Sensitivity analysis and model exploration

Perez Raphaël - adapted from R. Faivre [1]

December 3, 2020

1. Faivre, R., Iooss, B., Mahévas, S., Makowski, D., & Monod, H. (2016). *Analyse de sensibilité et exploration de modèles: application aux sciences de la nature et de l'environnement*. Editions Quae.

Definition of uncertainty and sensitivity analyses

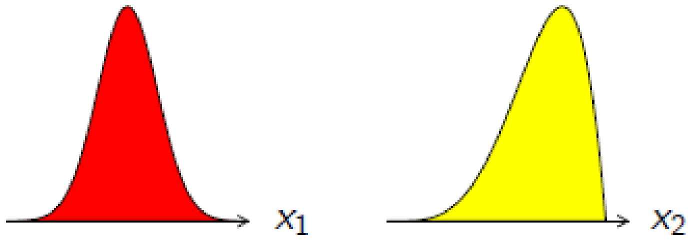
Uncertainty in a model **φ** may come from various sources:

$\varphi(z, p, d)$

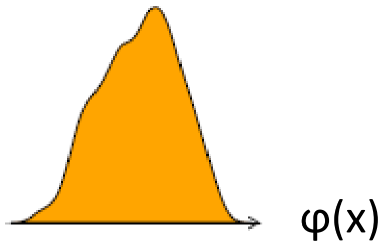
- d → Decisions: inaccuracy in decision rules
- p → Parameters: uncertainty in parameter estimations
- z → Variables : quality of measures, environmental variations
- φ → Equations : lack of knowledge (empirically determined)

Principle of uncertainty and sensitivity analyses

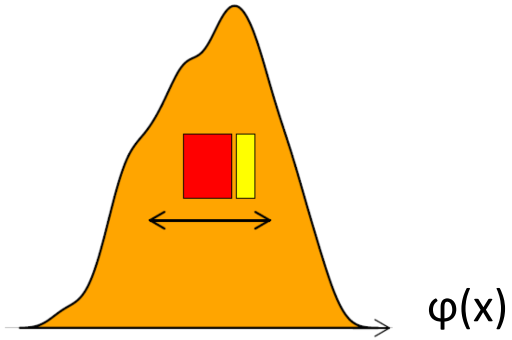
Uncertainty analysis



What is the uncertainty in φ(x) resulting from the uncertainty in x_1 and x_2?



Sensitivity analysis



What are the main sources of uncertainty in x_1,…x_k influencing φ(x)?

Variance of φ(x) = effect of x_1 + effect of x_2 +…

Interests and uses of uncertainty and sensitivity analyses

Uncertainty analysis

- ➢ Give information about uncertainty in model prediction
- ➢ Optimize decisive variables to reduce uncertainty

Sensitivity analysis

- ➢ Identify inputs (variables or parameters) that highly influence model predictions
  - → Which parameters need to be accurately estimate?
- ➢ Analyze model behavior

Uncertainty analysis

Q: What is the uncertainty in φ(x) resulting from the uncertainty in x_1 and x_2?

Example

$\varphi(x) = x_1 + 2\,x_2^2$

$X_1 \sim N(20,16)$ and $X_2 \sim N(60,64)$

→ Analytic expression: $\varphi(x) \sim N(140,272)$

But usually φ(x) cannot be define analitycally

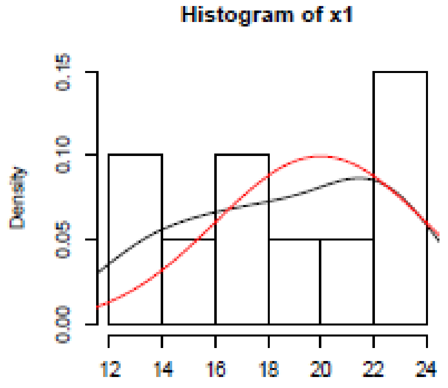
Procedure in 4 steps

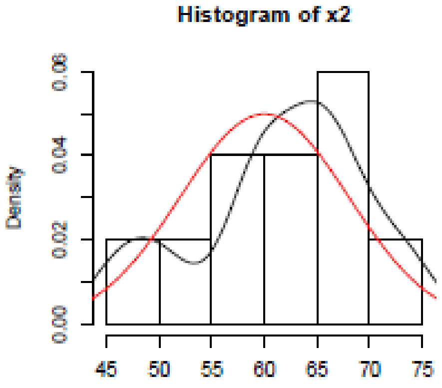
- Define distributions of the inputs $x_1,\ldots,x_k$
- Generate samples from the defined distributions
- Calculate φ(x) for each serie $x_1,\ldots,x_k$
- Estimate the distribution of φ(x)

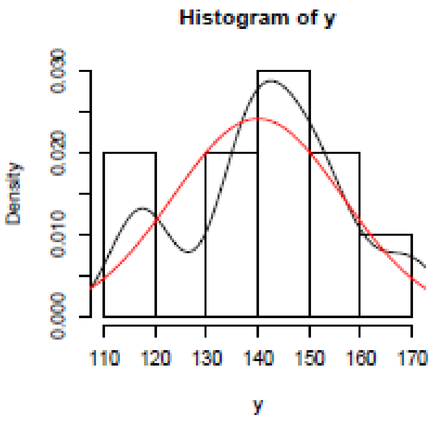
Uncertainty analysis

Application on a simple example

Sample size **N=10**







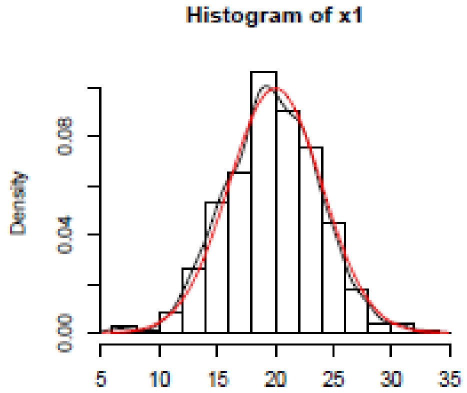
X_1~ N(20,16)

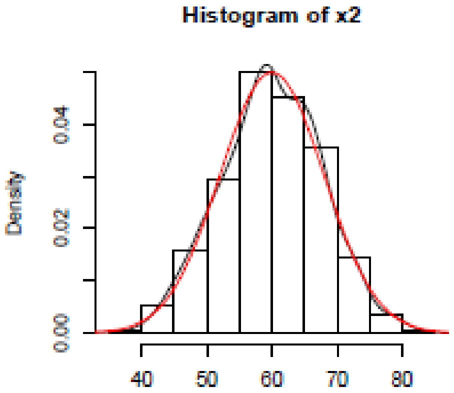
X_2~ N(60,64)

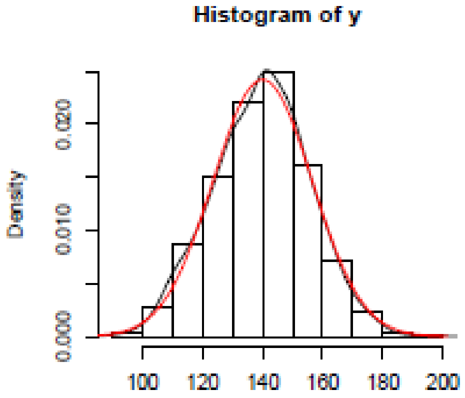
$\varphi(x) = x_1 + 2\ x_2^{\ 2}$

φ(x) ~ N(140,272)

Sample size **N=1000**







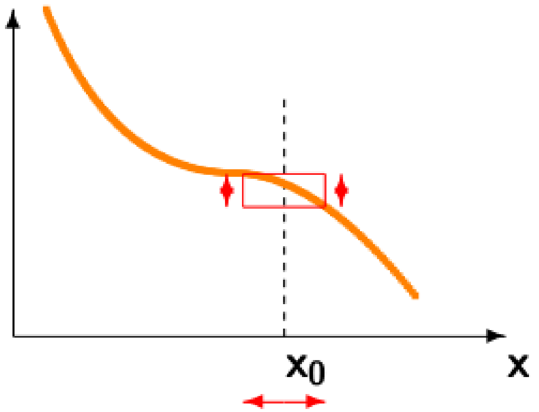
→ Accuracy of φ uncertainty depends on N, not k

Sensitivity analysis

Q: What are the main sources of uncertainty in $x_1,\ldots x_k$ influencing $\varphi(x)$?

Local sensitivity analysis

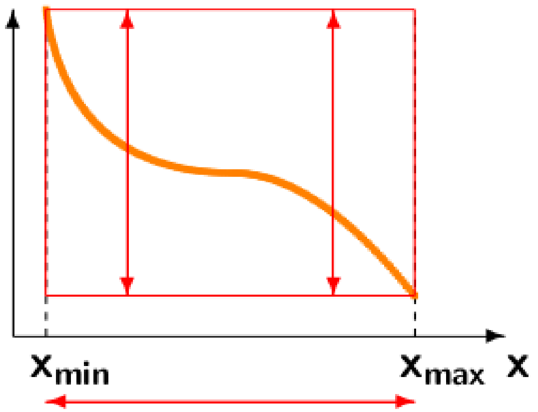
Variations of $\varphi(x)$ around x_0



- Estimation based on derivate

Global sensitivity analysis

Variations of $\varphi(x)$ over range of x



- Estimation based on indices

Global sensitivity analysis

Procedure in 4 steps

- Define distributions of the inputs $x_1,\dots, x_k$

- Generate samples from the defined distributions

- Calculate $\varphi(x)$ for each serie $x_1,\dots, x_k$

- Estimate sensitivity indices

Global sensitivity analysis

Defining indices based on analysis of variances

$$\underbrace{Var[\varphi(x)]}_{\text{Total variance of the output variable}} = \underbrace{Var_{x1} + Var_{x2} + Var_{x3}}_{\text{Principal effects of inputs}} + \ldots + \underbrace{Var_{x1.x2} + Var_{x1.x3}}_{\text{Interaction terms}} + \ldots$$

Total variance of the output variable

Principal effects of inputs

Interaction terms

First order index of x_i : $SI_i = Var_{xi} / Var[\varphi(x)]$

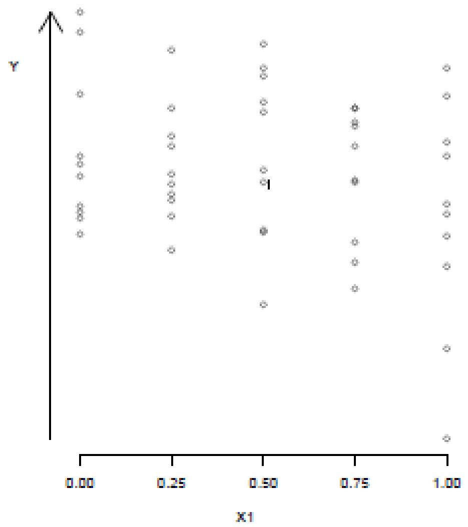
Total sensitivity index of x_i : $TSI_i = (Var_{xi} + Var_{xi.xj} + Var_{xi.xk} + \ldots) / Var[\varphi(x)]$

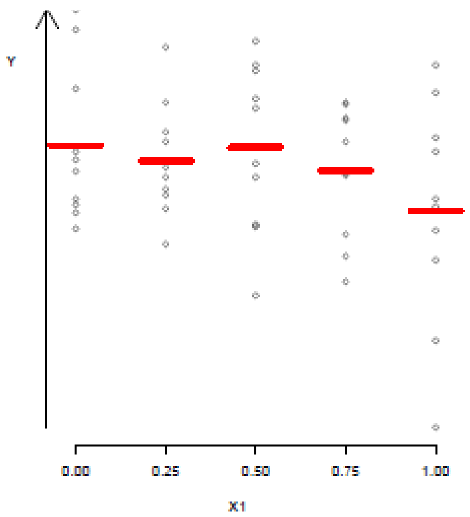
Global sensitivity analysis

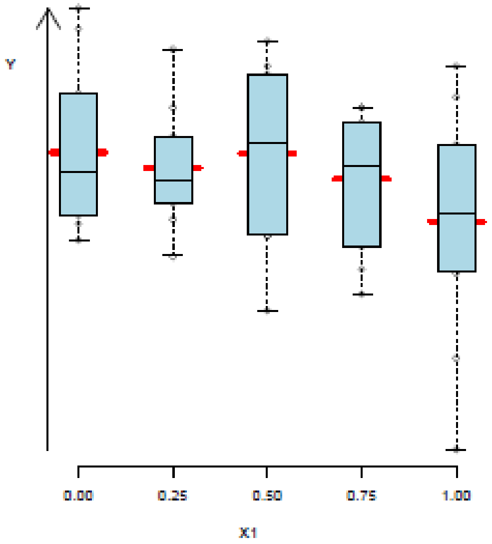
Visualisation of the indices

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

Repeated measurements of *Y* on 5 values of x_1







Theorem of total variance

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

→ **Accuracy depends on N and k**

$$\mathrm{SI}_i = \frac{Var[Esp(\mathcal{G}(\mathbf{x})|x_i)]}{Var[\mathcal{G}(\mathbf{x})]}$$

$$\mathrm{TSI}_i = \frac{Esp(Var[\mathcal{G}|\mathbf{x}^{(-i)}])}{Var[\mathcal{G}(\mathbf{x})]}$$

Experimental designs

Factorial designs:

- Each factor is discretized in levels
- Complete factorial design: all combinations of factors levels are tested

Example:

20 parameters with 3 levels→ 3^{20} = 3 486 784 401 runs

1 run = 0.01 s → 581 160,4 hours of calculations = 24 days → 11,6 K€!!

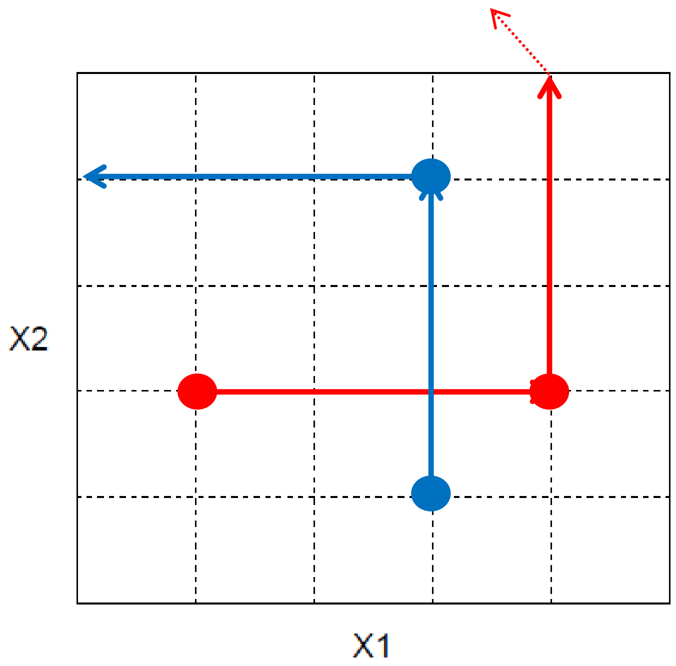
→ **Importance of selecting the appropriated numerical design**

The Morris' Method

An ingenious exploration of the space

- Space discretization
- One At a Time method (OAT): sequential change of inputs with one factor at a time
- Random starting point for each trajectory
- A trajectory is a K displacements (K+1 points)
- OAT design is repeated R times

(total: n= R*(k+1) experiments)



The Morris' Method

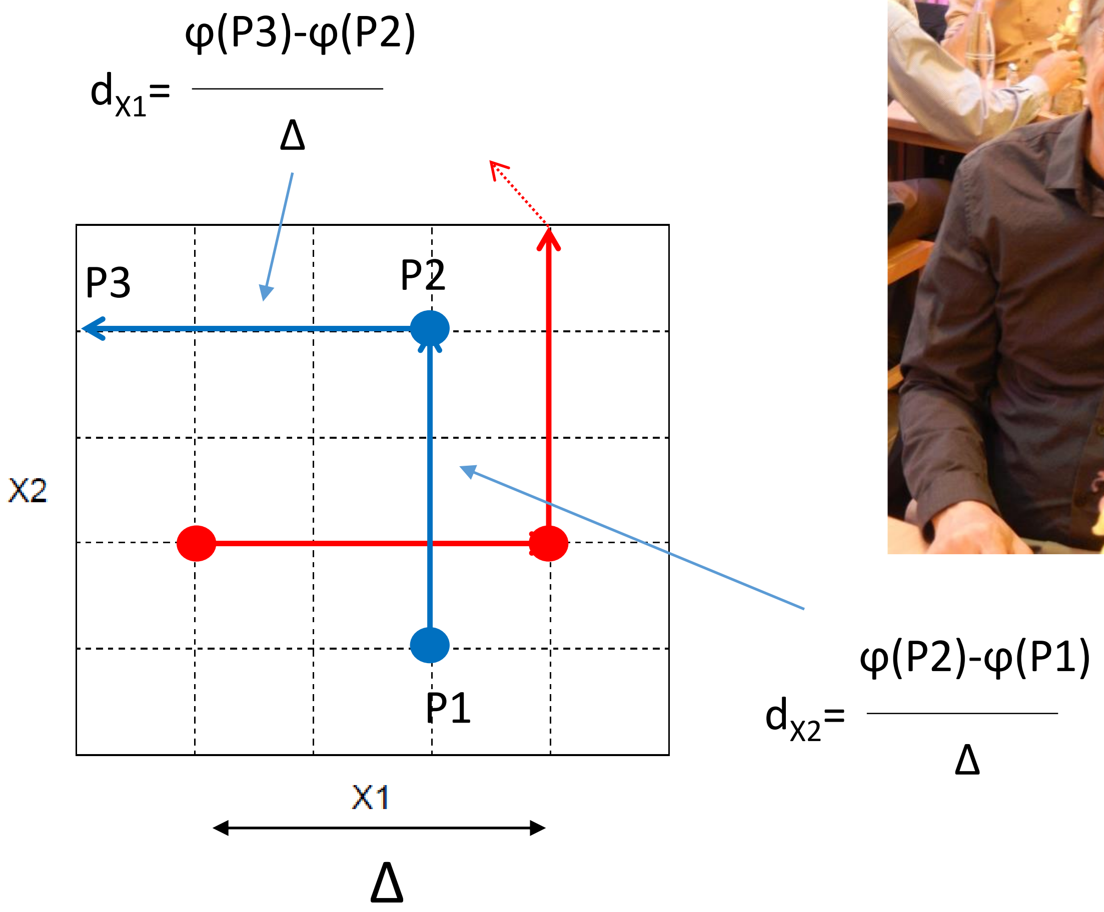
Sensitivity indices

- Estimation of elementary effect d_{Xi}

- R trajectories gives sensitivity measures of Xi:

Mean effect as a measure of importance: $\boldsymbol{\mu i^* = E(|d_{Xi}|)}$

Standard deviation as a measure of interaction / non linearity: $\boldsymbol{\sigma i = \sigma(d_{Xi})}$

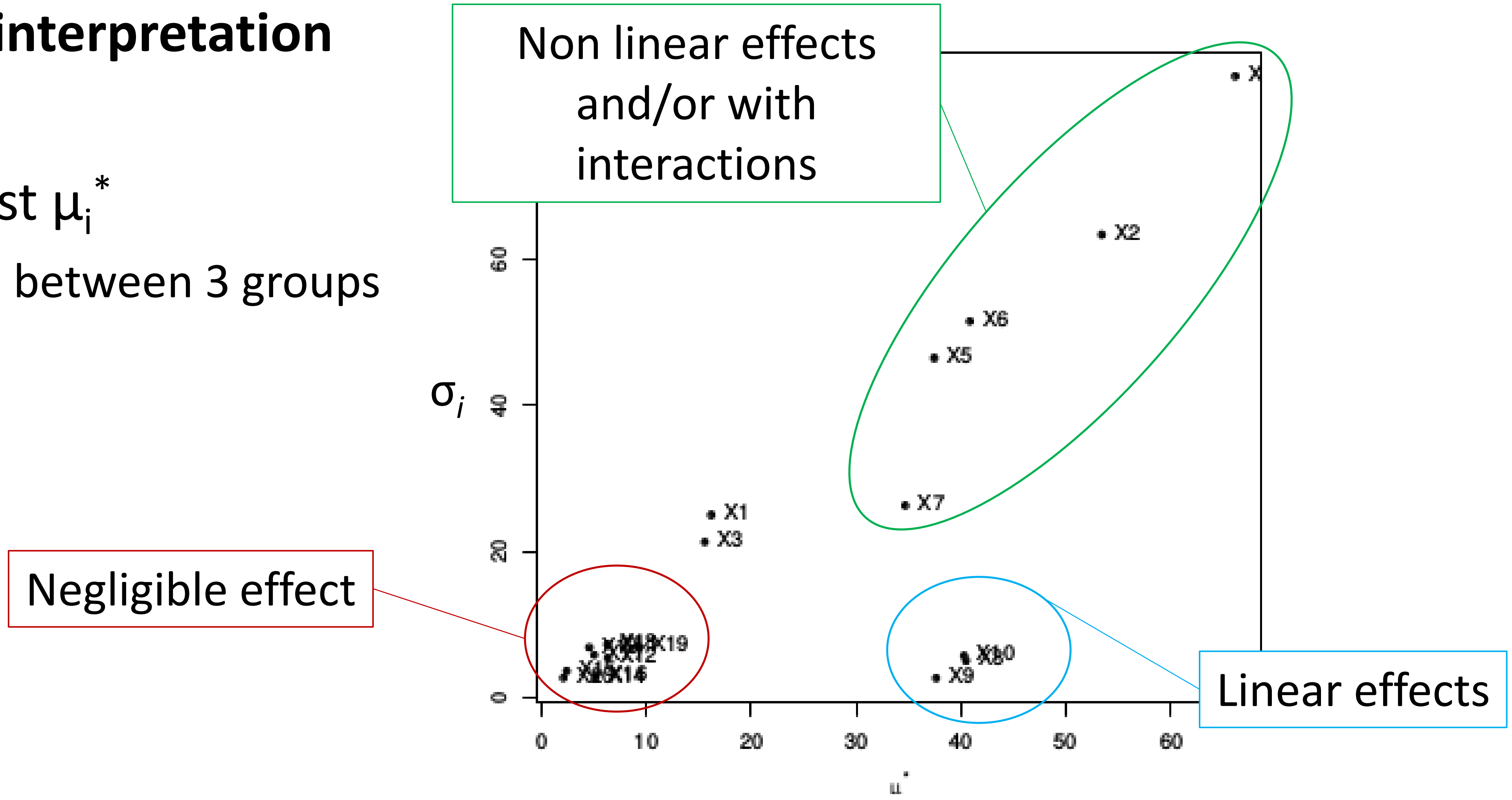


The Morris' Method

Example of interpretation

Plot σ_i against μ_i^*

- ➢ Distinction between 3 groups



μ_i^*

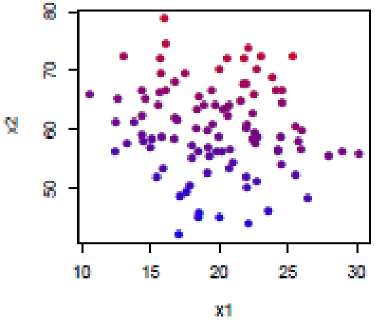
→ Screening method with qualitative information

Metamodelling

Definition and principle

- Mathematical function (polynomial, neural network, kriging, ...) representative of the computer model with negligible cpu cost
- Approximation from a design of experiments

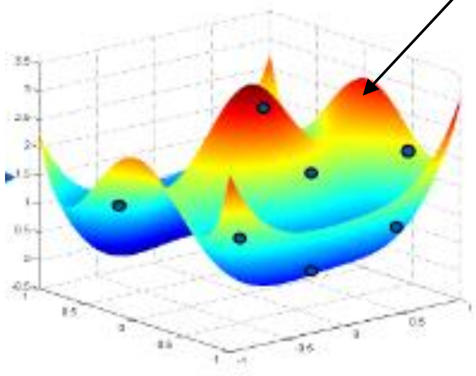
N samples of inputs



N simulations

of φ

M(x)



Iooss & Saltelli (2017).

Metamodelling

Example on FSPM

- MappleT (Coste et al, 2008)

- 4 architectural parameters
    - Leaf area
    - Internode length
    - Top shoot diameter
    - Branching angle

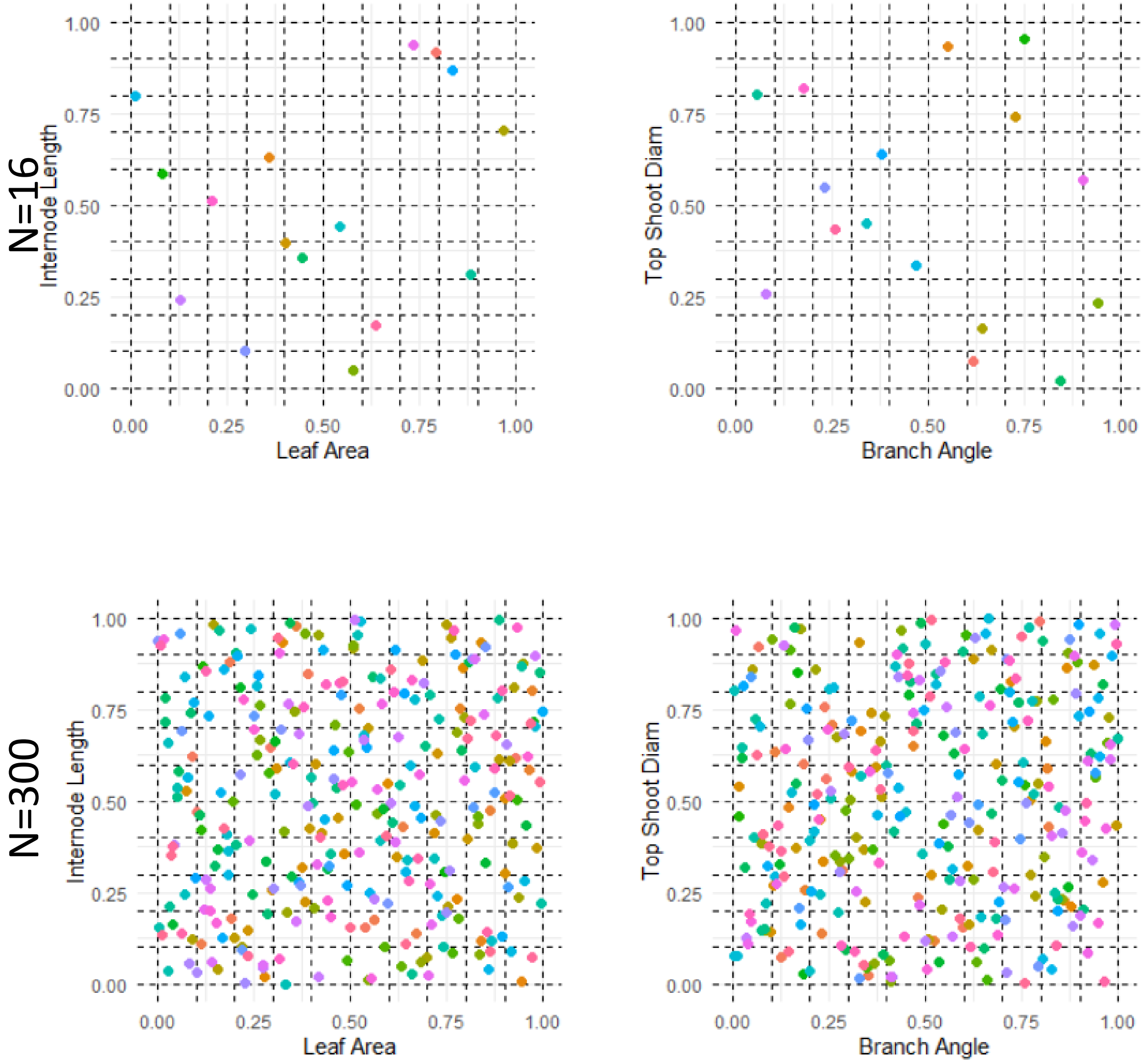
- Output = light interception

- Computational time: 1h per run

Metamodelling

Example on FSPM: experimental design

- ➢ Each parameter discretize in 10 levels
- ➢ Latin hypercube sample (lhs)
  Each parameter level sampled at least once
  N=300



Metamodelling

Example of Polynomial linear model

$$Y = \sum_{a=1}^{A} \beta_a \left(\prod_{k=1}^{K} X_k^{d_{a,k}} \right) + \eta$$

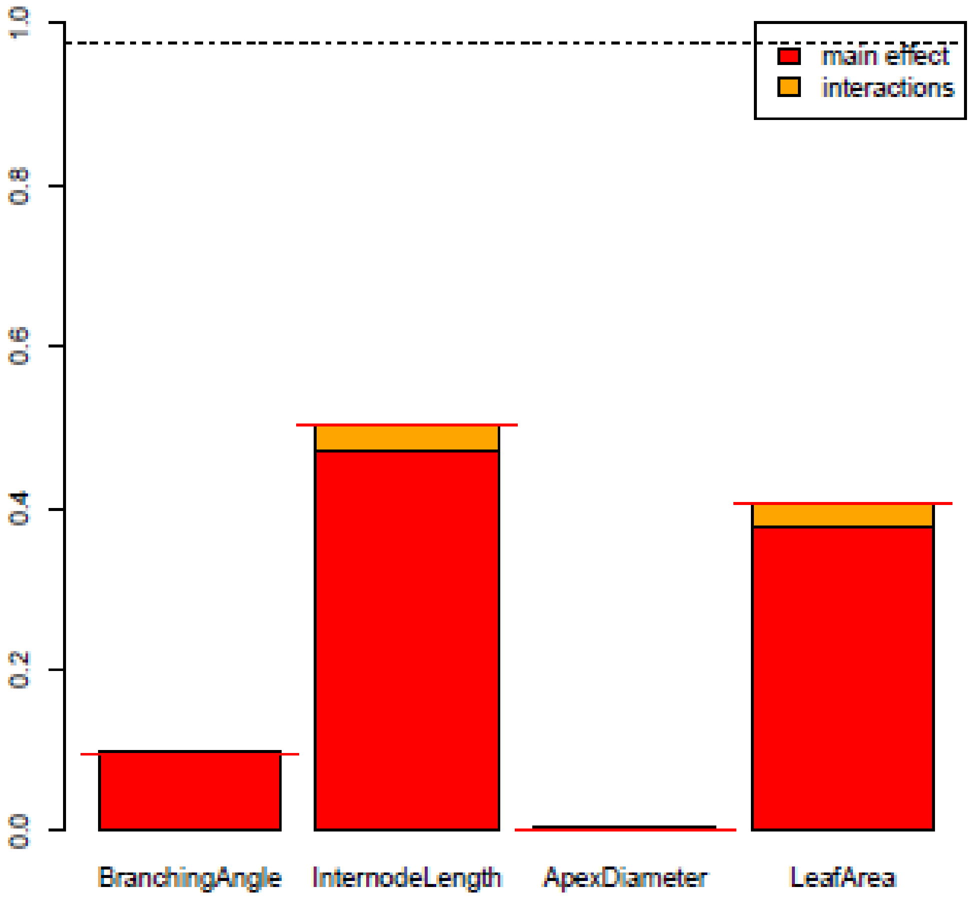
- K is the number of input parameters;
- $A = C_{K+D}^{D}$ is the number of cross product terms ($0 \leq \sum_k d_{a,k} \leq D$);
- D the maximal degree of the polynomial;
- η is a centred random term independent of the X_k variables.

Decomposition of the sources for X_1 (% of explained variance, R^2)
- Main effect : $x_1 + x_1^2 + x_1^3$
- Total : $x_1 + x_1^2 + x_1^3 + x_1 x_2 + x_1 x_2^2 + x_1^2 x_2 + x_1 x_3 + x_1 x_3^2 + x_1^2 x_3 + x_1 x_2 x_3 + \ldots$

Metamodelling

Example on FSPM: metamodel

- Multiple polynomial metamodel of degree 3
- 97.6% of variance in light interception explained
- High contributions of internode length and leaf area

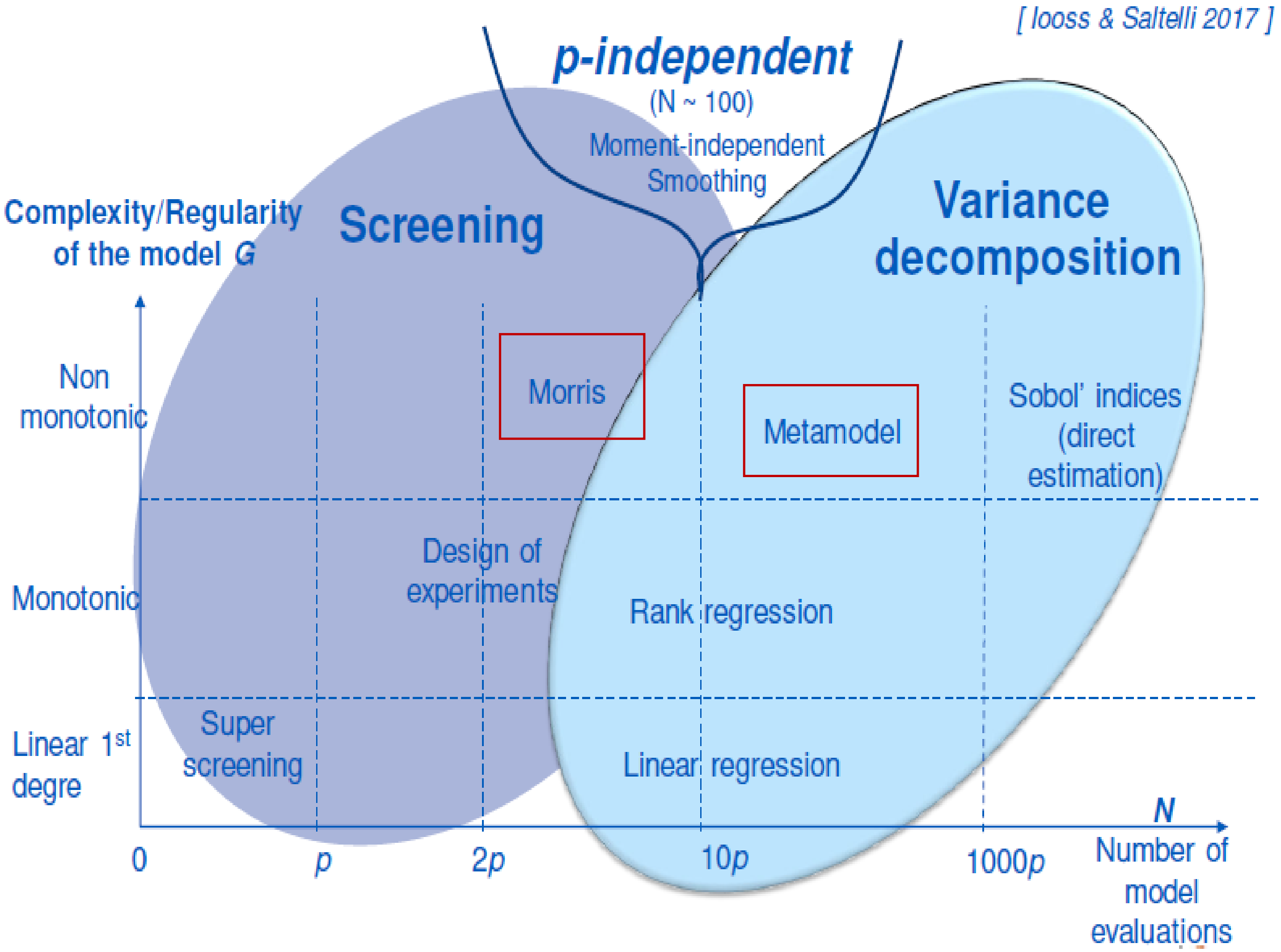


Da Silva et al,2014

Classification of methods

Choosing the method depends on:

- Requested information (qualitative/quantitative)
- Number of inputs
- Regularity of the model
- Computational cost
- Number of outputs



To keep in mind

Aim know the influence of model parameter on predictions for better harness model behavior

Selection of the method

- Number of parameter
- Computational cost
- Screening method (Morris) or quantitative method (metamodel)

Procedure

1. Define the experimental design
2. Run simulations
3. Calculate sensitivity indices
4. Interpret results

References

Reseau MEXICO (reseau-mexico.fr) et GdR Mascot-Num (www.gdr-mascotnum.fr)

Faivre, R., Iooss, B., Mahévas, S., Makowski, D., & Monod, H. (2016). *Analyse de sensibilité et exploration de modèles: application aux sciences de la nature et de l'environnement*. Editions Quae.

Costes E, Smith C, Renton M, Guedon 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. Inuence of the variation of geometrical and topological traits on light interception eciency of apple trees: sensitivity analysis and metamodelling for ideotype denition. Annals of Botany 114: 739{752, 2014.

Morris M. D., 1991, Factorial sampling plans for preliminary computational experiments, Technometrics, 33, 161{174.

Iooss, B., & Saltelli, A. (2017). Introduction to sensitivity analysis. *Handbook of uncertainty quantification*, 1103-1122.