

La simulation du comportement des produits industriels fait généralement appel à des théories déterministes : tous les paramètres des modèles sont supposés connus avec exactitude et, par conséquent, le résultat est lui aussi déterminé de façon exacte. Cependant, les produits réels et leurs environnements sont sujets à des variabilités et dispersions difficiles à modéliser de cette façon, et les paramètres des modèles sont toujours connus à « quelque chose près ». La simulation non-déterministe, encore peu répandue dans les bureaux d'études, désigne un ensemble de techniques permettant d'évaluer l'impact des variabilités et autres incertitudes sur le résultat ; nous nous proposons d'en présenter quelques-unes ici.

Cette ressource fait appel à quelques notions de base de théorie des probabilités : variables aléatoires, moyenne, écart-type, probabilité d'un événement, densité de probabilité d'une variable aléatoire.

1 - Les limites du point de vue déterministe

Les modèles utilisés pour simuler le comportement des produits industriels sont, la plupart du temps, **déterministes**. Cela signifie que chacun de leurs paramètres possède une valeur unique, supposée connue de manière exacte, et qu'ils conduisent de façon prévisible à des résultats eux aussi exacts et uniques, ce qui revient à supposer que les phénomènes ainsi modélisés sont *parfaitement reproductibles* et *connus avec une précision infinie*. Néanmoins, certains phénomènes se manifestant au sein des produits industriels ou de leurs environnements mettent à mal ce point de vue ; on peut les classer en trois catégories, qui ne sont pas exclusives les unes des autres.

1. La première catégorie est celle des **dispersions** ou **variabilités** (les deux termes sont synonymes), qui désigne des phénomènes non rigoureusement reproductibles. Ce défaut de reproductibilité peut se manifester :
 - **Au fil du temps** : c'est notamment le cas lorsque le produit vieillit et se dégrade, ou lorsque son environnement change ;
 - **Dans l'espace** : cela concerne notamment les dispersions dues à la **fabrication du produit** (deux exemplaires réputés identiques ne le sont en réalité pas) et l'hétérogénéité des défauts dans les matériaux fragiles, qui entraîne une forte **variabilité de leur résistance mécanique**, que ce soit en rupture statique ou en fatigue.

Ces dispersions peuvent théoriquement être prises en compte dans les simulations déterministes, car elles peuvent être observées et caractérisées. Malheureusement, cela nécessite d'une part de mesurer toutes les variations des grandeurs concernées dans le temps et/ou l'espace, et d'autre part de multiplier les simulations en conséquence, et ces opérations conduisent la plupart du temps à des coûts et/ou des délais rédhibitoires. On se contente donc souvent, en pratique, de fixer ces grandeurs à une *valeur nominale* dans les simulations.

2. La seconde catégorie est celle des **incertitudes**, qui désigne des phénomènes que l'on ne connaît pas parfaitement. Une source notable d'incertitudes réside dans les mesures des grandeurs obtenues expérimentalement.
3. Enfin, les **simplifications** délibérément opérées lorsque l'on modélise certains phénomènes (linéarisations, interactions négligées...) sont également à l'origine de variations sur les données des modèles : par exemple, si l'on dispose d'une courbe de traction qui n'est pas parfaitement rectiligne, il existe plusieurs façons d'y faire passer une droite afin d'identifier le module d'Young (figure 1). Les variations qui en résultent sont d'autant plus importantes que le modèle est éloigné de la réalité par rapport à laquelle on cherche à identifier ses paramètres.

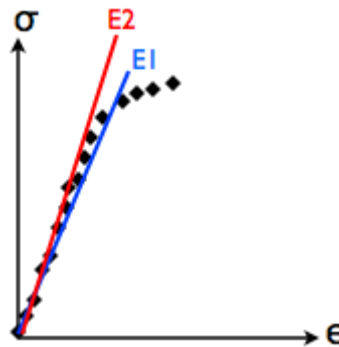


Figure 1 : Exemple d'incertitude : un module d'Young déterminé expérimentalement.

2 - L'objectif de la simulation non-déterministe

Les phénomènes ci-dessus (que nous désignons dans la suite de la ressource par le terme générique « *incertitudes* ») ont une influence sur les résultats recherchés : en effet, sans remettre en cause le caractère causal de la simulation (le résultat dépend des données d'entrée de façon déterministe), **les variations des données d'entrée devraient logiquement entraîner une variation du résultat**. En d'autres termes, les incertitudes affectent les performances, et sont donc susceptibles d'entraîner l'insatisfaction du client. Or, les méthodologies actuelles se contentent de « *simuler au nominal* » et ne permettent donc pas de prévoir les éventuelles variations des performances dues aux incertitudes.

Pour cette raison, les concepteurs cherchent généralement à « limiter la casse » en surdimensionnant leurs produits ou encore en les concevant de manière robuste (c'est-à-dire de sorte à minimiser l'effet de quelques variabilités, préalablement identifiées et sélectionnées, sur les performances). Malheureusement, ces pratiques ont un coût, et dégradent à leur tour d'autres performances ; les entreprises ont donc tout intérêt, pour rester compétitives, à être capables de **prévoir les variations du résultat en fonction des variations des données d'entrée** (figure 2). Tel est l'objet de la **simulation non déterministe**, qui est une pratique encore peu répandue dans les bureaux d'études bien que la démarche, en soi, ne soit pas nouvelle : les expérimentateurs la mettent couramment en œuvre, et les théories mathématiques étudiant des problèmes similaires sont légion (nous allons d'ailleurs en aborder quelques-unes dans la suite de cette ressource).

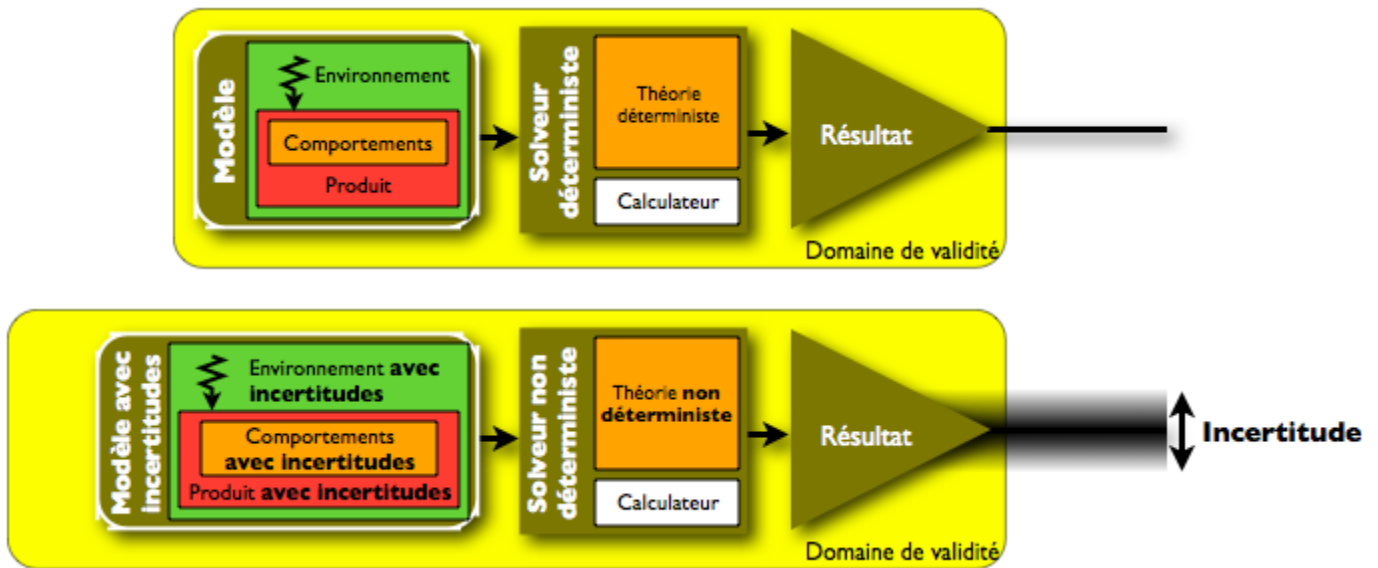


Figure 2 : Comparaison des simulations déterministe et non déterministe.

Il est important de ne pas confondre le non-déterminisme dont il est question ici, qui s'applique aux **données d'entrée** et non à leur relation avec le résultat, avec celui de certains types de problèmes mal posés, parfois improprement appelés « problèmes complexes », dont la solution mathématique n'est pas unique bien que toutes les données d'entrée le soient rigoureusement ; ces problèmes sont eux aussi non-déterministes, mais le sont par la **relation entre les données d'entrée et le résultat**, c'est-à-dire à cause des lois de leur théorie, et non par les données d'entrée elles-mêmes. **Nous nous limitons ici aux théories dans lesquelles la relation entre résultat et données d'entrées reste déterministe.**

Comme le montre la figure 2, la simulation non-déterministe nécessite quelques « ingrédients » supplémentaires par rapport à une simulation déterministe, et demande de :

1. **Choisir une théorie non déterministe**, permettant de représenter les variations des données des modèles et d'en déduire les variations du résultat ;
2. **Modéliser les incertitudes**, qu'elles concernent le modèle du produit, le modèle de l'environnement, ou des modèles de comportements ;
3. **Simuler la "propagation" de ces incertitudes**, c'est-à-dire leur influence sur le résultat, sans remettre en cause le lien de causalité entre les données d'entrée et le résultat.

La suite de cette ressource présente quelques techniques permettant de mener à bien cette démarche, en les classant par les théories dont elles relèvent. Ces théories sont ici au nombre de trois : la théorie des probabilités, la théorie des intervalles et la théorie des méconnaissances. Nous les illustrons sur l'exemple très simple d'une barre en traction dont la rigidité K est supposée mal connue, soumise à un effort F connu, dont nous cherchons à calculer l'allongement U (figure 3).



Figure 3 : Une barre en traction dont la rigidité est incertaine

L'objectif de la simulation non déterministe est donc de modéliser l'incertitude portant sur K , puis d'en déduire celle portant sur U .

3 - Les simulations basées sur la théorie des probabilités

Une première possibilité est de faire appel à la théorie des probabilités pour représenter les grandeurs incertaines. Ceci conduit à modéliser la rigidité et le déplacement par des variables aléatoires $K(\theta)$ et $U(\theta)$; la notation (θ) indique que la rigidité et le déplacement n'ont pas une valeur fixe donnée, mais sont des variables aléatoires.

Afin de décrire la connaissance que l'on a de ces quantités, il faut se donner une *loi de probabilité* pour $K(\theta)$; l'objectif de la simulation sera ensuite d'en déduire des informations sur la loi de probabilité de $U(\theta)$. Ici, nous supposons que la rigidité $K(\theta)$ suit une loi normale, également appelée **loi gaussienne** ; il s'agit d'une loi de probabilité très employée, car on l'obtient naturellement en additionnant un grand nombre de variables aléatoires indépendantes. Sa fonction densité de probabilité est :

$$f_k(k) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{k-\mu}{\sigma}\right)^2\right)$$

où μ représente la **valeur moyenne** de la rigidité, et σ son **écart-type**. Le sens de cette fonction est le suivant : plus la rigidité a de chances de prendre la valeur k , plus la valeur de $f_k(k)$ est élevée. Plus rigoureusement, la probabilité que la rigidité soit comprise dans un intervalle $[k1, k2]$ est par définition égale à *l'intégrale* de f_k sur cet intervalle.

Le graphe de cette fonction est représenté sur la figure 4 ; c'est la fameuse « courbe en cloche », très souvent employée en statistiques pour la raison évoquée ci-dessus. Pour la suite, nous choisissons $\mu = 1$ et $\sigma = 0,01$. Avec ces valeurs, la rigidité a plus de 99% de chances de se trouver entre 0,975 et 1,025 fois sa valeur nominale ; il s'agit d'un ordre de grandeur courant pour la dispersion des rigidités des pièces mécaniques réelles.

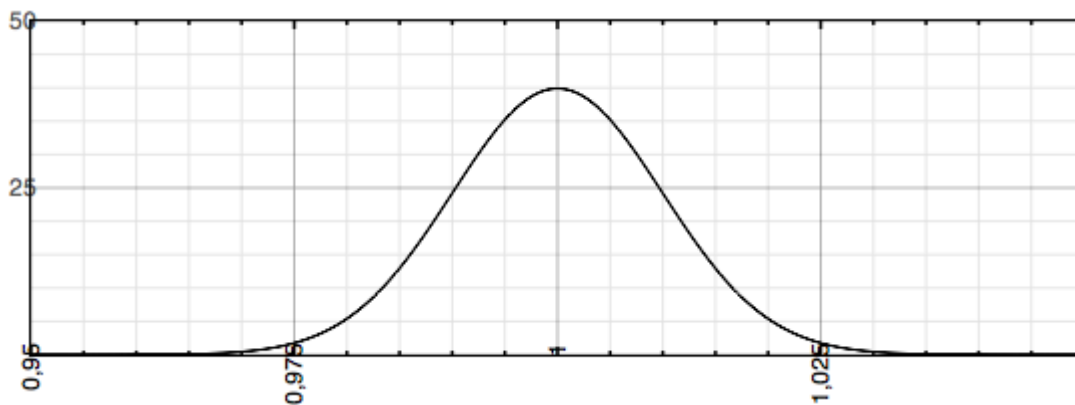


Figure 4 : Fonction densité de probabilité de la rigidité (loi gaussienne de moyenne 1 et d'écart-type 0,1).

L'objectif de la simulation non déterministe est de caractériser la variable aléatoire $U(\theta)$. Nous présentons ici trois méthodes.

3.1 - Résolution analytique

Ici, le problème étant très simple, il est possible de trouver directement « à la main » la densité de probabilité de $U(\theta)$ par une approche analytique. En effet, $U(\theta)$ s'exprime en fonction de $K(\theta)$ par :

$$U(\theta) = \frac{F}{K(\theta)}$$

Il est alors possible d'appliquer le **théorème dit du changement de variable**, qui donne directement la densité de probabilité de U en fonction de celle de K . Ce théorème stipule que s'il existe une fonction g strictement monotone telle que $U = g(K)$, alors :

$$f_U(u) = \left| \frac{1}{g'(g^{-1}(u))} \right| f_K(g^{-1}(u))$$

où g^{-1} représente la réciproque de g et g' sa dérivée. Ici, $g(K) = F/Kg$; cette fonction est strictement monotone sur le bon ensemble de définition (c'est-à-dire pour $k > 0$, puisque la rigidité n'est jamais négative ou nulle) et l'application du théorème donne donc :

$$f_U(u) = \frac{F}{u^2} f_K\left(\frac{F}{u}\right)$$

c'est-à-dire :

$$f_U(u) = \frac{1}{\sigma\sqrt{2\pi}u^2} \exp\left(-\frac{1}{2}\left(\frac{\frac{F}{u}-\mu}{\sigma}\right)^2\right)$$

Comme le montre l'expression ci-dessus, le **déplacement $U(\theta)$ ne suit plus une loi gaussienne**. Ce résultat est général : l'inverse d'une variable gaussienne n'est pas une variable gaussienne. La figure 5 représente la courbe de cette fonction pour $\mu=1$, $F=1$ et $\sigma=0,01$. Le tracé obtenu semble très proche d'une courbe en cloche ; cela n'est vrai que pour de très faibles valeurs de σ , et on obtiendrait une allure différente pour des dispersions plus importantes.

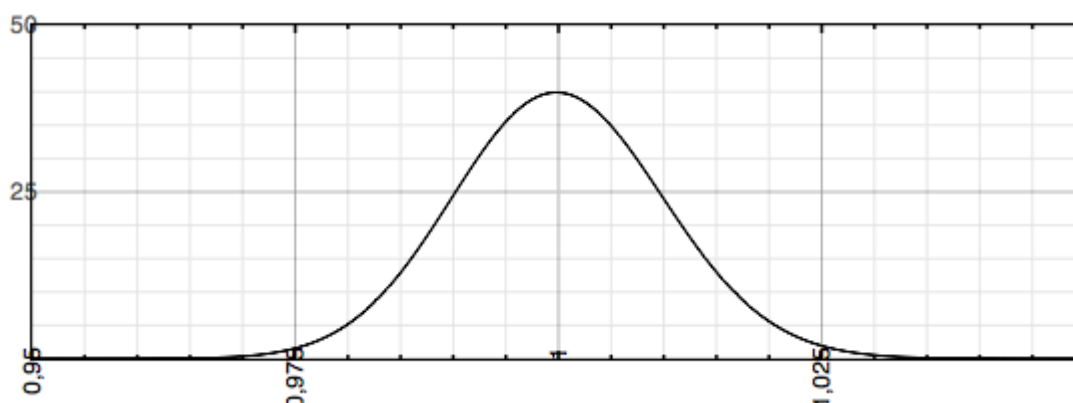


Figure 5 : Fonction densité de probabilité du déplacement obtenu lorsque la rigidité suit une loi normale.

Une telle résolution analytique n'est naturellement envisageable que sur des problèmes très simples ; la plupart du temps, il est impossible de trouver l'expression explicite de la densité de probabilité du résultat, et il faut alors procéder autrement.

3.2 - Tirages de Monte-Carlo

Parmi les techniques utilisables en l'absence de solution analytique connue, la plus simple et la plus utilisée d'entre elles est la méthode de Monte-Carlo, qui consiste à :

- Tirer un grand nombre de rigidités au hasard, suivant leur loi de probabilité,
- Pour chacune d'entre elles, réaliser une simulation déterministe,
- Une fois les simulations réalisées, faire une étude statistique de l'ensemble des résultats afin d'obtenir la moyenne, l'écart-type, de rechercher l'occurrence d'un certain événement (comme une probabilité de rupture)...

Le fait que l'on ait uniquement des simulations déterministes à effectuer rend la mise en œuvre de cette méthode très facile, et explique sa popularité. Néanmoins, la loi des grands nombres fait

qu'un grand nombre de simulations doit être effectué afin que l'étude statistique des réponses donne des résultats pertinents. Ainsi, la figure 6 représente les résultats obtenus sur l'exemple de la barre en traction, après programmation de la méthode ; on constate qu'avec 100 tirages de la rigidité (graphe de gauche), la répartition des déplacements est obtenue de manière relativement imparfaite, et il faut 5000 tirages (graphe de droite) afin d'obtenir une répartition proche de la solution analytique. Cela donne un ordre de grandeur du nombre de simulations déterministes qu'il faut réaliser pour obtenir une information représentative.

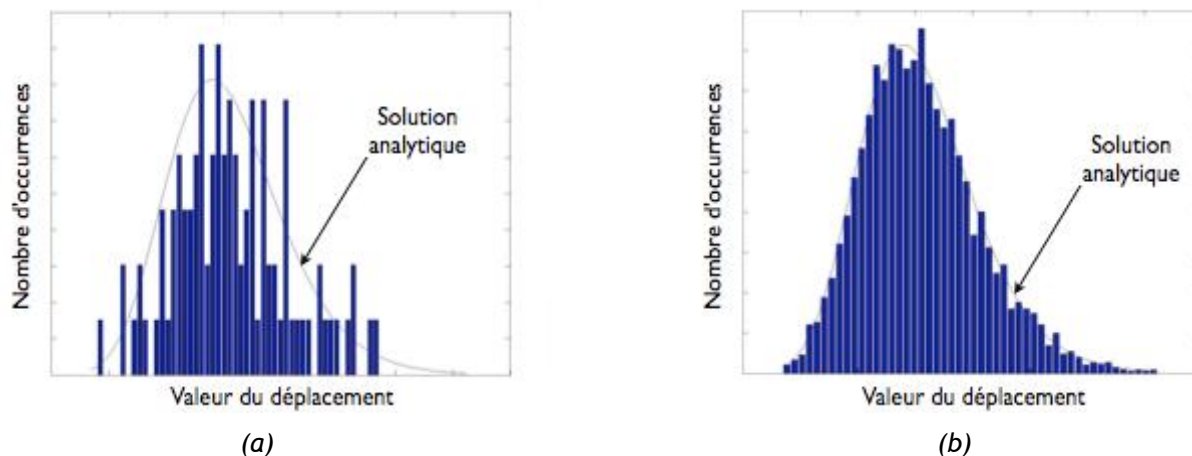


Figure 6 : Distribution des déplacements obtenue par la méthode de Monte-Carlo : (a) avec 100 tirages de la rigidité, (b) avec 5000 tirages.

Ainsi, l'utilisation de la méthode de Monte-Carlo pour obtenir une estimation pertinente des incertitudes peut entraîner des coûts très élevés : ceci est le principal frein à l'utilisation de cette technique.

3.3 - Linéarisation des incertitudes

Afin de réduire ces coûts, il est plus efficace d'évaluer directement l'influence des incertitudes, sans passer par des tirages aléatoires. Dans le cas où les grandeurs incertaines varient peu, ce qui est le cas ici, une approche simple est de les écrire sous la forme suivante :

$$U(\theta) = \frac{F}{K(\theta)}$$

où K_0 est la valeur moyenne de la rigidité, $\alpha(\theta)$ est une variable aléatoire centrée réduite (i.e. de moyenne nulle et de variance 1) et δ est son coefficient de variation (c'est-à-dire que l'écart-type de la rigidité vaut $K_0\delta$). Pour rechercher le déplacement $U(\theta) = K(\theta)^{-1}F$, nous linéarisons alors ses variations autour de sa moyenne à l'aide d'un développement limité, ce qui s'écrit :

$$U(\theta) = \frac{F}{K_0} \left(1 - \delta\alpha(\theta) + \delta^2\alpha(\theta)^2 \right)$$

A partir de ce développement, on peut alors caractériser les variations du déplacement, par exemple en calculant sa moyenne et son écart-type. L'avantage de cette technique est de très bien s'adapter aux simulations de grande taille, car aucun calcul analytique n'est nécessaire, et le volume de calculs numériques à effectuer est généralement bien moindre qu'avec la méthode de Monte-Carlo. Ainsi, ce procédé est notamment utilisé dans une classe de logiciels de simulation non déterministes par éléments finis, nommés S-FEM (pour Stochastic Finite Element Method, "stochastique" étant synonyme d'aléatoire).

Sous réserve que les variations de $K(\theta)$ soient effectivement faibles, il s'agit d'une technique performante. Ici, $\delta = 0,01$; le développement limité est donc tout à fait justifié et les résultats

obtenus sont quasiment identiques à la solution analytique. Naturellement, pour des variations plus importantes, l'écart serait lui aussi plus important.

4 - Les simulations basées sur la théorie des intervalles

Un inconvénient des solveurs probabilistes est que, pour les utiliser, il faut préalablement identifier les lois de probabilité auxquelles sont soumis les paramètres incertains. En pratique, cela peut nécessiter de faire beaucoup d'essais, et ce particulièrement lorsque les paramètres incertains sont nombreux. Pour cette raison, des techniques alternatives ont été proposées, basées sur des modélisations simplifiées de l'incertitude.

Un exemple d'une telle technique est la **théorie des intervalles**, qui consiste à représenter une grandeur incertaine en se donnant simplement deux bornes entre lesquelles elle varie. On cherche alors entre quelles bornes varie le résultat du calcul. Ici, par exemple, la rigidité est simplement modélisée par l'intervalle $[K^-, K^+]$; le déplacement évolue alors dans l'intervalle $[F/K^+, F/K^-]$.

L'avantage d'une telle représentation est que, contrairement aux approches probabilistes, **il n'y a pas besoin de savoir comment la rigidité se répartit entre ses deux bornes** : il suffit de connaître un encadrement des quantités de départ de la simulation pour obtenir un encadrement de celles d'arrivée, ce qui est plus simple à identifier qu'une loi de probabilité et donne des informations pertinentes pour l'ingénieur.

Cependant, la théorie des intervalles souffre d'un gros défaut : la plupart du temps, **elle surestime l'incertitude portant sur le résultat**. En effet, elle est basée sur des opérations arithmétiques consistant à rechercher le « pire » cas de figure possible afin d'encadrer le résultat de façon certaine (ce que l'on appelle « propager les intervalles »). Ainsi, si une quantité A est modélisée par l'intervalle $[A^-, A^+]$ et si une autre quantité B est modélisée par $[B^-, B^+]$, alors :

- La somme $A+B$ sera modélisée par $[A^-+B^-, A^++B^+]$
- La différence $A-B$ sera modélisée par $[A^- - B^+, A^+ - B^-]$
- Le produit, le quotient... seront définis de même, de sorte à toujours garantir l'encadrement.

Le problème est que, une fois programmée numériquement, la propagation des intervalles ne reconnaît pas les multiples instances d'une même grandeur : ainsi, en appliquant les règles ci-dessus, l'image de A par la fonction $f(x) = 2x - x$ sera égale à l'intervalle $[2A^- - A^+, 2A^+ - A^-]$, alors que cette fonction n'est autre que l'identité soit un intervalle 3 fois trop large. Un examen rapide des bornes de l'intervalle obtenu montre que chacune fait intervenir à la fois A^- et A^+ . Or, A ne peut pas être à la fois à son maximum et à son minimum ! Les bornes prévues ne seront donc jamais atteintes, et **l'application de la théorie des intervalles conduit ainsi à surestimer l'incertitude portant sur le résultat**, en considérant des combinaisons impossibles. Ce phénomène est d'autant plus marqué que la taille des systèmes d'équations à résoudre est importante.

Une illustration de ce dernier point sur l'exemple de la barre en traction est donnée dans le tableau ci-dessous ; la rigidité est supposée varier entre 0,977 et 1,023, ce qui revient à conserver 99% des réalisations pour la loi gaussienne de la figure 4. On constate que les bornes du déplacement donné par $U=F/K$ sont bien identiques à la solution analytique, correspondant à la loi de probabilité de la figure 5 ; cependant, si l'on discrétise la barre à l'aide de deux « demi-barres » en série, et que l'on applique la propagation des intervalles au système de deux équations à deux inconnues ainsi obtenu, alors on trouve une incertitude 3 fois trop grande sur le

déplacement : plus ou moins 7%, au lieu de 2,3% ! Lorsqu'il y a plusieurs milliers d'équations, les résultats deviennent ainsi rapidement inutilisables...

Méthode de simulation	U	U^+	Incertitude
Analytique probabiliste (bornes à 99%)	0,978	1,024	+/- 2,3%
Intervalles ($U=F/K$)	0,978	1,024	+/- 2,3%
Intervalles (2 demi-barres)	0,933	1,071	+/- 7%

5 - La théorie des méconnaissances : une approche hybride

La **théorie des méconnaissances** est une théorie récente, développée au LMT¹, qui emprunte des idées à la fois à la théorie des probabilités et à la théorie des intervalles afin de tenter de concilier leurs avantages respectifs. En une phrase, son idée essentielle est de décrire les grandeurs incertaines à l'aide d'**intervalles dont les bornes sont des variables aléatoires**.

Ainsi, pour l'exemple de la barre en traction, la rigidité est caractérisée de la manière suivante :

$$K_0(1 - m^-(\theta)) \leq K \leq K_0(1 + m^+(\theta))$$

où $m^-(\theta)$ et $m^+(\theta)$ sont deux variables aléatoires appelées **méconnaissances**, que l'on choisit généralement gaussiennes, et K_0 est une rigidité déterministe pouvant représenter une valeur nominale, une moyenne... Par rapport à une modélisation par intervalles, l'encadrement est moins strict puisque les bornes sont « floues » ; par rapport à une modélisation probabiliste, la description est différente puisqu'on ne s'intéresse pas à la moyenne, l'écart-type, la fonction de répartition... des rigidités (dont l'identification nécessite souvent de nombreux essais) mais simplement à des **bornes raisonnables**, plus simples à identifier et qui intéressent directement l'ingénieur. En d'autres termes, une modélisation avec méconnaissances contient une information du type « *la rigidité a 99% de chances de se trouver entre 0,97 et 1,02 fois sa valeur nominale* ».

La résolution du problème s'effectue alors en deux temps. On réalise d'abord la simulation déterministe correspondant aux données nominales ; ici, il s'agit simplement de calculer $U_0 = F/K_0$. On calcule ensuite, à l'aide de règles de « propagation » suivant la même idée que celles des intervalles, les méconnaissances portant sur le résultat, à savoir ici :

$$\frac{U_0}{(1 + m^+(\theta))} \leq U \leq \frac{U_0}{(1 - m^-(\theta))}$$

La différence essentielle avec la théorie des intervalles est qu'ici, les bornes étant des variables aléatoires, les règles de propagation tiennent compte de l'aspect probabiliste des bornes. Ainsi, **on peut obtenir une mesure pertinente de l'incertitude portant sur le résultat, sans surestimations** ; en contrepartie, l'encadrement n'est pas garanti dans 100% des cas (on considère généralement des « bornes à 99% », c'est-à-dire un encadrement qui a 99% de chances d'être vérifié). Ici, en propageant des méconnaissances de plus ou moins 2,5% sur la rigidité, on obtient une dispersion de plus ou moins 2,1% sur le déplacement (en termes de "bornes à 99%"), soit un résultat cohérent avec les méthodes précédentes.

¹ Laboratoire de Mécanique et Technologie, Ecole Normale Supérieure de Cachan

6 - Bilan et conclusions

Dans cette ressource, nous avons présenté quelques techniques non-déterministes permettant de **prendre en compte des incertitudes dans les données d'une simulation, et de quantifier leur influence sur le résultat**. La figure 7 propose un récapitulatif des trois théories sur lesquelles ces techniques sont basées :

- La **théorie des probabilités** fait appel à une loi de probabilité des quantités incertaines ;
- La **théorie des intervalles** fait appel à un encadrement (strict) des quantités incertaines ;
- La **théorie des méconnaissances** fait appel à un encadrement non strict (par exemple, ayant 99% de chances d'être vérifié) des quantités incertaines.

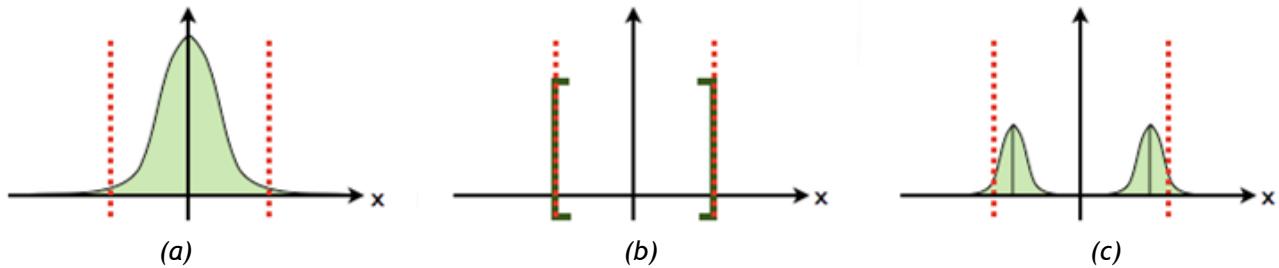


Figure 7 : Trois façons de modéliser les incertitudes :
(a) théorie des probabilités, (b) théorie des intervalles, (c) théorie des méconnaissances.

Sur un problème aussi simple que la barre en traction évoquée ci-dessus, il est difficile d'évaluer les performances de ces techniques. Sur une simulation de grande taille, on constate que les approches probabilistes ont tendance à encadrer correctement l'incertitude mais demandent d'identifier les lois de probabilité des différents paramètres, ce qui peut être long et coûteux en termes d'essais ; inversement, la théorie des intervalles fait appel à des données très simples à identifier mais tend à surestimer l'incertitude portant sur le résultat. La théorie des méconnaissances, qui emprunte des éléments à ces deux familles d'approches, a été conçue pour tenter de concilier leurs avantages.

Toutes ces techniques permettent de quantifier l'incertitude portant sur le résultat d'une simulation, et peuvent donc aider à en évaluer la pertinence.

Références :

[a]: Enjalbert P., La validation des modèles ou comment décrire le manque de connaissance, Technologies et Formations n° 138.

[b]: Enjalbert P., Sur la théorie des méconnaissances en conception robuste. Thèse de Doctorat, Ecole Normale Supérieure de Cachan, 2009.

[c]: Ghanem R. G. et Spanos P. D., Stochastic Finite Elements : a Spectral Approach, Springer, 1991.

[d]: Moore R., Methods and Applications of Interval Analysis, Studies in Applied Mathematics (SIAM), 1979.

Ressource publiée sur EDUSCOL-STI : <http://eduscol.education.fr/sti/si-ens-cachan/>