Statistical learning for computer experiment

Atelier GDR MASCOT NUM 2011

14:00 – 14:07 : F. Mangeant : Introduction

14:07 – 14:59 : G. Obozinski : Survey of statistical learning trends

15:00 – 15:45 : N. Rachdi : Application of statistical learning techniques

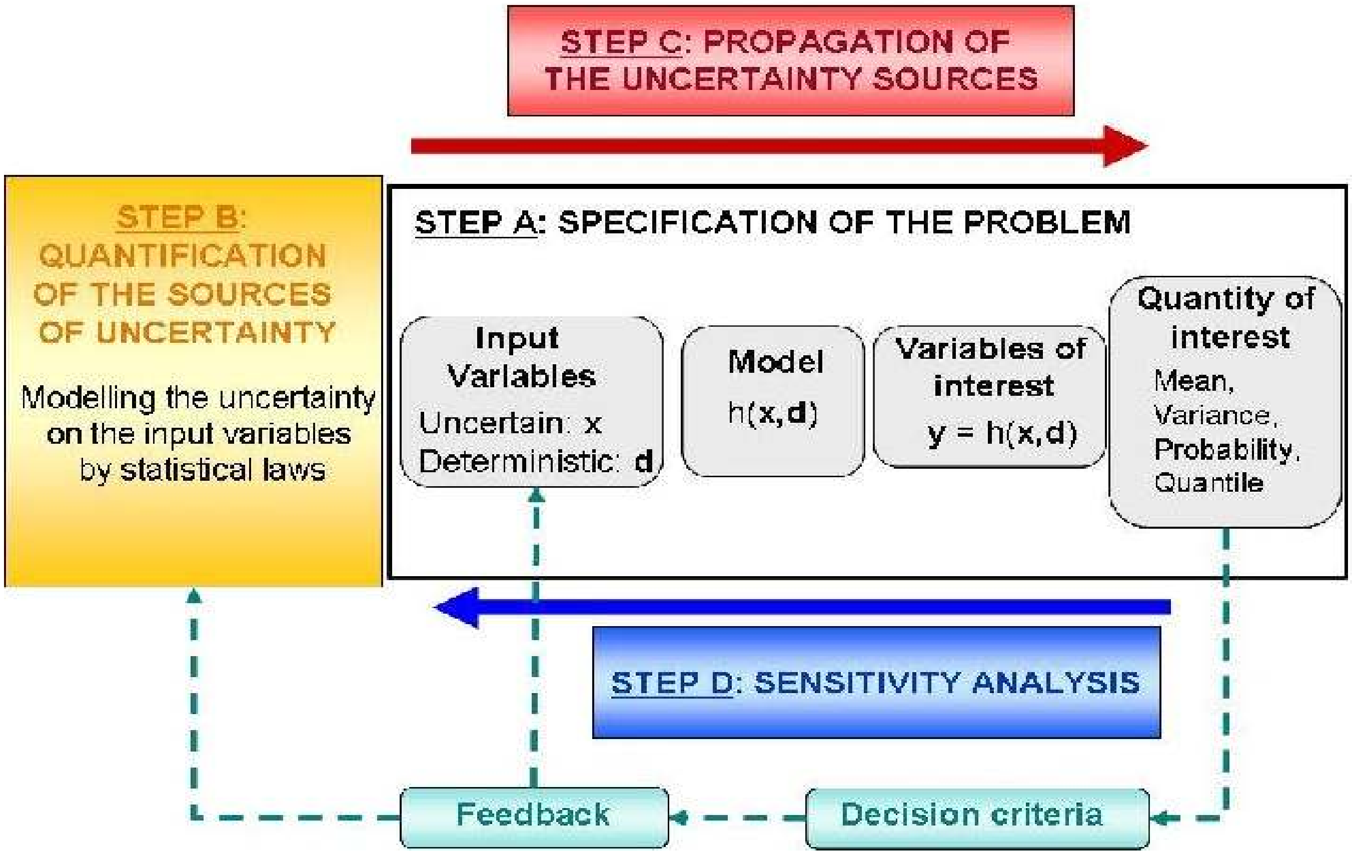
15:46 – 16:01 : Break

16:01 – 16:58 : Open discussion

Statistical learning for computer experiment

Atelier GDR MASCOT NUM 2011

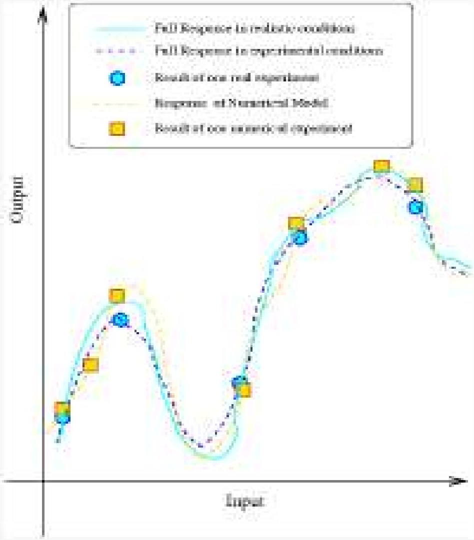
Once again, the engineer's point of view !



The "model" uncertainty

What are the components of model uncertainty?

- **Reference model h^***: Usually not accessible, expression of a natural or a complex technical object.
- **Theoretical model $\tilde{h}$**: Scientific expert activity (theoretical solution of a PDE system, ...), corresponding to the level of understanding and simplification of the problem.
- **Numerical model $\hat{h}$**: Numerical solution of the theoretical model (effects of meshing, choice of a numerical scheme)
- **Implementation model h**: Software implementation of the model on a given hardware architecture (computer accuracy, choice of coding rules, ..).



Statistical learning for computer experiment

Atelier GDR MASCOT NUM 2011

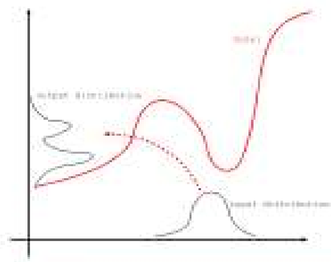
Other components of uncertainty

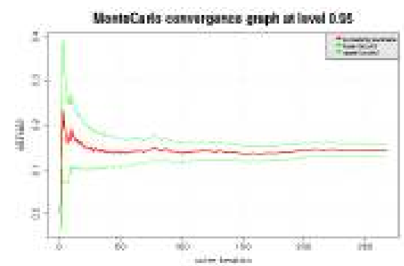
Uncertainty on the input parameters

Uncertainty due to the choice of probability measure $\mathbb{P}^{\mathbf{X}}$ on the input parameters $\mathbf{X}$ compared to $\mathbb{P}^{\mathbf{X}}_{*}$

Uncertainty on the stochastic convergence

Approximation of the criterion of interest $\rho(Y)$ by a stochastic computation $\hat{\rho}_N(Y)$





Is it possible to define a sort of contribution's metrics?

$$\Delta \leq \underbrace{\mathcal{N}_{\mathcal{S}}(h^*, \tilde{h})}_{\text{Scientific Validation}} + \underbrace{\mathcal{N}_{\mathcal{N}}(\tilde{h}, \hat{h})}_{\text{Numerical Validation}} + \underbrace{\mathcal{N}_{\mathcal{I}}(\hat{h}, h)}_{\text{Hardware / Software Validation}} + \underbrace{\mathcal{N}_{\mathcal{Q}}(\mathbb{P}^{\mathbf{X}}_{*}, \mathbb{P}^{\mathbf{X}})}_{\text{Statistical Validation}} + \underbrace{\mathcal{N}_{\mathcal{P}}(\rho(Y), \hat{\rho}_N(Y))}_{\text{Propagation Validation}}$$

Engineering basic metrics!

Statistical learning for computer experiment

Atelier GDR MASCOT NUM 2011

Specification of the model, input/output/criteria

- **To characterize the properties of the model**
  - **Dimension**: h is classically a real function belonging to $\mathcal{F}(\mathbb{R}^P \times \mathbb{R}^T, \mathbb{R}^Q)$. Even if the dimension of $\mathbf{x}$ can be large, most of the engineering problems we are focused on only contain $P \leq 50$ and $Q \leq 5$.
  - **Computational budget**: A single computation of h can be very expensive. The computational budget will represent the number of runs N affordable to solve the problem.
  - **Black box/white box**: h is either a black box (*the inner operations of the model are not accessible*), a grey box (*part of the inner operations is accessible*) or a white box (*all the operations of the model are accessible*).
  - **Mathematical properties**: the basic mathematical properties (regularity, monotony, linearity or non linearity towards certain parameters) may be unknown to the engineer.
  - **Domain of validity**: h should be delivered with its domain of validity $\mathcal{V}^{[\epsilon]} \subseteq \mathbb{R}^P \times \mathbb{R}^T$.

Statistical learning for computer experiment

Atelier GDR MASCOT NUM 2011

What kind of information is available in many of our computer experiment ?

- Couples (Xi, Yi)
- Database of {Xi} and database of {Yj} not obtained simultaneously
- Panoply of numerical models {h1, h2, h3}

  → ***Considered deterministic this afternoon***
- Different quantities of interest
  - Probability, quantile, moments
  - Probability density function, cumulative density function, …

Example :
h1: Interpolation in database Lookup table,
h2: Analytical model,
h3: 2D PDE model,
h4: 3D PDE solved by a numerical method

Statistical learning for computer experiment

Atelier GDR MASCOT NUM 2011

Requirements of the algorithms to be developed

- Best use of the information for a given objective (*ie* quantity of interest)

- Non asymptotic approaches (linked to the size of the database and the CPU budget)

- Adaptative building of surrogate models for a given objective

- Measure of accuracy of the estimator/algorithm

- Validation/Verification strategies

Statistical learning for computer experiment

Atelier GDR MASCOT NUM 2011

Challenge 1 – Metrics to selec the right level of fidelity for a given goal

In many applications, several numerical codes are available
for the same type of application *(Example : Rule of the thumb, excell sheet, 2D model, 3D FE model, …)*

A compromise has to be found between the « complexity » of the model and the « objective » of the simulation.

Statistical learning for computer experiment

Atelier GDR MASCOT NUM 2011

Challenge 2 – How to develop new surrogate models adapted to specific purpose ?

Kriging, polynomial chaos as they are built today for a L2 norm and may be not well fitted for some specific objectives

Challenge 3 – How to develop validation and verification strategies ?

To be able to evaluate the different types components of validation