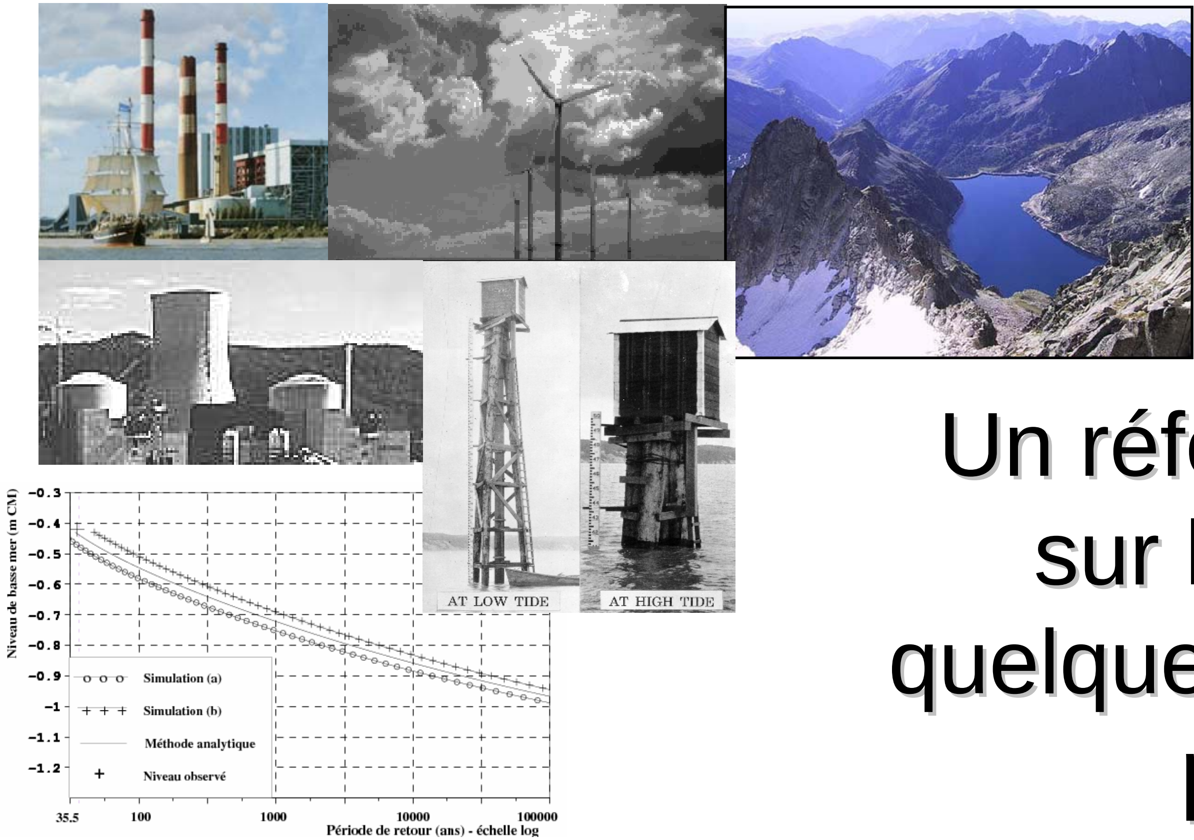


Un référentiel européen sur les incertitudes et quelques nouveaux défis pour la simulation

E. de Rocquigny

Mission transverse EDF R&D
« statistiques et incertitudes »

Chairman, ESReDA Uncertainties

European Safety, Reliability and Data Association

Plan de la conférence

- ESReDA : un *référentiel industriel européen* sur les incertitudes
- Nouveaux défis pour la simulation

ESReDA – a European industry-reference on uncertainty

- ESReDA = European Safety Reliability & Data Association
  - Established 1990, gathering some 52 members Europe-wide (80% industrial)
- Key goals of the Uncertainty Project Group (est. end 2004)
  - Provide a European industry-realistic guidelines on "Uncertainties"
  - Facilitate cultural dissemination for non-specialists
- Key steps
  - A book in early 2008 with John Wiley
  - Prepare through *intensely cooperative* working meetings

Key success features

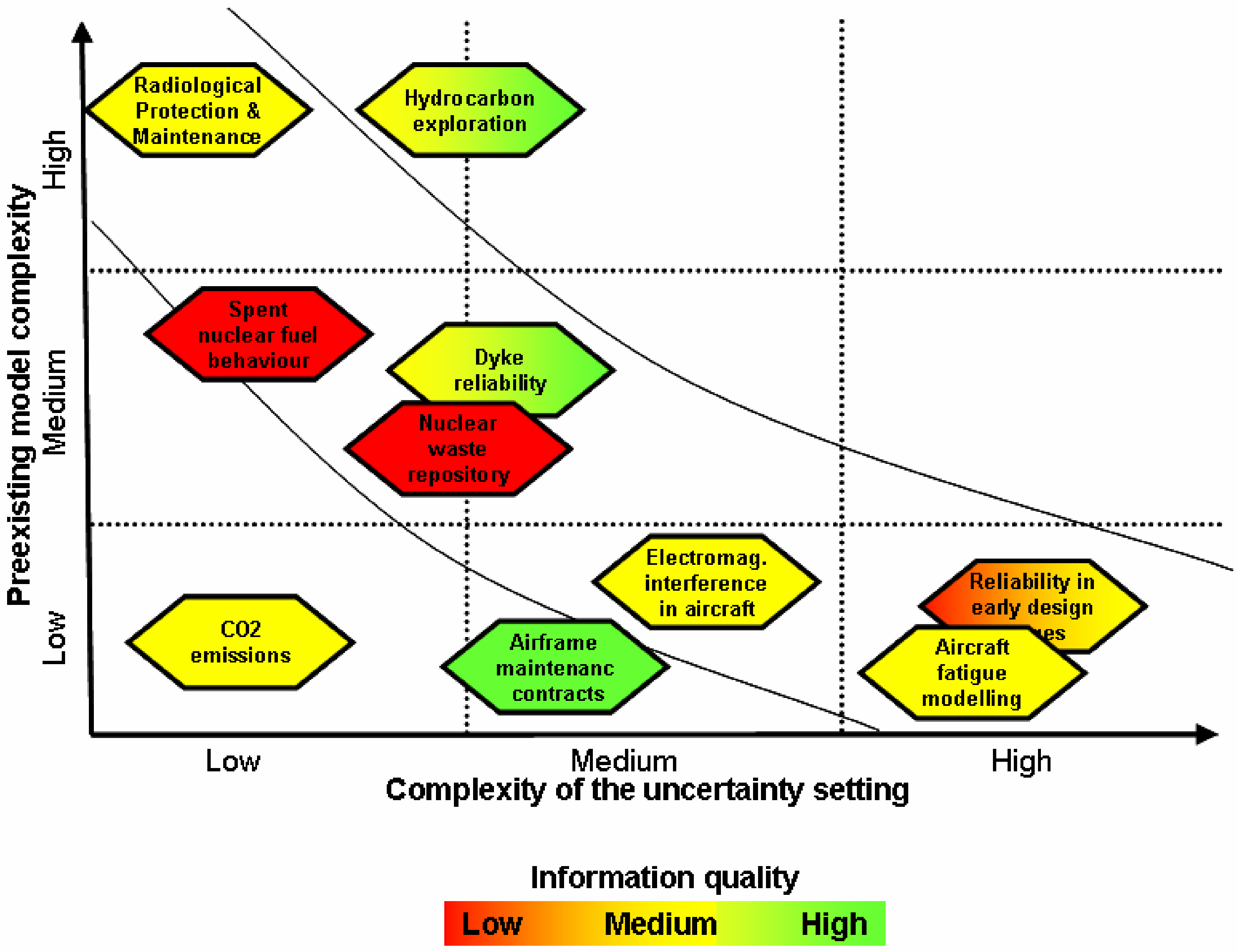
- Build on practical industrial examples
  - practice vs theory
- Be cross-disciplinary … *uncertainties in the real world do not restrict to one field*
  - Ex1 : in flood protection, use statistics on flow ; metrology and command/control uncertainties ; SRA on flood defence ; PSA in decision-making
  - Ex2 : external costs ; mix uncertainties on pollutant measurement, propagation, and uncertainties on health impacts + costs
- Refer to existing standards + best scientific references
  - Ex: IPCC guidelines, GUM/ISO in metrology, Sandia Labs etc.
- ***Not a standardisation effort as such, but link to on-going reg. (EU Comm.)***

Diversity of case studies

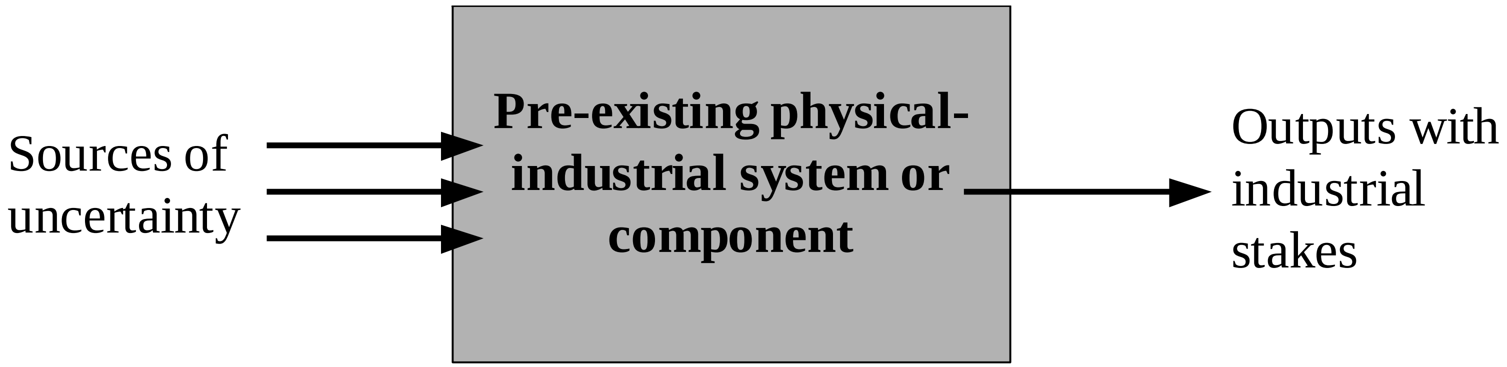
- Power (thermal) - CO2 emission control
- Nuclear – Maintenance and radioprotection
- Nuclear - Safety of Waste
- Nuclear – safety of interim storage
- Aerospace – EM compatibility – Personal Electrical Devices
- Aerospace – Fatigue modelling
- Aerospace - Maintenance contracts
- Reliability in early Design / Mechanical industry
- Civil engineering – safety of river dikes
- Oil industry – performance of oil exploration

Nationalities : France, Germany, Holland, Italy, Spain / European Commission JRCs

Diversity of case studies



Basic framework



Meteo. Randomness

Variability

Model imprecisions

….

Mechanical component

Waste storage facility

Maintenance program

…

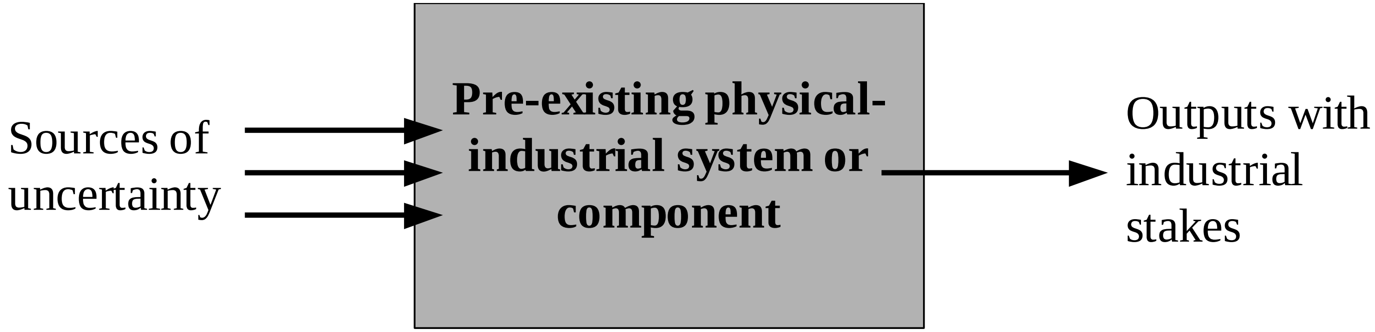
Mechanical margin to failure

Peak dose emitted

Net cost/benefit expected

….

Key items for an uncertainty study



Meteo. Randomness	*Mechanical component*	*Mechanical margin to failure*
Variability	*Waste storage facility*	*Peak dose emitted*
Model imprecisions	*Maintenance program*	*Net cost/benefit expected*
….	…	….

- One of 4 salient goal
- The role of the quantity(ies) of interest *(+ variable(s) of interest, decision criterion, uncertainty setting)*
- Key features of the pre-existing model
- A necessary feedback process

Common framework – key goals of an Uncertainty study

Four final goals for most uncertainty studies => a key difference for the methods :

<u>Understand</u> Influence of sources/ Rank Importance / Prioritize uncertainty

<u>Validate / Trust</u>, through model calibration / simplification

<u>Select</u> through comparison of relative performance (on quantities of interest) / Optimisation

<u>Comply</u> with threshold / criterion

Modelling framework and *quantities of interest*

Model inputs(factors) $\underline{x}$ *(uncertain)*

Model inputs(factors) $\underline{d}$ (fixed)
(design variables ,scenarios, “low importance factors” …)

System model
or
Pre-existing model

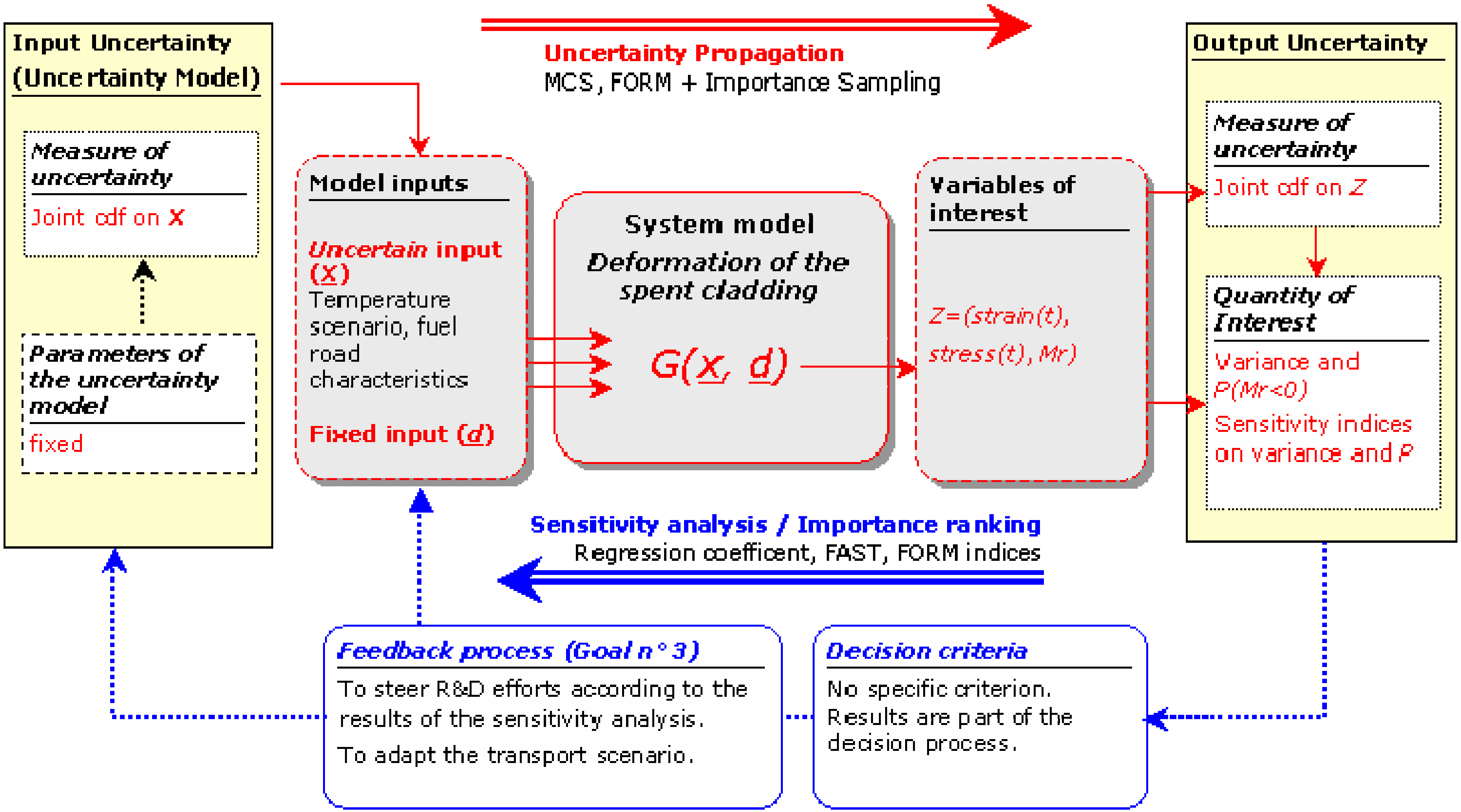
G($\underline{x}$, $\underline{d}$)

(Output) Variable(s) of Interest $\underline{Z}$, Z(t), Z(N) … = G($\underline{x}$, $\underline{d}$)

Quantity(ies) of interest

$P(z < z_s)$ (e.g. $z=R-S < 0$)
$C_v(z)$, $Var(z)$, f z (pdf)
$Bel\ (z < zs \setminus d_1) > Pl\ (z < zs \setminus d_2)$
$P(d_i = Argmax_d(G(X,d))$

Common conceptual framework



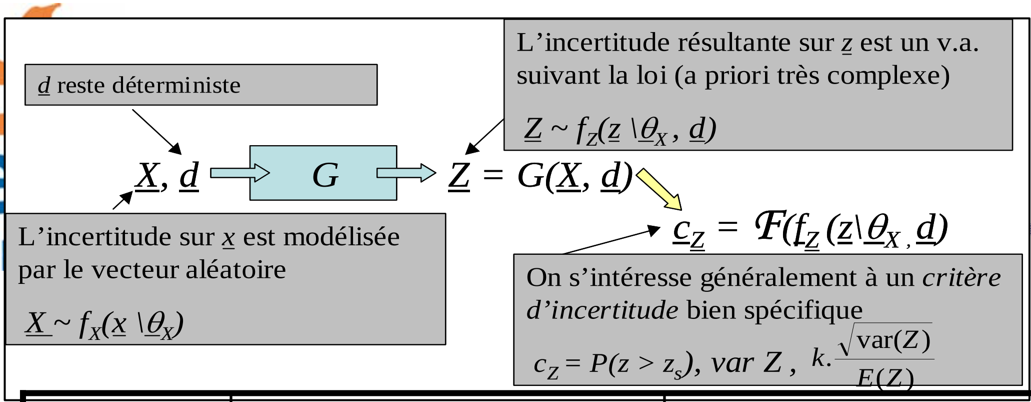
Common framework – goals and quantities of interest

Goals	Handling the q.o.I.	Comments
Comply	Demonstrate a threshold on a q.o.i., e.g. $P(Z<z_s \backslash d)<10^{-x}$, $cv(z)<b\%$ …	Mostly mixed deterministic / proba.
Select	Choose d to minimize a q.o.i. e.g. $E(G(x,d))$, $G(x,d)^{95\%}$	Mostly a few scenarios … *towards frontier optimisation*
Understand	Which x^i contributes most to the q.o.i. e.g. $Var[G(X,d)]$, $P(Z<z_s \backslash d)$ … ?	Large literature on variance or distribution analysis
Validate (calibrate / simplify …)	Identify X / measurements, or downscale X or G, provided control on q.o.i. …	Largely empirical … *towards complex stats.*

Common framework – key steps of an uncertainty study

- **Specify final goal, chain of model and choose uncertainty setting**
- **Build Uncertainty model**
- **Propagation**
- **Sensitivity analysis / importance ranking**

+ Feed-back process (New data ; Adjust X,d ; Simplify model ; Deterministic compliance)



Les grandes étapes

Etape **A**	Spécifier le problème	Choisir $\underline{x}$, $\underline{d}$, *G(.,.)* et un critère *($C_Z = P(Z > z_s)$* ou *var Z …)*	Sciences de la décision, analyse du risque, cadre classique ou bayésien …
Etape **B**	Modéliser /Quantifier les sources	Modéliser $\underline{X}$ en estimant $\underline{\theta}_X$ d'après $(X_j)_j$ ou expertise	Echantillonnage paramétrique classique ou bayésien *Mais aussi* : non-paramétrique, copules (loi jointe) …
Etape **C**	Propager	Estimer f_Z (ou plutôt C_Z)	Taylor, Monte-Carlo et accélérées, Form-Sorm, Plans d'Exp./Surfaces de Réponse …
Etape **C'**	Hiérarchiser l'importance des sources *Sensitivity analysis*	Estimer des S_X (selon C_Z) $S_X = \frac{Corr(Z, X^i)^2}{\sum_{i=1}^{N} Corr(Z, X^i)^2}, \frac{Var(E[Z \backslash X^i])}{Var(Z)}$	Idem + RCC/ PRCC, indices de Sobol estimés par MCS, quasi-MC … , FAST,

Characteristics of pre-existing models and choice of methods

The key characteristics to choose between methods (MCS, Form, response surface, chaos polynomials, …) :

- *Quantity of interest*
- *Data available*
- *Computing constraints on G(.)*
  - Dimension of X
  - Dimension of Z
  - Regularity of G(.)

Role of probabilistic settings ?

- Probabilistic/statistical settings : a central but non exclusive role
  - Proba/stat. modelling refines purely deterministic margins
  - Not necessarily applicable to all types of uncertainties – *mix deterministic/probabilistic - according to the case*
  - Separate epistemic / aleatory natures ? *Inventory is useful as a check-list … adapt to criterion / model etc.*
  - Mixed non proba. / proba. : *not yet industrially mature … to be followed on*

? Common framework –Uncertainty settings

Name	**Level 1** **uncertainty**	**Level 2** **Parameters of the uncertainty (« epistemic »)**
Determ.	X at « Worst value » (intervals)	Fixed (fixed bounds)
Standard proba.	X ~ pdf	Fixed parameters θ_X of the pdf
Proba + L2 Determ.	idem	Deterministic interval for θ_X
Double Proba	idem	θ_X becomes a random vector
Mixed proba. / DST	Idem (or not …)	θ_X becomes DST belief / plausibility df.

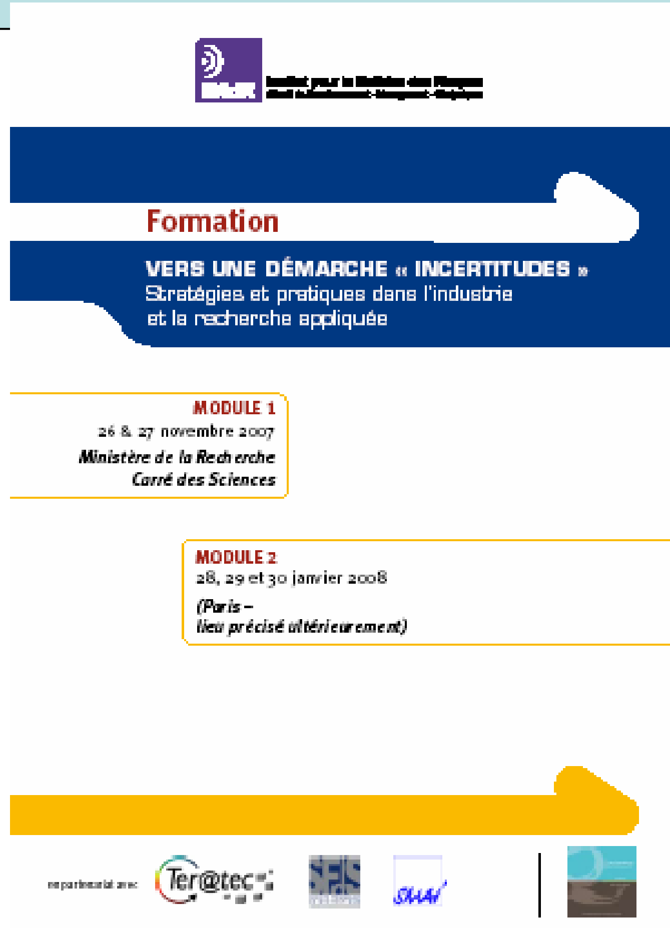
Further on …

- Book release around early 2008

« Uncertainty in Industrial Practice », Wiley

•*Formation IMdR – Teratec - SFdS – SMAI*

26/27 novembre 07 + 28/30 janvier 08



Nouveaux défis pour la simulation

- *Understand* >> Sensibilité (en grande dimension)
- *Comply* >> Propagation d’incertitudes (robuste, et de « niveau 2 »)
- *Validate* >> *Calibration / assimilation / méthodes inverses probabilistes*

Cumuler les « incertitudes des incertitudes »

$\underline{X}, \underline{d}$ → G → $\underline{Z} = G(\underline{X}, \underline{d})$

$\underline{c_Z} = \mathcal{F}(\underline{f_Z}(\underline{z} \backslash \underline{\theta_X}, \underline{d}))$

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

$\underline{X} \sim f_X(\underline{x} \backslash \underline{\theta_X})$

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

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

$c_Z = P(z > z_s)$, *var Z* ,

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

Ordres de grandeurs des défis pour la simulation

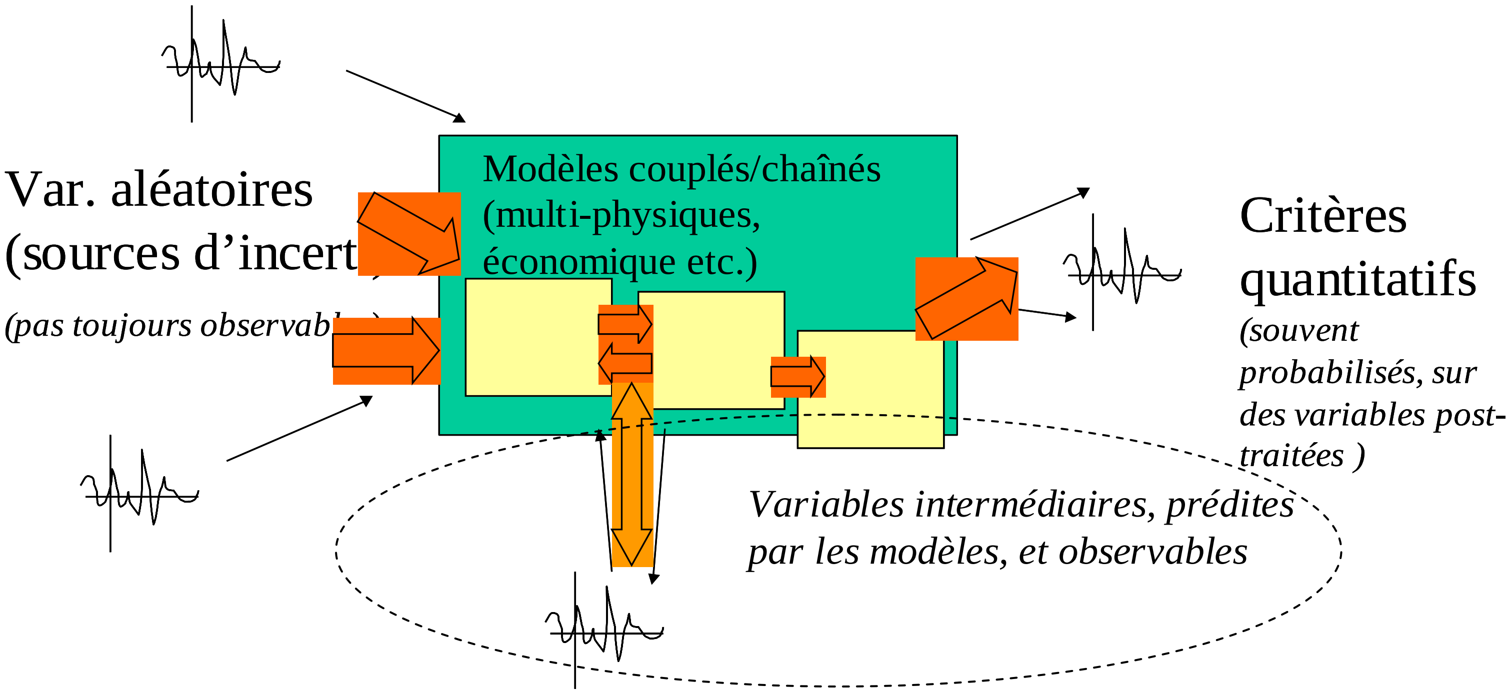
Z=G(X,d) où X~f(x\θ), dimX=p, dim θ=t (nb. optimal selon « régularité »)

- Propager
  - Niveau 1 - *Var Z* : n= 10^1 à 10^2 ;
  - Niveau 1 - quantile Z^{α} (ou P(Z>z_s) : n=10^1p (ou 10^2) à >10^4 selon α
  - *Niveau 2 – x10 (ou x3p) à x10^2*
- Hiérarchiser
  - Pour var Z : n= 3p à >10^3p (voire a ^{p}) selon régularité code
  - Pour quantile/proba dépassement : *mal connu*
- Calibrer / assimiler / identifier
  - 10^0p à >10^1p en calibration / assimilation classique
  - 10^1p à >10^4p en identification complète de la variabilité

(du linéaire/gaussien au cas général …)

Calibration de modèles et identification inverse d'incertitudes

Des observations et un modèle à calibrer

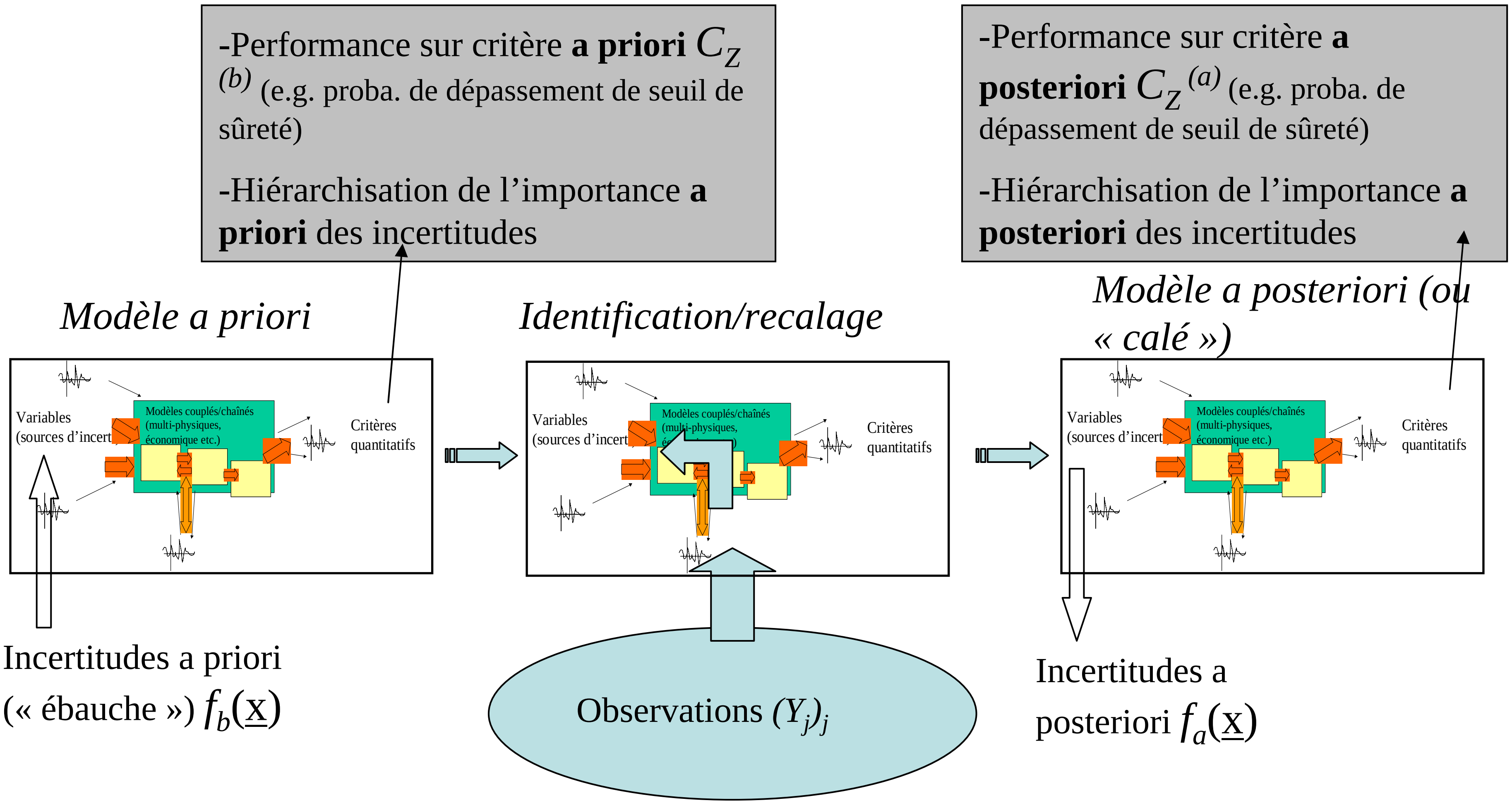


Des observations, non pas sur les paramètres $\underline{X}$ du modèle mais sur $\underline{Y}$ (via l'erreur modèle/mesures $\underline{U}$) $\underline{Y}_m = \underline{Y} + \underline{U}$

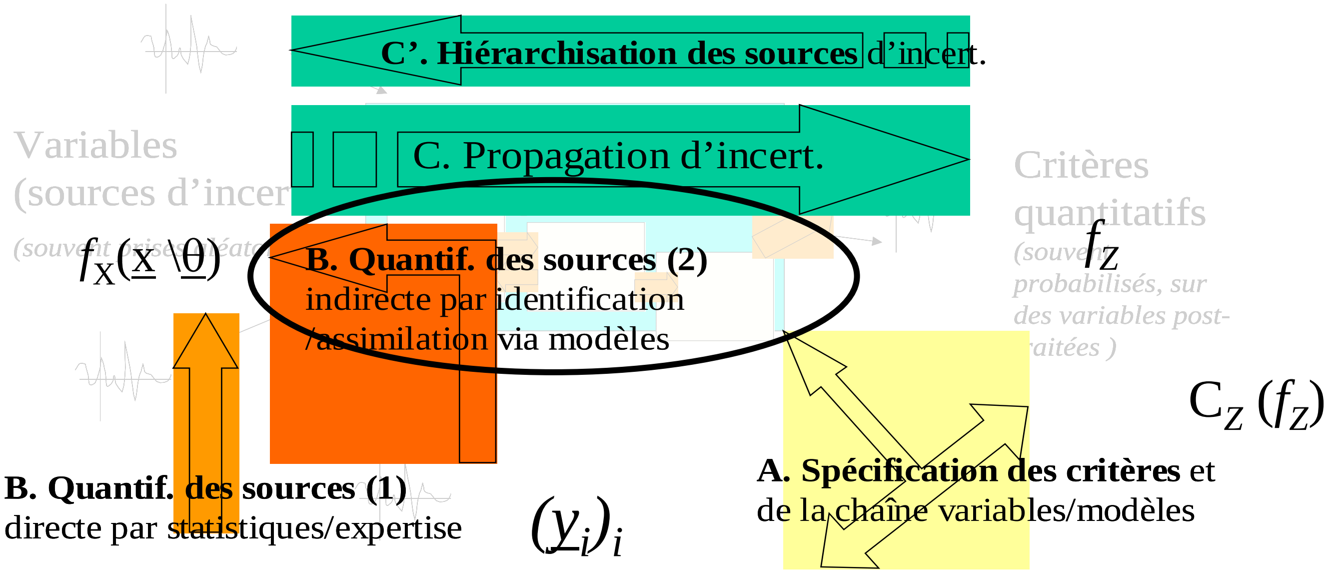
$\underline{x}$, $\underline{d}$ → $\underline{z} = G(\underline{x}, \underline{d})$

$\underline{y} = H(\underline{x}, \underline{d})$

L’étape inverse dans le processus « incertitudes »



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)

Formalisation du problème de calibration inverse (sous incertitude)

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

$\underline{y} = H(\underline{x}, \underline{d})$

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

$$\underline{Y}_i = H(\underline{X}^*_i, \underline{d}_i) + \underline{\varepsilon}_i^*, \quad i = 1..p$$

H est le modèle physique déterministe …éventuellement non-linéaire

On observe les $\underline{y}$ avec des bruits mesure-modèle incertains … typiquement aléatoires, indépendants … de loi parfois connue

$$\varepsilon_i^* \sim F_\varepsilon\left(R(d_i)\right)$$

Xi est inobservable … on cherche à identifier sa meilleure valeur … voire son incertitude (taille vectorielle = 3 – 15 …)

Une première approche : régression / assimilation de données

Depuis l'identification par méthodes inverses déterministes …

- Techniques d'optimisation : moindres carrés ou d'autres fonctions de coût …

$$x_a = \min_{x \in R^k} J(x) = \min_{x \in R^k} \sum_{i=1}^{N} \left\| y_i - H_i(x) \right\|_\alpha$$

- Avec parfois de la sensibilité « déterministe » / optimum

… généralisable par de « l'optimisation bruitée »

- En rajoutant simplement une matrice de coût

$$J(x) = \sum_{i=1}^{N} (y_i - H_i(x))^T R_i^{-1} (y_i - H_i(x))$$

- une interprétation proba. devient possible : variance de Xa ~

$$A = \left(\frac{1}{2} J''(x_a) \right)^{-1}$$

Ce sont les algorithmes de « l'assimilation de données » (i.e. régression, mais avec un modèle physique)

- **Largement pratiquée en météo., etc.**

Le contenu des algo. d'assimilation - cas linéarisé (1)

- Cadre linéarisé (en stationnaire) : les moindres carrés classiques …

$$Y = H(x_t) + \varepsilon$$

On linéarise le modèle

$$H(x) = H(x_l) + \mathbf{H}(x - x_l)$$

On observe les y avec une incertitude de mesure de covariance R (sans biais)

$$\varepsilon \sim F(0, R)$$

- L'estimateur BLUE est
- …. de variance

$$\hat{X}_a = x_l + (\mathbf{H}^T . R^{-1} . \mathbf{H})^{-1} . \mathbf{H}^T . R^{-1} . \big(Y - H(x_l)\big)$$

$$V(X_a) = A = (\mathbf{H}^T . R^{-1} . \mathbf{H})^{-1}$$

>> c'est simplement une régression linéaire multiple un peu généralisée, … en notant « *H* » par « *X* » et R=Id, on retrouve $(X'X)^{-1}$

Des limites à la calibration type « régression – assimilation » …

- Mais la variance estimée A ne représente pas toute l'incertitude possible …
  - **On montre que A -> 0 si N -> ∞**
    - **Avec N observ. à R et H constant**

$$Y_i = H(x_t, d_i) + \varepsilon_i = H(x_t) + \varepsilon_i \quad avec \quad \varepsilon_i \sim L(0, R)$$

$$A = \left(B^{-1} + \mathbf{H}^T . \left(\frac{R}{N} \right)^{-1} . \mathbf{H} \right)^{-1}$$

A->0 (quelque soit l'ébauche)

- **Ce qui reste vrai, même à essais/variances d'erreurs variables (démontré sous conditions algébriques raisonnables (injectivité de H – intersection nulle Ker Hi))**

- **Ce modèle probabiliste suppose l'absence de variabilité physique intrinsèque**
  - **On a modélisé en fait l'observation de :**

(x_t est inconnu mais déterministe)

$$Y_i = H(x_t, d_i) + \varepsilon_i = H_i(x_t) + \varepsilon_i$$

- **Là où, physiquement, on a parfois**

(X_i est inconnu ET varie aléatoirement pour i ≠ j

$$Y_i = H(X_i, d_i) + \varepsilon_i = H_i(X_i) + \varepsilon_i$$

Une classe plus générale pour la variabilité intrinsèque : algo. issus d'EM

- Partir du modèle probabiliste « intrinsèque » (cadre linéaire gaussien)

$$Y_i = H(X_i^*, d_i) + \varepsilon_i = H_i(X_i^*) + \varepsilon_i$$

$$X_i^* \sim \mathrm{N}(\mathrm{x_m}, \mathrm{V}) \quad \text{iid} \quad \mathrm{x_m}, \mathrm{V} \; \textit{inconnus}$$

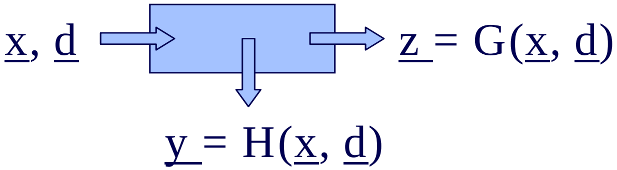
Différence majeure / régression-assimilation : Xi est non seulement méconnu mais physiquement variable (variance V inconnue)

- **Minimiser une nouvelle fonction (log-vraisemblance)**

$$LL((y_i)_i \backslash x_m, V, R_i) = -\frac{1}{2}\sum_i [y_i - H_i(x_m)]^T \left[H_i V H_i^T + R_i\right]^{-1} [y_i - H_i(x_m)] + \ldots$$

- Ce n'est plus du tout le « moindres carrés » classique : des algorithmes itératifs plus complexes sont alors nécessaires (type E-M Expectation-maximisation)

Formalisation du problème (3)



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

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

Linéaire gaussien $\quad LL(\underline{y}_{m\,j} \setminus \theta_X^{un}) = \ldots \sum_j (\underline{y}_{m\,j} - H_j \underline{x}_m)'(R + H_j \underline{\underline{V}} H_j')^{-1}(\underline{y}_{m\,j} - H_j \underline{x}_m)$

…Non linéaire non Gaussien

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

$$= \prod_j \int_{\underline{u}_j^*, \underline{x}^*} f_{U_j}(\underline{u}_j^*) f_X(\underline{x}^* \setminus \underline{\theta}_X^{un}) 1_{\left[\underline{y}_{m\,j} = H_j(\underline{x}^*) + \underline{u}_j^*\right]} d\underline{x}^* d\underline{u}_j^*$$

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

Classification de quelques algorithmes …

- X* est-il aléatoire ?
  - **oui** "Identification intrinsèque"
    - **Max. Vraisemblance**
      - **Gaussien (X et U)**
        - *Papiers CEA (Circé)*
      - **Param. non gaussien (en X)**
        - *Papier INRIA/EDF (SAEM) ?*
    - **Méthode des moments**
      - *Papier JUSSIEU*
  - **non** "Assimilation"
    - **Cadre linéaire**
      - **BLUE, 3/4DVar,KF ... (classique ou bayésien)**
    - **Cadre non linéaire**
      - *Papier IFP*

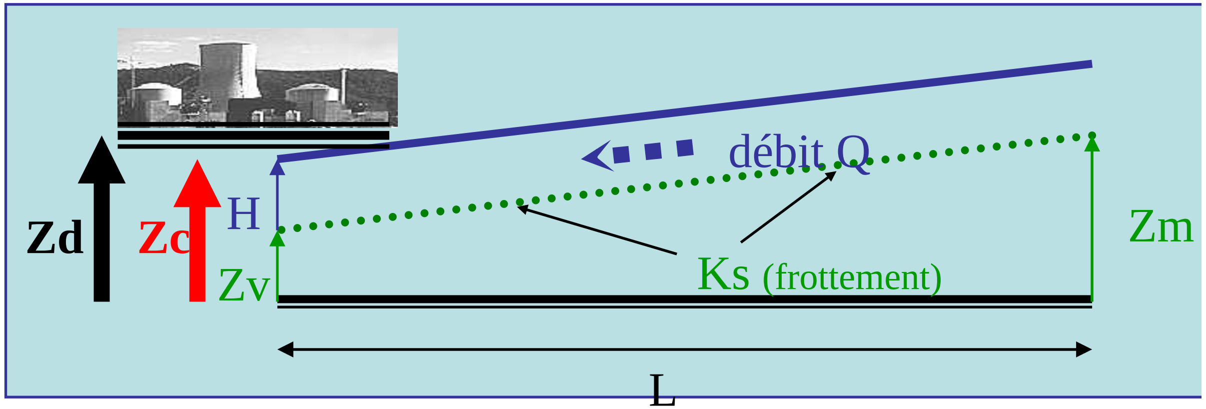
- *La 1ère question essentielle : en A.D., X* ne devient une v.a. que par erreur d'estimation : sa vraie valeur est unique, déterministe*

L'exemple numérique des crues

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

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

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



$$c_Z = 1_{\{P[Z<0]<10^{-3}\}}$$

- Critère d'intérêt = probabilité de débordement < 10-3

- On mesure (à bruits de variances connues) : $\underline{Y}_m = \begin{pmatrix} Z_c \\ V \end{pmatrix} = \begin{pmatrix} Z_c * \\ V * \end{pmatrix} + \underline{U} = H(\underline{X}, \underline{d}) + \underline{U} = \begin{pmatrix} Z_v + h \\ \frac{Q}{h.B} \end{pmatrix} + \begin{pmatrix} U_Z \\ U_V \end{pmatrix}$

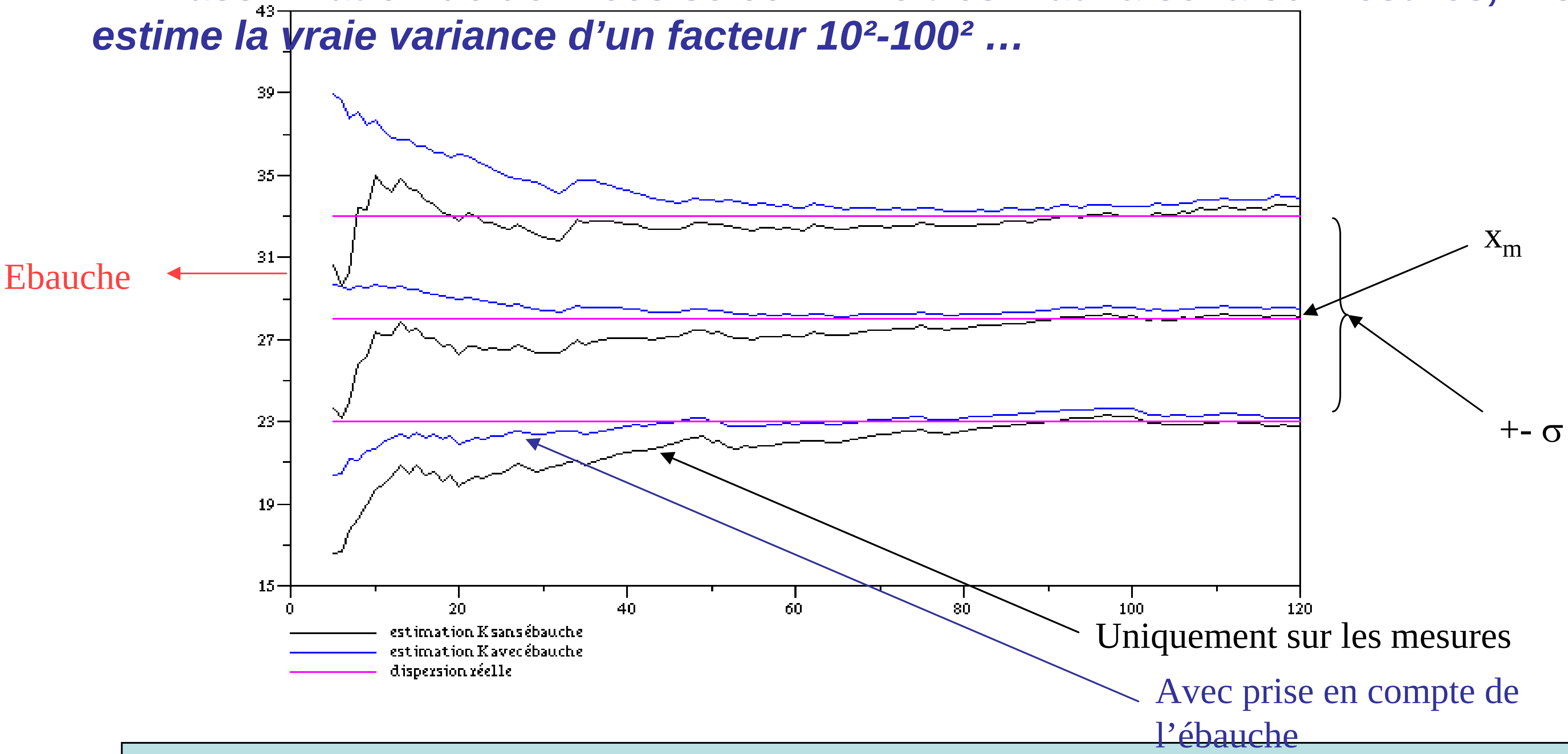
- *4 variables aléatoires, triangulaires ou gaussiennes ;* ***<u>3 v.a. de paramètres inconnus, à identifier</u>***

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

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

Résultats sur les crues en gaussien linéaire

•>> *l'assimilation de données se confirme très mauvaise :à 30 mesures, A sous-estime la vraie variance d'un facteur 10²-100² …*



Estimation raisonnable de l'incertitude, l'ébauche améliorant l'estimation à faible nb d'observations, mais détériorant un peu à N grand