Les ambitions de l'analyse de sensibilité Au cours de l'élaboration, de la construction ou de l'utilisation d'un modèle mathématique, l'analyse de sensibilité peut s'avérer être un outil précieux. En effet, en étudiant comment la réponse du modèle réagit aux variations de ses variables d'entrée, l'analyse de sensibilité permet de répondre à un certain nombre de questions.

1. Le modèle est-il bien fidèle au phénomène/processus modélisé ?
   En effet, si l'analyse exhibe une forte influence d'une variable d'entrée habituellement connue comme non influente, il sera nécessaire de remettre en cause la qualité du modèle ou (et) la véracité de nos connaissances sur l'impact réel des variables d'entrée.
2. Quelles sont les variables qui contribuent le plus à la variabilité de la réponse du modèle ?
   Si cette variabilité est synonyme d'imprécision sur la valeur prédite de la sortie, il sera alors possible d'améliorer la qualité de la réponse du modèle à moindre coût. En effet, la variabilité de la sortie du modèle pourra être diminuée en concentrant les efforts sur la réduction des variabilités des entrées les plus influentes. Il doit être précisé que cela n'est pas toujours possible, notamment lorsque la variabilité d'une variable d'entrée est intrinsèque à la nature de la variable et non due à un manque d'information ou à des imprécisions de mesures.
3. Quelles sont au contraire les variables les moins influentes ?
   Il sera possible de les considérer comme des paramètres déterministes, en les fixant par exemple à leur espérance, et obtenir ainsi un modèle plus *léger* avec moins de variables d'entrée. Dans le cas d'un code informatique, il sera possible de supprimer les parties de codes qui n'ont aucune influence sur la valeur et la variabilité de la réponse.
4. Quelles variables, ou quels groupes de variables, interagissent avec quelles (quels) autres ?
   L'analyse de sensibilité peut permettre de mieux appréhender et comprendre le phénomène modélisé, en éclairant les relations entre les variables d'entrée.

Bon nombre de publications sur le sujet explicitent et illustrent ces objectifs. On pourra se référer notamment aux travaux de Saltelli et al. [13, 14, 16].

La section suivante présente les indices de sensibilité définis pour des modèles à variables d'entrée indépendantes, ainsi que leur méthode d'estimation. La section 3 s'intéresse aux modèles à entrées non indépendantes, et présente deux types d'indices utilisables dans ce cas. Enfin, la section 4 présente deux applications simulées illustrant l'intérêt et l'interprétation des indices de sensibilités, dans le cas de modèles à entrées indépendantes et non indépendantes.

2 Indicateurs de sensibilité pour modèles à entrées indépendantes

Nous supposons dans cette section que les variables d'entrée $\mathbf{X} = (X_1, \ldots, X_p)$ du modèle sont indépendantes.

2.1 Préambule : cas du modèle linéaire

Supposons que le modèle étudié soit linéaire, et qu'il s'écrive sous la forme suivante :

$$Y = \beta_0 + \sum_{i=1}^{p} \beta_i X_i. \tag{2}$$

Comme les variables X_i sont supposées indépendantes, la variance de Y s'écrit alors :

$$\mathrm{V}(Y) = \sum_{i=1}^{p} \beta_i^2 \mathrm{V}(X_i),$$

où $\beta_i^2 \mathrm{V}(X_i)$ est la part de variance due à la variable X_i. La sensibilité de Y à X_i peut donc simplement être quantifié par le rapport de la part de variance due à X_i sur la variance totale. On définit ainsi l'indice de sensibilité SRC (*Standardized Regression Coefficient*) :

$$SRC_i = \frac{\beta_i^2 \mathrm{V}(X_i)}{\mathrm{V}(Y)}. \tag{3}$$

Il exprime la part de variance de la réponse Y due à la variance de la variable X_i. Cet indice SRC, toujours positif ($SRC \in [0, 1]$), est en outre le carré du coefficient de corrélation linéaire entre la réponse du modèle et ses variables d'entrée.

2.2 Les indices de Sobol

Plaçons nous désormais dans le cas d'une fonction f dont la forme analytique n'est pas connue. Pour apprécier l'importance d'une variable d'entrée X_i sur la variance de la sortie Y, nous étudions à combien la variance de Y décroît si on fixe la variable X_i à une valeur x_i^* : $\mathrm{V}(Y|X_i = x_i^*)$. Le problème de cet indicateur est le choix de la valeur x_i^* de X_i, que l'on résout en considèrant l'espérance de cette quantité pour toutes les valeurs possibles de x_i^* : $\mathrm{E}[\mathrm{V}(Y|X_i)]$. Ainsi, plus la variable X_i sera importante vis-à-vis de la variance de Y, plus cette quantité sera petite. Etant donné la formule de la variance totale $\mathrm{V}(Y) = \mathrm{V}(\mathrm{E}[Y|X_i]) + \mathrm{E}[\mathrm{V}(Y|X_i)]$, nous pouvons utiliser de façon équivalente la quantité

$$\mathrm{V}(\mathrm{E}[Y|X_i]),$$

qui sera d'autant plus grande que la variable X_i sera importante vis-à-vis de la variance de Y. Afin d'utiliser un indicateur normalisé, nous définissons l'indice de sensibilité de Y à X_i :

$$S_i = \frac{\mathrm{V}(\mathrm{E}[Y|X_i])}{\mathrm{V}(Y)}. \tag{4}$$

Cet indice est appelé **indice de sensibilité de premier ordre** par Sobol [17], *correlation ratio* par McKay [6], ou encore *importance measure*. Il quantifie la sensibilité de la sortie Y à la variable d'entrée X_i, ou encore la part de variance de Y due à la variable X_i.

Remarque. *Dans le cas du modèle linéaire (2), cet indice de sensibilité est égal à l'indice SRC, puisque $V(E[Y|X_i]) = V(\beta_i X_i) = {\beta_i}^2 V(X_i)$.*

Sobol [17] a introduit cet indice de sensibilité en décomposant la fonction f du modèle en somme de fonctions de dimensions croissantes :

$$Y = f(X_1, ..., X_p) = f_0 + \sum_{i=1}^{p} f_i(X_i) + \sum_{1 \le i < j \le p} f_{ij}(X_i, X_j) + \ldots + f_{1,\ldots p}(X_1, ..., X_p). \tag{5}$$

où

$$\begin{aligned}
f_0 &= \mathrm{E}[Y], \\
f_i(X_i) &= \mathrm{E}[Y|X_i] - \mathrm{E}[Y], \\
f_{i,j}(X_i, X_j) &= \mathrm{E}[Y|X_i, X_j] - \mathrm{E}[Y|X_i] - \mathrm{E}[Y|X_j] + \mathrm{E}[Y], \\
f_{i,j,k}(X_i, X_j, X_k) &= \mathrm{E}[Y|X_i, X_j, X_k] - \mathrm{E}[Y|X_i, X_j] - \mathrm{E}[Y|X_i, X_k] - \mathrm{E}[Y|X_j, X_k] \ldots
\end{aligned}$$

La variance de Y, V, peut alors se décomposer selon le théorème suivant.

Théorème. *Décomposition de Sobol de la variance.*
La variance du modèle à entrées indépendantes (1) se décompose en :

$$V = \sum_{i=1}^{p} V_i + \sum_{1 \le i < j \le p} V_{ij} + \ldots + V_{1\ldots p}, \tag{6}$$

où

$$\begin{aligned} V_i &= V(E[Y|X_i]), \\ V_{ij} &= V(E[Y|X_i, X_j]) - V_i - V_j, \\ V_{ijk} &= V(E[Y|X_i, X_j, X_k]) - V_{ij} - V_{ik} - V_{jk} - V_i - V_j - V_k, \\ &\dots \\ V_{1\dots p} &= V - \sum_{i=1}^{p} V_i - \sum_{1\le i<j\le p} V_{ij} - \dots - \sum_{1\le i_1<\dots<i_{p-1}\le p} V_{i_1\dots i_{p-1}} \end{aligned}$$

Sobol se base sur cette décomposition pour définir des indices de sensibilité d'ordre supérieur à un. Les **indices de sensibilité d'ordre deux** :

$$S_{ij} = \frac{V_{ij}}{V}$$

expriment la sensibilité de la variance de Y à l'interaction des variables X_i et X_j, c'est-à-dire la sensibilité de Y aux variables X_i et X_j qui n'est pas prise en compte dans l'effet des variables seules. Les indices de sensibilité d'ordre trois :

$$S_{ijk} = \frac{V_{ijk}}{V}$$

expriment la sensibilité de la variance de Y aux variables X_i, X_j et X_k qui n'est pas prise en compte dans l'effet des variables seules et des interactions deux à deux. Et ainsi de suite jusqu'à l'ordre p.
L'**interprétation de ces indices est facile**, puisque grâce à (6), **leur somme est égale à 1**, et étant tous positifs, plus l'indice sera grand (proche de 1), plus la variable aura d'importance.

Le nombre d'indices de sensibilité ainsi construit, de l'ordre 1 à l'ordre p, est égale à $2^p - 1$. Lorsque le nombre de variables d'entrée p est trop important, le nombre d'indices de sensibilité explose. L'estimation et l'interprétation de tous ces indices deviennent vite impossible. Homma et Saltelli [2] ont alors introduit des indices de sensibilité totaux, qui expriment la sensibilité totale de la variance Y à une variable, c'est-à-dire la sensibilité à cette variable sous toutes ses formes (sensibilité à la variable seule et sensibilité aux interactions de cette variable avec d'autres variables).
L'**indice de sensibilité total** S_{T_i} à la variable X_i est défini comme la somme de tous les indices de sensibilité relatifs à la variable X_i :

$$S_{T_i} = \sum_{k \# i} S_k. \tag{7}$$

où $\#i$ représente tous les ensembles d'indices contenant l'indice i.
Exemple : pour un modèle à trois variables d'entrée $S_{T_1} = S_1 + S_{12} + S_{13} + S_{123}$.

2.3 Estimation des indices de Sobol

Estimation de Monte Carlo Dans beaucoup de problèmes scientifiques, on est amené à calculer une intégrale du type

$$I = \int_D f(\mathbf{x}) d\mathbf{x},$$

où D est un espace de plus ou moins grande dimension, et f une fonction (intégrable). Soit $x_1, \dots, x_N$ la réalisation d'un N-échantillon d'une variable aléatoire uniforme sur D. Nous supposons cet échantillon pris de manière totalement aléatoire (échantillonnage aléatoire). Une approximation de I par la méthode de Monte Carlo est faite par :

$$\hat{I}_N = \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{x}_i).$$

La convergence (presque sûre) de I_N vers I découle directement de la loi forte des grands nombres. Cette méthode d'estimation permet alors d'estimer l'espérance de toute fonction d'une variable aléatoire

de densité quelconque par

$$\hat{\mathrm{E}}[f(X)] = \frac{1}{N}\sum_{i=1}^{N} f(x_i),$$

où $(x_i)_{i=1..N}$ est un N-échantillon de réalisations de la variable aléatoire X. Le taux de convergence d'une méthode de Monte Carlo est en $\mathcal{O}(N^{-\frac{1}{2}})$.

Bon nombre de méthodes alternatives ont été proposées pour améliorer la convergence, parmi lesquelles les méthodes de simulation pseudo-probabilistes[1], comme l'échantillonnage stratifié ou par hypercube latin (*LHS*) [7], les méthodes de Quasi-Monte Carlo [8], ou encore les méthodes de Quasi-Monte Carlo Randomisé [10]. L'échantillonnage stratifié consiste à découper l'espace des variables d'entrée en petits espaces disjoints, puis à échantillonner au sein de chacun de ces sous espaces. L'échantillonnage *LHS* est basé sur le même principe, en s'assurant que le découpage a défini des espaces équiprobables, et que chaque espace est bien échantillonné ; le quadrillage se fait dans le cube unité, pour un tirage aléatoire d'échantillon uniforme, puis ces échantillons sont transformés via la fonction de répartition inverse. Les méthodes de Quasi-Monte Carlo sont des versions déterministes des méthodes de Monte Carlo. Ces méthodes définissent des séquences d'échantillons déterministes qui ont une discrépance plus faibles que les séquences aléatoires, c'est-à-dire qu'elles ont une meilleure répartition uniforme dans l'espace des variables d'entrée. Ces méthodes de quasi-Monte Carlo permettent d'obtenir une convergence plus rapide en $\mathcal{O}(N^{-1}(logN)^{p-1})$ (sous des conditions relativement faibles de régularité de f). Les méthodes de Quasi-Monte Carlo Randomisé, sous certaines conditions peu restrictives sur f, ont un taux de convergence en $\mathcal{O}(N^{-\frac{3}{2}}(logN)^{\frac{p-1}{2}})$, et permettent une approximation de l'erreur d'estimation. Owen [9] présente ces méthodes comme une ré-randomisation des séquences utilisées dans les méthodes de quasi-Monte Carlo : on prend les séquences déterministes a_i de ces dernières, et on les transforme en variables aléatoire x_i. Cette transformation se fait par exemple par $x_i = a_i + U \mod 1$, où $U \sim U[0,1]^p$.

Estimation des indices de sensibilité par Monte Carlo

Considérons un N-échantillon $\tilde{X}_{(N)} = (x_{k1}, \ldots, x_{kp})_{k=1..N}$ de réalisations des variables d'entrée $(X_1, \ldots, X_p)$. L'espérance de Y, $\mathrm{E}[Y] = f_0$, et sa variance, $\mathrm{V}(Y) = V$, sont estimées par :

$$\hat{f}_0 = \frac{1}{N}\sum_{k=1}^{N} f(x_{k1}, \ldots, x_{kp}), \qquad \text{et} \qquad \hat{V} = \frac{1}{N}\sum_{k=1}^{N} f^2(x_{k1}, \ldots, x_{kp}) - \hat{f}_0{}^2. \tag{8}$$

L'estimation des indices de sensibilité nécessite l'estimation d'espérance de variance conditionnelle. Nous présentons une technique d'estimation due à Sobol [17].
L'estimation des indices de sensibilité de premier ordre (4) consiste à estimer la quantité :

$$V_i = \mathrm{V}(\mathrm{E}[Y|X_i]) = \underbrace{\mathrm{E}[\mathrm{E}[Y|X_i]^2]}_{U_i} - \mathrm{E}[\mathrm{E}[Y|X_i]]^2 = U_i - \mathrm{E}[Y]^2,$$

la variance de Y étant estimée classiquement par (8). Sobol propose d'estimer la quantité U_i, c'est-à-dire l'espérance du carré de l'espérance de Y conditionnellement à X_i, comme une espérance classique, mais en tenant compte du conditionement à X_i en faisant varier entre les deux appels à la fonction f toutes les variables sauf la variable X_i. Ceci nécessite deux échantillons de réalisations des variables d'entrée, que nous notons $\tilde{X}^{(1)}_{(N)}$ et $\tilde{X}^{(2)}_{(N)}$:

$$\hat{U}_i = \frac{1}{N}\sum_{k=1}^{N} f\left(x^{(1)}_{k1}, \ldots, x^{(1)}_{k(i-1)}, x^{(1)}_{ki}, x^{(1)}_{k(i+1)}, \ldots, x^{(1)}_{kp}\right) f\left(x^{(2)}_{k1}, \ldots, x^{(2)}_{k(i-1)}, x^{(1)}_{ki}, x^{(2)}_{k(i+1)} \ldots, x^{(2)}_{kp}, \right).$$

Les indices de sensibilité de premier ordre sont alors estimés par :

$$\hat{S}_i = \frac{\hat{V}_i}{\hat{V}} = \frac{\hat{U}_i - \hat{f}_0{}^2}{\hat{V}}.$$

[1] «pseudo» puisqu'elle consiste en un échantillonnage non totalement aléatoire

Pour les indices de sensibilité de second ordre $S_{ij} = \frac{V_{ij}}{V}$, où :

$$V_{ij} = \mathrm{V}(\mathrm{E}[Y|X_i, X_j]) - V_i - V_j = U_{ij} - \mathrm{E}[Y]^2 - V_i - V_j,$$

nous estimons les quantités $U_{ij} = \mathrm{E}[\mathrm{E}[Y|X_i, X_j]^2]$ de la même manière, en faisant varier entre les deux appels à la fonction toutes les variables sauf X_i et X_j :

$$\begin{aligned}\hat{U}_{ij} = \frac{1}{N}\sum_{k=1}^{N} & f\left(x_{k1}^{(1)}, \ldots, x_{k(i-1)}^{(1)}, x_{ki}^{(1)}, x_{k(i+1)}^{(1)}, \ldots, x_{k(j-1)}^{(1)}, x_{kj}^{(1)}, x_{k(j+1)}^{(1)}, \ldots, x_{kp}^{(1)}\right) \\ \times & f\left(x_{k1}^{(2)}, \ldots, x_{k(i-1)}^{(2)}, x_{ki}^{(1)}, x_{k(i+1)}^{(2)}, \ldots, x_{k(j-1)}^{(2)}, x_{kj}^{(1)}, x_{k(j+1)}^{(2)}, \ldots, x_{kp}^{(2)}\right).\end{aligned}$$

L'indice S_{ij} est alors estimé par :

$$\hat{S}_{ij} = \frac{\hat{U}_{ij} - \hat{f}_0^{\,2} - \hat{V}_i - \hat{V}_j}{\hat{V}}.$$

Et ainsi de suite pour les indices de sensibilité d'ordre supérieur.

Remarque. *L'estimation des indices de sensibilité d'ordre i, ($1 < i \leq p$), nécessite l'estimation des indices de sensibilité d'ordre 1 à $i-1$.*

Par contre, les indices de sensibilité totaux peuvent être estimés directement. En effet, on remarque facilement que l'indice de sensibilité total peut s'écrire

$$S_{T_i} = 1 - \frac{\mathrm{V}(\mathrm{E}[Y|X_{\sim i}])}{\mathrm{V}(Y)} = 1 - \frac{V_{\sim i}}{V}.$$

où $V_{\sim i}$ est la variance de l'espérance de Y conditionnellement à toutes les variables sauf X_i. $V_{\sim i}$ est alors estimée comme V_i, sauf qu'au lieu de faire varier toutes les variables sauf X_i, nous ne faisons varier uniquement X_i.
Ainsi, pour estimer $V_{\sim i} = \mathrm{E}[\mathrm{E}[Y|X_{\sim i}]^2] - \mathrm{E}[\mathrm{E}[Y|X_{\sim i}]]^2 = U_{\sim i} - \mathrm{E}[Y]^2$, on estime $U_{\sim i}$ par :

$$\hat{U}_{\sim i} = \frac{1}{N}\sum_{k=1}^{N} f\left(x_{k1}^{(1)}, \ldots, x_{k(i-1)}^{(1)}, x_{ki}^{(1)}, x_{k(i+1)}^{(1)}, \ldots, x_{kp}^{(1)}\right) f\left(x_{k1}^{(1)}, \ldots, x_{k(i-1)}^{(1)}, x_{ki}^{(2)}, x_{k(i+1)}^{(1)} \ldots, x_{kp}^{(1)},\right),$$

et on obtient

$$\hat{S}_{T_i} = 1 - \frac{\hat{U}_{\sim i} - \hat{f}_0^{\,2}}{\hat{V}}.$$

Quels indices estimer : stratégie à adopter En utilisant une taille d'échantillon de Monte Carlo de N, le nombre réel de simulations des variables d'entrée nécessaires à l'estimation des indices de sensibilité est $2N$, puisque cette estimation nécessite deux jeux de simulations. Le nombre d'appels à la fonction du modèle est alors $N \times (k+1)$, où k est le nombre d'indices estimés. Pour un modèle à p variables d'entrée, l'estimation de tous les indices de sensibilité nécessite $N \times (2^p)$ appels à la fonction. En revanche, n'estimer que les indices de premier ordre et les indices totaux ne demande que $N \times (2p+1)$ appels.
Il conviendra donc d'estimer dans un premier temps les indices de permier ordre et les indices totaux. S'il existe des écarts importants entre ces deux indices, c'est que la part des interactions est non négligeable et il peut être utile d'estimer les indices d'ordres intermédiaires. Dans le cas contraire, l'effet des variables d'entrée sera principalement de premier ordre et il ne sera pas utile de s'intéresser aux indices d'ordres intermédiaires.
En pratique, une taille d'échantillon de l'ordre de 10000 sera suffisante pour estimer les indices de sensibilité d'un modèle comportant une dizaine de variables d'entrée. En outre, il sera possible d'estimer la variabilité des estimateurs obtenus par boostrap. Lorsque le modèle demande un temps d'exécution important, il est illusoire de vouloir utiliser de telle taille d'échantillon en un temps raisonnable. On a en général recourt à une approximation de la fonction f (surface de réponse), permettant de faire des

simulations intensives et donc d'estimer les indices de sensibilité. Le lecteur interessé par une revue des méthodes de surface de réponse pour l'analyse de sensibilité pourra se référer à [3] par exemple.

2.3.1 La méthode de McKay

La méthode d'estimation des indices de sensibilité de premier ordre proposée par McKay, [6], se base sur l'échantillonnage par hypercube latin répliqué (*r-LHSampling*). À partir d'un N-échantillon créé selon le plan d'échantillonnage par hypercube latin (N premières lignes de la matrice ci-dessous), on crée r réplications (paquet de N lignes) en permutant indépendamment et aléatoirement les N valeurs de chaque variable (i.e. colonne). La réunion de ces r réplications donnera $N \times r$ échantillons pour chaque variable. Ce schéma d'échantillonnage par hypercube latin répliqué peut être représenté par la figure 1.

$$
\underset{r \text{ réplications}_k \times N \text{ valeurs}_j}{} \quad
\begin{bmatrix}
\begin{pmatrix} x_{11}^1 \\ \vdots \\ x_{N1}^1 \end{pmatrix} & \begin{pmatrix} \vdots \\ \vdots \\ \vdots \end{pmatrix} \\
\begin{pmatrix} x_{1k}^1 \\ \vdots \\ x_{jk}^1 \\ \vdots \\ x_{Nk}^1 \end{pmatrix} & N \left\{ \begin{pmatrix} \vdots \\ \vdots \\ x_{jk}^i \\ \vdots \\ \vdots \end{pmatrix} \right. \\
\begin{pmatrix} x_{1r}^1 \\ \vdots \\ x_{Nr}^1 \end{pmatrix} & \begin{pmatrix} \vdots \\ \vdots \\ \vdots \end{pmatrix}
\end{bmatrix}
\begin{matrix} \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{matrix}
\Longrightarrow
\begin{bmatrix} \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{bmatrix}
\begin{matrix} \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{matrix}
\begin{matrix} \\ \\ \\ \\ y_{jk} \\ \\ \\ \\ \\ \end{matrix}
$$

$\text{entrées}^1 \qquad\qquad \text{entrées}^i$

$1 \leq j \leq N$: N valeurs des variables d'entrée (prises dans des intervalles équiprobables),
$1 \leq k \leq r$: r permutations des N-vecteurs de simulations des entrées,
$1 \leq i \leq p$: p paramètres.

FIG. 1 – Échantillonnage par hypercube latin répliqué.

Les moyennes suivantes sont alors définies :

$$\overline{y}_{j.} = \frac{1}{r} \sum_{k=1}^{r} y_{jk} \qquad \overline{y} = \frac{1}{N} \sum_{j=1}^{N} \overline{y}_{j.},$$

où $\overline{y_{j.}}$ est la moyenne *inter* réplications et $\overline{y}$ est la moyenne sur toutes les valeurs de y.

L'estimation de l'indice de sensibilité de premier ordre de la variable X_i, défini par (4) nécessite l'esti-

mation des quantités $\mathrm{V}(\mathrm{E}[Y|X_i])$ et $\mathrm{V}(Y)$. La variance totale $\mathrm{V}(Y)$ peut être estimée par :

$$\widehat{\mathrm{V}^{(*)}(Y)} = \frac{1}{r}\sum_{k=1}^{r}\underbrace{\frac{1}{N}\sum_{j=1}^{N}(y_{jk}-\overline{y}_{.k})^2}_{\widehat{\mathrm{V}_k(Y)}}, \tag{9}$$

où $\overline{y}_{.k} = \dfrac{1}{N}\sum_{j=1}^{N} y_{jk}$ et $\widehat{\mathrm{V}_k(Y)}$ sont les estimations de la moyenne et de la variance de Y au sein de la réplication k (*intra* réplications). En utilisant la formule classique de l'analyse de la variance, pour une somme de carrés *intra* et *inter* réplications, qui s'écrit :

$$\begin{aligned}\sum_{k=1}^{r}\sum_{j=1}^{N}(y_{jk}-\overline{y})^2 &= \underbrace{\sum_{k=1}^{r}\sum_{j=1}^{N}(\overline{y}_{.k}-\overline{y})^2}_{\textit{inter}} + \underbrace{\sum_{k=1}^{r}\sum_{j=1}^{N}(y_{jk}-\overline{y}_{.k})^2}_{\textit{intra}} \\ &= N\sum_{k=1}^{r}(\overline{y}_{.k}-\overline{y})^2 + \sum_{k=1}^{r}\sum_{j=1}^{N}(y_{jk}-\overline{y}_{.k})^2,\end{aligned}$$

on a :

$$\widehat{\mathrm{V}^{(*)}(Y)} = \frac{1}{Nr}\sum_{k=1}^{r}\sum_{j=1}^{N}(y_{jk}-\overline{y})^2 - \frac{1}{r}\sum_{k=1}^{r}(\overline{y}_{.k}-\overline{y})^2.$$

Or, pour un échantillonnage *LHS*, comme $\mathrm{E}[(\overline{y}_{.k}-\overline{y})^2]$ est en $\frac{1}{N}$, le dernier terme de cette égalité peut être considéré comme négligeable pour une taille d'échantillon N suffisamment grande. McKay propose alors l'estimation de la variance totale suivante :

$$\widehat{\mathrm{V}(Y)} = \frac{1}{Nr}\sum_{j=1}^{N}\sum_{k=1}^{r}(y_{jk}-\overline{y})^2.$$

Soient $\overline{Y}_{j.}$ et $\overline{Y}$ les variables aléatoires dont $\overline{y}_{j.}$ et $\overline{y}$ sont les réalisations sur notre matrice d'échantillonnage. Comme :

$$\mathrm{E}\left[(\overline{Y}_{j.}-\overline{Y})^2\right] \simeq \mathrm{V}(\overline{Y}_{j.}) = \mathrm{V}(\mathrm{E}[\overline{Y}_{j.}|X_i]) + \mathrm{E}[\mathrm{V}(\overline{Y}_{j.}|X_i)] = \mathrm{V}(\mathrm{E}[Y|X_i]) + \frac{1}{r}\mathrm{E}[\mathrm{V}(Y|X_i)],$$

le terme $\mathrm{V}(E[Y|X_i])$ est estimé par :

$$\frac{1}{N}\sum_{j=1}^{N}(\overline{y}_{j.}^{(i)}-\overline{y})^2 - \frac{1}{r}\frac{1}{Nr}\sum_{j=1}^{N}\sum_{k=1}^{r}(y_{jk}^{(i)}-\overline{y}_{j.})^2,$$

où $\dfrac{1}{Nr}\sum_{j=1}^{N}\sum_{k=1}^{r}(y_{jk}^{(i)}-\overline{y}_{j.})^2$ est l'estimateur de $\mathrm{E}[\mathrm{V}(Y|X_i)]$, avec $y_{jk}^{(i)}$ et $\overline{y}_{j.}^{(i)}$ obtenus en fixant, pour la variable X_i, les r réplications, (x_{jk}^i constant sur k, c'est-à-dire $x_{j1}^i = x_{j2}^i = ... = x_{jr}^i$ pour tout $1 \leq j \leq N$).
L'indice de sensibilité de premier ordre de la variable X_i, défini par (4) est alors estimé par :

$$S_i = \frac{r\sum_{j=1}^{N}(\overline{y}_{j.}^{(i)}-\overline{y})^2 - \frac{1}{r}\sum_{j=1}^{N}\sum_{k=1}^{r}(y_{jk}^{(i)}-\overline{y}_{j.})^2}{\sum_{j=1}^{N}\sum_{k=1}^{r}(y_{jk}-\overline{y})^2}.$$

3 Modèles à entrées dépendantes

L'hypothèse de l'indépendance des facteurs d'entrée faite précédemment est nécessaire pour garantir l'interprétabilité des indices (un indice d'ordre un n'exprime plus la sensibilité à une unique variable si cette dernière est corrélée avec d'autres) et la validité de leur méthode d'estimation par Monte-Carlo (les intégrales multidimensionnelles sont évaluées comme des produits d'intégrales unidimensionnelles).
Nous présentons dans cette section les stratégies possibles pour réaliser une analyse de sensibilité sur un modèle à variables d'entrée non indépendantes.

3.1 Indices multidimensionnels

Lorsque toutes les variables d'entrée ne sont pas dépendantes, mais qu'elles peuvent être regroupées en clusters de variables dépendantes (les variables au sein d'un cluster sont dépendantes mais les variables de différents clusters sont indépendantes), il est possible de considérer des indices de sensibilité mutlidimensionnels [5] qui expriment la sensibilité de la variance de Y à un cluster de facteurs.
Si par exemple les deux variables X_i et X_j sont dépendantes, mais indépendantes du reste des autres variables, la sensibilité à la variable bidimensionnelle (X_i, X_j) sera exprimé par l'indice multidimensionnel

$$S_{\{i,j\}} = \frac{V(E[Y|X_i, X_j])}{V(Y)}.$$

Il est possible de définir des indices d'ordre supérieur exprimant la sensibilité de Y à l'interaction entre cette variable bidimensionnelle (X_i, X_j) et d'autres variables uni ou multidimensionnelles. Les clusters de variables étant indépendants entre eux, l'interprétabilité (et en particulier la sommation des indices de tout ordre à 1) est conservée.
L'estimation de ces indices peut être faite par Monte Carlo avec une approche similaire à celle utilisée pour estimer les indices de sensibilité de Sobol classiques (cf. [5] pour plus de détails).

3.2 Utilisation des indices de Sobol d'ordre 1

Lorsque l'analyse de sensibilité est menée dans le but de savoir quelle variable ou quel groupe de variables qui, une fois fixé, conduit à la plus grande réduction de la variance de Y, Saltelli et Tarantola [15] expliquent que les indices de sensibilité d'ordre un sont toujours les indicateurs à utiliser en présence de corrélation. En effet, si en présence de corrélation l'indice d'ordre un S_i n'exprime plus uniquement la sensibilité à une variable X_i mais également une partie de sensibilité aux variables avec lesquelles elle est corrélée, fixer X_i conduit à également jouer sur la distribution des variables avec lesquelles elle est corrélée, et donc conduit à réduire d'autant plus la variance de la réponse du modèle.
Si l'estimation de Monte-Carlo des indices de premier ordre présentée précédemment (section 2.3) n'est plus valable en l'absence d'indépendance entre les variables d'entrée, la méthode de McKay (section 2.3.1) est toujours valable. Néanmoins, cette méthode d'estimation est très gourmande en nombre d'évaluations de la fonction f du modèle, ce qui peut être problématique lorsque l'évaluation de f est coûteuse en temps de calcul. Nous présentons ci-après une méthode d'estimation par polynômes locaux réduisant considérablement ce nombre d'évaluations [1].

Estimation par polynômes locaux La méthode d'estimation des indices de sensibilité d'ordre 1 de Da Veiga [1] consiste à estimer dans un premier temps l'espérance de Y conditionnellement à chaque variable d'entrée X_i, puis dans un second temps à estimer la variance de cette espérance conditionnelle pour obtenir l'estimateur de l'indice de sensibilité. L'avantage principal de cette méthode est qu'elle ne fait appel qu'à un nombre réduit d'appel à la fonction, contrairement à la méthode de McKay précédente.
Notons $m_i(x) = E[Y|X_i = x]$. On approche $m(x)$ localement par un polynôme

$$m_i(z) \simeq \sum_{j=0}^{p} \beta_j (z - x)^j \qquad \forall z \in \mathcal{V}(x)$$

où $\mathcal{V}(x)$ un voisinage de x, symbolisé par une fonction noyau K (de paramètre d'échelle h) pondérant l'estimation par moindres carrés :

$$\beta = argmin_\beta \sum_{j=1}^{n} \left(Y^j - \sum_{j=0}^{p} \beta_j (X_i^j - x)^j \right)^2 K\left(\frac{X_i^j - x}{h} \right),$$

avec $(X_i^j, Y^j)_{j=1,n}$ un échantillon de réalisations du couple (X_i, Y). Utilisant un second échantillon $(\tilde{X}_i^j)_{j=1,n'}$ de réalisations de la variable X_i, indépendant du premier, on peut estimer classiquement la variance de $m_i(x)$ par :

$$U_i = \frac{1}{n'-1} \sum_{j=1}^{n'} (m_i(\tilde{X}_i^j) - \bar{m_i})^2$$

où $\bar{m_i} = \sum_{j=1}^{n'} m_i(\tilde{X}_i^j)/n'$. Il suffit alors de diviser par l'estimation de la variance de Y pour obtenir une estimation de l'indice de sensibilité d'ordre un S_i.

4 Outil logiciel sous R et illustrations numériques

Dans cette section, après avoir précisé les packages **R** permettant des réaliser des analyses de sensibilité, nous présentons deux analyses de sensibilité de modèles simulés, dans le cas d'entrées indépendantes puis non indépendantes.

4.1 Outil logiciel sous R

Package `sensitivity` Le package `sensitivity` [11] du logiciel **R**, disponible sur le site du CRAN[2] permet de calculer les indices de sensibilité de Sobol présentés dans ce document, lorsque les variables d'entrée sont indépendantes.
La fonction `sobol` permet de calculer les indices de tout ordre, tandis que la fonction `sobol2002` permet d'estimer les indices de premier ordre et d'ordre total à partir d'un nombre d'échantillons plus réduit que la fonction `sobol`. Ces deux fonctions retournent des intervalles de confiance estimés par bootstrap.

Package `sensitivity-dependent` Un package `sensitivity-dependent` pour le logiciel **R**, disponible sur le site de l'auteur[3], permet de calculer les indices de sensibilité multidimensionnels (fonction `sobol_multi`) et les indices de sensibilité de premier ordre par la méthode de McKay (fonction `mckay`) en présence de variables d'entrée dépendantes. La fonction `sobol_multi` fournit en outre une estimation de la variabilité des estimations par boostrap.

4.2 Illustration de modèles à entrées indépendantes

Nous considérons trois modèles à entrées indépendantes :
- le modèle linéaire $Y = 9X_1 + 6X_2 + 3X_3 + X_4$ avec $X_i \sim \mathcal{U}[0,1]$,
- le benchmark d'Ishigami [12] :$Y = \sin(X_1) + 7\sin^2(X_2) + \frac{X_3^4}{10}\sin(X_1)$ où $X_i \sim \mathcal{U}[-\pi, \pi]$ pour $i = 1, 2, 3$,
- le benchmark de Sobol [12] : $Y = \prod_{j=1}^{8} \frac{|4*X[,j]-2|+a[j]}{1+a[j]}$ avec $a = [0, 1, 4.5, 9, 99, 99, 99, 99]$ et $X_i \sim \mathcal{U}[0,1]$.

Les indices de sensibilité de premier ordre et totaux, estimés à l'aide de la fonction `sobol2002` du package `sensitivity`, sont donnés dans la table 1. La taille d'échantillon utilisée est 10000, et les intervalles de confiance sont obtenus par 100 réplications bootstrap.

[2] http ://cran.r-project.org/web/packages/sensitivity/
[3] http ://labomath.univ-lille1.fr/∼jacques/

variable	indice ordre 1	intervalle de confiance	indice total	intervalle de confiance
modèle linéaire				
X_1	0.593	[0.523,0.661]	0.684	[0.603,0.753]
X_2	0.252	[0.200,0.298]	0.300	[0.250,0.345]
X_3	0.071	[0.055,0.097]	0.068	[0.045,0.089]
X_4	0.009	[0,0.019]	0.007	[0,0.016]
Ishigami benchmark				
X_1	0.305	[0.269,0.338]	0.578	[0.546,0.607]
X_2	0.4356	[0.403,0.462]	0.428	[0.402,0.452]
X_3	$\simeq 0$	[0,0.011]	0.254	[0.232,0.282]
Sobol benchmark				
X_1	0.759	[0.701,0.813]	0.769	[0.718,0.807]
X_2	0.146	[0.123,0.175]	0.290	[0.258,0.316]
X_3	0.025	[0.017,0.034]	0.038	[0.022,0.053]
X_4	0.003	[0,0.010]	0.019	[0.010,0.029]
X_5 à X_8	$\simeq 0$	$\simeq 0$	$\simeq 0$	$\simeq 0$

TAB. 1 – Indices de sensibilité de premier ordre et totaux pour les modèles linéaire, d'Ishigami et de Sobol.

Lecture des résultats Nous présentons ci-dessous un exemple de lecture des résultats pour le modèle d'Ishigami :

- La variable qui a le plus d'influence sur la variance de la sortie (au sens de l'indice total, c'est-à-dire en prenant en compte les interactions avec les autres variables), est la variable X_1, avec un indice total de 0.6 et près de 30% de la variance de Y expliquée à elle seule.
- La variable X_2 n'intervient que seule (indice d'ordre un équivalent à indice total), en expliquant près de 40% de la variance de Y.
- La variable X_3 n'a aucune influence seule, mais a une influence relativement importante en interaction (avec X_1), avec un indice total d'environ 0.3.
- On en déduit que la variance de Y est due pour 40% à X_2, 30% à X_1 et 30% à l'interaction entre X_1 et X_3.

Notons également que dans cet exemple, l'interaction entre X_1 et X_3 est due à une relation non additive entre ces deux variables dans l'expression du modèle.

Interprétation des résultats Afin d'intepréter les valeurs des indices de sensibilité, nous fixons tour à tour chaque variable du modèle à son espérance et examinons l'impact que cela a sur la distribution de la sortie Y. La figure 2 présente sous la forme de boîte à moustaches les résultats obtenus, pour les trois modèles linéaire, d'Ishigami et de Sobol (de gauche à droite). Sur chaque graphique, la première boîte (à gauche) correspond à la distribution initiale de Y. Les boîtes suivantes correspondent aux distributions de Y lorsque les variables sont fixées une à une, en les ordonnant de gauche à droite selon leur ordre décroissant d'importance (au sens de l'indice de sensibilité total).

Comme attendu, la plus grande réduction de variance est obtenue en fixant la variable ayant l'indice de sensibilité total le plus important. Mais il faut avoir à l'esprit que modifier la distribution d'une variable d'entrée (pour réduire sa variance) n'agît pas uniquement sur la variance de Y : en effet, sauf dans le cas d'un modèle linéaire, une modification de la distribution des entrées influe également la position centrale de la distribution. Il conviendra donc de s'assurer avant de modifier la distribution d'une variable d'entrée dans le but d'améliorer le pouvoir prédictif du modèle, que celle-ci est bien justifiée et réaliste.

4.3 Illustration de modèles à entrées dépendantes

Dans cette seconde illustration numérique, nous considérons le modèle :

$$Y = X_1 + X_2 X_3 + X_4^2,$$

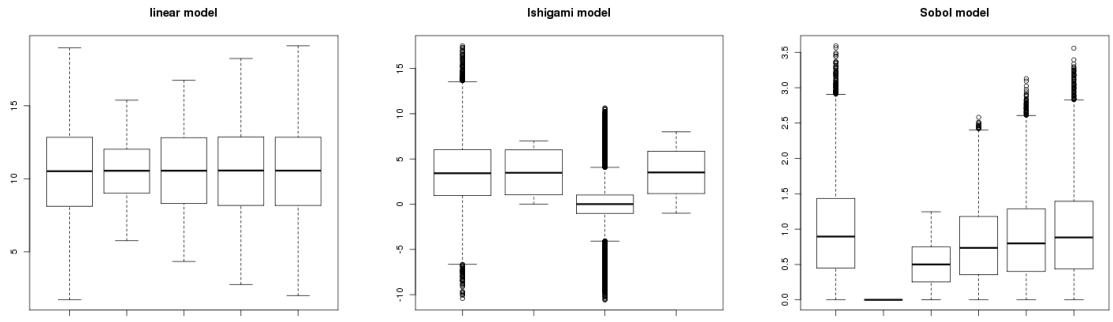

FIG. 2 – Distributions de Y en fonction de la variable d'entrée fixée, pour les trois modèles linéaire, d'Ishigami et de Sobol.

où $(X_1, X_2, X_3, X_4)^t$ est un vecteur gaussien d'espérance $(1, 1, 1, 1)^t$ et de matrice de variance

$$\Sigma = \begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 4 & 2 \\ 0 & 0 & 2 & 3 \end{pmatrix}.$$

Les indices de sensibilités multidimensionnels de premier ordre et totaux sont estimés à l'aide de la fonction `sobol_multi` du package `sensitivity-dependent` avec une taille d'échantillon de 10000. Les résultats (estimation moyenne et écart-type sur 100 réplications bootstrap) sont donnés par la table 2 et la figure 3. La table 2 présente également les résultats d'estimation des indices d'ordre un par la méthode de McKay (20 réplications de l'échantillonnage LHS), obtenus par la fonction `mckay` du package `sensitivity-dependent`.

variable	indice ordre 1 (McKay)	indices multidimensionnels ordre 1	indices multidimensionnels total
X_1	0.073	0.061 (0.014)	0.069 (0.008)
X_2	0.060	0.049 (0.013)	0.298 (0.016)
X_3	0.078	0.634 (0.016)	0.882 (0.012)
X_4	0.526		

TAB. 2 – Indices de sensibilité de premier ordre et totaux du modèle d'Ishigami.

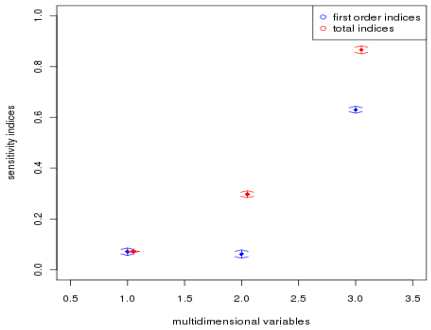

FIG. 3 – Indices de sensibilité multidimensionnels de premier ordre et totaux

Sur cet exemple, les indices d'ordre un estimés par McKay nous indiquent une importance prépondérante de la variable X_4, et aucune influence des autres variables (seules). Le calcul des indices multidimensionnels nous permet d'aller plus loin dans l'interprétation :

- la variable X_2 a également une influence significative grâce à des interactions avec d'autres variables,
- les variables X_3 et X_4 (et non pas uniquement X_4) expliquent certes à elles seules une grande partie de la variance de Y (environ 60%), mais également une autre partie non négligeable de cette variance à travers l'interaction avec la variable X_2.

5 Discussion

L'analyse de sensibilité globale a pour objectif de déterminer l'impact des variables d'entrée sur la variabilité de la sortie d'un modèle mathématique. Dans le cas de modèles à entrées indépendantes (cas le plus fréquement abordé dans la littérature mais pas forcément le plus répandu en pratique), les indices de sensibilité expriment la part de variance de la sortie due à chaque variable d'entrée. De nombreux travaux ont permis de développer des méthodes d'estimation efficaces et simples à mettre en oeuvre, que le praticien pourra intégrer sans problème à ses propres codes de calcul. Nous attirons néanmoins l'attention du praticien quant à l'interprétation et l'utilisation des résultats de sensibilité, comme nous l'avons illustré précédemment : modifier la variance de la variable d'entrée la plus influente pour diminuer l'incertitude de prédiction du modèle n'influe pas uniquement sur la variance de la sortie. L'hypothèse d'indépendance des entrées faite précédemment est primordiale et le praticien ne doit surtout pas s'aventurer à des analyses de sensibilité classiques lorsque son modèle ne respecte pas cette hypothèse. Dans une telle situation, il dispose soit des indices de sensibilité multidimensionnels lorsque les entrées ne sont pas toutes dépendantes entre elles, soit des indices de sensibilité d'ordre un classiques mais estimés par des méthodes spécifiques au cas d'entrées dépendantes : méthode de McKay, facile d'implémentation, ou de Da Veiga, plus complexe mais beaucoup moins gourmande en évaluations du modèle.

Références

[1] S. Da Veiga, F. Wahl, and F. Gamboa. Local polynomial estimation for sensitivity analysis on models with correlated inputs. *Technometrics*, 51(4) :452–463, 2009.

[2] T. Homma and A. Saltelli. Importance measures in global sensitivity analysis of non linear models. *Reliability Engineering and System Safety*, 52 :1–17, 1996.

[3] B. Iooss. Revue sur l'analyse de sensibilité globale de modèles numériques. *Journal de la Société Française de Statistique*, 152 :1–23, 2011.

[4] J. Jacques. *Contributions à l'analyse de sensibilité et à l'analyse discriminante généralisée*. Thèse de l'Université Joseph Fourier, 2005.

[5] J. Jacques, C. Lavergne, and N. Devictor. Sensitivity analysis in presence of model uncertainty and correlated inputs. *Reliability Engineering and System Safety*, 91 :1126–1134, 2006.

[6] M.D. McKay. Evaluating prediciton uncertainty. Technical Report NUREG/CR-6311, US Nuclear Regulatory Commission and Los Alamos National Laboratory, 1995.

[7] M.D. McKay, R. Beckman, and W. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2) :239–245, 1979.

[8] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*. Philadelphia : SIAM, 1992.

[9] A. Owen. *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, chapter Randomly Permuted (t,m,s)-Nets and (t,s)-Sequences. New York : Springer-Verlag, Niederreiter,H. and Shiue,P.J.-S. (Eds), 1995.

[10] A. Owen. Monte carlo extension of quasi-monte carlo. In *1998 Winter Simulation Conference*, Washington (DC, USA), 1998.

[11] G. Pujol and B. Iooss. Package sensitivity : Sensitivity analysis. Technical report, **R** software, 2008.

[12] A. Saltelli, K. Chan, and E.M. Scott, editors. *Sensitivity Analysis*. Wiley, 2000.

[13] A. Saltelli and E.M. Scott. Guest editorial : The role of sensitivity analysis in the corroboration of models and its link to model structural and parametric uncertainty. *Reliability Engineering and System Safety*, 1997.

[14] A. Saltelli and S. Tarantola. Sensitivity analysis : a prerequisite in model building ? *Foresight and Precaution*, 2000.

[15] A. Saltelli and S.. Tarantola. On the relative importance of input factors in mathematical models : safety assessment for nuclear waste disposal. *Journal of the American Statistical Association*, 97(459) :702–709, 2002.

[16] A. Saltelli, S. Tarantola, and F. Campolongo. Sensitivity analysis as an ingredient of modeling. *Statistical Science*, 15(4) :377–395, 2000.

[17] I.M. Sobol. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiments*, 1 :407–414, 1993.

[18] T. Turanyi. Sensitivity analysis of complex kinetic system, tools and applications. *Journal of Mathematical Chemistry*, 5 :203–248, 1990.