

ÉTIENNE DE ROCQUIGNY

La maîtrise des incertitudes dans un contexte industriel. 1re partie : une approche méthodologique globale basée sur des exemples

Journal de la société française de statistique, tome 147, n° 3 (2006), p. 33-71

http://www.numdam.org/item?id=JSFS_2006__147_3_33_0

© Société française de statistique, 2006, tous droits réservés.

L'accès aux archives de la revue « Journal de la société française de statistique » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

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

LA MAÎTRISE DES INCERTITUDES DANS UN CONTEXTE INDUSTRIEL

1^{re} PARTIE : UNE APPROCHE MÉTHODOLOGIQUE GLOBALE BASÉE SUR DES EXEMPLES

Etienne de ROCQUIGNY *

RÉSUMÉ

Un industriel appuyant son processus de décision sur des modèles quantitatifs (sûreté de fonctionnement, dimensionnement physique des installations ou procédés, optimisation économique, impact environnemental *etc.*) est vite confronté à une multitude d'incertitudes, imprécisions, erreurs et aléas de nature très variable affectant données et modèles. À travers quatre exemples introductifs qui résument plusieurs années de pratique industrielle dans différents métiers, on propose une approche générique des incertitudes dans un contexte industriel, en dépassant en particulier des variations terminologiques largement explicables par des compartimentages historiques entre les domaines impliqués (métrologie, fiabilité, statistique, analyse numérique, ...). Cette approche consiste notamment à formaliser de grandes étapes de traitement à partir des finalités principales liées à la nature de la réglementation et à la mise en œuvre d'un critère de décision (étape A) : la quantification des sources d'incertitudes (étape B), la validation inverse associée (B'), leur propagation à travers un modèle physico-industriel pré-existant (étape C), la hiérarchisation d'importance (C') voire l'optimisation qui en découle. La revue des méthodes de modélisation statistique, physique et numérique associées fait l'objet d'une seconde partie suivant cet article. L'approche se veut ouverte vis-à-vis de la pluralité des choix épistémologiques sous-jacents, notamment quant à la délicate question de la différenciation de la modélisation suivant la nature des incertitudes. Il s'agit de cadrer de manière cohérente et industrielle la modélisation mathématique *pratique* des incertitudes, en se restreignant bien entendu à celles de nature quantitative et quantifiable. Ce domaine fait récemment l'objet d'un fort intérêt de la part du monde industriel.

ABSTRACT

An industrial decision process supported by quantitative modelling (safety and reliability, physical design of facilities or processes, economic optimisation, environmental impact, *etc.*) quickly faces a wide diversity of uncertainties, imprecisions, errors or randomness affecting all data or models. Through four introductory examples that summarise years of industrial practice in different sectors, this paper evidences that a generic and applied approach to uncertainty in industry can surpass a terminological heterogeneity that is largely explicable by the historical separation of the fields

* Coordinateur du Réseau des projets/applications «Incertainitudes» d'EDF, Ingénieur senior EDF R&D MRI, 6 quai Watier, 78401 Chatou Cedex.
Courriel : etienne.derocquigny@edf.fr

involved (such as metrology, reliability, statistics, numerical analysis, ...). In relation with the applicable regulation and standards and a relevant decision criterion (step A), this approach involves in particular the proper identification of key steps such as : the quantification (or modelling) of the sources of uncertainty (step B), possibly involving an inverse approach (B'), their propagation through a pre-existing physical-industrial model (step C), the ranking of importance or sensitivity analysis (step C') and sometimes a subsequent optimisation step. A review of the corresponding statistical, physical and numerical methods is the subject of a second part following the present paper. This approach is intended to be amenable to a variety of underlying epistemological choices, including the tricky choice of differentiating the modelling according to nature of uncertainty. It aims at giving a consistent and industrially-realistic framework for *practical* mathematical modelling, assumingly restricted to quantitative and quantifiable uncertainty, a domain of considerable recent interest from the industrial sector.

1. Introduction – Enjeux industriels et problématique scientifique

Le caractère risqué des procédés, sites et activités industrielles est largement lié à l'omniprésence des incertitudes : intrants et conditions de fonctionnement incertaines dans les processus ; modèles physiques de conception ou d'optimisation imprécis ; aléas opérationnels ; variabilité environnementale imprévisible *etc.* Corrélativement, de nombreux enjeux conduisent l'industriel à tenter de maîtriser l'incertitude : sûreté de fonctionnement, responsabilité environnementale, optimisation de marché *etc.*

Dans ce contexte, EDF a été ces dernières années confronté à un nombre croissant de dossiers techniques traitant des « incertitudes » avec un apparent foisonnement de méthodes provenant de la métrologie, de la fiabilité, des probabilités et statistiques, de l'analyse numérique ... Il s'est néanmoins avéré possible peu à peu d'induire à partir d'exemples concrets dans différents domaines, de la réglementation et de la pratique industrielle associée, un cadre conceptuel commun : c'est le parti pris du présent article qui détaille et approfondit le « Tutoriel Incertitudes » présenté en octobre 2004 à la conférence $\lambda\mu 14$ ([16]).

Les notions de risque et d'incertitude sont étroitement liées, elles recouvrent de vastes acceptions selon les auteurs et leur appréhension théorique soulève d'importants débats au plan épistémologique, sociologique ou de la théorie de la décision. Le propos de l'article n'est pas de trancher sur ces délicates problématiques, mais de proposer un cadre cohérent et industriel de *modélisation mathématique* des incertitudes de nature *quantitative et qui sont effectivement quantifiables*. Plutôt que partir de concepts *a priori*, qui s'avèrent moins facilement discernables dans la pratique industrielle que dans un cadre idéalisé (*e.g.* des jeux de hasard), le texte suggérera qu'en fait, c'est surtout dans les choix d'usage de telle ou telle méthode et l'interprétation des résultats que le choix épistémologique réside. Par exemple la mise en œuvre, large-

ment débattue dans la littérature, de distinctions effectives entre incertitudes de nature «aléatoire» (irréductible, intrinsèque, variabilité ...) et de nature «épistémique» (réductible, liée à de la méconnaissance...) peut se faire de plusieurs façons dans le cadre d'une même approche méthodologique globale, selon la réglementation ou la pratique industrielle (cf. § 3.3).

À partir d'une revue détaillée de quatre exemples au § 2, l'approche méthodologique globale sera présentée au § 3 : Elle autorise *a priori* tout autant le choix d'un «paradigme» déterministe que probabiliste, possibiliste, flou. La seconde partie de l'article développera ensuite plus particulièrement les méthodes de modélisation statistique correspondant à un paradigme mixte déterministe/probabiliste qui joue un rôle central dans la pratique industrielle. Elle exposera quelques questions ouvertes.

On trouvera à la fin de cet article un rappel des principales notations et abréviations utilisées.

2. Quatre exemples industriels

Nous partons d'exemples concrets : trois d'entre eux correspondant à des études effectuées à un niveau industriel (à EDF entre autres). Le dernier est un exemple simplifié mais représentatif et très utilisé en formation à EDF R&D. Telle a été en effet la démarche à l'origine de la méthodologie globale : c'est en levant les compartimentages historiques constatés dans les applications industrielles récentes entre approches métrologique, mécano-fiabiliste, physico-statistique que des problématiques génériques d'incertitudes ont été formulées, puis replacées dans un cadre mathématique cohérent.

2.1. Maîtriser la métrologie d'un procédé pour garantir les émissions environnementales

Un premier exemple industriel provient du domaine de la *métrologie*; le calcul d'incertitudes s'y avère nécessaire (et normalisé) pour couvrir le risque associé par exemple à une *déclaration environnementale*. Il est également utile pour l'optimisation de la chaîne de mesure d'un procédé.

Parmi les nombreuses réglementations environnementales affectant les procédés d'un industriel comme EDF, il y a le contrôle des gaz à effet de serre. En application des engagements européens vis-à-vis du protocole de Kyoto pour la part industrielle des émissions, un système de permis d'émissions s'est mis en place, via des quotas nationaux d'abord, puis par entreprise. Dans ce contexte les incertitudes affectant le mesurage du tonnage (par exemple du CO₂) effectivement émis par un acteur peut perturber l'atteinte de ses engagements, voire même biaiser le marché financier de permis d'émissions basé sur les écarts entre émissions réelles et droits acquis. C'est ce qui a conduit les acteurs institutionnels (Commission Européenne, Gouvernements nationaux) à exiger la déclaration des incertitudes (cf. [20], [34]). L'industriel devient alors intéressé à optimiser le choix de la chaîne de mesure pour réduire les incertitudes, car celles-ci représentent en fait un coût.

Physiquement, dans le cas du CO_2 émis par une centrale thermique au charbon (Fig. 1 et 2), la mesure directe par un capteur unique de l'émission s'avère peu fiable. Plusieurs options sont possibles pour mesurer le nombre de tonnes de CO_2 ($t\text{CO}_2$) émises : (a) débitmètres à bascules intégratrices, (b) inférence depuis la puissance électrique produite via une consommation spécifique, (c) inventaire des stocks et intrants. Elles impliquent toutes l'agrégation, par le moyen d'opérations analytiques simples, de plusieurs mesures élémentaires : du combustible consommé (Cc) en tonnes; du pouvoir calorifique inférieur (PCI) en téra-joules par tonne (TJ/t); du facteur d'émission (FE_{CO_2}) en tonnes de CO_2 par téra-joule, ou du facteur d'oxydation (FE_{oxy} , sans unité), comme le montre la formule suivante :

$$\text{CO}_2(t\text{CO}_2) = \sum_{i=1}^n CC_i(t) * PCI_i(TJ/t) * FE_{\text{CO}_2(i)}(t\text{CO}_2/TJ) * FE_{\text{oxy}(i)}(-)$$

Dans l'exemple de l'option (c) la consommation de combustible Cc est elle-même déduite d'un bilan entre tonnages approvisionnés et variation de stocks mesurés par cubatures.



FIG 1. — Centrale thermique et rejets atmosphériques.

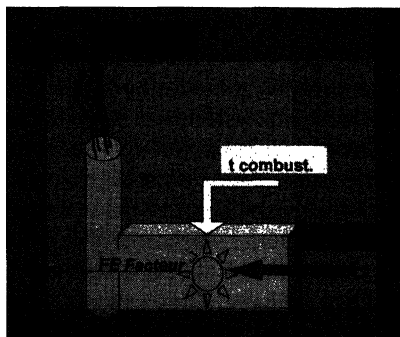


FIG 2. — Procédé d'émission.

Mathématiquement, il s'agit donc d'agrèger les incertitudes associées à chaque variable x^i mesurée par chaque i -ème opération ou capteur élémentaire, qui sont généralement évaluées par les fournisseurs à la conception, au sein de la fonction analytique $z = G(x^1, \dots, x^p) = G(\underline{x})$ (ou « modèle physico-industriel ») exprimant la valeur finale à mesurer z (mesurande). La variable z désigne par exemple le tonnage annuel de CO₂.

C'est un exemple typique de la métrologie de procédés industriels, qu'elle soit environnementale ou non : en la matière, il y a de longue date une culture probabiliste encadrée par des normes internationales, et notamment le « Guide pour l'expression de l'incertitude de mesure » (GUM, [48]). Plutôt que d'agrèger de façon déterministe les incertitudes élémentaires, elle se base sur l'hypothèse standard selon laquelle l'incertitude¹ de chaque capteur ou opération de mesure présente, par rapport à une valeur de référence, un caractère aléatoire efficacement modélisé par le calcul des probabilités.

Pour être précis sur la notion d'incertitude par rapport à une *valeur de référence*, le GUM recommande de se référer plutôt à la « meilleure valeur disponible » qu'à la « valeur vraie » physique. En effet d'une part la spécification ou la définition de la grandeur à mesurer n'est jamais infiniment précise, d'autre part la valeur vraie n'est jamais observable, mais seulement approchée par des capteurs se référant à des étalons eux-mêmes incertains. Le GUM recommande donc de ne pas travailler sur « l'erreur », définie comme écart absolu entre le résultat de mesure et la valeur vraie (inobservable), mais sur « l'incertitude » représentant la fluctuation du résultat d'un mesurage supposé avoir corrigé raisonnablement tous les effets systématiques par rapport à un étalon. La variable aléatoire représentant cette incertitude comportera éventuellement un « biais » (qui reste donc relatif) si son espérance est non nulle, même si certains réservent le terme « d'incertitude » à la variable aléatoire d'espérance nulle, restreinte à la fluctuation après déduction des biais.

Considérant que ce cadre est accepté, un critère de type *probabiliste* s'impose alors aux études réglementaires d'incertitudes. Il consiste à déterminer une « incertitude élargie », définie par un intervalle de confiance à $\alpha = 95\%$ autour de la masse $t\text{CO}_2$ mesurée, le pourcentage de la demi étendue relative à la grandeur Z , représentant une « incertitude relative de référence », ne devant pas excéder un seuil. Ce seuil limitera par exemple à 3 % l'incertitude de référence ainsi définie :

$$\%inc_z = \frac{1}{2 \cdot E(Z)} (z^{\frac{1+\alpha}{2}} - z^{\frac{1-\alpha}{2}}) \quad (1)$$

où l'on note le quantile $z^\beta = F_Z^{-1}(\beta)$, F_Z désignant la fonction de répartition de Z . On accepte souvent la formule approchée suivante, plus facile à déterminer :

$$\%inc_z \approx k \cdot \frac{\sqrt{\text{var}(Z)}}{E(Z)} \quad (2)$$

1. Incluant les bruits affectant le fonctionnement du capteur, la fluctuation locale du point de mesure par rapport à la quantité physique moyenne recherchée, l'erreur de lecture d'un compteur ou d'arrondi de l'affichage, etc.

le «facteur d'élargissement» k étant pris à 2, 3 ou encore une valeur intermédiaire suivant les cas.

De manière classique, on considère alors que les incertitudes élémentaires X^i , formant le vecteur aléatoire \underline{X} des sources d'incertitudes, sont de nature gaussienne : cette hypothèse très courante en métrologie est parfois argumentée par la symétrie des erreurs et l'existence de multiples phénomènes physiques additifs sous-jacents. Le fait que les incertitudes soient bornées, par exemple à cause d'un dispositif de contrôle-commande, peut conduire néanmoins à un choix de loi uniforme, voire triangulaire. L'espérance est généralement prise égale au résultat de la mesure bien calibrée et la variance est fournie par les caractéristiques du capteur.

La propagation des incertitudes élémentaires (on parle aussi d'agrégation ou de cumul des incertitudes) se fait par l'une des deux méthodes prévues par la norme internationale, présentées dans le tableau 1 :

- en estimant le critère approché (2) via la traditionnelle formule de cumul quadratique d'approximation de la variance et de l'espérance de Z ,
- en estimant le critère approché (2) ou le critère complet (1) par simulation de Monte-Carlo (*Monte-Carlo Sampling*, noté *MCS*).

TABLEAU 1. — Méthodes élémentaires de propagation et de hiérarchisation d'incertitude.

méthode	estimation de l'incertitude élargie	estimation de l'importance relative des sources d'incertitudes
cumul quadratique (« loi de propagation de l'incertitude » selon GUM)	$E(Z) \approx G(E(\underline{X}))^2$ $Var Z \approx \sum_{i=1}^n \left(\frac{\partial G}{\partial X^i}\right)^2 Var X^i + \sum_{i=1}^n \sum_{j=1}^n \left(\frac{\partial G}{\partial X^i}\right) \left(\frac{\partial G}{\partial X^j}\right) Cov(X^i, X^j) \quad (3)$ $inc_z \approx k \cdot \sqrt{var(Z)} \quad \%inc_z \approx k \cdot \frac{\sqrt{var(Z)}}{E(Z)}$	$C_1(Inc_{x_i} / Inc_z) = \frac{\left(\frac{\partial G}{\partial X^i}\right)^2 inc_{x_i}^2}{inc_z^2} \quad (4)$
Simulation de Monte-Carlo (MCS)	$\hat{G}(\underline{X}) = \frac{1}{n} \sum_{j=1}^n G(\underline{X}_j) \quad \hat{F}_n(z) = \frac{1}{n} \sum_{j=1}^n I_{(z)}(\underline{X}_j) \quad \hat{Z}^p = \hat{F}_n^{-1}(\beta)$ $inc_z \approx 1/2 \cdot (\hat{Z}^{\frac{1+\alpha}{2}} - \hat{Z}^{\frac{1-\alpha}{2}}) \quad \%inc_z \approx \frac{1}{2G(\underline{X})} (\hat{Z}^{\frac{1+\alpha}{2}} - \hat{Z}^{\frac{1-\alpha}{2}}) \quad (5)$	$C_2(Inc_{x_i} / Inc_z) = \frac{\hat{C}orr(r_{x_i}, r_z)^2}{\sum_{i=1}^n \hat{C}orr(r_{x_i}, r_z)^2} \quad (6)$ (corrélations de rangs « Spearman »)

Ces deux méthodes permettent également d'estimer l'importance relative des sources d'incertitude en l'absence de corrélations entre sources, comme indiqué dans la troisième colonne du Tableau 1. La première méthode est très utilisée car elle suppose peu de calculs de $G(\cdot)$ et les formules (3) et (4) sont facilement interprétables par un décideur non-statisticien : pour le résultat final, une source d'incertitude n'est importante que dans la mesure où son incertitude propre inc_{X^i} est forte et où le modèle physique y est sensible (via $\partial G / \partial X^i$). Théoriquement, il s'agit en fait d'un simple développement limité (D.L.) au premier ordre de \underline{x} autour de $E(\underline{X})$ qui est strictement équivalent à la Delta-méthode en statistique, et qui repose sur l'hypothèse que les incertitudes sont faibles au regard des non-linéarités de $z = G(x^1, \dots, x^p) = G(\underline{x})$. Son utilisation

2. Valable au 1^{er} ordre, un terme correctif apparaissant au second ordre, impliquant la hessienne : $(\partial^2 G / \partial X^i \partial X^j)_{i,j}$

pour déterminer le critère (1) suppose de plus le caractère gaussien de Z : c'est d'ailleurs le cas si, outre la faible non-linéarité de $G(\cdot)$, les X^i forment un vecteur gaussien.

Notons tout de suite que ces avantages font que la même méthode est utilisée dans un domaine industriel radicalement différent, le calcul scientifique. $G(\cdot)$ y représente alors un grand code numérique (e.g. un calcul neutronique dans un cœur nucléaire), ses paramètres d'entrée étant affectés par toutes sortes d'incertitudes ; on recherche un ordre de grandeur de « précision » sur le vecteur z des résultats. Les stratégies numériques déployées pour calculer à moindre coût les dérivées qui apparaissent dans la formule conduisent à lui donner d'autres noms : par dérivée directe = *Direct Sensitivity Analysis (DSA)*, ou adjointe = *Adjoint Sensitivity Method (ASM)* cf. [8] ; ou encore la « méthode des perturbations » ou le « calcul des moments »³. Mais du point de vue probabiliste c'est *le même traitement d'incertitudes*.

Les deux hypothèses restrictives de la méthode de cumul quadratique sont en fait levées par la seconde méthode. Il s'agit de la simulation de Monte-Carlo (MCS), c'est-à-dire du tirage d'un n -échantillon $(\underline{X}_j)_{j=1\dots n}$ des sources d'incertitude, suivi d'autant de calculs par le modèle physico-industriel déterministe des $Z_j = G(\underline{X}_j)$. Mais celle-ci suppose bien évidemment $G(\cdot)$ assez rapidement calculable, et le degré d'approximation ainsi gagné est à relativiser compte tenu de l'erreur éventuelle liée au choix de loi des X^i : faute de données dans ces approches simples, la normalité, tout à fait conventionnelle, n'est souvent pas testable.

Notons enfin la question délicate de la quantification des sources d'incertitudes : elle consiste dans ce cas simplement en la détermination des écarts-type σ_{X^i} pour chaque X^i , et des corrélations entre X^i . Selon la disponibilité des données, les normes autorisent *a priori soit* le jugement d'expert *soit* une statistique (souvent basée sur le $s_{X^i}^2$, variance empirique de X^i)⁴. Le processus d'audit interne ou externe souvent associé à la déclaration d'incertitudes abordera légitimement la question de la représentativité, vis-à-vis de la variabilité en situation réelle des capteurs, des incertitudes mesurées par le fournisseur : on retrouve le débat traditionnel entre « répétabilité » et « reproductibilité » si les conditions opératoires varient.

Par ailleurs, la quantification des sources d'incertitude génère elle-même de nouvelles incertitudes. On estime par exemple l'écart-type d'une source d'incertitude σ_{X^i} en se basant sur un écart-type empirique s_X observé sur un petit échantillon de données. Cette statistique est sujette à une inévitable fluctuation, d'autant plus forte que le nombre d'essais est faible. Cette fluctuation peut ne pas être négligeable en métrologie, et la norme impose parfois de l'inclure dans la description des sources d'incertitudes

3. Ces dénominations englobent parfois aussi la propagation déterministe d'incertitudes, car elle fait appel aux mêmes dérivées, ou alternativement l'estimation des autres moments probabilistes par développements différentiels d'ordre strictement supérieur à 1.

4. Certains qualifient l'évaluation de « type B » s'il y a jugement d'expert et de « type A » s'il y a statistique ; mais cette terminologie, courante surtout en métrologie, se chevauche parfois ailleurs avec un classement selon la nature « aléatoire » ou « épistémique » des incertitudes : elle ne sera pas utilisée ci-après.

avant calcul de propagation. Pour tenir compte de cela, il arrive que l'on pénalise la variance estimée d'un facteur⁵ supérieur à 1, et décroissant avec le nombre n de mesures ayant servi à l'estimer, au lieu de prendre simplement la valeur brute estimée : celle-ci risque, par effet de fluctuation d'échantillonnage, de sous-estimer la vraie variance inconnue. La pratique est hétérogène, notamment car n n'est pas toujours connu par l'industriel final : le § 2.2 de la 2nde partie montrera qu'il s'agit d'une première façon, simple mais non sans limites, de traiter le cumul d'incertitudes de nature « épistémique » et « aléatoire ». Cette question s'avère particulièrement importante quand le coût des essais nécessaires à la détermination des sources d'incertitudes est élevé parce qu'ils impliquent une opération élémentaire lourde. Pour EDF, la réception d'assemblages combustibles avant recharge en fait partie. Le mesurage est très onéreux, mais le risque associé à une trop forte incertitude est important : on craint d'accepter à tort un assemblage dont les dimensions excèdent en réalité le gabarit, rendant finalement indisponible l'installation.

Dans l'exemple du CO₂ (cf. [7]), quelques dizaines de sources d'incertitudes ont été propagées au sein des trois modèles physico-industriels qui représentaient les différentes options de mesure du CO₂. Dans ce cas, le caractère approximativement gaussien de Z était apparent sur le résultat de 20 000 tirages aléatoires : cela justifiait *a posteriori* l'usage de l'approximation du cumul quadratique pour évaluer le critère $\%inc_Z$ malgré la non-linéarité de $G(\cdot)$ liée aux opérations de multiplication. Le choix de l'une des chaînes de mesure pour contrôler ultérieurement le CO₂ a été fait en minimisant le critère $\%inc_Z$, mais également, via la hiérarchie d'importance des sources, en s'assurant que les plus fortes contributions correspondent à des sources traçables en matière d'assurance qualité, et si possible réductibles.

2.2. Évaluer la fiabilité d'une structure mécanique liée à ses marges (l'exemple de la cuve d'un réacteur nucléaire)

Le deuxième exemple provient de l'analyse de fiabilité des structures et de la mécanique probabiliste ([21], [53]). L'enjeu y est de garantir la fiabilité d'une structure-clé dans un procédé industriel, ici la cuve d'un réacteur nucléaire, en contrôlant les paramètres de conception ou d'exploitation, ici les conditions de fonctionnement du cœur du réacteur. Certaines marges sont prises pour se prémunir d'une défaillance importante, comme la rupture brutale, malgré les inévitables aléas, variabilités et imprécisions diverses. En simplifiant physiquement l'exemple de la cuve (Fig. 3), la défaillance redoutée peut, hypothétiquement, se produire sous l'effet d'un comportement anormal en pression-température du fluide primaire, subissant par exemple une dépressurisation à la suite d'une brèche ailleurs dans le circuit. La défaillance a lieu si les contraintes induites sur la paroi de la cuve dépassent, au droit d'un défaut pré-existant, la marge permise par la ténacité du matériau

5. e.g. un quantile 95 % de la loi de l'estimateur :

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2 \text{ (gaussien à distance finie) ou } \frac{s^2}{\sigma^2} \sim N\left(1, \frac{2}{n}\right) \text{ (asymptotique).}$$

avant la rupture brutale. On s'en tiendra parfois au simple amorçage du défaut, en adoptant une approche que l'on qualifiera de «conservative» au sens où elle englobe sûrement au *minimum* la véritable défaillance (cf.[52]).

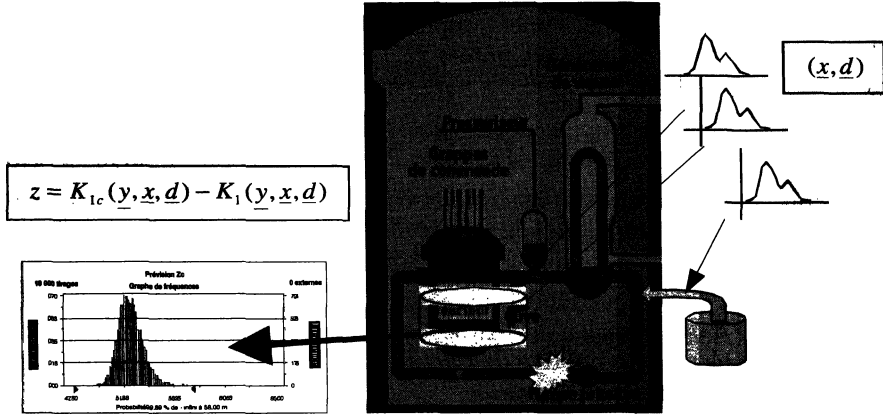


FIG 3. — Incertitudes et Fiabilité de la cuve d'un réacteur nucléaire.

L'objet d'intérêt pour l'industriel est donc l'ensemble thermomécanique, que l'on modélise de façon déterministe par une chaîne associant modèles physiques de type éléments finis et formules analytiques phénoménologiques. En reprenant la terminologie courante dans le domaine fiabiliste, la variable finale d'intérêt z est généralement la marge à défaillance : elle est construite à partir des résultats bruts de modèle $\underline{y} = M(\cdot)$ de type champ de contraintes/températures, en soustrayant un facteur d'intensité de contraintes (noté K_1) à une fonction de ténacité (notée K_{1c}). On constitue ainsi la «fonction de défaillance» ou «fonction d'état limite» $G(\cdot)$ qui doit rester positive.

$G(\cdot)$ dépend de nombreux paramètres $(x^1 \dots x^p) = \underline{x}$ qui font l'objet de multiples incertitudes : les propriétés des matériaux, les caractéristiques des défauts, la thermodynamique des transitoires accidentels, l'irradiation subie selon la durée et la fragilisation induite etc. $G(\cdot)$ dépend également des conditions contrôlées de conception ou d'exploitation rassemblées dans le vecteur \underline{d} :

$$\begin{aligned} \underline{y} &= M(x^1, \dots, x^p, \underline{d}) = M(\underline{x}, \underline{d}) \\ z &= K_{1c}(\underline{y}, \underline{x}, \underline{d}) - K_1(\underline{y}, \underline{x}, \underline{d}) = g(\underline{y}, \underline{x}, \underline{d}) = g(M(\underline{x}, \underline{d}), \underline{x}, \underline{d}) = G(\underline{x}, \underline{d}) \end{aligned} \quad (7)$$

Le domaine de la fiabilité des structures est largement réglementé dans les industries à risque comme le nucléaire. Selon les composants considérés, les critères spécifient une exigence de fiabilité, c'est-à-dire l'absence de défaillance pendant une période et pour un scénario donnés, comme un accident conventionnel avec des hypothèses \underline{d} sur la structure. Cela revient généralement à assurer que :

$$z = G(\underline{x}, \underline{d}) > 0 \quad (8)$$

Contrairement au domaine métrologique, la culture probabiliste y est encore peu présente, tout au moins dans la réglementation française. Ainsi, pour garantir l'équation (8), la voie traditionnelle consiste à « pénaliser » les variables sources d'incertitudes x par des marges de sécurité, en appliquant des coefficients multiplicatifs f^i (également appelés facteurs de sécurité) aux valeurs « best-estimate » des paramètres, notées \underline{x}_{be} , et en vérifiant la fiabilité par un calcul déterministe effectué avec ces paramètres pénalisés, cf. [53].

$$\underline{x}_{pn} = (x_{be}^1 \cdot f^1, \dots, x_{be}^p \cdot f^p) \quad z_{pn} = G(\underline{x}_{pn}, \underline{d}) > 0 \quad (9)$$

Cette approche peut être vue comme un traitement « déterministe » élémentaire des sources d'incertitudes ou, si l'on préfère, comme un traitement « par scénario pénalisé » : elle suppose bien évidemment :

- d'une part qu'il y ait monotonie de $G(\cdot)$. Si ce n'est pas le cas, un algorithme d'optimisation devient en toute rigueur nécessaire, multipliant considérablement le nombre de calculs à réaliser : on parlera alors d'une approche « déterministe forte »,
- d'autre part qu'une borne supérieure (ou inférieure suivant le sens de la monotonie) de l'incertitude de chaque x^i ait un sens et une valeur « réaliste ». En pratique, cela relève souvent l'expertise physique informelle ; mais il arrive que certaines bornes x_{pn}^i , dénotant la valeur pénalisée de la variable x^i , correspondent à des quantiles approximatifs $x^{i\alpha^i}$ de chaque paramètre *implicitement* modélisé comme une variable aléatoire X^i sur la base de quelques données. La question de l'incertitude épistémique générée par la faible taille des échantillons, évoquée dans l'exemple métrologique ci-dessus, peut donc se poser à nouveau.

Formellement on peut voir (9) comme un critère mixte déterministe-probabiliste basé sur le conditionnement intégral de \underline{X} :

$$(9) \Leftrightarrow z = G(\underline{X}/(X^i = x^{i\alpha^i})_i, \underline{d}) > 0 \Leftrightarrow P[G(\underline{X}/X^i = x^{i\alpha^i})_i, \underline{d}] > 0 = 1 \quad (10)$$

Une importante littérature continue à débattre de l'intérêt et du conservatisme inhérent à ce genre d'approche dite « déterministe », qui cumule en fait d'une façon difficilement contrôlable des marges reflétées par les quantiles α^i , qui sont éventuellement hétérogènes. Depuis de nombreuses années, le débat est alimenté par leur comparaison avec des approches dites « probabilistes » : celles-ci modélisent explicitement les sources d'incertitudes par des variables aléatoires X^i , et s'intéressent à la « probabilité de défaillance » du scénario définie par $P_f(\underline{d}) = P[G(\underline{X}, \underline{d}) < 0]$ en cherchant à assurer qu'elle est très proche de 0. En fait il est fréquent que l'on ne « probabilise » qu'une partie des sources d'incertitudes (indicées par $i = 1, \dots, m$) ce qui revient à un conditionnement partiel aux valeurs pénalisées des autres sources (indicées par $i = m + 1, \dots, p$). *In fine*, on comparera P_f à un seuil, ou à la probabilité d'un scénario de référence \underline{d}^0 , considéré comme assez fiable ou sûr, si un seuil absolu en 10^{-b} est trop délicat à spécifier. Formellement on peut donc lire ces approches dites « probabilistes » comme un traitement d'incertitudes mixte

déterministe-probabiliste associé au nouveau critère (11) qui remplace (10) :

$$P_f(\underline{d}) = P[G(X/(X^i = x_{pn}^i)_{i=m+1,\dots,p}, \underline{d}) < 0] < 10^{-b} \quad (11)$$

ou $P_f(\underline{d}) = P[G(X/X^i = x_{pn}^i)_{i=m+1,\dots,p}, \underline{d}) < 0] < P_f(\underline{d}^0)$

L'étude d'incertitudes face à ce critère pourrait-elle employer l'une des deux méthodes de *propagation* déjà mentionnée dans l'exemple métrologique du § 2.1 ? Notons d'abord que le fort compartimentage historique entre métrologie et fiabilité des structures n'a pas rendu naturelle cette question ; mais notre formalisme a pour but précisément de dépasser ces héritages. Plus profondément, la différence vient essentiellement de ce que l'on s'intéresse, en fiabilité des structures, à des « événements de défaillance » (c'est-à-dire $\{X \text{ tels que } G(\underline{X}, \underline{d}) < 0 \text{ avec } (X^i = x_{pn}^i)_{i=m+1,\dots,p}\}$ par définition) qui sont rares, c'est-à-dire qui correspondent à la queue de la distribution de Z . De plus, ces événements sont souvent associés à un modèle physico-industriel $G(\cdot)$ beaucoup plus long à calculer qu'une formule analytique métrologique. Dans ce contexte les hypothèses gaussiennes et linéaires associées à la méthode de cumul quadratique s'avèrent souvent fausses. Quant à MCS, la méthode nécessite un trop grand nombre de simulations pour stabiliser l'estimateur de P_f (cf. Tableau 3 de la 2^{de} partie).

Outre l'usage de méthodes de réduction de la variance de MCS (réduction de la dimension, tirage d'importance), des stratégies de calcul plus spécifiques ont été historiquement développées pour l'évaluation de la probabilité de dépassement d'un seuil. Elles sont connues en fiabilité sous le nom de First (respectivement Second) Order Reliability Method, abrégées en FORM (resp. SORM). Elles réduisent grandement, si les approximations sous-jacentes le permettent, le nombre d'appels au modèle déterministe $G(\cdot)$ pour évaluer une P_f très faible (cf. § 2.4.2 de la 2^{de} partie). Ces méthodes génèrent également des « facteurs d'importance » hiérarchisant l'importance des sources d'incertitudes. Mais cette hiérarchisation est faite vis-à-vis d'un critère de dépassement de seuil et non pas vis-à-vis de l'incertitude moyenne (coefficient de variation) de z , ce qui s'avère différent des mesures d'importance (4) et (6) présentées ci-dessus (cf. § 2.4.4 de la 2^{de} partie). Dans l'exemple de la cuve, des probabilités conditionnelles⁶ de l'ordre de 10^{-4} ont pu être évaluées en quelques dizaines de simulations, là où MCS classique aurait nécessité au moins 10^5 - 10^6 simulations cf. [52]. Enfin, d'autres stratégies remplacent le modèle physique originel $G(\cdot)$ par une « surface de réponse », c'est-à-dire un autre modèle déterministe « proche » de $G(\cdot)$ et beaucoup plus rapidement calculable, et donc simulable par MCS (cf. § 2.4.3 de la 2^{de} partie et [15] dans l'exemple de la cuve).

6. On gardera en mémoire qu'il s'agit de probabilités *conditionnelles* à de nombreuses hypothèses, elles-mêmes peu probables, dont notamment des transitoires de fonctionnement accidentels dont la fréquence conventionnelle est typiquement moindre que 10^{-4} par an et par réacteur.

2.3. Maîtriser la robustesse d'une analyse de risque/de sûreté de fonctionnement (l'exemple des incertitudes dans les eps)

Notre troisième exemple est tiré du domaine de l'analyse de risque/de la sûreté de fonctionnement. L'enjeu y est typiquement d'évaluer la fréquence d'un événement redouté sur un système : il peut s'agir de l'endommagement d'un cœur d'un réacteur nucléaire ou encore de l'indisponibilité d'une usine, d'une fréquence d'accidents sur un système de transport etc.. Dans de nombreux cas, les analyses quantitatives consistent à construire un modèle de type *fiabilité des systèmes* (par arbres de défaillance cf. Fig. 4, ou par arbres d'événements, voire par réseaux bayésiens etc.) des causalités ou enchaînements d'événements élémentaires probabilisés conduisant logiquement à l'événement redouté. On calcule ainsi la fréquence ou la probabilité d'occurrence⁷ sur une durée donnée. En manipulant des événements dont l'occurrence et les liens causaux sont souvent mal connus car (heureusement) rares, ce type d'évaluation est sujet à d'évidentes incertitudes, qui sont d'ailleurs une préoccupation de longue date : dans le cas des « études probabilistes de sûreté » (E.P.S.) qui modélisent la fréquence d'endommagement du cœur d'une centrale nucléaire, les études d'incertitudes ont été initiées dès les années 70 aux USA (cf. [56]). L'enjeu associé aux incertitudes y est de maîtriser la précision sur le risque prévu, et de garantir typiquement la robustesse d'une conclusion basée sur une analyse de risque. On s'intéressera par exemple au fait que le risque associé à un scénario ou à un système reste inférieur à un seuil donné ou à celui d'un système de référence suffisamment sûr.



$$F_{fusion} = G_{déterministe}[(\lambda_i)(\gamma_j), (init_k) \dots]$$

FIG 4. — Fiabilité des systèmes par arbres de défaillances

En fait, même si l'analyse de risque manipule des grandeurs ayant un sens probabiliste, *formellement vis-à-vis des incertitudes* l'objet ressemble beaucoup aux exemples précédents. Notons z la quantité d'intérêt – ici la fréquence annuelle d'endommagement du cœur d'un réacteur – et $(x^1 \dots x^p)$ les paramètres

7. Les deux notions sont théoriquement distinctes si l'événement peut survenir plus d'une fois sur la période : la fréquence d'occurrence d'un événement sur une durée T peut être supérieure à 1 tandis que la probabilité qu'il arrive au moins une fois sur T reste évidemment bornée à 1. Mais elles sont numériquement très proches pour des risques rares, pour lesquels la probabilité d'observer plus d'une occurrence sur T reste négligeable.

incertains des événements élémentaires, ici la fréquence annuelle d'un événement initiateur (brèche, séisme, grand froid), la probabilité conditionnelle de défaillance d'un composant suite à tel événement *etc.* L'analyse de risque conduit à une *relation fonctionnelle* entre z et $(x^1 \dots x^p)$, et qui dépend également des paramètres contrôlés (et donc non incertains) de conception ou d'exploitation du système, que l'on rassemblera dans le vecteur \underline{d} :

$$z = G(x^1, \dots x^p, \underline{d}) = G(\underline{x}, \underline{d})$$

Cette relation fonctionnelle est identique en tout point à une relation « déterministe » physique, même si z représente une probabilité d'occurrence ou une fréquence d'un événement redouté, et même si $G(\cdot)$ n'est plus un modèle physique, mais un modèle de fiabilité des systèmes. Les critères associés à l'étude d'incertitudes sont alors typiquement :

- la donnée d'un intervalle de confiance $(z^{\frac{1-\alpha}{2}}, z^{\frac{1+\alpha}{2}})$ autour de la fréquence moyenne prédite (ou « best-estimate ») z_{be} , e.g. $1.5 \cdot 10^{-9} < z < 3.0 \cdot 10^{-9}$ à 95 % de confiance,
- la comparaison à un seuil absolu $z < z_s$: e.g. $z < 10^{-8}$ à une confiance supérieure à 99 %,
- la comparaison relative à un système de référence $z(\underline{d}) < z(\underline{d}^0)$ c'est-à-dire : dans la configuration \underline{d} , le système est significativement (e.g. à 95 %) plus fiable/sûr que dans une configuration de référence jugée acceptable \underline{d}^0 ,
- plus généralement enfin, le critère final peut être un classement par rangs entre scénarii (ou systèmes, types de risques...) restant significatif même sous incertitudes, c'est-à-dire robuste : par exemple le système étudié reste significativement le plus sûr parmi 5 systèmes de référence, malgré les incertitudes. Celles-ci peuvent d'ailleurs affecter conjointement plusieurs des systèmes comparés, auquel cas il faut veiller à les représenter par des variables aléatoires dépendantes dans l'étude d'incertitudes.

On conçoit ainsi que de nombreux travaux aient décliné des études d'incertitudes qui apparaissent formellement comparables aux exemples précédents : soit par approche déterministe, en pénalisant les paramètres $\underline{x} = \underline{x}_{pn}$; soit par simulation MCS des incertitudes, ce qui est notamment le cas pour les E.P.S.. Comme $G(\cdot)$ est souvent rapidement calculable, l'utilisation de MCS permet d'estimer :

- un intervalle de confiance $(z^{\frac{1-\alpha}{2}}, z^{\frac{1+\alpha}{2}})$ sur la fréquence d'endommagement,
- ou la probabilité de dépassement d'un seuil de fréquence inacceptable $P(Z = G(\underline{X}, \underline{d}) > z_s)$.

MCS donne également une hiérarchie d'importance des sources d'incertitudes par la même méthode qu'en métrologie (cf. (6) supra). Ce dernier résultat est d'importance majeure compte tenu du très grand nombre de sources d'incertitudes, \underline{X} étant un vecteur à plusieurs centaines de dimensions dans les E.P.S.

Il faut prendre garde à ne pas confondre cette simulation MCS utilisée pour propager les incertitudes avec la simulation MCS éventuellement utilisée pour calculer le résultat du modèle fiabiliste $G(\underline{x})$ en un point donné \underline{x} , c'est-à-dire sans incertitudes. En effet, dans le cas où $G(\underline{x})$ est un modèle de fiabilité dynamique, il se peut que l'absence de solution analytique conduise à utiliser une simulation MCS. Dans ce cas, l'étude d'incertitudes par Monte-Carlo conduit à un double emboîtement de simulations MCS. De même, la hiérarchie d'importance des sources d'incertitudes ne doit pas être confondue avec les facteurs d'importance des événements élémentaires ou séquences vis-à-vis de la fréquence de l'événement redouté : il s'agit là de caractéristiques structurelles du modèle fiabiliste représenté par la « fonction déterministe » $G(\cdot)$ que l'on étudie dans une E.P.S. même en l'absence d'incertitudes sur x , et qui s'apparentent à des « sensibilités » $G(\underline{x}d)/\Delta x^i$. Quant à elle, la mesure d'importance des sources vis-à-vis des incertitudes autour de la fréquence de fusion dépend donc de cette notion de facteurs d'importance « fiabilistes », mais s'en distingue : pour avoir une importance notable au titre des incertitudes, il faut que l'événement élémentaire, dont la fréquence ou probabilité conditionnelle est x^i , ait un facteur d'importance élevé, mais aussi que la source d'incertitudes sur x^i soit forte (cf. l'équation simplifiée (4)).

Souvent dans les calculs industriels d'incertitudes dans les E.P.S. (cf. [58]), des hypothèses systématiques de loi log-normale ont été faites sur les X^i . Cette même habitude a d'ailleurs été prise dans certaines études d'incertitudes en économie de l'environnement cf. [54], où l'on manipule également un grand nombre de sources X^i en se basant principalement sur le jugement d'expert. Ces hypothèses log-normales présentent deux avantages pratiques : un support exclusivement positif ; la facilité, pour les experts, de s'exprimer en ordres de grandeur multiplicatif d'incertitude en catégorisant le très grand nombre de variables par des facteurs d'erreur (par exemple un facteur 10). Un facteur d'erreur multiplicatif est traduit, via un intervalle de confiance centré, en écart-type de $\ln X^i$.

Ces hypothèses ne sont néanmoins en rien nécessaires. Remarquons que certains paramètres x^i sont estimés d'après des données de fiabilité comme par exemple la fréquence de défaillance d'un composant sur une durée T , d'après retour d'expérience sur le fonctionnement d'un parc. La statistique mathématique peut alors parfois indiquer d'autres choix de lois pour leur incertitude d'estimation comme par exemple celui d'une distribution du χ^2 pour l'estimateur du paramètre d'une loi exponentielle sur des durées de vie. Si l'on tient à rester pour des raisons pratiques dans un cadre log-normal, on assurera le recouvrement d'un I.C. à 95 % entre la loi issue de l'estimation statistique et le modèle log-normal : cette approximation est supposée acceptable à condition que l'on ne recherche pas des quantiles trop rares sur l'incertitude de la variable finale z .

Même si l'étude d'incertitudes d'un modèle d'analyse de risques $G(\cdot)$ paraît ainsi comparable à celle d'un modèle physique, le fait que les grandeurs sous-jacentes \underline{x} et z aient un sens probabiliste conduit parfois à des interprétations spécifiques. Assez naturellement l'aléa modélisé par le modèle d'analyse

de risque, conduisant à une fréquence ou à une probabilité d'occurrence d'un événement redouté $z = G(\underline{x}, \underline{d})$, est distingué de l'incertitude dite « épistémique » affectant les paramètres x^i et donc cette fréquence z : si le traitement d'incertitudes est probabiliste, ceci conduit à un deuxième niveau de probabilité.

Mathématiquement, il est d'ailleurs possible de présenter l'étude dans un formalisme théorique bayésien modélisant ces deux niveaux de probabilité. Désignons par x^i l'état déterministe de l'initiateur/du composant, c'est-à-dire la variable booléenne $x^i = 1$ si l'événement est initié ou défaillant, $x^i = 0$ sinon, et non plus la fréquence/probabilité de l'événement d'initiation ou défaillance, c'est-à-dire $x^i = P$ (*i*-ème événement ou état composant = 1). De même désignons par z l'état de l'événement redouté et non plus sa fréquence annuelle, soit $z = 1$ si le réacteur est endommagé et $z = 0$ sinon. Le nouveau modèle $z = G(\underline{x})$ représente le même modèle logique (booléen) que précédemment mais calculé désormais avec des 0 et des 1 au lieu des valeurs réelles représentant les fréquences/probabilités élémentaires. On plonge alors dans un univers incertain bayésien avec des v.a. $\underline{X} \setminus \underline{\theta}$ représentant la loi de probabilité associée aux états (*i.e.* une loi booléenne) conditionnelle à la connaissance de la probabilité d'initiation / défaillance associée $\underline{\theta}$, qui est elle-même une v.a. suivant la distribution d'incertitudes. La probabilité $P(Z = 1 \setminus \underline{\theta})$ correspond alors au calcul « ponctuel » de l'EPS en supposant $\underline{\theta}$ connu ; la propagation d'incertitudes, par tirage des $\underline{\theta}_i$, génère la distribution des $P(Z = 1 \setminus \underline{\theta}_i)$. Le calcul reste donc tout à fait identique au formalisme présenté dans le texte ci-dessus, même si l'espérance de la loi d'incertitudes de Z s'interprète ici comme l'espérance inconditionnelle, ou « probabilité finale d'endommagement » intégrant les deux niveaux de probabilités « épistémique » et « aléatoire » (*cf.* aussi § 3.2 de la 2^{nde} partie). Ce formalisme est également utile dans le cas de la fiabilité dynamique, où l'estimation $P(Z = 1 \setminus \underline{\theta})$ à $\underline{\theta}$ fixé (sans incertitudes) requiert parfois déjà un premier MCS. Comme indiqué ci-dessus, on retrouve dans ce cas un double MCS emboîté : ce cadre théorique est plus lourd.

La distinction « épistémique » versus « aléatoire », qui semble facile à manier au premier abord, et qui est tout à fait traditionnelle dans le cadre EPS, a néanmoins ses limites (voir discussion au § 3.3).

Notons enfin que les incertitudes affectant l'analyse de risque ne se limitent pas aux « paramètres » d'entrée des modèles (*e.g.* occurrence des événements élémentaires, durées *etc.*), mais affectent également la structure des modèles eux-mêmes (*e.g.* non exhaustivité des séquences décrites, liens causaux mal connus, approximation de la fiabilité statique...). Il arrive même que ces « incertitudes de modèle » soient considérées *a priori* comme prépondérantes, alors que les études d'incertitudes se limitent aux incertitudes dites « paramétriques ». Toutefois, rien n'empêche formellement d'inclure dans le vecteur \underline{x} également des « méta-paramètres » représentant ces incertitudes de modèle, x_i quantifiant par exemple par 0 ou 1 l'inclusion d'un lien causal entre événements, voire indiquant par un nombre le passage d'un modèle à

un autre dans une collection de modèles concurrents. Ensuite, le traitement d'incertitudes peut se faire :

- (i) de façon mixte en pénalisant ces sources (*i.e.* en choisissant le modèle concurrent le plus pessimiste) tout en quantifiant des lois de distributions pour les sources classiques (« paramètres »),
- (ii) de façon intégralement probabiliste en quantifiant par des probabilités l'incertitude sur le choix du « bon modèle » (proposé par *e.g.* [1]).

Néanmoins, le traitement probabiliste (ii) de l'incertitude de modèle n'est utilisé que très rarement. Outre le fait qu'il est controversé dans la littérature (*e.g.* [2], [47]), il reste en effet très difficile à mettre en pratique car il est irréaliste ou trop coûteux de construire complètement une collection de modèles vraisemblables. On restera généralement sur l'option (i), sans toutefois expliciter formellement les « méta-paramètres ». Cela revient à un traitement pessimiste vis-à-vis de l'incertitude de modèle. On gardera toutefois ci-après le cadre formel général où \underline{x} peut inclure *a priori* toutes les sources d'incertitudes.

2.4. Un exemple pédagogique – le cas des crues

Un dernier exemple analytique à caractère pédagogique a été proposé (*cf.* [16]) par EDF R&D, de manière à illustrer et tester l'ensemble des méthodes utilisées en calcul d'incertitudes⁸. L'objet physico-industriel représente dans ce cas une rivière bordée par une digue de protection d'un site industriel. Les variables finales d'intérêt (vecteur \underline{z}) comprennent (Fig. 5) :

- d'une part la surverse $s = z_c - z_d$ de la crue (z_c) par rapport à la cote de digue (z_d) au droit d'un site industriel, c'est-à-dire la hauteur d'eau de crue débordant de la crête de la digue. La surverse dépend, via le modèle physique : du débit de crue de la rivière (q); de l'état incertain de la rivière, via le coefficient de frottement k_s dit « de Strickler » (avec la convention inverse, classique en hydraulique, selon laquelle plus k_s décroît plus l'écoulement frotte sur le fond); et des cotes du fond z_m à l'amont et z_v au droit du site inondable;
- d'autre part le coût complet c_c incluant : le coût d'investissement c_i , dépendant de la variable de contrôle $\underline{d} = z_d$, la cote de dimensionnement de la digue; et le coût de dommages c_d s'il y a eu débordement, incluant les dégâts et l'indisponibilité induite. Ces coûts sont calculés par un modèle économique qui intègre une nouvelle source d'incertitude, celle sur le coût à surverse donnée c_m : ce dernier coût est très variable selon l'état d'activité du site, et difficile à chiffrer précisément *a priori*.

Deux enjeux, la sûreté et l'optimisation économique, conduisent à deux critères différents :

- (i) le non dépassement de la cote seuil par la cote de crue (surverse négative),

⁸. Cet exemple pédagogique, même s'il a un sens physique, n'est pas représentatif des actions et études menées pour la protection des sites EDF contre les inondations : le modèle physique est trop schématique.

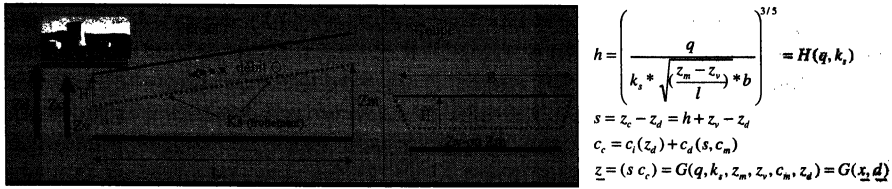


FIG 5. — Modèle simplifié des crues au droit d'un site industriel

- (ii) l'optimisation du coût complet. L'optimisation concerne ici le choix de la hauteur de la digue, vis-à-vis de la distribution du coût complet c_c : on peut élémentairement en minimiser l'espérance, ce qui est bien entendu assez risqué, ou plutôt minimiser la probabilité d'un dommage catastrophique (e.g. c_c dépassant un seuil de solvabilité) ce qui dénote une plus grande aversion au risque. On peut enfin effectuer l'optimisation à partir de tout autre critère composite, même non-linéaire en probabilité (cf. la théorie de la décision, par exemple [4]).

Plusieurs approches de traitement d'incertitudes sont alors possibles pour l'un ou l'autre de ces critères (voir tableau 2). Le caractère aléatoire des crues étant reconnu de longue date, elles ont en commun de requérir au préalable un traitement des données de type « statistiques extrêmes » pour estimer un quantile de crue de référence (typiquement millénal), via un modèle paramétrique sur le débit. On exclut ici une méthode « déterministe élémentaire » qui se baserait sur un débit pénalisé sans aucune étude statistique, par exemple le maximum historique observé ou un débit d'expert : une telle approche est très rare dans le cas des crues, où l'on se réfère toujours à une notion de quantile de débit, même s'il y a très peu de données. On s'appuie pour cela sur les méthodes classiques maxima par blocs (annuels) ou par dépassement de seuil (renouvellement), en estimant également un intervalle de confiance, souvent normal asymptotique, sur les quantiles d'intérêt (cf. [45]).

Après cela, les approches possibles sont schématiquement les suivantes, leurs caractéristiques étant comparées dans le tableau 2 :

- L'approche dite « déterministe » classique (cf. (10) supra) : elle se base sur des valeurs pénalisées de chaque source. Pour l'état de la rivière, il s'agira d'une valeur pessimiste choisie par jugement d'expert. Pour le débit on prendra généralement une borne supérieure de l'intervalle de confiance d'estimation du quantile conventionnel (e.g. millénal à 70 %). Un seul calcul physique est alors nécessaire.
- L'approche mixte « probabiliste sur quantile » (cf. (11) supra) : les incertitudes sur l'état de la rivière sont représentées de façon probabiliste, au même titre que l'incertitude d'estimation du débit millénal. Plusieurs calculs physiques sont alors nécessaires pour estimer une borne supérieure de confiance (e.g. 70 %) de la cote à débit millénal. Cela s'effectue par simulation MCS ou parfois seulement par le cumul quadratique.

- L'approche probabiliste « directe » (cf. aussi (11) supra) : dans cette dernière approche, l'aléa sur le débit de crue est représenté au même niveau que les incertitudes sur l'état de la rivière, et il faut alors estimer un quantile millénal de z , c'est-à-dire une cote millénale et non plus cote à débit millénal. Beaucoup de calculs sont alors nécessaires, surtout si l'on se contente de la simulation MCS. Mais cette approche est la seule à permettre de simuler, et donc d'optimiser, la dimension économique, qui dépend de la totalité de la distribution des crues.

TABLEAU 2. — Diverses approches possibles du traitement d'incertitudes dans les crues

	Approche classique dite « déterministe »	Approche « mixte » (classique + traitement probabiliste d'incertitudes)	Approche probabiliste « directe »
Enjeux	Sûreté – « protection assurée sur crue conventionnelle »	Sûreté – « protection assurée sur crue conventionnelle, avec niveau de confiance »	Sûreté – « protection pour un niveau de risque global donné » ou Economie – « investissement optimal en moyenne »
Variable finale (z)	s surverse	s surverse	$\bar{z}=(s, c_r)$ surverse et coût
Variables sources (x)	q_{1000}, k_s, z_m, z_v	q_{1000}, k_s, z_m, z_v	$q, k_s, z_m, z_v, c_m \dots$
Var. contrôlées (d)	z_d	Idem	Idem
Critère	$S_{pn} = G(\bar{x}_{pn}, d) < 0$	$P[S = G(\bar{X}, Q_{1000}, d) > 0] < \dots$	$P[S = G(\bar{X}, Q, d) > 0] < 10^{-b}$ ou $Min[E(Cc(X, d))]$
Quantification des sources			
var. pénalisées	Q_{1000}, K_s, Z_m, Z_v	aucune	aucune
var. aléatoires	(Q_{1000})	Q_{1000}, K_s, Z_m, Z_v	Q, K_s, Z_m, Z_v, C_m
Propagation des incertitudes	1 seul calcul	Cumul quadratique MCS (n faible)	MCS (n grand) MCS accéléré, Form-Sorm ...
Hiérarchisation des sources	non	oui, vis-à-vis de l'incertitude centrale inc_z	oui, vis-à-vis (i) d'un quantile rare de S ou (ii) de l'espérance de C_c

La justification du choix de l'une ou l'autre des méthodes n'est pas l'objet du présent article : elle relève d'un débat avec les autorités de contrôle, propre au domaine des risques d'inondations. Outre la différence de temps de calcul et de complexité d'études, remarquons simplement que ces approches répondent à des critères différents et/ou à des choix de modélisation différents de l'incertitude. Les deux premières approches séparent explicitement d'un côté « l'aléa » que traduit le quantile millénal du débit et de l'autre côté les incertitudes restantes (affectant la connaissance de l'état de la rivière et l'estimation statistique du débit). Ces incertitudes restantes sont traitées soit par pénalisation déterministe (1^{ère} approche, l'incertitude d'estimation de q_{1000} provenant néanmoins d'une méthode statistique cf. (10) supra), soit par cumul probabiliste (2^{ème} approche). La troisième approche plonge l'aléa du débit et les autres incertitudes dans une seule et même expérience aléatoire. Il est tentant de relire ces différences à travers le distinguo célèbre entre le type « aléatoire » (pour le débit, à information parfaite) et le type « épistémique » (pour le reste), mais la réalité est plus complexe. L'état de la rivière recèle à la fois une variabilité irréductible, due par exemple aux phénomènes d'érosion/sédimentation

non prévisibles par un modèle déterministe à l'échelle de temps inter-annuelle des crues extrêmes, et une méconnaissance importante, due à des mesures insuffisantes et bruitées, un maillage topographique simplifié *etc.* Il n'est pas simple en pratique de séparer ces deux aspects.

Remarquons enfin que cet exemple illustre une situation courante dans les questions d'incertitudes en mécanique des fluides : de l'incertitude « de modèle » est inévitablement incluse dans la source d'incertitude K_s dans la mesure où le coefficient de Strickler provient en fait d'une modélisation macroscopique décrivant de façon très intégrée les nombreux phénomènes de dissipation (éventuellement turbulente) sous-jacents à l'écoulement (cf. [41]). De ce fait, une phase de « calage » du modèle se fait généralement pour quantifier cette variable qui n'a pas en réalité de « vraie » valeur physique observable. Ce calage consiste à ajuster $X^i = K_s$ pour faire « coller » au mieux des mesures $(Y_{mj})_{j=1\dots n}$ (e.g. sur la hauteur d'eau mesurée pour un débit connu) aux prédictions du modèle Y . Ce calage est effectué classiquement de façon déterministe : selon le cas, une valeur « best-estimate » est obtenue en minimisant les écarts modèle-mesure $U_j = Y_{mj} - Y_j$ (en norme quadratique par exemple) ou bien une valeur « pénalisée » est obtenue en assurant la « couverture » de tous les essais (écarts modèle-mesure toujours positifs ou négatifs selon le sens de la sécurité). Ce calage peut être effectué de façon probabiliste, en identifiant une distribution en K_s expliquant au mieux, au sens d'une certaine norme probabiliste, la « proximité » entre les réalisations des v.a. Y et Y_m . On peut voir ce « calage probabiliste » comme une méthode de quantification de la source d'incertitude sur K_s par voie indirecte (cf. 2.3 de la 2^{nde} partie). Notons que cette méthode conduit aussi à représenter une partie de l'incertitude de modèle dans certaines composantes X^i du vecteur \underline{X} : cette pratique est plus couramment acceptée dans la littérature de mécanique des fluides que dans l'exemple de fiabilité des systèmes présenté au § 2.3.

3. Approche méthodologique globale

3.1. Bien identifier le modèle physico-industriel et les finalités de l'étude d'incertitudes adaptées aux enjeux

Les exemples présentés permettent désormais de proposer un cadre générique d'approche des questions d'incertitudes *quantitatives*. Le tableau 3 synthétise les caractéristiques des exemples présentés au § 2 : de très nombreux autres exemples issus de la pratique industrielle récente pourraient en être rapprochés, cf. en particulier [10], [11], [24], [25], [30], [44], [52], [63], [64], [65].

LA MAÎTRISE DES INCERTITUDES DANS UN CONTEXTE INDUSTRIEL

TABLEAU 3. — Bilan comparatif des exemples industriels pour le traitement d'incertitudes

	Météorologie des émissions environnementales (cas du CO ₂)	Fiabilité de la Cuve	Analyse de risque - EPS	Protection contre les inondations
Enjeu décisionnel /finalité	1. Justification vis-à-vis de la réglementation environnementale 2. Optimisation d'un système météorologique	Justification vis-à-vis du risque de dépassement d'un seuil de sûreté (e.g. marge à la rupture). Hiérarchisation des variables importantes, et donc des efforts de R&D	Contrôle du risque d'occurrence d'un événement redouté : hiérarchisation des incertitudes face à ce risque moyen.	Justification vis-à-vis du risque de débordement par dessus la protection d'un site. Optimisation de l'investissement de protection
Critères	% « incertitude de référence » (i.e. _ I.C. à 95%) inférieure à un seuil sur la variable réglementée (rejet en tCO ₂ /an)	Probabilité de dépassement d'un seuil de sûreté	Intervalle de confiance autour du risque « best-estimate »	Probabilité d'inondation –ou espérance/ Probabilité de coûts importants
Domaine et objet « physico-industriel »	Météorologie d'un procédé industriel réglementé – émissions environnementales d'un site industriel	Dimensionnement mécanique de la cuve du réacteur sous conditions accidentelles	Fiabilité des systèmes – ensemble des scénarii de défaillance d'un système	Dimensionnement contre des risques naturels – zone inondable+digue
Source de risque / d'incertitude	Erreurs de mesure Variabilité de paramètres de fonctionnement de 5 à 50 variables incertaines	Incertitude des paramètres (ou entrées) du modèle 4-8 variables incertaines	Epistémique paramétrique Erreur de modèle plusieurs centaines de variables.	Aléa météo. Incertitude de modèle. Méconnaissance de la physique 4-6 variables incertaines
Modèle physico-industriel et variable finale	Procédé industriel à l'origine de l'émission Modèle analytique représentant la chaîne météorologique aboutissant à la variable réglementée d'émission	La marge à rupture, fonction d'un modèle thermo-mécanique et matériaux à complexité variable (0D, 1D, 2D ...)	Modèle de fiabilité des systèmes/analyse des risques Fiabilité/disponibilité du système	Modèle hydraulique de complexité variable (0D, 1D, 2D ...) >> surverse Modèle économique>> coût
Temps CPU typique pour un calcul de G(.) ⁹	< 10 ⁻³ s	1 mn – 10 h selon modèle	< 1 s	de 10 ⁻² s à 10 h – selon modèle
Paradigme de traitement d'incertitudes et quantification des sources	Modélisation probabiliste Incertitudes généralement gaussiennes déterminées par avis d'expert ou par statistiques de calibrage par le fournisseur	Modélisation mixte déterministe-probabiliste. Lois déterminées par statistiques sur les matériaux, essais etc.	Modélisation probabiliste Hypothèses de lois lognormales. Quelques statistiques de durées de vie	Plusieurs mixages possibles déterministe-probabiliste Statistiques extrêmes sur débit – avis d'expert Quantification inverse - « calage probabiliste »
Méthode de « calcul » (propagation hiérarchisation)	Cumul quadratique ou simulation MCS standard Corrélations de rangs entrées/sortie	Méthodes accélérées (Form, MCS accélérée) Facteurs d'importance	MCS standard Corrélations de rang, voire méthodes plus complexes	MCS standard Corrélations de rang ou f. d'importance Form
Autres exemples comparables	Météorologie de procédés industriels / d'essais de réception	Génie civil – Accidents thermo-hydrauliques – stockage de déchets	Analyse de risque par réseau bayésien	Séisme. Vents extrêmes

On retrouve en effet à chaque fois (cf. Fig. 6) :

- un objet physico-industriel au cœur de l'étude, représenté par une chaîne de modèles (G(.))
- des sources d'incertitudes affectant ces modèles (notées x),
- des enjeux industriels (sûreté/sécurité, responsabilité environnementale, optimisation économique et financière) qui pilotent l'étude d'incertitude,
- un ou des critères portant sur une ou des variables finales z prédites par la chaîne de modèle pour une ou plusieurs finalités parmi les suivantes (cf. aussi [17]) :

9. Ordres de grandeurs sur des PC de 2006, sans efforts spécifiques d'optimisation informatique. Noter qu'il faut souvent évaluer beaucoup de scénarii différents avec pour chacun un traitement d'incertitudes; le temps est donné pour une valeur des variables incertaines et un scénario.

LA MAÎTRISE DES INCERTITUDES DANS UN CONTEXTE INDUSTRIEL

- o la justification (ou « certification », vis-à-vis de la réglementation),
- o la hiérarchisation (des efforts de mesure, de modélisation ou de R&D), menée souvent simultanément avec la première finalité,
- o l'optimisation (des choix de conception, d'exploitation/maintenance...) ultérieure éventuelle,
- o la validation (i.e. calage/qualification d'un modèle numérique, d'un procédé de mesure...) parfois menée en amont des autres finalités.

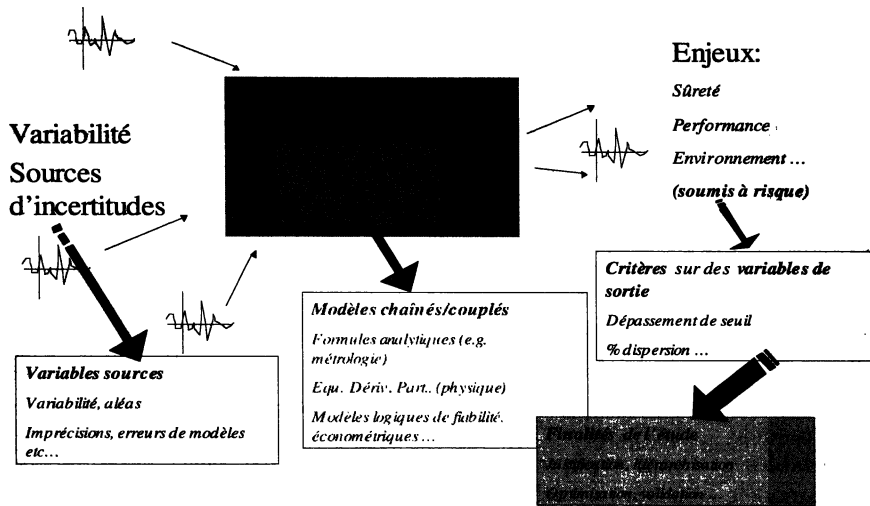


FIGURE 6. — Cadre général et enjeux d'un problème industriel sous incertitudes

Notons tout de suite que, pour un objet physico-industriel donné, le choix de la variable finale et de la « bonne » chaîne de modèles dépend de l'enjeu industriel, et qu'à leur tour les sources d'incertitudes varient suivant le choix de modèles. **Cette question s'avère en pratique un point-clé quand la question des incertitudes est soulevée dans un nouveau domaine industriel.**

Dans l'exemple des crues, l'optimisation économique induit un changement de variable finale et une chaîne de modèle hydro-économique plus complète que la justification de sûreté, avec de nouvelles incertitudes; par ailleurs la sophistication supplémentaire du modèle physico-numérique (e.g. 3D au lieu de 2D ou 1D) multiplie souvent le nombre de sources d'incertitudes, sans toujours s'avérer plus adaptée pour supporter une justification réglementaire : à information de calibration constante, l'incertitude d'estimation risque plutôt de croître.

Comme les exemples l'ont montré, le « modèle physico-industriel » peut recouvrir une très grande variété de nature de modèle, depuis la formule analytique

simple jusqu'aux modèles physiques couplés à bases d'équations aux dérivées partielles non stationnaires avec post-traitements analytiques d'un critère à rupture, ou aux modèles intrinsèquement probabilistes de type sûreté de fonctionnement, ou encore à des modèles économiques ... Formellement il suffit qu'il relie, par une fonction déterministe $\underline{z} = \underline{G}(\underline{x}, \underline{d})$:

- des variables finales $\underline{z} = (z^l)_{l=1\dots r}$ importantes pour les enjeux (souvent en petit nombre, $r = 1$ à 5),
- à un certain nombre de variables ou paramètres d'entrée, continues ou discrètes $\underline{x} = (x^i)_{i=1\dots p}$ sur lesquels existent des aléas, imprécisions et généralement des sources d'incertitude ($p = 3$ à plusieurs centaines),
- et à d'autres variables considérées comme bien connues, rassemblées dans le vecteur \underline{d} , éventuellement optimisables ultérieurement : par exemple les conditions de fonctionnement ou d'essais de l'installation ou du produit destiné à être certifié. On peut également y adjoindre les variables d'entrée sujettes à des incertitudes dont l'importance est négligée, ou fixées par convention.

Le calcul du modèle physico-industriel $G(\cdot)$ en une valeur ponctuelle $(\underline{x}, \underline{d})$ peut demander un temps CPU très variable (de 10^{-4} s à quelques jours de calcul selon les cas!).

L'éventuelle phase de calage du modèle consiste à ajuster certaines variables sources en comparant certains résultats *intermédiaires* du modèle à des essais ou à des mesures : certaines des variables (ou degrés de liberté) calculables par le modèle – notées $\underline{y} = (y^k)_{k=1\dots q} = \underline{H}(\underline{x}, \underline{d})$ car souvent différentes des variables \underline{z} – sont typiquement observables (via un écart expérimental mesure-modèle \underline{u}) selon $\underline{y}_m = \underline{H}(\underline{x}, \underline{d}) + \underline{u}$ dans des conditions d'essais \underline{d} bien connues (éventuellement très différentes des conditions de fonctionnement \underline{d}' ultérieures). D'où le schéma général pour la **structure du modèle physico-industriel déterministe** en figure 7 :

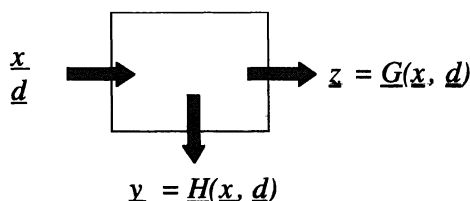


FIGURE 7. — Structure et notations pour le modèle physico-industriel déterministe (correspondant au problème industriel de la figure 6)

3.2. Plonger le modèle physico-industriel dans un « univers incertain » répondant bien aux critères décisionnels

Le modélisation des incertitudes conduit alors à plonger ce modèle déterministe dans un univers incertain, en **associant à chacune de ces variables**

$(\underline{x}, \underline{y}, \underline{z})$ des « mesures d'incertitude » $(\underline{X}, \underline{Y}, \underline{Z})$. Ces mesures d'incertitudes peuvent être, selon le choix du paradigme déterministe, probabiliste ou possibiliste : soit des intervalles, soit des variables aléatoires, soit encore des distributions non probabilistes (nombres flous, distributions possibilistes, de Dempster-Shafer etc., cf. [9], [32], [67]), comme l'indique le tableau 4.

Quelque soit le choix effectué à ce stade, quantifier la mesure d'incertitude de \underline{Z} supposera un arsenal mathématique et des outils numériques pour propager la mesure d'incertitude \underline{X} à travers le modèle physico-industriel, elle-même devant être quantifiée selon des paramètres $\underline{\theta}_X$ d'après toutes les informations disponibles (jugement d'expert, observations directes ou indirectes via \underline{Y}_m).

Dans l'exemple du cadre probabiliste standard, les variables sources \underline{x} évoluent dans un espace \mathcal{X} (généralement une partie de \mathbb{R}^p mais certaines composantes peuvent être discrètes) que l'on munit d'une structure d'espace probabilisé où $\underline{X} = (X^i)_{i=1\dots p} \sim f_X(\underline{X} = \underline{x}/\underline{\theta}_X)$, f_X désignant la densité de probabilité (pdf) jointe du vecteur des sources \underline{X} . Les paramètres de cette densité sont rassemblés dans le vecteur $\underline{\theta}_X$. Ces paramètres peuvent inclure par exemple les paramètres de position, d'échelle et de forme ou moments des lois marginales, les coefficients de corrélations, les paramètres d'un modèle de copule cf. [46], ou par extension les coefficients d'un modèle non-paramétrique à noyaux. Des mesures images sur les espaces \mathcal{Y} et \mathcal{Z} (généralement des parties de \mathbb{R}^q et \mathbb{R}^r) s'en déduisent, sous hypothèse de mesurabilité des modèles physico-industriels déterministes $H(\cdot)$ et $G(\cdot)$.

TABLEAU 4. — Paradigmes et nature des mesures et paramètres des incertitudes

	Déterministe	Probabiliste	Non probabiliste
Mesure d'incertitude ($\underline{X}, \underline{Y}$ ou \underline{Z})	intervalles min/max	loi jointe de probabilité	Ensemble de nombre flou, distribution de Dempster-Shafer [14], pdf inférieure et supérieure, distributions de possibilités ...
Paramètres des mesures d'incertitude ($\underline{\theta}_x, \underline{\theta}_y$ ou $\underline{\theta}_z$)	bornes ou facteurs de sécurité	paramètres de la loi jointe	paramètres du graphe des nombres flous, paramètres de la distribution de Dempster-Shafer, des densités de probabilités inférieure et supérieure...

Comme on l'a vu sur les exemples les paradigmes déterministe et probabiliste peuvent être formellement réunis dans une approche « mixte déterministe-probabiliste », la partie déterministe apparaissant formellement comme un conditionnement (partiel ou total) du probabiliste ; par ailleurs le § 3.2 de la 2^{de} partie montrera qu'il y a en fait plusieurs paradigmes probabilistes possibles, notamment par conditionnement hiérarchique dans un cadre bayésien.

La modélisation des incertitudes débouche ainsi sur l'évaluation d'un critère associé aux enjeux industriels : les critères cz sont formellement des quantités d'intérêt, comme un multiple du coefficient de variation ou une probabilité de dépassement de seuil, ou des tests, comme le fait que ce coefficient de variation soit inférieur à 3 %. Ces critères se déduisent tous de la mesure d'incertitude de la variable finale (ou du vecteur) \underline{Z} . Le tableau 5 reprend d'une manière systématique les différentes quantités d'intérêt ou tests pouvant constituer les critères d'une étude d'incertitudes. Ils ont été, pour la plupart, introduits dans les exemples ci-dessus :

TABEAU 5. — Exemples de critères envisageables pour une étude d'incertitudes

Type de critère	quantité d'intérêt	version test à seuil « absolu »	version test à seuil « comparatif »
valeur pénalisée	$z_{pn} = G(x_{pn}, \underline{d})$	$z_{pn} < z_s$	$z_{pn}(\underline{d}) < z_{pn}(\underline{d}^0)$
probabilité (conditionnelle) de dépassement de seuil	$P[Z = G(X / (X^i = x_{pn}^i)_{i=m+1...p}, \underline{d}) > z_s]$	$P[Z > z_s] < Pf_s$	$P[Z > z_s] < P[Z(\underline{d}^0) > z_s]$
quantile (conditionnel)	$z^\alpha = F_z^{-1}(\alpha)$ où $F_z(z) = P[Z = G(X / (X^i = x_{pn}^i)_{i=m+1...p}, \underline{d}) < z]$	$z^\alpha < z_s$	$z^\alpha(\underline{d}) < z^\alpha(\underline{d}^0)$
« incertitude absolue »	$inc_z = k \cdot \sqrt{\text{var}(Z)}$ ou $inc_z = \frac{\frac{1-\alpha}{2} - z^\alpha}{2}$	$inc_z < i_s$	$inc_z(\underline{d}) < inc_z(\underline{d}^0)$
« incertitude relative »	$\%inc_z = k \cdot \frac{\sqrt{\text{var}(Z)}}{E(Z)}$ ou $\%inc_z = \frac{\frac{1-\alpha}{2} - z^\alpha}{2 \cdot E(Z)}$	$\%inc_z < \%i_s$	$\%inc_z(\underline{d}) < \%inc_z(\underline{d}^0)$
espérance	$E(Z)$	Sert plutôt en optimisation $Min[E(Z(\underline{d}))]$ qu'en test à seuil	
largeur floue	intervalle flou sur Z	intervalle flou < seuil	intervalles flous disjoints
intervalle d'imprécision d'une proba de dépassement de seuil	$[Pf_{inf}, Pf_{sup}]$	intervalle de probabilité < probabilité absolue	intervalles de proba disjoints ou < intervalle de référence

On notera que le seuil « absolu » ou comparatif est souvent implicite, de sorte que l'on a tendance à réduire le critère à la seule quantité d'intérêt : « indiquer l'incertitude relative » sous-entend par exemple « montrer qu'elle est faible, ou moindre qu'ailleurs ».

Plus rarement, il arrive que le critère porte sur le vecteur \underline{z} : il associe alors logiquement les tests ci-dessus pour chaque composante, selon plusieurs variantes possibles comme par exemple $P[Z^1 > z_s^1 \text{ ou } Z^2 > z_s^2] < Pf_s$ ou $P[Z^1 > z_s^1 \text{ et } Z^2 > z_s^2] < Pf_s$. Dans cette situation, le vocabulaire fiabiliste parle d'un « événement de défaillance multiple », l'événement de défaillance étant la réunion ou l'intersection des $\{z^l \text{ t.q. } z_s^l - z^l < 0\}$ pour $l = 1, \dots, r$.

Mathématiquement, on pourra toujours écrire ces critères sous la forme d'une fonctionnelle de \underline{Z} , $c_z = \mathcal{F}[\underline{Z}]$, qu'il s'agisse de grandeurs simples $c_z = inc_z$, $c_z = E(\underline{Z})$... ou de tests $c_z = 1_{\{test(\underline{Z})\}}$, où $\{test(\underline{Z})\} = \{z_{pn}(\underline{d}) < z_{pn}(\underline{d}^0)\}$ ou $\{inc_z < i_s\}$... : en effet, dans le cas probabiliste par exemple, c_z est calculée à partir d'intégrales impliquant la densité jointe de \underline{Z} , comme :

$$\begin{aligned}
 inc_z &= k \cdot \sqrt{\text{var}(Z)} = k \cdot \sqrt{\int_z Z^2 f_Z(z) \cdot dz - \left(\int_z Z f_Z(z) \cdot dz\right)^2} \\
 &= k \cdot \sqrt{\int_{\underline{x}} G(\underline{x}, \underline{d})^2 f_{\underline{X}}(\underline{x}/\underline{\theta}_X) \cdot d\underline{x} - \left(\int_{\underline{x}} G(\underline{x}, \underline{d}) f_{\underline{X}}(\underline{x}/\underline{\theta}_X) \cdot d\underline{x}\right)^2}
 \end{aligned}$$

C'est donc bien une fonctionnelle de la loi jointe f_Z (ou plus généralement de la mesure d'incertitude \underline{Z}) de \underline{z} , et non pas une fonction déterministe de \underline{z} . Et cette fonctionnelle apparaît en fait, à travers la propagation de la mesure de \underline{X} dans le modèle physique $G(\cdot)$, comme une fonction (très complexe, mais

théoriquement déterministe) des paramètres $\underline{\theta}_X$ de la mesure d'incertitude de \underline{X} et des variables fixées \underline{d} :

$$c_z = \mathcal{F}[\underline{Z}] = \mathcal{F}[f_{\underline{Z}}] = C_z(\underline{\theta}_X, \underline{d}).$$

Le choix du paradigme de modélisation des incertitudes occupe un pan entier de la littérature sur l'incertitude et le risque (cf. [1], [2], [3], [31], [32], [33], [40], [43], [60], [61] ...). Cette discussion, de nature épistémologique mais aussi liée à la théorie de la décision, est souvent associée à des distinctions suivant la nature des incertitudes (cf. le § 3.3). Dans une perspective économique, elle fait notamment référence à la façon dont les acteurs peuvent appréhender un avenir incertain plus ou moins probabilisable, avec une appréhension contrastée de l'utilité (cf. [4], [36], [66] et § 3.1 de la 2^{nde} partie).

Dans les applications industrielles de calcul d'incertitudes, il s'agit d'abord :

- a) d'être cohérent avec ce qui est imposé par la réglementation :

Les exemples montrent que la réglementation et la pratique des dossiers associés imposent généralement le type de critère, qui s'avère souvent de nature mixte déterministe/probabiliste¹⁰. De plus, elles recommandent souvent des hypothèses spécifiques sur la description des incertitudes de certaines composantes X^i , comme une pénalisation déterministe forfaitaire sur l'incertitude de modèle, une analyse de sensibilité déterministe sur certains paramètres physiques mal connus, ou le choix d'un « scénario pénalisant » *etc.*

- b) d'être réaliste vis-à-vis de l'information disponible sur les sources \underline{x} et des contraintes de propagation vis-à-vis du modèle physico-industriel :

L'information sur certains X^i est parfois très pauvre : le support des valeurs *possibles* est lui-même mal connu (et donc a fortiori les valeurs *probables* en sont mal connues) : c'est le cas typiquement de l'incertitude de modèle. Dans ce cas la représentation probabiliste, qui s'applique dans une approche fréquentielle classique à des variables observables de façon répétée, peut apparaître inacceptable; même si théoriquement le paradigme bayésien permet de dépasser largement ce cadre, l'approche fréquentielle reste de bon aloi dans une démarche opérationnelle auprès d'autorités de contrôle¹¹ : une pénalisation déterministe peut être jugée préférable dans ce genre de cas. Néanmoins, le choix des bornes reste tout aussi problématique à justifier dans l'absolu.

Par ailleurs, une modélisation déterministe du vecteur \underline{x} , dans le cas général où le modèle $G(\cdot)$ n'a pas de raison d'être monotone ou linéaire, requiert pour être propagée rigoureusement une optimisation non linéaire

10. « assurer que l'agression externe de récurrence millénaire reste contenue par les parades (...) dans un scénario pénalisé » induit un critère mixte déterministe-probabiliste de type **(10)** (cf. supra), tandis qu'une formulation « (...) avec une confiance de 95 % » autorise un critère probabiliste du type de l'approche mixte du § 2.4

11. Il n'est pas évident de justifier par exemple l'emploi d'un tirage aléatoire sur l'erreur numérique de résolution du modèle, ou sur les approximations de maillage, même si cela est proposé théoriquement par certains (cf. [49]).

à plusieurs variables. Cela pose rapidement un challenge numérique supérieur à celui de la simulation probabiliste (cf. 2.3 de la 2nde partie).

Les paradigmes «extra-probabilistes», comme la théorie des possibilités ou la théorie Dempster-Shafer (DST) présentent une grande souplesse pour aller au-delà des limites du calcul probabiliste quand l'information est très limitée et hétérogène ([14], [22]). Néanmoins, ils conduisent souvent à des besoins de calculs beaucoup trop lourds : ils cumulent typiquement des calculs d'optimisation déterministe à de la simulation probabiliste, ce qui multiplie immensément le nombre d'appels au code physique quand \underline{x} est de dimension importante. Par ailleurs, les mesures d'importance pour hiérarchiser les sources d'incertitudes y sont encore peu développées. Enfin les résultats paraissent plus délicats à communiquer à des autorités réglementaires.

Dans beaucoup d'applications industrielles récentes, le paradigme est en fait «mixte déterministe-probabiliste» (de type (11) *supra*) ; les exemples montrent qu'il offre déjà des choix assez puissants et contrastés de représentation des incertitudes, en choisissant si besoin composante X^i par composante le mode de traitement (déterministe ou probabiliste).

3.3. La question de la nature des incertitudes

Il est classique dans la littérature de mettre en rapport le choix de modélisation des incertitudes avec une typologie des sources d'incertitudes, qui distinguera par exemple (cf. [1], [29], [31], [43]) :

- «l'aléatoire» de «l'épistémique» (*lack of knowledge*), selon le cadre irréductible ou réductible de l'incertitude, vis-à-vis de l'augmentation de l'information / la connaissance. On peut détailler encore en distinguant parmi celles qui sont théoriquement réductibles, celles qui sont *industriellement* irréductibles pour des contraintes opérationnelles ou financières données.
- la «variabilité» par opposition à «l'incertitude» : cette distinction, assez connexe à la précédente, est souvent utilisée quand le problème «mélange» en entrée une population d'objets (ou de scénarios) ou une répartition spatiale ou temporelle de propriétés dans un objet. Dans ce cas, la variation des propriétés d'un objet (respectivement d'un point ou d'un instant) à l'autre au sein de la population est distinguée de l'éventuelle incertitude attachée aux propriétés d'un objet (*resp.* en un point ou un instant donné).
- «l'incertitude épistémique» de «l'erreur» selon que la méconnaissance est inévitable ou «délibérée» malgré la disponibilité de la connaissance. Par exemple on peut «choisir» d'utiliser un modèle plus simple ou un schéma numérique plus rapide mais moins précis. Plus insidieusement, on peut laisser passer des erreurs d'études involontaires.
- L'incertitude «paramétrique», ou associée aux «données» d'un modèle considéré comme représentatif, de l'incertitude «de modèle».

Certains réservent théoriquement par exemple le traitement probabiliste aux seules incertitudes « aléatoire » et « paramétrique ».

Les exemples industriels montrent que :

- ces distinctions apparaissent naturellement dans une étude et sont utiles pour comprendre : distinguer par exemple l'aléatoire du débit de crue ou de l'incertitude d'un capteur, de l'épistémique associée à l'estimation statistique du débit millénal et de la variance du capteur. Pour autant, elles ne sont pas toujours simples à discerner quantitativement sur certaines sources compte tenu de l'information très limitée réellement disponible (e.g. l'état de la rivière). Dans la réalité physique, on n'est jamais dans une situation aussi idéalisée que celle d'une urne à boules de couleur de proportions connues (resp. inconnues), qui est l'un des exemples classiques « d'incertitude probabilisable » (resp. de méconnaissance) en théorie de la décision : les proportions sont partiellement connues, et il n'est même jamais certain, faute d'observations, que l'on tire effectivement de manière répétitive dans une même population.
- dans le cas extrême, ce qui paraît très clairement aléatoire, comme les variables météorologiques¹², est généralement traité d'un point de vue probabiliste. Même dans une approche dite « déterministe », les valeurs pénalisées seront adossées à des quantiles. Ce qui est « radicalement méconnu » est traité de façon déterministe, comme par exemple l'incertitude de modèle associée à des séquences très rares en analyse de risque. Mais dans les cas plus mélangés, comme l'incertitude d'échantillonnage, qui est de « l'épistémique quantifiable », la pratique réglementaire combine ou alterne déjà couramment des traitements de nature probabiliste et de nature déterministe sans se restreindre d'une manière absolue et définitive à un type unique de traitement par nature d'incertitudes.
- ces distinctions dépendent de la nature de l'objet et du modèle physico-industriel considéré :
 - o ainsi dans le cas de l'analyse de risque, même en l'absence de traitement d'incertitudes, le modèle $G(\cdot)$ représente déjà l'aléatoire : on est tenté de concevoir que toute l'incertitude qui s'y attache est en fait épistémique,
 - o sur l'exemple de la crue, selon que la finalité est la justification de sûreté conventionnelle ou l'optimisation économique, l'incertitude sur la composante associée au débit représente de « l'épistémique » (autour d'un aléa conventionnel) ou de « l'aléatoire » (avec toute la distribution des crues possibles).
- l'incertitude « de modèle » est incluse au même titre que les incertitudes « paramétriques » en mécanique des fluides, ou plus généralement quand il y a du calage de modèle physique par identification de paramètres de

12. Mais les probabilités ne restent-elles pas, même dans ce cas, plutôt un *modèle* jugé satisfaisant par expérience qu'une propriété de la nature ? Y-a-t il en effet assurance que la « mécanique » sous-jacente, que certains identifieront au chaos déterministe, satisfasse une axiomatique de Kolmogorov ? (cf. par exemple [27]).

modélisation sur des essais ; en analyse de risque, elle en est généralement séparée.

- quant à la variabilité, sa représentation dépend surtout de la façon de définir le critère d'incertitude : par exemple, « à 95 % de confiance l'individu *le pire* vérifiera la propriété P » (resp. « parmi la population, 95 % *en moyenne* vérifieront P ») exclut (resp. inclut) la variabilité vraie du périmètre de probabilisation de l'incertitude associée à un individu donné.

Les distinctions *a priori* selon la nature d'incertitudes sont donc essentielles pour comprendre et identifier les variables sources d'incertitudes. Néanmoins, elles ne nous apparaissent pas encore comme une base opératoire systématique pour définir, de façon générique sur les applications concrètes, la façon de modéliser les incertitudes. La recherche en ce sens doit être poursuivie (cf. [32], les résultats de l' « Epistemic Uncertainty Project » des Sandia National Laboratories, USA).

Une autre catégorisation nous apparaît en revanche incontournable dès lors qu'on s'engage dans une étude de traitement de l'incertitude. Elle est proposée dans le tableau 6.

TABLEAU 6. — Catégorisation des incertitudes relative à leur traitement dans une étude donnée

Catégories d'incertitudes liées au traitement	Grandeur mathématique concernée	Exemples
1. Sources d'incertitudes (ou incertitudes « d'entrée »)	\underline{X} (à θ_X donné)	Crue : Q_{1000} dans l'approche mixte mais Q dans l'approche directe ; Métrologie : incertitude du capteur
2. Incertitude d'estimation (des sources)	incertitude sur l'estimation θ_X par $\hat{\theta}_X$	incertitude d'estimation de la loi de Q dans l'approche directe ; aucune dans l'approche mixte. Métrologie : incertitude d'estimation de la variance d'un capteur
3. Incertitude d'observation (des sources)	μ bruitant l'observation des \underline{Y}	incertitude sur la mesure de hauteur, utilisée pour caler le modèle de crue
4. Incertitude de propagation (des sources au sein de $G(\cdot)$)	incertitude sur l'estimation pratique de c_z par \hat{c}_z (à θ_X donné)	incertitude sur l'estimation de la probabilité de surverse due à la simulation MCS (à n fini), à l'approximation Form etc...

Les sources d'incertitudes ou *incertitudes d'entrée* portant sur les variables \underline{x} sont représentées par la mesure d'incertitude \underline{X} : elles sont au cœur de l'étude. Les exemples ont montré que l'on peut (au moins formellement) y inclure toutes les incertitudes affectant le modèle de l'objet physico-industriel, qu'elles soient paramétriques ou de modèle, épistémique ou aléatoire¹³ etc. et quelque soit d'ailleurs leur mode de traitement déterministe ou probabiliste. Le périmètre couvert par les incertitudes d'entrée est lié au choix de modèle

13. ... sauf bien entendu les incertitudes qui sont l'objet du modèle lui-même comme l'aléa en analyse de risque.

et de variable d'entrée x : dans l'exemple des crues, l'incertitude d'estimation du q_{1000} en fait partie dans l'approche « mixte » (pour laquelle q_{1000} est une variable d'entrée) mais non pas dans l'approche « directe » (pour laquelle seule q est une variable d'entrée).

Modéliser les incertitudes sur x par une mesure X induit néanmoins inévitablement une incertitude « de second niveau », l'incertitude d'estimation des paramètres θ_X de l'incertitude X , car l'information d'estimation est toujours limitée. Dans un cadre statistique, c'est un intervalle de confiance d'estimation des paramètres de la loi jointe. Dans un cadre déterministe, c'est le choix incertain des bornes d'incertitudes. Même si cette catégorie est plus subtile et souvent implicite, elle est parfois clairement identifiée dans les normes, comme dans le facteur d'élargissement en métrologie. Naturellement l'estimation statistique de l'incertitude d'un estimateur reste elle-même souvent approximative parce qu'elle est par exemple estimée dans un cadre asymptotique : cela induit théoriquement un niveau d'incertitudes supplémentaire, lui-même à nouveau estimé *etc.* On se bornera au premier niveau, qui est déjà mentionné dans certains critères réglementaires très avancés, tout en gardant en mémoire qu'il pourrait lui-même être théoriquement re-développé. Mais il est en pratique incontournable de s'arrêter à un certain niveau.

Par ailleurs, les données servant à l'estimation des mesures X , par voie directe ou par recalage du modèle, sont éventuellement bruitées par de l'erreur de mesure liée au capteur (indépendamment du modèle $G(.)$: elles ne sont donc pas des *incertitudes d'entrée* de $G(.)$). Bien que distincte de la deuxième catégorie de par son origine, cette troisième catégorie pourrait en pratique être fusionnée avec la deuxième parce qu'elle concourt *in fine* à perturber l'estimation des θ_X .

Enfin, une dernière catégorie apparaît dans la mise en œuvre des méthodes de traitement des incertitudes. Le calcul d'un critère c_z passe en effet sur des cas réels par de fréquentes approximations de calcul, de nature numérique ou probabiliste suivant le mode de propagation. Cette *incertitude de propagation* n'a rien à voir avec les autres catégories en termes *décisionnels*. Le § 2.4.1 de la 2nde partie montrera que son *estimation numérique* pourra passer par les mêmes algorithmes statistiques que la seconde catégorie d'incertitudes (calcul asymptotique de l'intervalle de confiance, calcul à distance finie par la formule de Wilks ...). Comme la seconde catégorie, l'explicitation réglementaire de cette dernière catégorie peut conduire à formuler certains critères en introduisant un second niveau de probabilité, comme le quantile 95 % (ou millénal), à 95 % (ou 70 %) de confiance. Néanmoins les expériences aléatoires sous-jacentes sont très différentes : pour la seconde catégorie, il s'agit d'une variabilité induite par l'observation des données ou de l'expertise ; pour cette dernière catégorie, il s'agit d'une variabilité induite par les expériences numériques de calcul de propagation.

Ces catégories nous semblent apparaître dans tout processus de traitement d'incertitudes, quelque soit d'ailleurs le paradigme choisi. On verra que pour certains critères très exigeants la question du cumul de ces catégories d'incertitudes sur le résultat final reste ouverte (*cf.* § 3.2 de la 2nde partie).

Les exemples montrent que la correspondance entre cette catégorisation et la typologie classique n'est pas univoque, mais qu'elle dépend des choix de l'étude elle-même quant à l'objet physico-industriel, au modèle et aux variables d'entrée considérées. Notons par exemple que :

- les trois dernières catégories relèvent par exemple toujours de « l'épistémique » (ou de « l'erreur »),
- de « l'épistémique » peut aussi se retrouver incluse dans la première catégorie, mélangée avec de « l'aléatoire » selon les choix de l'étude,
- l'incertitude « de modèle » peut se retrouver dans la première ou la deuxième catégorie ou bien être exclue des incertitudes que l'on quantifie.

3.4. Les étapes essentielles d'une étude d'incertitudes

En résumé, l'étude d'incertitudes pour un objet physico-industriel donné comprend les étapes essentielles suivantes (cf. Fig. 8) :

A/ la spécification du problème : le choix de la variable finale (éventuellement vectorielle) z , de l'enchaînement logique des données et modèles adaptés pour les évaluer, du critère décisionnel, et du paradigme de modélisation des incertitudes.

B/ la quantification (ou modélisation) des sources d'incertitudes : la détermination des mesures d'incertitudes X par l'estimation de leurs paramètres θ_X , qui peut impliquer deux voies différentes :

B : *quantification directe* par expertise et/ou traitement d'observations directes

B' : *quantification indirecte* quand les observations ou l'expertise portent sur d'autres variables, par méthodes inverses. Cette étape peut être aussi assimilée à la phase de *calage/qualification* du modèle, ou, dans un paradigme déterministe, à de l'*identification de paramètres*.

C/ la « propagation d'incertitudes » : le calcul de la mesure d'incertitude Z de la variable finale z et l'estimation des critères c_z .

C'/ la hiérarchisation d'importance des sources d'incertitudes (ou « analyse de sensibilité ») : le classement de l'importance des composantes X^i des sources vis-à-vis des critères c_z .

Cette classification séquentielle des étapes a été proposée essentiellement à titre *pédagogique* (cf. Tutoriel [16]) pour clarifier l'expression des besoins industriels. En réalité, les aller-retours sont essentiels entre ces différentes étapes, par exemple parce que la hiérarchisation d'importance des (innombrables) sources d'incertitudes redonne des priorités pour l'étape B, voire conduit à modifier le modèle et son calage (B') avant de re-propager (C), etc.

Selon les métiers, d'autres termes sont rencontrés. Pour la phase C par exemple, on rencontrera les formulations suivantes : « propagation d'erreurs », « calcul d'incertitude », « cumul d'incertitudes », « calcul probabiliste », « calcul

fiabiliste» ou «couplage mécano-fiabiliste» en fiabilité des structures, dans la mesure où l'algorithme de propagation (par exemple MCS ou Form) doit être couplé au modèle physique (mécanique). Quant au terme «analyse de sensibilité», nous ne le recommandons pas, bien qu'il soit très courant notamment dans la littérature anglaise où *uncertainty analysis* et *sensitivity analysis* désignent à peu près les phases C et C'. L'usage en est en effet très ambigu dans le monde industriel, dans la mesure où il peut se référer à au moins trois notions distinctes :

- (i) à la sensibilité du modèle physico-industriel à une variation déterministe des sources x c'est-à-dire $\Delta G(\underline{x}, d)/\Delta x_i$,
- (ii) au résultat de la phase C', c'est-à-dire l'une des mesures d'importance $C_z(inc_{X_i}/inc_z)$ de la source X_i selon c_z ,
- (iii) à la sensibilité du critère c_z à une variation des paramètres θ_X de la mesure d'incertitude des sources, c'est-à-dire $\Delta C_z(\theta_X, d)/\Delta \theta_X^i$.

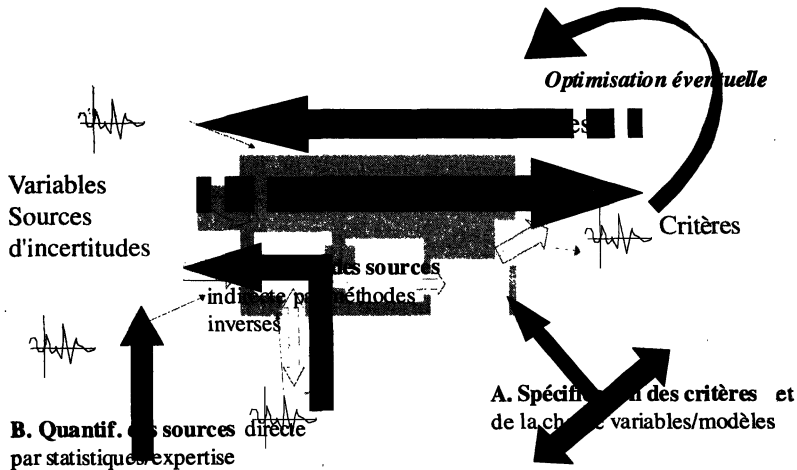


FIGURE 8. — Grandes étapes d'une étude d'incertitudes

La Figure 8 positionne, en référence à la Figure 6, les différentes étapes d'une étude d'incertitudes en surimpression du cadre général d'un problème industriel. Le tableau 7 remet en perspective les quatre finalités généralement associées à une étude, et le séquençement des étapes associées.

Dans certaines phases industrielles, il arrive qu'après une première étude complète de référence, une simplification opérationnelle soit nécessaire pour décliner ultérieurement l'étude sur un grand nombre de scénarios de conception ou d'exploitation de la même famille, ce qui consiste formellement à faire varier les variables \underline{d} . On substituera aux phases B et C, coûteuses et affaires de spécialistes, un unique calcul déterministe moyen, souvent qualifié de «best-estimate», que l'on comparera à un critère décisionnel synthétique. Ce critère, généralement de type «marge déterministe enveloppe», aura été

défini d'après l'étude complète de référence. La pertinence de cette simplification, que l'on trouve par exemple en thermo-hydraulique nucléaire ou dans les Eurocodes de calcul des structures (« coefficients (partiels) de sécurité »), dépend bien entendu de la similarité réelle des problèmes. Le choix d'un critère déterministe *enveloppe* couvrant une large gamme de problèmes peut induire des « conservatismes » implicites, c'est-à-dire la surestimation des incertitudes.

TABLEAU 7. — Finalités généralement associées à une étude d'incertitudes, et étapes à réaliser

Finalité	Résultat formel attendu	Etapes à réaliser (a minima)	Commentaires
Justification	Assurer le respect d'un critère $C_z(\theta_x, \underline{d}) = 1$ où par exemple : $C_z(\theta_x, \underline{d}) = (P_f < 10^6)$	A puis B (et B') puis C ce que l'on notera : $A > B$ (et $B') > C$	Sur de grandes études, le processus peut être cyclique : $A > B$ élémentaire $> C > C' > B$ affiné ... $> C$
Hierarchisation	Ranger l'importance des X_i par rapport à c_z	$A > B$ (et $B') > C > C'$	Idem
Optimisation	Optimiser une partie des vecteurs \underline{d} ou θ_x suivant c_z i.e. $Min_{\underline{d}, \theta_x} \{C_z(\theta_x, \underline{d})\}$ ¹⁴	$A > B$ (et B') puis optimiser \underline{d} par itérations de C	Idem, avec re-quantification des incertitudes qui peuvent éventuellement changer selon \underline{d}
Validation	Contrôler $Y(\theta_x) - Y_m$ en ajustant éventuellement des composantes de θ_x	$A > B > C > B'$	Souvent des cycles $A > B > C > C'$ avec plusieurs modèles pour choisir le bon modèle

4. Conclusions de la première partie

Le propos de la première partie de cet article était de montrer par l'exemple industriel qu'un cadre générique d'appréhension des incertitudes est possible, malgré les compartimentages historiques entre les domaines impliqués (métrologie, fiabilité, statistiques, analyse numérique, ...). Il consiste notamment à discerner de grandes étapes comme la quantification des sources d'incertitudes (étape B), mobilisant éventuellement une quantification inverse/validation (B'), leur propagation à travers un modèle physico-industriel pré-existant (étape C), la hiérarchisation d'importance (C') voire l'optimisation qui en découle. Nous montrons que le choix des méthodes les plus pertinentes, qui associent de façon étroite les mathématiques appliquées et l'analyse physico-industrielle, doit être avant tout pensé en fonction du problème posé, qui dépend de la nature de la réglementation ou du critère de décision choisi (étape A), et non pas du « métier » spécifique du problème industriel. Si dans cette même approche globale d'autres paradigmes mathématiques sont possibles, le cadre mixte déterministe-probabiliste apparaît central dans les applications industrielles actuelles : de nombreux problèmes de modélisation statistique et de calcul scientifique en découlent, comme l'exposera la seconde partie de cette article.

14. Des programmes plus complexes peuvent bien sûr impliquer de minimiser un critère C_{z1} sous contrainte d'un autre C_{z2} , e.g. minimiser l'espérance du coût de la crue sous contrainte que la probabilité de défaillance reste au moins inférieure à 10^{-3}

5. Abréviations, sigles et acronymes utilisés

CPU	Central Processing Unit – en référence au temps de calcul d'une machine informatique
D.L.1	Développement limité d'ordre 1
E.P.S.	Etudes probabilistes de Sûreté – études probabilistes de type « fiabilité des systèmes » d'une centrale nucléaire vis à vis du risque de sûreté nucléaire
EVD	Extreme Value Distribution – famille de distributions associée à une méthode d'estimation de variables extrêmes de type « maxima par blocs »
FAST	Fourier Amplitude Sensitivity Test – méthode de hiérarchisation d'importance des sources d'incertitudes basée sur la transformée de Fourier
FE	Facteur d'émission
Form/Sorm	First (respectivement Second) Order Reliability Method – méthode d'approximation du premier (resp. second) ordre d'une probabilité de dépassement de seuil
GPD	Generalised Pareto Distribution – famille de distributions associée à une méthode d'estimation de variables extrêmes de type « pic au-dessus de seuil »
GUM	Guide for expression of uncertainty of measurement – Guide pour l'expression de l'incertitude de mesure – norme ISO/AFNOR en métrologie
I.C.	Intervalle de confiance
i.i.d.	Indépendants et identiquement distribués
LHS	Latin Hypercube Sampling – méthode de tirage aléatoire suivant les hypercubes latins
MCS	Monte-Carlo Sampling – méthode de simulation de Monte-Carlo (en version simple)
MV	(Méthode du) Maximum de vraisemblance
PCI	Pouvoir calorifique inférieur
pdf	Probability distribution function – densité de probabilité d'une variable (ou d'un vecteur) aléatoire
v.a.	Variable aléatoire

6. Principales notations

$\underline{x} = (x^i)_{i=1\dots p}$	Vecteur (de dimension p) des variables d'entrée sources d'incertitudes (continues ou discrètes)
\underline{d}	Vecteur (de dimension d) des variables d'entrée fixes, car considérées comme bien connues (continues ou discrètes) : scénarios de référence, conditions de fonctionnement ou d'essais, variables de décision, sources d'incertitudes d'importance négligée, <i>etc.</i>
\underline{d}°	Vecteur de variables fixes dont la valeur est prise dans un scénario de référence.
$\underline{z} = (z^l)_{l=1\dots r}$	Vecteur (de dimension r) des variables finales d'intérêt (continues ou discrètes) pour l'étude d'incertitudes, intervenant dans la définition du critère
$\underline{y} = (y^k)_{k=1\dots q}$	Vecteur (de dimension q) des variables intermédiaires calculées par le modèle, et observables modulo un bruit de mesure éventuel
$\underline{u}, \underline{u}_{mod}, \underline{u}_{mes}$	Vecteurs (de dimension q) représentant l'écart modèles-mesure à l'observation des variables intermédiaires : \underline{u} représente l'écart total, \underline{u}_{mes} , la composante éventuelle liée au bruit de mesure, \underline{u}_{mod} , la composante éventuelle d'erreur de modèle : $\underline{u} = \underline{u}_{mod} + \underline{u}_{mes}$
$\underline{y}_m = (\underline{y}_m^k)_{k=1\dots q}$	Vecteur (de dimension q) des variables intermédiaires observables. On a $\underline{y}_m = \underline{y} + \underline{u}$
$\underline{G}(\cdot)$	Fonction déterministe, éventuellement très complexe, représentant le modèle physico-industriel vu des variables finales d'intérêt : $\underline{z} = \underline{G}(\underline{x}, \underline{d})$. On allègera fréquemment la notation en écrivant $G(\cdot)$
$\underline{H}(\cdot)$	Fonction déterministe, éventuellement très complexe, représentant le modèle physico-industriel vu des variables intermédiaires : $\underline{y} = \underline{H}(\underline{x}, \underline{d})$. On allègera fréquemment la notation en écrivant $H(\cdot)$
$\underline{X}, \underline{Y}, \underline{Z}, \underline{U} \dots$	Vecteurs de variables incertaines munis de leur mesure d'incertitude correspondants aux vecteurs $\underline{x}, \underline{y}, \underline{z}, \underline{u} \dots$. Dans le cadre probabiliste, ce sont des vecteurs aléatoires. En notations allégées, le soulignement est omis : $X, Y, Z, U \dots$
$f_X(\underline{x}/\underline{\theta}_X)$	Densité jointe du vecteur aléatoire \underline{X} , paramétrée par $\underline{\theta}_X$. En notation allégée, le vecteur de paramètre est omis : $f_X(\underline{x})$.
$F_Z(z)$	Fonction de répartition de la variable scalaire z .
$(\underline{X}_j)_{j=1\dots n}$	Echantillon (de taille n), généralement i.i.d., de vecteurs aléatoires suivant la loi de \underline{X}

$(\underline{Y}_{mj})_{j=1\dots n}$	Echantillon (de taille n), non nécessairement i.i.d., de vecteurs aléatoires d'observations \underline{y}_m sur n essais
$(\underline{d}_j)_{j=1\dots n}$	Matrice des n conditions d'essais correspondant aux observations $(\underline{Y}_{mj})_{j=1\dots n}$
inc_z	Quantité d'intérêt sur une variable finale scalaire z représentant une dispersion absolue autour de la valeur moyenne (un multiple donné de l'écart-type)
$\%inc_z$	Quantité d'intérêt sur une variable finale scalaire z représentant une dispersion relative autour de la valeur moyenne (un multiple donné du coefficient de variation)
z_s	Valeur représentant un seuil à ne pas dépasser par z (scalaire) dans un critère à seuil
$\underline{\theta}_X, \underline{\theta}_U$	Vecteurs de paramètres de la mesure d'incertitude de \underline{X} ou de \underline{U} : dans le cas probabiliste, ce sont les paramètres d'une loi jointe. En notations allégées, on se contentera de noter θ_X, θ_U
$\underline{\theta}_X^{kn}, \underline{\theta}_X^{un}$	Vecteurs représentant les composantes connues (kn) et inconnues (un) du vecteur $\underline{\theta}_X$, dans une approche inverse probabiliste.
c_z	Critère, c'est-à-dire une quantité scalaire ou vectorielle, de composantes entières ou réelles, définie comme une fonctionnelle de la mesure d'incertitude des variables finales $c_z = \mathcal{F}[\underline{Z}]$. Par exemple : la dispersion inc_z , le test booléen comparant cette dispersion moyenne à un seuil, la probabilité de dépassement d'un seuil <i>etc.</i>
$C_z(\underline{\theta}_X, \underline{d})$	Fonction déterministe (très complexe) associant aux paramètres de la mesure d'incertitudes des sources et aux variables fixes le résultat égal au critère $c_z = C_z(\underline{\theta}_X, \underline{d})$

Pour les variables : l'exposant désigne la composante du vecteur (lettres latines : x^i) ou le niveau d'un quantile (lettres grecques : x^α) ; l'indice désigne la place dans l'échantillon (x_j) sauf cas particulier où il dénote une valeur spéciale (e.g. pénalisée : x_{pn} , seuil : z_s) ; la majuscule (X) désigne la variable (ou le vecteur) aléatoire (ou plus généralement incertain) par opposition à la valeur déterministe, notée en minuscule (x) ; le soulignement (\underline{x}) désigne une quantité vectorielle, mais celui-ci est omis dans les notations allégées. Un chapeau (\hat{C}_z) dénote classiquement un estimateur associé à une quantité déterministe inconnue (c_z).

7. Références

- [1] APOSTOLAKIS G. (1999), *The distinction between aleatory and epistemic uncertainties is important; an example from the inclusion of aging effects in the PSA*, PSA'99, Washington DC.
- [2] AVEN T. (2003), *Foundations of Risk Analysis*, Wiley.
- [3] BECK M.B.(1987), *Water Quality Modeling : A Review of the Analysis of Uncertainty*, Wat. Res. Research, Vol. 23, n° 8.
- [4] BEAUDOUIN F., MUNIER B., SERQUIN Y. (1997), *Multiattribute utility theory : Towards a more general framework*, Proc. of the 1997 ESReDA seminar on decision analysis and its applications in safety and reliability.
- [5] BESTION D. (2003), *Final Report -§ 4.2 State of the Art on Uncertainty Evaluation*, European Project for Future Advances in Sciences and Technology for Nuclear Engineering Thermal-Hydraulics (EUROFASTNET).
- [6] BOYACK B.E. (1990), *Quantifying reactor safety margins-part I : an overview of the code scaling, applicability and uncertainty evaluation methodology*, Nucl. Eng. Des. 119, 1-15.
- [7] BUSSAC R., BAILLY P., THOMAS Ph., de ROCQUIGNY E. (2004), *CO₂ emissions : estimating uncertainties in practice for power plants*, "Implementation of European Environmental Regulation in Fossil-Fired Power Plant" of EDF Group" Seminar Sep. 2004, Gdansk, Polska.
- [8] CACUCI D.G., *et al.* (1980), *Sensitivity Theory for General Systems of Nonlinear Equations*, Nucl. Sc. & Eng. 75.
- [9] CAMBIER S. (2000), *Approche probabilité pour la prise en compte de la dispersion de paramètres mécaniques - Application à la fatigue vibratoire de réseaux de tuyauteries*, Thèse ENSAM Paris.
- [10] CAMBIER S., GUIHOT P., COFFIGNAL G. (2002), *Computational methods for accounting of structural uncertainties, applications to dynamic behavior prediction of piping systems*. Structural Safety, 24 : 29-50.
- [11] CORRE B. (2003), *Problématique des incertitudes en exploration-production pétrolière : application des plans d'expériences*. Total Fina Elf, Séminaire «Plans d'Expériences Numériques», Ec. Mines St-Etienne.
- [12] CUKIER R.I., LEVINE H.B., SCHULER K.E. (1978). *Non-linear sensitivity analysis of multi-parameter model systems*. Journal of computational physics 26, p. 1-42.
- [13] DACUNHA-CASTELLE D. (1996). *Chemins de l'aléatoire. Le hasard et le risque dans la société moderne*. Flammarion, Paris.
- [14] DEMPSTER A. P. (1967). *Upper and lower probabilities induced by a multivalued mapping*, Annals of Math. Statistics 38 : 325-339.
- [15] DE ROCQUIGNY E. (2005), *Couplage mécano-probabiliste pour la fiabilité des structures - un cas industriel où la robustesse d'une surface de réponse est démontrable*. Actes du 17^{ème} Congrès Français de Mécanique. Troyes.
- [16] DE ROCQUIGNY E. (2004), *Tutoriel Incertitudes*, présenté dans la Conférence Lambda-Mu 14, Bourges, Octobre 2004.
- [17] DEVICTOR N. (2004), *Intérêt des méthodes statistiques pour la prise en compte des incertitudes dans les processus de décision*, Phoebus N°31.
- [18] DEVICTOR N. (1996), *Fiabilité et mécanique : méthodes FORM/SORM et couplages avec des codes d'éléments finis par surfaces de réponse adaptative*. PhD Université Blaise Pascal, Clermont-Ferrand.

- [19] DEVICTOR N., BOLADO LAVIN R. (eds) (2005), *Proc. of the Workshop on the Use of Expert Judgment in Decision-Making*, CEA & Eur. Comm. JRC Petten, Aix-en-Provence.
- [20] Directive 2003/87/EC of the European Parliament and of the Council of 13 October 2003 establishing a *scheme for greenhouse gas emission allowance trading within the Community & Commission* – Decision of 29 January 2004 (C(2004) 130) establishing *guidelines for the monitoring and reporting of greenhouse gas emissions pursuant to Directive 2003/87/EC*.
- [21] DITLEVSEN O. & MADSEN H.O. (1996), *Structural reliability Methods*, John Wiley & Sons.
- [22] DUBOIS D., PRADE H. (1988), *Possibility theory : an approach to computerized processing of uncertainty*, New York, Plenum Press.
- [23] DUCKSTEIN L., PARENT E. (1994), *Engineering Risk in Natural Resources Management*, NATO ASI Series.
- [24] DUTFOY A. (2000), *Uncertainty propagation in radionuclide transport modelling for performance assessment of a nuclear waste repository*, Proc. of PSAM5, Japan.
- [25] DUTFOY A., RITZ J.B. (2001), *Uncertainty propagation in a 3D thermal code for performance assessment of a nuclear waste disposal*, Proceedings of the SIAM'6 Conference on Mathematical and Computational Issues in the Geosciences, Boulder, Colorado.
- [26] EFRON B., TIBSHIRANI R.J. (1993), *An introduction to the Bootstrap*, Chapman & Hall.
- [27] EKELAND I. (1991), *Au Hasard – La chance, la science et le monde*, Seuil.
- [28] FREY H.C., RHODES D.S. (2005), *Quantitative Analysis of Variability and Uncertainty in with Known Measurement Error : Methodology and Case Study*, Risk Analysis Vol. 25, No3.
- [29] GRANGER MORGAN M., HENRION M. (1990), *Uncertainty – A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge University Press.
- [30] HELTON J. (1993), *Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal*, Rel. Eng. & Syst. Saf., 42, 327-367.
- [31] HELTON J.C., BURMASTER D.E. et al. (1996), *Treatment of Aleatory and Epistemic Uncertainty*, Special Issue of Rel. Eng. & Syst. Saf., vol. 54 n°2 et 3.
- [32] HELTON J.C., OBERKAMPF W.L. (2004), *Alternative Representations of Epistemic Uncertainty*, Special Issue of Rel. Eng. & Syst. Saf., vol. 85 n°1-3.
- [33] HOFFMAN F.O., HAMMONDS J.S. (1994), *Propagation of Uncertainty in Risk Assessments : the Need to Distinguish between Uncertainty Due to Lack of Knowledge and Uncertainty due to Variability*, Risk Analysis, Vol 14. n°5.
- [34] INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (IPCC) 2000, *Good practice guidance and Uncertainty Management in National Greenhouse Gas Inventories*.
- [35] KENDALL MG. & STUART (1943-1979), A, *The Advanced Theory of Statistics* (2 vol), Griffin & Co., London.
- [36] KNIGHT F.H. (1921), *Risk, Uncertainty and Profit*, Hart, Schaffner & Marx.
- [37] KUSCHEL N., RACKWITZ R. (2000), *Optimal design under time-variant reliability constraints*, Structural Safety, Volume 22, Issue 2, p. 113-127.

- [38] LANNON A., *et al.* (1994), *Méthodes avancées d'analyse des bases de données du retour d'expérience industriel*, Coll. D.E.R.EDF, n°86, Eyrolles.
- [39] LEBRUN R. (2006), *Modelling dependency with copulas in reliability analysis, a new approach to the FORM and SORM methods*, submitted to 8th PSAM Conference.
- [40] LEMMER J.F., KANAL L.N. (eds) (1986), *Uncertainty in Artificial Intelligence*, Elsevier.
- [41] LENCASTRE A. (1999), *Hydraulique générale*, Eyrolles.
- [42] MCKAY M.D. (1996), *Application of Variance-Based Methods to NUREG-1150 Uncertainty Analyses*, USNRC 1996.
- [43] MCKONE T.E. (1994), *Uncertainty and Variability in Human Exposures to Soil Contaminants Through Home-Grown Food : A Monte-Carlo Assessment*, Risk Analysis Vol 14, N°4.
- [44] MERRICK J.R., *et al.* (2005), *Assessing uncertainty in Simulation-Based Maritime Assessment*, Risk Analysis Vol. 25, N°3.
- [45] MIQUEL J. (1984), *Guide pratique d'estimation des probabilités de crues*, Eyrolles.
- [46] NELSEN R.B. (1999), *An introduction to copulas*, SPRINGER.
- [47] NILSEN T., AVEN T. (2003), *Models and model uncertainty in the context of risk analysis*, Rel. Eng. Sys. Saf. 79, 309-317.
- [48] Norme AFNOR XP X 07-020 (1996) : *Normes Fondamentales : Guide pour l'expression de l'incertitude de mesure* – AFNOR (1999) : *Guide pour l'expression de l'incertitude de mesure*, NF ENV 13005.
- [49] OBERKAMPF W.L. *et al.* (2002), *Error and uncertainty in modelling and simulation*, Rel. Eng. & Sys. Saf., 75.
- [50] PATÉ-CORNELL M.E. (1996), *Uncertainty in risk analysis : six levels of treatment*, Rel. Eng. & Syst. Saf. 54 (2-3), 95-111.
- [51] PENDOLA M. (2000), *Fiabilité des Structures en contexte d'incertitudes statistiques et d'écarts de modélisation*, Thèse de l'Université Clermont II.
- [52] PERSOZ M., HUGONNARD-BRUYÈRE S., VENTURINI V., MEISTER E. (2000), *Deterministic and probabilistic assessments of the reactor pressure vessel structural integrity with user-friendly software*, ASME Press. Vess. & Piping.
- [53] PROCACCIA H., MORILHAT P. (1996), *Fiabilité des structures des installations industrielles*, Coll. D.E.R. EDF, n°94, Eyrolles.
- [54] QUIGGIN J. (1982), *A theory of anticipated utility*, Journal of Economic Behavior and Organization, vol 3, pp. 323-343.
- [55] RABL A., SPADARO J.V. (1999), *Damages and Costs of Air Pollution : an Analysis of Uncertainties*, Environment International Vol 25(1), 29-46.
- [56] RASMUSSEN *et al.* (1975), *Reactor Safety Study : An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants*, Nuclear Regulatory Commission, NUREG-75/014 (WASH-1400), Washington D.C.
- [57] REISS R.D., THOMAS M. (2001), *Statistical Analysis of Extreme Values*, Ed. Birkhauser.
- [58] RELCON A.B., Risk Spectrum – Theory Manual.
- [59] ROYSET J.O., DER KIUREGHIAN A., and POLAK E. (2001), *Reliability-based optimal structural design by the decoupling approach*, Rel. Eng. & Syst. Saf., Vol. 73, Issue 3, P. 213-221.
- [60] ROWE W.D. (1994), *Understanding Uncertainty*, Risk Analysis, Vol 14, n°5.

- [61] SAVAGE L.H. (1954 & 1972), *The Foundations of Statistics*, Dover, New York.
- [62] SHANNON C.E., (1948), *A mathematical theory of communication*, Bell Systems Technical Journal 27.
- [63] SOURGET F., *et al.* (2004), *Protection de la ligne TGV-Méditerranée face aux vents traversiers*, Actes de la Conf. Lambda-Mu 14, Bourges.
- [64] SUDRET B., DEFAUX G., PENDOLA M. (2005), *Time-variant finite element reliability analysis - Application to the durability of cooling towers*, Structural Safety, Vol. 27, pp. 93-112.
- [65] SUDRET B., GUÉDÉ Z., LEMAIRE M. (2005), *Probabilistic assessment of thermal fatigue in nuclear components*, Nucl. Eng. Des., Vol. 235, pp. 219-235.
- [66] VON NEUMANN J., MORGENSTERN O., (1944), *Theory of games and economic behavior*, Princeton University Press.
- [67] WALLEY P. (1991), *Statistical Reasoning with Imprecise Probabilities*, London, Chapman and Hall.