

Global 感度解析—Sobol' 法

本間 俊充

分散に基づくグローバル感度解析の代表的手法の一つとして、Sobol' 法について述べる。Sobol' 法の感度指標は、分散低減に基づくパラメータの“重要度指標”を一般化したもので、入力変数の不確かさに起因するモデル出力の不確かさ、すなわち分散に対する入力変数の寄与を表す。本稿では、その数学的定式化と実際の計算手順および適用例について解説する。

キーワード：グローバル感度解析，分散低減，Sobol' 法，総感度指標

1. 分散に基づく感度解析手法

モデル出力の感度解析の目的は、出力変数の値を決定する際の各入力変数（パラメータ）の相対的重要度を定量化することにある。グローバル感度解析では、入力変数がとり得るパラメータ空間（不確かさ）の全範囲にわたる出力の不確かさに着目する。各入力変数の不確かさの範囲は、我々の知識不足に起因する不確かさを代表する。ここで、感度解析はどの入力変数の不確かさが最も出力の不確かさに寄与するか、キーとなるパラメータを見いだす有効な手がかりを与える。したがって、研究の優先度決定に利用でき、最終的には未知のパラメータのよりよい定義や不確かさ範囲の低減に役立つ。

モンテカルロ法に基づく入力変数の不確かさ伝播解析では、モデル出力の不確かさの推定とともに、どの入力変数がモデル出力の不確かさ低減に最も寄与するかを見いだすための感度解析の指標として、いくつかの“不確かさ重要度指標 (uncertainty importance measures)”が提案された[1]~[3]。これは、着目するパラメータ x_i の不確かさ低減によるモデル出力の不確かさ低減を尺度とするもので、例えば以下のように定義される。

$$I_i = V[y] - E\{V[y|x_i]\} \quad (1)$$

ここで、 $V[y]$ はモデル出力 y の分散、 $E\{V[y|x_i]\}$ は、 x_i を固定したときの $V[y]$ の期待値を表す。期待値は x_i の不確かさ分布を考慮して計算される。実際、 I_i は x_i の不確かさが除かれた場合のモデル出力

の分散低減の期待値で、各パラメータがモデル出力の分散に及ぼす影響の重要度となる。

(1)式の右辺は、 $V[y] = V\{E[y|x_i]\} + E\{V[y|x_i]\}$ を利用すると $V\{E[y|x_i]\}$ と表すことができるので、これを推定される分散で除した

$$\frac{V\{E[y|x_i]\}}{V[y]} \quad (2)$$

を相関比 (correlation ratio) と呼んで、各パラメータの重要度指標として定義する場合もある[4][5]。この不確かさ重要度指標や相関比は、以下で解説するように Sobol' 法の1次の感度指標や本特集号の劉氏による別稿 FAST の感度指標と同等であることが分かる。

2. Sobol' 法の感度指標

Sobol' によって提案されたグローバル感度解析手法は、上記のモンテカルロ法による分散に基づく手法を一般化したものである[6]。ここで、モデル関数 $f(\mathbf{x}) = f(x_1, x_2, \dots, x_k)$ を k 次元の入力変数に関する単位超立方体空間 Ω^k で定義する。

$$\Omega^k = \{\mathbf{x} | 0 \leq x_i \leq 1; i=1, \dots, k\} \quad (3)$$

Sobol' の方法は、関数 $f(\mathbf{x})$ を以下のように、入力変数の次元の増加に従って分解した項の展開で考える。

$$\begin{aligned} f(x_1, \dots, x_k) \\ = f_0 + \sum_{i=1}^k f_i(x_i) + \sum_{1 \leq i < j \leq k} f_{i,j}(x_i, x_j) \\ + \dots + f_{1,2,\dots,k}(x_1, \dots, x_k) \end{aligned} \quad (4)$$

(4)式が成立するためには、 f_0 は定数で、すべての分解項は含まれるいかなる入力変数についての積分も0でなければならない。すなわち、

$$\int_0^1 f_{i_1, \dots, i_s}(x_{i_1}, \dots, x_{i_s}) dx_{i_k} = 0 \quad 1 \leq k \leq s \quad (5)$$

でなければならない。その結果、(4)式の各分解項は直

交性を持つことが分かる。すなわち、 $(i_1, \dots, i_s) \neq (j_1, \dots, j_s)$ のとき、以下の(6)式および(7)式が成立する。

$$\int_{\Omega^k} f_{i_1, \dots, i_s}(x_{i_1}, \dots, x_{i_s}) f_{j_1, \dots, j_s}(x_{j_1}, \dots, x_{j_s}) dx = 0 \quad (6)$$

$$f_0 = \int_{\Omega^k} f(x) dx \quad (7)$$

(4)式の分解項への展開は $f(x)$ が Ω^k で積分可能ならば一意に決まり、分解の各項は多重積分によって以下のように求まることが示された[6]。入力変数に関する1次の項 $f_i(x_i)$ は、(4)式を着目する入力変数 x_i 以外のすべての入力変数で積分すると、(5)式を用いて以下のように示される。

$$\int_0^1 \dots \int_0^1 f(x) dx_{-i} = f_0 + f_i(x_i) \quad (8)$$

ここで、 $\int \dots \int dx_{-i}$ は、 x_i を除くすべての入力変数についての積分を表す。同様に、2次の項 $f_{i,j}(x_i, x_j)$ も以下のように示される。高次の項も同様である。

$$\int_0^1 \dots \int_0^1 f(x) dx_{-ij} = f_0 + f_i(x_i) + f_j(x_j) + f_{i,j}(x_i, x_j) \quad (9)$$

Sobol'法の感度指標は分散に基づく指標で、以下のように総分散と(4)式の各項の偏分散を考える。

$$D = \int_{\Omega^k} f^2(x) dx - f_0^2 \quad (10)$$

$$D_{i_1, \dots, i_s} = \int_0^1 \dots \int_0^1 f_{i_1, \dots, i_s}^2(x_{i_1}, \dots, x_{i_s}) dx_{i_1} \dots dx_{i_s} \quad (11)$$

ここで、 $1 \leq i_1 < \dots < i_s \leq k, s=1, \dots, k$ である。(4)式を2乗して空間 Ω^k について積分し、(6)式の直交性を用いると(4)式と同様に総分散は次元の増加に従った偏分散の和に分解でき、以下が得られる。

$$D = \sum_{i=1}^k D_i + \sum_{1 \leq i < j \leq k} D_{i,j} + \dots + D_{1,2, \dots, k} \quad (12)$$

ここから、総分散 D に対する入力変数 x_{i_1}, \dots, x_{i_s} の寄与を表す感度指標として、Sobol'は以下を定義した。

$$S_{i_1, \dots, i_s} = \frac{D_{i_1, \dots, i_s}}{D}, \quad i \leq i_1 < \dots < i_s \leq k \quad (13)$$

ここで、 S_i は入力変数 x_i に対する1次オーダーの感度指標であり、 $f(x)$ の分散に対する x_i の寄与として主効果 (main effect) とも呼ばれる。 $S_{i,j}$ ($i \neq j$) は2次オーダーの感度指標で、 x_i と x_j の各々単独の効果の和で説明できない2因子の交互効果 (interaction effect) を表す。同様に、高次オーダーの感度指標を考えると、感度指標の総和は以下のように1となる。

$$\sum_{i=1}^k S_i + \sum_{1 \leq i < j \leq k} S_{i,j} + \dots + S_{1,2, \dots, k} = 1 \quad (14)$$

なお、Sobol'法は分散分析 (ANOVA) を(4)式のHDMR (High Dimensional Model Representation) に適用して得られる展開で、ANOVA-HDMR と呼ばれるものの一例である[7][8]。詳しくは、本特集号の香田氏の別載を参照されたい。

3. 総感度指標

(13)式で定義した Sobol'法の1次オーダーの感度指標は、本質的に本特集号の劉氏による別稿のFAST手法による指標と同等である。それを Sobol'は高次の感度指標まで一般化した。以下4節の計算手順で示すとおり、高次の項をすべて求めるには計算量が膨大である。そこで、着目する入力変数が含まれるすべての次元の感度指標の和として定義される総感度指標 (total sensitivity indices) S_T という非常に便利な感度指標が導入された[9]。例えば、対象とするモデルに3つの入力変数が存在するとする。入力変数1がモデル出力の総分散に及ぼす寄与の合計を S_{T1} とすると、 $S_{T1} = S_1 + S_{1,2} + S_{1,3} + S_{1,2,3}$ で与えられる。

総感度指標の導出では、まず関与する k 個のすべての入力変数を2つのグループに分割する。例えば、入力変数 x を u と v 、すなわち $u = (x_1, \dots, x_t), v = (x_{k-t+1}, \dots, x_k)$ に分割すれば、 $f(x)$ のANOVA分解 (decomposition) は以下のように与えられる。

$$f(x) = f_0 + f_u(u) + f_v(v) + f_{u,v}(u, v) \quad (15)$$

ここで、 $f_u(u)$ は入力変数 u のみに依存する項すべての和、 $f_v(v)$ は入力変数 v のみに依存する項すべての和で、 $f_{u,v}(u, v)$ は残りの項を表す。ANOVAの定義から $\int f_u du = \int f_v dv = \int f_{u,v} du = \int f_{u,v} dv = 0$ および、

$$D_u = \int f_u^2 du, D_v = \int f_v^2 dv, D_{u,v} = \int f_{u,v}^2 du dv \quad (15)$$

式を2乗し $du dv$ で積分すると、総分散に対する以下が得られる。

$$D = D_u + D_v + D_{u,v} \quad (16)$$

この式を用い、入力変数 u に依存するすべての分散 D_{Tu} は

$$D_{Tu} = D_u + D_{u,v} = D - D_v \quad (17)$$

と表すことができ、入力変数 u に対する総感度指標は以下のように定義できる。

$$S_{Tu} = D_{Tu} / D \quad (18)$$

したがって、入力変数 x_i に対する総感度指標は、 $u = (x_i), v = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k)$ として、以下のように表すことができる。

$$S_{T_i} = \frac{D - D_{\sim i}}{D} = 1 - S_{\sim i} \quad (19)$$

ここで、 $S_{\sim i}$ は i を含まないすべての S_{i_1, \dots, i_s} 項の和である。上記の3つの入力変数の場合、入力変数1の総感度指標 S_{T_1} は、 $S_{T_1} = 1 - (S_2 + S_3 + S_{2,3})$ となる。

総感度指標は入力変数 x_i に対する高次の個別の感度を表すわけではないが、1次の感度指標に比べ、次節で示すように同じ計算量ですべての高次の交互作用を含んだ影響を評価できる点で優れている[9]。

4. 感度指標の計算手順

Sobol'法の感度指標は、モデル関数 $f(\mathbf{x})$ のANOVA分解である(4)式が未知でも、(10)および(11)式の総分散および偏分散が計算できれば評価ができる点で簡便である。(7)、(10)および(11)式の積分はモンテカルロ積分で実行できる。 N 組の入力変数 \mathbf{x} を空間 Ω^k で生成し、以下の(20)および(21)式で総分散が推定できる。

$$\hat{f}_0 = \frac{1}{N} \sum_{m=1}^N f(\mathbf{x}_m) \quad (20)$$

$$\hat{D} = \frac{1}{N} \sum_{m=1}^N f^2(\mathbf{x}_m) - \hat{f}_0^2 \quad (21)$$

(11)式の偏分散の計算では、各偏分散の計算ごとにモンテカルロサンプルを生成する必要はない。空間 Ω^k での m 番目のサンプルを \mathbf{x}_m とし、

$$\mathbf{x}_{(\sim i)m} = (x_{1m}, x_{2m}, \dots, x_{(i-1)m}, x_{(i+1)m}, \dots, x_{km}) \quad (22)$$

とすると、添字(1)および(2)で示す2つのモンテカルロサンプルを生成し、以下の式で1次の感度指標の計算に必要な偏分散が計算できる。

$$\hat{D}_i = \frac{1}{N} \sum_{m=1}^N f(\mathbf{x}_{(\sim i)m}^{(1)}, x_{im}^{(1)}) f(\mathbf{x}_{(\sim i)m}^{(2)}, x_{im}^{(2)}) - \hat{f}_0^2 \quad (23)$$

式(23)から分かるように、偏分散 \hat{D}_i の計算では添字(1)の行列で表される点に対応するモデル関数 $f(\mathbf{x})$ の値に、添字(2)の別の行列で着目する入力変数 x_i の列の値だけを添字(1)と同一にしたモデル関数 $f(\mathbf{x})$ の値を乗ずる。したがって、入力変数 x_i が寄与の大きな因子であれば、関数値 f の大きな値同士を乗じた結果が偏分散の値となる。一方、入力変数 x_i の寄与が重要でない場合は、関数値 f は大小の値をランダムにとり、結果として偏分散は小さな値となる。

(11)式で定義される高次の感度指標の計算手順については文献[9]を参照されたい。Sobol'法の感度指標の計算は、上記のようにモデル関数を直接計算する簡便さはあるが、(10)式の f_0 を含めすべての感度指標を計算するには 2^k のモンテカルロ積分が必要になる。これは、例えば k の値が小さくとも膨大な計算となる。 k

個あるすべての1次感度指標の計算に、 $N(k+1)$ のモデル関数の計算が必要となる。高次の交互効果を含む総感度指標(19)式の計算は、(23)式のモンテカルロ積分と同様に、以下で推定できる。

$$\hat{D}_{\sim i} = \frac{1}{N} \sum_{m=1}^N f(\mathbf{x}_{(\sim i)m}^{(1)}, x_{im}^{(1)}) f(\mathbf{x}_{(\sim i)m}^{(2)}, x_{im}^{(2)}) - \hat{f}_0^2 \quad (24)$$

(24)式の計算は、 f_0 の計算も含め同様に $N(k+1)$ のモデル関数の計算で可能である。

Sobol'は感度指標の計算に必要なモンテカルロ積分に、擬似乱数ではなく自身が提案した準乱数を用いる準モンテカルロ法を推奨した[4][10]。Sobol'の準乱数 LP_r 列は、超一様分布列 (low-discrepancy sequences) と呼ばれ定積分を高速に求めるために工夫された特殊な無限列の一つである[11]。Sobol'の準乱数 LP_r 列の有効性については文献[12]を参照されたい。

5. 適用事例

本節では、モデル独立なSobol'法の感度指標の有効性を、良く用いられる回帰係数との比較で見てみる。ここに示すモデルは、放射性廃棄物の地層処分に伴う公衆の被ばく線量評価に用いられるものである。対象とする複雑なシステムを非常に理想化し、単純化したものに置き換えているが、その本質は変わらない[13]。

地層中の処分施設から放射性核種が長期間を経て放出され、地層中では吸着平衡が成立し、放射性核種は移流拡散によって生態圏すなわち地下水に移行する。その地下水を人が飲むことで放射性核種を体内に取り込み内部被ばくを受けるというシナリオである。処分施設からの放射性核種の漏洩は、固化体から一定時間 T 後に浸出率 k で起こるとする。固化体中の放射性核種の存在量は以下で与えられる。

$$\frac{dM}{dt} = -\lambda M + kM, \quad t \geq T \quad (25)$$

放射性核種の初期存在量 $M(0) = M_0$ とする。固化体からの漏洩フラックスは、以下で与えられる。

$$S(t) = kM, \quad t \geq T \quad (26)$$

地層を移行中の放射性核種のフラックス $F(x, t)$ は、移流拡散、吸着平衡、放射性崩壊のメカニズムによる以下の式で与えられる。

$$R \frac{\partial F}{\partial t} + v \frac{\partial F}{\partial x} - dv \frac{\partial^2 F}{\partial x^2} = -\lambda RF \quad (27)$$

ここで、 R 、 v 、 d 、 λ は各々遅延係数、流速、分散長、崩壊定数を表す。初期条件

$$F(x, 0) = 0 \quad (28)$$

および、境界条件を

$$F(0, t) = S(t) \quad (29)$$

$$F(\infty, t) = 0 \quad (30)$$

として(27)式を解くと、地下水へ移行した放射性核種の人の摂取による線量 $D(t)$ は以下の解析解で与えられる。

$$D(T+t) = \frac{1}{2} \beta \frac{W}{W} k M_0 e^{-\lambda(T+t)} e^{1/2d} e^{-Rl^2/4dvt} e^{-vt/4dR} \times [\phi(\gamma) + \phi(\gamma')] \quad (31)$$

$$\gamma = \left(\frac{Rl^2}{4dvt} \right)^{1/2} + \left(\frac{vt}{4dR} - kt \right)^{1/2} \quad (32)$$

$$\gamma' = \left(\frac{Rl^2}{4dvt} \right)^{1/2} - \left(\frac{vt}{4dR} - kt \right)^{1/2} \quad (33)$$

$$\phi(z) = e^{z^2} \text{erf}(z) \quad (34)$$

本事例では、放射性核種として¹²⁹Iを対象として6個の入力パラメータの不確かさを表1の確率分布で与えた。6個の入力パラメータの感度指標の計算には、Sobol'の準乱数列を用いて1,024のサンプルを生成した。(31)~(34)式のモデル出力である線量について、3時点における $2^6 - 1 = 63$ 個の S_{i_1, \dots, i_6} を計算した結果を表2に示す。表2から明らかのように、3時点すべてにおいて、主効果だけでなく2次および3次の高次のパラメータの組み合わせがモデル出力である線量の総

表1 テストケースにおける入力データ

記号	内容	分布	範囲	単位
v	地層中の流速	対数一様	$/10^{-3}, 10^{-1}/$	$m y^{-1}$
W	地下水流量	対数一様	$/10^5, 10^7/$	$m^3 y^{-1}$
R	ヨウ素の遅延係数	一様	$/1, 5/$	-
l	地層の長さ	一様	$/100, 500/$	m
T	密封時間	一様	$/100, 1000/$	y
k	ヨウ素の浸出率	対数一様	$/10^{-3}, 10^{-2}/$	y^{-1}

表2 3時点における線量に対する感度指標 $\hat{S}_{i_1, \dots, i_6}$ の値

	$t = 10^4 (y)$	$t = 8 \times 10^4 (y)$	$t = 7 \times 10^5 (y)$
vW	0.203	vWR 0.286	v 0.281
$vWRI$	0.197	vWI 0.231	vWR 0.154
l	0.131	v 0.119	vWI 0.129
vl	0.111	W 0.109	vRI 0.119
vRI	0.109	vW 0.102	RI 0.109
R	0.104		
v	0.100		

分散に寄与している。その寄与の和をとると総分散に対して80%以上の割合であり、高次の影響は明らかである。

本事例のモデルは、パラメータに対して特徴のある非線型性を有している。図1にはモンテカルロ法に基づくパラメータの不確かさ伝播解析で得られたモデル出力の実値および順位数に対する回帰モデルの決定係数を示した。決定係数は、標準回帰係数(SRC)あるいは標準順位回帰係数(SRRC)に基づく線型回帰モデルがどの程度モデル出力を再現するかの尺度となる。図1に示すように、実値による回帰モデルの小さな決定係数はモデルの非線型性を示している。また、 $t = 8 \times 10^4 (y)$ 付近では、順位数の回帰モデルの決定係数も小さく非線型、非単調性を示している。

表3には、3時点における各入力パラメータの総感度指標 S_{Ti} とSRRCを示した。SRRCの絶対値はパラメータの重要度を、正負は入出力の相関関係を示している。地層中の流速 v のSRRCを見ると、 $t = 8 \times 10^4 (y)$ を境にそれ以前では、モデル出力である線量に対して正の相関を持つが、この時点以後では負の相関を示している。これは、流速 v が地層中の放射性核種の移行時間に関連し、比較的早い時間にはモデル出力である線量に対して正の相関を持つが、遅い時間では負の相関を持っていることを表している。また、 $t = 8 \times 10^4 (y)$ 付近ではモデルの非単調性により、SRRCによるパラメータの感度の順位付けは意味を成していない。これに対し、モデル独立でモデル構造に依存しない S_{Ti} は $t = 8 \times 10^4 (y)$ を含め、地層中の流速 v が線量の総分散に最も寄与の大きなパラメータであることを示すことに成功している。

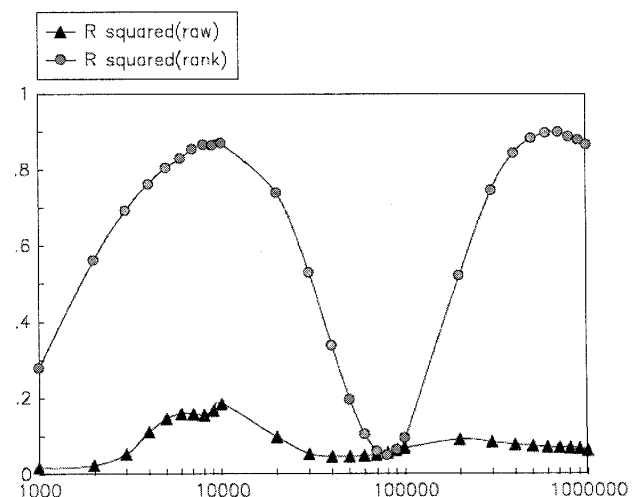


図1 テストケースにおける決定係数

表3 3時点における線量に対する総感度指標 S_{T_i} と $SRRC$ の値

R^2	$t=10^4$ (y)		$t=8 \times 10^4$ (y)		$t=7 \times 10^5$ (y)	
	S_{T_i}	$SRRC$	S_{T_i}	$SRRC$	S_{T_i}	$SRRC$
	0.869		0.050		0.898	
v	0.79	0.82	0.99	-0.04	0.87	-0.92
W	0.42	-0.05	0.51	-0.12	0.44	-0.04
R	0.45	-0.38	0.48	-0.18	0.61	0.07
l	0.40	-0.22	0.40	0.04	0.47	0.23
T	-0.02	0.00	-0.02	0.00	0.00	0.00
k	0.03	0.00	-0.02	0.00	0.00	0.00

6. おわりに

本稿では、代表的なグローバル感度解析手法の一つである Sobol' 法を紹介した。Sobol' 法は FAST と同様、モデル構造によらない、すなわちモデル独立な手法で、入力変数の主効果だけでなく総感度指標を計算することで、入力変数のすべての次元の交互作用を考慮することができ非常に魅力的な手法である。ただし、すべての1次オーダーおよび総感度指標を計算するのに $N(2k+1)$ のモデル関数の計算が必要である。したがって、計算時間がかかるモデル関数ではメタモデルの使用か、あるいはスクリーニングにより入力変数の重要度をあらかじめ推定し、重要でない入力変数を固定化する等の何らかの工夫が必要になる[14]。

参考文献

- [1] Iman, R. L. (1987). A matrix-based approach to uncertainty and sensitivity analysis for fault trees. *Risk Analysis*, 7(1), 21-33.
- [2] Ishigami, T. and Homma, T. (1989). An importance quantification technique in uncertainty analysis for computer models. JAERI-M 89-111, Japan Atomic Energy Research Institute Report.
- [3] Iman, R. L. and Hora, S. C. (1990). A robust measure of uncertainty importance for use in fault tree system analysis. *Risk Analysis*, 10(3), 401-406.
- [4] Krzykacz, B. (1990). A computerprogram for the derivation of empirical sensitivity measures of results from large computermodels. GRS-A-1700, Gesellschaft fur Reaktorsicherheit (GRS) mbH, Garching, Germany.
- [5] Mckay, M. D. (1997). Nonparametric variance-based methods for assessing uncertainty importance. *Reliability Engineering and System Safety* 57, 267-279.
- [6] Sobol', I. M. (1990). Sensitivity estimates for nonlinear mathematical models. *Matem Modelirovanie*, 2(1), 112-118 [in Russian]. English translation: (1993). *Math. Modelling Comput. Exp.*, 1(4), 407-414.
- [7] Archer, G. E. B., Saltelli, A. and Sobol', I. M. (1997). Sensitivity measures, ANOVA-like techniques and use of bootstrap. *J. Statist. Comput. Simul.*, 58, 99-120.
- [8] Sobol', I. M. (2003). Theorems and examples on high dimensional model representation, *Reliability Engineering and System Safety*, 79, 187-193.
- [9] Homma, T. and Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering and System Safety*, 52, 1-17.
- [10] Sobol', I. M. (1998). On quasi-Monte-Carlo integrations. *Math. Comput. Simul.*, 47, 103-112.
- [11] Bratley, P., Fox, B. and Niederreiter, H. (1992). Implementation and tests of low-discrepancy sequences. *ACM Trans. Model Comput. Simul.*, 2(3), 195-213.
- [12] Homma, T. and Saltelli, A. (1995). Use of Sobol's quasirandom sequence generator for integration of modified uncertainty importance measure. *Journal of Nuclear Science and Technology*, 32, 1164-1173.
- [13] Robinson, P. C. and Hodgkinson, D. P. (1987). Exact solutions for radionuclide transport in the presence of parameter uncertainty. *Radioactive Waste Management and the Nuclear Fuel Cycle*, 8, 283-311.
- [14] Sobol', I. M., Tarantola, S., Gatelli, D., Kucherenko, S. S. and Mauntz, W. (2007). Estimating the approximation error when fixing unessential factors in global sensitivity analysis. *Reliability Engineering and System Safety*, 92, 957-960.