

DE LA RECHERCHE À L'INDUSTRIE



www.cea.fr

Introduction aux approches V&V-UQ

G. Perrin

CEA/DAM/DIF, Arpajon, France

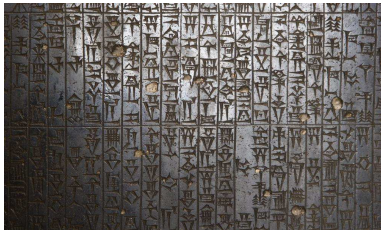
ETSN 2018 - Cargèse |

23 Avril 2018

- Acteur majeur d'aide à la prise de décision, le rôle de la simulation ne cesse de se renforcer pour la **conception** et la **garantie** de nouveaux objets.
- En effet, avec des puissances de calcul toujours plus importantes, des progrès constants en modélisation des phénomènes physiques, la simulation ne se limite plus à décrire et comprendre ce que l'on a observé, mais à **prédire** ce qui n'a pas encore été observé.
- Dans une perspective de minimisation des coûts, la simulation doit alors permettre de mieux adapter/justifier/réduire les marges de sécurité.



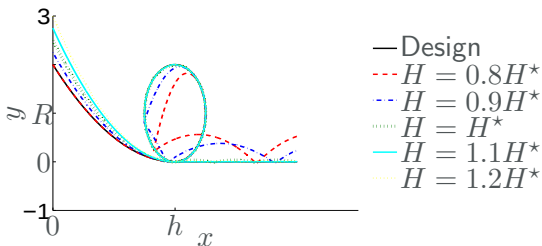
(a) Statue d'Hammurabi, roi de Babylone vers 2000 ans av. JC



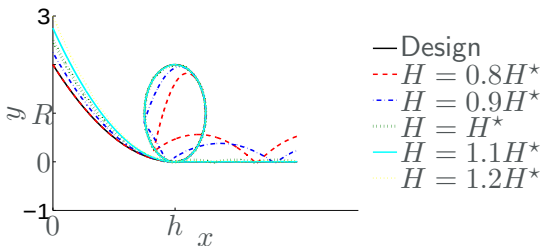
(b) Extrait du code d'Hammurabi

[...] Si un architecte a construit pour un autre une maison, et n'a pas rendu solide son œuvre, si la maison construite s'est écroulée, et a tué le maître de la maison, cet architecte est passible de mort. [...]

[...] Si c'est la fortune mobilière qu'il a détruite, il restituera tout ce qu'il a détruit, et parce qu'il n'a rendu solide la construction, et qu'elle s'est effondrée, il restaurera la maison ruinée, à ses propres frais. [...]



- Quels paramètres retenir pour la conception du looping ?
- Sur quels modèles (physiques, mathématiques...) se baser pour le dimensionnement ?



- Quels paramètres retenir pour la conception du looping ?
- Sur quels modèles (physiques, mathématiques...) se baser pour le dimensionnement ?
- En négligeant les frottements, on peut montrer par conservation d'énergie que la hauteur minimale de lancement du wagon est égale à $H^* = 5R/2$. Commenter.

- Pour pouvoir être prédictive, la simulation se doit de prendre en considération les différentes sources d'incertitudes potentiellement présentes.
- Une fois les incertitudes considérées, l'information fournie est plus **riche**. Au lieu de fournir une **unique** valeur déterministe, la simulation est en mesure:
 - de fournir une **confiance** pour des prédictions de quantités d'intérêt ;
 - d'associer une **probabilité d'occurrence** à différents scénarii.

Exemples

- indices de confiance pour les prévisions météorologiques ;
- probabilité de succès pour une opération médicale ;
- risque de ruine d'une structure...

Les approches de vérification/validation et de quantification des incertitudes (V&V-UQ) doivent ainsi être vues comme des outils efficaces pour cet enrichissement des résultats de simulation, en

- fournissant des méthodes permettant d'**intégrer au mieux l'information disponible** (avis d'experts, mesures expérimentales directes ou indirectes) pour construire des modèles adaptés des incertitudes ;
- proposant une **reformulation** (bien posée) **des problèmes classiques** pour leur permettre d'intégrer les incertitudes ;
- mettant à disposition des méthodes permettant de **résoudre efficacement** ces problèmes soumis à des incertitudes.

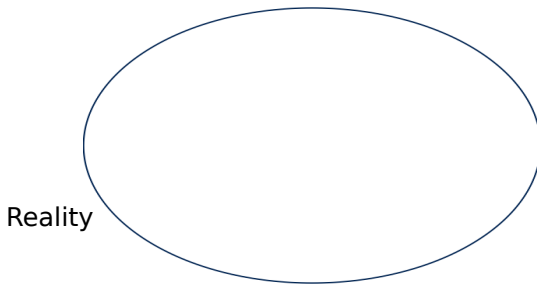
- 1 Introduction
- 2 Notations et définitions
- 3 L'approche ABC pour la prise en compte des incertitudes
- 4 Calibration, conception et garantie sous incertitudes
- 5 Quelques difficultés et questions ouvertes
- 6 Conclusion

Dans cette présentation, on notera:

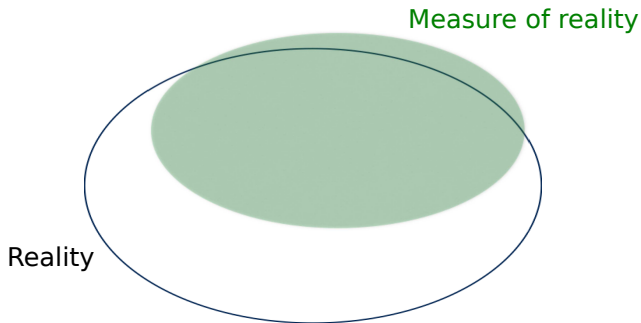
- $\mathcal{S}$ le système d'intérêt (physique, mécanique, économique, biologique,...), que l'on suppose pouvoir être décrit par un vecteur $x \in \mathbb{X}$ (dimensions, propriétés matériaux, équations d'état, conditions limites...).
- $y(x) \in \mathbb{Y}$ la quantité d'intérêt (QoI) permettant d'analyser le bon fonctionnement de $\mathcal{S}$ (on peut penser au maximum d'un effort, d'une pression, d'une accélération).
- x et y peuvent correspondre à des scalaires, des vecteurs, des fonctions...
- Les ensembles $\mathbb{X}$ (des entrées) et $\mathbb{Y}$ (des sorties) sont la plupart du temps des sous-ensembles de $\mathbb{R}^d$ et $\mathbb{R}$ respectivement.

L'application $x \mapsto y(x)$ est *a priori* **non-linéaire**.

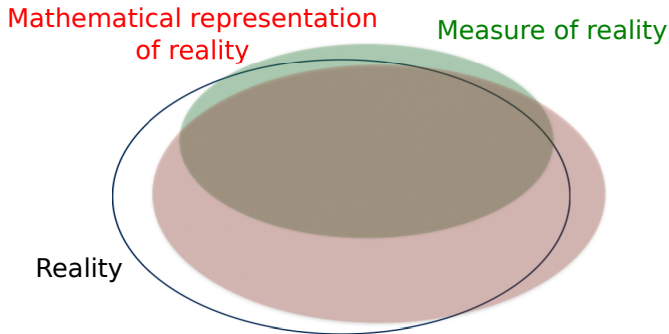
Différentes versions de la réalité conduisent à différentes sources d'incertitudes.



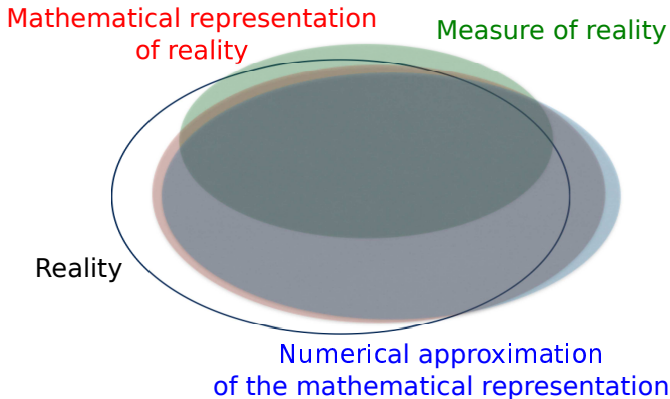
Différentes versions de la réalité conduisent à différentes sources d'incertitudes.



Différentes versions de la réalité conduisent à différentes sources d'incertitudes.



Différentes versions de la réalité conduisent à différentes sources d'incertitudes.



- On note y^{real} et y^{mes} les valeurs réelles et mesurées de y .
- On note $y^{\text{mod}}(\cdot; \beta)$ et $y^{\text{sim}}(\cdot; \beta, \delta)$ la représentation (mathématique) paramétrique de y et son approximation numérique (par discrétisation des équations par exemple).

Plusieurs sources d'incertitudes sur les sorties

- Contrôler l'erreur de mesure, $\varepsilon^{\text{mes}} := y^{\text{mes}} - y^{\text{real}}$, est du ressort de la **métrologie** (plus de détails dans le Guide to the expression of Uncertainty in Measurement (GUM)).
- Choisir δ afin de contrôler l'erreur numérique, $\varepsilon^{\text{num}} := y^{\text{mod}}(\cdot; \beta) - y^{\text{sim}}(\cdot; \beta, \delta)$, est l'objectif de la **vérification**.
- Identifier la valeur de β est l'objectif de la **calibration**.
- Contrôler l'erreur de simulation, $\varepsilon^{\text{sim}} := y^{\text{real}} - y^{\text{sim}}(\cdot; \beta, \delta)$, est le but de la **validation**.

Vérification

- Définition : le processus de déterminer si un modèle numérique est conforme au modèle mathématique qu'il est censé résoudre $\leftrightarrow$ "solving the equations right".
- Cela fait référence à de l'analyse numérique, en se basant classiquement sur des solutions de références et l'étude d'algorithmes.

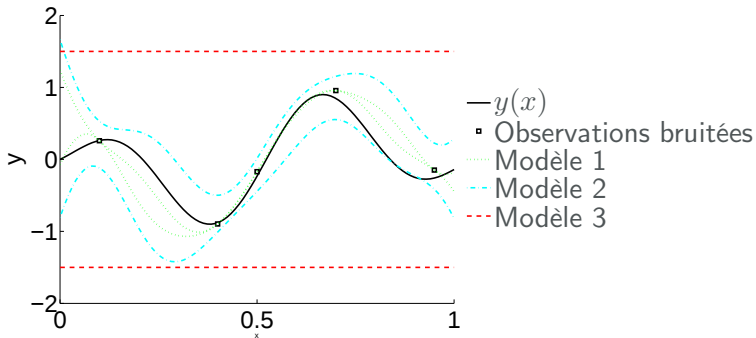
Validation

- Définition: le processus de déterminer si un modèle est représentatif de la réalité, dans son périmètre d'utilisation $\leftrightarrow$ "solving the right equations".
- Cela suppose classiquement de la confrontation à des mesures et l'utilisation de tests statistiques (test du χ^2 ...).

- Un code peut être **"valide" pour une application** et ne pas l'être pour une autre (d'où l'importance du choix de l'ensemble "représentatif").
- Un code peut être **"valide" pour la restitution de certains observables** et ne pas l'être pour d'autres.
- L'identification de l'erreur se fait sur un **ensemble d'apprentissage** et est vérifiée sur un ensemble de **validation** (le plus proche possible des conditions d'utilisation globale de l'objet).

Notations et objectifs

Attention au sur-apprentissage



- Quel modèle est "valide" ?
- Quel est le meilleur modèle ?

- Si l'on utilise la simulation pour analyser des systèmes pour lesquels peu ou pas d'expériences globales sont disponibles (déraillement de train, incident nucléaire, accident de voiture), le processus de validation est plus difficile.
- Dans ce cas, on effectue classiquement la validation d'un maximum de **sous-systèmes**, pour lesquels des mesures sont disponibles, et l'idée est ensuite d'accumuler systématiquement des "preuves" de bon fonctionnement du système global.
- Plusieurs guides (principalement aux USA) ont ainsi été proposés :
 - le CSAU (Code Scaling, Applicability and Uncertainty) est utilisé par la US Nuclear Regulatory Commission ;
 - le CAS (Credibility Assessment Scale) est développé par la NASA ;
 - le PMI (Predictive Maturity Index) est développé par Los Alamos ;
 - le PCMM (Predictive Capability Maturity Model) est mis-au-points au Sandia National Lab.

La valeur de x peut aussi être incertaine. On distingue généralement deux sources d'incertitudes pour x :

- les incertitudes **épistémiques** (réductibles) : des composantes déterministes de x ne sont pas parfaitement connues ;
- les incertitudes **aléatoires** (non-réductibles) : des composantes de x sont aléatoires par nature (dispersion de fabrication, vent, radioactivité naturelle...).

La **théorie des probabilités** est généralement considérée pour modéliser à la fois les incertitudes sur les entrées et les sorties.

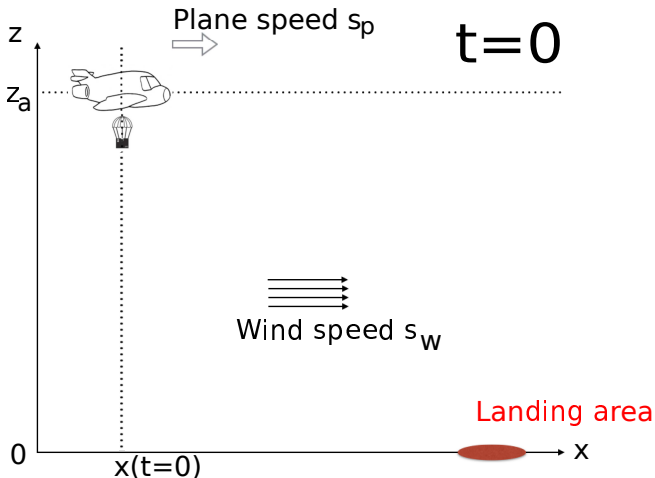
Cela conduit le plus souvent à des modèles probabilistes **hiérarchiques** pour $y^{\text{sim}}(x; \beta, \delta)$, du fait que :

- x est aléatoire ;
- $y^{\text{sim}}(x; \beta, \delta) | x$ est également aléatoire.

Notations et objectifs

L'exemple du parachutage d'un colis

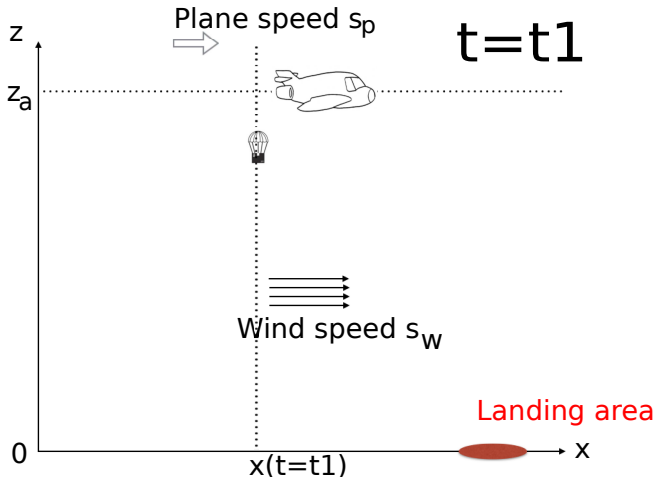
Comment localiser avec la simulation la position d'atterrissage d'un colis parachuté ?



Notations et objectifs

L'exemple du parachutage d'un colis

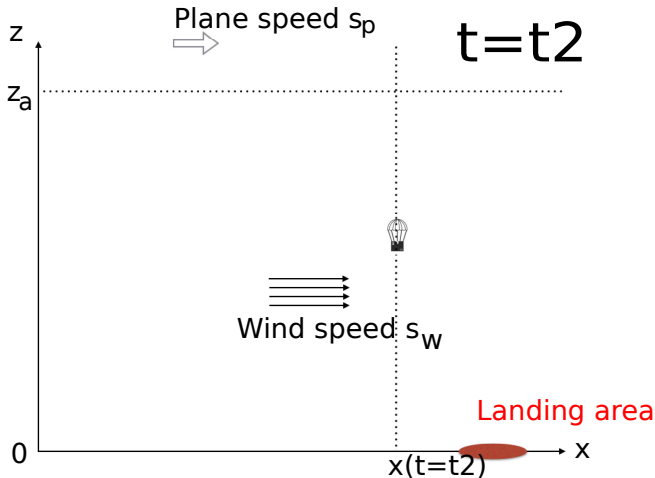
Comment localiser avec la simulation la position d'atterrissage d'un colis parachuté ?



Notations et objectifs

L'exemple du parachutage d'un colis

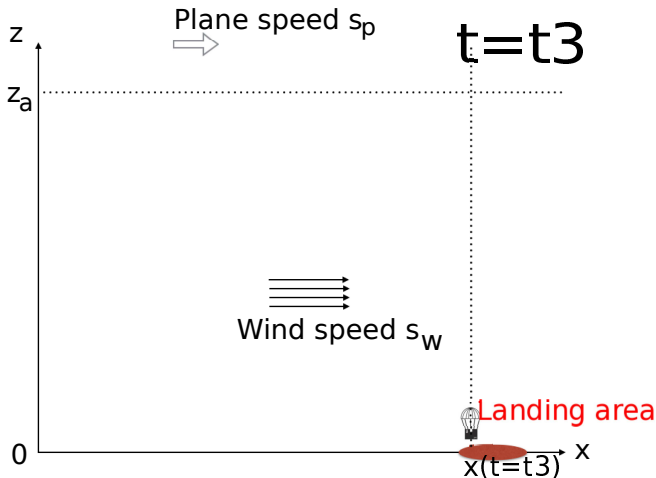
Comment localiser avec la simulation la position d'atterrissage d'un colis parachuté ?



Notations et objectifs

L'exemple du parachutage d'un colis

Comment localiser avec la simulation la position d'atterrissage d'un colis parachuté ?



Sorties

- y^{real} est la "vraie" position d'atterrissage.
- En notant $\beta = g$, la position "modélisée", $y^{\text{mod}}(\cdot; \beta)$, peut être vue comme la composante horizontale de $x(t)$ lorsque $\langle x(t), e_z \rangle = 0$:

$$\begin{cases} m\ddot{x} = mg - A(\dot{x} - s_w) \|\dot{x} - s_w\|, \\ \dot{x}(t=0) = s_p, \quad x(t=0) = (x_0, z_a). \end{cases}$$

- En notant $\delta = dt$, la position "simulée", $y^{\text{sim}}(\cdot; \beta, \delta)$, peut être vue comme la composante horizontale de $\hat{x}(t)$ lorsque $\langle \hat{x}(t), e_z \rangle = 0$:

$$\begin{cases} \hat{x}(t + dt) = \hat{x}(t) + \dot{\hat{x}}(t)dt, \\ \dot{\hat{x}}(t + dt) = \dot{\hat{x}}(t) + \left(g - \frac{1}{m} A(\dot{\hat{x}}(t) - s_w(t)) \|\dot{\hat{x}}(t) - s_w(t)\| \right) dt, \\ \dot{\hat{x}}(t=0) = s_p, \quad \hat{x}(t=0) = (x_0, z_a). \end{cases}$$

Notations et objectifs

L'exemple du parachutage d'un colis

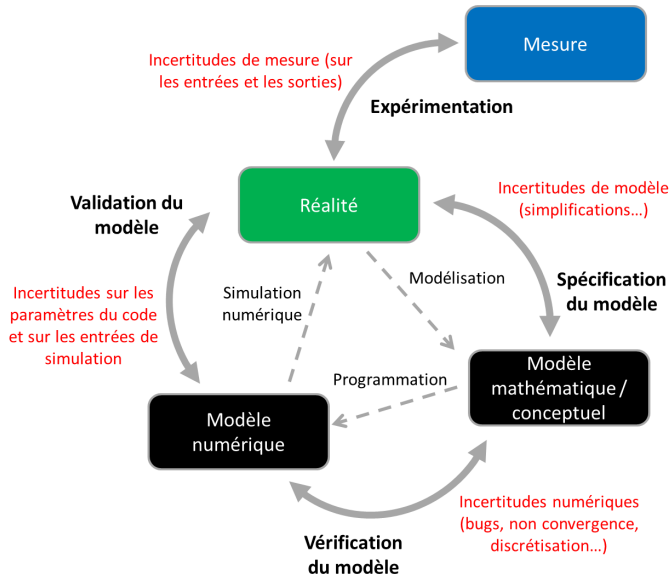
Entrées : $\mathbf{x} = (x_0, z_a, \mathbf{s}_p, \mathbf{s}_w, m, \mathbf{A})$

- (x_0, z_a) , la position de largage (épistémique).
- $\mathbf{s}_p$, la vitesse de l'avion à $t = 0$ (épistémique).
- $\mathbf{s}_w$, la vitesse du vent, qui dépend du temps (aléatoire).
- m , la masse du colis (épistémique).
- $\mathbf{A}$, la matrice des forces aérodynamiques (épistémique).

- **Vérification** $\leftrightarrow$ à $\beta, \mathbf{x}$ fixés, est-ce que le schéma numérique choisi permet à $y^{\text{sim}}(\mathbf{x}; \beta, \delta)$ de converger vers $y^{\text{mod}}(\mathbf{x}; \beta)$ lorsque dt tend vers 0 / comment choisir dt ?
- **Validation** $\leftrightarrow$ si $\mathcal{I}_\alpha$ désigne un intervalle de prédiction de $y^{\text{sim}}(\mathbf{x}; \beta, \delta)$ (intégrant les incertitudes sur $\mathbf{x}, \beta$ et l'erreur de convergence numérique), avec quelle confiance peut-on assurer que y^{real} soit dans $\mathcal{I}_\alpha$? / A-t-on besoin d'ajouter une erreur de modèle? Si oui, quelle structure d'erreur proposer ?

Notations et objectifs

Le "cercle des incertitudes" de Sargent



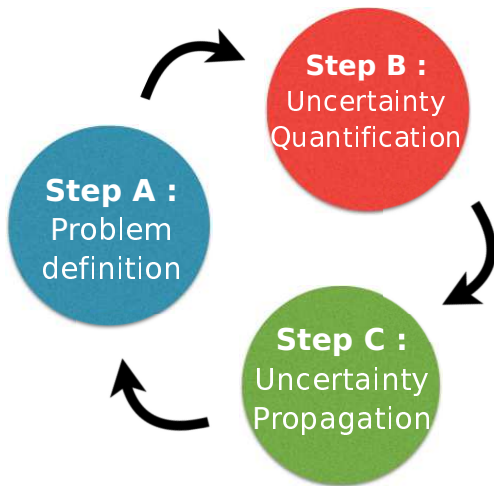
R. P. Feynman, "Seeking New Laws"

In general we look for a new law by the following process.

- First we guess it.
- Then we compute the consequences of the guess to see what would be implied if this law that we guessed is right.
- Then we compare the result of the computation to nature, with experiment or experience, compare it directly with observation, to see if it works.

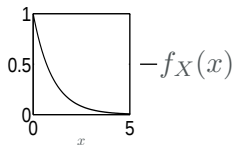
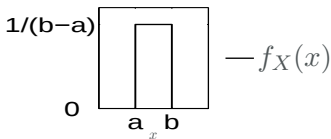
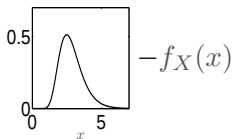
If it **disagrees with experiment it is wrong**. In that simple statement is the key to science. It does not make any difference how beautiful your guess is. It does not make any difference how smart you are, who made the guess, or what his name is. **If it disagrees with experiment it is wrong**. That is all there is to it.

- 1 Introduction
- 2 Notations et définitions
- 3 L'approche ABC pour la prise en compte des incertitudes
- 4 Calibration, conception et garantie sous incertitudes
- 5 Quelques difficultés et questions ouvertes
- 6 Conclusion

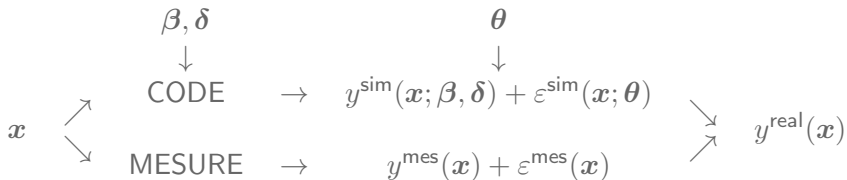


- 1 Quelle est la question ?
- 2 Quelles sont les quantités d'intérêt (QoI) ?
- 3 Quels modèles utiliser pour calculer ces QoI et répondre à la question ?
- 4 Quelles entrées et sorties pour ces modèles ?
- 5 Quels paramètres doivent être fixés pour pouvoir lancer ces modèles ?
- 6 Qu'est-ce qui peut manquer dans ces modèles ?
- 7 A-t-on des informations sur la structure de l'erreur de modèle qui peut affecter les sorties considérées ?
- 8 Quelle méthode est utilisée pour résoudre les équations des modèles ?
- 9 Sur quels paramètres sont basés ces solveurs ?
- 10 Peut-on évaluer l'erreur numérique de ces solveurs ?

- 1 Quels sont les différentes sources d'incertitudes (sur les entrées et les sorties) ?
- 2 De quelle information dispose-t-on sur ces incertitudes ?
 - jugements d'expert (contraintes de positivité, bornes physiques,...) ;
 - mesures expérimentales directes et/ou indirectes ?
- 3 Quel est le modèle le "plus adapté" pour caractériser ces incertitudes, et notamment la structure de dépendance (inférence statistique) ?
 - modèles paramétriques (Gaussien, uniforme, Beta, ...) ;
 - modèles non paramétriques (représentations à noyaux) ;
 - théories alternatives (ensembles flous, P-box, fonctions de croyance,...).



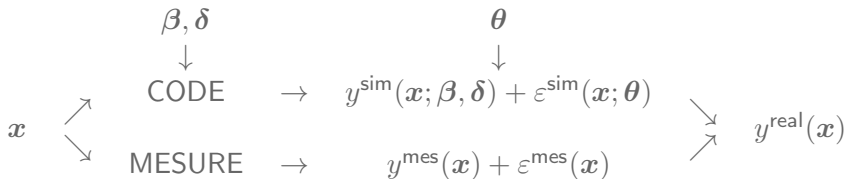
- 1 A-t-on accès aux équations sur lesquelles se base le modèle ?
 - oui : on peut considérer des méthodes intrusives ou non ;
 - non : seules les méthodes non-intrusives peuvent être considérées. On parle dans ce cas de cas "boîte noire".
 - 2 Le modèle est-il linéaire (ou faiblement non-linéaire) ? Si oui, on pourra partir sur des techniques de développement de Taylor.
 - 3 Le modèle est-il monotone ?
 - 4 A-t-on une quelconque information sur la régularité du modèle ?
- ⇒ Les techniques d'échantillonnage (telles que les approches Monte Carlo / Markov Chain Monte Carlo) sont généralement utilisées pour propager les incertitudes.
- ⇒ Moins le modèle est régulier, plus un grand nombre d'appels aux codes sont généralement nécessaires pour atteindre une précision donnée sur les résultats.



Problèmes directs

Connaissant x, β, δ, θ , on aimerait estimer des quantités statistiques de y^{real} , telles que:

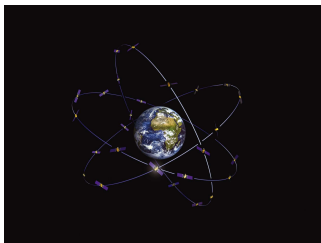
- sa moyenne et sa variance ;
- des probabilités de dépassement de seuil (**certification**) ;
- sa loi de probabilité complète (pour la propager dans d'autres modèles par exemple...).



Problèmes inverses

Connaissant x et y , on aimerait

- identifier les paramètres β, δ, θ (**vérification** et **validation**) ;
- classer l'influence de chaque entrée (ou groupe d'entrées) sur la variabilité de y (**analyses de sensibilité**) ;
- trouver la "meilleure" valeur de x vis-à-vis d'un certain critère sur y (**optimisation, conception**).



- 1 Connaissant la position de 5 satellites (incluant des incertitudes), comment calculer la position moyenne d'un récepteur GPS $\mathbf{R}$?
- 2 Quel serait le rayon de la sphère Ω centrée sur cette position moyenne telle que $\mathbb{P}(\mathbf{R} \in \Omega) = 90\%^1$?

¹Inspiré du concours 2014-2015 proposé par la fédération française des jeux mathématiques.

Step A : définition du problem

- On dispose de 5 satellites, et chaque satellite S_k envoie un signal, qui est analysé par R .
- Les horloges des satellites sont supposées parfaites et synchronisées (c'est le "temps absolu").
- L'horloge du récepteur peut différer de ce temps absolu d'une quantité τ , appelée "time shift" dans la suite.
- Des signaux envoyés par les satellites, le récepteur déduit les temps $t_k = \tau + \|S_k - R\| / c$, ou les distances $D_k = c\tau + \|S_k - R\|$, avec c la vitesse de la lumière (supposée connue).

Remarque: pour ce problème, on n'a pas besoin de codes numériques.

Step B : sources d'incertitudes

- τ peut être positif ou négatif, et est inconnu. On le suppose uniformément distribué sur $\mathbb{T}$.
- On suppose que chaque satellite renvoie une PDF pour D_k , $f_D = f_{D_1} \times \cdots \times f_{D_5}$. En pratique, cette PDF est donnée sous une forme discrète, et définie sur un sous-ensemble compact $\mathbb{D} \subset \mathbb{R}^5$.
- Les positions des satellites sont aussi incertaines. Les positions S_k sont néanmoins supposées indépendantes et distribuées (non-uniformément) dans des sphères $\mathbb{S}_k$, de centres et de rayons connus. Soient f_{S_k} les PDFs de S_k .

Remarque: toutes les incertitudes sont épistémiques, il n'y a pas d'aléatoire ici, juste du manque d'information...

Step C : propagation des incertitudes

Par construction:

- $\mathbb{E}[\mathbf{R}] = \int_{\mathbb{R}^3} \mathbf{r} f_{\mathbf{R}}(\mathbf{r}) d\mathbf{r},$
- $\mathbb{P}(\mathbf{R} \in \Omega) = \int_{\Omega} f_{\mathbf{R}}(\mathbf{r}) d\mathbf{r}.$

⇒ l'étape "clé" est la construction de $f_{\mathbf{R}}$.

⇒ Pour cela, on peut remarquer que pour tout $\mathbf{r}$:

$$\begin{aligned} f_{\mathbf{R}}(\mathbf{r}) &= \int_{\mathbb{T} \times \mathbb{S}_1 \times \dots \times \mathbb{S}_5} f_{\mathbf{R}, \tau, \mathbf{S}_1, \dots, \mathbf{S}_5}(\mathbf{r}, t, \mathbf{s}_1, \dots, \mathbf{s}_5) dt d\mathbf{s}_1 \dots d\mathbf{s}_5, \\ &= \int_{\mathbb{T} \times \mathbb{S}_1 \times \dots \times \mathbb{S}_5} f_{\mathbf{D}}(ct + \|\mathbf{r} - \mathbf{s}_1\|, \dots, ct + \|\mathbf{r} - \mathbf{s}_5\|) dt d\mathbf{s}_1 \dots d\mathbf{s}_5, \\ &= \mathbb{E} \left[\frac{f_{\mathbf{D}}(c\tau + \|\mathbf{r} - \mathbf{S}_1\|, \dots, c\tau + \|\mathbf{r} - \mathbf{S}_5\|)}{f_{\tau}(\tau) f_{\mathbf{S}_1}(\mathbf{S}_1) \dots f_{\mathbf{S}_5}(\mathbf{S}_5)} \right]. \end{aligned}$$

Step C: propagation des incertitudes - résolution pratique

- Le support des PDFs de $\mathbf{S}_k$, τ et $\mathbf{D}$ étant compact, le support de $f_{\mathbf{R}}$ est aussi compact.
- On utilise alors des techniques d'échantillonnage pour évaluer la PDF en un nombre bien choisi de points de $\mathbb{R}^3$.
- La PDF peut ensuite être post-traitée pour calculer $\mathbb{E}[\mathbf{R}]$ et trouver Ω .

Application numérique

- L'incertitude sur τ est souvent de $\pm 10^{-9}s$, le rayon des sphères pour la position des satellites est souvent choisi égal à $2m$, tandis que l'incertitude sur la distance entre le récepteur et les satellites est souvent de l'ordre de $\pm 10m$ (pour des distances supérieures à $20000km$).
- Pour ces valeurs, on trouve que le rayon de Ω , tel que $\mathbb{P}(\mathbf{R} \in \Omega) = 90\%$, est d'environ $2m$.

- 1 Introduction
- 2 Notations et définitions
- 3 L'approche ABC pour la prise en compte des incertitudes
- 4 Calibration, conception et garantie sous incertitudes
- 5 Quelques difficultés et questions ouvertes
- 6 Conclusion

Equation fondamentale

$$y^{\text{real}}(\mathbf{x}) = y^{\text{sim}}(\mathbf{x}; \boldsymbol{\beta}, \boldsymbol{\delta}) + \varepsilon^{\text{sim}}(\mathbf{x}) = y^{\text{mes}}(\mathbf{x}) + \varepsilon^{\text{mes}}(\mathbf{x}).$$

Information disponible

Un ensemble de mesures, $y^{\text{mes}}(\mathbf{x}^{(n)})$, $1 \leq n \leq N$, avec les incertitudes associées, que l'on regroupe dans un vecteur $\mathbb{Y}$.

Problématique

Dans quelle mesure les problématiques de calibration, de conception et de garantie sont impactées par la prise en compte des incertitudes ?

Formalisme déterministe

- Dans une calibration déterministe, on s'intéresse (le plus souvent) à une **unique** valeur de β .
- Classiquement, cette valeur est cherchée comme minimum de :

$$\beta^* \approx \arg \min_{\beta} \mathcal{C}(y^{\text{mes}}(x^{(n)}), y^{\text{sim}}(x^{(n)}; \beta, \delta)),$$

où $\mathcal{C}$ est une fonction coût à définir (par exemple les moindres carrés), pouvant intégrer les erreurs de mesure sous la forme de pondération.

Formalisme bayésien

- Les paramètres inconnus β sont supposés aléatoires pour modéliser le fait qu'ils sont *a priori* inconnus.
- L'idée est alors de rechercher les **plages de valeurs vraisemblables** pour β sachant les observations (inférence bayésienne) :

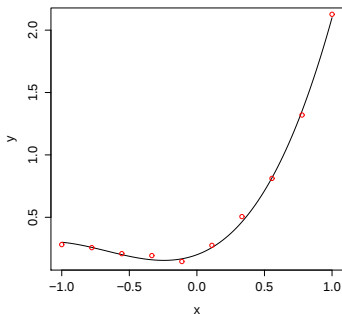
$$\pi[\beta|Y] \propto \pi[\beta]\pi[Y|\beta].$$

Exemple

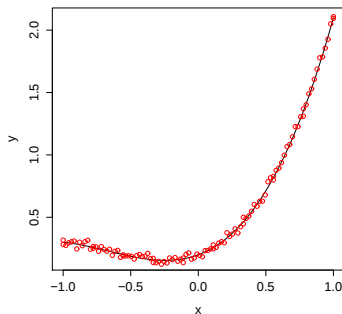
$$y^{\text{real}}(x) = 0.2 + 0.4x + x^2 + 0.5x^3, \quad -1 \leq x \leq 1, \quad \sigma_{\text{mes}} = 0.02.$$

$$y^{\text{sim}}(x, \beta) = \mathbf{h}(x)^T \beta, \quad \beta \sim \mathcal{N}(\mathbf{0}, [I_4]), \quad y^{\text{mes}}(x) \sim \mathcal{N}(y^{\text{real}}(x), \sigma_{\text{mes}}^2).$$

$N = 10$



$N = 100$



Exemple de calage sans erreur de modèle

$$y^{\text{real}}(x) = 0.2 + 0.4x + x^2 + 0.5x^3, \quad -1 \leq x \leq 1, \quad \sigma_{\text{mes}} = 0.02.$$

	β_1	β_2	β_3	β_4
référence	0.2	0.4	1	0.5
$\hat{\beta}, N = 10$	0.19 ± 0.019	0.45 ± 0.050	1.0 ± 0.035	0.45 ± 0.064
$\hat{\beta}, N = 100$	0.20 ± 0.006	0.40 ± 0.017	1.0 ± 0.013	0.51 ± 0.025

Table: Sans erreur de modèle: $\mathbf{h} = (1, x, x^2, x^3)$ (IC 95%)

	β_1	β_2	β_3	β_4
référence	0.2	0.4	1	0.5
$\hat{\beta}, N = 10$	0.21 ± 0.019	0.77 ± 0.019	0.98 ± 0.035	
$\hat{\beta}, N = 100$	0.20 ± 0.0059	0.71 ± 0.0067	1.0 ± 0.013	

Table: Avec erreur de modèle : $\mathbf{h} = (1, x, x^2)$ (IC 95%)

Avec erreur de modèle, $\hat{\beta}$ ne converge pas forcément vers β !

Exemple de calage avec erreur de modèle

$$y^{\text{real}}(x) = 0.2 + 0.4x + x^2 + 0.5x^3, \quad -1 \leq x \leq 1, \quad \sigma_{\text{mes}} = 0.02.$$

	β_1	β_2	β_3	β_4
référence	0.2	0.4	1	0.5
$\hat{\beta}, N = 10$	0.26 ± 0.52	0.74 ± 0.55	0.81 ± 0.82	
$\hat{\beta}, N = 100$	0.25 ± 0.51	0.72 ± 0.52	0.85 ± 0.79	

Table: $\mathbf{h} = (1, x, x^2)$, covariance stationnaire (IC 95%)

	β_1	β_2	β_3	β_4
référence	0.2	0.4	1	0.5
$\hat{\beta}, N = 10$	0.19 ± 0.027	0.53 ± 0.097	1.1 ± 0.31	
$\hat{\beta}, N = 100$	0.20 ± 0.0096	0.40 ± 0.051	0.99 ± 0.23	

Table: $\mathbf{h} = (1, x, x^2)$, covariance non stat. en $(x^3)^2$ (IC 95%)

Avec une "mauvaise" structure d'erreur de modèle, $\hat{\beta}$ ne converge pas forcément vers β !

Identification de l'erreur de modèle

- Pour bien estimer β , il est nécessaire de bien caractériser l'erreur de modèle $\Rightarrow$ dans ce cas, on cherche à **minimiser les compensations d'erreur**.
- De manière générale, cette erreur est **inconnue** et doit être identifiée à partir d'avis d'expert et de mesures les plus représentatives des conditions d'utilisation du système.
- L'erreur de modèle peut être placée sur les entrées ou sur les sorties.
- Dans ce cas, il est primordial de ne pas "mélanger" les estimations de β et de l'erreur de modèle : attention aux problèmes d'identifiabilité (séparer l'information disponible en données pour la calibration et données pour la validation) !
- Attention, pour un meilleur modèle prédictif, on peut être **intéressé par des compensations d'erreur** !

Conception déterministe

D'un point de vue général, concevoir un objet revient à résoudre un problème d'optimisation résultant des étapes A et B :

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathbb{X}} y(\mathbf{x}), \quad \text{avec :}$$

- $\mathbb{X}$ est un sous-ensemble donné de $\mathbb{R}^d$, qui peut intégrer différents types de contraintes (sur les entrées ou les sorties) ;
- la QoI y est une fonction coût déterministe adaptée au problème, associée à un ou plusieurs codes donnés.

Conception sous incertitudes

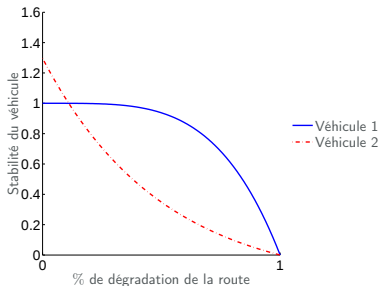
- Lorsque l'on considère des incertitudes, $y(x)$ devient aléatoire.
- Au lieu de chercher son minimum, on se concentre généralement sur une quantité statistique de y (sa moyenne par exemple).



Quel est le véhicule **le plus robuste** ?

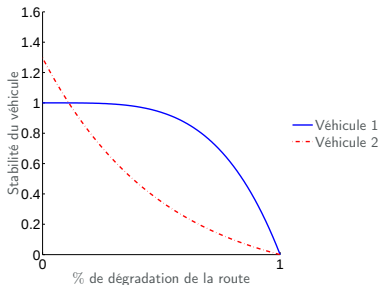
Qu'est ce qu'on appelle un système robuste ?

⇒ Se dit d'un système capable de bien se comporter dans des conditions **non nominales** (en raison d'incertitudes notamment).



Quel est le véhicule **le plus robuste** ?

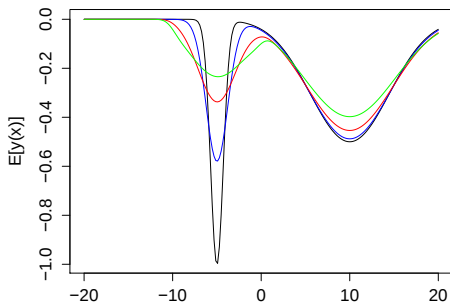
Quand chercher des structures robustes ?



Quel est le "meilleur" véhicule ?

Exemple 1 : incertitudes sur les entrées seulement

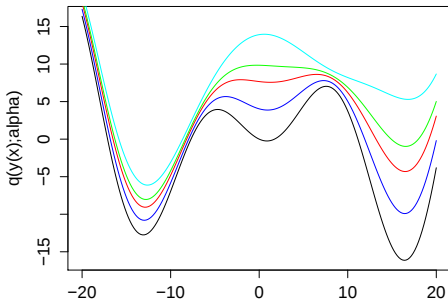
$$x^* = \arg \min_{x \in [-20, 20]} \mathbb{E} [y(x + \delta)], \quad \delta \sim \mathcal{N}(0, \sigma^2)$$



Noir : $\sigma = 0$ / Bleu : $\sigma = 1$ / Rouge : $\sigma = 2$ / Vert : $\sigma = 3$.

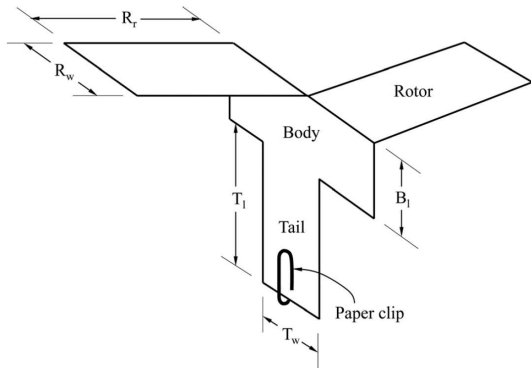
Exemple 2 : incertitudes sur les sorties seulement

$$x^* = \arg \min_{x \in [-20, 20]} q[y(x); \alpha],$$



Noir : $\alpha = 50\%$ / Bleu : $\alpha = 75\%$ / Rouge : $\alpha = 90\%$ / Vert : $\alpha = 95\%$
/ Cyan : $\alpha = 99\%$.

Exemple 3 : incertitudes sur les entrées et les sorties



Paramètres:

- Dimensions R_r , R_w , T_l , B_l , T_w .
- Masses de l'hélicoptère et du trombone.

En partant d'une feuille A4 standard, on aimerait maximiser le temps de chute d'un hélicoptère lâché d'une hauteur de 5m, portant un trombone.

Quelles sources d'incertitudes? Quel design pour quel objectif?

Garantie déterministe

- D'un point de vue **déterministe**, garantir le bon fonctionnement du système peut se ramener à garantir que $y^{\text{réel}}(x^{\text{réel}}) \leq S$, où la valeur de S est une donnée du problème (fournie par un client, une norme...).
- En se basant sur un code aux entrées/sorties incertaines, cela revient à garantir que $\mathbb{P}(y(x) > S) = 0$, ce qu'il n'est quasiment jamais possible de faire en pratique.

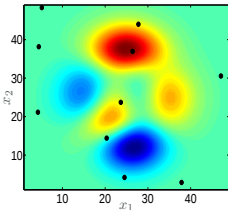
Garantie probabiliste

- Pour un risque $0 \leq \alpha \leq 1$, une version **probabiliste** plus *réaliste* de cette garantie consiste à démontrer que : $P_f := \mathbb{P}(y(x) > S) \leq \alpha$.

- 1 Introduction
- 2 Notations et définitions
- 3 L'approche ABC pour la prise en compte des incertitudes
- 4 Calibration, conception et garantie sous incertitudes
- 5 Quelques difficultés et questions ouvertes
- 6 Conclusion

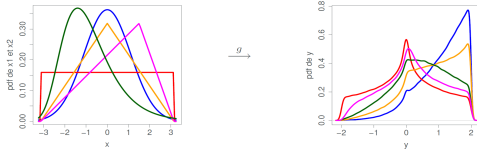
(1/4) Coûts numériques

- La plupart des méthodes de propagation des incertitudes sont basées sur le traitement d'un grand nombre de simulations ($\sim 10^4 - 10^6$).
- Lorsque le coût numérique de chaque évaluation est élevé (de plusieurs minutes à plusieurs jours), il est nécessaire de coupler ces approches à des techniques de métamodélisation.
- En remplaçant le code par un métamodèle, on introduit une nouvelle source d'incertitude, qui doit elle-même être précisément contrôlée.



(2/4) Modélisation des incertitudes des entrées/sorties

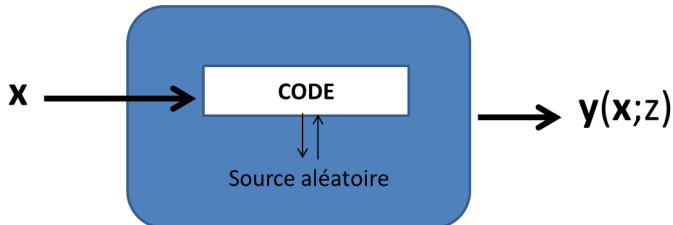
- La propagation des incertitudes dépend complètement de la modélisation des sources d'incertitudes.
- Dans la plupart des applications industrielles, ces incertitudes sont difficiles à caractériser. C'est particulièrement vrai lorsque l'information disponible est réduite, et que l'on considère des entrées et sorties de grandes dimensions (grands vecteurs, champs aléatoires...).
- Cette incertitude supplémentaire sur les lois de probabilité peut également être prise en compte dans les processus de validation.



$$g(x) = \cos(x_1) + \cos(x_2)$$

(3/4) Codes stochastiques

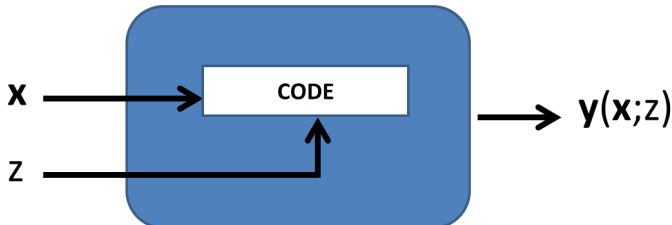
Cas 1 : le **simulateur stochastique**.



- La source aléatoire n'est pas accessible directement.
- Seul x peut être choisi: $y(x; z)$ est observé sans contrôle de z .
- La loi de z n'est pas connue explicitement.

(3/4) Codes stochastiques

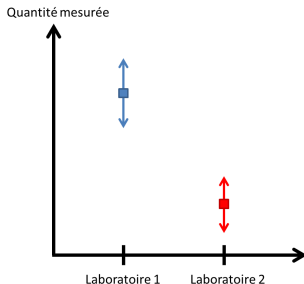
Cas 2 : la présence de **variables d'environnement**.

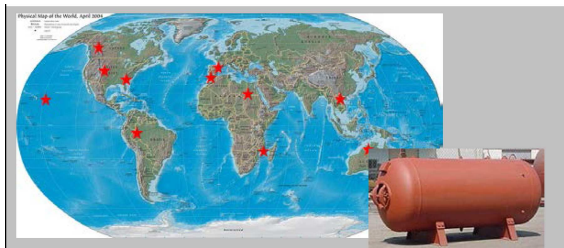


- Le simulateur reste déterministe.
- La loi de z est connue.
- Il est possible de spécifier à la fois les valeurs de x et z .

(4/4) Agrégation de données

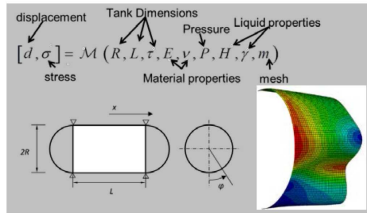
Dans les procédures V&V-UQ, une dernière difficulté importante vient de l'agrégation d'informations issues d'expériences différentes, d'experts différents, de codes numériques différents, donnant parfois l'impression d'être incohérentes...





- Plusieurs conteneurs contenant un "mystérieux liquide" sont répartis sur le globe.
- Aucun conteneur n'a jamais fui, mais durant une opération de surveillance, il s'avère qu'un conteneur dépasse une spécification de sécurité.
- ⇒ Y a t-il un risque de fuite pour ces conteneurs ? Doit-on tous les retirer pour éviter un accident ?
- ⇒ Comment exploiter des données de simulation et d'expériences pour guider la décision finale ?

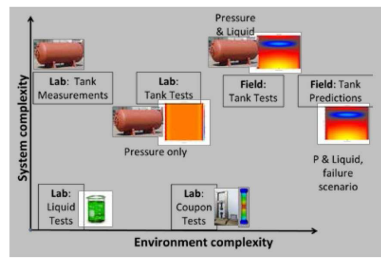
Modèle numérique



4 maillage éléments finis de la même géométrie sont fournis, basés sur 4 discrétisations différentes.

- ⇒ Il n'y a pas de "vraie" réponse.
- ⇒ Le cheminement logique proposé est souvent plus intéressant que le résultat final.

Données expérimentales



Des dispersions de fabrication sont également fournies.

- 1 Introduction
- 2 Notations et définitions
- 3 L'approche ABC pour la prise en compte des incertitudes
- 4 Calibration, conception et garantie sous incertitudes
- 5 Quelques difficultés et questions ouvertes
- 6 Conclusion

- Le formalisme et les notations des approches V&V-UQ basées sur la théorie des probabilités a été présenté et illustré sur une série d'exemples.
- Il existe des théories alternatives pour construire des intervalles de prédiction.
- Les approches V&V-UQ conduisent souvent à davantage de questions et de discussion que des réponses tranchées, et doivent être perçues comme des outils d'aide à la décision.
- Les approches V&V-UQ n'ont pas pour objectif de remplacer le modèle (physique, mécanique, économique...), mais de maximiser l'information que l'on peut en extraire.
- Bien caractériser les incertitudes ne signifie pas forcément les réduire...

Merci pour votre attention!

