Paradigme bayésien et traitement des incertitudes

A. PASANISI

La problématique « Incertitudes » à EDF-R&D

Traitement des incertitudes à EDF-R&D

Problématique :

- $\underline{z}$: variable d'intérêt → enjeux décisionnels (ex. débit de crue d'une rivière, émissions CO_2 d'une centrale thermique, prix demain à 7h d'1 MWh électrique …)
- $\underline{z}$ est issue d'un modèle déterministe (physique, économique …) dont certaines variables d'entrée ($\underline{x}$) sont incertaines
- Objectif : quantifier proprement l'incertitudes sur la variable d'intérêt $\underline{z}$

Modèle déterministe :

$$\underline{z} = G(\underline{x}, \underline{d})$$

$\underline{z}$: variable d'intérêt → décision

$\underline{x}$: variables d'entrée incertaines (en général, v.a.)

$\underline{d}$: variables d'entrée non incertaines

Cadre méthodologique

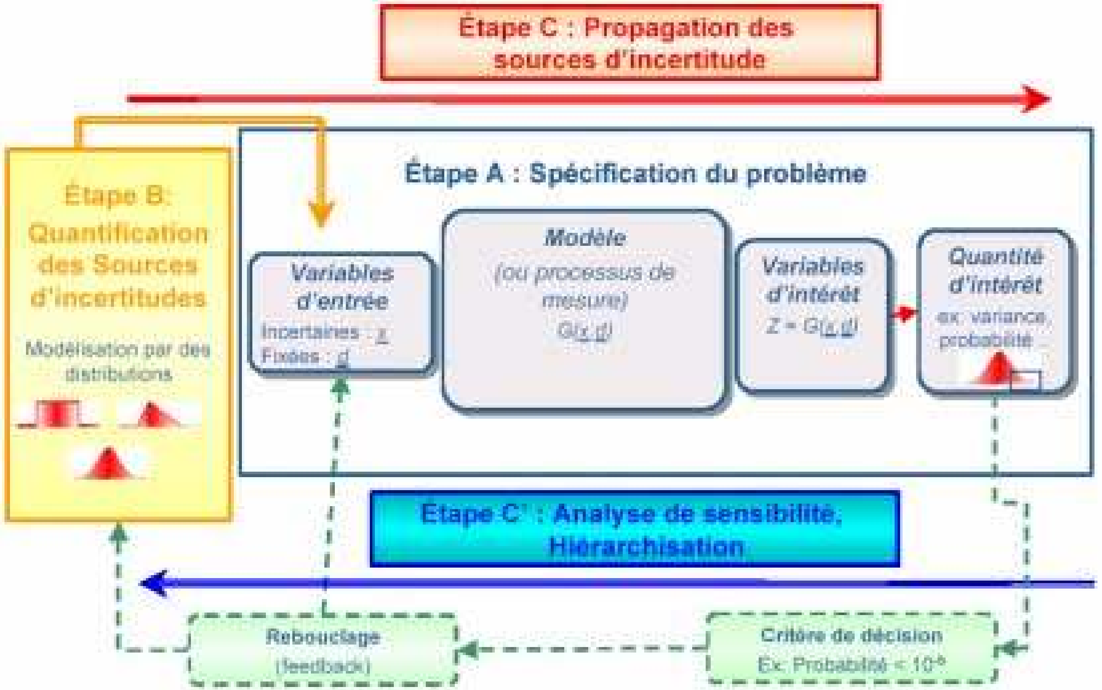
Grandes étapes d'une étude d'incertitudes :

a) Spécification du problème : Quel modèle G ? Quelle variable d'intérêt ? Quelle quantité associée à la variable $\underline{z}$ (moyenne, éc. type, prob. de dépassement …) ?

b) Modélisation des sources d'incertitudes → établir la loi de probabilité de $\underline{x}$

c) Propagation des sources d'incertitudes de $\underline{x}$ vers $\underline{z}$ (méthodes Monte-Carlo adaptées …)

d) Analyse de sensibilité, hiérarchisation des sources d'incertitudes



Un sujet de recherche très actif

CSDL – Complex System Design Lab

Thalès, DA, Renault, EADS, MDBA, INRIA, ONERA, ECP, SUPELEC, École des Mines

DICE consortium

Ecole Nationale Supérieure des Mines SAINT-ETIENNE

- Univ. Paul Cézanne
- Univ. d'Orsay
- Univ. Joseph Fourier
- Univ. of Florida

- TOTAL
- IRSN
- ONERA
- RENAULT
- EDF

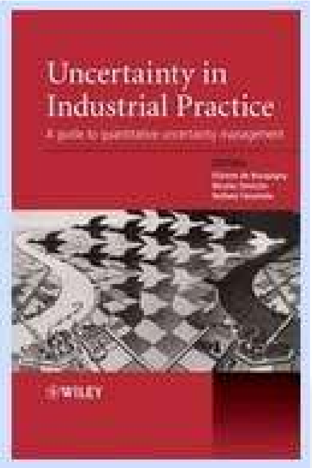
Groupement De Recherche MASCOT NUM

Méthodes d'Analyse Stochastique pour les COdes et Traitements NUMériques

IMdR Institut pour la Maîtrise des Risques
Sûreté de Fonctionnement - Management - Cindyniques

ESReDA European Safety, Reliability & Data Association

Un sujet de recherche très actif



Ouvrage ESReDA : traitement des incertitudes dans la pratique industrielle

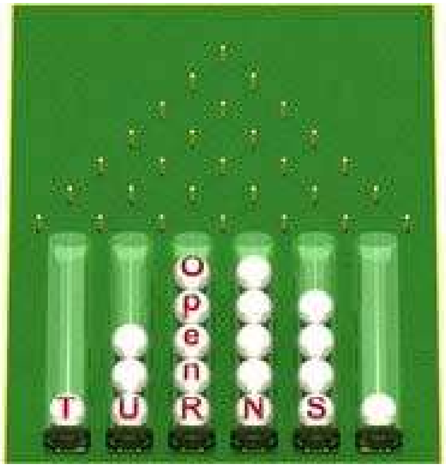


Plate-forme de calcul probabiliste et propagation d'incertitudes Open TURNS (EDF-EADS-Phimeca)

R&D « amont » (thèses, Post-doc, publications scientifiques)

- Méthodes Monte-Carlo avancées pour l'estimation de faibles probabilités
- Modélisation de la dépendance statistique par la théorie des copules
- Modélisation inverse probabiliste
- Application de la théorie de Dempster-Schäfer
- Chaos polynomial

Paradigme bayésien et incertitudes : quelques idées clés

Paradigme bayésien

Elements de base :

• Données x (réalisations d'une v.a. observable)

• Un modèle statistique paramétré par θ

• Les paramètres θ sont inconnus mais on dispose d'une connaissance préalable (a priori) exprimée sous la forme d'une distribution de probabilité $\pi(\boldsymbol{\theta})$

• La formule de Bayes donne la distribution de probabilité de θ conditionnellement aux données observées $\pi(\theta|x)$

$$\underbrace{\pi(\theta \mid x)}_{\text{Loi a posteriori de } \boldsymbol{\theta}} = \frac{\overbrace{\pi(\theta)}^{\text{Loi a priori de } \boldsymbol{\theta}} \cdot \overbrace{f(x \mid \theta)}^{\text{Vraisemblance}}}{\underbrace{\int \pi(\theta) \cdot f(x \mid \theta) \mathrm{d}\theta}_{\text{Constante de normalisation}}}$$

Paradigme bayésien et incertitudes

- La formule de Bayes donne la distribution de probabilité a posteriori $\pi(\theta|x)$ des paramètres du modèle.
- Cette incertitude se propage naturellement sur la variable observable x et sur toute autre v.a. issue d'un modèle paramétré par θ

$$\underbrace{\pi(\tilde{x} \mid x)}_{\substack{\text{Distribution} \\ \text{« prédictive » de la} \\ \text{v.a observable x}}} = \int \pi(\tilde{x}, \theta \mid x) d\theta = \int \underbrace{\pi(\tilde{x} \mid \theta)}_{\substack{\text{Modèle} \\ \text{statistique}}} \cdot \underbrace{\pi(\theta \mid x)}_{\substack{\text{Distribution a} \\ \text{posteriori de } \theta}} d\theta$$

La distribution prédictive intègre :
- L'incertitude d'échantillonnage (pour θ fixées la valeur de **x** est incertaine)
- L'incertitude d'estimation sur les paramètres θ, décrite par la distrib. a posteriori

L'incertitude décrite par la loi prédictive peut être propagée sur toute variable z, liée à x par un modèle déterministe : $z = G(x, d)$

Paradigme bayésien et incertitudes

Le cadre bayésien est particulièrement adapté au traitement et à la propagation des incertitudes :

- **L'approche bayésienne est, par essence, une approche de niveau 2 :**
  - Incertitude sur les variables observables X (aléatoire).
  - Incertitude sur les paramètres θ des lois de probabilités de X (épistémique).
  - Éventuellement, il est possible de rajouter un troisième niveau d'incertitude. sur le modèle statistique (facteurs de Bayes, cf. exemple).
  - Ces différents niveaux d'incertitudes se propagent **naturellement** par marginalisation d'une loi jointe de probabilité.

- **L'approche bayésienne est adaptée aux applications industrielles :**
  - Souvent, peu de données mais expertise disponible.
  - Historique important à EDF-R&D (MRI) dans le domaine de la fiabilité des composants (Lannoy, Procaccia, Clarotti, Celeux ...)

Perspectives

Une thèse est déjà prévue (début d'un stage en février, puis thèse en septembre 2009)

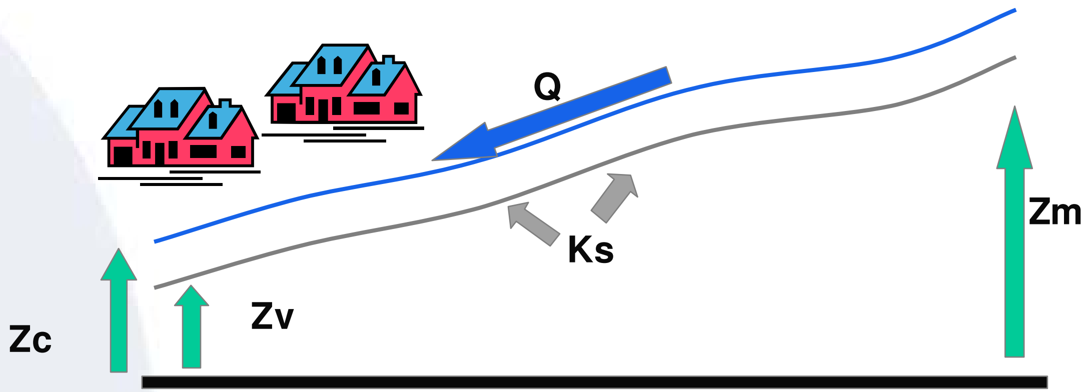
Montrer par l'exemple la pertinence de l'approche bayésienne dans un contexte industriel

- Analyse et étude de 3 cas issus de différents champs d'application (ex. fiabilité, hydraulique, thermique des bâtiments …).
- Déploiement des méthodes et outils de l'analyse bayésienne (élicitation, inférence par méthodes Monte-Carlo, choix de modèle, propagation).
- Comparaison avec les résultats obtenus avec une approche fréquentiste de niveau 2
- …

Quelques résultats d'une première étude (Andrianov, Fisher, Nabahane, 2008)

Un modèle analytique simple

Problème : protection d'un centre habité vis-à-vis du risque « inondation »



- **Zc : côte maximale atteinte par la crue (variable d'intérêt)**
- Zm et Zv : côtes du fond du cours d'eau (v.a.)
- Q : débit du cours d'eau (v.a)
- Ks : coefficient de frottement de Strickler (v.a.)
- B, L : Largeur et longueur du tronçon du cours d'eau (déterministe)

$$Zc = \left(\frac{Q}{K_s \cdot \sqrt{\frac{Z_m - Z_v}{L}} \cdot B} \right)^{3/5}$$

Problème

On veut évaluer la probabilité que la côte Zc dépasse la hauteur Zd (fixée) d'une digue de protection

L'incertitude sur Zm et Zv est décrite par des lois triangulaires.
L'incertitude sur Ks est décrite par une loi normale.
Pour décrire l'incertitude sur Q, on teste plusieurs variables :

- Gumbel
- Log-Normale
- Normale (on choisit volontairement une loi qui n'est pas adaptée)

Dans ce cas, aucune loi n'est rejetée par les tests d'ajustement classiques ! En théorie, on serait tenté de prendre le résultat le plus conservatif !

Avec la technique des facteurs de Bayes, il est possible d'évaluer la « probabilité a posteriori » associée à chacun des 3 modèles en compétition

Probabilités à posteriori des modèles

Ensemble de n modèles en compétition M_1, M_2 , ... , M_n paramétrés par θ_1, θ_2 , ... , θ_n

$$B(M_i, M_j) = \frac{P(M_i)}{P(M_j)} \cdot \frac{\int f(D|\theta_i, M_i) \cdot \pi(\theta_i|M_i) d(\theta_i)}{\int f(D|\theta_j, M_j) \cdot \pi(\theta_j|M_j) d(\theta_j)}$$

Facteurs de Bayes

$$\frac{P(M_i) \cdot \int f(D|\theta_i, M_i) \cdot \pi(\theta_i|M_i) d(\theta_i)}{\sum_j P(M_j) \cdot \int f(D|\theta_j, M_j) \cdot \pi(\theta_j|M_j) d(\theta_j)}$$

Probabilités a posteriori des n modèles

Deux façons d'utiliser ces probabilités :

- On choisit le modèle le plus probable
- On introduit une incertitude "de niveau 3" sur le modèle → mélange des 3 modèles selon les probabilités respectives (Bayesian Model Averaging, Raftery)

Quelques résultats

Probabilités de surverse pour une valeur fixée de la hauteur de la digue de protection (espérance a posteriori)

Zd (m)	Q modelling	P(S<0|Zd, Q)
57.8	Gumbel	$1.18*10^{-3}$
57.8	Normal	$7.4*10^{-4}$
57.8	LogNormal	$2.58*10^{-3}$
57.8	Bayesian MA (gaussian prior)	$1.40*10^{-3}$
57.8	Bayesian MA (uniform prior)	$1.13*10^{-3}$

L'utilisation du modèle log-normal aurait mené à une surestimation de la prob. de défaillance, ou à un surdimensionnement de la digue (à probabilité fixée)