

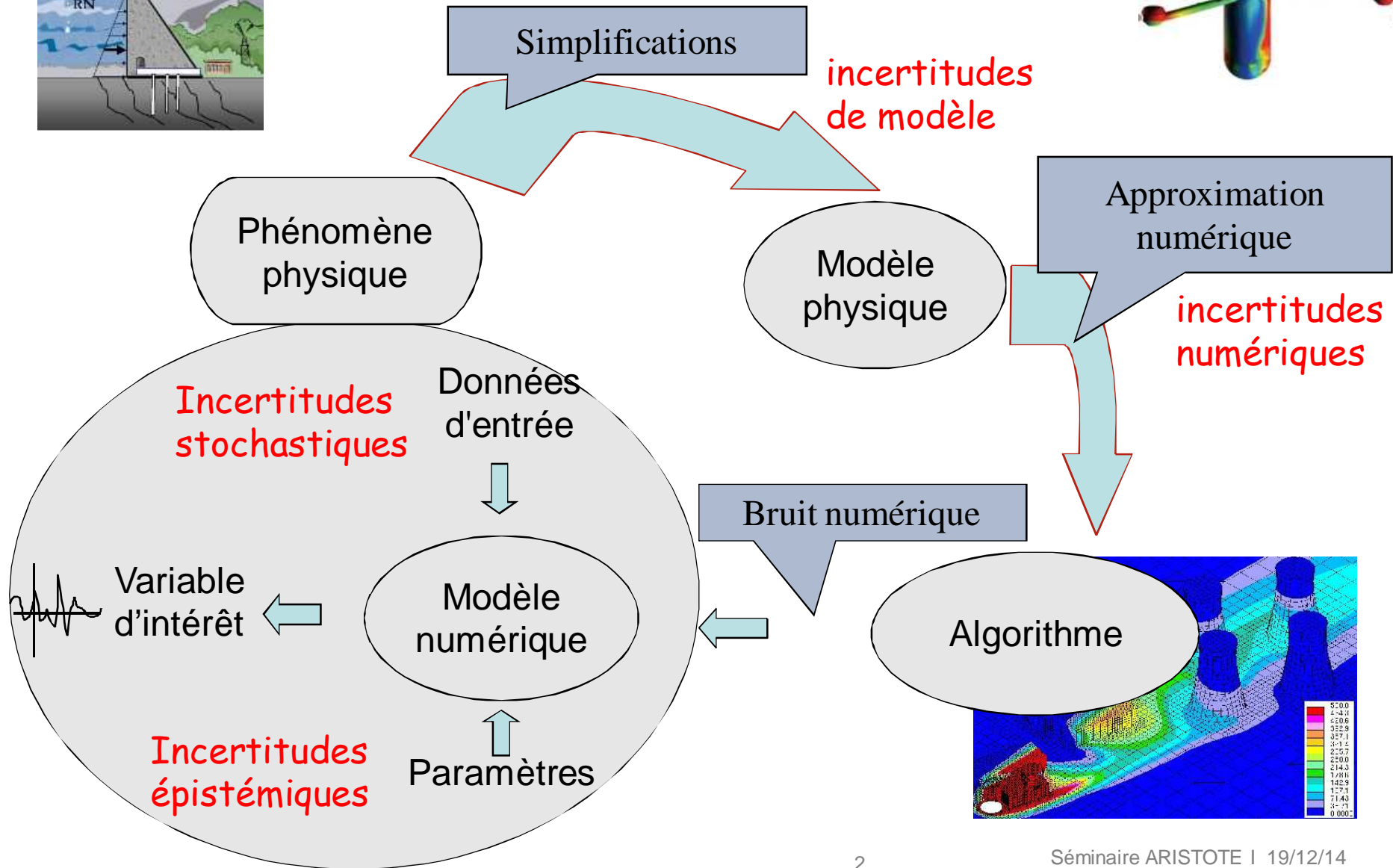
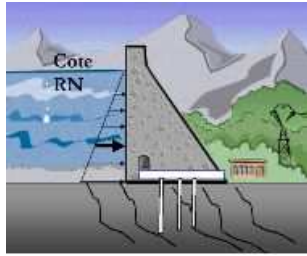
Traitement des incertitudes en simulation numérique pour la maîtrise de procédés industriels

Séminaire ARISTOTE
Ecole Polytechnique
19 décembre 2014

Bertrand looss

EDF R&D, Chatou
Département de Management des Risques Industriels

Crédibilité des résultats d'un modèle physico-numérique



Les incertitudes : objectifs génériques

Étude exploratoire : comprendre un **phénomène**, un processus expérim/indust.

Étude de sûreté : calculer un **risque** de défaillance (marges, événements rares)

Étude de conception : optimiser et maîtriser **les performances** d'un système

Incertitudes

- Variables environnementales
- Paramètres physiques
- Paramètres du processus

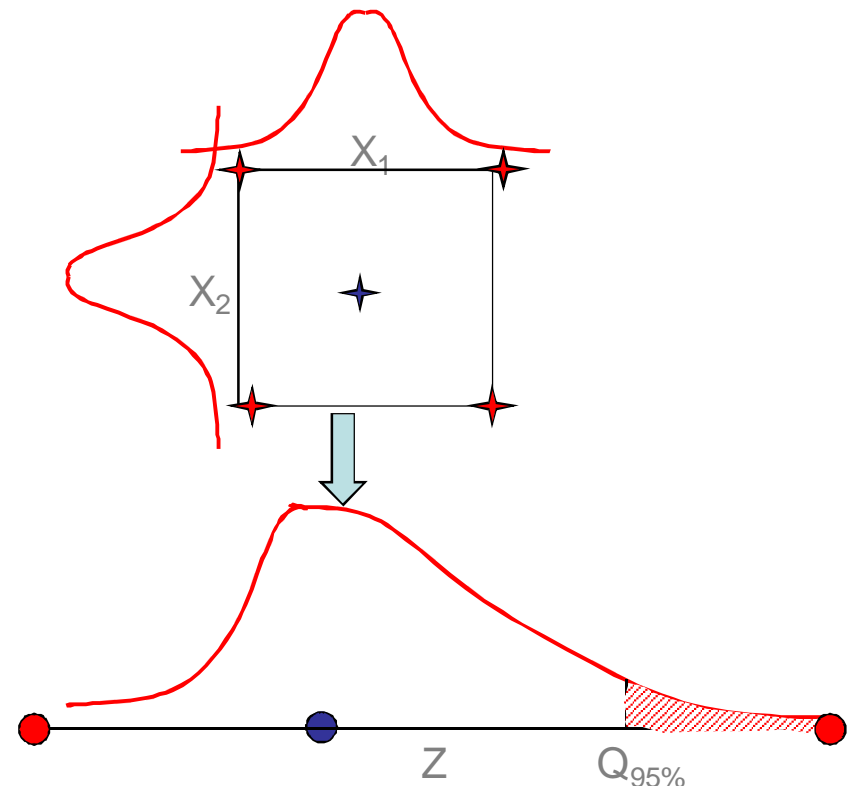
Plan d'expériences

**Code de calcul
ou
Expérience**

Métamodèle éventuel

- Distributions des réponses
- Probabilité de défaillance
- « Entrées » les plus influentes

Cadre déterministe / probabiliste



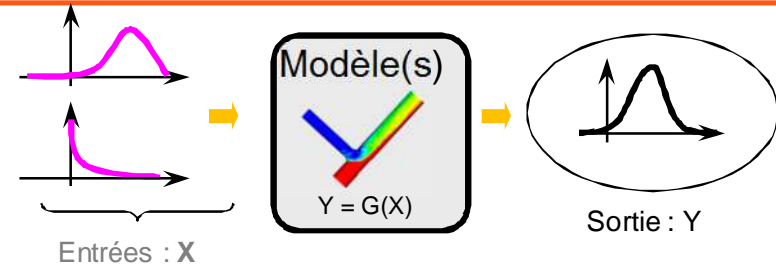
Analyses d'incertitudes et de sensibilité

Représentation boîte noire : $Y = G(\mathbf{X})$

Représentation EDPéiste : $D(\mathbf{X}, u(\mathbf{X})) = f(\mathbf{X}) \Rightarrow Y = G(\mathbf{X}) = \ell(u(\mathbf{X}))$

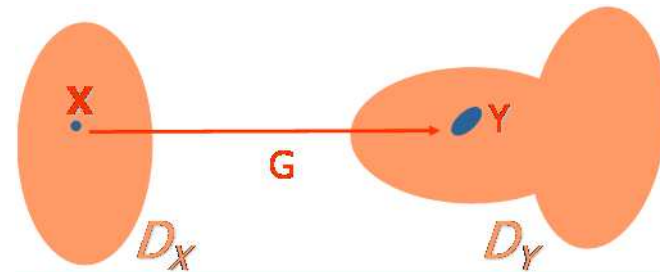
Analyse d'incertitude (problématique générale)

Estimation et étude de la distribution des sorties d'un code induite par les distributions de probabilités des variables d'entrée



Analyse de sensibilité

La sortie Y est-elle plus ou moins variable lorsque l'on fixe une des entrées X_i ?
Quelles variables d'entrée sont les plus influentes ?

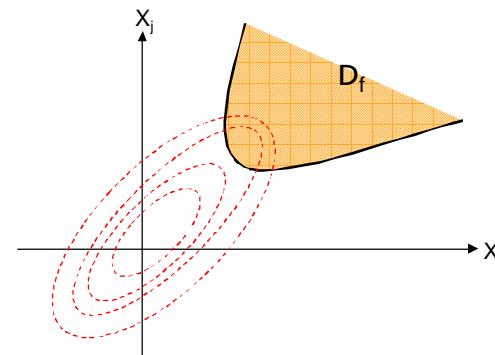


Analyse de fiabilité (des structures)

Quelle est la probabilité que la variable Y dépasse un seuil donné ?

$$p_f = \int \mathbb{1}_{\{G(x) \leq 0\}} p(x) dx = \int_{D_f} p(x) dx$$

Autres problématiques : estimation de quantiles, inversion du domaine de défaillance





Enjeu typique : justifier un critère de sureté

Uncertainty propagation: Main principles

Propagate uncertainties from X to Y , via the deterministic function G

- Conceptually simple problem, but with sometimes a complex implementation
- Choice of method strongly depends on the quantity of interest,

Two kinds of problems :

- Central tendency (mean) or dispersion (var)  Analytical methods (e.g. variance summation) sometimes applicable
- High quantile, probability of failure p_f
→ justification of a safety criterion  Numerical methods

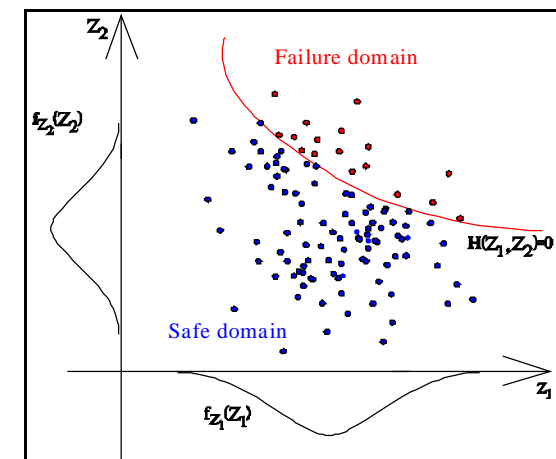
Statistical approach (Monte Carlo sampling)

N samples of X  N outputs Y

Drawback: High computational cost for low p_f

 Specific methods to reduce the cost n

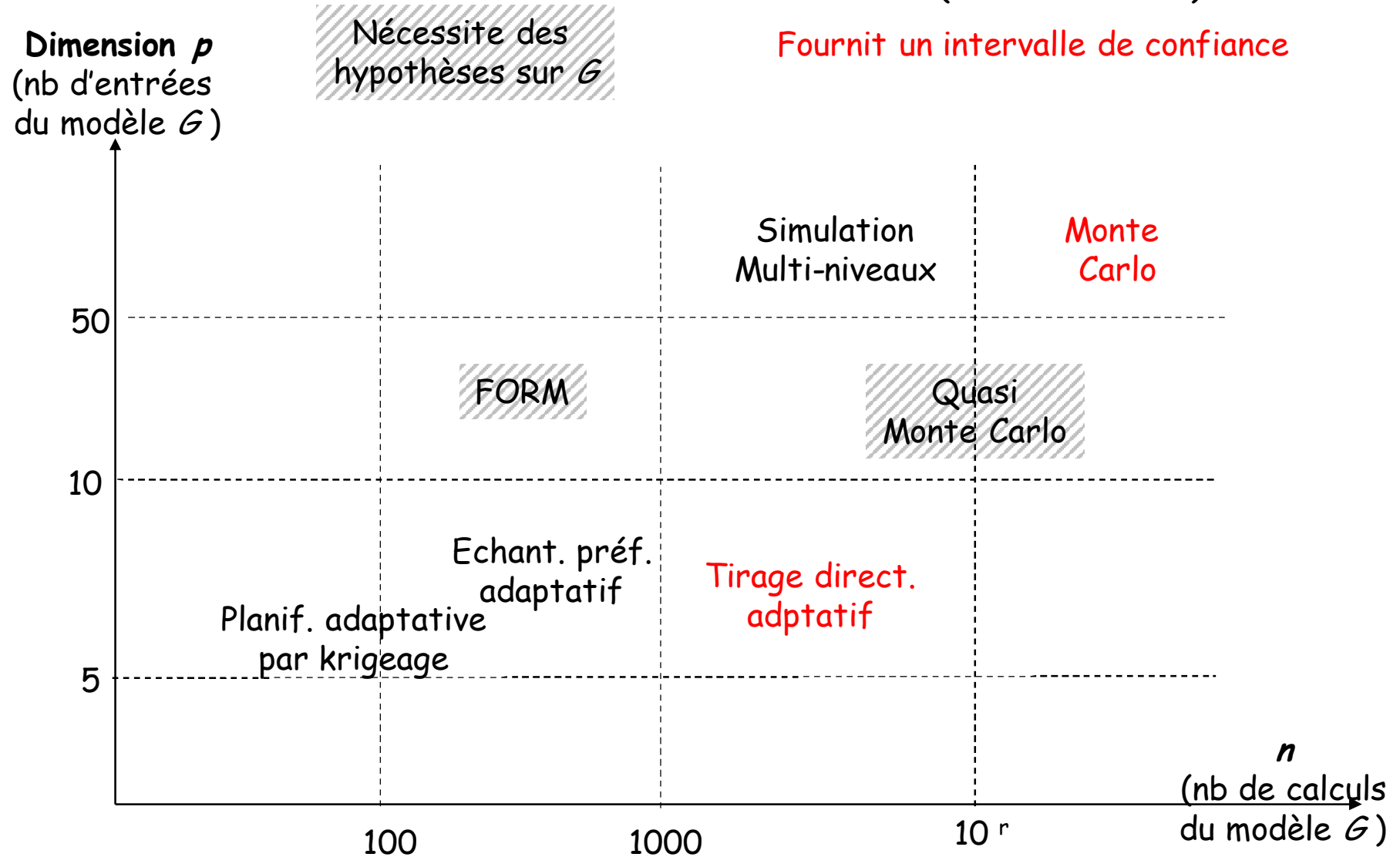
One solution: developing a metamodel



Classification des méthodes d'événements rares

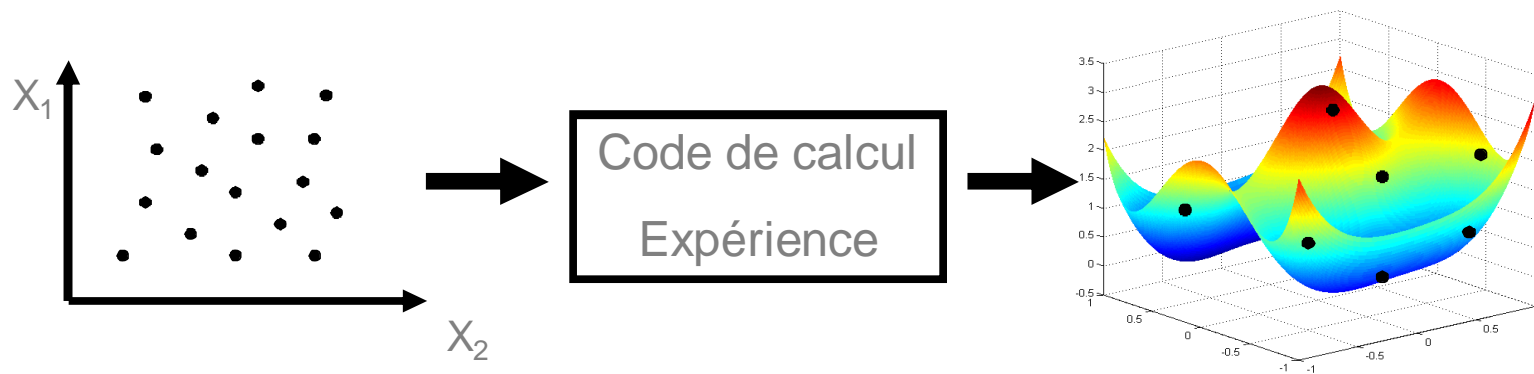
Probabilité (événement rare) $\sim 10^{-r}$

Fournit un intervalle de confiance



Utilisation d'un métamodèle

- Fonction statistique représentative du code de calcul (polynôme, splines, chaos polynomial, réseau de neurones, krigeage, ...) avec un tps de calcul négligeable
- Approximation à partir de n simulations du code ($n \sim 10 p$)
- Important de contrôler les **qualités d'approximation et de prédiction**



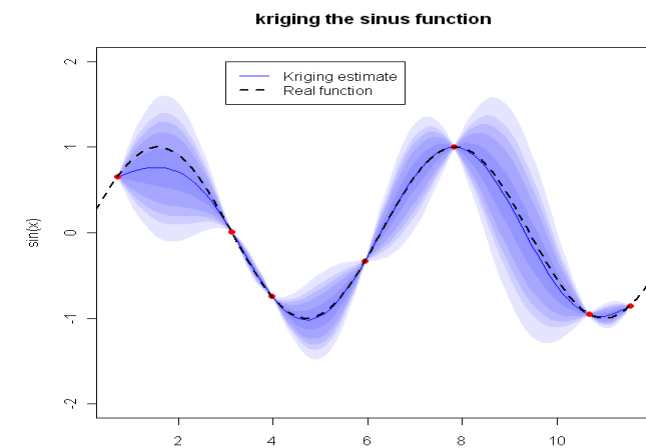
Métamodèle du krigeage

Interpolation par un processus gaussien stationnaire

Avantage : fournit en plus d'une prédiction, une estimation de l'erreur (loi gaussienne en chaque point)

=> Outil algorithmique pour **spécifier des calculs**

pertinents (vis-à-vis de la quantité d'intérêt) : **planification adaptative**



Exemple : simulation thermohydraulique

Benchmark OCDE/NEA : **Calculs de sûreté des réacteurs nucléaires**

[de Crecy et al. 2008]

Scénario :

Accident de Perte de Réfrigérant Primaire due à une Grosse Brèche

Variable d'intérêt (sortie) :

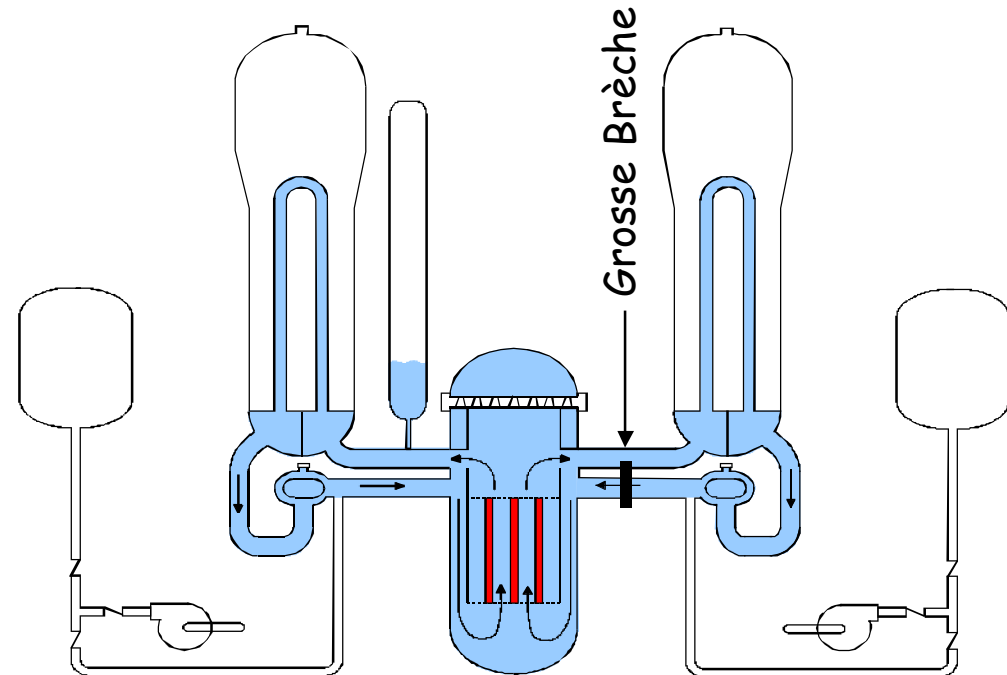
Max (sur le temps) de la température de la gaine

Critère de sûreté :

Non dépassement d'une valeur critique

Variables d'entrée :

Paramètres matériaux, géométriques, conditions initiales, ...



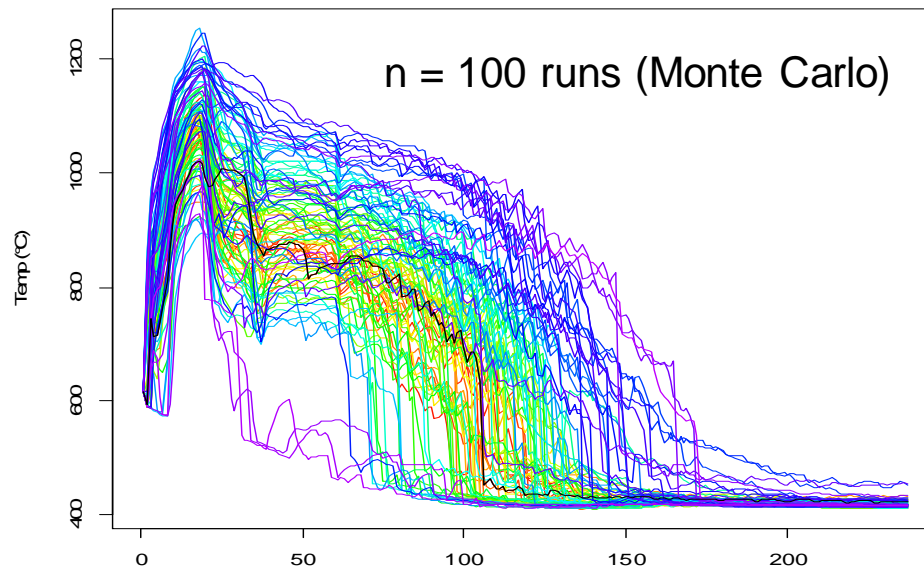
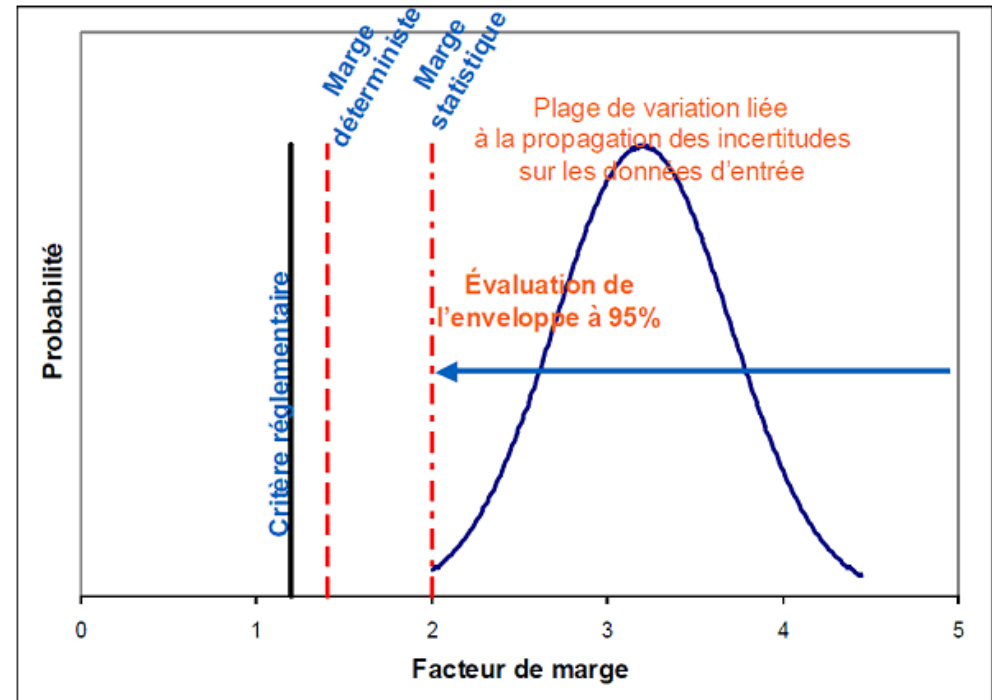
Modèle physique :

Code de calcul « thermohydraulique système » CATHARE (1 run ~ 1h)

Utilisation d'un critère statistique

Calcul support à un dossier de sûreté :
 montrer que le quantile à α % de
 la sortie du code de calcul
 est sup. à un critère réglementaire
 (1477°K) avec un haut niv. de confiance

Problème majeur : coût CPU du code



Exemple de problème :
 Obtention du quantile à 95% de
 $Y = T_{gaine_max}$ à l'aide de $n = 200$ calculs

Statistical estimates of a code quantile – Via a metamodel

Find $\hat{Y}(\alpha, n)$ estimate of y_α / $P(Y \leq y_\alpha) = \alpha$ from a n -sample (Y_1, \dots, Y_n)

$Y = G(\mathbf{X})$ is a real computer code output, $\mathbf{X} = (X_1, \dots, X_d)$ a random vector

- Empirical estimate: $\hat{Y}_{EE}(\alpha, n) = Y_{(\lfloor \alpha n + 1 \rfloor)}$, large asymptotic variance $\sigma_{EE}^2 = \frac{\alpha(1-\alpha)}{p^2(y_\alpha)}$

with the ordered sample $(Y_{(1)}, \dots, Y_{(n)})$ such that $Y_{(1)} \leq \dots \leq Y_{(n)}$

- Wilks estimate: $P(\hat{Y}_{\alpha, n} \geq y_\alpha) \geq \beta$

The variance of this estimator has been also proved to be rather large

- Direct metamodel estimate: empirical estimate of the metamodel (large n)
Bad estimation because the metamodel has a smoothing effect

- Controlled stratification [Cannamela et al. 2009]:

Intensive use of the metamodel $Z = f_r(\mathbf{X})$ to find X samples (fixed size n) in interesting zones (by quantile stratification of the output values), then computation of the “true” code $Y = f(\mathbf{X})$ and applying an adequate estimator

Results

| | Estimation method ($n = 200$) | Q_2 | ρ_I | Estimated quantile | St. Dev. |
|-------------|--|------------|------------|-----------------------|----------|
| Reference | EE with code f ($N=1e3$) | | | 928 | 6 |
| | EE with code f | | | 926 | 12 |
| Dangerous | EE ($N= 10^6$) with Gaussian Process metamodel | 66% | 37% | 912 | ~ 0 |
| Best | CS with linear metamodel (trained on 30 points) | 10% | 54% | 928 | 7 |

Q_2 : metamodel predictivity coefficient (global)

ρ_I : metamodel correlation coefficient around the quantile
(between $1_{Y \leq y_\alpha}$ et $1_{Z \leq z_\alpha}$)

Conclusions :

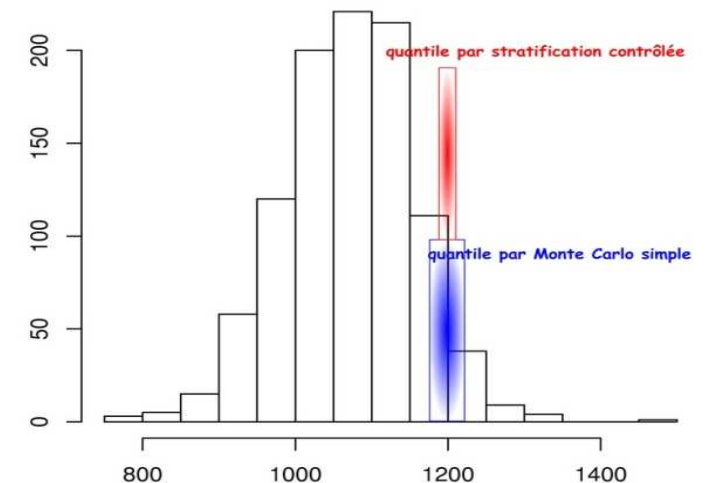
CS is well adapted when using coarse simplified model

ρ_I is the criterion to judge the CS method success

Solution:

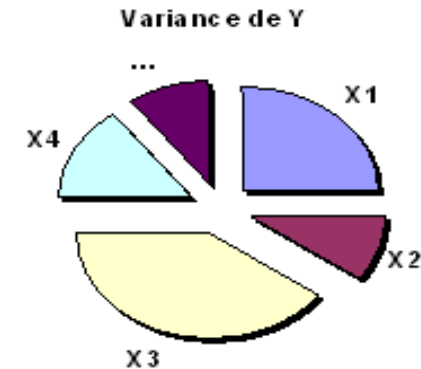
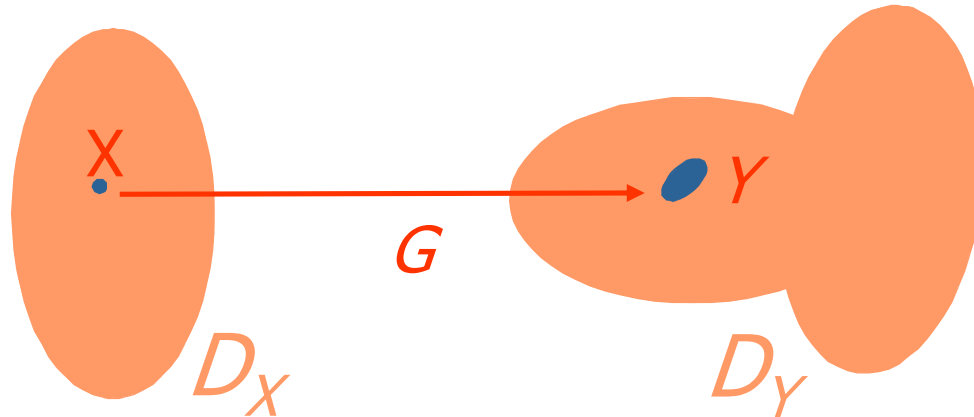
Maximize ρ_I while fitting the metamodel (maximize the fit around y_α)

=> **Sequential fitting of the Gaussian process metamodel** [Picheny 2009, Bect et al. 2012]



Analyse de sensibilité : objectifs/méthodes

- Hiérarchiser les sources d'incertitudes afin de réduire la variabilité de la sortie



- Simplification du modèle

- Déterminer les variables aléatoires non-influents, qui peuvent être considérées déterministes sans influence la variabilité de la sortie
- Construire un modèle simplifié, un métamodèle

Méthodes d'analyse de sensibilité :

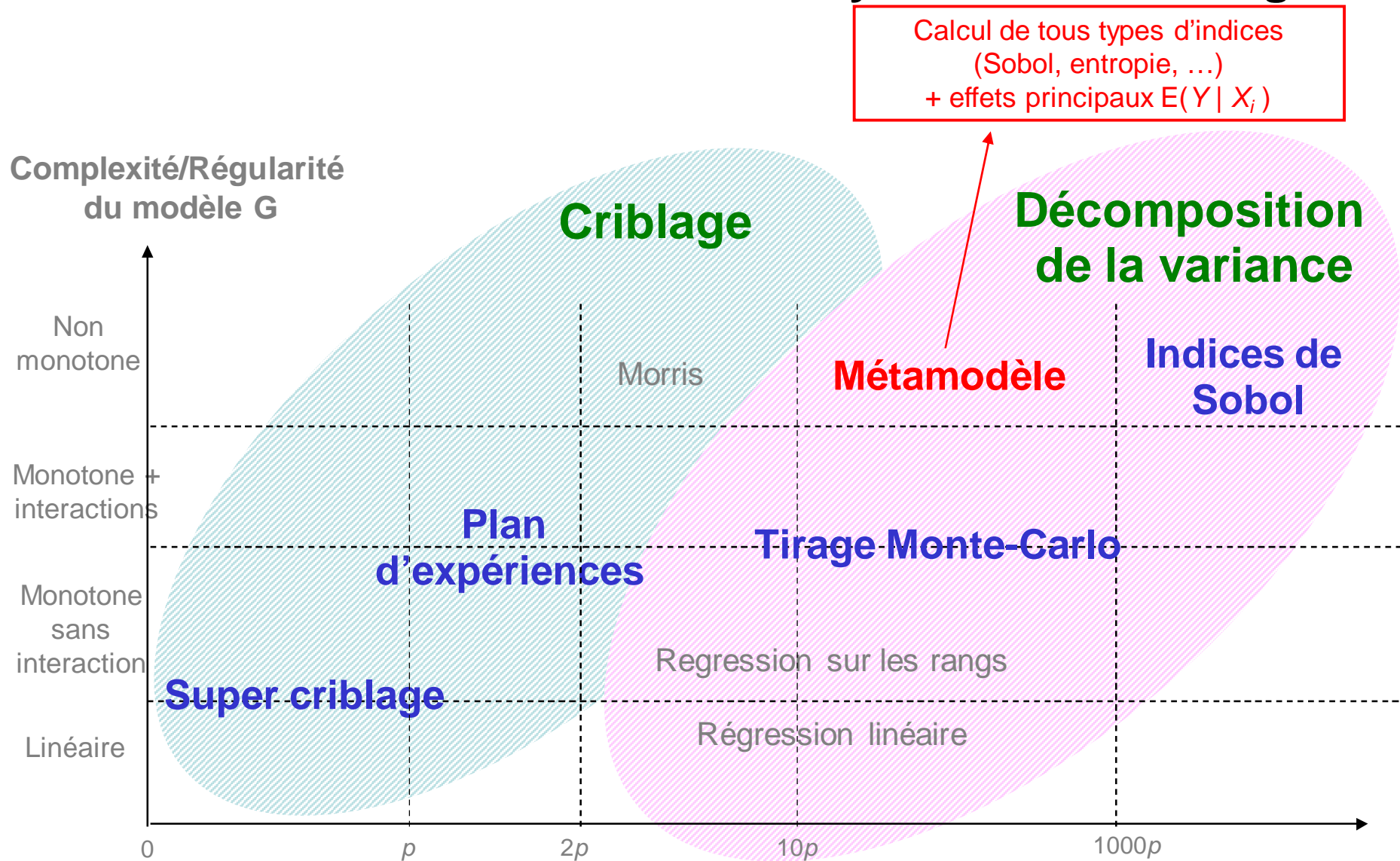
- Locale $\frac{\partial G}{\partial X_i} \Big|_{\mathbf{x}_0}$

- Globale : indices de Sobol

$$S_i = \frac{\text{Var}[E(Y|X_i)]}{\text{Var}(Y)}$$

$$S_{Ti} = 1 - \frac{\text{Var}[E(Y|X_{\sim i})]}{\text{Var}(Y)}$$

Classification des méthodes en analyse de sensibilité globale



(p = nombres de variables en entrée du modèle)

[looss 2011]

- Estimateur empirique par échantillonnage direct de

$$\text{Var}[E(Y|X_i)] = \text{Cov}[G(X_i, X_{\sim i}), G(X_i, X'_{\sim i})]$$

$$\text{Var}[E(Y|X_{\sim i})] = \text{Cov}[G(X_i, X_{\sim i}), G(X'_i, X_{\sim i})]$$

avec $(X_i^{(j)})_{i=1,\dots,p;j=1,\dots,n}$ et $(X'_i{}^{(j)})_{i=1,\dots,p;j=1,\dots,n}$ i.i.d.

$$N = \text{coût (Indices 1^{er} ordre + indices totaux)} = n (p + 2)$$

- Echantillon Monte Carlo : $n \sim 10^4$
- Echantillon Quasi-Monte Carlo, méthode FAST : $n \sim 10^3$
- Les indices d'ordre faible peuvent être estimés à moindre coût ($N \sim 100$) à l'aide de méthodes de lissage (paramétrique/non param) [Da Veiga et al. 2009; Storlie & Helton 2008]
- La difficulté principale réside alors en l'estimation des indices totaux**
 - => approximations par utilisation de métamodèles [Sudret 2008; Le Gratiet et al. 2014]
 - => **Bornes sup**

Derivative-based global sensitivity measures (DGSM)

[Sobol & Greshman 1995; Kucherenko et al. 2009]

$$v_i = \int_{\mathcal{X}} \left(\frac{\partial G(\mathbf{X})}{\partial X_i} \right)^2 d\mu(\mathbf{X}) \quad \text{où } \mathbf{X} \in \mathcal{X} \subset \mathfrak{R}^p \text{ et } \frac{\partial G}{\partial X_i} \in L_2$$

avec $\mu(\mathbf{X})$ la densité jointe de \mathbf{X}

Synthèses des tests numériques :

1. DGSM est moins coûteux à évaluer que l'indice de Sobol total
Cas faciles (faible dimension effective) : $N \sim 100 p$; cas + difficiles : $N \sim 1000 p$
 2. DGSMs permettent de distinguer les X_i non influents et influents (**criblage**),
mais peuvent se tromper dans la hiérarchisation des influents
- ⇒ Pour les codes avec adjoint, on peut envisager de réaliser une analyse de sensibilité globale (via les DGSM) avec $p \gg 100$

Mais comment interpréter les DGSM ?

[Lamboni et al. 2012; Roustant et al. 2014]

- Si les X_i sont indépendants, $G \in L^2(\mathfrak{X})$, $dG/dX_i \in L^2(\mathfrak{X})$, et si la distribution de X_i est continue, on a

$$S_{Ti} \leq \frac{C(\mu_i)}{\text{Var}(Y)} \nu_i \quad \text{avec } C(\mu_i) = 4 \left[\sup_{x_i \in \mathbb{R}} \frac{\min(F_i(x), 1 - F_i(x))}{\mu_i(x)} \right]^2$$

où $C(\cdot)$ est appelé constante de Poincaré (ou Cheeger) [Bobkov 1999; Fougères 2005]

=> Liens entre approches stochastique et déterministe

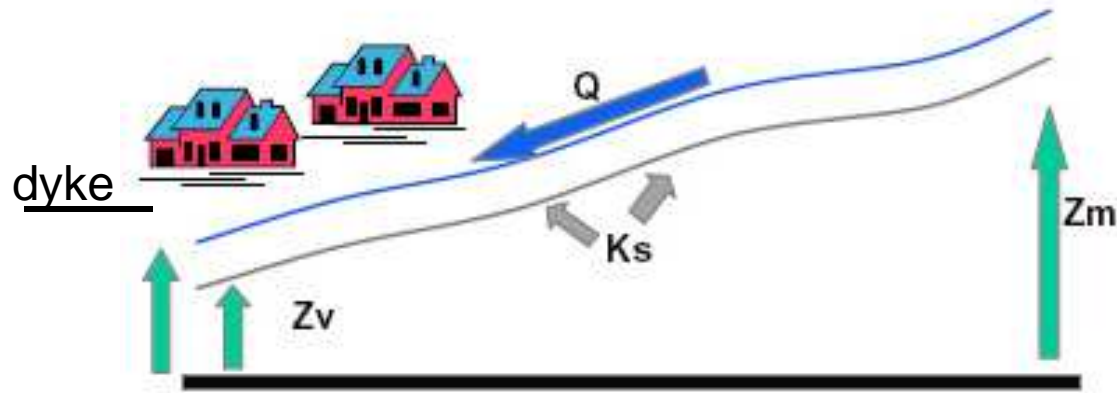
- Preuve : **inégalité de Poincaré** $\|g\|_{L^2} \leq C \|\nabla g\|_{L^2}$ appliquée sur

$$g(x_j, \mathbf{x}_{\sim j}) = \sum_{u \ni j} f_u(\mathbf{x}_u) \quad (\text{décomp. fctelle unique : } G(\mathbf{x}) = f_0 + g(x_j, \mathbf{x}_{\sim j}) + h(\mathbf{x}_{\sim j}))$$

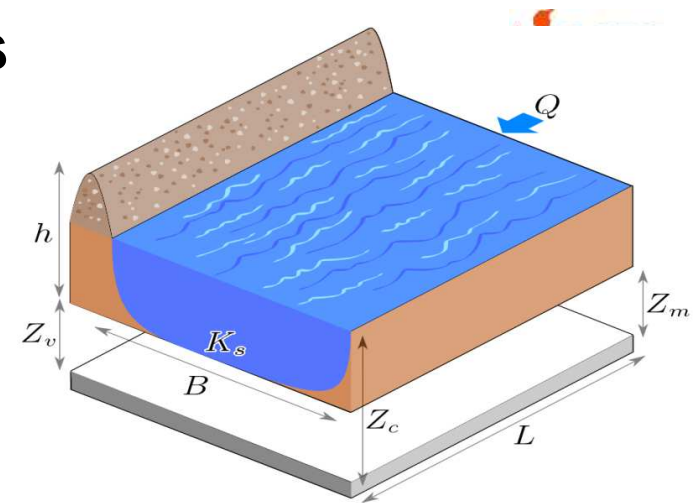
$$S_{Ti} = \int g^2(x_i, \mathbf{x}_{\sim i}) d\mu(\mathbf{x}) / \text{Var}(Y) \quad \text{et } \nu_i = \int \left(\frac{\partial g(\mathbf{x})}{\partial x_i} \right)^2$$

- *Résultat similaire à l'ordre deux (dérivées croisées et indice total d'interaction)*

Application « jouet » : modèle de crues



Modèle physique simplifié (hydraulique)



Variables d'entrée

Q = débit ~ G[1013,558]

Ks = coef. frottement ~ N(30,8)

Zv = hauteur du lit aval T[49,51]

Zm = hauteur lit amont T[54,56]

Hd = hauteur dique ~ U[7,9]

Cb = côte de la berge ~ T[55,56]

L=longueur tronçon ~ T[4990,5010]

B = largeur rivière ~ T[295,305]

S = surverse

$$S = Z_v + H - H_d - C_b \text{ avec } H = \left(\frac{Q}{BK_s \sqrt{\frac{Z_m - Z_v}{L}}} \right)^{0.6}$$

Résultats sur le modèle de crues



| Input | S_j | S_{T_j} | ν_j | Υ_j | → Borne de S_{T_j} |
|-------|-------|-----------|-----------|--------------|----------------------|
| Q | 0.343 | 0.353 | 1.296e-06 | 2.807 | |
| K_s | 0.130 | 0.139 | 3.286e-03 | 0.198 | |
| Z_v | 0.185 | 0.186 | 1.123e+00 | 0.561 | |
| Z_m | 0.003 | 0.003 | 2.279e-02 | 0.011 | |
| H_d | 0.276 | 0.276 | 8.389e-01 | 0.340 | |
| C_b | 0.036 | 0.036 | 8.389e-01 | 0.105 | |
| L | 0.000 | 0.000 | 2.147e-08 | 0.000 | |
| B | 0.000 | 0.000 | 2.386e-05 | 0.000 | |

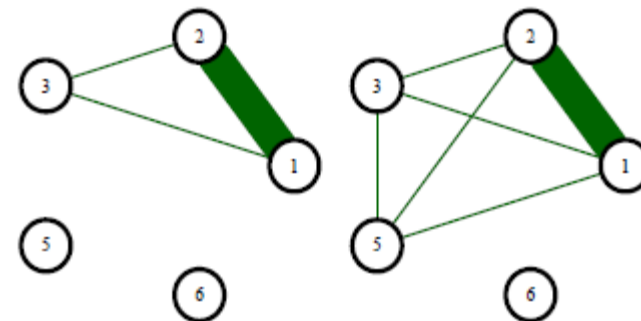
Sobol (MC) : $n = 1^e5$, 20 répliques $\Rightarrow N = 2^e7$ - DGSM (QMC) : $n = 1^e4 \Rightarrow N = 9^e4$

Interactions

$X_1=Q$, $X_2=Ks$, $X_3=Zv$, $X_5=Hd$, $X_6=Cb$

| | $\hat{S}_{i,j}^{super}$ | $\hat{U}_{i,j}$ | $C(\mu_i)C(\mu_j)$ |
|-------------|-------------------------|-----------------|--------------------|
| $X_1 : X_2$ | 0.008235 | 1.837809 | 819887392.342 |
| $X_1 : X_3$ | 0.000178 | 0.011691 | 2165585.265 |
| $X_1 : X_5$ | 0.000000 | 0.001947 | 877678.649 |
| $X_1 : X_6$ | 0.000000 | 0.000000 | 541396.316 |
| $X_2 : X_3$ | 0.000070 | 0.003758 | 378.599 |
| $X_2 : X_5$ | 0.000000 | 0.000000 | 153.440 |
| $X_2 : X_6$ | 0.000000 | 0.000000 | 94.650 |
| $X_3 : X_5$ | 0.000000 | 0.000000 | 0.405 |
| $X_3 : X_6$ | 0.000000 | 0.000000 | 0.250 |
| $X_5 : X_6$ | 0.000000 | 0.000000 | 0.101 |

FANOVA graph



Superset importance

Crossed DGSM

Conclusion : grandes problématiques identifiées



Estimation proba. de défaillance, quantiles rares, domaines de défaillance, mesures de risque

Étape C : Propagation des sources d'incertitude

Modélisation de la dépendance et des extrêmes

Incertitudes et décision, cadre bayésien

Étape B : Quantification des sources d'incertitudes

Modélisation probabiliste (loi jointe)

Méthodes directes, statistique, expertise

Étape A : Spécification du problème

Données d'entrée

Variables : x
Paramètres : θ
Fixées : d

Modèle numérique

$G(x, \theta, d)$

Variables d'intérêt

$Z = G(x, \theta, d)$
 $Z^R = G^R(x, d)$
 $= G(x, \theta, d) + b(x, d)$

Quantité d'intérêt

Ex : variance, probabilité

Quantifier l'incertitude de modèle

Étape B' : Calibration / Validation

Variables observées

$Z_{obs}(x, d)$
 $= Z^R + \varepsilon(x)$

Inversion probabiliste

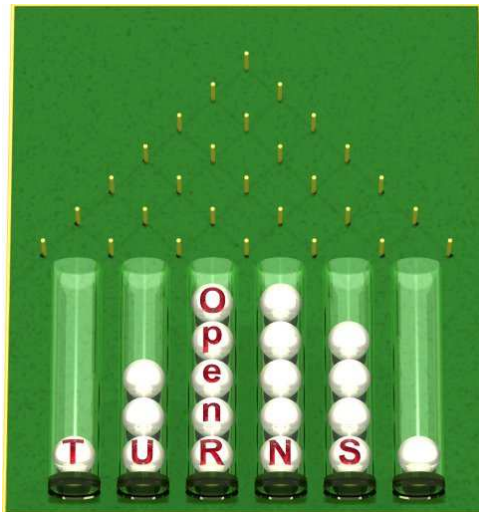
Étape C' : Analyse de sensibilité

Analyse de sensibilité en grande dimension et sur événements rares

Revision des modèles

Critère de décision Ex: Probabilité < 10^{-b}

Fin – Pub - Merci



Logiciel OpenTURNS : Treatments of
Uncertainties, Risk'n Statistics

Open source

<http://www.openturns.org/>

Groupement De Recherche MASCOT NUM

Méthodes d'Analyse
Stochastique
pour les COdes et
Traitements NUMériques

Événements

Publications, softs,
Offres stages/theses/...

www.gdr-mascotnum.fr