

Congrès SMAI 2011

Risques rares
*Est-il raisonnable de
modéliser des probabilités
aussi faibles ?*

E. de Rocquigny
Ecole Centrale Paris

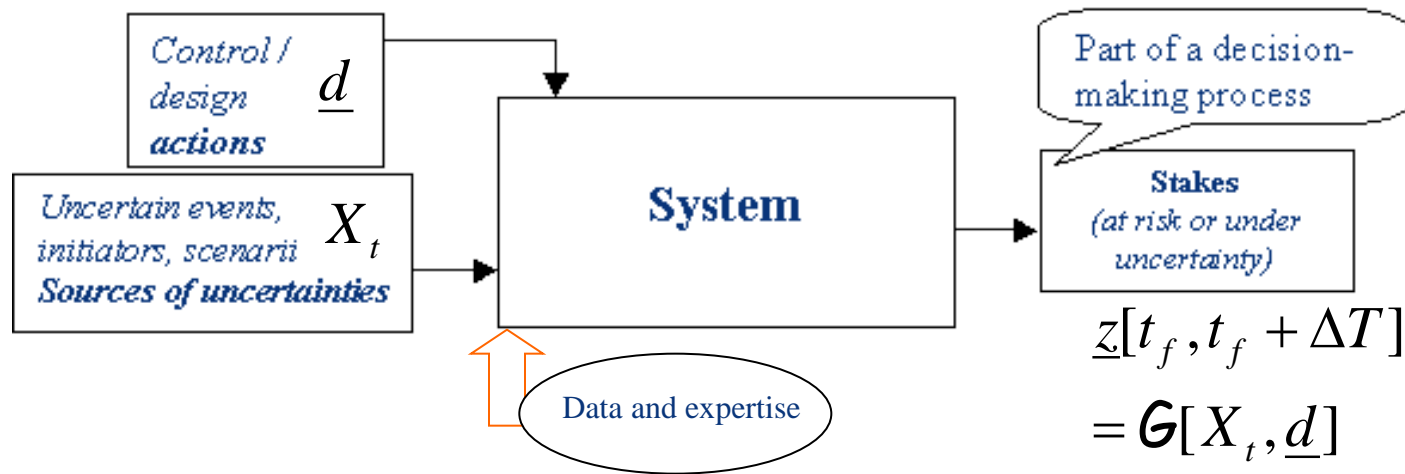
Plan

- ◎ Motivations – quelle rareté ?
- ◎ Calcul d'un risque rare (mesure de risque)
- ◎ Estimation des paramètres incertains (facteurs de risque)
- ◎ Calcul du couple risque / incertitude et des sensibilités
- ◎ Modélisation et réglementation – a-t-on le choix ?

Plan

- ◎ Motivations – quelle rareté ?
- ◎ Calcul d'un risque rare (mesure de risque)
- ◎ Estimation des paramètres incertains (facteurs de risque)
- ◎ Calcul du couple risque / incertitude et des sensibilités
- ◎ Modélisation et réglementation – a-t-on le choix ?

Décision à base de modèle (sous risque & incertitude)



◎ Décision à base de modèle:

- Conditionnellement à \underline{d} (actions, scénarios, ...)
- Evaluer – et assurer « pour une vraisemblance donnée » - que \underline{z} respecte une certaine propriété \mathcal{C}_z sur la période ΔT

Modélisation et mesure du risque c_Z

- Comparer / optimiser des alternatives risquées

$$C_Z : \{c_Z(\underline{d}_A) \leq c_Z(\underline{d}_B)\}$$

- Contrôler le respect d'une limite (d'un budget) de risque

$$C_Z : \{c_Z(\underline{d}) \leq c_s\}$$

où:

$$c_Z(\underline{d}) = \mathcal{F}[Z_t, \underline{d}]$$

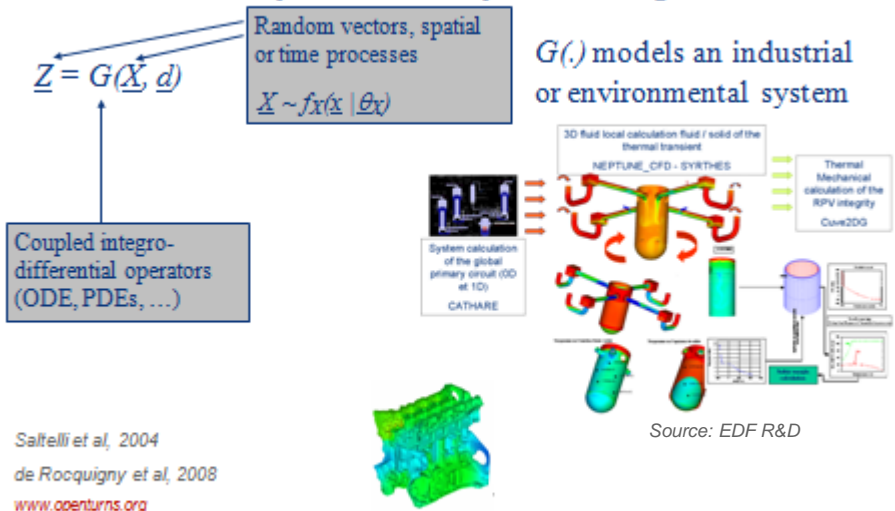
$$\begin{cases} c_Z(\underline{d}) = E1_{\text{Max}_{[t_f, t_f + \Delta T]} Z_t(\omega) \leq z_s} = P[\text{Max}_{[t_f, t_f + \Delta T]} Z_t \geq z_s | \underline{d}] \\ c_Z(\underline{d}) = E\left[\sum_{T_k} C(Z_{T_k}, T_k) \middle| (T_k)_k = \{t \in [t_f, t_f + \Delta T] | Z_t \geq z_s\}\right] \\ \dots \end{cases}$$

◎ Exemples de modèles de systèmes

$$\underline{z}[t_f, t_f + \Delta T] = G[X_t, \underline{d}]$$

Probabilistic physics & engineering

Uncertainty & sensitivity modelling



Saltelli et al, 2004

de Rocquigny et al, 2008

www.openturns.org

Reliability

Static risk & reliability models

- Failure/event trees, BDD, BBN etc. + lifetime statistics

$$e = \sum_i \left[e^{m_i} \cdot \prod_{j, (\underline{E}_i)} e^{n_j} \right] = G(e^{m_i}, e^{n_j}, \underline{d})$$

$$E^j = I_{t > T_j} \quad T_j \sim f_T(t_j | \underline{\theta}_T)$$

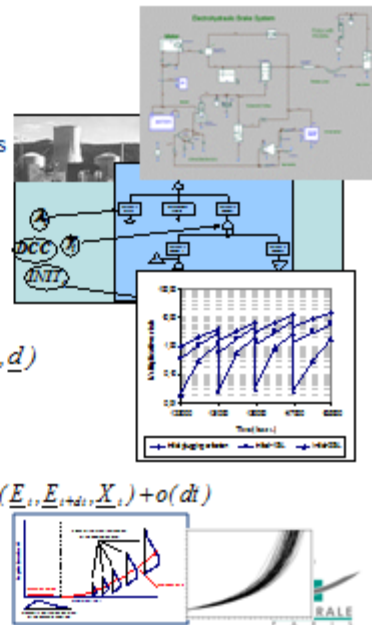
Dynamic reliability models

- « Chained » renewal processes $Z_t = \sum_j I_{E^j} g_j(\underline{X}_j, \underline{d})$

e.g. Probabilistic life cycle management

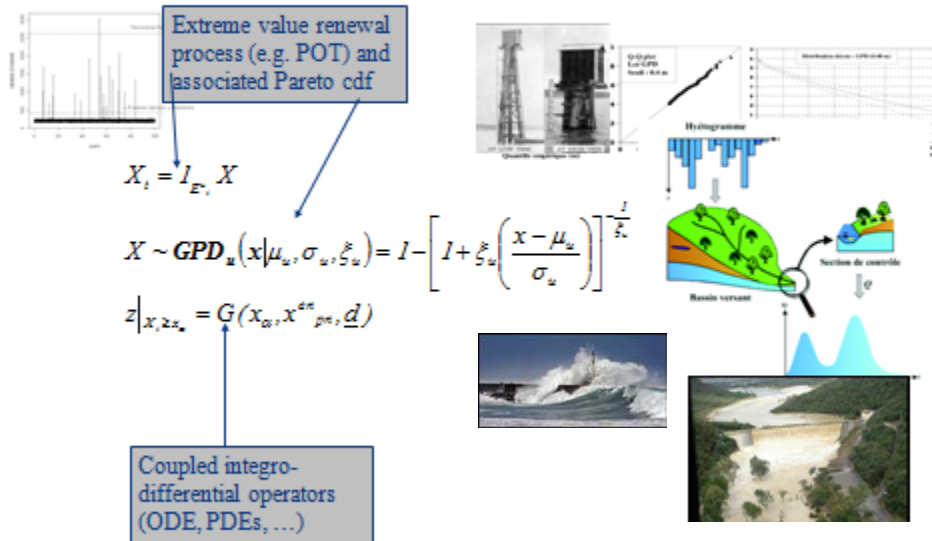
- « Coupled » dynamic processes given $\underline{X}_t = F[\underline{X}_{t-1}, \underline{E}_t]$

(MP, PDMP ...) or $P(\underline{E}_{t+dt} | \underline{E}_t) = h(\underline{E}_t, \underline{E}_{t+dt}, \underline{X}_t) + o(dt)$



Natural risk / extreme events

(External aggressions, ...)



Phenomenological risk models

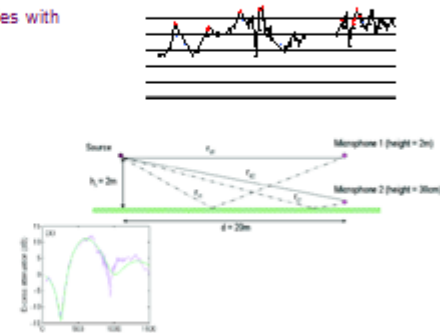
(environmental / biological warning systems)

- Regression or ANOVA-based closed-form functions
- Calibrated on explicative (possibly auto-regressive) variables with explicit residual error

$$Z_{i+1} = G(x_i, \underline{x}_{i-1}, \dots, \underline{\theta}, \underline{d}) + U$$

$$E_{i+1} = \sum_i \left[\prod_j e^{y_j} I_{M_j}(\underline{z}_i, \underline{z}_{i-1}, \dots, \underline{E}_i, \underline{d}) + U, \leq 0 \right]$$

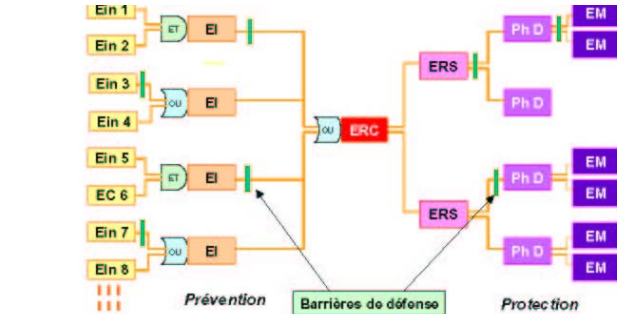
...



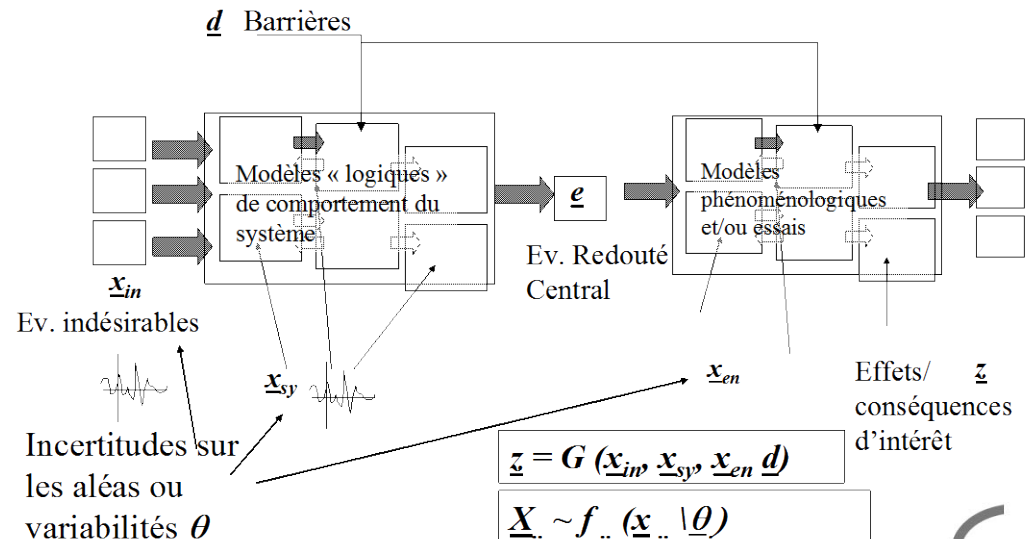
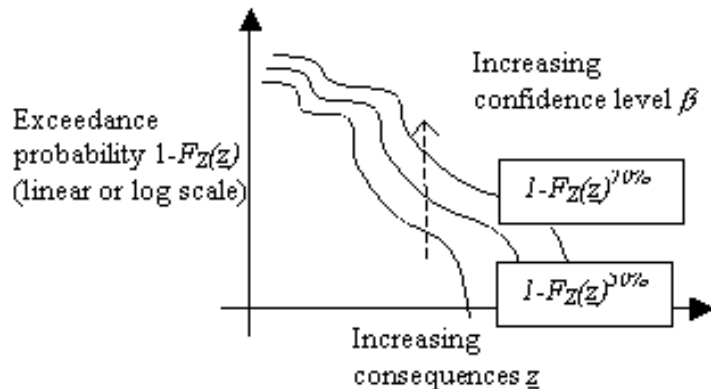
Modèles intégrés IPRA (Integrated Probabilistic Risk Assessment)

- Couplant physique probabiliste, agressions externes, fiabilité et modèles phénoménologiques
- Nucléaire Niveau 2 – EPS
- Etude intégrée d'impact environnemental

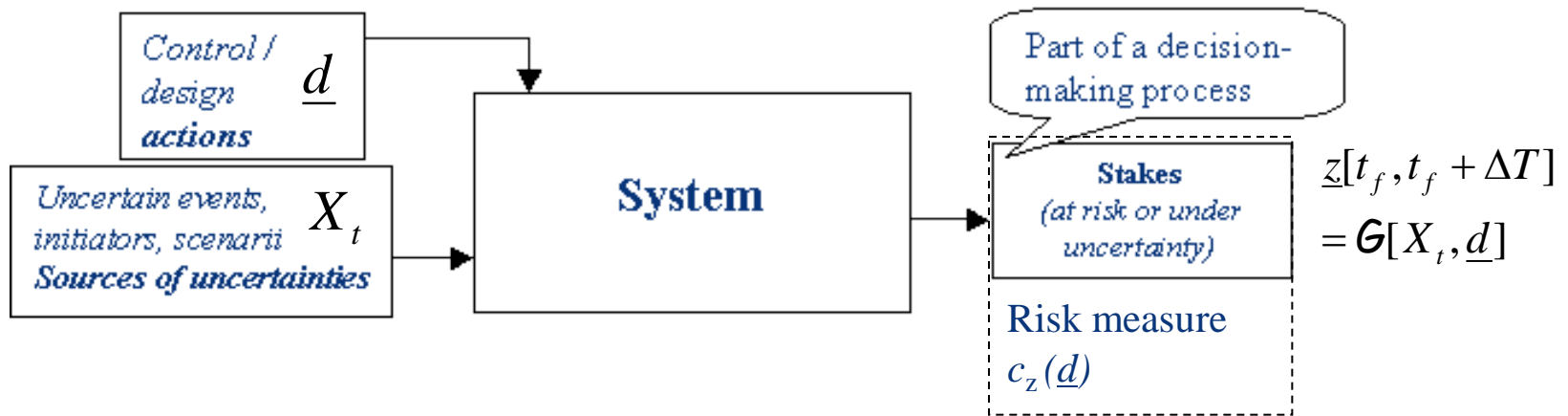
Helton, 1994



Source: ARAMIS / INERIS



Les étapes-clé d'un modèle de risque / d'incertitude



- *Théorie de la décision - Spécifier la mesure de risque*
- *Modélisation - Estimation/sélection des modèles $G(\cdot)$ et $f_{\underline{X}}(\underline{X}_t)$*
- *Calcul - de c_z et analyse de sensibilité $s_{x_i}(c_z)$*
- *Optimisation - de c_z via \underline{d}*

Le « budget » de risque rare

$$c_Z(\underline{d}) = P\left[\text{Max}_{[t_f, t_f + \Delta T]} Z_t \geq z_s \mid \underline{d}\right] \leq 10^{-r} = c_s$$

A quel point $c_s = 10^{-r}$ sont-ils rares ?

10^{-r}	Domaine
$5 \cdot 10^{-2}$	Nucléaire / APRP, défaillance conditionnelle
10^{-2}	Protection tempête du réseau électrique
$10^{-2} - 10^{-3}$	Civil / protection crue d'une ville
$2 \cdot 10^{-3}$	Civil / protection sismique de bâtiments normaux
$10^{-2} - 10^{-4}$	Nucléaire / rupture conditionnelle de la cuve
10^{-4}	Rupture de grand barrage
10^{-5}	Nucléaire / endommagement du cœur
$10^{-8} - 10^{-9}$	Aérospatial / fiabilité en vol

Des raretés comparables ?

10^r	Domaine	Détails
$5 \cdot 10^{-2}$	Nucléaire / défaillance conditionnelle APRP	(γ) – conditional probability on a rare initiator $< 10^{-4}$
$10^{-2} - 10^{-3}$	Civil / protection d'une ville contre les crues	(p_f) – probability over a <i>year</i>
10^{-5}	Nucléaire / endommagement du cœur	(λ) – overall cumulated <i>yearly</i> frequency
$10^{-8} - 10^{-9}$	Aérospatial / fiabilité en vol	(λ) – overall cumulated <i>hourly</i> frequency

Comment apprécier la rareté ?

Rareté – deux aspects

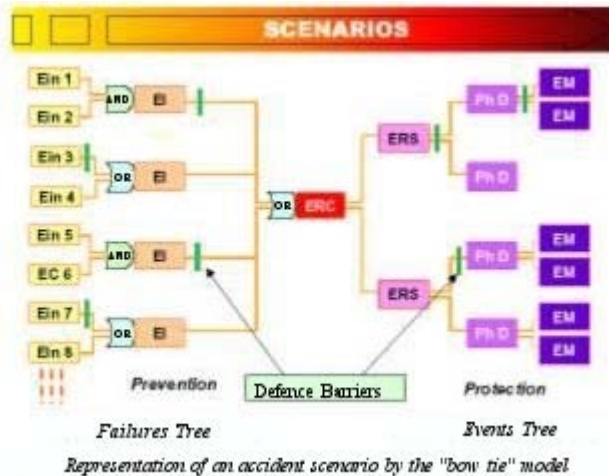
- *Vis-à-vis des données (incertitudes d'estimation)*
- *Vis-à-vis des calculs (incertitudes/erreurs de propagation)*

$$R.I. = \text{Log} \frac{1}{c_s} - \text{Log} N$$

N = nb de d.d.I (i.i.d. ou équivalents-independents)

Rarity index	Domaine	
$R.I. \leq -2$	Fréquent	Quasi-asymptotique
$-2 \leq R.I. \leq -1$	Moderément rare	Domaine des méthodes robustes
$-1 \leq R.I. \leq 0$	Rare	
$0 \leq R.I. \leq 1$	Très rare	Contrôle dégradé ou absent
$1 \leq R.I.$	Extrême	

Indices composés



Source : ARAMIS / INERIS

$$k = 1, \dots, n_k$$

$$E^{j_i(k+1)} = \bigcup_{i(k)} \left[E^{i(k)} \bigcap_{j_i(k)} E^{j_i(k)} \right]$$

$$e^{j_i(k+1)} = \sum_{i(k)} \left[e^{i(k)} \prod_{j_i(k)} e^{j_i(k)} \right]$$

...

$$e^f = \underline{e}^{j_i(n_k)} = G(\underline{e}^{i(k)}, \underline{e}^{j_i(0)})$$

Décomposer R.I. pour l'estimation / le calcul des évènements élémentaires

$$R.I.(j_i(k)) \ll R.I.(c_s)$$

Plan

- ◎ Motivations – quelle rareté ?
- ◎ Calcul d'un risque rare (mesure de risque)
- ◎ Estimation des paramètres incertains (facteurs de risque)
- ◎ Calcul du couple risque / incertitude et des sensibilités
- ◎ Modélisation et réglementation – a-t-on le choix ?

Calcul de la mesure du risque

$$\left\{ \begin{array}{l} c_Z(\underline{d}) = P\left[\text{Max}_{[t_f, t_f + \Delta T]} Z_t \geq z_s \mid \underline{d}\right] \\ = \int \mathbf{1}_{\text{Max}_t G(X_t, \underline{d}) \geq z_s} d(F_X) \\ c_Z(\underline{d}) = E\left[\sum_{T_k} C(Z_{T_k}, T_k) \mid (T_k)_k = \{t \in [t_f, t_f + \Delta T] \mid Z_t \geq z_s\}\right] \\ = \int C(X_t, \underline{d}) d(F_X) \end{array} \right.$$

$$\text{Estimer } c_Z(\underline{d}) = \mathcal{F}[Z_t, \underline{d}] = \int Q(X_t, \underline{d}) d(F_X)$$

Monte-Carlo asymptotique

$$q = \int_{\underline{w} \in \Omega} Q(\underline{w}) d(F_{\underline{W}})$$

$$\hat{q}_N = \frac{1}{N} \sum_{j=1..N} Q(\underline{w}_j)$$

varQ < +∞ mais Q éventuellement non-différentiable, non-continue, ...

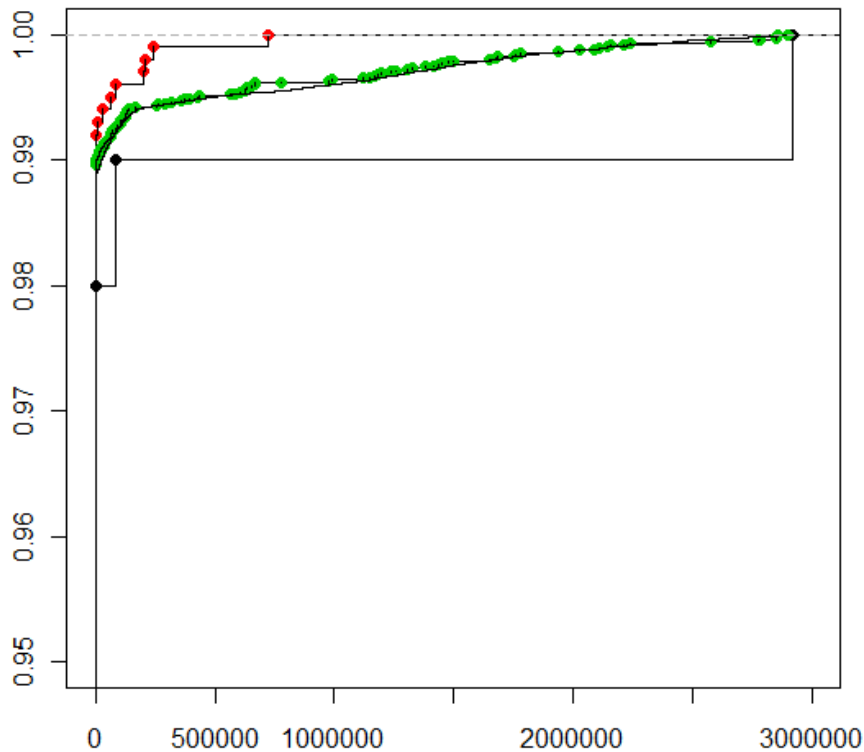
$$(1) \quad \hat{q}_N \xrightarrow{a.s.} q$$

$$(2) \quad \sqrt{N}(\hat{q}_N - q) \xrightarrow{L} \mathbf{N}(0, \text{Var}(Q(\underline{W})))$$

$$\hat{s}_Q^2 = \frac{1}{N(N-1)} \sum_{j=1..N} [Q(\underline{w}_j) - \hat{q}_N]^2 \xrightarrow{a.s.} \text{Var}(Q(\underline{W}))$$

Monte-Carlo asymptotique (2)

Modèle typique « high-loss low-probability » : $\text{var } Q(\cdot)$ est grande



	q (en M€/yr)				
N	10^2	10^3	10^4	10^5	10^6
q_N	18 ± 192%	11 ± 82%	9.8 ± 38%	8.8 ± 12%	9.1 ± 3%
s_N	176 ± 14%	140 ± 4%	120 ± 1%	121 ± 0.4%	126 ± 0.1%
$P(q > 0)$	2%	1.2%	1.4%	1.1%	1.2%
$R.I.$	~ 0	-1	-2	-3	-4

Queue de distribution de q = coût total de dommages (modèle complexe physique/économique)

$\text{VaR}_{98\%} = 0$; $\text{ES}_{98\%} = 450 \text{ M€}$; $C_V = 1380\%$

Monte-Carlo asymptotique (3)

Une propriété restant vraie si $\text{var } G(.) \sim +\infty$ (high-loss low-probability)

$$Q(\underline{w}) = 1_{G(\underline{w}, \underline{d}) > z_s}$$

$$Cv_N = \frac{\sqrt{\text{Var}(\hat{q}_N)}}{E(\hat{q}_N)} = \frac{\sqrt{\frac{1}{N} p_f (1 - p_f)}}{p_f} \approx \sqrt{\frac{1}{N p_f}}$$

$$P \left\{ q \in \left[\hat{q}_N \pm \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{\frac{\hat{q}_N (1 - \hat{q}_N)}{N}} \right] \right\} \longrightarrow \alpha$$

Noter que R.I. = $-2 \text{Log}(Cv_N)$

R.I. < -2 en général pour une convergence acceptable

Monte-Carlo non-asymptotics

Une propriété non-asymptotique property (applicable à var $G(\cdot) \sim +\infty$)

$$\hat{F}_{Q_N}(z) = \frac{1}{N} \sum_{j=1..N} 1_{G(\underline{W}_j, \underline{d}) > z}$$

$$P\left(\hat{F}_{Q_N}(z) = \frac{k}{N}\right) = C_N^k F(z)^k (1 - F(z))^{N-k}$$

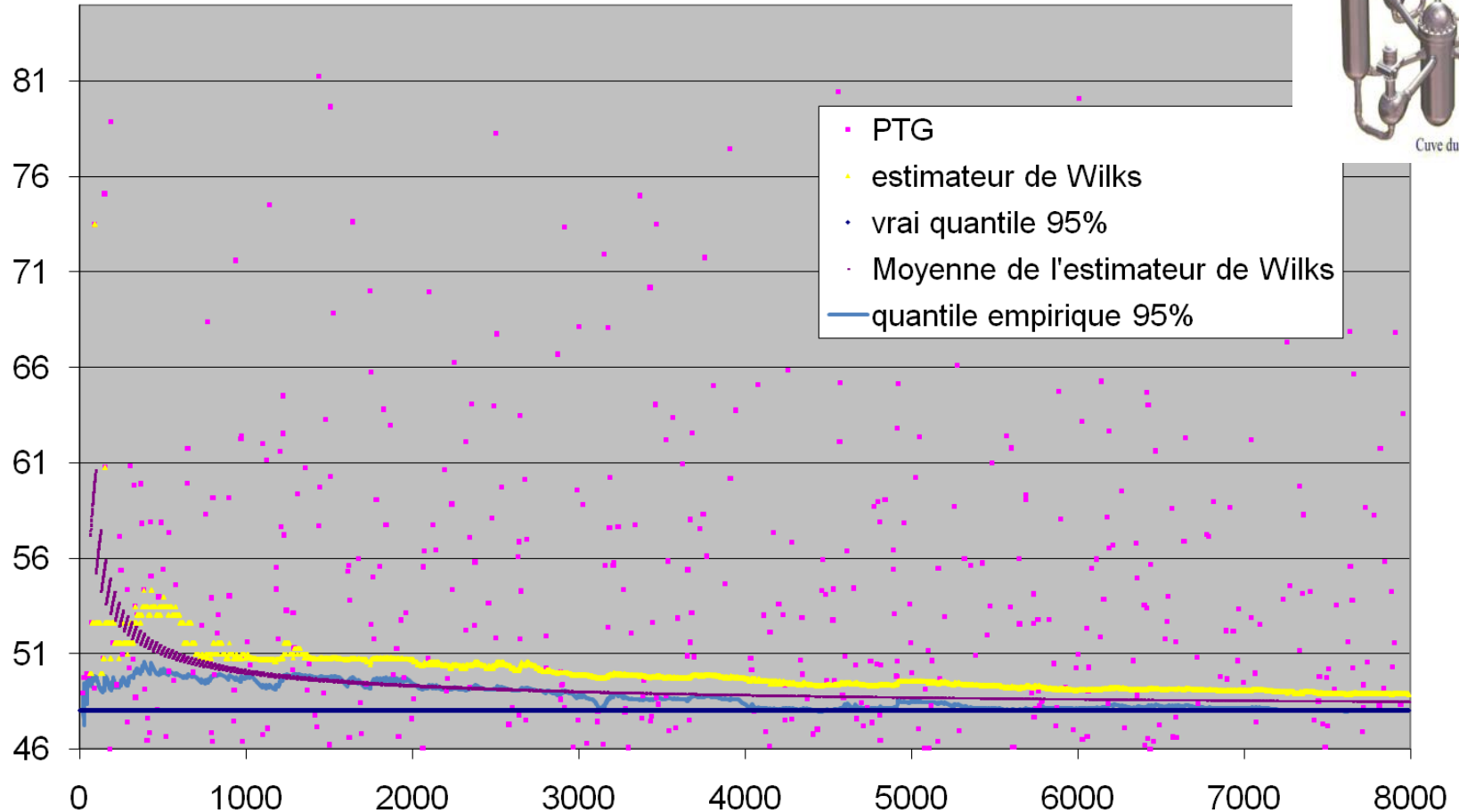
D'où un estimateur β -conservatif

$$N \geq N_m(\alpha, \beta) = \text{Min} \left\{ N \left| \sum_{k=N-m+1}^N C_N^k \alpha^k (1 - \alpha)^{N-k} < 1 - \beta \right. \right\}$$

$$\Rightarrow P(z^\alpha < Z^{N-m+1}) > \beta$$

Adapté à $R.I. \in [-1, -0.5]$ et beaucoup plus robuste que MC asymptotique

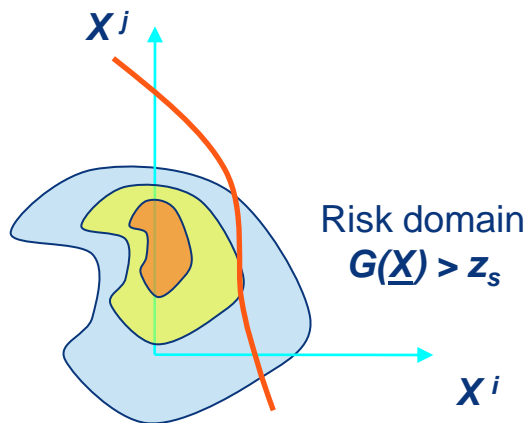
Monte-Carlo non-asymptotique



Propriété physique de pic (HPC: 25-an équivalent de CPU)

R.I. < - 1 pour N > 200

Alternatives à Monte-Carlo



Calculer efficacement le volume pondéré du domaine risqué

F(S)ORM > descente + méta-modèle hyperplan/quadratique après transformation isoprobabiliste

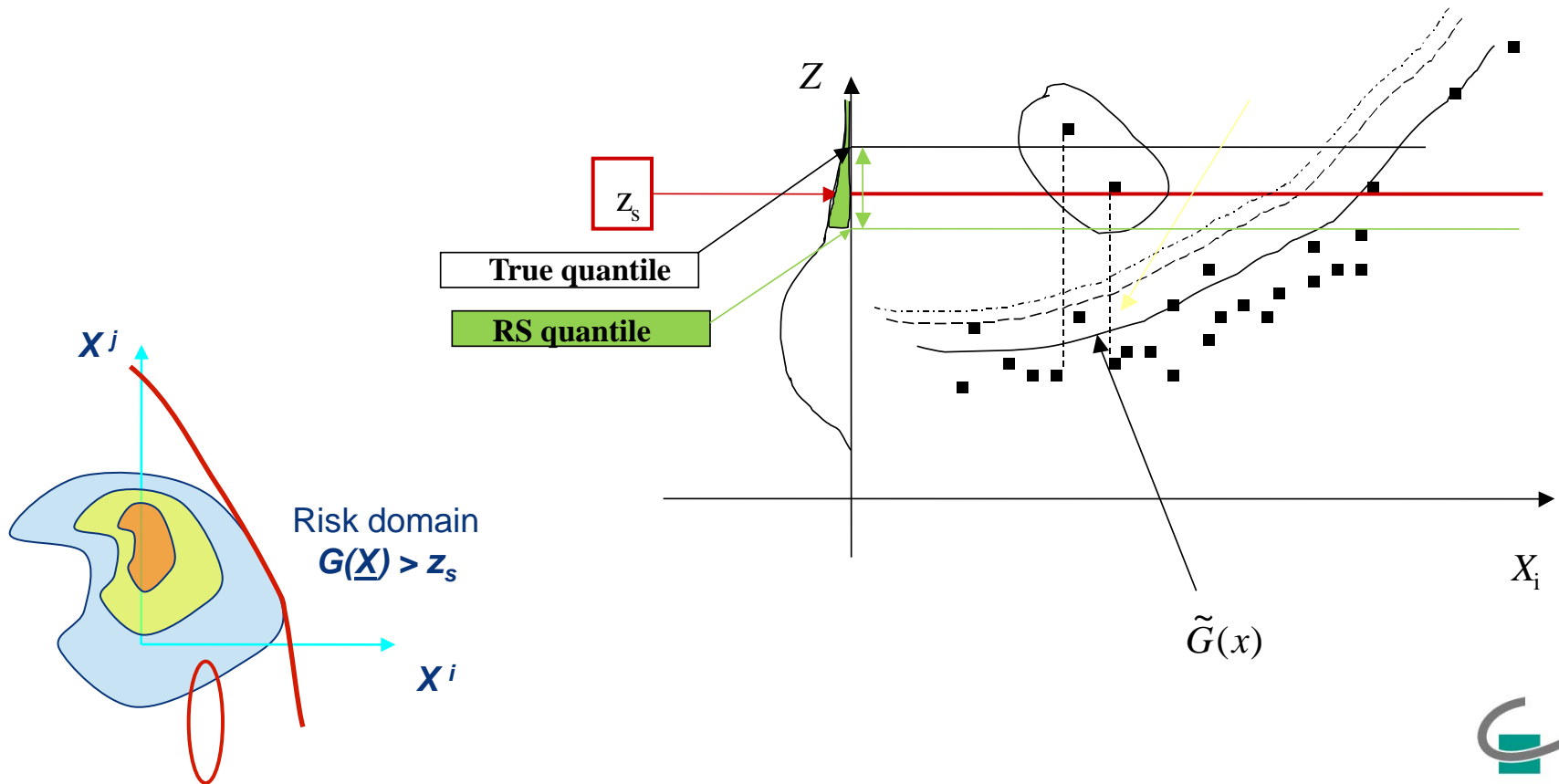
(pas de limite en R.I., mais sans contrôle d'erreur)

IS > biaiser les tirages pour intensifier dans le volume (inconnu) – espace transformé Gaussien-cartésien–directionnel; par apprentissage via FORM, MM ou processus gaussiens

(R.I. $\in [-1,3]$ sur des exemples publiés, - dépend de la forme du domaine)

Un résidu risqué ?

$$\underline{Z} = G(\underline{X}, \underline{d}) = \sum_{l=1..P} a^l(\underline{d}) \cdot \psi^l(\underline{X}) = \sum_{l=1..P} a^l(\underline{d}) \cdot \psi^l(\underline{X}) + u(\underline{X}, \underline{d})$$



Alternatives

Nombreux travaux >>> GdR Mascot-NUM

Quelques-uns visant un contrôle non-asymptotique en modifiant le tirage

➤ *Tirage du vrai modèle biaisé par méta-modèle* Cannamela C, Garnier J and Iooss B. (2008)

➤ *(control variate): MM pondère MCS*

➤ *(controlled stratification): MM constitue un filtre d'acceptation/rejet de MCS*

➤ *(controlled importance sampling): MM façonne la densité d'importance*

➤ *Tirage adaptatif à base d'importance directionnelle* Munoz Zuniga, et al., (2009)

➤ *MRM: bornes robustes monotones et accélération du tirage*

de Rocquigny (2009) ; Limbourg et al (2010)

L'approche des bornes monotones robustes

Monotonie des modèles:

- *Intuitif: les facteurs de risques dominants ont un impact monotone (par morceaux) sur l'évènement extrême*
- *Généralisable par l'approche conservative de l'enveloppe monotone*
- *Présent dans la réglementation effective (mécanique de la rupture, agressions externes etc.)*

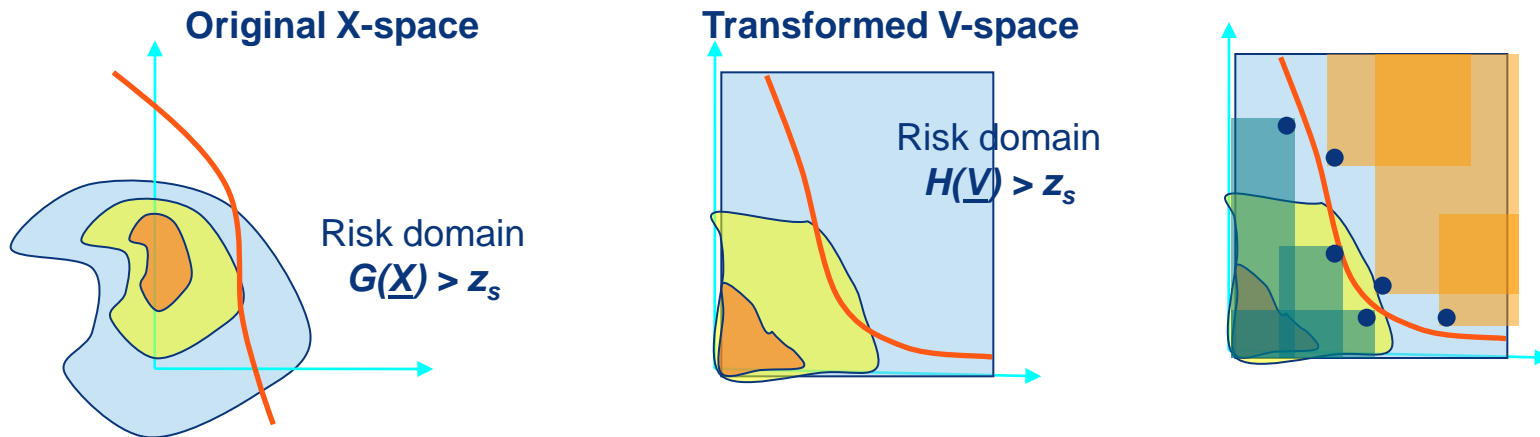
$$G(\underline{x}, \underline{d}) = g_R(x_k^1 / \gamma_R^1, \dots, x_k^r / \gamma_R^r) - g_S(x_k^{r+1} \gamma_S^{r+1}, \dots, x_k^p \gamma_S^p)$$

$$C(\underline{d}) = \{G(\underline{x}, \underline{d}) > 0\}$$

L'approche MRM

Sous monotonie: un intervalle 100%-certain sans appel additionnel de G(.)

$$\int_{\bigcup_{\underline{x}_j \in \Xi_n^f} \{\underline{x} \geq \underline{x}_j\}} f_x(\underline{x}) d\underline{x} = P(E_n^+) \leq \int_{\underline{x}} \mathbf{1}_{G(\underline{x}, \underline{d}) \geq z_s} f_x(\underline{x}) d\underline{x} \leq 1 - P(E_n^-) = 1 - \int_{\bigcup_{\underline{x}_j \in \Xi_n^s} \{\underline{x} \leq \underline{x}_j\}} f_x(\underline{x}) d\underline{x}$$

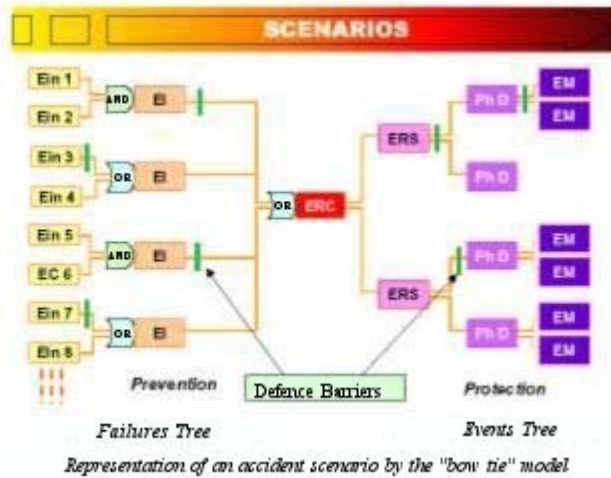


Un estimateur adaptatif très robuste Monte-Carlo

$$\hat{P}_f^{mis} = \frac{1}{n} \sum_{j=1..n} \left[P(E_{j-1}^+) + \mathbf{1}_{G(\underline{x}_j) \leq 0} \left[1 - P(E_{j-1}^+) - P(E_{j-1}^-) \right] \right]$$

Cas publiés R.I.>2 – fort fléau de la dimension en dimX

Indices composés et CPU



Source : ARAMIS / INERIS

$$k = 1, \dots, n_k$$

$$E^{j_i(k+1)} = \bigcup_{i(k)} \left[E^{i(k)} \cap E^{j_i(k)} \right]$$

$$e^{j_i(k+1)} = \sum_{i(k)} \left[e^{i(k)} \prod_{j_i(k)} e^{j_i(k)} \right]$$

...

$$e^f = \underline{e}^{j_i(n_k)} = G(\underline{e}^{i(k)}, \underline{e}^{j_i(0)})$$

Pour un budget CPU, optimiser la sous-structuration conditionnelle pour réduire les appels MCS et/ou le coût de chaque appel

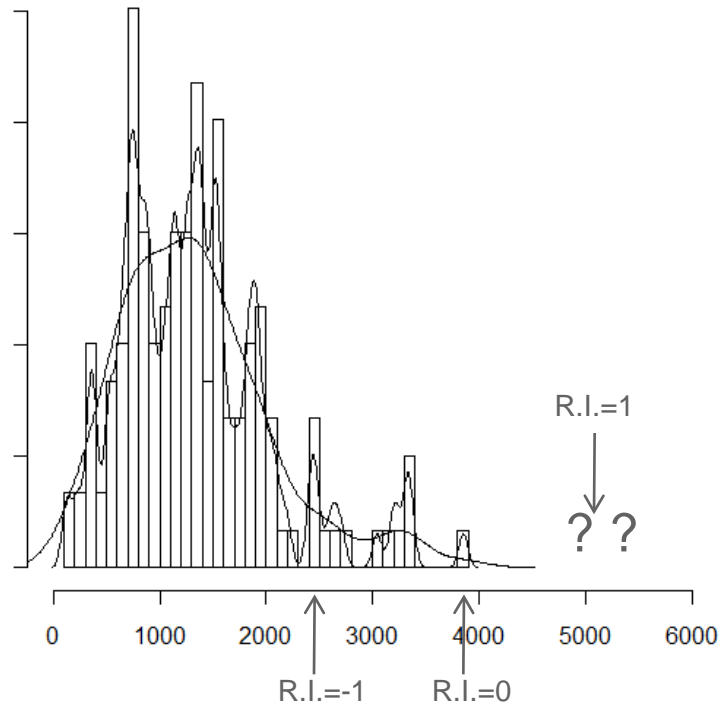
Scalabilité rapide si le risque est « sous-conditionnel »

$$R.I.(j_i(k)) \approx \alpha \frac{R.I.(c_s)}{n_k}$$

Plan

- ◎ Motivations – quelle rareté ?
- ◎ Calcul d'un risque rare (mesure de risque)
- ◎ Estimation des distributions d'entrées (facteurs de risque)
- ◎ Calcul du couple risque / incertitude et des sensibilités
- ◎ Modélisation et réglementation – a-t-on le choix ?

Estimation des queues de distribution



- ⊙ Estimateurs aux noyaux non robustes pour $R.I. > -0.5$
- ⊙ Modèle paramétrique nécessaire

L'approche de la théorie des valeurs extrêmes

- Les distributions max-stables, *nécessairement* de la forme suivante

$$GEV(x|\mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}$$

- Sont les seules limites non-dégénérées d'une maximisation (Fischer-Tippett, 1928 – Gnedenko 1943 – Gumbel 1958 ...)

$$F_n(z) = P[Z = \text{Max}_{i=1..n} Z_i \leq z] = F_Z(z|\theta)^n$$

$$P[\text{Max}_{i=1..n} Z_i \leq \sigma_n x + \mu_n] \xrightarrow{L} GEV(x)$$

- NB : pour Fréchet, k-moments infinis pour $\xi \geq 1/k$*
- Applicabilité à l'estimation réelle (non-asymptotique, etc.) ?

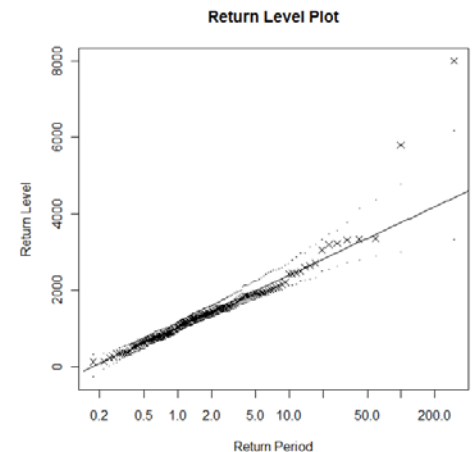
Incertitudes d'estimation

- Maximum de vraisemblance - as.G. avec incertitude d'estimation calculable sur les quantiles

$$\hat{\underline{\theta}}_X = \underset{\underline{\theta}}{\text{Arg max}} \left(LL[\underline{\Xi}_n = (x_j)_{j=1..n}, \underline{\theta}_X] \right) \quad \text{Var} \hat{\underline{\theta}}_X \approx \left[-\frac{\partial^2}{\partial \underline{\theta} \partial \underline{\theta}'} LL[\underline{\Xi}_n, \underline{\theta}_X] \Big|_{\hat{\underline{\theta}}_X} \right]^{-1}$$

- D'autres techniques pour des i.c. robustes sur les quantiles: bootstrap paramétrique, stat. d'ordres, bayésienne*

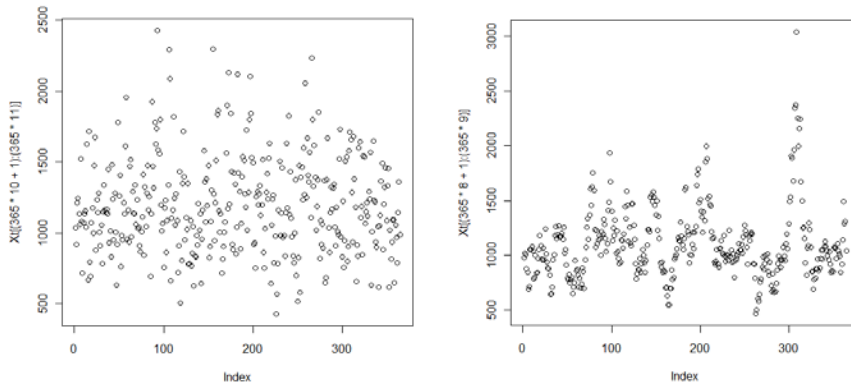
$$\pi_1(\underline{\theta}_X | \underline{\Xi}_n) = \frac{L[\underline{\Xi}_n, \underline{\theta}_X] \pi_0(\underline{\theta}_X | \zeta)}{\int_{\underline{\theta}_X} L(\underline{\Xi}_n | \underline{\theta}_X) \pi_0(\underline{\theta}_X | \zeta) d\underline{\theta}_X}$$



- Couverture de toute l'incertitude épistémique ?

Convergence suffisamment rapide ?

- ⊙ Gumbel ($\xi=0$) est le domaine d'attraction de processus courants : normal + ses transformations monotones (logn, ...), exp, ..
- ⊙ Des taux de convergence anciens dans le cas iid (e.g. Dziubdziela, 1978 etc.) + étendus à la dépendance faible, ... mais la convergence peut être très lente pour les échantillons réels



Auto-correlated flow-like lognormal time series: daily sampling over 1 year (left – $\tau=1dy$, right – $\tau=10dy$) –
 $\log X_{t+1} = \phi \cdot \log X_t + \mu(1-\phi) + U_t$

	M=10	M=365	Reference
$\tau=0$	4400 G 3600 LN	4330 G 3930 LN	4300
$\tau=10$	4630 G 4070 LN	4070 G 3600 LN	4000
$\tau=100$	4220 G 3890 LN	4040 G 3660 LN	3800

1/1000 estimated of Gumbel vs Lognormal estimated on Block Maxima of increasing size M over 150yrs of data

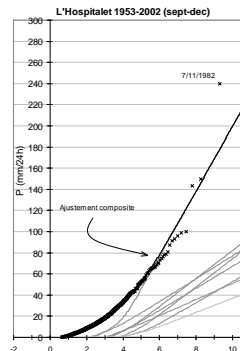
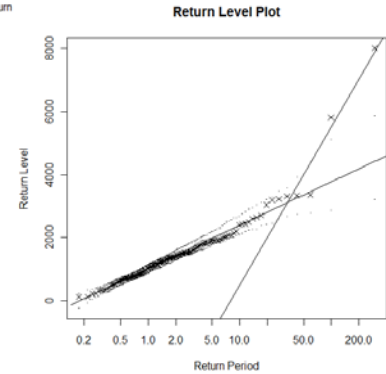
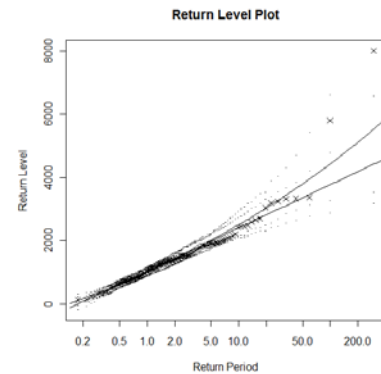
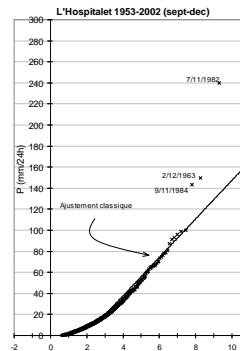
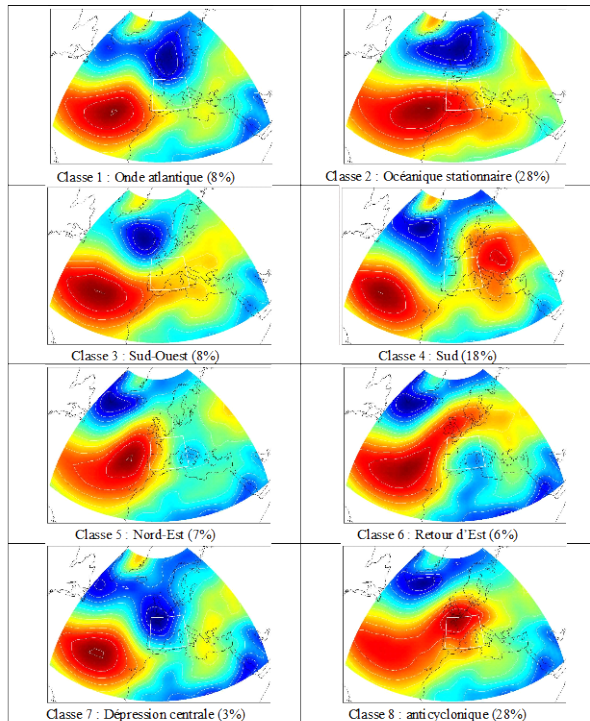
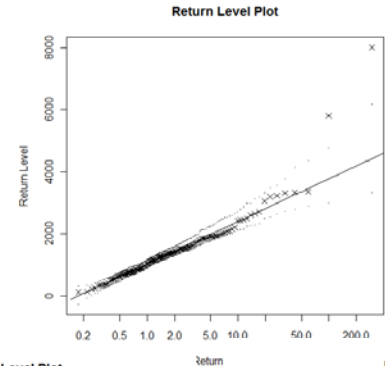
Outliers et incertitudes des distributions

○ Affiner le modèle :

- Paramétrique $F_Z = GEV(z|\mu, \sigma, \xi)$

- Mélange

$$F_Z = \sum_k \alpha_k GEV(z|\mu_k, \sigma_k, 0)$$



NB : raffiner la paramétrisation accroît l'inc épistémique ... sauf si l'on rajoute de l'information physique

Fléau de la multi-dimensionalité

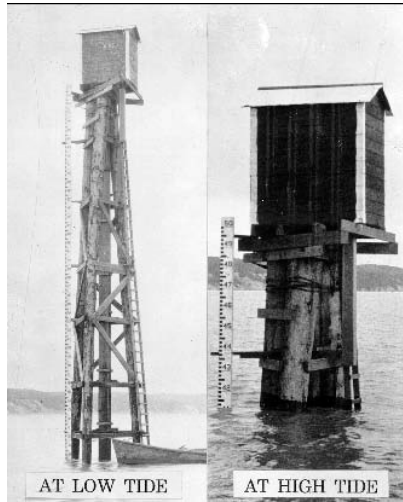
- ⊙ Modèle de copule et dépendance de queue

$$F_X(\underline{x}|\underline{\theta}_X = (\underline{\theta}_m, \underline{\theta}_c)) = P[X^1 < x^1, \dots, X^p < x^p | \underline{\theta}_X = (\underline{\theta}_m, \underline{\theta}_c)] \\ = C[F_{X^1}(x^1 | \underline{\theta}_{m^1}), \dots, F_{X^p}(x^p | \underline{\theta}_{m^p}), \underline{\theta}_c]$$

$$\lambda_U = \lim_{v \rightarrow 1} \left[P(X^1 > F_{X^1}^{-1}(v) | X^2 > F_{X^2}^{-1}(v)) \right]$$

- ⊙ Des espaces fonctionnels très grands ; des échantillons multi-variés très rares
- ⊙ Alternative : modélisation du sous-jacent physique $\underline{X} = h(\underline{W})$
 - Forçage océanologique : description jointe des surcotes, houles et vents.
 - Modèle géophysique sismique spatial + tsunami

Niveaux marins extrêmes



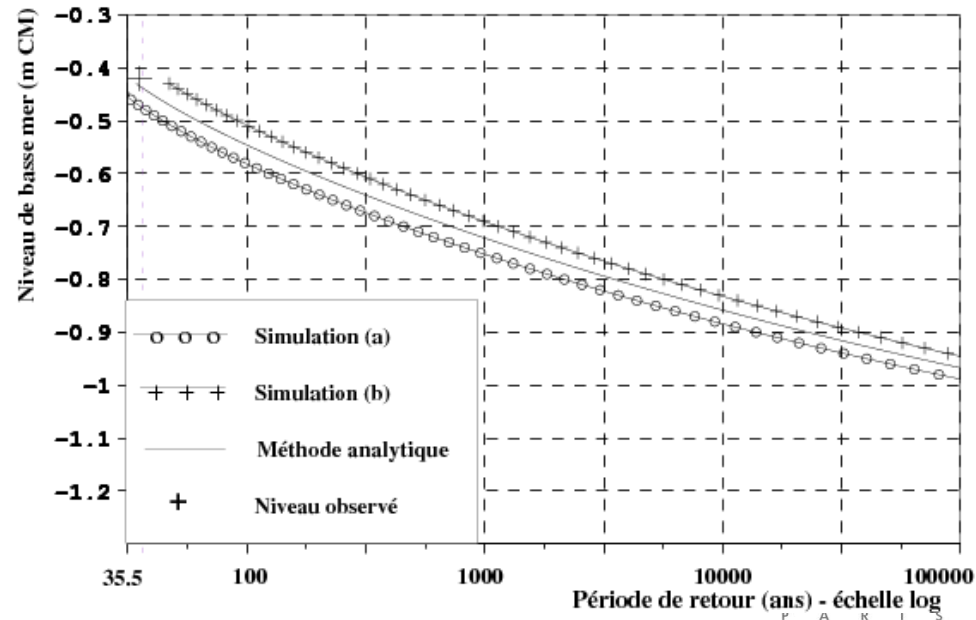
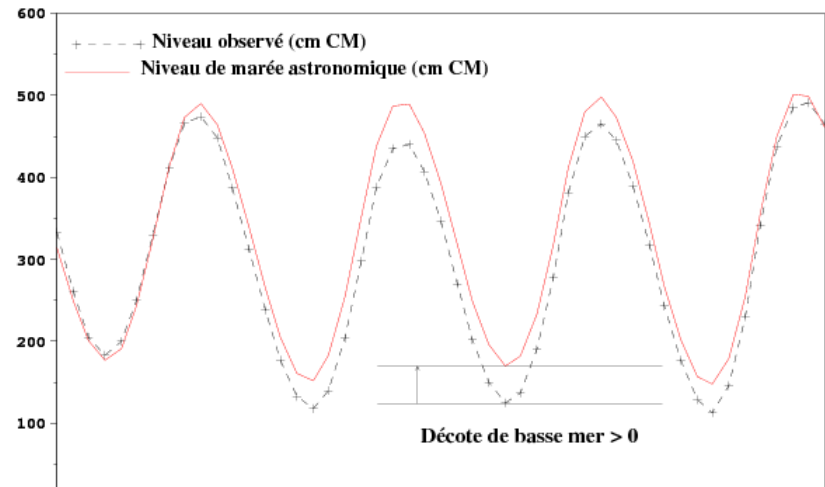
$$Z_t = x(t) - Y_t$$

Z_t = niveau combiné de la mer

$x(t)$ = marée astronomique

Y_t = décote à basse mer

Lafon & de Rocquigny, 2006



Estimation généralisée – inverse probabiliste

$$\underline{Y}_{mj} = H(\underline{X}_j, \underline{d}_j) + \underline{U}_j$$

$$\underline{X}_j | \underline{\theta}_x \sim f_X(\cdot | \underline{\theta}_x) \quad \underline{U}_j | \underline{\theta}_u \sim f_U(\underline{u} | \underline{\theta}_u) \quad \underline{\theta} = (\underline{\theta}_x, \underline{\theta}_u) \sim \pi(\underline{\theta} | \zeta)$$

identify $\{f_X(\cdot | \underline{\theta}_x) f_U(\cdot | \underline{\theta}_u) \pi(\cdot | \zeta)\}$ knowing $\Xi_n = \left(\underline{y}_{mj}, \underline{d}_j \right)_{j=1..n}$

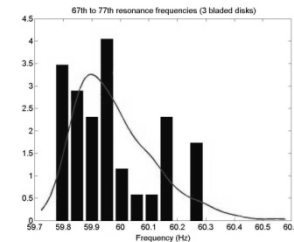
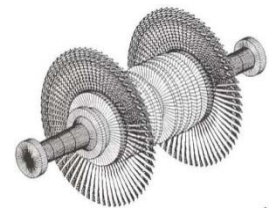
- Identifiabilité

$$(\mathbf{Y}, \mathbf{H}) = \left(\underline{Y}_{mj}, \nabla_{(\underline{x}_j, \underline{u}_j)} H \Big|_{d=d_j} \right)_j$$

- EMV Non-iid

$$\hat{\underline{\theta}} = \text{Arg max} \prod_j E_{\underline{X}} \left[f_{\underline{U}_j}(\underline{y}_{mj} - H(\underline{X}, \underline{d}_j) | \underline{\theta}_{uj}) | \underline{\theta}_x \right]$$

- Fort coût CPU en général (produit de n intégrales p -dim.)

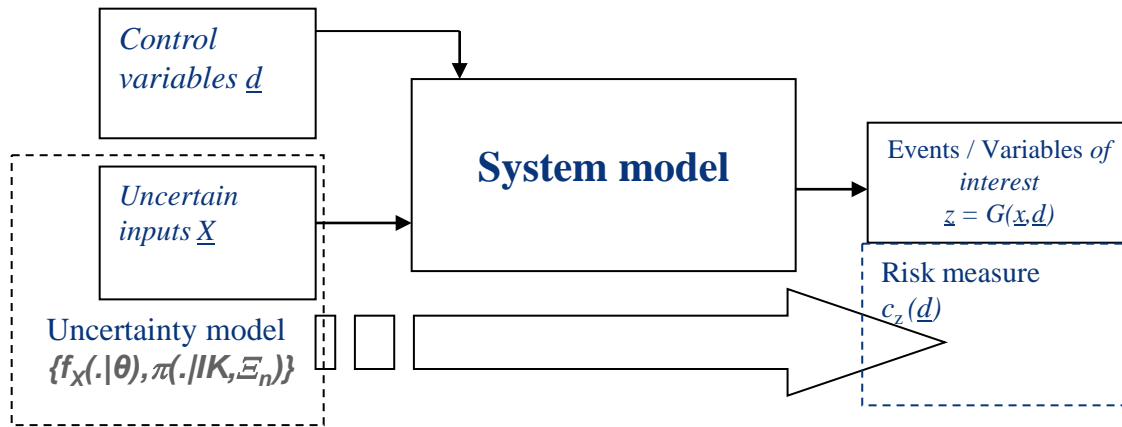


de Rocquigny & Cambier 2009
Bernardara et al, 2010
Barbillon et al, 2011

Plan

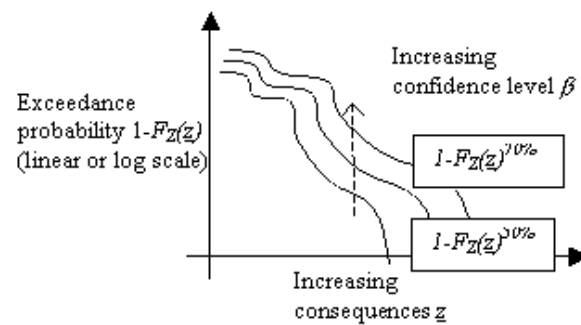
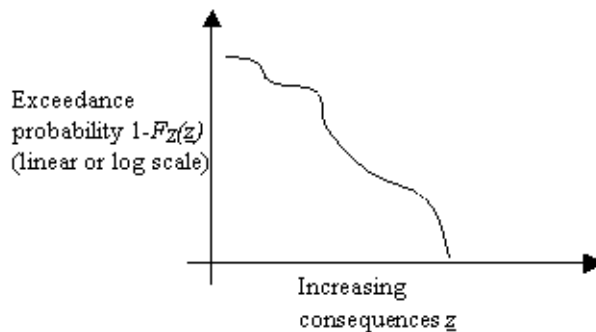
- ◎ Motivations – quelle rareté ?
- ◎ Calcul d'un risque rare (mesure de risque)
- ◎ Estimation des paramètres incertains (facteurs de risque)
- ◎ Calcul du couple risque / incertitude et des sensibilités
- ◎ Modélisation et réglementation – a-t-on le choix ?

Calcul du couple risque / incertitude



$$(\underline{X}, \underline{\Theta}_X) \sim f_X(\cdot | \underline{\Theta}_X) \cdot \pi$$

$$(\underline{Z}, \underline{\Theta}_X) \sim f_Z(\cdot | \underline{\Theta}_X) \cdot \pi$$



Calcul du couple risque/incertitude (2)

- ◉ *Mesure fondamentale – la ccdf combinée aleatoire/épistémique α - β*

$$c_z = P_{\Theta} [P_{\underline{X}|\underline{\theta}_X} (Z > z_s | \underline{\Theta}_X) < \alpha]$$

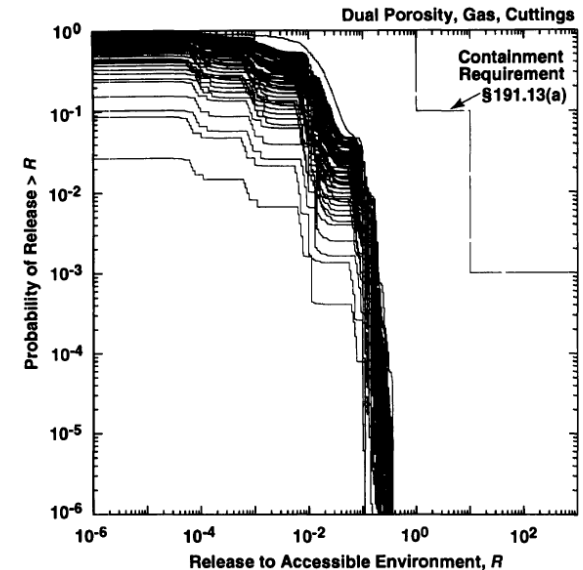
$$= \int_{\theta} \mathbf{1}_{\left[\int_{\underline{x}} \mathbf{1}_{G(\cdot) > z_s} d[F_{\underline{X}}(\cdot | \underline{\Theta}_X)] \right] < \alpha} \pi_1(\underline{\theta}_X | \underline{\zeta}) d\underline{\theta}_X$$

- ◉ *Robustesse nécessaire pour décrire le niveau 2 (R.I. < -2 + Log[k.(1- β)] i.e. R.I. < -3 au mieux)*

$$c_z = \frac{1}{N_1} \sum_{j=1..N_1} Q_e \left[\frac{1}{N_2} \sum_{k=1..N_2} [Q_a(\underline{X}_{j,k})] \right]$$

- ◉ *Alternative dédiées à MC :*

- ◉ *MRM niveau 2 : cas publiés à R.I. = -0.5*



Helton, 1994

Limbourg et al, 2010

Analyse de sensibilité - risque

⊙ Questions typiques

- « Comment réduire le risque ? »
- « Comment réduire l'incertitude sur le niveau de risque ? »

⊙ Modification des méthodes classiques de SA, dont les indices de Sobol:

$$S_i = \frac{\text{var}[E(G(\underline{X}, \underline{d})|X^i)]}{\text{var} G(\underline{X}, \underline{d})}$$

$$S_{Qi} = \frac{\text{var}[E(Q \circ G(\underline{X}, \underline{d})|X^i)]}{\text{var} Q \circ G(\underline{X}, \underline{d})} \quad e.g. \quad Q = 1_{G(\underline{X}, \underline{d}) > z_s} \cdot C(\underline{X}, \underline{d})$$

⊙ Un calcul robuste: $N=a^p$ pour le cas de la variance standard ... i.e. *R.I.* < -3 to -7 pour la SA combinée risque/incertitude.

⊙ Des indices dédiés pour aller au-delà des indices FORM

Munoz-Zuniga et al, 2011; Borgonovo et al, subm. 2011

Plan

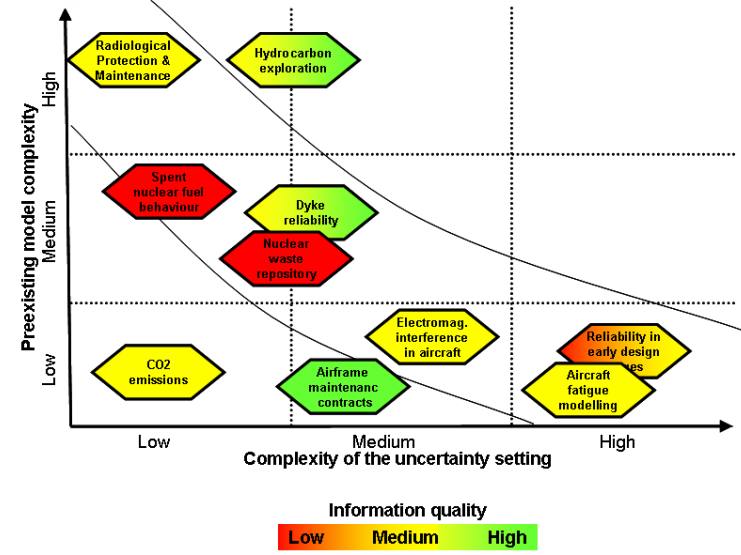
- ◎ Motivations – quelle rareté ?
- ◎ Calcul d'un risque rare (mesure de risque)
- ◎ Estimation des paramètres incertains (facteurs de risque)
- ◎ Calcul du couple risque / incertitude et des sensibilités
- ◎ **Modélisation et réglementation – a-t-on le choix ?**

Résumé – évaluation trop risquée ?

- ⊙ Des limites sur les données à $RI > 1$ (courant pour les grands risques) voire $RI > -1$ en multi-D
 - Modélisation phénoménologique pour (i) populations mélangées sous-jacentes, (ii) décomposer le RI par sous-conditionnement (iii) facteurs de risque multi-dimensionnels
- ⊙ Des défis de calcul
 - Robustesse au-delà de $RI \sim -1$ à 0 pour le calcul standard
 - Robustesse au-delà de $RI \sim -3$ s'il faut du niveau 2 ou de la sensibilité
 - HPC: naturel

HPC challenges

- 3 levels of greediness to optimise
 - Underlying deterministic system model >> e.g. CFD
- A layer of probabilistic sampling
 - Fully distributed if standard Monte-Carlo, but more generally mixed sequential / parallel >> e.g. adaptive importance sampling
- A layer of optimisation
 - Because of *mixed deterministic-proba.* risk criteria
 - Because calibrating the uncertainty model requires *inverse techniques*
 - Because the final goal is to *optimise under risk*, not just *compute risk*



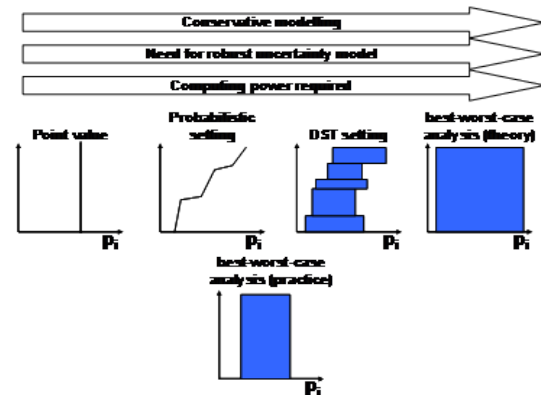
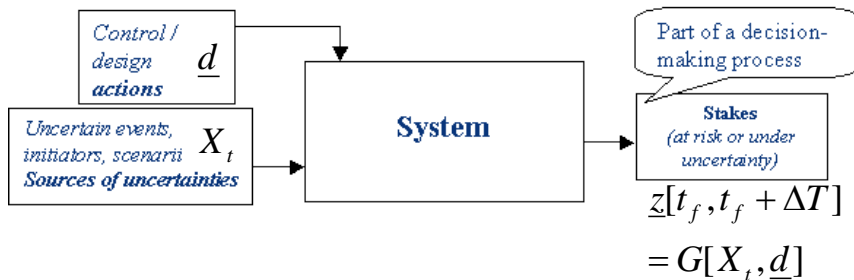
De Rocquigny et al, 2008



How do we efficiently allocate computing power in 3 such layers ?

A-t-on le choix ?

- Probabilités déraisonnables pour $RI > 0$? Quelles alternatives pour la *mesure de risque*?
- $F_X(X_t|\theta_X)$ peut être une mesure déterministe (un sous-ensemble de IR^p) ou une mesure DST (couple de plausibility/belief cdfs),



De Rocquigny et al, 2008

- Quels impacts sur les données et le calcul ?

A-t-on le choix (2)

- DempsterShaferTh relâche la sous-additivité

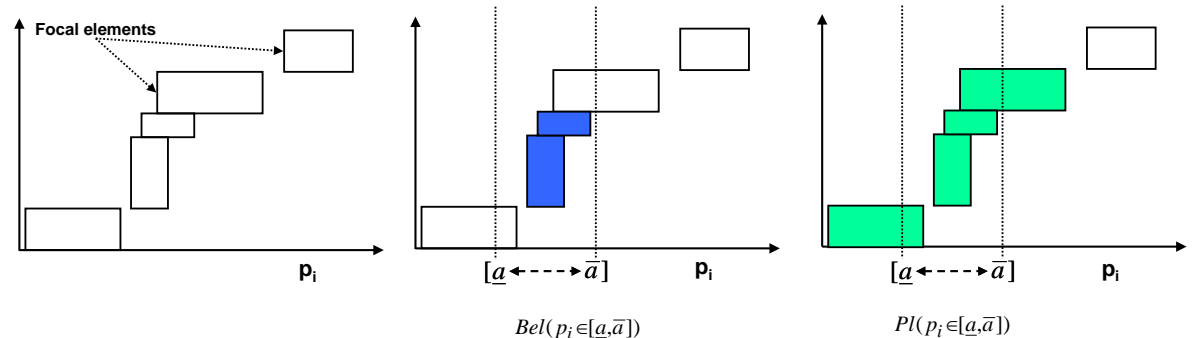
$$m : \wp(\Omega) \rightarrow [0,1]$$

$$m(\emptyset) = 0 \quad \sum_i m(A_i) = 1$$

$$Bel(x \in B) = \sum_{A_i \subseteq B} m(A_i) \quad Pl(x \in B) = \sum_{A_i \cap B \neq \emptyset} m(A_i)$$

$$Bel(B) + Bel({}^c B) \leq 1 \quad Pl(B) + Pl({}^c B) \geq 1$$

- Une interprétation possible : espace probabilisé incomplètement-spécifié / avec proba. supérieures et inférieures.
- Le manque de connaissance est plus finement modélisé – mais entraîne l'indécidabilité



De Rocquigny et al, 2008

« Le pire à l'exclusion du reste ? »

- ⊙ Approches déterministes
 - Un risque de sous-estimation des bornes absolues
 - Non-décidabilité : on perd l'ordre complet des options
 - Potentiellement inévitables pour les événements trop incertains
- ⊙ Modéliser le risque, c'est finalement comparer

$$C_Z : \{c_Z(\underline{d}_A) \leq c_Z(\underline{d}_B)\}$$
$$c_Z(\underline{d}) = EU(Z|\underline{d}) \quad \text{e.g.} \quad E1_{\text{Max}_{[t_1, t_1+\Delta T]} Z_t(\omega) \leq z_s}$$

- ⊙ Probabilités – éventuellement combinées avec des scénarios conditionnels déterministes – est inévitable quand il faut comparer la « plausibilité » et allouer des ressources

References

- © Helton J.C. (1994), Treatment of Uncertainty in Performance Assessments for Complex Systems, *Risk Analysis*, **14**, 483-511.
- © Saltelli, A., Tarantola, S., Campalongo, F., Ratto, M. (2004), *Sensitivity analysis in practice: a guide to assessing scientific models*, Wiley
- © Lafon F., de Rocquigny E. (2006), Niveaux marins extrêmes : contrôle d'une méthode analytique par simulation d'un processus couplé, 38èmes Journ. Franc. de Stat., Clamart.
- © Paquet E., Gailhard J., Garçon R. (2006), Evolution de la méthode du Gradex: approche par type de temps et modélisation hydrologique, *Colloque Société Hydrologique de France "Valeurs rares et extrêmes de débit "* – Lyon – Mars 2006
- © de Rocquigny E., Devictor N., Tarantola ed. (2008), *Uncertainty in industrial practice – A guide to Quantitative Uncertainty Management*, John Wiley & Sons
- © Cannamela C, Garnier J and Iooss B. (2008) Controlled stratification for quantile estimation. *Annals of Applied Statistics*, 2:1554-1580
- © de Rocquigny E. (2009), Structural reliability under Monotony : Properties of Form, simulation or response surface methods and a new class of monotonous reliability methods (MRM), *Structural Safety* vol. 31, pp 363-374.
- © de Rocquigny E., Cambier S. (2009) Inverse probabilistic modeling of the sources of uncertainty : a non-parametric simulated-likelihood method with application to an industrial turbine vibration assessment, *Inverse Problems in Science and Engineering* vol.17 (7), pp. 937-959.

References

- ⊙ Munoz-Muniga M., Garnier J., Remy E., de Rocquigny E. (2009), Adaptive Directional Stratification – An adaptive directional sampling method on a stratified space, 10th International Conference on Structural Safety and Reliability ICOSSAR, Osaka, 2009
- ⊙ Bernardara P., de Rocquigny E., Goutal N., Arnaud A., Passoni G. (2010) Flood risk and uncertainty analysis: joint assessment of the hydrological & hydraulic components, *Canadian Journal of Civil Engineering*, vol. 37(7), pp. 968-979
- ⊙ Limbourg P., de Rocquigny E. (2010) Uncertainty analysis using Evidence theory - confronting level-1 and level-2 approaches with data availability and computational constraints, *Reliability Engineering and System Safety* vol. 95 (5), pp. 550-564
- ⊙ Limbourg P., de Rocquigny E., Andrianov G. (2010), Accelerated uncertainty propagation in two-level probabilistic studies under monotony, *Reliability Engineering & System Safety* vol. 95 (2010) 998–1010
- ⊙ Barbillon P., Celeux G., Grimaud A., Lefebvre Y., de Rocquigny E. (2011) – Non linear methods for inverse statistical problems, *Computational Statistics and Data Analysis*, vol. 55(1), pp. 132-142
- ⊙ Munoz-Muniga M., Garnier J., Remy E., de Rocquigny E. (2011), Adaptive Directional Stratification for controlled estimation of the probability of a rare event, Accepted in *Reliability Engineering & System Safety*
- ⊙ Borgonovo E., Zentner I., Tarantola S., de Rocquigny E. (subm), Importance Measures for Seismic Probabilistic Fragility Assessment, Submitted to *Risk Analysis*