

Airbus Group Innovations

Current engineering practices in UQ&M in aeronautics and associated challenges



$\mathsf{UQ}\&\mathsf{M}$ in aero-engineering practices

Outline

1 Introduction

- Mathematical notations and link with statistical learning
- Uncertainty related to "modelling" activities
- Review of some UQ challenges
 Six UQ challenges

3 Conclusions and perspectives

- A collaborative platform to share experience
- Lessons learnt



Outline

1 Introduction

- Mathematical notations and link with statistical learning
- Uncertainty related to "modelling" activities
- Review of some UQ challengesSix UQ challenges
- 3 Conclusions and perspectives
 - A collaborative platform to share experience
 - Lessons learnt



What kind of information do we manipulate in a mono-disciplinary context?

Elements of information

- A reference database $(\mathbf{Y}_1^*, \cdots, \mathbf{Y}_n^*)$ or $((\mathbf{X}_1^*, \mathbf{Y}_1^*), \cdots, (\mathbf{X}_n^*, \mathbf{Y}_n^*))$ that is enriched during the design cycle.
- A set of risk measures ($\rho_1(\mathbf{Y}^*), \cdots, \rho_d(\mathbf{Y}^*)$) built upon \mathbf{Y}^* to be estimated during the design cycle.
- A panoply of numerical models $\mathcal{H} = \{h_1, \dots, h_D\}$ representing the phenomenon at different levels of fidelity/adequacy $(\mathbf{Y}^* \approx h_i(\mathbf{X}^{[i]}, \theta^{[i]})$
- A quantification of the uncertainties attached to the inputs of the numerical models represented by a statistical law $\mathbb{P}_{X^{[r]}}$ that is enriched during the design cycle
- A global computational budget ${\cal B}$



Our motivation in terms of computer experiments

Goal

Let \mathbb{Q} be the unknown probability measure associated to the real random variable \mathbf{Y}^* defined over $(\mathbb{R}^Q, \mathcal{B}(R^Q), \mathbb{Q})$. Our goal is to predict one or several features $\rho(\mathbb{Q}) \in \mathbb{F}$ of the distribution \mathbb{Q} (also abusively noted $\rho(\mathbf{Y}^*)$). This feature corresponds to the measure of risk over our variable of interest \mathbf{Y}^* .

Examples of probabilistic measures of risk $ ho(\mathbf{Y}^*)$				
Mean:	$ ho(\mathbf{Y}^*) = \mathbb{E}\left[\mathbf{Y}^* ight]$	$\in \mathbb{F} = \mathbb{R}$		
Variance:	$ \rho(\mathbf{Y}^*) = \operatorname{Var}\left[\mathbf{Y}^*\right] $	$\in \mathbb{F} = \mathbb{R}_+$		
Quantile:	$ ho(\mathbf{Y}^*)=q_r(\mathbf{Y}^*)$	$\in \mathbb{F} = \mathbb{R}_+$		
Probability:	$ ho(\mathbf{Y}^*) = \mathbb{P}\left(\mathbf{Y}^* \in \mathcal{D}_P ight)$	$\in \mathbb{F} = [0,1]$		
CDF:	$ ho(\mathbf{Y}^*) = \mathbb{P}\left(\mathbf{Y}^* \leq \mathbf{y}^* ight)$	$\in \mathbb{F} = \mathcal{F}_{cdf}(\mathbb{R}^Q, [0,1])$		
PDF:	$ ho(\mathbf{Y}^*) = f_{\mathbf{Y}^*}(\mathbf{y}^*)$	$\in \mathbb{F} = \mathcal{F}_{pdf}(\mathbb{R}^Q, \mathbb{R}_+)$		

Our motivation in terms of computer experiment

Properties of a numerical model h

- **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 x can be large, most of the engineering problems we are focused on only contain $P \leq 100$ and $Q \leq 20$.
- **Computational budget**: A single computation of *h* can be very expensive. The computational budget *B* will be represented by the number *m* of runs 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 $\mathcal{V}^{[\epsilon]} \subseteq \mathbb{R}^{P} \times \mathbb{R}^{T}$.



Uncertainty attached to "modelling" activities

"Model" uncertainty in a mono disciplinary context

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

 $h^* \rightsquigarrow h$



Uncertainty attached to "modelling" activities

Parametric input uncertainty

- For a given numerical model h: (x, θ) ∈ X × Θ → y = h(x, θ) ∈ Y, we consider an uncertainty attached to the input variables X modelled by a statistical law P^{*}_X.
- In practical contexts, it is often difficult to build \mathbb{P}_{X}^{*} due to scarsity of data, heterogeneous database, lack of information on the dependency, ... As a matter of fact, one has to work with an **approximate statistical law** \mathbb{P}_{X} .

$\mathbb{P}^*_{\boldsymbol{\mathsf{X}}} \rightsquigarrow \mathbb{P}_{\boldsymbol{\mathsf{X}}}$

Computational budget &

In many situations, it is difficult to compute analytically the risk measures $\rho(h(\mathbf{X}, \theta))$. Numerical methods $\mathcal{M}(\mathcal{B}, \varepsilon, h(\mathbf{X}, \theta))$ (either stochastic or not) are required using a fixed computational budget \mathcal{B} for a given accuracy ε

 $\rho(h(\mathbf{X}, \theta)) \rightsquigarrow \mathcal{M}(h(\mathbf{X}, \theta), \mathcal{B}, \varepsilon)$



How to manage all the components of the error?

Recap of the errors in a mono disciplinary context

- **1** Building of the model: $\mathcal{N}_{\mathcal{S}}(h^*, h_{th})$
- **2** Numerical approximation: $\mathcal{N}_{\mathcal{N}}(h_{th}, h_{num})$
- **3** Hardware/Software implementation: $\mathcal{N}_{\mathcal{I}}(h_{num}, h)$
- 4 Model paramaters uncertainty: $\mathcal{N}_{\mathcal{Q}}(\mathbb{P}^*_{\mathbf{X}}, \mathbb{P}_{\mathbf{X}})$
- **5** Uncertainty propagation error: $\mathcal{N}_{\mathcal{P}}(\rho(h(\mathbf{X}, \theta)), \mathcal{M}(h(\mathbf{X}, \theta), \mathcal{B}, \varepsilon))$

Naive form of the total error

 $\Delta \leq$

$$\underbrace{ \begin{array}{l} \underbrace{\mathcal{N}_{\mathcal{S}}(h^{*}, h_{th})}_{Scientific \ Validation} \\ + \underbrace{\mathcal{N}_{\mathcal{N}}(h_{th}, h_{num})}_{Numerical \ Validation} + \underbrace{\mathcal{N}_{\mathcal{I}}(\hat{h}, h)}_{Hardware / Software \ Validation} \\ + \underbrace{\mathcal{N}_{\mathcal{Q}}(\mathbb{P}^{X}_{*}, \mathbb{P}^{X})}_{Statistical \ Validation} + \underbrace{\mathcal{N}_{\mathcal{P}}(\rho(Y), \hat{\rho}_{\mathcal{B}}(Y))}_{Propagation \ Validation} \end{array}$$



UQ&M in aero-engineering practices

Outline

1 Introduction

- Mathematical notations and link with statistical learning
- Uncertainty related to "modelling" activities

2 Review of some UQ challengesa Six UQ challenges

3 Conclusions and perspectives

- A collaborative platform to share experience
- Lessons learnt



Challenge C1: To develop a "global" UQ approach (1/3)

Prediction of the feature $\rho(\mathbb{Q}) = \rho(\mathbf{Y}^*)$ thanks to a pre defined model $h(\mathbf{x}, \theta) = \mathbf{y}$, a statistical law $\mathbb{P}_{\mathbf{X}}$ and a numerical method $\mathcal{M}(q, \varepsilon)$

• Model uncertainty: The probability measure \mathbb{Q} being unknown, it is approximated by the composition of a model h, defined over $\mathcal{X} \times \Theta$ and a statistical law $\mathbb{P}_{\mathbf{X}}$. Thus, it is possible to approximate the feature $\rho(\mathbf{Y}^*)$ by $\rho(h(\mathbf{X}, \theta))$.

$$\rho(\mathbf{Y}^*) \approx \rho(h(\mathbf{X}, \hat{\theta}_n))$$

 θ is calibrated to $\hat{\theta}_n$ thanks to the reference database $(\mathbf{Y}_1^*, \cdots, \mathbf{Y}_n^*)$ or $((\mathbf{X}_1^*, \mathbf{Y}_1^*), \cdots, (\mathbf{X}_n^*, \mathbf{Y}_n^*))$

Quantification uncertainty: The statistical law $\mathbb{P}_{\mathbf{X}}$ may be biased for statistical reasons if only *p* values are available. Thus, $\rho(h(\mathbf{X}, \theta))$ is obtained through:

$$\rho(h(\hat{\mathbf{X}}_{p},\hat{\theta}_{n})) \xrightarrow{p \to \infty} \rho(h(\mathbf{X},\hat{\theta}_{n}))$$

Propagation uncertainty: As it is quite rare to compute exactly $\rho(h(\mathbf{X}, \theta))$, it is approximated by either a deterministic or a stochastic numerical method \mathcal{M} . \mathcal{M} is characterized by its accuracy ε for a given budget of computations \mathcal{B} .

$$\mathcal{M}(h(\mathbf{X},\theta),\mathcal{B},\varepsilon) \xrightarrow{\mathcal{B} \to \infty, \varepsilon \to \mathbf{0}} \rho(h(\mathbf{X},\theta))$$



Challenge C1: To develop a "global" UQ approach (2/3)

Prediction of the feature $\rho(\mathbb{Q}) = \rho(\mathbf{Y}^*)$ thanks to a pre defined model $h(\mathbf{x}, \theta) = \mathbf{y}$, a statistical law $\mathbb{P}_{\mathbf{X}}$ and a numerical method $\mathcal{M}(q, \varepsilon)$

- Surrogate modelling uncertainty:
 - Strategy 1 "Build & replace strategy": The computational budget \mathcal{B} is used in two steps. First, a surrogate model \tilde{h} is built upon a budget $\mathcal{B}' < \mathcal{B}$. Then, the model h is replaced by \tilde{h} to be used by the propagation method \mathcal{M}' . It yields to :

$$\|\mathcal{M}'\left(ilde{h}_{\mathcal{B}'}(\mathsf{X},\eta),\mathcal{B}-\mathcal{B}',arepsilon
ight)-
ho(h(\mathsf{X}, heta)\|\leq\|\mathcal{M}(h(\mathsf{X}, heta),\mathcal{B},arepsilon)-
ho(h(\mathsf{X}, heta)\|$$

Strategy 2 : "Build & collaborate strategy": The final approximation with numerical method *M*' should be more accurate in a certain sense |||| than the previous one:

$$\|\mathcal{M}'\left(h(\mathsf{X},\theta),\tilde{h}_{\mathcal{B}'}(\mathsf{X},\eta),\mathcal{B},\varepsilon\right) - \rho(h(\mathsf{X},\theta)\| \leq \|\mathcal{M}(h(\mathsf{X},\theta),\mathcal{B},\varepsilon) - \rho(h(\mathsf{X},\theta)\|$$



Challenge C1: To develop a "global" UQ approach (3/3)

Summary approach

Example:

• Estimators In the particular case when the budget \mathcal{B} is used for N number of computations of h and M computations of the surrogate model:

$$\hat{\mathcal{E}}_1(M, N, p, n) = \hat{\rho}_M(\tilde{h}_N(\hat{\mathbf{X}}_p, \hat{\theta}_n))$$
$$\hat{\mathcal{E}}_2(N, p, n) = \hat{\rho}_N(h(\hat{\mathbf{X}}_p, \hat{\theta}_n))$$

Cost model

$$C(\mathcal{B}, p, n) = \alpha_{simu}(\mathcal{B}) + \beta_{input \ data}(p) + \gamma_{input \ ref}(n)$$

Classically,

$$lpha_{simu}(\mathcal{B}) < eta_{input \ data}(p) << \gamma_{input \ ref}(n)$$



Challenge C2: To build efficient goal-oriented surrogate models (1/3)



Challenge C2: To build efficient goal-oriented surrogate models (2/3)

Surrogate modelling

Implementation Improvements in OpenTURNS						
In order to improve the performance of the SPCE algorithm, we implemented the following improvements:						
 Parallelized the evaluation of the polynomial basis as well as the iso-probabilistic transformation that maps the 						
probability distribution of the uncertainties into the measure associated to the orthogonal polynomial basis						
 Switched to highly efficient OpenBLAS that enables for a parallel approach for the linear algebra part of the 						
computations						
• Switched to the normal equation instead of the QR decomposition to solve the least-squares problems (normal						
equation becomes more well-conditioned when the size of the database increases)						
	OpenTURNS 1.2	OpenTURNS 1.3	OpenTURNS 1.4			
Tim	• (s) 43040	19089	2332			
Spee	d-up 1	2.25	18.45			

An additional Kriging approach has been investigated to model the residuals

- · For the global behaviour only a small quality increase is achieved at relatively high costs
- A significant gain can only be observed near the stall region (which is out of our scope here)



Challenge C2: To build efficient goal-oriented surrogate models (3/3)

Surrogate modelling

Integrate the objective of the computation in the building phase of the surrogate model: optimization, reliability analysis, sensitivity towards a specific risk measure.



Challenge C3: To develop a goal-oriented sensitivity analysis

Sensitivity analysis to the choice of predefined model $h(\mathbf{x}, \theta) = \mathbf{y}$ and the statistical law $\mathbb{P}_{\mathbf{X}}$ on the prediction of the feature $\rho(\mathbb{Q}) = \rho(\mathbf{Y}^*)$

The probability measure \mathbb{Q} being unknown, it is approximated by the composition of a model *h*, defined over $\mathcal{X} \times \Theta$ and a statistical law $\mathbb{P}_{\mathbf{X}}$. Thus, it is possible to approximate the feature $\rho(\mathbf{Y}^*)$ by $\rho(h(\mathbf{X}, \theta))$.

$$\rho(\mathbf{Y}^*) \approx \rho(h(\mathbf{X}, \theta))$$

Influence of the group of input variables $\mathbf{X}^{\mathcal{K}}$ ($\mathcal{K} \subseteq \{1, \dots, P\}$) on the feature of interest $\rho(h(\mathbf{X}, \theta))$:

$$\mathbb{E}\left[\rho(h(\mathbf{X},\theta)|\mathbf{X}^{K}=\mathbf{x}^{K})\right]=\rho(h(\mathbf{X},\theta))??$$

Influence of the statistical model \mathbb{P}_{X} on the feature of interest $\rho(h(X, \theta))$

$$\rho(h(\mathbb{P}^{1}_{\mathsf{X}},\theta)) = \rho(h(\mathbb{P}^{2}_{\mathsf{X}},\theta)) ??$$

Influence of the choice of model h_i among the panoply of model $\mathcal{H} = \{h_1, \cdots, h_D\}$

$$\rho(h_i(\mathbf{X}^{(i)}, \theta^{(i)})) = \rho(h_i(\mathbf{X}^{(i)}, \theta^{(i)})) ??$$



Description of the situation

- A target *T* is given to the variable **y**^{*}. This target can evolve during the time of the design.
- These performances are uncontrolled for many reasons (lack of knowledge, variability, approximation, dependency, ...).
- The amount of available information \mathcal{I} for each variable y_i^* evolves during the time of the design (either over the knowledge of the input variables, parameters, mesurements, availability of numerical models).
- At a given time of the design, these technical performances must be estimated with a level of confidence.





Figure: Evolution of a performance during the design phase



Figure: An uncertainty study at a given time of the design







Objectives in a mathematical framework

In a probabilistic framework, two main goals can be identified:

- **I** To control the stochastic behaviour of the performances \mathbf{y}^* to reach the initial or adapted target \mathcal{T} .
- 2 To estimate on-demand some measures of risks ρ(Y*) during the time of the design.



Definition

A margin M is a quantity that aims at covering the risk that a given performance of a system measured by $\mathbf{Y}_{capable}$ does not reach the given target $\mathbf{Y}_{required}$

$$M = \mathcal{R}_{\mathcal{C}} \left(d(\mathbf{Y}_{capable}, \mathbf{Y}_{required}) \| \mathcal{I}, \mathcal{K} \right)$$

where

- $\blacksquare \ \mathcal{R}$ is the risk measure defined over the set of configurations \mathcal{C}
- d() measures the "distance"
- $\blacksquare \ \mathcal{I}$ is the information available when assessing the margin
- $\blacksquare \ \mathcal{K}$ is the knowledge available when assessing the margin

Examples

$$egin{aligned} M &= \max_{\mathcal{C}} \| \mathbf{Y}_{capable} - \mathbf{Y}_{required}) \| \ M &= q_{\mathcal{C}}^{lpha} (\mathbf{Y}_{capable} - \mathbf{Y}_{required}) \end{aligned}$$



November 23, 2016

Challenge C5: To develop Robust Optimization (1/4)

Goal:

Describe what has to be solved

Information:

Cost/Objective function(s) to minimize/maximize

$$\mathcal{J}: \quad (\mathbf{x}, \boldsymbol{\xi}) \in \mathcal{X} imes \Xi \mapsto \mathcal{J}(\mathbf{x}, \boldsymbol{\xi}) \in \mathbb{R}^k$$

- $\textbf{x} \in \mathcal{X} \subset \mathbb{R}^{\textit{P}} \textbf{=}$ parameters to optimize
- $\boldsymbol{\xi}\in\Xi\subset\mathbb{R}^{\mathcal{Q}}=$ variables supposed to be dispersive
- Set of Constraint functions $\mathbf{C} = (C_1, ..., C_p)$ on $\mathcal{X} \times \Xi$, Requirements \mathcal{R} and induced feasible Domain \mathcal{D}

$$\mathcal{D} = \{ \mathbf{x} \in \mathcal{X}, \, \mathbf{C}(\mathbf{x}, \boldsymbol{\xi}) \in \mathcal{R} \}$$

 Mathematical/Computational complexity of Objective/Constraint function(s) (Nonlinearity, smoothness, gradients availability, cost, approximations, etc.)

Challenge C5: To develop Robust Optimization (2/4)

Step RO-1: Problem Specification

Goal:

Describe what has to be solved. Information:

- Uncertainties identification:
 - What are the variables that would be subjected to uncertainties ?...
- Needs in terms of Robustness:
 - Which behaviour (w.r.t uncertainties) do we want to avoid in the cost function ?
- Needs in terms of Risk/Reliability:
 - What behaviour (w.r.t uncertainties) do we want to avoid in the constraints ?...



Challenge C5: To develop Robust Optimization (3/4)

Step RO-2: Robustness & Risk formulation

Goal:

Describe formally a robust and reliable version of the initial problem **Information**:

- **RO-2(1):** Quantify the sources of uncertainties on $\xi \in \Xi$, x, \mathcal{J} and C
- RO-2(2): Define a Robustness Measure ρ_J for Objective function(s)
- **RO-2(3):** Define a Risk Measure ρ_{C} for Constraint functions, an associated confidence region \mathcal{R}_{α} .



RBUS

Challenge C5: To develop Robust Optimization (3/4)

Step RO-3: Resolution

Goal:

Define efficient algorithmic strategy to solve the robust/reliable formulation problem

$$\widehat{\mathbf{x}} = \operatorname*{Argmin}_{\mathbf{x} \in \mathcal{D}_{\mathbb{P}^{\xi}, \rho_{\mathbf{C}}, \mathcal{R}_{\mathbf{\alpha}}}} \rho_{\mathcal{J}}(\mathcal{J}(\mathbf{x}, \boldsymbol{\xi}))$$

Information:

- What are the software resources (memory, cores, etc.) ?
- What is the error tolerance allowed ?
- Define a numerical strategy to compute the robust version of the objective function ρ_J(J(x, ξ))
- Define a numerical strategy to explore the domain $\mathcal{D}_{\mathbb{P}^x,\rho_{\mathsf{C}},\mathcal{R}_{\alpha}} \subset \mathcal{X}$

November 23, 2016

Challenge C5: To develop Robust Optimization (4/4)

Step RO-3: Resolution





C6: To develop Multi Disciplinary Uncertainty Quantification

Elements of information in a multi-disciplinary context

• A reference database $\left(\mathbf{Y}_{1}^{[j,*]},\cdots,\mathbf{Y}_{n}^{[j,*]}\right)$ or

 $\left((\mathbf{X}_{1}^{[j,*]}, \mathbf{Y}_{1}^{[j,*]}), \cdots, (\mathbf{X}_{n}^{[j,*]}, \mathbf{Y}_{n}^{[j,*]}) \right)$ that is enriched during the design cycle for each discipline $j \in J$.

- A set of risk measures $(\rho_1(\mathbf{Y}^{[j,*]}), \cdots, \rho_{d_j}(\mathbf{Y}^{[j,*]}))$ built upon $\mathbf{Y}^{[j,*]}$ to be estimated during the design cycle.
- A panoply of numerical models H^[j] = {h^[j]₁, · · · , h^[j]_{D_j}} that is enriched during the design cycle.
- A quantification of the uncertainties attached to the inputs of the numerical models represented by a statistical law $\mathbb{P}_{\mathbf{X}^{[j]}}$ that is enriched during the design cycle
- A definition of the target *T*^[j] and its associated level of confidence α^[j] to be reached that is enriched during the design cycle.
- A global computational budget B^[j] that can be allocated at different times of the design cycle.

C6: To develop Multi Disciplinary Uncertainty Quantification

How to go from deterministic MDO to probabilistic MDO ?





UQ&M in aero-engineering practices

Outline

1 Introduction

- Mathematical notations and link with statistical learning
- Uncertainty related to "modelling" activities
- Review of some UQ challenges
 Six UQ challenges
- 3 Conclusions and perspectives
 - A collaborative platform to share experience
 - Lessons learnt



UQ&M in aero-engineering practices Our collaborative platform : the Open TURNS software platfo





Lessons learnt

- This topic has emerged 10 years ago in our environment, mainly coming from academic side.
- Benefits are expected when coordinated approach is available @ industrial process lelvel and not only @ disciplinary level.
- Difficulty to transfer this technology: training !
- Still a lot of R&T topics to take into account the engineering usages of uncertainty quantification
- Many R&T communities to dialog with !



BACK-UP SLIDES



UQ&M in aero-engineering practices

November 23, 2016

Sources of uncertainty in this use-case (1/3)



Sources of uncertainty: system design variables

- Variability linked to thermal parameters (composite materials, junction, installation parameters).
- Lack of knowledge of the detailed behaviour of some electronic equipments (Printed Circuit Board, chips, ...).
- Complexity of the system definition.



Sources of uncertainty in this use-case (2/3)



Sources of uncertainty: environmental variables

- Likelihood of occurence of a lightning strike in a given area.
- Variability of the lightning strike (current level, signal shape, ...).
- Zoning of the attachment zone on the aircraft.



UQ&M in aero-engineering practices

November 23, 2016



- Approximation linked to the choice of the Navier-Stokes equation.
- Choice of the 3D <u>numerical scheme</u> (Finite Volume, nodal methods Method, ...).
- Coupling inside the zones and with control/command systems.
- <u>Non linear behaviour</u> of the radiating effects.

Sources of uncertainty: test uncertainties

- Representativity of on-ground test facilities.
- Reproductibility of test set-ups.
- <u>Calibration</u> of test devices.



....