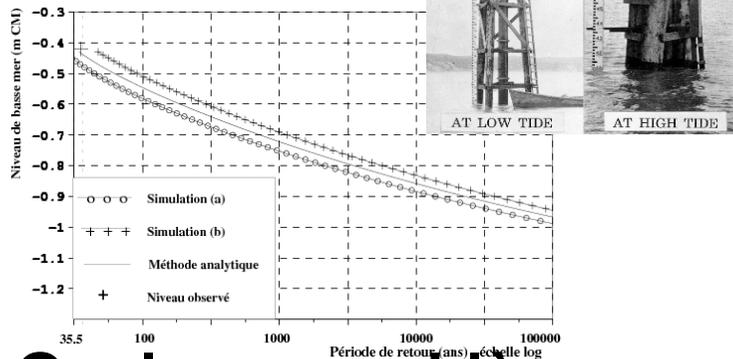




# Incertitudes, risques et environnement



## Quelques problèmes mathématiques pour un industriel

SMAI – 7ème Rencontres Math-Industrie « Mathématiques et Environnement »

E. de Rocquigny, EDF R&D

Ing. Senior « Statistiques, simulation dans les risques et l'environnement »

Coordinateur du Programme « Incertitudes » / réseau européen ESReDA



# Incertitudes, risques et environnement

- Des exemples « point de départ »
- Cadre de modélisation général et panorama des problématiques
- Deux problématiques particulières
  - Estimation de risque via propagation d'incertitudes
  - Identification de variabilité par approche inverse probabiliste / assimilation

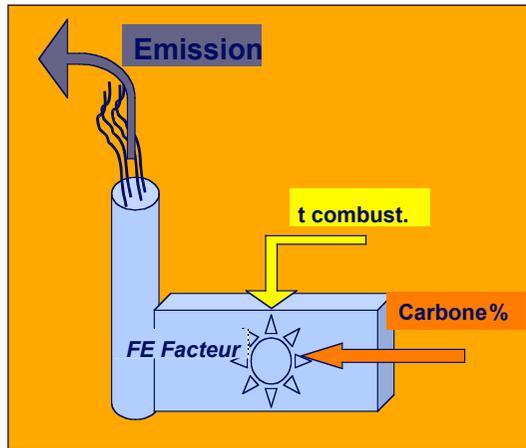
# Exemples

Enjeu principal	Exemple	Domaine
Justification maîtrise d'impacts	<b>Maîtrise des émissions industrielles (e.g. CO2)</b>	Métronologie
Justification risques liés à l'environnement	<b>Niveaux marins extrêmes et calage des centrales</b>	Stat. Extrêmes pour Dimensionnement d'une protection
Optimisation de l'exploitation / contexte environnemental	<b>Productible éolien</b>	Optimisation technico-économique sous incertitude

Thermique	Nucléaire	Renouvelable / hydraulique
-----------	-----------	----------------------------

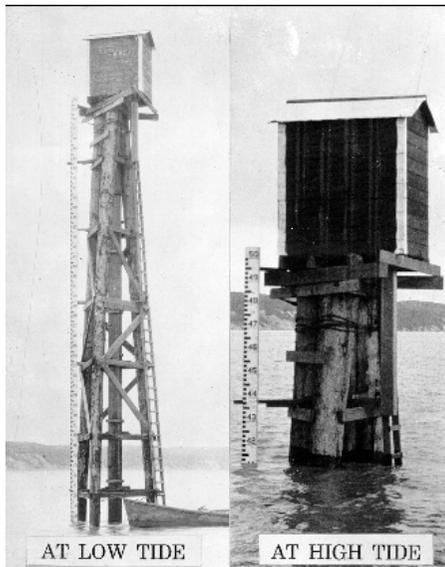
SMAI – 7ème Rencontres Math-Industrie « Mathématiques et Environnement », 2007

# Thermique classique: Émissions de CO<sub>2</sub>



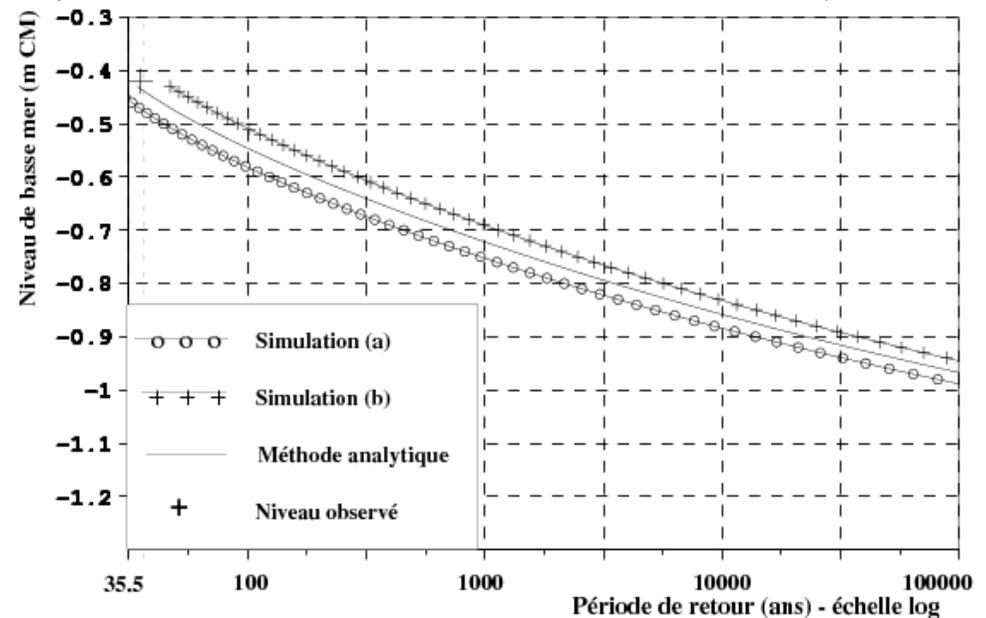
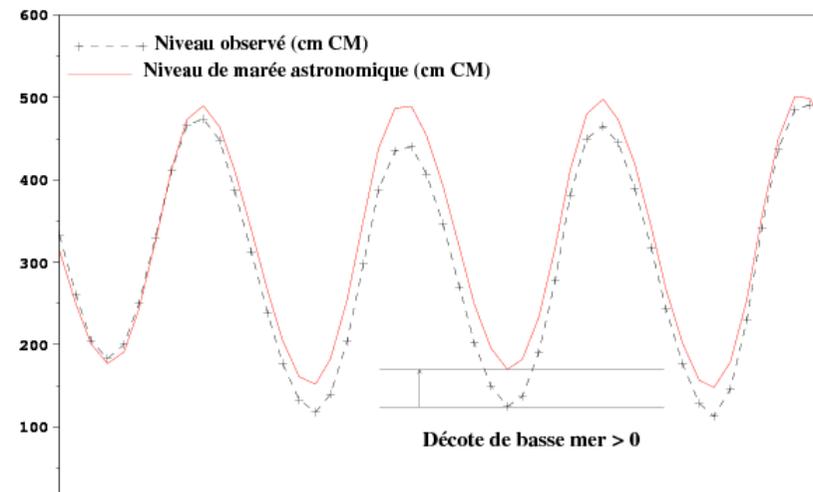
<p><u>Finalité</u></p>	<p>Déclarer rejets polluants d'un site industriel (ex : t<sub>CO2</sub> annuelles) avec x% de précision</p> <p>&gt;&gt; <i>justification</i> réglementaire + <i>optimisation</i> chaîne/marché d'émissions</p>
<p><u>Critère</u></p>	<p>%inc<sub>t CO2</sub> &lt; x %</p> $\%inc_z = \frac{1}{2.E(Z)} \left( z^{\frac{1+\alpha}{2}} - z^{\frac{1-\alpha}{2}} \right) \approx k \cdot \frac{\sqrt{\text{var}(Z)}}{E(Z)}$
<p><u>Sources/mo dèle</u></p>	<p>Erreurs capteurs – variabilité naturelle charbon – méconnaissance du process</p> <p><i>Modèle phys. = formule analytique</i></p>
<p><u>Propagation</u></p>	$Inc_Y ? \sqrt{\sum_{i=1}^N \left( \frac{\partial Y}{\partial X_i} \right)^2 Inc_{X_i}^2} \leq$ <p style="text-align: center;"> <span style="background-color: green; color: black; padding: 2px;">Sensibilité de la variable d'intérêt à la source n°i</span> <span style="background-color: green; color: black; padding: 2px; margin-left: 100px;">Source d'incertitude n°i</span> </p>

# Niveaux marins extrêmes et centrales

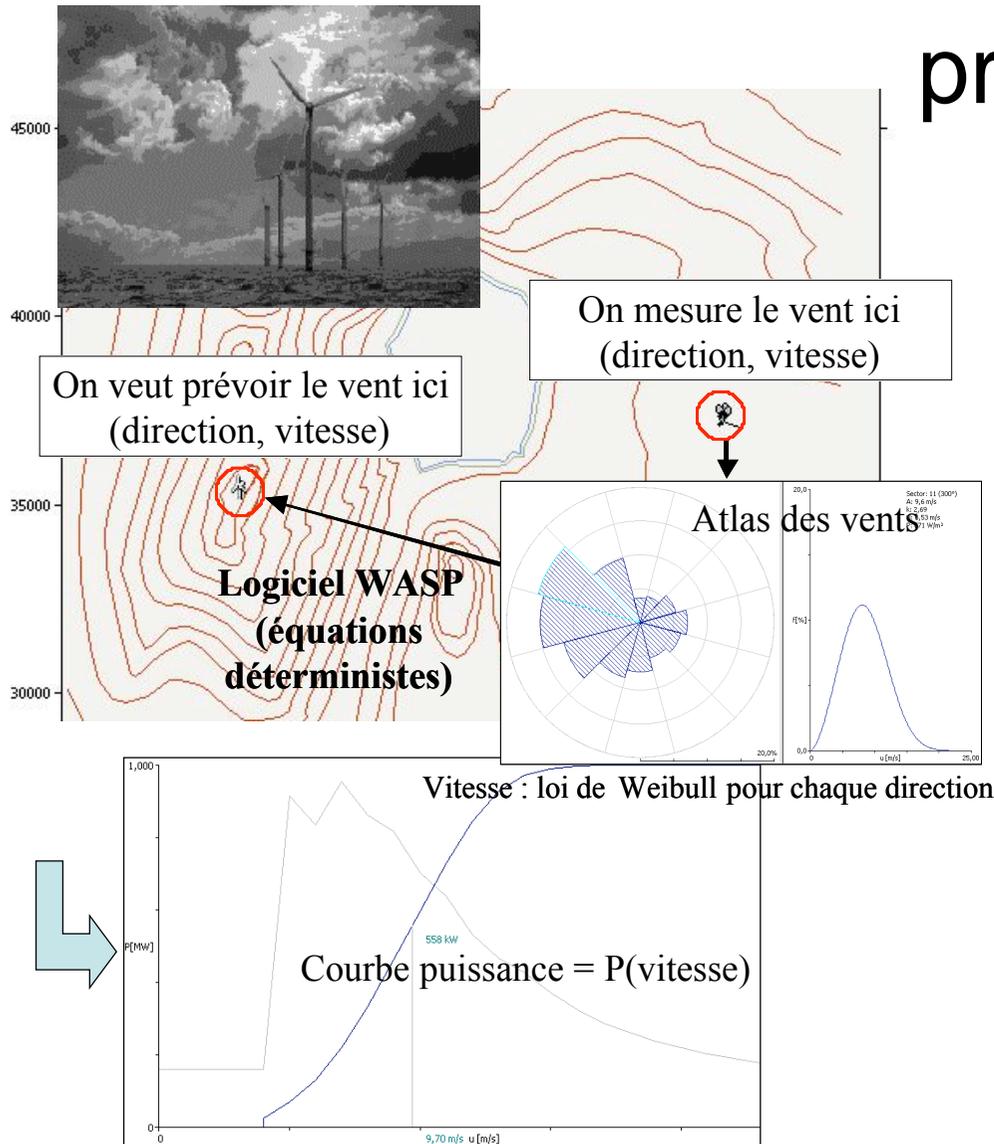


$$Z_t = x(t) - Y_t$$

$Z_t$  = niveau de mer complet  
 $x(t)$  = marée astro. déterministe  
 $Y_t$  = décote de basse mer (aléatoire)  
 ⇒ **Estimer les niveaux de retour rares de la variable Z superposant un signal déterministe complexe et une variable aléatoire.**  
**(stat. extrêmes + simulation)**



# Variabilité du vent et production éolienne



<u>Finalité</u>	<b>Comprendre</b> le risque de mésestimation du productible – <b>hiérarchiser</b> les incertitudes – <b>optimiser</b> le lieu des éoliennes
<u>Critère</u>	%inc sur productible
<u>Sources /modèle</u>	Erreurs d'estimation du vent (erreur spatiale, échantill. Temporel ...); Inc. de modèle Modèle intrinsèquement statistique du vent / du productible
<u>Quantification / propagation</u>	<b>Estimation statistique (moments, max. vraisemblance ...)</b> <b>Propagation par cumul quadratique simple</b>

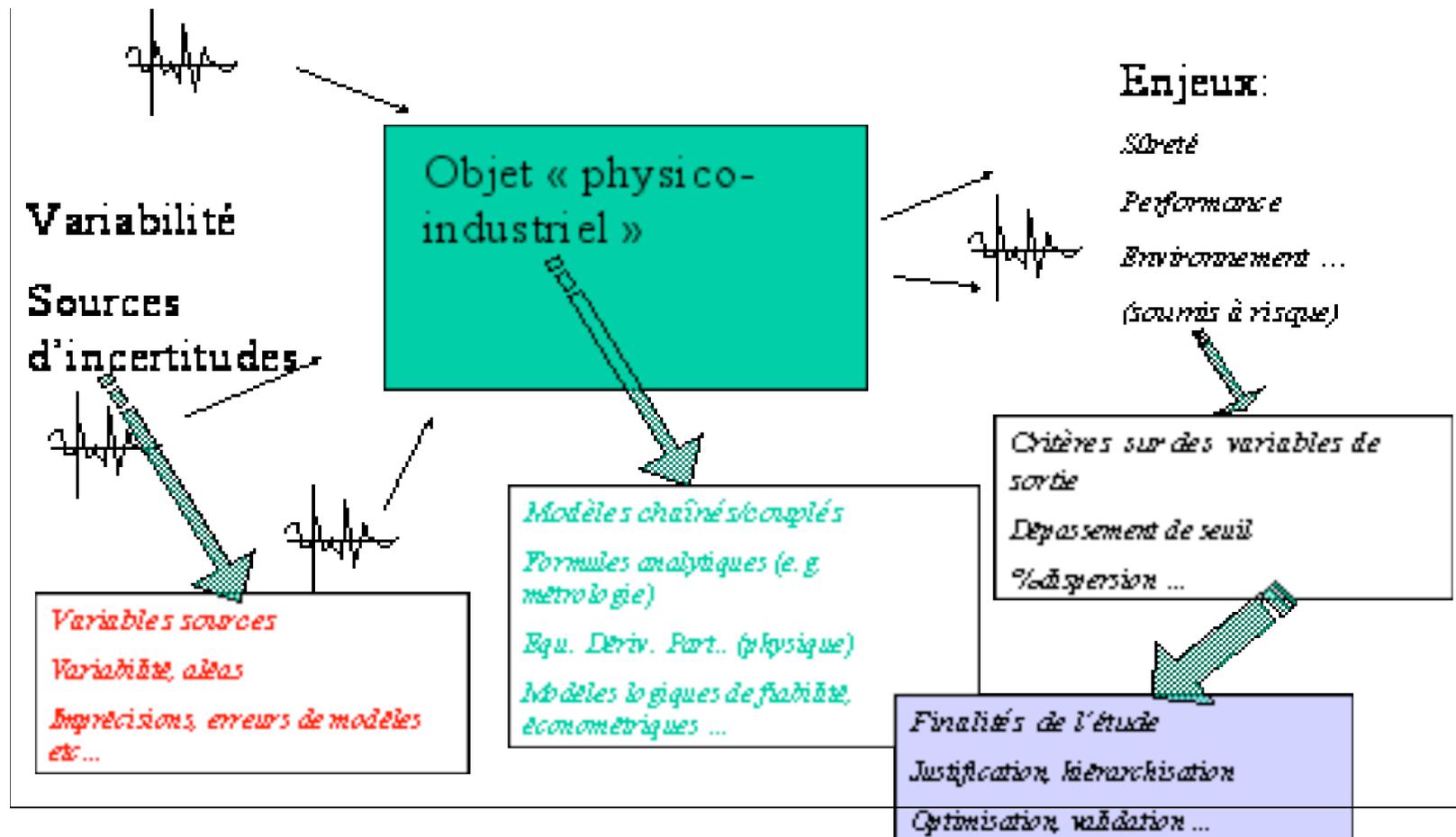
# Cadre commun de modélisation

SMAI – 7ème Rencontres Math-Industrie « Mathématiques et Environnement », 2007

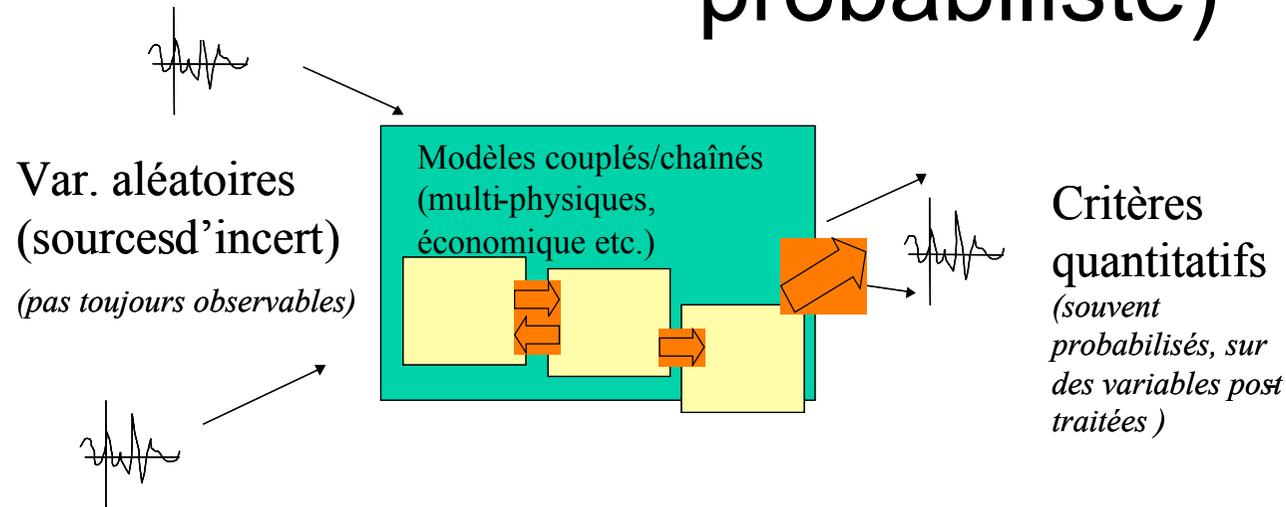
© EDF, E. de Rocquigny



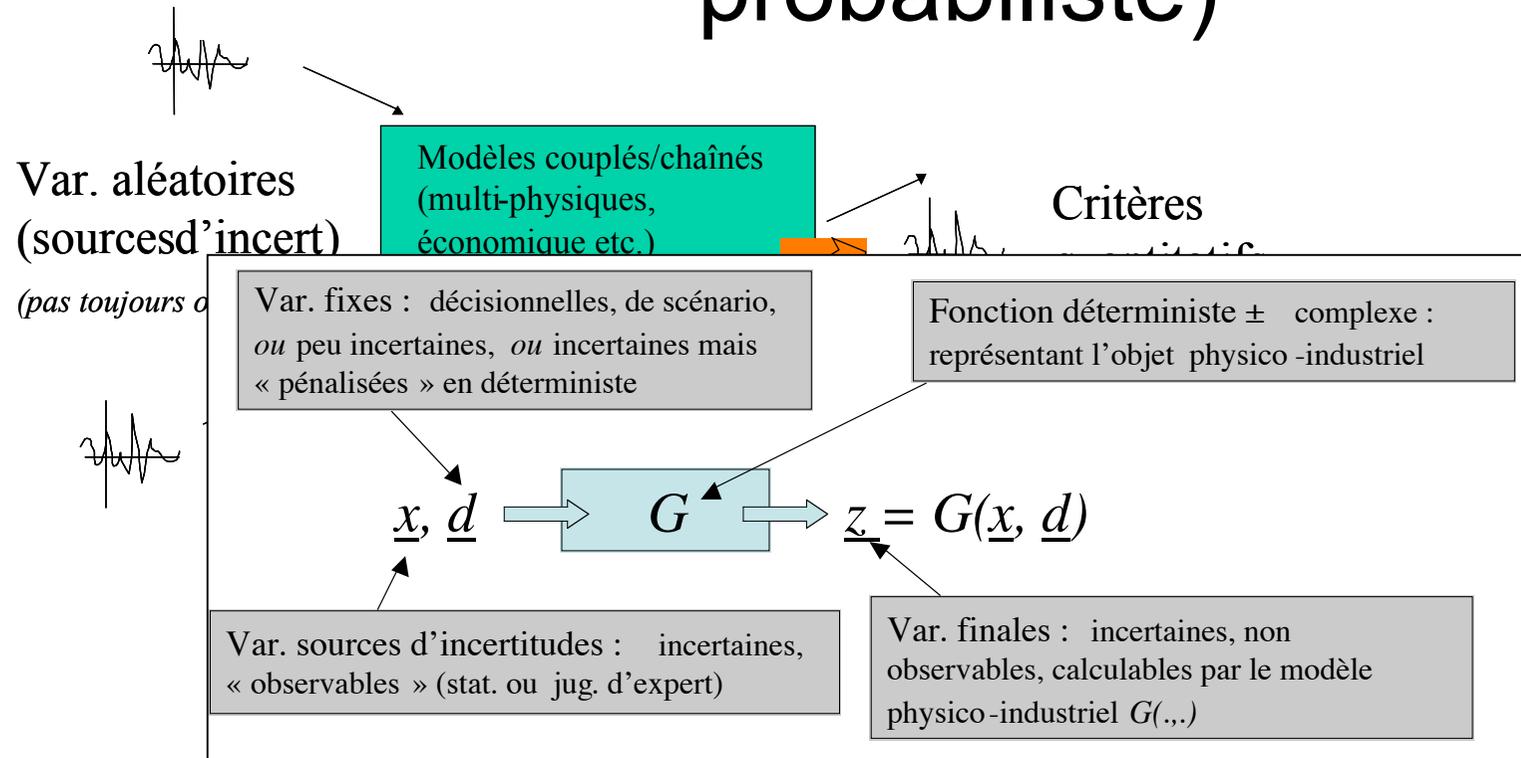
# Vers un cadre méthodologique commun ...



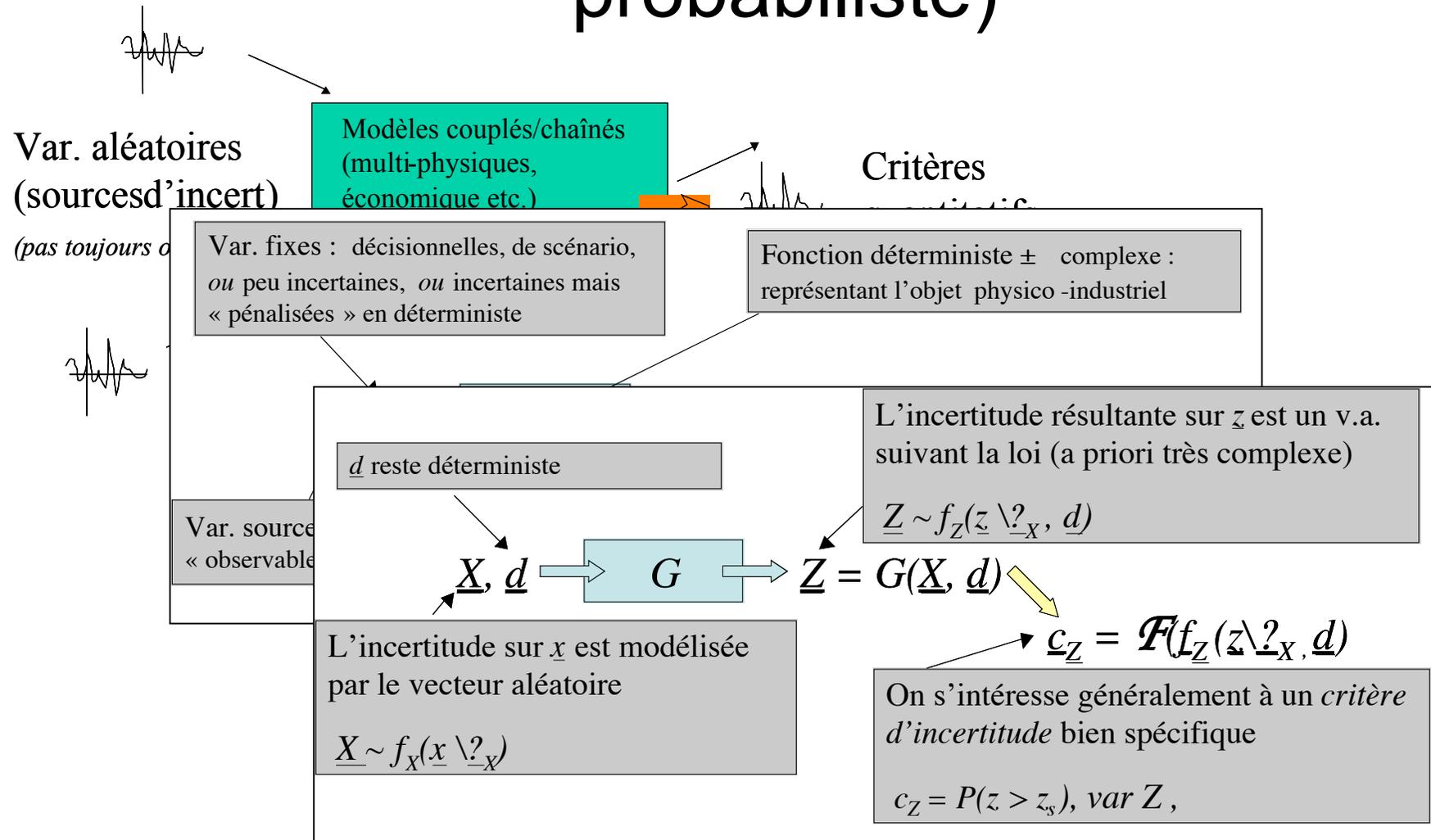
# Cadre de modélisation (physico-probabiliste)



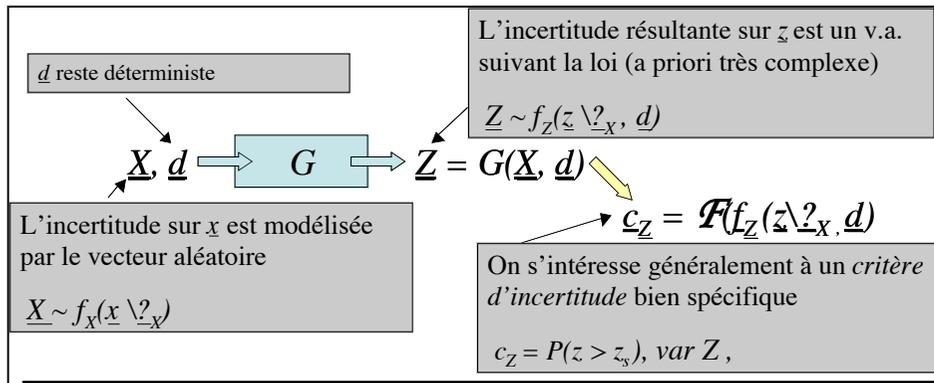
# Cadre de modélisation (physico-probabiliste)



# Cadre de modélisation (physico-probabiliste)



# Les grandes étapes

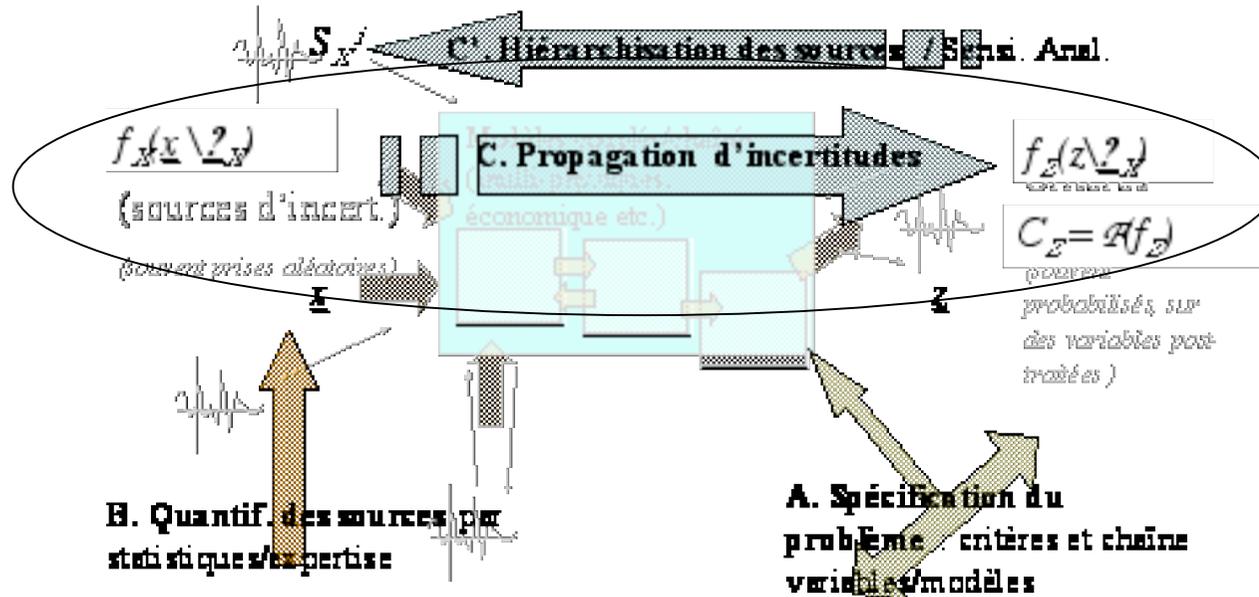


Etape <b>A</b>	Spécifier le problème	Choisir $\underline{x}, \underline{d}, G(.,.)$ et un critère ( $C_Z = P(Z > z_s)$ ou $\text{var } Z \dots$ )	Sciences de la décision, analyse du risque, cadre classique ou bayésien ...
Etape <b>B</b>	Quantifier les sources <i>Uncertainty modelling</i>	Modéliser $\underline{X}$ en estimant $\underline{\theta}_X$ d'après $(X_j)_j$ ou expertise	Echantillonnage paramétrique classique ou bayésien <i>Mais aussi</i> : non-paramétrique, copules (loi jointe) ...
Etape <b>C</b>	Propager <i>U. Propagation / U. Analysis</i>	Estimer $f_Z$ (ou plutôt $C_Z$ )	Taylor, Monte-Carlo et accélérées, Form-Sorm, Plans d'Exp./Surfaces de Réponse ...
Etape <b>C'</b>	Hierarchiser l'importance des sources <i>Sensitivity analysis</i>	Estimer des $S_X$ (selon $C_Z$ ) $S_X = \frac{\text{Corr}(Z, X^i)? \cdot \frac{\text{Var}(E[Z \setminus X^i])}{\text{Var}(Z)}}{\sum_{i=1}^N \text{Corr}(Z, X^i)?}$	Idem + RCC/ PRCC, indices de Sobol estimés par MCS, quasi-MC ... , FAST,

# Deux problématiques

- Sujet n° 1 – Estimation de risque via propagation d'incertitudes
- Sujet n° 2 – Identification de variabilité par approche inverse probabiliste / assimilation

# Estimation de risque via propagation d'incertitudes



- Si  $C_Z = \text{var}(Z)$ , « facile » : Monte-Carlo, Approx. Différentielle, surfaces de réponses « standard », chaos polynômial ...
- Si  $C_Z = P[Z > z_s]$  « rare » et  $Z = G(X)$  « lourd » >> difficile (d'être robuste)

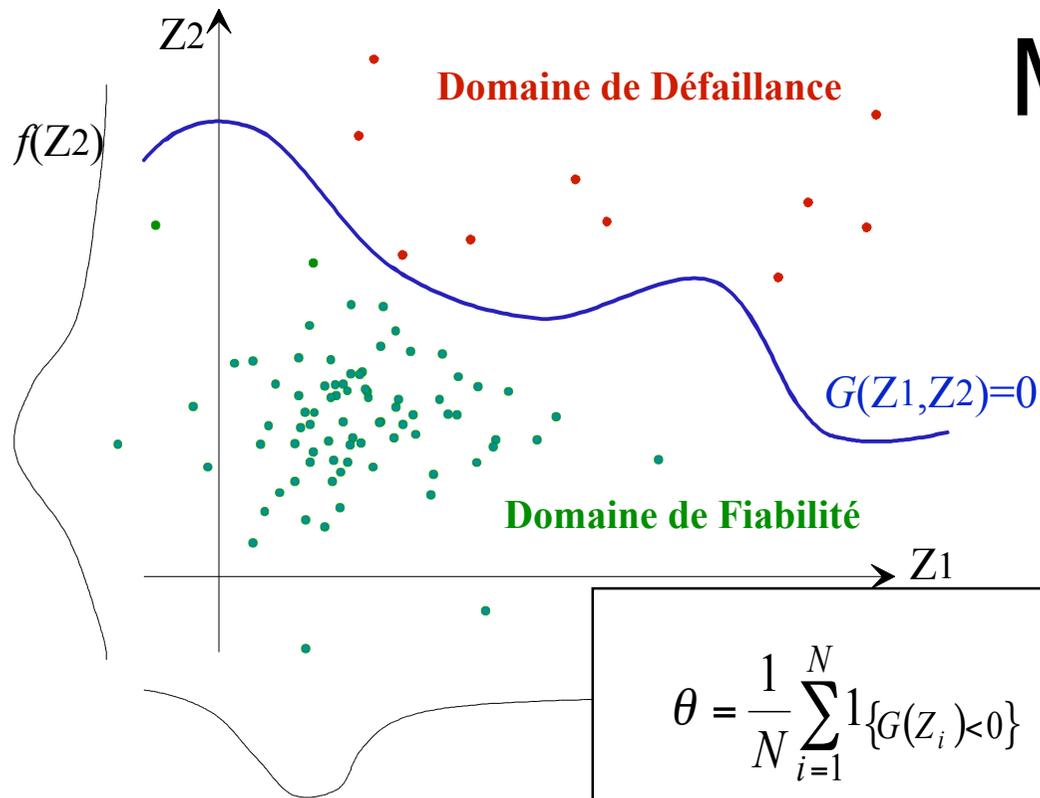
# Estimation de risque via propagation d'incertitudes

- Estimer une intégrale (ultra-complexe)

$$Pf = P(z = G(\underline{X}, \underline{d}) > z_s) = \int_{\underline{X}} 1_{G(\underline{X}, \underline{d}) > z_s} f_x(\underline{X}) d\underline{X}$$

- Alternatives classiques
  - Intégration numérique (pb. de la dimension de X et du contrôle d'erreur)
  - Simulation (Monte-Carlo, importance sampling ...)
  - Optimisation Form-Sorm et dérivées
  - Surfaces de réponse (dont E.F. stochastiques)
- *Vrai défi : justifier robustesse selon régularité de G ...*

# Monte-Carlo



$$Pf = \int_{\underline{X}} 1_{G(\underline{X}, d) > z_s} f_x(\underline{X}) d\underline{X}$$

$$\theta = \frac{1}{N} \sum_{i=1}^N 1_{\{G(z_i) < 0\}}$$

nombre moyen de défaillances  
obtenues parmi  $N$  tirages  
(somme de 0 et de 1)

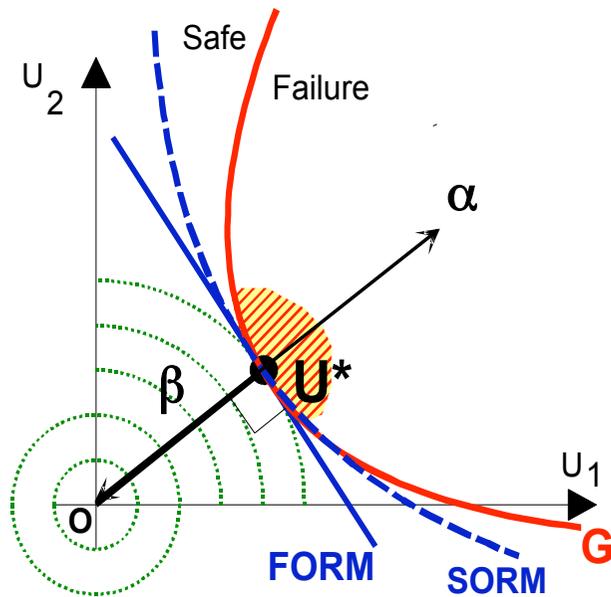
Variance de l'estimateur :

$$Var(\theta_N) = \frac{P_F(1 - P_F)}{N}$$

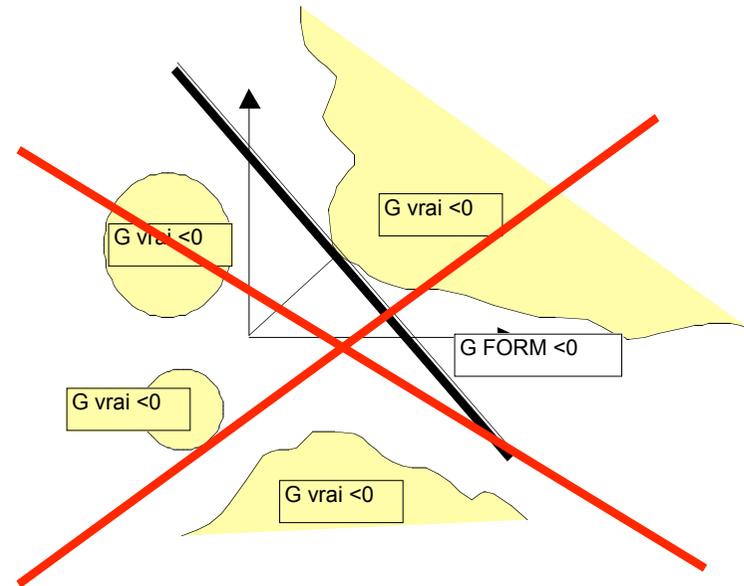
Importance Sampling : on biaise la densité de tirage

Mais la convergence est moins rapide / contrôlable ...  $\hat{P}_f = \frac{1}{n} \sum_{j=1..n} 1_{G(\underline{X}_j, d) > z_s} \frac{f_{\underline{X}}(\underline{X}_j)}{\tilde{f}_{\underline{X}}(\underline{X}_j)}$

# Form –Sorm et ses limites



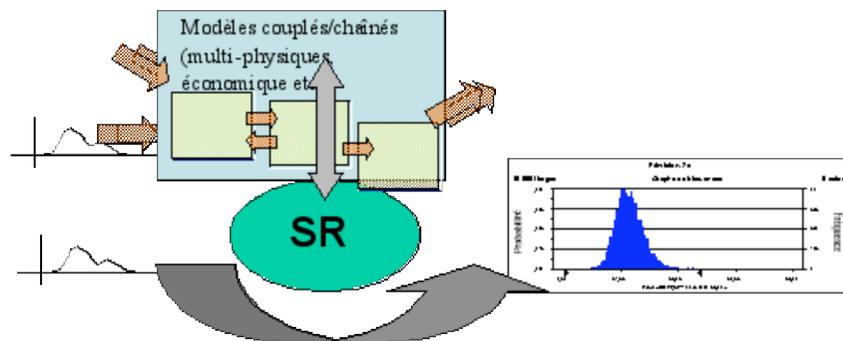
Principe : trouver par optimisation le point de dépassement de seuil (i.e.  $G(X,d) = z_s$ ) le plus proche de l'origine  $E(X)$  ... puis approximer par une surface régulière



SMAI – 7ème Rencontres Math-Industrie « Mathématiques et Environnement », 2007

# La stratégie des surfaces de réponse

- On remplace  $G(\cdot)$  (lourd en CPU) par une « surface de réponse »  $\tilde{G}(\cdot)$
- ... i.e.  $\tilde{G}(\cdot)$  approximant bien  $G(\cdot)$  et se calcule vite
  - polynômes, chaos polynômial, réseaux de neurone ...



$$G(x) = \tilde{G}(x) + u(x) = \sum_l \tilde{G}^l(x) \cdot \beta^l + \mu$$

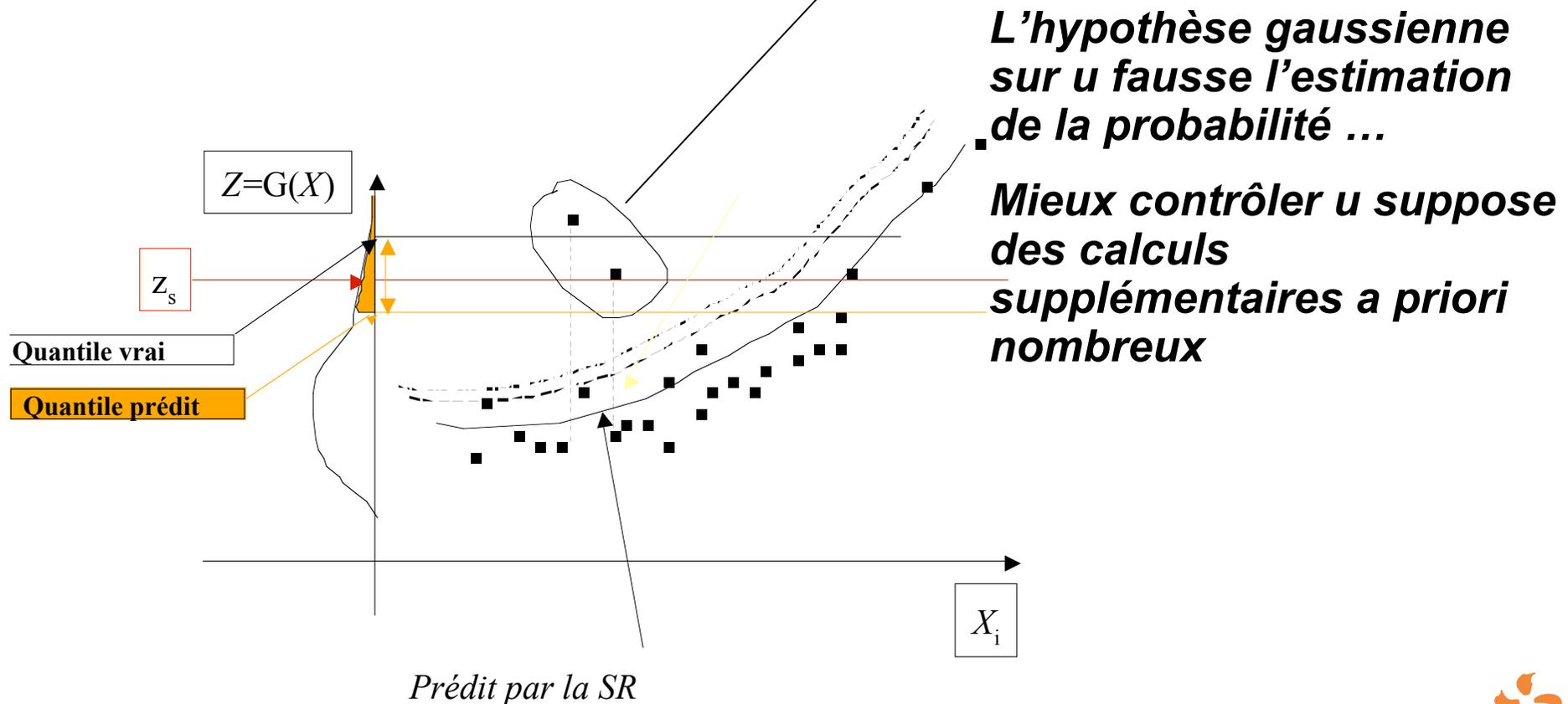
Estimation  $\hat{\beta}^l$  aux OMS / GLM

• **(Rarement mentionné) bien contrôler l'incertitude du résidu pour un calcul conservatif de Pf demande statistiquement beaucoup de calculs**

SMIAI – 7ème Rencontres Math-Industrie « Mathématiques et Environnement », 2007

# ... le cas difficile de physiques irrégulières

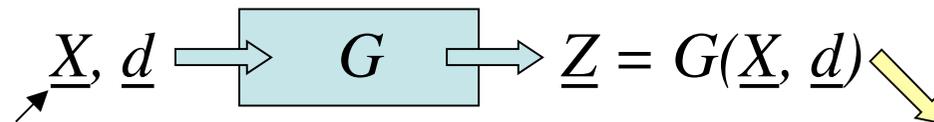
$$G(x) = \tilde{G}(x) + u(x) = \sum \tilde{G}^l(x) \cdot \beta^l + u$$



SMAI – 7ème Rencontres Math-Industrie « Mathématiques et Environnement », 2007

# Au-delà de la propagation simple ...

Cumuler les « incertitudes des incertitudes »



La *source d'incertitude* sur  $\underline{x}$  est modélisée par le vecteur aléatoire

$$\underline{X} \sim f_X(\underline{x} \setminus \underline{\theta}_X)$$

L'*estimation* de  $\underline{\theta}_X$  à données finies produit  $\Rightarrow$  variance non négligeable sur  $\hat{\theta}_X$

$$\underline{c}_Z = \mathcal{F}(f_Z(\underline{z} \setminus \underline{\theta}_X, \underline{d}))$$

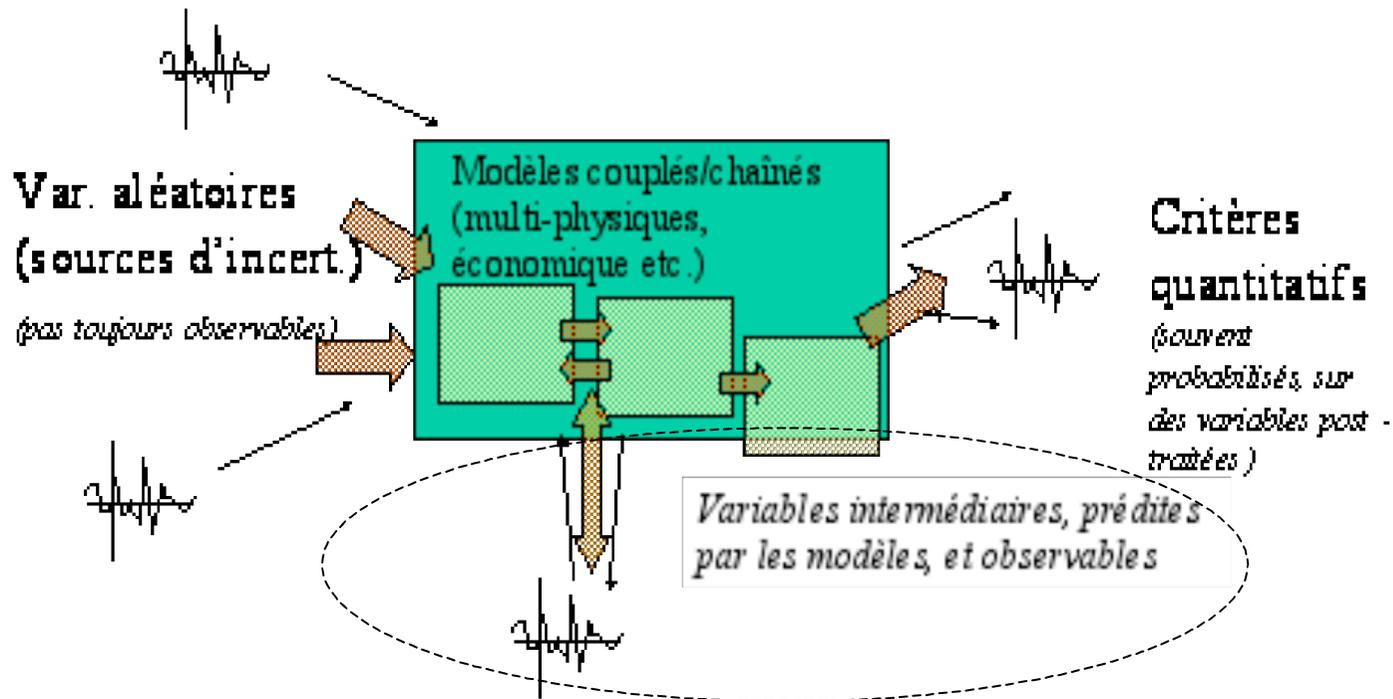
On s'intéresse généralement à un *critère d'incertitude* bien spécifique

$$c_Z = P(z > z_s), \text{ var } Z,$$

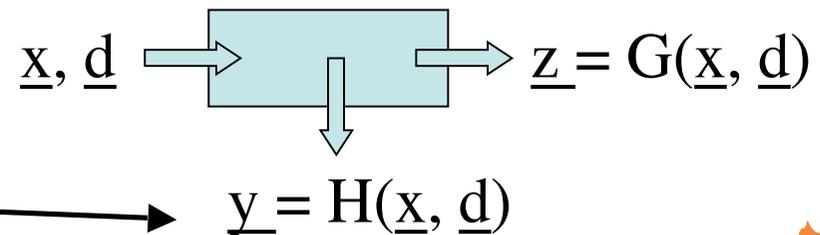
Même à  $\underline{\theta}_X$  parfaitement connu, contraintes CPU de *propagation*  $\Rightarrow$  variance d'estimation  $\hat{C}_z(\theta_X)$  parfois non négligeable

# Le problème n°2 : *inverse*

## Cadre de modélisation

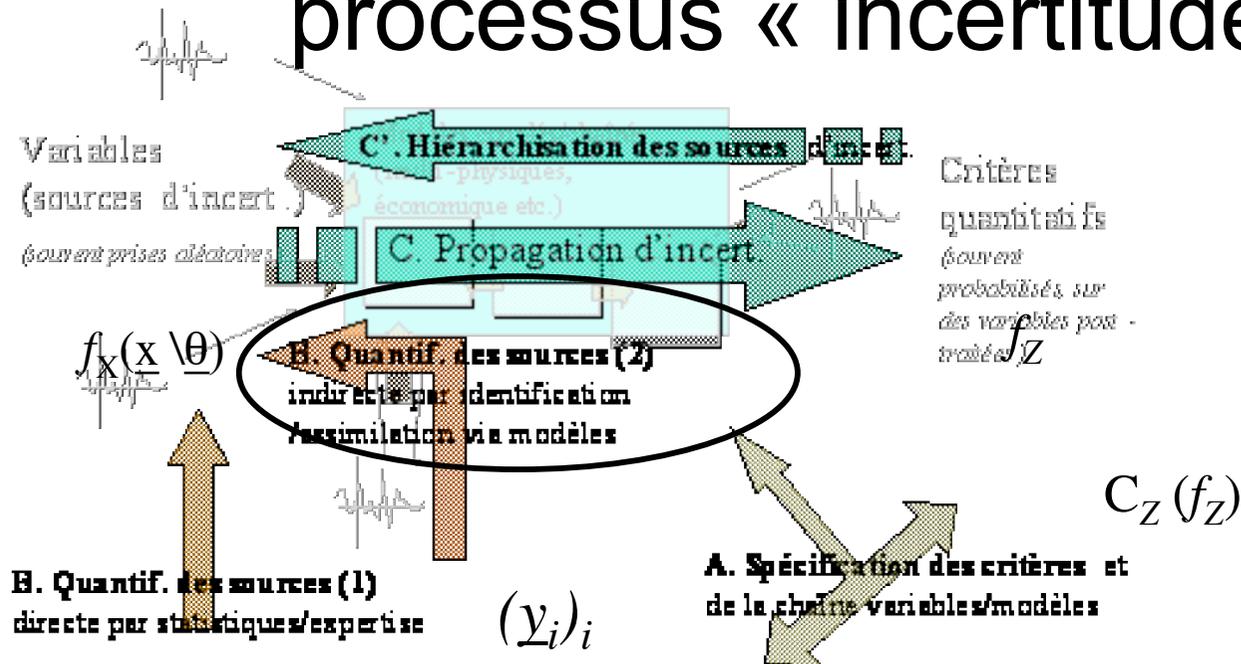


On a des observations,  
non pas sur  $\underline{X}$ , mais sur  $\underline{Y}$   
(via le bruit  $\underline{U}$ )  $\underline{Y}_m = \underline{Y} + \underline{U}$



SMAI – 7ème Rencontres Math-Industrie « Mathématiques et Environnement », 2007

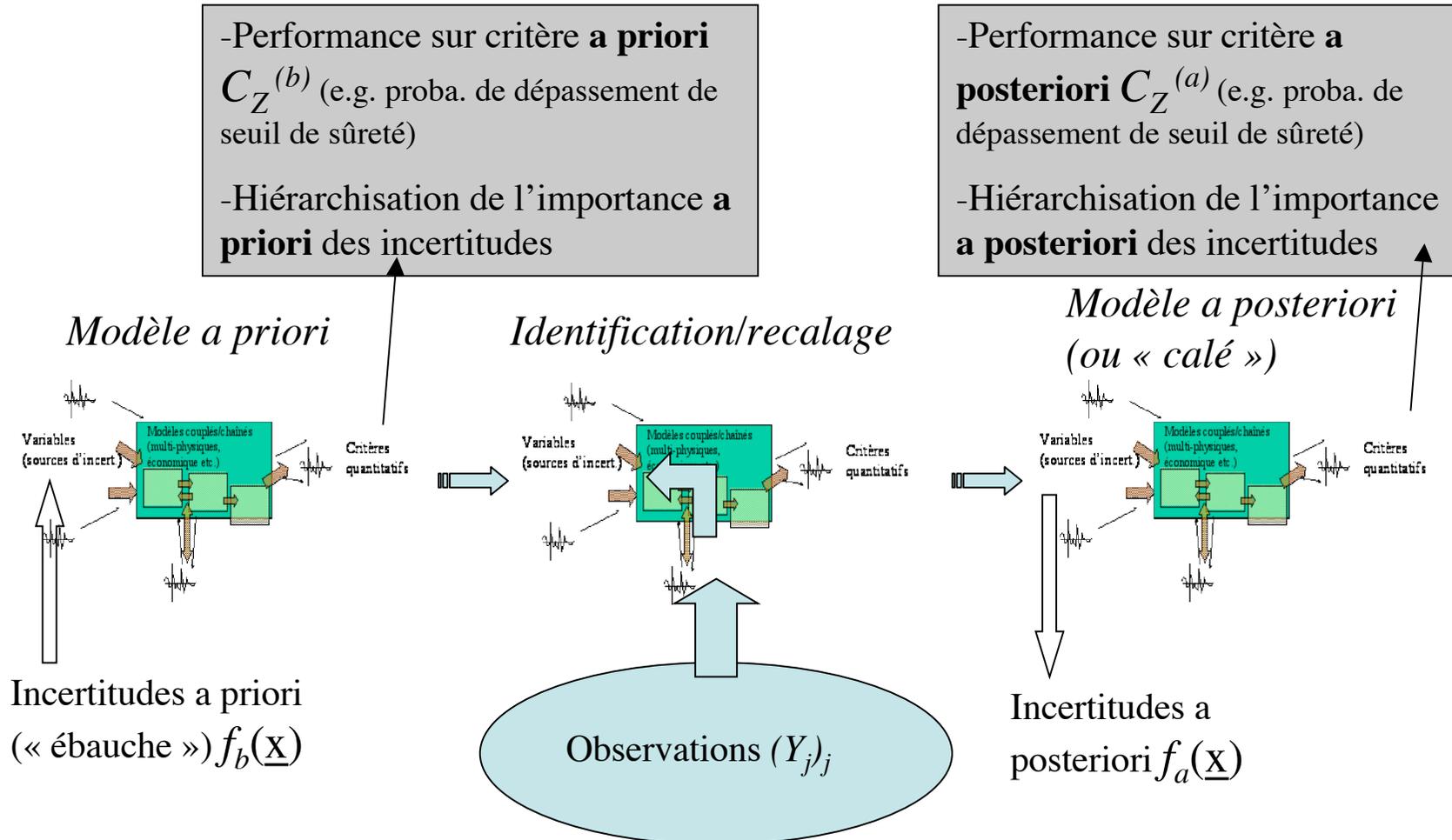
# L'étape inverse dans le processus « incertitudes »



• B, C et C' sont relativement « classiques » pour les spécialistes d'incertitudes

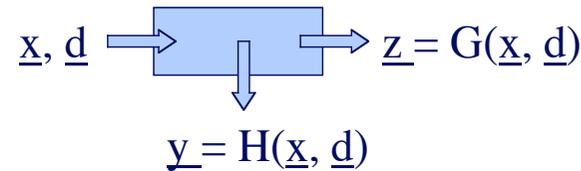
• B' l'est beaucoup moins ... elle a un potentiel industriel énorme ! (maîtrise dynamique des incertitudes)

# L'étape inverse dans le processus « incertitudes » (stationnaire)



SMIAI – 7ème Rencontres Math-Industrie « Mathématiques et Environnement », 2007

# Formalisation du problème (1)



$$\underline{Y}_{mj} = \underline{Y}_j^* + \underline{U}_j^* = H(\underline{X}_j^*, \underline{d}_j) + \underline{U}_j^*, j = 1..n$$

$d_j$  = conditions de fonctionnement / d'essais variant de façon déterministe et connue selon les n observations

$H$  est déterministe et connu, mais lourd à calculer (e.g. Elém. Finis) généralement non-linéaire

On observe les  $\underline{y}$  avec des écarts mesure-modèle incertains (supposés additifs) mais indép. et de lois connues (e.g. centrée de variance  $R$ , mais évent. dépendante de  $d_j$  et non gaussienne) ou non

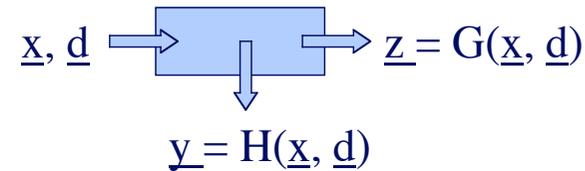
$$\underline{U}_j^* \sim f_U(u_j | \theta_U(d_j))$$

$X_j^*$  est inobservable ... on cherche sa pdf  
 Sa taille vectorielle = 3 - 15 ...

Une partie du vecteur  $X_j^*$  est de loi éventuellement connue ... manque  $\theta_X^{un}$

$$X_j^* = \begin{pmatrix} X^{kn}_j \\ X^{un}_j \end{pmatrix} \quad \begin{matrix} X^{kn}_j \sim f_{X^{kn}}(x^{kn}_j / \theta_X^{kn}) \\ X^{un}_j \sim f_{X^{un}}(x^{un}_j / \theta_X^{un}) \end{matrix}$$

# Formalisation du problème (2)

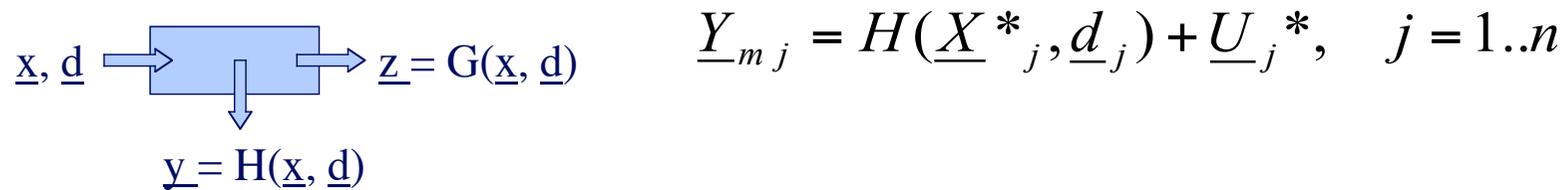


$$\underline{Y}_{m_j} = H(\underline{X}^*_{j}, \underline{d}_j) + \underline{U}_j^*, \quad j = 1..n$$

Connaissant  $(\underline{Y}_{m_j}, \underline{d}_j, \theta_{U_j})_{j=1..n}$  ainsi que  $\underline{\theta}_X^{kn}$  et  $H(\underline{x}, \underline{d})$  par calcul numérique (lourd)

Identifier  $\underline{\theta}_X^{un}$  de sorte que  $\underline{Y}_{m_j} \cong H(\underline{X}^*_{j}, \underline{d}_j) + \underline{U}_j^*$

# Formalisation du problème (3)



Par exemple, maximiser la vraisemblance (en ajustant  $\underline{\theta}_X^{un}$ )

Linéaire gaussien  $LL(\underline{y}_{mj} \setminus \underline{\theta}_X^{un}) = \dots \sum_j (\underline{y}_{mj} - H_j \underline{x}_m)' (R + H_j V H_j')^{-1} (\underline{y}_{mj} - H_j \underline{x}_m)$

...Non linéaire non Gaussien

$$\begin{aligned}
 Lik \left[ \left( \underline{y}_{mj} \right)_{j=1..n} \setminus \underline{\theta}_X^{un} \right] &= \prod_j \int_{\underline{u}_j^*} f_{U_j}(\underline{u}_j^*) \left[ \int_{\underline{x}^*} f_X(\underline{x}^* \setminus \underline{\theta}_X^{un}) 1_{\left[ \underline{y}_{mj} = H_j(\underline{x}^*) + \underline{u}_j^* \right]} d\underline{x}^* \right] d\underline{u}_j^* \\
 &= \prod_j \int_{\underline{u}_j^*, \underline{x}^*} f_{U_j}(\underline{u}_j^*) f_X(\underline{x}^* \setminus \underline{\theta}_X^{un}) 1_{\left[ \underline{y}_{mj} = H_j(\underline{x}^*) + \underline{u}_j^* \right]} d\underline{x}^* d\underline{u}_j^*
 \end{aligned}$$

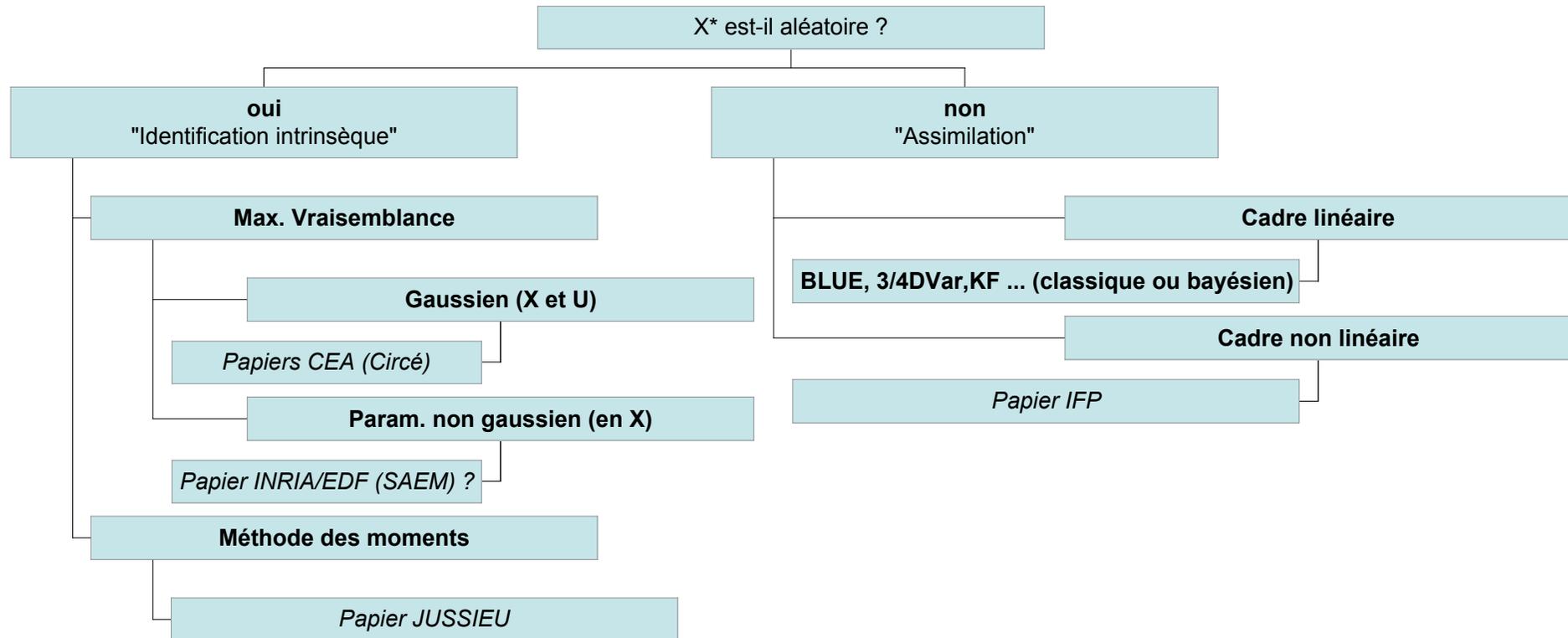
# Algorithmes de résolution

- Max. Vraisemblance trop « gros » en général à optimiser en direct
- Idée : insérer un PE/SR adaptatif dans les algorithmes itératifs MV
  - Linéarisation (adaptative) puis E-M gaussien (Circé)
  - SAEM
  - MCMC ...

Cf. [www.jds2006.fr/prob-ouverts.php](http://www.jds2006.fr/prob-ouverts.php)

SMAI – 7ème Rencontres Math-Industrie « Mathématiques et Environnement », 2007

# Classification de quelques algorithmes ...



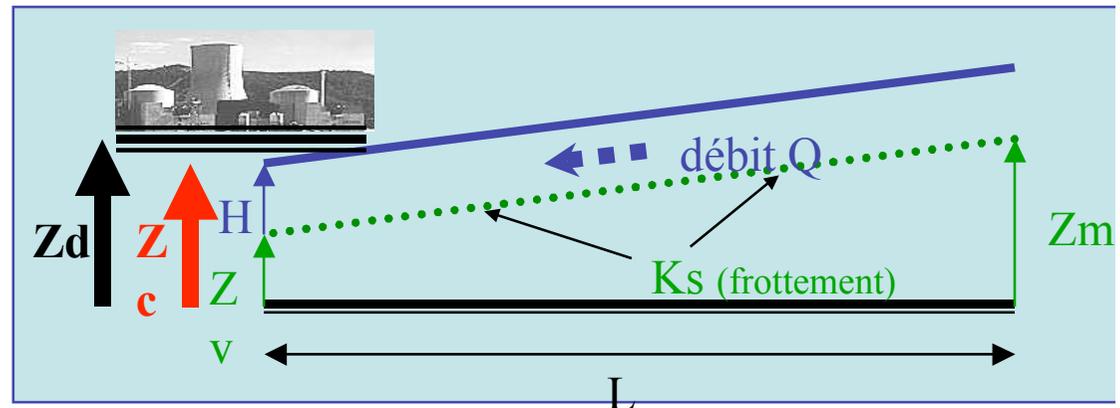
- *La 1ère question essentielle : en A.D.,  $X^*$  ne devient une v.a. que par erreur d'estimation : sa vraie valeur est unique, déterministe*

# L'exemple numérique des crues

- Un modèle physique simplifié (analytique, à 4 var. incertaines) ... mais à l'image d'un modèle physico-probabiliste industriel

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

$$Z = G(\underline{X}, \underline{d}) = Z_{seuil} - (Z_v + h)$$



- Critère d'intérêt = probabilité de débordement  $< 10^{-3}$   $c_Z = 1 \{p[Z < 0] < 10^{-3}\}$
- On mesure (à bruits de variances connues) :  $\underline{Y}_m = \begin{pmatrix} Z_c \\ V \end{pmatrix} = \begin{pmatrix} Z_c^* \\ V^* \end{pmatrix} + \underline{U} = H(\underline{X}, \underline{d}) + \underline{U} = \begin{pmatrix} Z_v + h \\ \frac{Q}{h \cdot B} \end{pmatrix} + \begin{pmatrix} U_Z \\ U_V \end{pmatrix}$
- 4 variables aléatoires, triangulaires ou gaussiennes : **3 v.a. de paramètres inconnus, à identifier**

$$\underline{X} = (X_Q \quad K_s \quad Z_v \quad Z_m) \quad X^{kn} = X_Q \quad \underline{d} = (Q_e \quad L \quad B)$$

$$Q = Q_e \cdot (1 + X_Q)^{-1}$$

SMIAI – 7ème Rencontres Math-Industrie « Mathématiques et Environnement », 2007

# Pour aller plus loin ...

- Problème ouvert d'intérêt industriel –  
« Statistiques, mesures et calcul  
scientifique inverse »,  
<http://www.jds2006.fr/prob-ouverts.php>
- de Rocquigny, La maîtrise des incertitudes  
dans un contexte industriel – Journ. Soc.  
Fran. Stat. n°147, 2006, pp.33-106
- GdR CNRS : MASCOT
- Groupe « Incertitudes et Industrie » de  
l'Institut de Maîtrise des Risques :  
[www.imdr-asso.fr](http://www.imdr-asso.fr)

SMAI – 7ème Rencontres Math-Industrie « Mathématiques et Environnement », 2007