

Nathalie SAINT-GEOURS
Doctorante UMR TETIS

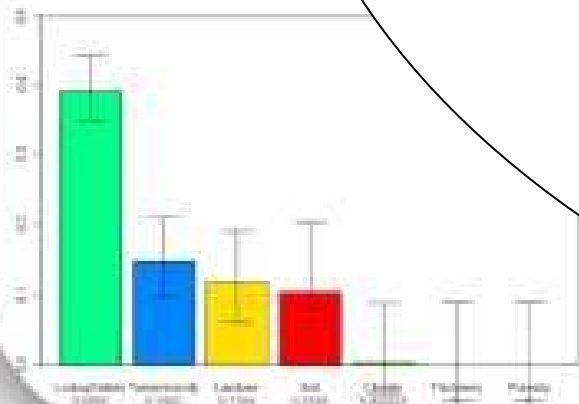
Global sensitivity analysis with spatial input information



Manaaki Whenua
Landcare Research



CEA Cadarache
12 janvier 2010



Contexte

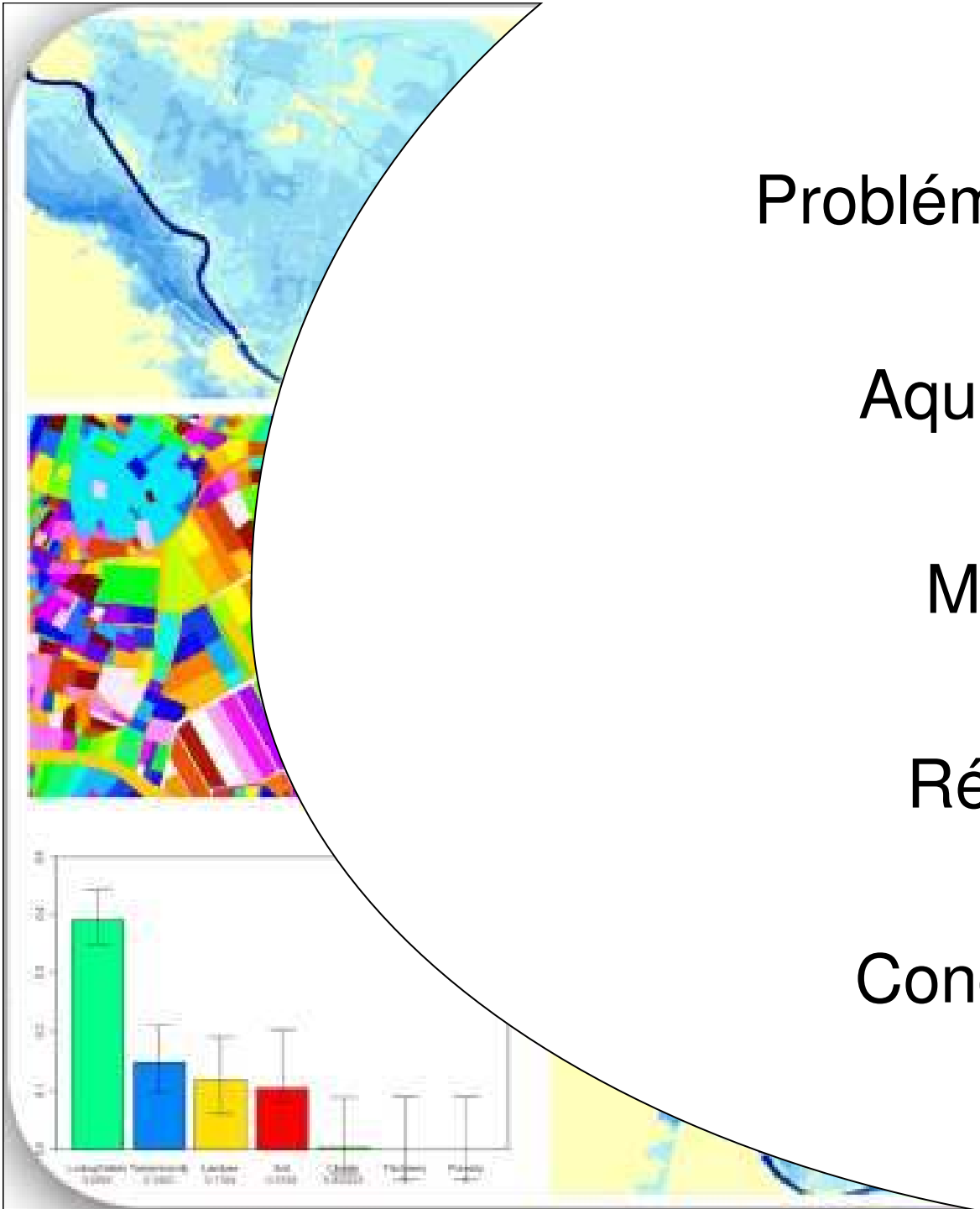
- Thèse AgroParisTech

Contexte

- Thèse AgroParisTech
- UMR TETIS à Montpellier
Territoires, Environnement, Télédétection et Information Spatialisée

Contexte

- Thèse AgroParisTech
- UMR TETIS à Montpellier
Territoires, Environnement, Télédétection et Information Spatialisée
- Travail de Mastère réalisé avec Linda Lilburne
 - Nouvelle-Zélande – Landcare Research
 - Analyse de sensibilité d'un modèle de transport de polluants



Problématique

AquiferSim

Méthode

Résultats

Conclusion

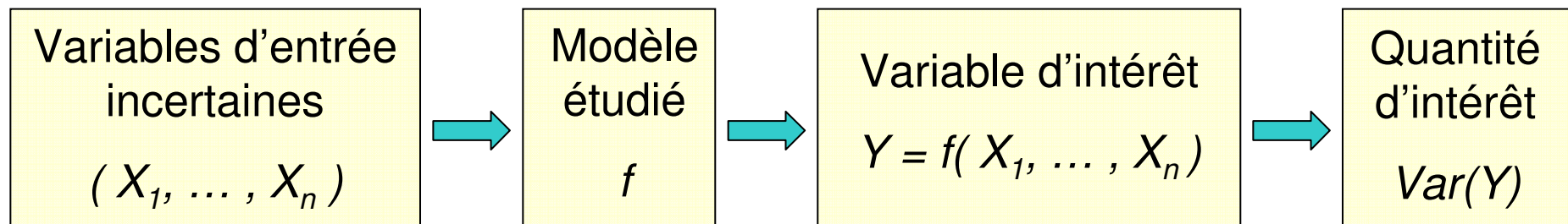


AS globale et variables d'entrées spatialisées

- Analyse de sensibilité globale basée sur la variance

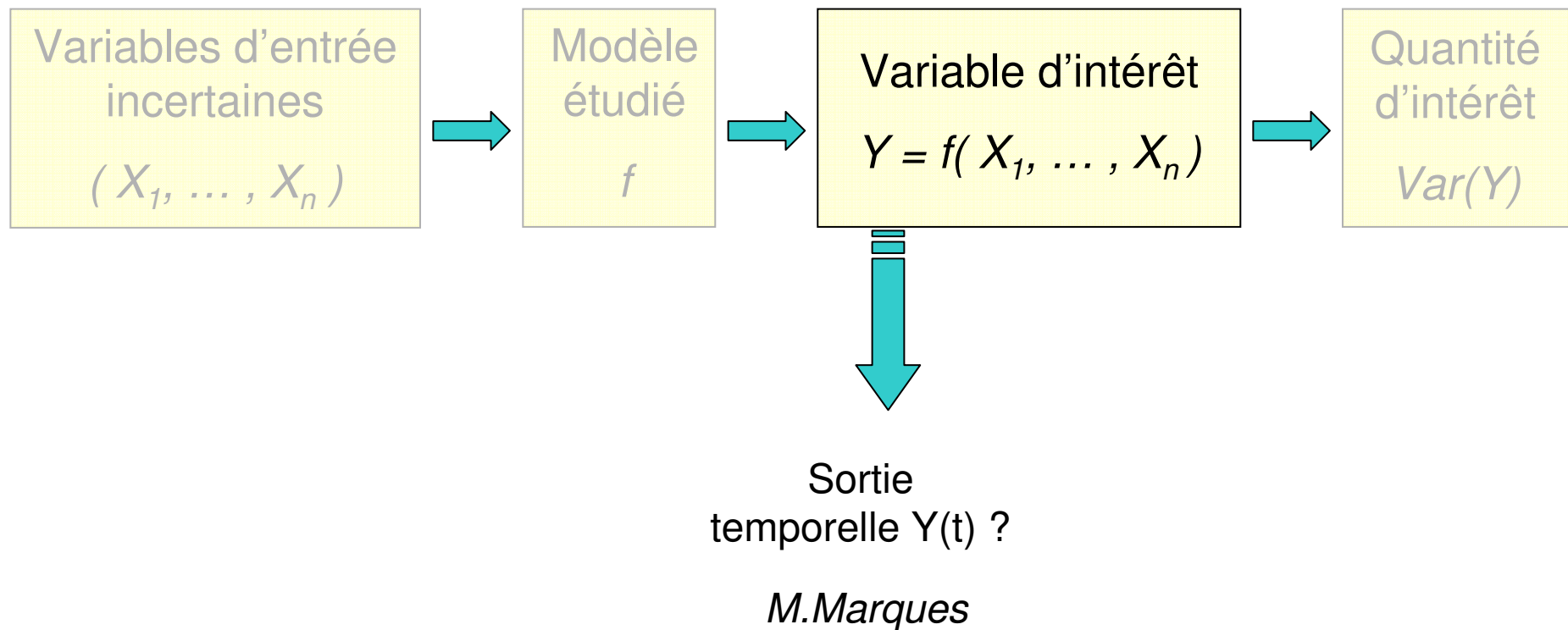
AS globale et variables d'entrées spatialisées

- Analyse de sensibilité globale basée sur la variance



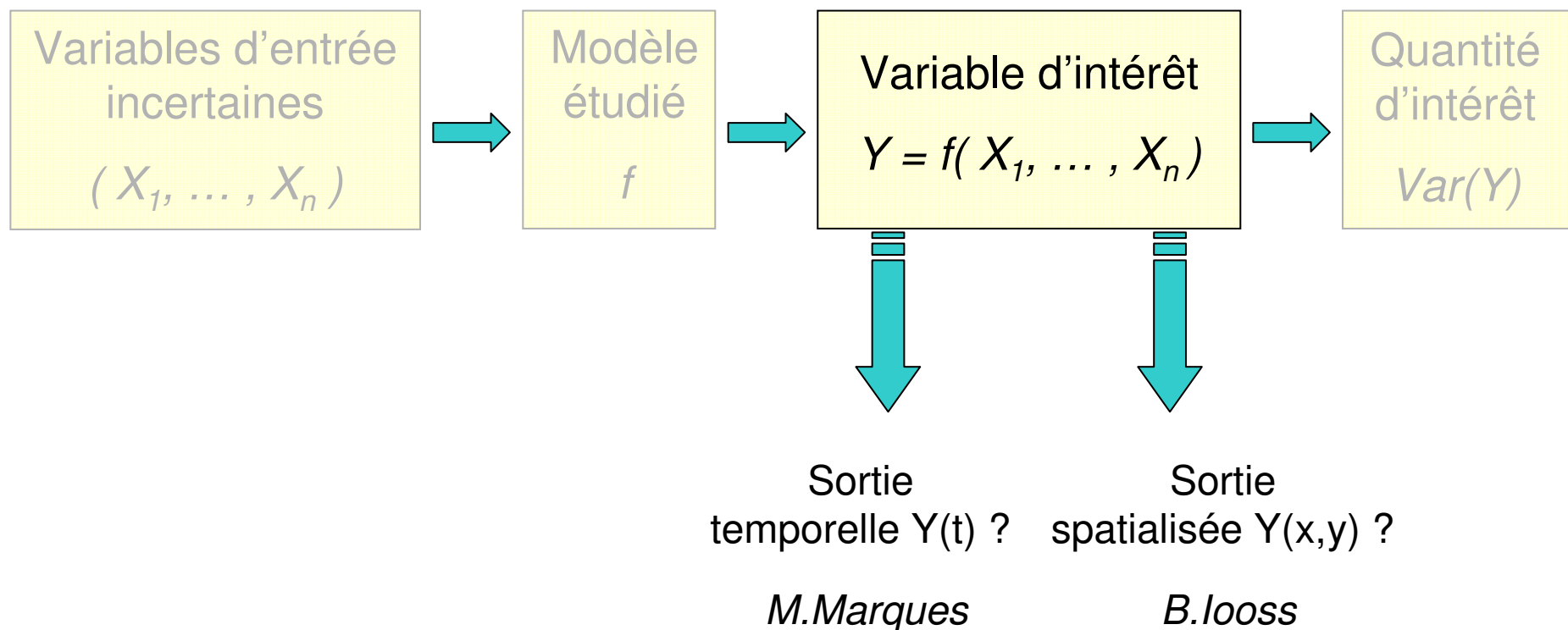
AS globale et variables d'entrées spatialisées

- Analyse de sensibilité globale basée sur la variance



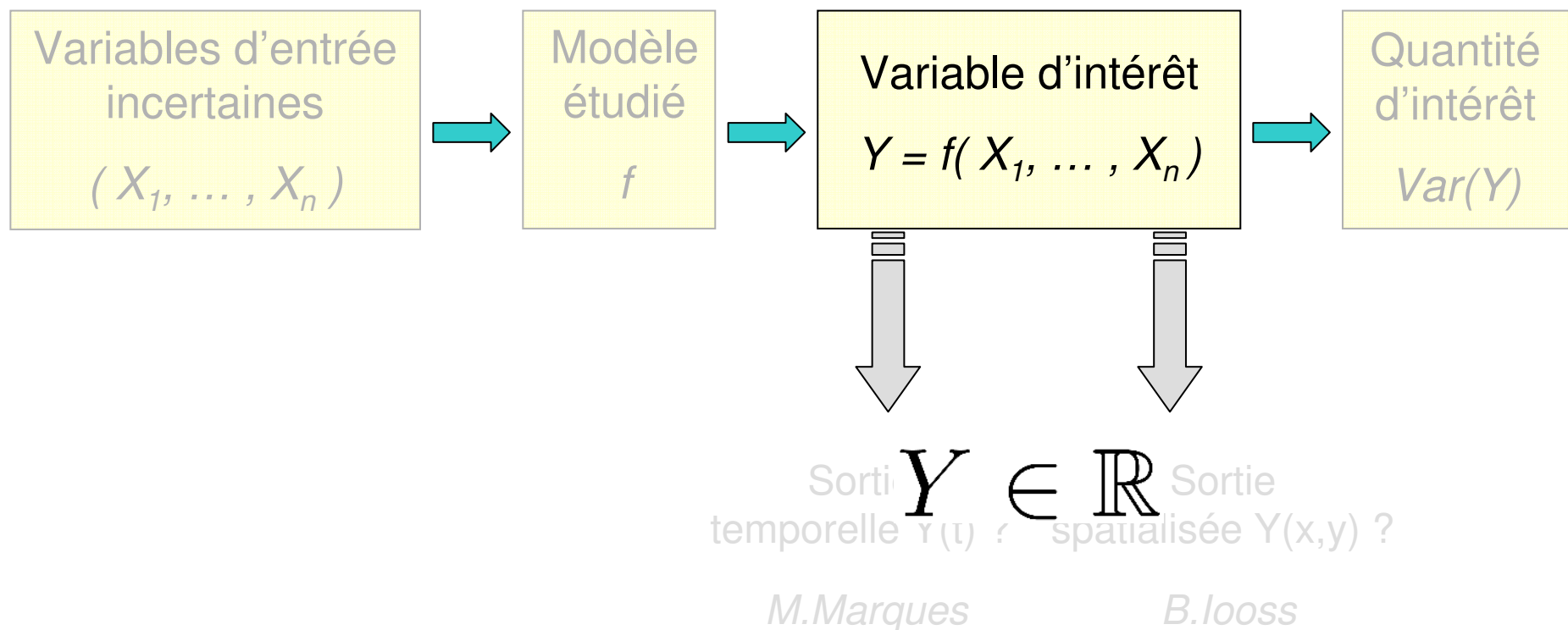
AS globale et variables d'entrées spatialisées

- Analyse de sensibilité globale basée sur la variance



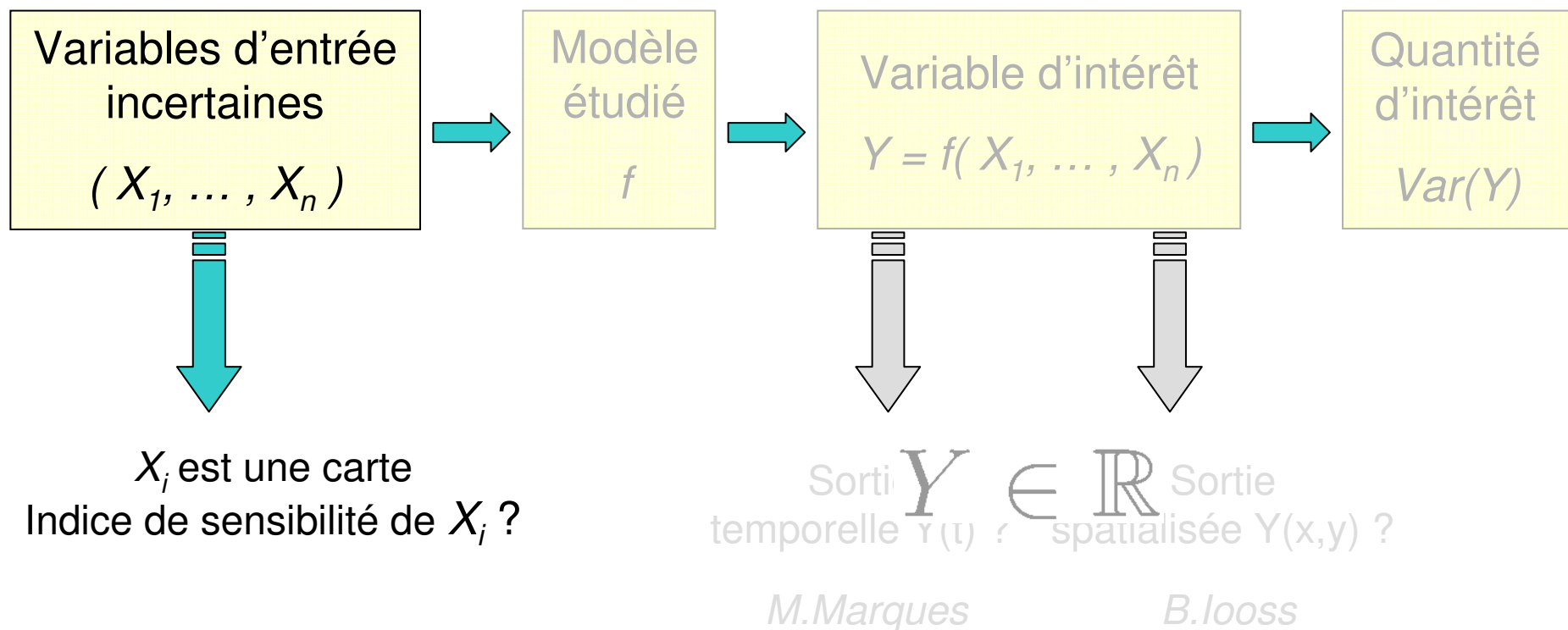
AS globale et variables d'entrées spatialisées

- Analyse de sensibilité globale basée sur la variance



AS globale et variables d'entrées spatialisées

- Analyse de sensibilité globale basée sur la variance



Plusieurs approches déjà envisagées

- Méthode du « macroparamètre » *[Jacques et al., 2006]*
- Méthode du paramètre « trigger » *[Tarantola, 2002]*
- Méta-modélisation par « Joint GAM » *[looss & Ribatet, 2009]*
- Scénarios de carte *[Ruffo & al., 2006 ; Lilburne & Tarantola, 2009]*

Modèle AquiferSim

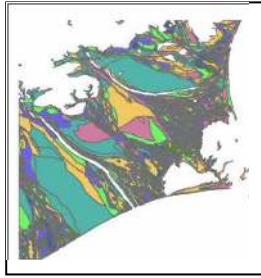
Modèle AquiferSim

- Modèle d'écoulement des eaux souterraines et de transport de contaminants en régime permanent
 - analyse de la pollution agricole diffuse « de la parcelle à l'aquifère »
 - échelle régionale (3500 km², maille de 100m)
 - comparaison de différentes politiques d'occupation du sol

Modèle AquiferSim

- Modèle d'écoulement des eaux souterraines et de transport de contaminants en régime permanent
 - analyse de la pollution agricole diffuse « de la parcelle à l'aquifère »
 - échelle régionale (3500 km², maille de 100m)
 - comparaison de différentes politiques d'occupation du sol

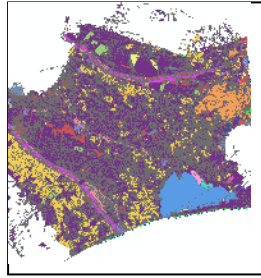
- Objectif : aide à la décision pour l'aménagement du territoire



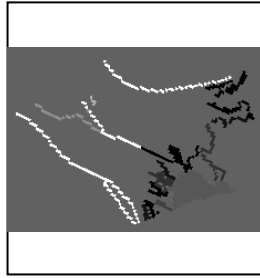
Nature des sols



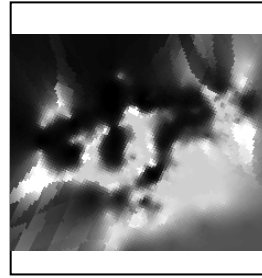
**Précipitations
annuelles**



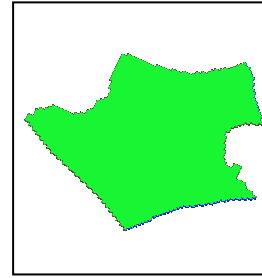
Occ. sol



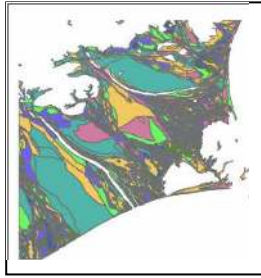
**Réseau
hydrologique**



**Transmissivité
aquifère**



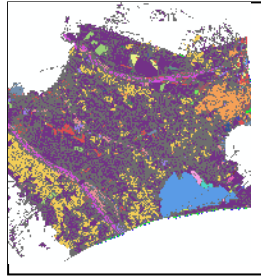
**Conditions
aux limites**



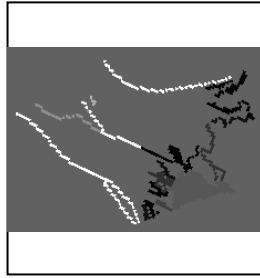
Nature des sols



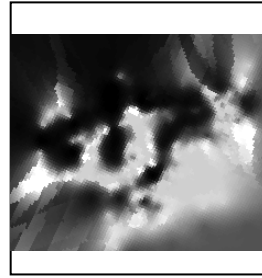
Précipitations annuelles



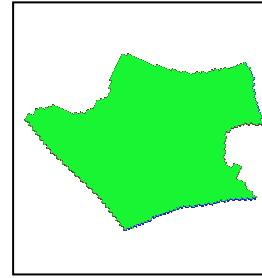
Occ. sol



Réseau hydrologique



Transmissivité aquifère



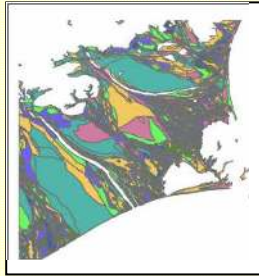
Conditions aux limites

$$(X_1, \dots, X_n)$$

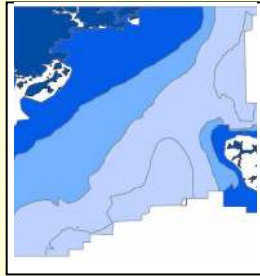
Paramètres scalaires

Soil	LU	Clim	Nit
1	510	1	1,22
2	511	1	2,15
3	510	1	7,51
4	589	1	4,15
5	741	2	6,24
6	125	2	4,12
7	632	2	8,12
8	512	2	7,13
9	514	2	3,54
10	123	2	0,11

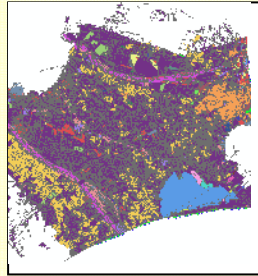
Données tabulaires



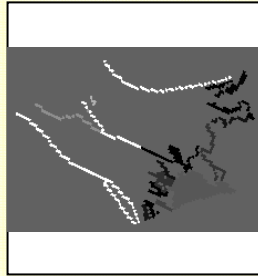
Nature des sols



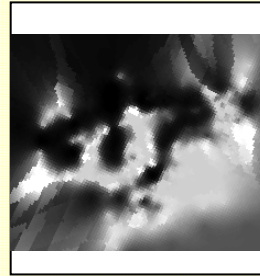
Précipitations annuelles



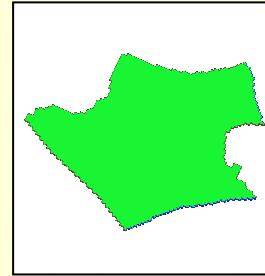
Occ. sol



Réseau hydrologique



Transmissivité aquifère



Conditions aux limites



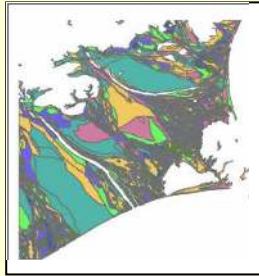
Modèle AquiferSim
*(interface SIG + modèle éléments finis
2D/3D « en tranches »)*

$$(X_1, \dots, X_n)$$

**Paramètres
scalaires**

Soil	LU	Clim	Nit
1	510	1	1,22
2	511	1	2,15
3	510	1	7,51
4	589	1	4,15
5	741	2	6,24
6	125	2	4,12
7	632	2	8,12
8	512	2	7,13
9	514	2	3,54
10	123	2	0,11

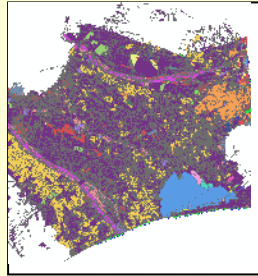
Données tabulaires



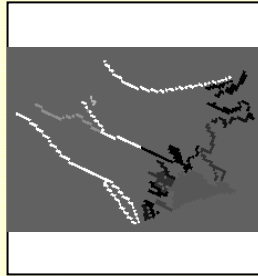
Nature des sols



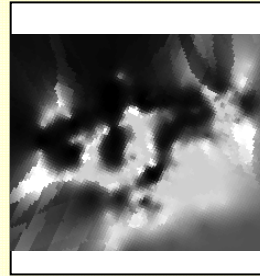
Précipitations annuelles



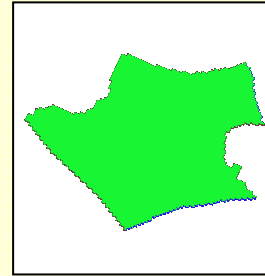
Occ. sol



Réseau hydrologique



Transmissivité aquifère



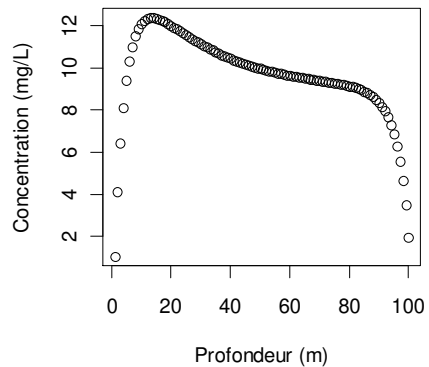
Conditions aux limites



Modèle AquiferSim
(interface SIG + modèle éléments finis
2D/3D « en tranches »)



Concentrations en nitrate N(x,y,z)

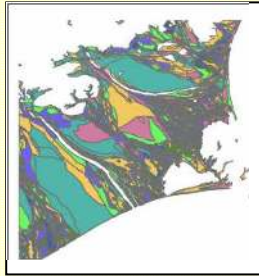


$$(X_1, \dots, X_n)$$

Paramètres scalaires

Soil	LU	Clim	Nit
1	510	1	1,22
2	511	1	2,15
3	510	1	7,51
4	589	1	4,15
5	741	2	6,24
6	125	2	4,12
7	632	2	8,12
8	512	2	7,13
9	514	2	3,54
10	123	2	0,11

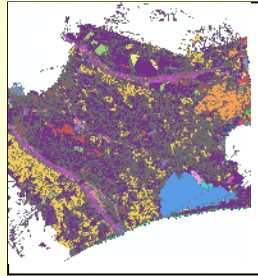
Données tabulaires



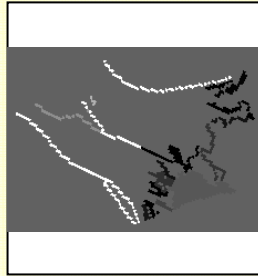
Nature des sols



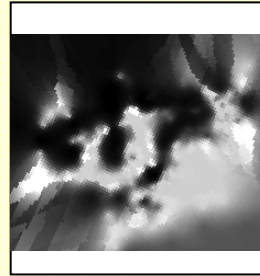
Précipitations annuelles



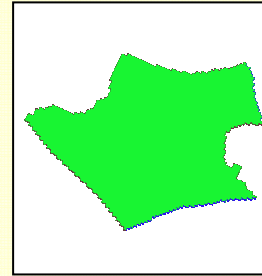
Occ. sol



Réseau hydrologique



Transmissivité aquifère



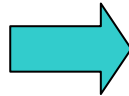
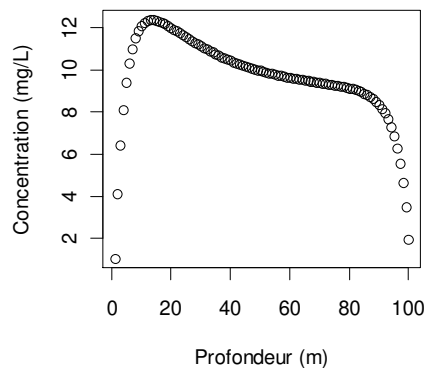
Conditions aux limites



Modèle AquiferSim
(interface SIG + modèle éléments finis
2D/3D « en tranches »)



Concentrations en nitrate N(x,y,z)



$$(X_1, \dots, X_n)$$

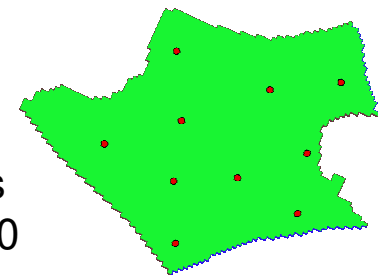
Paramètres scalaires

Soil	LU	Clim	Nit
1	510	1	1,22
2	511	1	2,15
3	510	1	7,51
4	589	1	4,15
5	741	2	6,24
6	125	2	4,12
7	632	2	8,12
8	512	2	7,13
9	514	2	3,54
10	123	2	0,11

Données tabulaires

Variable d'intérêt

Y = moyenne des concentrations maximales observées à l'aplomb de 10 points de contrôle



0. Spécification du problème

0. Spécification du problème

- N = 5 variables d'entrée incertaines

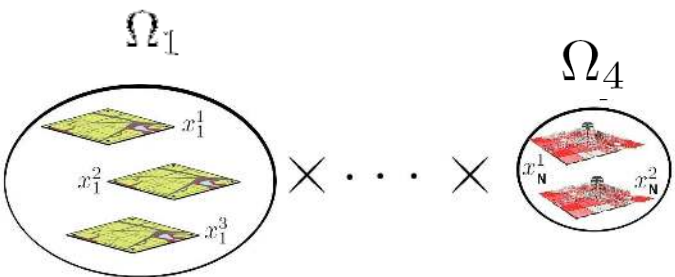
Variable	Notation	Nature
Nature du sol	C_1	Couche vectorielle SIG – 7 classes
Occupation du sol	C_2	Couche vectorielle SIG – 30 classes
Précipitations annuelles	C_3	Champ 2D continu (grille 1025 x 1025)
Transmissivité de l'aquifère	C_4	Champ 2D continu (grille 1025 x 1025)
Epaisseur de l'aquifère	X_5	Scalaire

0. Spécification du problème

- N = 5 variables d'entrée incertaines

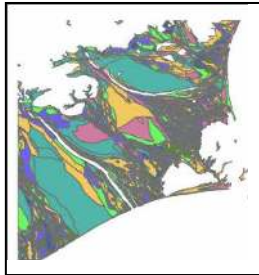
Variable	Notation	Nature
Nature du sol	c_1	Couche vectorielle SIG – 7 classes
Occupation du sol	c_2	Couche vectorielle SIG – 30 classes
Précipitations annuelles	c_3	Champ 2D continu (grille 1025 x 1025)
Transmissivité de l'aquifère	c_4	Champ 2D continu (grille 1025 x 1025)
Epaisseur de l'aquifère	x_5	Scalaire

- Variable d'intérêt Y :

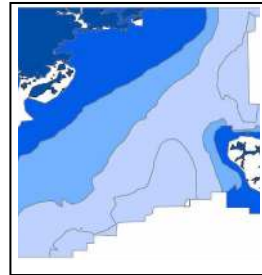
$$Y : \left(\Omega_1 \times \dots \times \Omega_4 \right) \times \mathbb{R} \rightarrow \mathbb{R}$$


$$(c_1, c_2, c_3, c_4, x_5) \mapsto f(c_1, c_2, c_3, c_4, x_5)$$

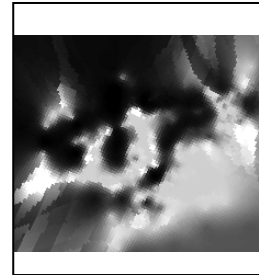
1. Quantifier et simuler les incertitudes



x



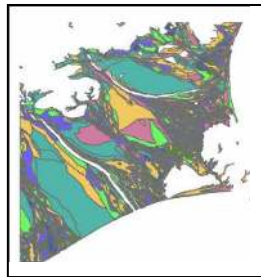
x



x

 X_5 scalaire

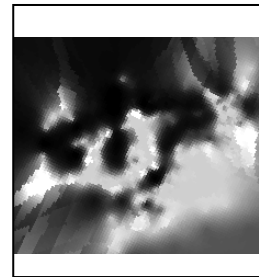
1. Quantifier et simuler les incertitudes



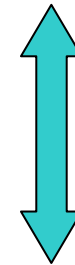
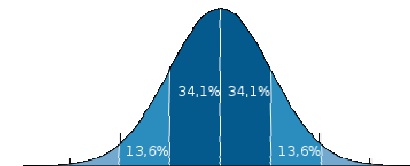
x



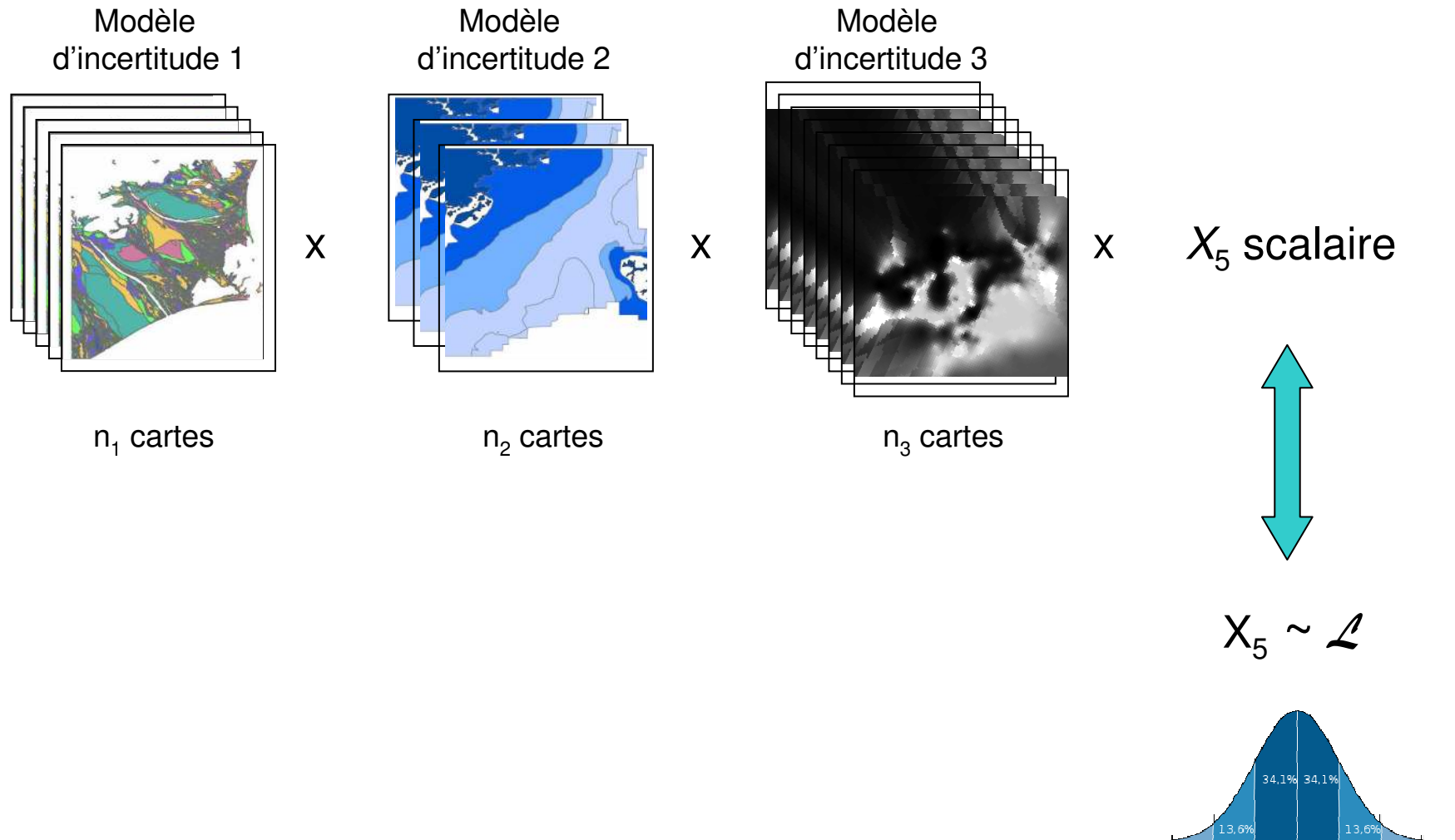
x



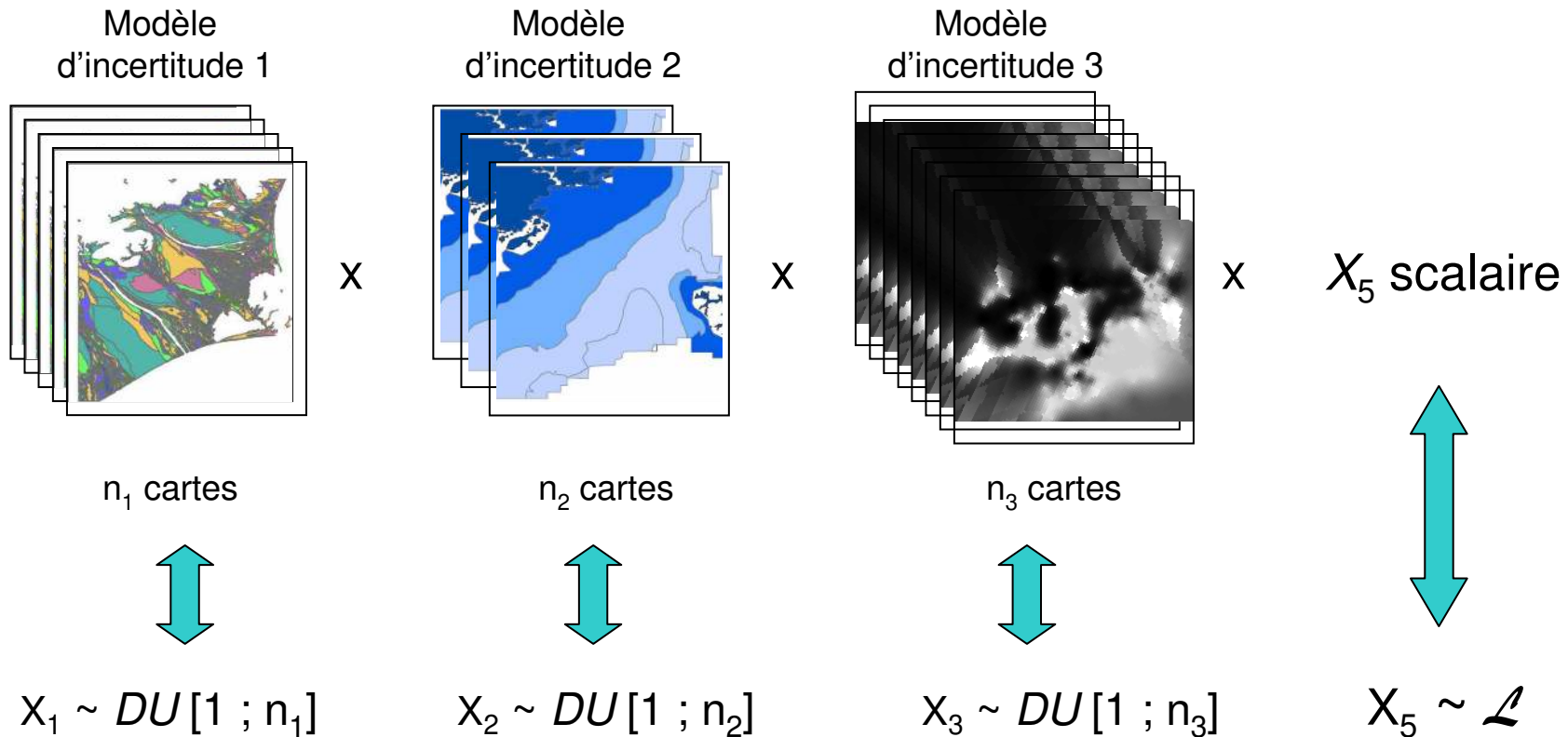
x

 X_5 scalaire $X_5 \sim \mathcal{L}$ 

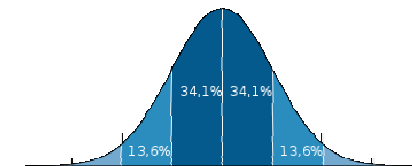
1. Quantifier et simuler les incertitudes



1. Quantifier et simuler les incertitudes



« Indice de carte »
 [Ruffo, 2006] [Lilburne & Tarantola, 2009]



Modèles d'incertitudes

Variable	Notation	Modèle d'incertitude	n_i
Nature du sol	C_1	Simulation de Monte-Carlo à partir d'une matrice de confusion	100
Occupation du sol	C_2	Simulation de Monte-Carlo à partir d'une matrice de confusion	100
Précipitations annuelles	C_3	Simulation gaussienne séquentielle	100
Transmissivité de l'aquifère	C_4	3 alternatives	3
Profondeur de l'aquifère	X_5	Triangulaire(200 ; 350 ; 500)	-

➡ Possibilité de décrire l'incertitude de manière complexe

Modèles d'incertitude : carte du sol

- Couche vectorielle
- 7 classes de nature de sol
- Confusion entre classes

	C10	C11	C12	C13	C14	C15	C16
C10	60	10	10	10	5	5	5
C11	10	60	10	10	5	5	5
C12	10	10	60	10	5	5	5
C13	10	10	10	60	5	5	5
C14	10	10	10	10	70	5	5
C15	0	0	0	0	5	65	10
C16	0	0	0	0	5	10	65

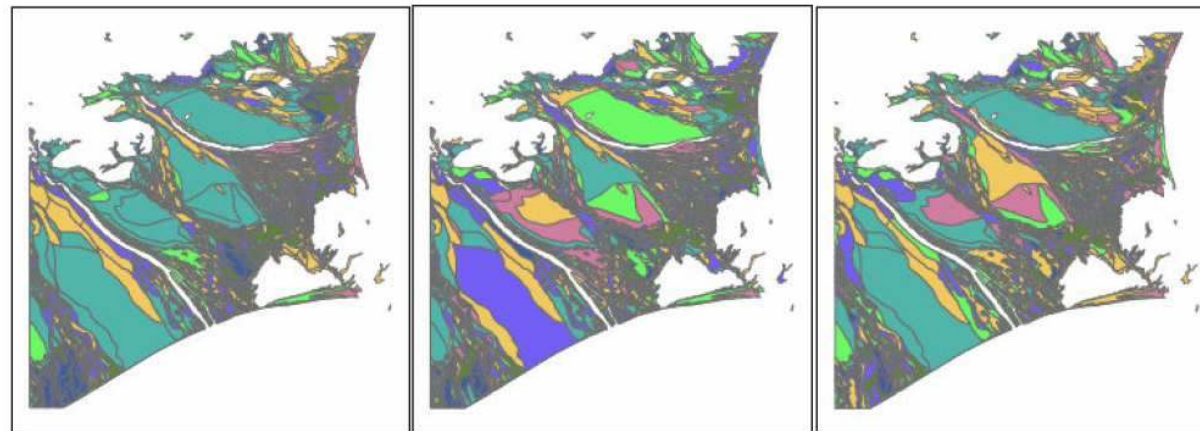
Matrice de confusion entre classes

Modèles d'incertitude : carte du sol

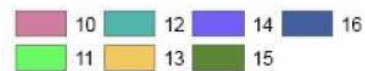
- Couche vectorielle
- 7 classes de nature de sol
- Confusion entre classes

	C10	C11	C12	C13	C14	C15	C16
C10	60	10	10	10	5	5	5
C11	10	60	10	10	5	5	5
C12	10	10	60	10	5	5	5
C13	10	10	10	60	5	5	5
C14	10	10	10	10	70	5	5
C15	0	0	0	0	5	65	10
C16	0	0	0	0	5	10	65

Matrice de confusion entre classes



Soil classes



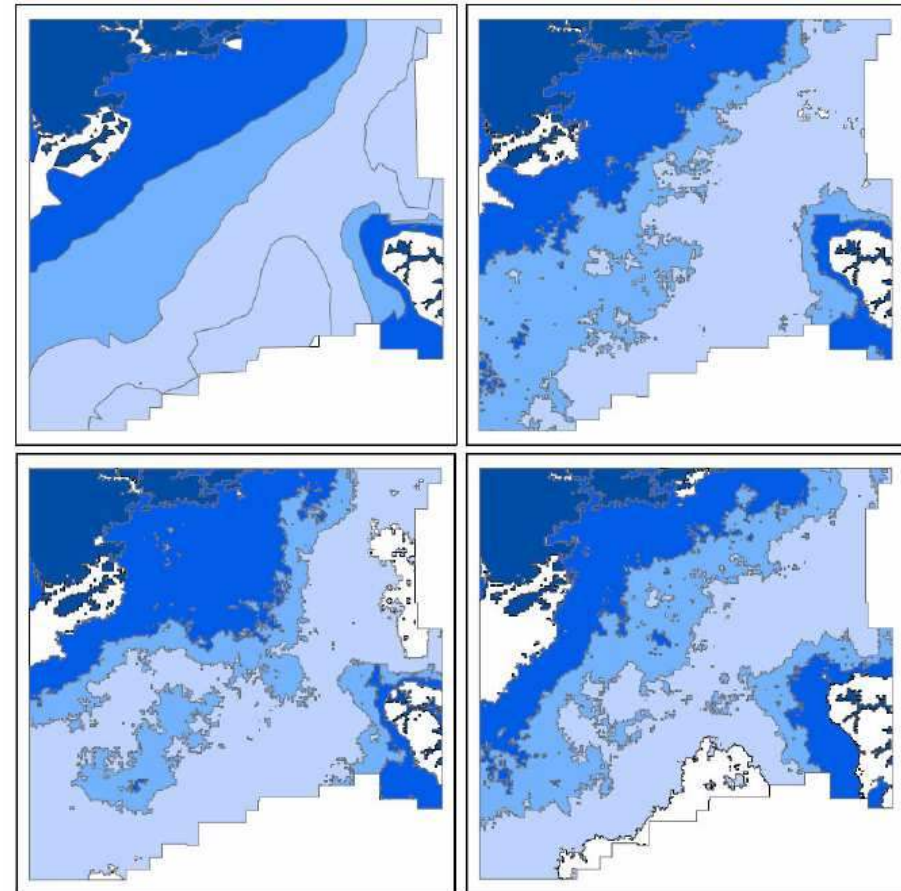
Soil map simulations

(left map is the original one)

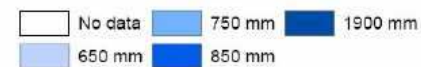
100 réalisations
aléatoires

Modèles d'incertitude : précipitations

- Précipitations annuelles
 - Raster 1025 * 1025
 - Résolution 100m
 - Interpolation par spline



Mean annual rainfall total



0 15 30 Kilometers

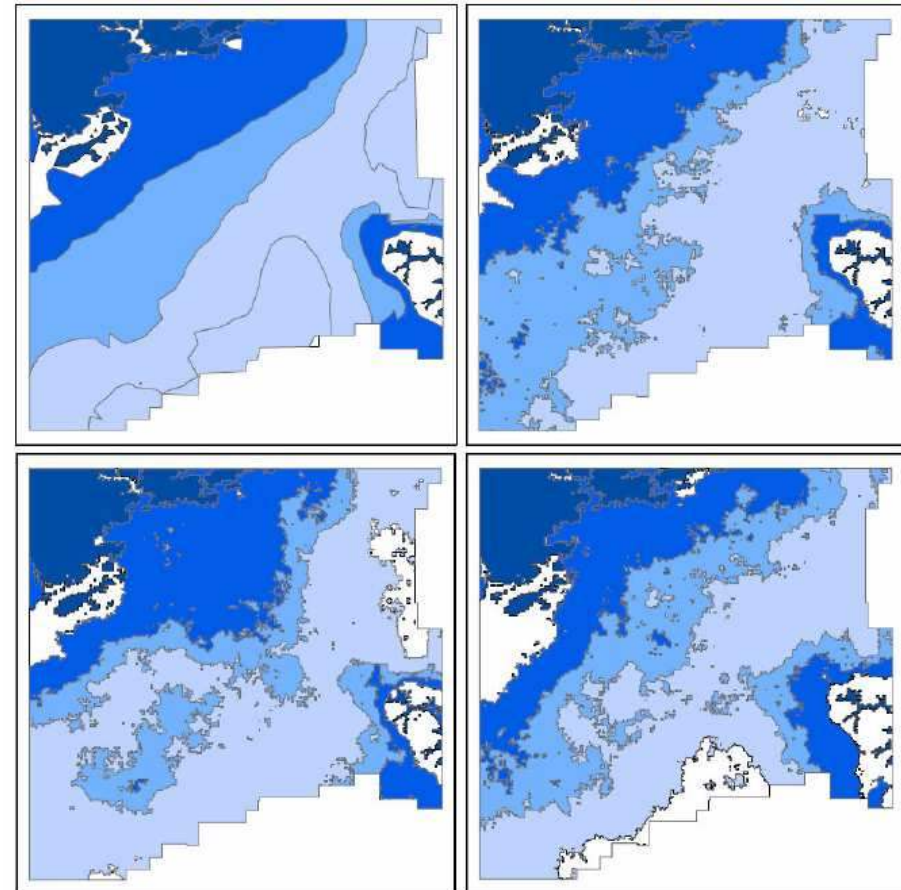
The upper left map is the original one.

Modèles d'incertitude : précipitations

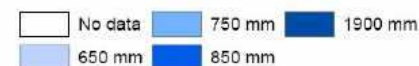
- Précipitations annuelles
 - Raster 1025 * 1025
 - Résolution 100m
 - Interpolation par spline
- Génération de champs d'erreurs
 - Gaussien de moyenne nulle
 - Sim. Gaussienne Séquentielle
 - SGeMS (bibliothèque GSLib)



100 réalisations aléatoires



Mean annual rainfall total



0 15 30 Kilometers

The upper left map is the original one.

2. Propager les incertitudes

2. Propager les incertitudes

X1	X2	X3	X4
1	2	45	5,36
5	3	12	2,1
1	7	78	4,59
8	1	49	7,21
4	5	63	8,12
9	1	70	4,61
6	4	26	5,09
7	2	35	4,85

X1	X2	X3	X4
9	4	50	5,12
5	3	83	8,38
9	-1	17	5,89
2	5	46	3,27
6	1	32	2,36
1	5	25	5,87
4	2	69	5,39
3	4	60	5,63

Tirage de deux échantillons quasi-aléatoires
de taille $L = 512$ [Sobol, 1995]

2. Propager les incertitudes

X1	X2	X3	X4
1	2	45	5,36
5	3	12	2,1
1	7	78	4,59
8	1	49	7,21
4	5	63	8,12
9	1	70	4,61
6	4	26	5,09
7	2	35	4,85

X1	X2	X3	X4
9	4	50	5,12
5	3	83	8,38
9	-1	17	5,89
2	5	46	3,27
6	1	32	2,36
1	5	25	5,87
4	2	69	5,39
3	4	60	5,63

Tirage de deux échantillons quasi-aléatoires
de taille $L = 512$ [Sobol, 1995]



$$A = \begin{bmatrix} x_1^{(1)} & \dots & x_j^{(1)} & \dots & x_k^{(1)} \\ x_1^{(2)} & \dots & x_j^{(2)} & \dots & x_k^{(2)} \\ \vdots & & \vdots & & \vdots \\ x_1^{(N)} & \dots & x_j^{(N)} & \dots & x_k^{(N)} \end{bmatrix} \quad
 B = \begin{bmatrix} x_1^{(N+1)} & \dots & x_j^{(N+1)} & \dots & x_k^{(N+1)} \\ x_1^{(N+2)} & \dots & x_j^{(N+2)} & \dots & x_k^{(N+2)} \\ \vdots & & \vdots & & \vdots \\ x_1^{(2N)} & \dots & x_j^{(2N)} & \dots & x_k^{(2N)} \end{bmatrix}$$

$$D_i = \begin{bmatrix} x_1^{(1)} & \dots & x_j^{(N+1)} & \dots & x_k^{(1)} \\ x_1^{(2)} & \dots & x_j^{(N+2)} & \dots & x_k^{(2)} \\ \vdots & & \vdots & & \vdots \\ x_1^{(N)} & \dots & x_j^{(2N)} & \dots & x_k^{(N)} \end{bmatrix} \quad
 C_i = \begin{bmatrix} x_1^{(N+1)} & \dots & x_j^{(1)} & \dots & x_k^{(N+1)} \\ x_1^{(N+2)} & \dots & x_j^{(2)} & \dots & x_k^{(N+2)} \\ \vdots & & \vdots & & \vdots \\ x_1^{(2N)} & \dots & x_j^{(N)} & \dots & x_k^{(2N)} \end{bmatrix}$$

Création d'un nouvel
échantillon de taille $2(N+1)L$
par permutations [Tarantola
& Lilburne, 2009]

2. Propager les incertitudes

X1	X2	X3	X4
1	2	45	5,36
5	3	12	2,1
1	7	78	4,59
8	1	49	7,21
4	5	63	8,12
9	1	70	4,61
6	4	26	5,09
7	2	35	4,85

X1	X2	X3	X4
9	4	50	5,12
5	3	83	8,38
9	-1	17	5,89
2	5	46	3,27
6	1	32	2,36
1	5	25	5,87
4	2	69	5,39
3	4	60	5,63

Tirage de deux échantillons quasi-aléatoires
de taille $L = 512$ [Sobol, 1995]



$$A = \begin{bmatrix} x_1^{(1)} & \dots & x_j^{(1)} & \dots & x_k^{(1)} \\ x_1^{(2)} & \dots & x_j^{(2)} & \dots & x_k^{(2)} \\ \vdots & & \vdots & & \vdots \\ x_1^{(N)} & \dots & x_j^{(N)} & \dots & x_k^{(N)} \end{bmatrix} \quad
 B = \begin{bmatrix} x_1^{(N+1)} & \dots & x_j^{(N+1)} & \dots & x_k^{(N+1)} \\ x_1^{(N+2)} & \dots & x_j^{(N+2)} & \dots & x_k^{(N+2)} \\ \vdots & & \vdots & & \vdots \\ x_1^{(2N)} & \dots & x_j^{(2N)} & \dots & x_k^{(2N)} \end{bmatrix}$$

$$D_i = \begin{bmatrix} x_1^{(1)} & \dots & x_j^{(N+1)} & \dots & x_k^{(1)} \\ x_1^{(2)} & \dots & x_j^{(N+2)} & \dots & x_k^{(2)} \\ \vdots & & \vdots & & \vdots \\ x_1^{(N)} & \dots & x_j^{(2N)} & \dots & x_k^{(N)} \end{bmatrix} \quad
 C_i = \begin{bmatrix} x_1^{(N+1)} & \dots & x_j^{(1)} & \dots & x_k^{(N+1)} \\ x_1^{(N+2)} & \dots & x_j^{(2)} & \dots & x_k^{(N+2)} \\ \vdots & & \vdots & & \vdots \\ x_1^{(2N)} & \dots & x_j^{(N)} & \dots & x_k^{(2N)} \end{bmatrix}$$

Création d'un nouvel
échantillon de taille $2(N+1)L$
par permutations [Tarantola
& Lilburne, 2009]



X1	X2	X3	X4
1	2	45	5,36
5	3	12	2,1
1	7	78	4,59
8	1	49	7,21
2	5	46	3,2725
9	1	70	4,61
6	4	26	5,09
7	2	35	4,85
9	4	50	5,1225
5	3	83	8,3825
6	4	26	5,09
2	5	46	3,2725
6	1	32	2,3625
1	5	25	5,8725
4	2	69	5,3925
3	4	60	5,6325
6	4	26	5,09
2	5	46	3,2725

2. Propager les incertitudes

X1	X2	X3	X4
1	2	45	5,36
5	3	12	2,1
1	7	78	4,59
8	1	49	7,21
4	5	63	8,12
9	1	70	4,61
6	4	26	5,09
7	2	35	4,85

X1	X2	X3	X4
9	4	50	5,12
5	3	83	8,38
9	-1	17	5,89
2	5	46	3,27
6	1	32	2,36
1	5	25	5,87
4	2	69	5,39
3	4	60	5,63

Tirage de deux échantillons quasi-aléatoires
de taille $L = 512$ [Sobol, 1995]



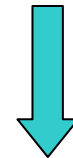
$$A = \begin{bmatrix} x_1^{(1)} & \dots & x_j^{(1)} & \dots & x_k^{(1)} \\ x_1^{(2)} & \dots & x_j^{(2)} & \dots & x_k^{(2)} \\ \vdots & & \vdots & & \vdots \\ x_1^{(N)} & \dots & x_j^{(N)} & \dots & x_k^{(N)} \end{bmatrix} \quad
 B = \begin{bmatrix} x_1^{(N+1)} & \dots & x_j^{(N+1)} & \dots & x_k^{(N+1)} \\ x_1^{(N+2)} & \dots & x_j^{(N+2)} & \dots & x_k^{(N+2)} \\ \vdots & & \vdots & & \vdots \\ x_1^{(2N)} & \dots & x_j^{(2N)} & \dots & x_k^{(2N)} \end{bmatrix}$$

$$D_i = \begin{bmatrix} x_1^{(1)} & \dots & x_j^{(N+1)} & \dots & x_k^{(1)} \\ x_1^{(2)} & \dots & x_j^{(N+2)} & \dots & x_k^{(2)} \\ \vdots & & \vdots & & \vdots \\ x_1^{(N)} & \dots & x_j^{(2N)} & \dots & x_k^{(N)} \end{bmatrix} \quad
 C_i = \begin{bmatrix} x_1^{(N+1)} & \dots & x_j^{(1)} & \dots & x_k^{(N+1)} \\ x_1^{(N+2)} & \dots & x_j^{(2)} & \dots & x_k^{(N+2)} \\ \vdots & & \vdots & & \vdots \\ x_1^{(2N)} & \dots & x_j^{(N)} & \dots & x_k^{(2N)} \end{bmatrix}$$

Création d'un nouvel
échantillon de taille $2(N+1)L$
par permutations [Tarantola
& Lilburne, 2009]



Evaluations multiples
du modèle AquiferSim



X1	X2	X3	X4	Y
1	2	45	5,36	1,727
5	3	12	2,1	1,344
1	7	78	4,59	1,957
8	1	49	7,21	1,814
2	5	46	3,2725	1,75
9	1	70	4,61	1,927
6	4	26	5,09	1,614
7	2	35	4,85	1,689
9	4	50	5,1225	1,833
5	3	83	8,3825	1,997
6	4	26	5,09	1,614
2	5	46	3,2725	1,75
6	1	32	2,3625	1,617
1	5	25	5,8725	1,567
4	2	69	5,3925	1,905
3	4	60	5,6325	1,861
6	4	26	5,09	1,614
2	5	46	3,2725	1,75

2. Propager les incertitudes

X1	X2	X3	X4
1	2	45	5,36
5	3	12	2,1
1	7	78	4,59
8	1	49	7,21
4	5	63	8,12
9	1	70	4,61
6	4	26	5,09
7	2	35	4,85

X1	X2	X3	X4
9	4	50	5,12
5	3	83	8,38
9	-1	17	5,89
2	5	46	3,27
6	1	32	2,36
1	5	25	5,87
4	2	69	5,39
3	4	60	5,63

Tirage de deux échantillons quasi-aléatoires
de taille $L = 512$ [Sobol, 1995]



$$A = \begin{bmatrix} x_1^{(1)} & \dots & x_j^{(1)} & \dots & x_k^{(1)} \\ x_1^{(2)} & \dots & x_j^{(2)} & \dots & x_k^{(2)} \\ \vdots & & \vdots & & \vdots \\ x_1^{(N)} & \dots & x_j^{(N)} & \dots & x_k^{(N)} \end{bmatrix} \quad
 B = \begin{bmatrix} x_1^{(N+1)} & \dots & x_j^{(N+1)} & \dots & x_k^{(N+1)} \\ x_1^{(N+2)} & \dots & x_j^{(N+2)} & \dots & x_k^{(N+2)} \\ \vdots & & \vdots & & \vdots \\ x_1^{(2N)} & \dots & x_j^{(2N)} & \dots & x_k^{(2N)} \end{bmatrix}$$

$$D_i = \begin{bmatrix} x_1^{(1)} & \dots & x_j^{(N+1)} & \dots & x_k^{(1)} \\ x_1^{(2)} & \dots & x_j^{(N+2)} & \dots & x_k^{(2)} \\ \vdots & & \vdots & & \vdots \\ x_1^{(N)} & \dots & x_j^{(2N)} & \dots & x_k^{(N)} \end{bmatrix} \quad
 C_i = \begin{bmatrix} x_1^{(N+1)} & \dots & x_j^{(1)} & \dots & x_k^{(N+1)} \\ x_1^{(N+2)} & \dots & x_j^{(2)} & \dots & x_k^{(N+2)} \\ \vdots & & \vdots & & \vdots \\ x_1^{(2N)} & \dots & x_j^{(N)} & \dots & x_k^{(2N)} \end{bmatrix}$$

Création d'un nouvel
échantillon de taille $2(N+1)L$
par permutations [Tarantola
& Lilburne, 2009]



Evaluations multiples
du modèle AquiferSim



X1	X2	X3	X4	Y
1	2	45	5,36	1,727
5	3	12	2,1	1,344
1	7	78	4,59	1,957
8	1	49	7,21	1,814
2	5	46	3,2725	1,75
9	1	70	4,61	1,927
6	4	26	5,09	1,614
7	2	35	4,85	1,689
9	4	50	5,1225	1,833
5	3	83	8,3825	1,997
6	4	26	5,09	1,614
2	5	46	3,2725	1,75
6	1	32	2,3625	1,617
1	5	25	5,8725	1,567
4	2	69	5,3925	1,905
3	4	60	5,6325	1,861
6	4	26	5,09	1,614
2	5	46	3,2725	1,75

2. Propager les incertitudes

- Coût CPU important :

Coût d'une simulation	10 min
Nombre de variables incertaines	5
Taille L des échantillons de base	512
Nombre de simulations	$2.L.(N+1) = 6144$
Puissance de calcul	10 CPU de 2.66 GHz
Temps de calcul	5 j.

3. Calcul des indices de Sobol

3. Calcul des indices de Sobol

- Variance-based Sensitivity Analysis : indices de Sobol

$$S_i = \frac{V_i}{V} = \frac{V_{X_i} (E_{X_{-i}} (Y|X_i))}{V} \quad S_{Ti} = 1 - \frac{V_{X_{-i}} (E_{X_i} (Y|X_{-i}))}{V}$$

3. Calcul des indices de Sobol

- Variance-based Sensitivity Analysis : indices de Sobol

$$S_i = \frac{V_i}{V} = \frac{V_{X_i}(E_{X_{-i}}(Y|X_i))}{V}$$

$$S_{Ti} = 1 - \frac{V_{X_{-i}}(E_{X_i}(Y|X_{-i}))}{V}$$

- Estimateurs proposés par [Tarantola & Lilburne, 2009]

$$\hat{S}_i^I = \frac{y_A \cdot y_C - y_A \cdot y_B}{y_A \cdot y_A - y_A \cdot y_B}$$

$$\hat{S}_i^{II} = \frac{y_A \cdot y_C - y_A \cdot y_B}{y_B \cdot y_B - y_A \cdot y_B}$$

$$\hat{S}_i^{III} = \frac{y_A \cdot y_C - y_C \cdot y_D}{y_C \cdot y_C - y_C \cdot y_D}$$

$$\hat{S}_i^{IV} = \frac{y_A \cdot y_C - y_C \cdot y_D}{y_D \cdot y_D - y_C \cdot y_D}$$

- Précision estimée par bootstrap (100 répliques)

$$\hat{S}_i^V = \frac{y_B \cdot y_D - y_A \cdot y_B}{y_A \cdot y_A - y_A \cdot y_B}$$

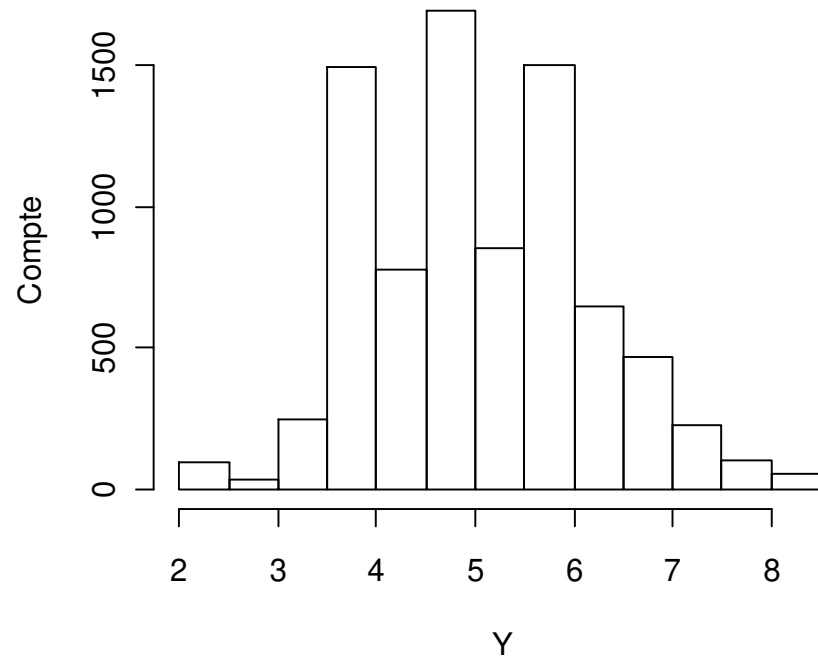
$$\hat{S}_i^{VI} = \frac{y_B \cdot y_D - y_A \cdot y_B}{y_B \cdot y_B - y_A \cdot y_B}$$

$$\hat{S}_i^{VII} = \frac{y_B \cdot y_D - y_C \cdot y_D}{y_C \cdot y_C - y_C \cdot y_D}$$

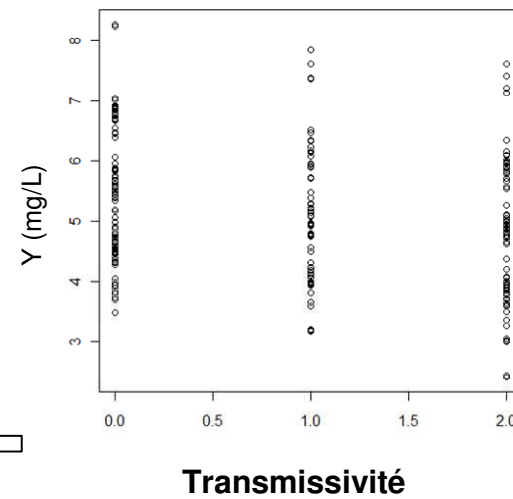
$$\hat{S}_i^{VIII} = \frac{y_B \cdot y_D - y_C \cdot y_D}{y_D \cdot y_D - y_C \cdot y_D}$$

Résultats

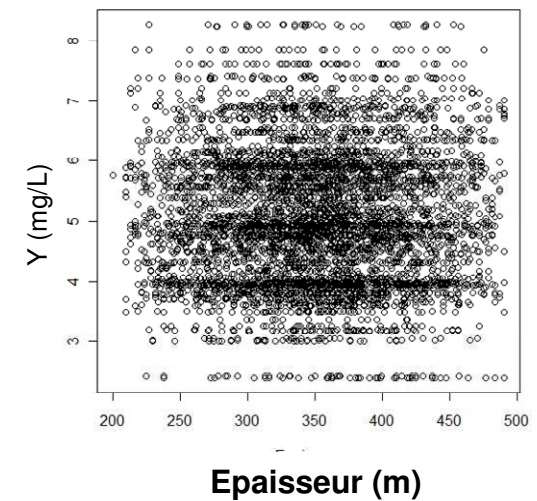
Distribution de Y



Influence de la Transmissivité



Influence de l'épaisseur

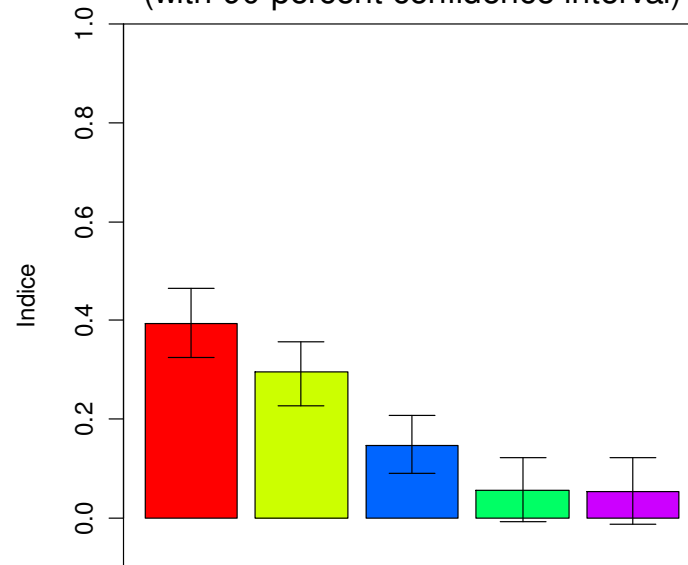


- Histogramme empirique de Y
- Scatterplots : discussion graphique sur l'influence de chaque paramètre
- Attention : pas de notion d'ordre dans les valeurs des X_i !

Résultats

First order sensitivity indices

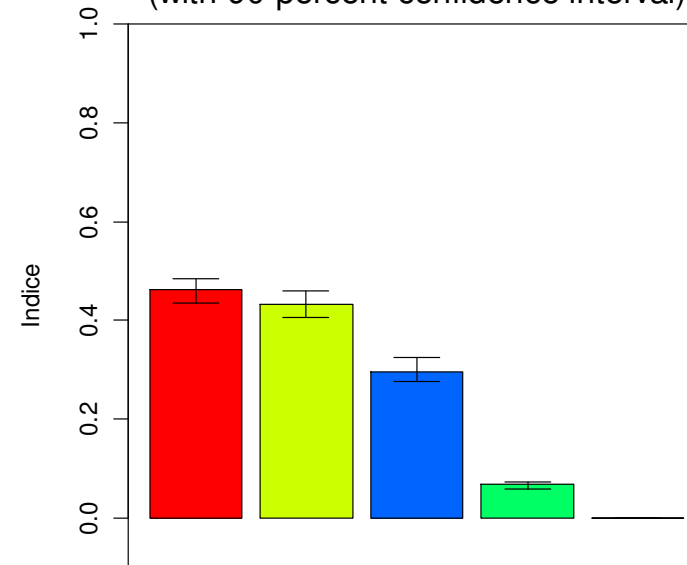
(with 90 percent confidence interval)



Sol Occ.Sol Trans. Precip. Epaiss.
0.3931 0.2972 0.1461 0.05635 0.05504

Total order sensitivity indices

(with 90 percent confidence interval)

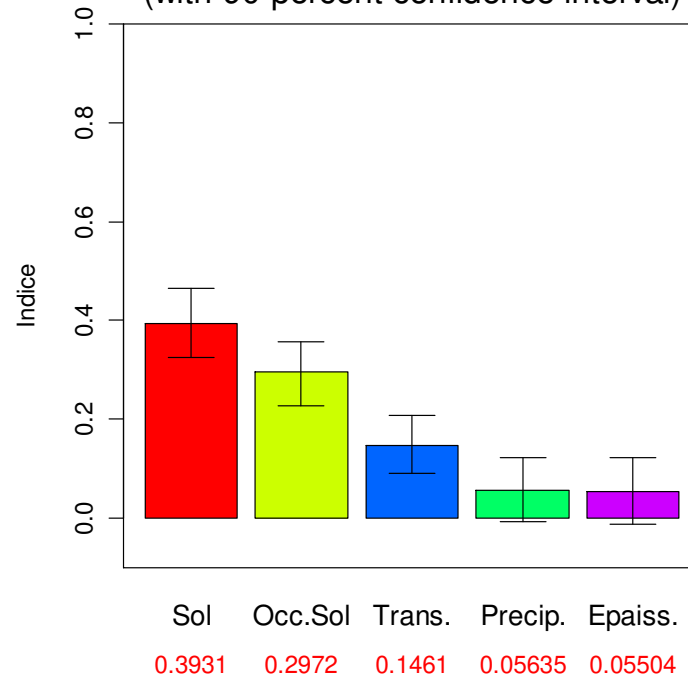


Sol Occ.Sol Trans. Precip. Epaiss.
0.4613 0.4321 0.2961 0.06765 0

Résultats

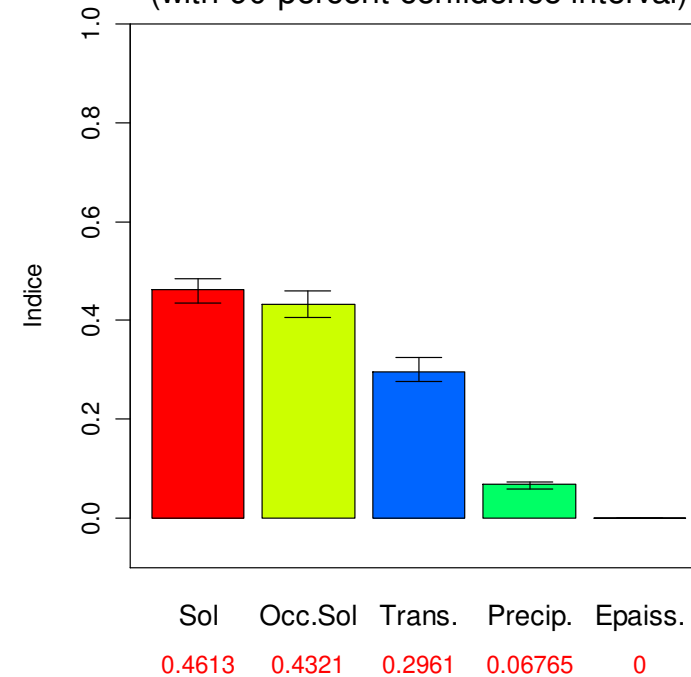
First order sensitivity indices

(with 90 percent confidence interval)



Total order sensitivity indices

(with 90 percent confidence interval)

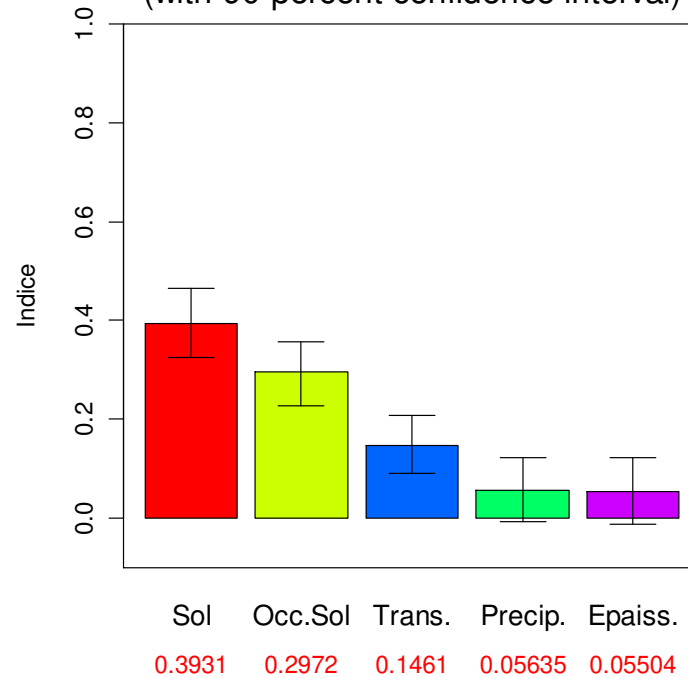


- Précision des indices suffisante pour tirer des conclusions

Résultats

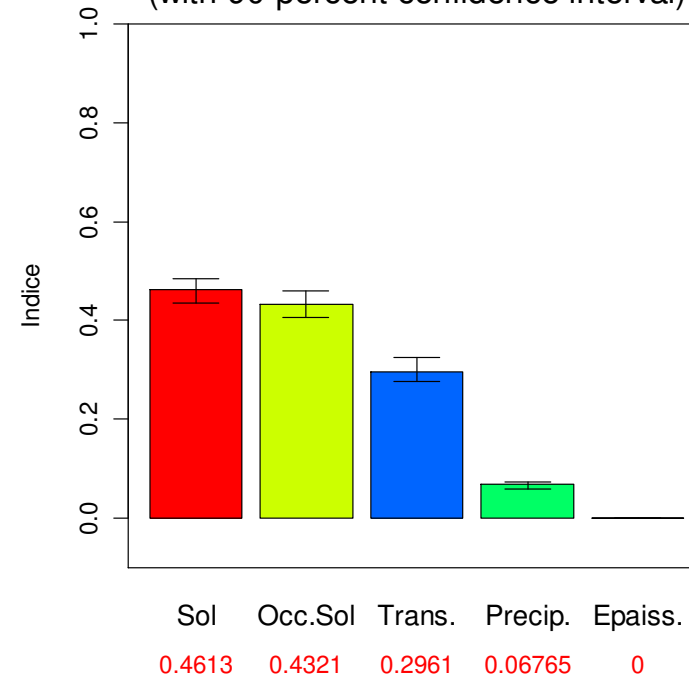
First order sensitivity indices

(with 90 percent confidence interval)



Total order sensitivity indices

(with 90 percent confidence interval)

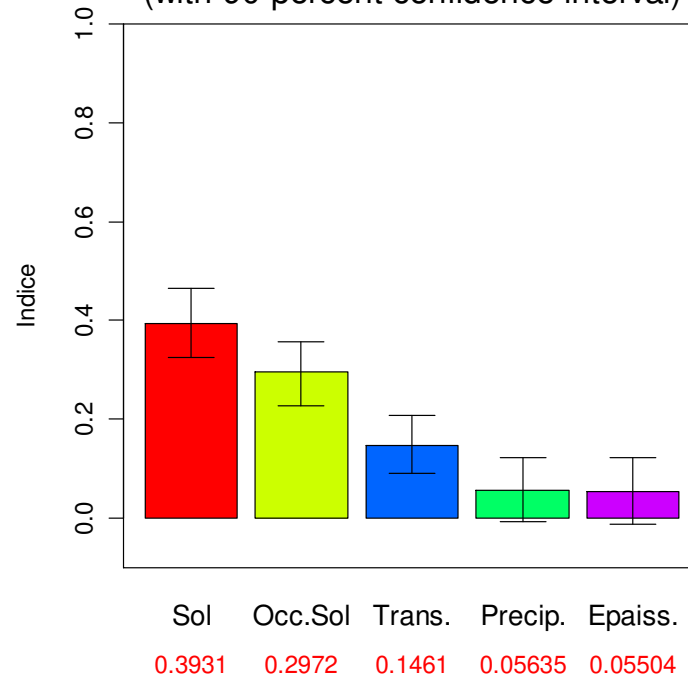


- Précision des indices suffisante pour tirer des conclusions
- Influence prépondérante de deux facteurs : nature et occupation du sol

Résultats

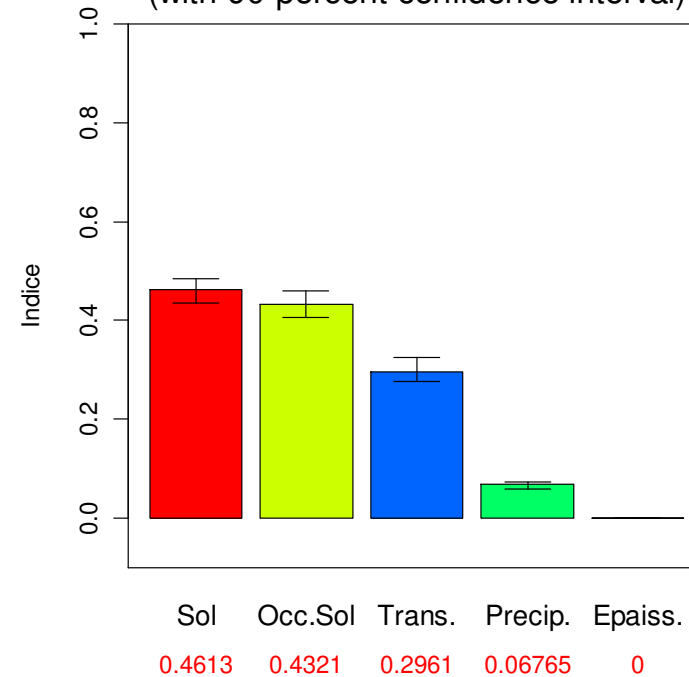
First order sensitivity indices

(with 90 percent confidence interval)



Total order sensitivity indices

(with 90 percent confidence interval)



- Précision des indices suffisante pour tirer des conclusions
- Influence prépondérante de deux facteurs : nature et occupation du sol
- Deux facteurs négligeables : épaisseur de l'aquifère et carte des précipitations

Résultats : interactions entre variables

Résultats : interactions entre variables

- Estimation de la part des interactions entre X_i dans la variance de Y

$$\sum_i s_i = 95 \%$$

Résultats : interactions entre variables

- Estimation de la part des interactions entre X_i dans la variance de Y

$$\sum_i \mathbf{s}_i = 95 \% \quad \longrightarrow \quad \textit{Approximation linéaire possible ?}$$

Résultats : interactions entre variables

- Estimation de la part des interactions entre X_i dans la variance de Y

$$\sum_i s_i = 95 \% \rightarrow \text{Approximation linéaire possible ?}$$

- Estimation des interactions pour chaque variable X_i : $ST_i - S_i$

	S_i	ST_i	$ST_i - S_i$
Nature du sol	0.39	0.46	0.07
Occupation du sol	0.30	0.43	0.13
Précipitations	0.06	0.07	0.01
Transmissivité	0.15	0.30	0.15
Profondeur	0.06	0.00	-0.06

Difficilement exploitable
du fait des larges
intervalles de confiance

Approfondissements

Approfondissements

- D'autres sorties d'intérêt **Y**
 - Indices de sensibilité en plusieurs points de la zone d'étude
 - Autres indicateurs synthétiques : max, moyenne, seuil...
 - Indices de sensibilité très différents !
 - ➔ Nécessité de bien définir la fonction objectif

Approfondissements


- D'autres sorties d'intérêt Y

- Indices de sensibilité en plusieurs points de la zone d'étude
- Autres indicateurs synthétiques : max, moyenne, seuil...
- Indices de sensibilité très différents !

 Nécessité de bien définir la fonction objectif

- Influence des tailles n_i des jeux de simulations

- Besoin de représenter correctement la variabilité de chaque carte
- Résultats non robustes pour n_i trop petits

 n_i doivent être « suffisamment grands »

Approfondissements

- D'autres sorties d'intérêt Y

- Indices de sensibilité en plusieurs points de la zone d'étude
- Autres indicateurs synthétiques : max, moyenne, seuil...
- Indices de sensibilité très différents !



Nécessité de bien définir la fonction objectif

- Influence des tailles n_i des jeux de simulations

- Besoin de représenter correctement la variabilité de chaque carte
- Résultats non robustes pour n_i trop petits



n_i doivent être « suffisamment grands »

- D'autres modèles d'incertitude

- Forme de la distribution de X_5 (scalaire) : impact négligeable
- Différents modèles géostatistiques pour la carte des précipitations : impact faible

Approfondissements

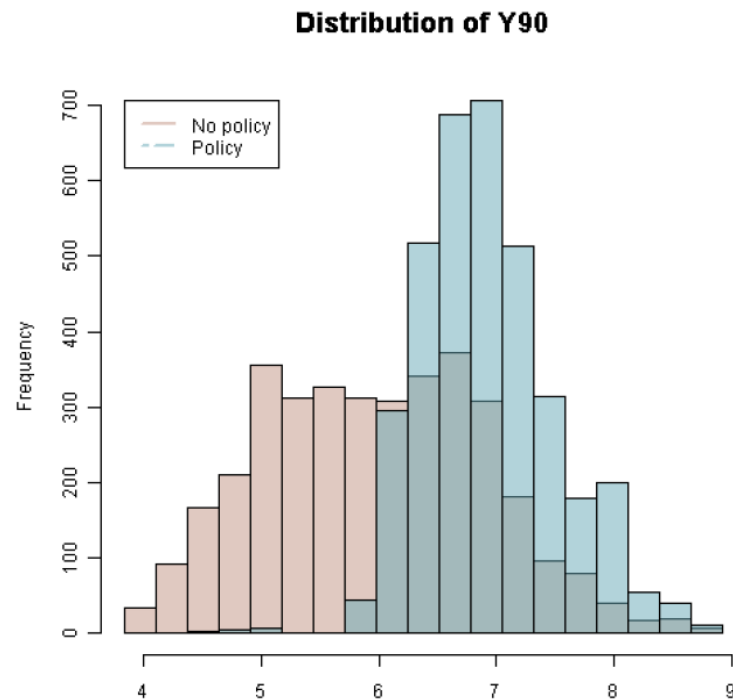
- Comparaison de deux politiques d'occupation du sol
- *Est-il possible de distinguer les deux politiques malgré la présence de nombreuses incertitudes ?*

Approfondissements

- Comparaison de deux politiques d'occupation du sol
- *Est-il possible de distinguer les deux politiques malgré la présence de nombreuses incertitudes ?*
- Variable binaire : « Policy switch »

Approfondissements

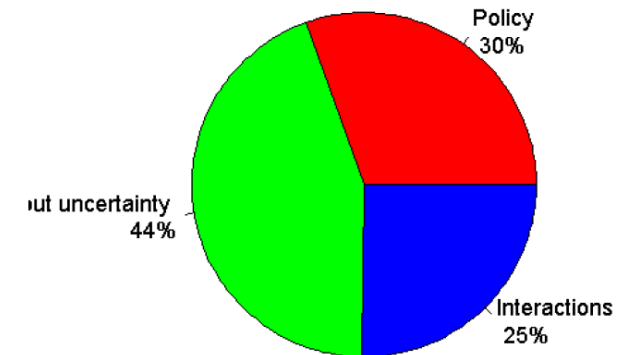
- Comparaison de deux politiques d'occupation du sol
- *Est-il possible de distinguer les deux politiques malgré la présence de nombreuses incertitudes ?*
- Variable binaire : « Policy switch »



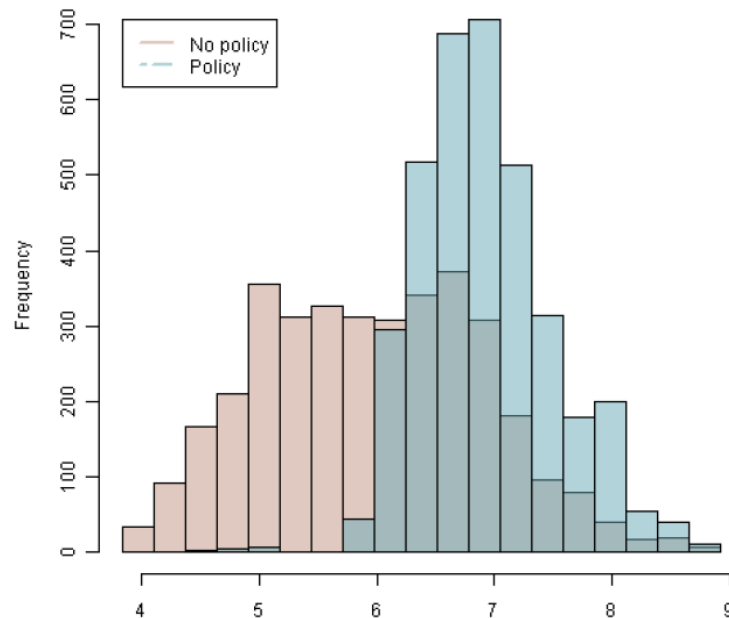
Approfondissements

- Comparaison de deux politiques d'occupation du sol
- *Est-il possible de distinguer les deux politiques malgré la présence de nombreuses incertitudes ?*
- Variable binaire : « Policy switch »

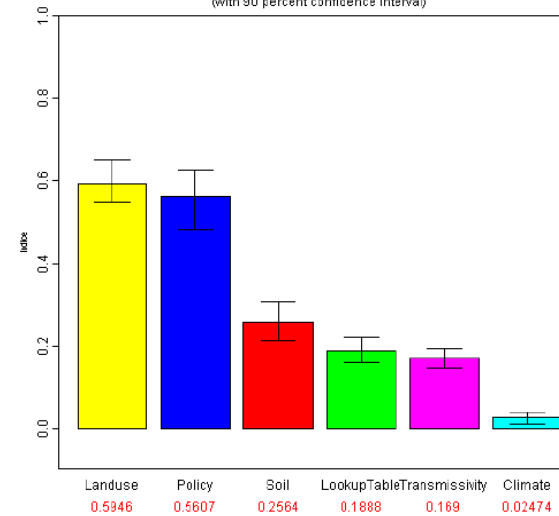
Variance of Y90 explained



Distribution of Y90



Total order sensitivity indices
(with 90 percent confidence intervals)



Approfondissements : à venir

Approfondissements : à venir

- Variables locales/globales et effets d'échelle
 - Influence de la taille de la zone d'étude sur la valeur des indices de sensibilité
 - Passage local/global

Approfondissements : à venir

- Variables locales/globales et effets d'échelle
 - Influence de la taille de la zone d'étude sur la valeur des indices de sensibilité
 - Passage local/global

- Géostatistique
 - Meilleure stratégie d'échantillonnage pour les simulations géostatistiques ?
 - Meilleure convergence des indices de sensibilité ?

Approfondissements : à venir

- Variables locales/globales et effets d'échelle
 - Influence de la taille de la zone d'étude sur la valeur des indices de sensibilité
 - Passage local/global

- Géostatistique
 - Meilleure stratégie d'échantillonnage pour les simulations géostatistiques ?
 - Meilleure convergence des indices de sensibilité ?

- Sortie spatialisée

Conclusion

Conclusion

- Intérêts de la démarche

- Pas de contraintes sur les modèles d'incertitude utilisés
- Généralisable à d'autres types de variables d'entrée (tabulaire, fonctionnelle, etc...)
- Mêmes estimateurs des indices de Sobol que pour une AS « scalaire »
- Éléments qualitatifs sur les interactions entre X_i
- Peut rendre compte d'effets fortement non-linéaires
- Facile à mettre en œuvre (bibliothèque « Sensitivity » sous R, SimLab...)

Conclusion

- Intérêts de la démarche

- Pas de contraintes sur les modèles d'incertitude utilisés
- Généralisable à d'autres types de variables d'entrée (tabulaire, fonctionnelle, etc...)
- Mêmes estimateurs des indices de Sobol que pour une AS « scalaire »
- Éléments qualitatifs sur les interactions entre X_i
- Peut rendre compte d'effets fortement non-linéaires
- Facile à mettre en œuvre (bibliothèque « Sensitivity » sous R, SimLab...)

- Limites

- Coût CPU élevé
- Biais introduit par le pré-échantillonnage des réalisations aléatoires de cartes ? (possibilité de simuler « à la volée » ?)

Comparaison

Méthode	Méta-modèle « Joint GAM »	Macro paramètre	Paramètre « Trigger »	Indices de cartes
Coût CPU	FAIBLE	FORT	FORT	FORT
Modèles d'incertitude complexes ?	OUI	NON	OUI	OUI
Indices de Sobol	Variables scalaires : S_i et ST_i Variable spatialisée : ST_i	S_i et ST_i	S_i et ST_i	S_i et ST_i
Plusieurs variables spatialisées ?	NON	OUI	OUI	OUI
Description des interactions entre X_i	Qualitatif (présence des X_i scalaires dans le modèle de dispersion)	Qualitatif (comparaison S_i / ST_i pour chaque variable)	Qualitatif (comparaison S_i / ST_i pour chaque variable)	Qualitatif (comparaison S_i / ST_i pour chaque variable)

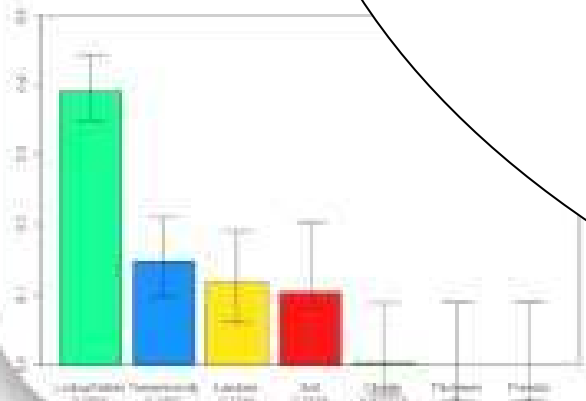
Bibliographie

- Fisher, P. F. (1991), 'Modelling soil map-unit inclusions by Monte Carlo simulation', *International Journal of Geographical Information Science* **5**(2), 193-208.
- Iooss, B. & Ribatet, M. (2009), 'Global sensitivity analysis of computer models with functional inputs', *Reliability Engineering and System Safety* **94**, 1194 - 1204.
- Lilburne, L. & Tarantola, S. (2009), 'Sensitivity analysis of spatial models', *International Journal of Geographical Information Science* **23:2**, 151-168.
- Ruffo, P.; Bazzana, L.; Consonni, A.; Corradi, A.; Saltelli, A. & Tarantola, S. (2006), 'Hydrocarbon exploration risk evaluation through uncertainty and sensitivity analyses techniques', *Reliability Engineering and System Safety* **91**(10-11), 1155-1162.
- Saltelli, A.; Ratto, M.; Andres, T.; Campolongo, F.; Cariboni, J.; Gatelli, D.; Saisana, M. & Tarantola, S. (2008), *Global Sensitivity Analysis - The Primer*, Wiley.
- Tarantola, S. (2008), 'Global Sensitivity Analysis of complex models', Seminar Series 2008, Centre for Process Systems Engineering Imperial College London.

Nathalie SAINT-GEOURS
Doctorante UMR TETIS

Merci pour
votre attention

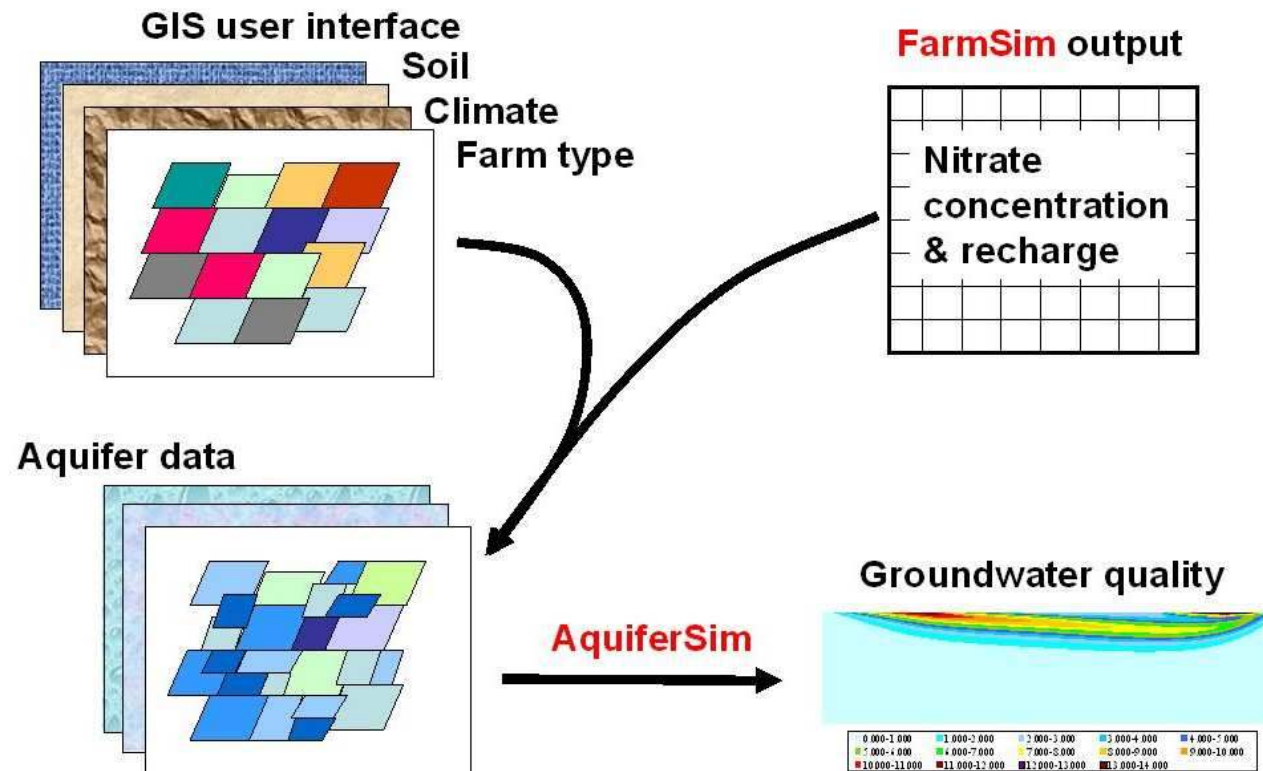
CEA Cadarache
12 janvier 2010



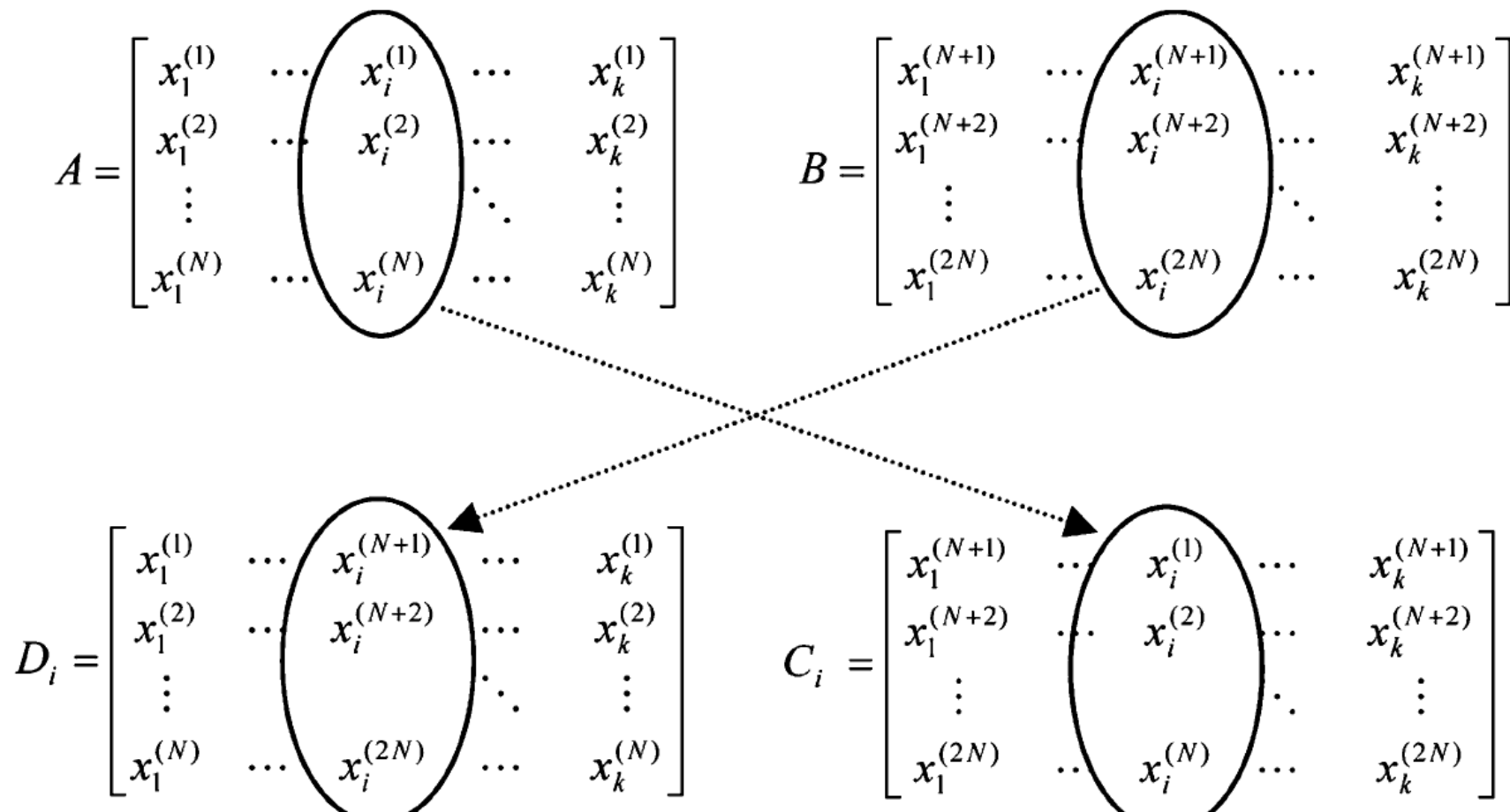
AquiferSim

- Contamination des eaux souterraines par la pollution agricole diffuse

How data and **models** are combined



Indices de Sobol



Permutation des échantillons pour la $i^{\text{ème}}$ variable d'entrée

Indices de Sobol

Estimateurs des indices de premier ordre :

$$\hat{S}_i^I = \frac{y_A \cdot y_C - y_A \cdot y_B}{y_A \cdot y_A - y_A \cdot y_B}$$

$$\hat{S}_i^{II} = \frac{y_A \cdot y_C - y_A \cdot y_B}{y_B \cdot y_B - y_A \cdot y_B}$$

$$\hat{S}_i^{III} = \frac{y_A \cdot y_C - y_C \cdot y_D}{y_C \cdot y_C - y_C \cdot y_D}$$

$$\hat{S}_i^{IV} = \frac{y_A \cdot y_C - y_C \cdot y_D}{y_D \cdot y_D - y_C \cdot y_D}$$

$$\hat{S}_i^V = \frac{y_B \cdot y_D - y_A \cdot y_B}{y_A \cdot y_A - y_A \cdot y_B}$$

$$\hat{S}_i^{VI} = \frac{y_B \cdot y_D - y_A \cdot y_B}{y_B \cdot y_B - y_A \cdot y_B}$$

$$\hat{S}_i^{VII} = \frac{y_B \cdot y_D - y_C \cdot y_D}{y_C \cdot y_C - y_C \cdot y_D}$$

$$\hat{S}_i^{VIII} = \frac{y_B \cdot y_D - y_C \cdot y_D}{y_D \cdot y_D - y_C \cdot y_D}$$

Indices de Sobol

Estimateurs des indices totaux :

$$\hat{S}_{Ti}^I = \frac{y_A \cdot y_A - y_A \cdot y_D}{y_A \cdot y_A - L \times \left(\frac{1}{L} \sum y_A \right)^2}$$

$$\hat{S}_{Ti}^{II} = \frac{y_D \cdot y_D - y_A \cdot y_D}{y_D \cdot y_D - L \times \left(\frac{1}{L} \sum y_D \right)^2}$$

$$\hat{S}_{Ti}^{III} = \frac{y_B \cdot y_B - y_B \cdot y_C}{y_B \cdot y_B - L \times \left(\frac{1}{L} \sum y_B \right)^2}$$

$$\hat{S}_{Ti}^{IV} = \frac{y_C \cdot y_C - y_B \cdot y_C}{y_C \cdot y_C - L \times \left(\frac{1}{L} \sum y_C \right)^2}$$

Réalisations aléatoires « à la volée »

Echantillon A

X4
5,36
2,1
4,59
7,21
8,12
4,61
5,09
4,85


Echantillon B

X4
5,12
8,38
5,89
3,27
2,36
5,87
5,39
5,63

Réalisations aléatoires « à la volée »

Echantillon A

X4
5.36
2,1
4,59
7,21
8,12
4,61
5,09
4,85



 Y_A

Echantillon B



X4
5,12
8,38
5,89
3,27
2,36
5,87
5,39
5,63

Réalisations aléatoires « à la volée »

Echantillon A



X4			Y_A
5.36			
2,1			
4,59			
7,21			
8,12			
4,61			
5,09			
4,85			

Echantillon B



X4			Y_B
5.12			
8,38			
5,89			
3,27			
2,36			
5,87			
5,39			
5,63			


Réalisations aléatoires « à la volée »

Echantillon A

X4			Y_A
5,36			
2,1			
4,59			
7,21			
8,12			
4,61			
5,09			
4,85			



Echantillon B

X4			Y_B
5,12			
8,38			
5,89			
3,27			
2,36			
5,87			
5,39			
5,63			



5,12			Y_C
------	------------------------------------------------------------------------------------	------------------------------------------------------------------------------------	-------

Réalisations aléatoires « à la volée »



Echantillon A

X4	Y _A
5,36	 
2,1	
4,59	
7,21	
8,12	
4,61	
5,09	
4,85	

Echantillon B



X4	Y _B
5,12	 
8,38	
5,89	
3,27	
2,36	
5,87	
5,39	
5,63	

5,12	 
------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------


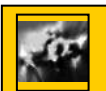
5,12	 
------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Réalisations aléatoires « à la volée »



Echantillon A

X4	Y _A
5,36	 
2,1	
4,59	
7,21	
8,12	
4,61	
5,09	
4,85	

Echantillon B

X4	Y _B
5,12	 
8,38	
5,89	
3,27	
2,36	
5,87	
5,39	
5,63	

5,12	 
------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------

5,12	 
------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------

- Implémentation plus fastidieuse
- Pas de biais lié au pré-échantillonnage
- Pas de stockage de jeux de simulations
- Temps de calcul plus élevé