

# Modélisation Structure-Fonction des Plantes

## Méthodes d'Analyse de sensibilité

Robert Faivre

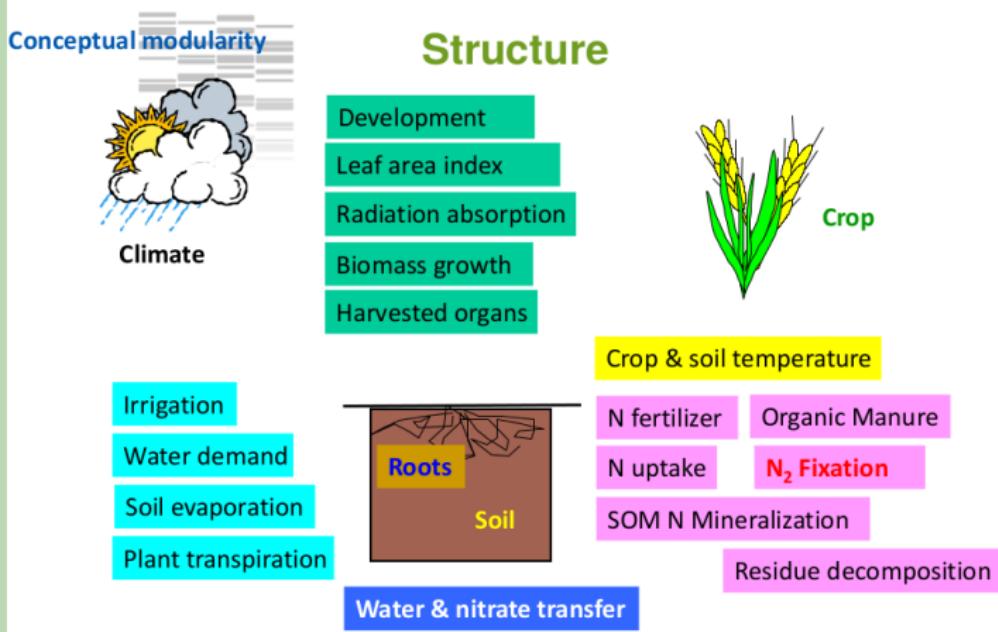
INRA - Mathématiques et Informatique Appliquées - Université Fédérale de Toulouse



Montpellier, 16 mars 2018

# Quelques types de modèles utilisés en AgroBioSciences

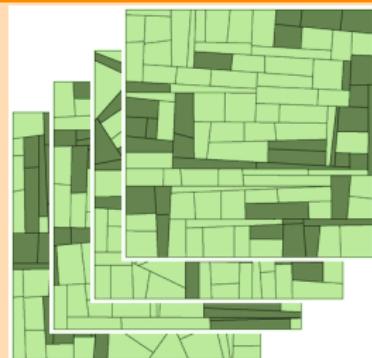
## Modèles de culture



### ► STICS - Simulateur multidisciplinaire pour les cultures standard

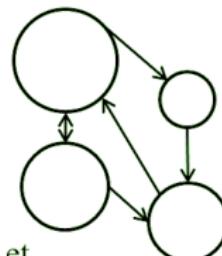
# Quelques types de modèles en AgroBioSciences

## Modèles épidémiologiques sur paysage

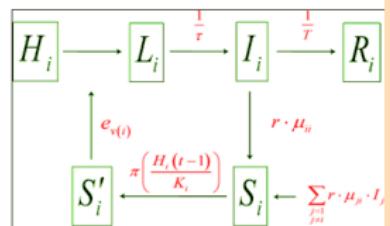


simulateur de paysages

algorithme Califlopp  
découplage des échelles de gestion et de dynamique de populations



modèle de dynamique de populations



- CaliFloPP - Calculation of the Integrated Flow of Particles between Polygons (Adamczyk et al.)

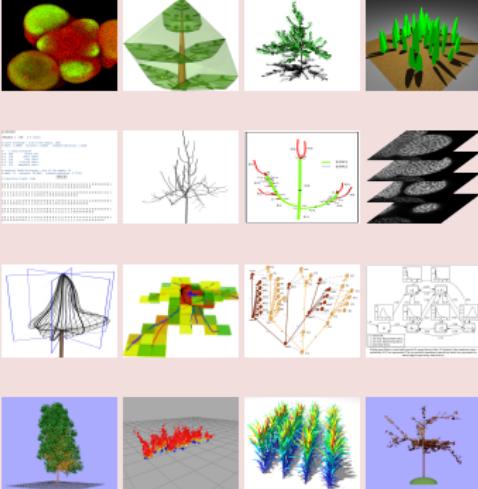
# Quelques types de modèles en AgroBioSciences

## Modèles structure-fonction d'architecture de plantes

OpenAlea in a nutshell Applications Conclusions

### Background: plant modelling at a glance

- Different scales:
  - cell
  - branch
  - tree
  - forest
- Heterogeneous data:
  - raw data
  - digitized data
  - tree databases
  - 3D images
- Many tools required:
  - topology
  - Geometry
  - simulation
- Many models possible:
  - theoretical
  - mechanical
  - probabilist



Thomas Cokelaer IHC 2010, Lisbon, 25 August

- ▶ OpenAlea - Software Environment for Plant Modeling (Pradal et al.)

# Spécificités des modèles en AgroBioSciences

- ▶ Un grand nombre de processus donc de paramètres ( $\text{Stics} \approx 400$ ), certains corrélés
- ▶ Des processus dynamiques (équations différentielles ou aux différences) ou spatiaux
- ▶ Des entrées temporelles (météo par exemple) ou spatiales (paysage)
- ▶ Des pratiques culturales adaptatives (optimisées parfois)
- ▶ Des fonctions à seuil (stress, phénologie)
- ▶ Des processus parfois stochastiques (diffusion de spores, position/orientation d'organes)
- ▶ Temps de calcul modèle-dépendant ( $< 1s$  à  $1h$  voire plus)
- ▶ Un grand nombre de variables de sortie

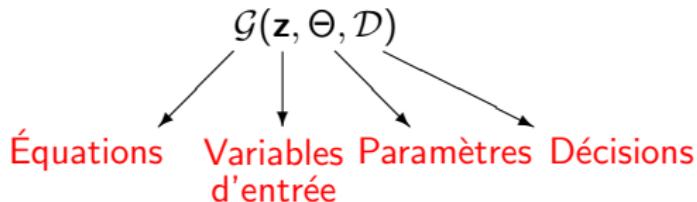
# Grandes questions

- ▶ Estimation - Calibration des paramètres (optimisation)
- ▶ Qualité prédictive (analyse d'incertitude)
- ▶ Simplification du modèle (analyse de sensibilité, criblage, métamodélisation)
- ▶ Conception de pratiques (optimisation)
- ▶ Couplage de modèles (sensibilité, criblage)

# Principes des analyses d'incertitude et de sensibilité

## Définition et Objectifs

- ▶ Sources d'incertitude dans un modèle



- ▶ Types d'incertitude
  - ▶ Manque de connaissance  
ex: équation identifiée empiriquement sur un jeu de données
  - ▶ Variabilité de l'environnement du système  
ex: " température moyenne journalière " variable entre années
  - ▶ Imprécision  
ex: paramètre de croissance estimé sur un jeu de données
  - ▶ Règle de décision mal maîtrisée  
ex: seuil de déclenchement de l'irrigation

# Principes des analyses d'incertitude et de sensibilité

## Notations

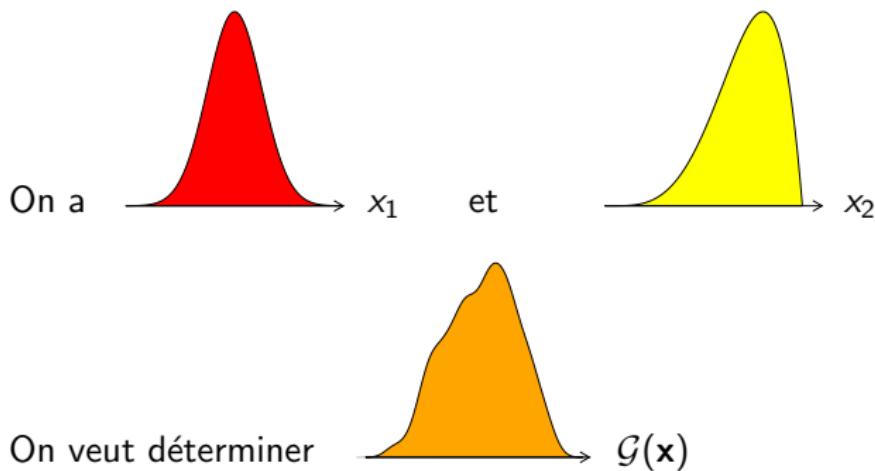
$$\begin{aligned}\mathbf{y} &= \mathcal{G}(\mathbf{x}) \\ &= \mathcal{G}(x_1, \dots, x_K)\end{aligned}$$

- ▶  $x_1, \dots, x_K$  = variables d'entrée et paramètres incertains
- ▶  $x_1, \dots, x_K$  les  $K$  facteurs incertains
- ▶  $\mathbf{x} = (x_1, \dots, x_K) \in \mathbb{R}^K$
- ▶  $\mathbf{y}$  peut être multidimensionnel

# Principes de l'analyse d'incertitude

## Objectif

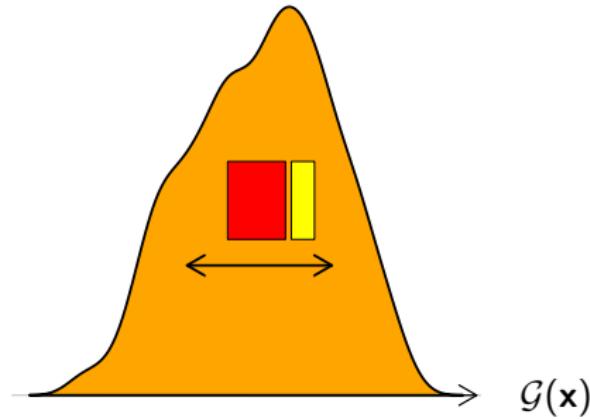
- ▶ Répondre à la question suivante
  - ▶ Quel est le niveau d'incertitude de  $\mathcal{G}(\mathbf{x})$  qui résulte de l'incertitude dans les  $x_k$  ?



# Principes de l'analyse de sensibilité

## Objectif

- ▶ Répondre à la question
  - ▶ Quelles sont les principales sources d'incertitude parmi  $x_1, \dots, x_K$  qui influencent  $\mathcal{G}(\mathbf{x})$  ?



Variance de  $\mathcal{G}(\mathbf{x}) = \text{Effet de } x_1 + \text{Effet de } x_2 + \dots$

## Analyse d'incertitude

- ▶ donner des informations sur l'incertitude associée aux prédictions d'un modèle
- ▶ optimiser des variables décisionnelles pour réduire l'incertitude

## Analyse de sensibilité

- ▶ identifier les paramètres et les variables d'entrée qui ont une forte influence sur les sorties d'un modèle  
→ **Important de les connaître avec précision**
- ▶ identifier les paramètres et les variables d'entrée qui ont une influence moindre sur les sorties  
→ **Moins important de les connaître avec précision**
- ▶ analyser le comportement du modèle

# 1. Analyse d'incertitude

# Analyse d'incertitude

## Un exemple très simple

- ▶ Équation :  $\mathcal{G}(x_1, x_2) = x_1 + 2x_2$
- ▶ Incertitude sur  $x_1$  et  $x_2$  :  $x_1 \sim \mathcal{N}(20, 16)$  et  $x_2 \sim \mathcal{N}(60, 64)$

Si  $X_1$  et  $X_2$  sont deux variables indépendantes de distribution gaussienne alors  $\alpha X_1 + \beta X_2$  suit une distribution gaussienne

$$\begin{aligned} E(\alpha X_1 + \beta X_2) &= \alpha E(X_1) + \beta E(X_2) \\ \text{Var}(\alpha X_1 + \beta X_2) &= \alpha^2 \text{Var}(X_1) + \beta^2 \text{Var}(X_2) \end{aligned}$$

$$\mathcal{G}(x_1, x_2) \sim \mathcal{N}(140, 272)$$

# Analyse d'incertitude

En général, c'est plus dur !

- ▶ Equations plus complexes, relation non linéaire entre  $\mathcal{G}(x)$  et  $x$   
→ Impossible de déterminer l'expression analytique de la distribution de  $\mathcal{G}(x)$
- ▶ La distribution de  $x$  n'est pas toujours connue  
→ Choix subjectif
- ▶ Temps de calcul parfois long avec certains modèles  
→ Le nombre de simulations peut être limitant

# Analyse d'incertitude

## Procédure en quatre étapes

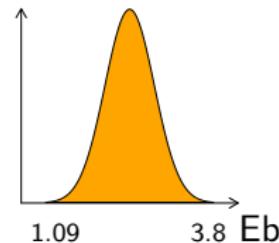
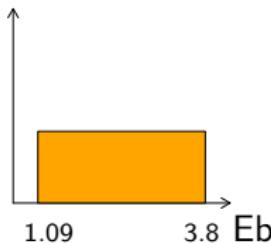
1. Définir les distributions de  $x_1, \dots, x_K$
2. Générer des échantillons à partir des distributions définies à l'étape 1
3. Calculer  $\mathcal{G}(\mathbf{x})$  pour chaque série de  $x_1, \dots, x_K$  générée
4. Estimer la distribution de  $\mathcal{G}(\mathbf{x})$

## Étape 1 - Définition des distributions

Les distributions de probabilité des facteurs incertains (paramètres ou variables d'entrée) peuvent être définies en utilisant :

- ▶ La littérature scientifique et l'expertise
- ▶ Des séries de mesures (série climatique ...)
- ▶ Les valeurs des paramètres estimées

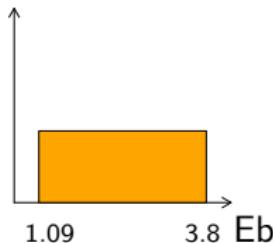
Exemple : d'après un article publié par Jeuffroy et Recous en 1999 dans EJA, l'efficacité d'utilisation de rayonnement intercepté varie entre 1.09 et 3.8 g.MJ<sup>-1</sup> pour le blé



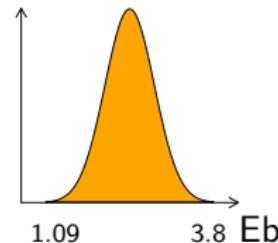
# Analyse d'incertitude

## Étape 2 - Génération d'échantillons à partir des distributions de $x_1, x_2, \dots, x_K$

- Il faut générer suffisamment de valeurs de  $x_1, x_2, \dots, x_K$
- Différentes méthodes d'échantillonnage peuvent être utilisées : échantillonnage aléatoire (Monte Carlo), échantillonnage en hypercube latin, ...
- En pratique, on utilise un logiciel pour générer  $N$  valeurs de  $x_1, x_2, \dots, x_K$  (ex:  $N = 20000$ ).



1.26 3.36 1.90 1.18 1.47  
2.05 2.44 1.30 2.86 2.48



1.83 2.23 2.01 2.44 2.60  
2.66 1.71 2.71 2.20 2.80

# Analyse d'incertitude

## Étape 3 - Calcul de $\mathcal{G}(\mathbf{x})$ pour chaque série $x_1, x_2, \dots, x_K$ générée

- ▶ Exécuter le modèle  $N$  fois, pour toutes les valeurs générées des facteurs d'entrée  $x_1, x_2, \dots, x_K$

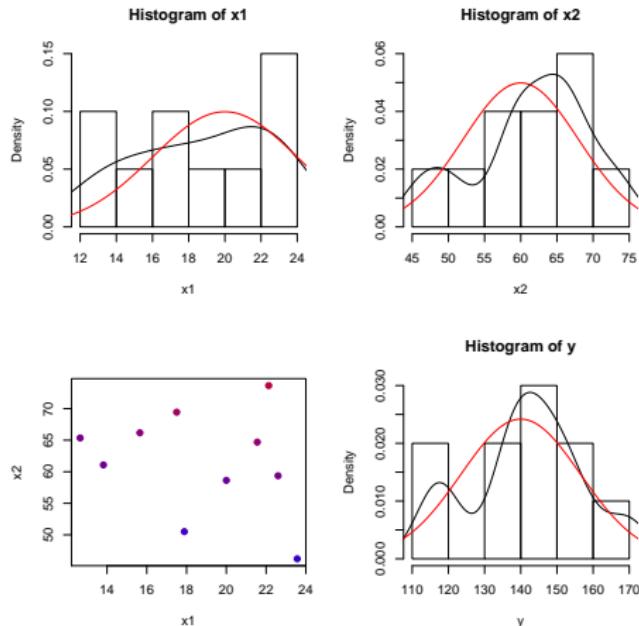
## Étape 4 - Approximation de la distribution de $\mathcal{G}(\mathbf{x})$

- ▶ Décrire les  $N$  valeurs de  $\mathcal{G}(\mathbf{x})$  calculées à l'étape 3
- ▶ Différentes approches possibles
  - ▶ calcul de la moyenne et de la variance,
  - ▶ calcul de quantiles (quartiles, déciles, ...),
  - ▶ histogramme,
  - ▶ fonction de distribution cumulée,
  - ▶ box plot...

# Analyse d'incertitude

## Application au modèle simple

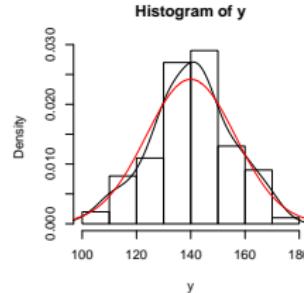
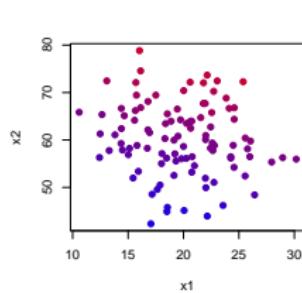
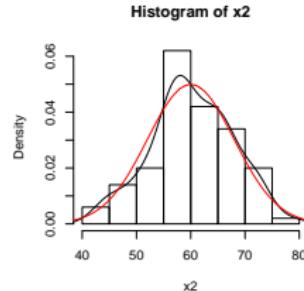
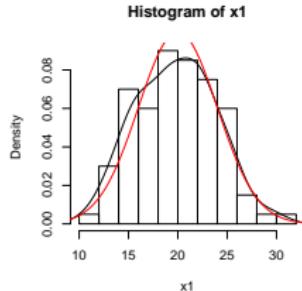
►  $N = 10$



# Analyse d'incertitude

## Application au modèle simple

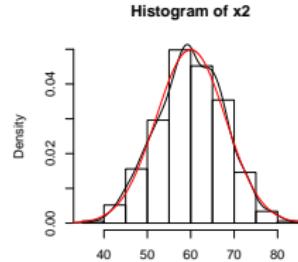
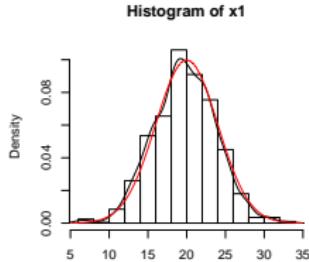
►  $N = 100$



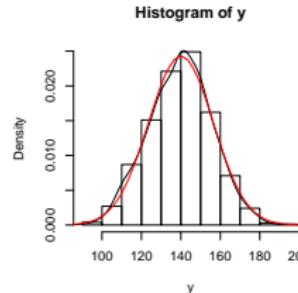
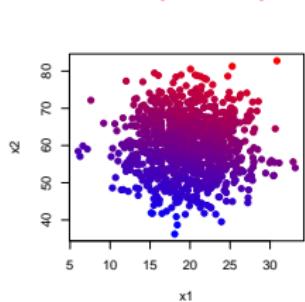
# Analyse d'incertitude

## Application au modèle simple

►  $N = 1000$



La précision ne dépend que de  $N$  et non de la dimension  $K$



## 2. Analyse de sensibilité

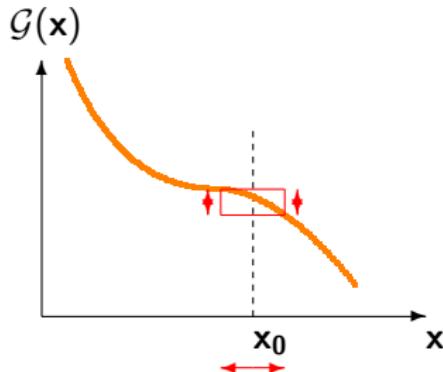
# Analyse de sensibilité

## Question

- ▶ Quelles sont les principales sources de variation qui influencent  $\mathcal{G}(x)$  ?

### Analyse de sensibilité locale

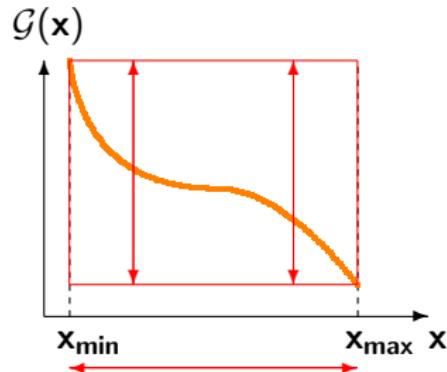
Variation de  $\mathcal{G}(x)$  autour de  $x_0$



basée sur le calcul de dérivées

### Analyse de sensibilité globale

Variation globale de  $\mathcal{G}(x)$  quand  $x$  varie dans son domaine d'incertitude



basée sur le calcul d'indices

# Analyse de sensibilité globale

Elle consiste à

- ▶ Définir des indices de sensibilité
- ▶ Calculer ces indices en faisant varier les facteurs incertains  $x_1, \dots, x_K$  sur leurs domaines

Procédure en quatre étapes

1. Définir les distributions de  $x_1, \dots, x_K$
2. Générer des échantillons à partir des distributions définies à l'étape 1
3. Calculer  $\mathcal{G}(\mathbf{x})$  pour chaque série de  $x_1, \dots, x_K$  générée
4. Calculer les indices de sensibilité

# Analyse de sensibilité globale

Indices basés sur une décomposition de la variance

$$\text{Var}[\mathcal{G}(\mathbf{x})] = \underbrace{\text{Var}_{x_1} + \text{Var}_{x_2} + \text{Var}_{x_3} + \dots}_{\begin{array}{c} \downarrow \\ \text{Variance totale} \\ \text{de la variable} \\ \text{de sortie} \end{array}} + \underbrace{\text{Var}_{x_1.x_2} + \text{Var}_{x_1.x_3} + \dots}_{\begin{array}{c} \downarrow \\ \text{Effets principaux des} \\ \text{facteurs incertains} \end{array}} + \underbrace{\dots}_{\begin{array}{c} \downarrow \\ \text{Termes d'interactions} \end{array}}$$

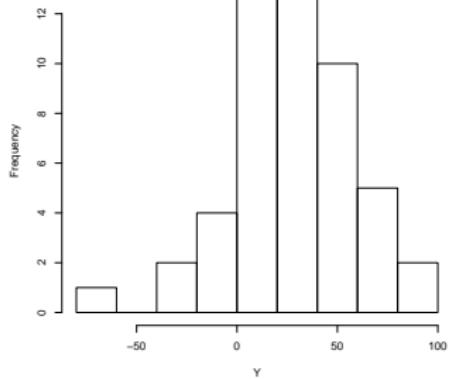
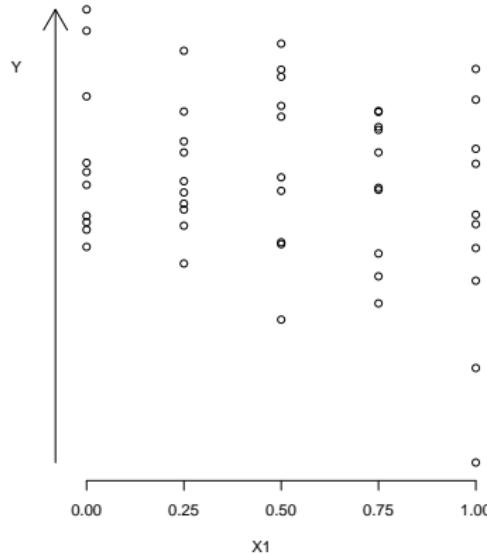
$$\text{Indice de premier ordre de } x_1 \text{ (SI}_1) = \frac{\text{Var}_{x_1}}{\text{Var}[\mathcal{G}(\mathbf{x})]}$$

$$\text{Indice de sensibilité totale de } x_i \text{ (TSI}_1) = \frac{\text{Var}_{x_1} + \text{Var}_{x_1.x_2} + \text{Var}_{x_1.x_3} + \dots}{\text{Var}[\mathcal{G}(\mathbf{x})]}$$

# Analyse de sensibilité globale

## Principe de la décomposition de la variance

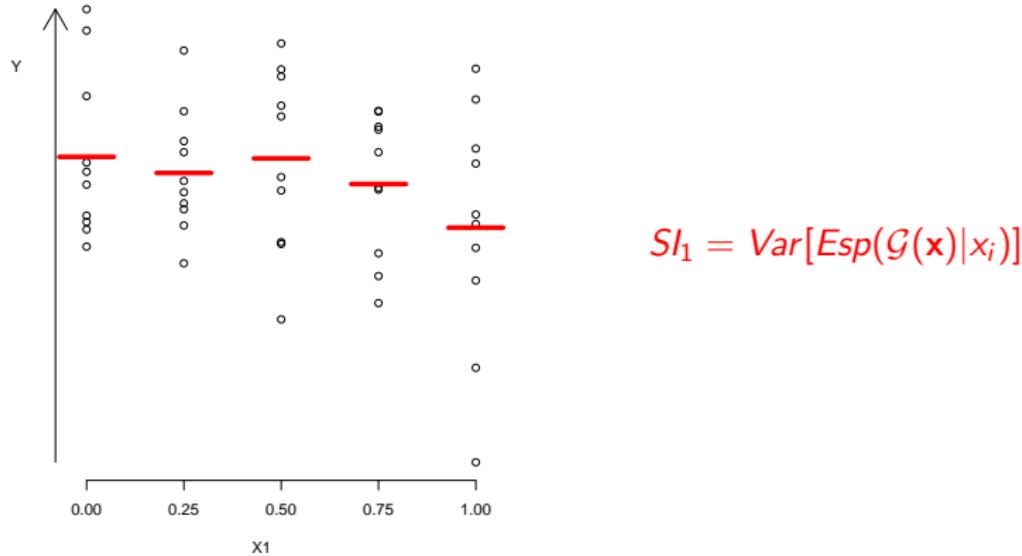
On dispose de mesures répétées sur 5 valeurs de  $x_1$



# Analyse de sensibilité globale

## Visualisation des indices

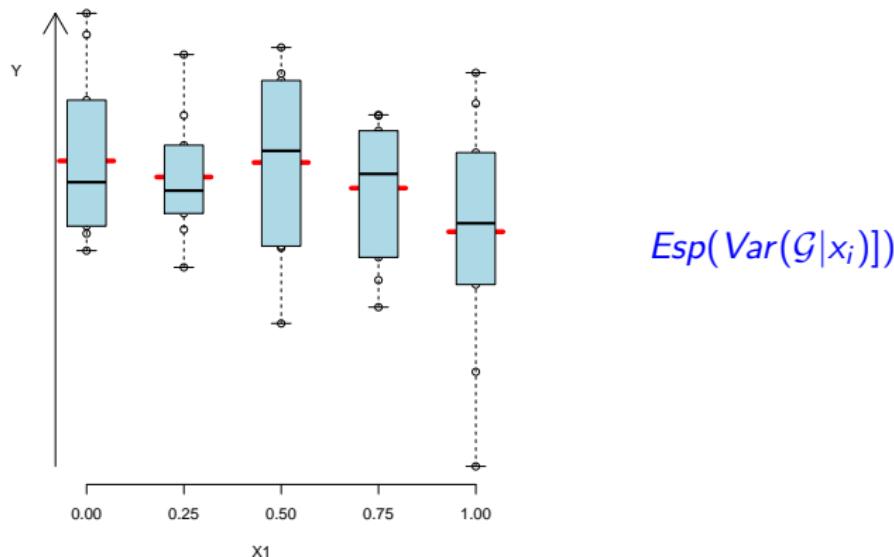
On calcule la variance des moyennes estimées sur les 5 points.



# Analyse de sensibilité globale

## Visualisation des indices

On calcule la moyenne des variances.



Théorème de la variance totale

$$Var[\mathcal{G}(\mathbf{x})] = Var[Esp(\mathcal{G}|x_i)] + Esp(Var(\mathcal{G}|x_i)))$$

# Analyse de sensibilité globale

## Autre formulation des indices

- ▶  $SI_i = \frac{Var[Esp(\mathcal{G}(\mathbf{x})|x_i)]}{Var[\mathcal{G}(\mathbf{x})]}$
- ▶  $TSI_i = \frac{Esp(Var[\mathcal{G}|\mathbf{x}^{(-i)}])}{Var[\mathcal{G}(\mathbf{x})]}$

## Interprétation

- ▶  $SI_i$  : pour chaque valeur du facteur  $x_i$ , on fait bouger les autres facteurs et on calcule les moyennes en chaque valeur. Ensuite, on regarde si la moyenne (espérance) bouge.
- ▶  $TSI_i$  : pour tous les n-uplets de valeurs de tous les facteurs  $(\mathbf{x}^{(-i)})$  sauf  $x_i$ , on fait bouger le facteur  $x_i$  et on calcule la variabilité. On en fait ensuite la moyenne.

→ Principe des méthodes d'analyse de variance sur plan factoriel ou des méthodes dites de Sobol' par échantillonnage alternatif, méthodes nécessitant beaucoup d'évaluations du modèle.

La précision dépend de N mais aussi de la dimension K

# Analyse de sensibilité globale

## Méthode par plan factoriel complet

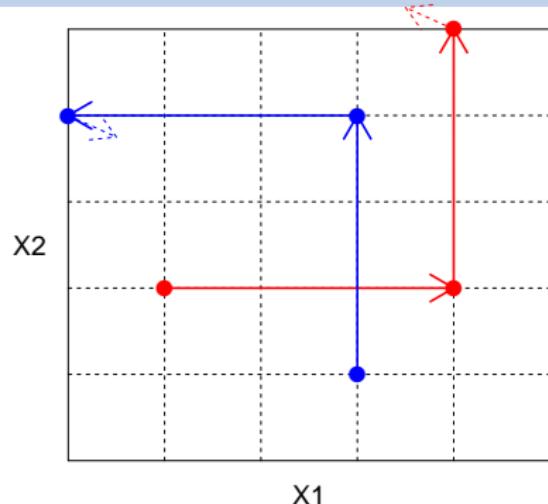
- ▶ chaque facteur est décomposé en un certain nombre de niveaux
  - ▶ plan factoriel complet : croisement de tous les niveaux
- 
- ▶ Prenons un modèle avec 20 paramètres (facteurs) incertains.
  - ▶ Pour un plan factoriel complet de 20 facteurs à 3 niveaux, cela nécessite  $3^{20} = 3\ 486\ 784\ 401$  évaluations du modèle.
  - ▶ Si une évaluation prend environ 0.01 seconde, cela nécessite 581 130.7 heures de calcul.
  - ▶ Avec une ferme de calcul de 1000 processeurs, cela représente environ 581 heures, soit 24.2 jours.
  - ▶ Pour une puissance de 200 Watts par unité centrale, la consommation électrique correspond à 116 226.1 kWh, soit la consommation annuelle de 16 familles françaises environ et un coût de 11.6 K€ (10 c. / kWh).

Une planification raisonnée, un enjeu pour le Développement Durable

# Analyse de sensibilité globale

## Méthode de Morris

Une exploration astucieuse de l'espace



### Grands principes

- ▶ Discrétisation de l'espace
- ▶ Un déplacement sur un seul facteur à la fois (méthode OAT One At a Time)
- ▶ Un point de départ aléatoire
- ▶ Une trajectoire correspond à K déplacements (K+1 points)

# Analyse de sensibilité globale

## Méthode de Morris

### Analyse

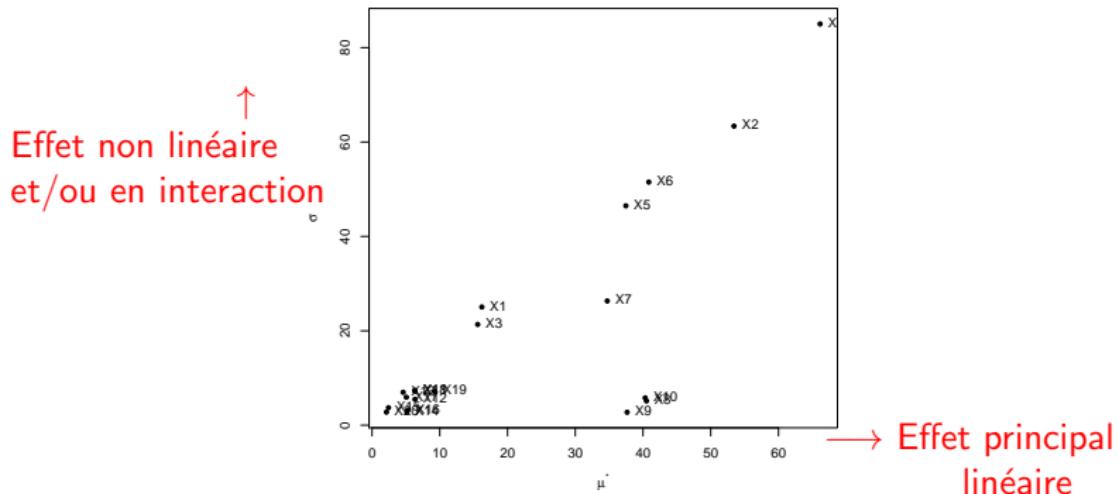
Mesures d'importance du facteur  $x_i$  évaluées sur  $R$  trajectoires  $T^r$  avec le segment  $[\mathcal{G}(T_{(i-1)}^r), \mathcal{G}(T_{(i)}^r)]$  correspondant au saut du facteur  $x_i$

- ▶ effet élémentaire :  $\Delta_i^r = \frac{\mathcal{G}(T_{(i-1)}^r) - \mathcal{G}(T_{(i)}^r)}{\delta}$ ,  $\delta$  la taille du saut
- ▶ effect moyen :  $\mu_i = \frac{1}{R} \sum_{r=1}^R \Delta_i^r$
- ▶ effect absolu moyen :  $\mu_i^* = \frac{1}{R} \sum_{r=1}^R |\Delta_i^r|$
- ▶ écart-type des effets élémentaires :  $\sigma_i = \sqrt{\frac{1}{R-1} \sum_{r=1}^R (\Delta_i^r - \mu_i)^2}$

# Analyse de sensibilité globale

## Méthode de Morris

Résultat du criblage sur la fonction morris, package sensitivity de R



```
morris(model = morris.fun, factors = 20, r = 4, design = list(type = "oat",
levels = 6, grid.jump = 3))
```

## Applications

- ▶ Effet de paramètres génétiques sur l'interception lumineuse de pommiers (MAppleT, Da Silva et al.)

## MappleT : Markov Apple Tree

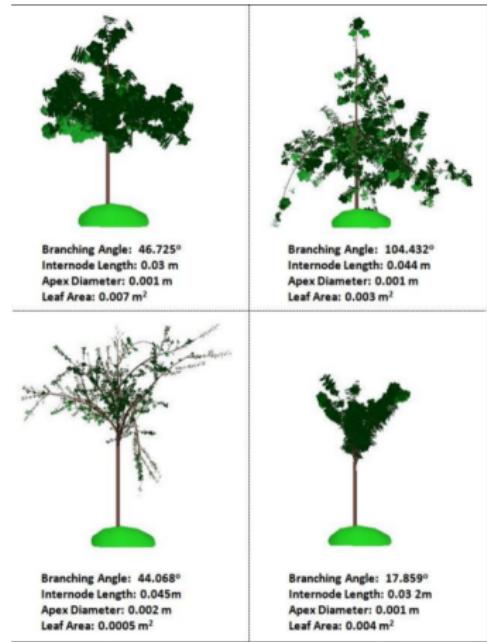
MappleT (Costes *et al.*, 2008)

Manipulation de l'architecture des arbres  
(paramètres génétiques)

Optimisation de l'interception lumineuse

Quatre paramètres : leaf area, internode length, top shoot diameter, branching angle

Temps de calcul ( $\sim 1h$  pour un arbre de 5 ans)

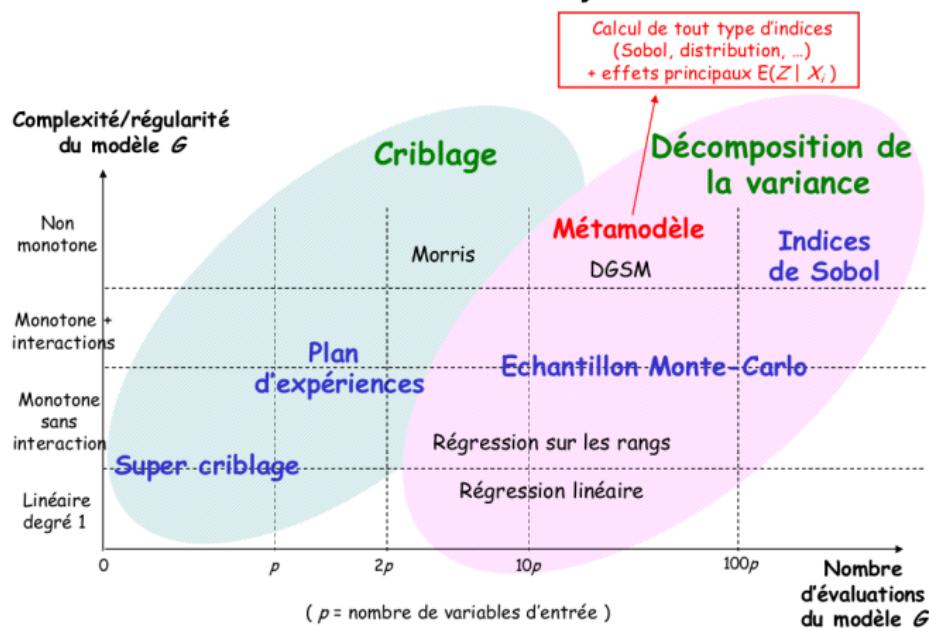


# Analyse de sensibilité globale

## Grille de choix des méthodes d'analyse

Bertrand Iooss (EDF R&D)

### Classification des méthodes d'analyses de sensibilité



## Métamodélisation

$$\begin{aligned}\mathbf{y} &= \mathcal{G}(\mathbf{x}) \\ &\approx \mathcal{M}(x_1, \dots, x_K)\end{aligned}$$

### Métamodèle Polynomial

$$y = \sum_{a=1}^A \beta_a \left( \prod_{k=1}^K x_k^{d_{a,k}} \right) + \varepsilon$$

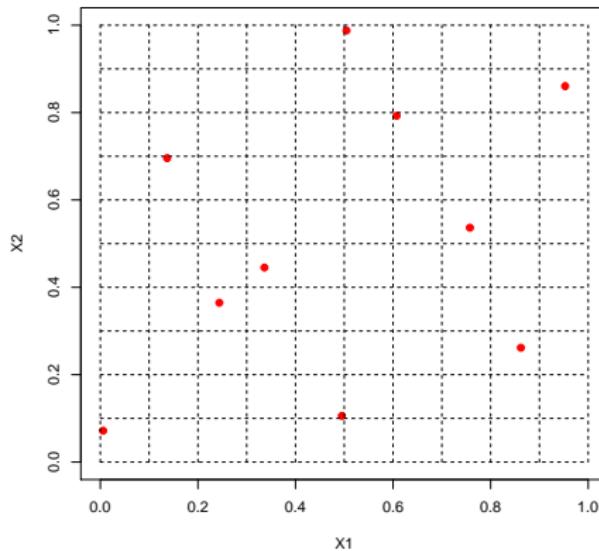
- $K$  nombre de paramètres ;
- $A = C_{K+D}^D$  nombre de termes croisés ( $0 \leq \sum_k d_{a,k} \leq D$ ) ;
- $D$  degré maximal des polynomes ;
- $\varepsilon$  terme d'erreur d'espérance nulle, indépendant des variables  $x_k$ .

Décomposition des sources de variation de  $x_1$  (% de variance expliquée,  $R^2$ )

- Effet principal :  $x_1 + X_1^2 + X_1^3$
- Total :  $x_1 + x_1^2 + x_1^3 + x_1 x_2 + x_1 x_2^2 + x_1^2 x_2 + x_1 x_3 + x_1 x_3^2 + x_1^2 x_3 + x_1 x_2 x_3 + \dots$

## Analyse de sensibilité

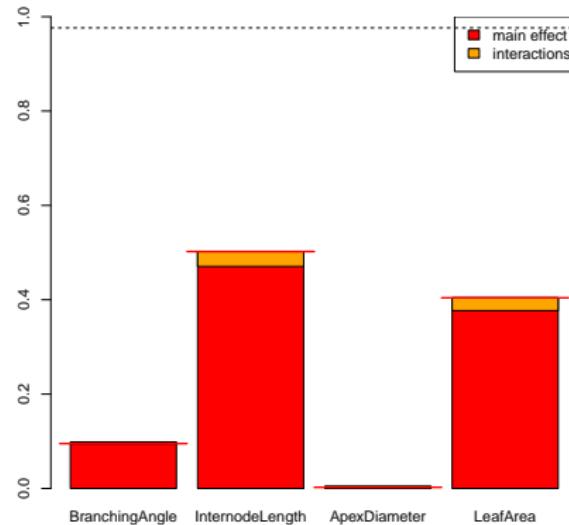
### Plan d'expérience de type hypercube latin (LHS)



un point par segment sur chaque paramètre

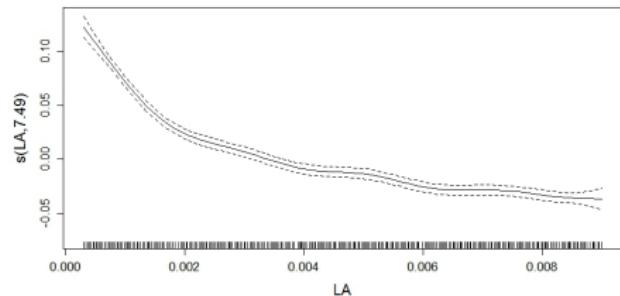
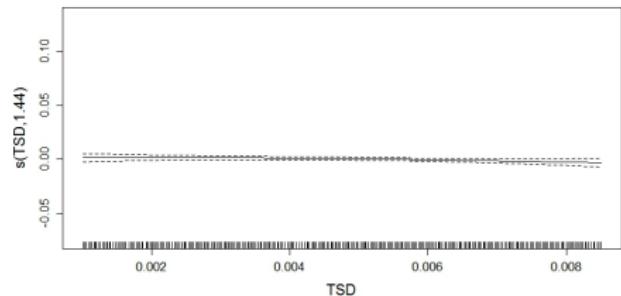
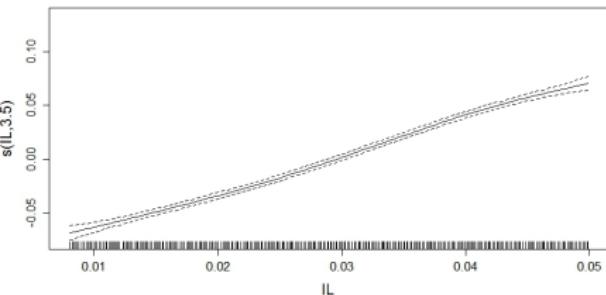
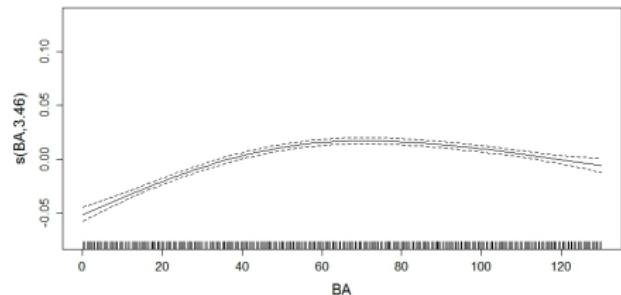
## Analyse de sensibilité

Métamodèle Polynomial de degré 3 évalué sur un plan LHS de taille 300



interception lumineuse : 97.6 % de variance expliquée

## Métamodèle GAM, modèle additif généralisé, $y \approx \sum_{k=1}^K g_k(x_k)$



# Références

- ▶ Réseau MEXICO ([reseau-mexico.fr](http://reseau-mexico.fr)) et GdR Mascot-Num ([www.gdr-mascotnum.fr](http://www.gdr-mascotnum.fr))
- ▶ STICS, <http://www6.paca.inra.fr/stics/>
- ▶ Adamczyk K., Kiêu K., Monod H., Lannou C., Papaïx J. <http://www.breves-de-maths.fr/modifier-un-paysage-agricole-pour-limiter-la-propagation-des-epidemis>
- ▶ Costes E, Smith C, Renton M, Guédon Y, Prusinkiewicz P, Godin C. 2008. MAppleT : simulation of apple tree development using mixed stochastic and biomechanical models. *Functional Plant Biology* 35: 936–950.
- ▶ Da Silva D., Han L., Faivre R., Costes E., 2014. Influence of the variation of geometrical and topological traits on light interception efficiency of apple trees: sensitivity analysis and metamodelling for ideotype definition. *Annals of Botany* 114: 739–752, 2014.
- ▶ Faivre R., Iooss B., Mahévas S., Makowski D., Monod H., editors, 2013. Analyse de sensibilité et exploration de modèles. Applications aux modèles environnementaux. Collection "Savoir Faire", Quae, 2013, 352p.
- ▶ Jeuffroy M.H., Recous S., 1999. Azodyn: a simple model simulating the date of Nitrogen deficiency. *European Journal of Agronomy*, 10, 129-144.
- ▶ Morris M. D., 1991, Factorial sampling plans for preliminary computational experiments, *Technometrics*, 33, 161–174.
- ▶ Pradal C., Dufour-Kowalski S., Boudon F., Fournier C., Godin C. (2008). OpenAlea: a visual programming and component-based software platform for plant modelling. *Func. plant bio.*, 35(10), 751-760.
- ▶ Pujol G., Iooss B., Janon A., 2014. Package 'sensitivity'
- ▶ Saltelli A., Chan K., Scott E.M., editors., 2000. *Sensitivity Analysis*. Wiley Series in Probability and Statistics, Wiley, New York.
- ▶ Saltelli A., Ratto M., Andres T., Campolongo F., Cariboni J., Gatelli D., Saisana M., Tarantola S., 2008. *Global Sensitivity Analysis : The Primer*. Wiley, New York.