

PROPOSITION de SUJET de THESE (2009)

Bertrand Iooss

CEA, Direction de l'Energie Nucléaire, Centre de Cadarache, Département d'Etude des réacteurs
Service d'Etudes des Systèmes Innovants, Laboratoire de Conduite et Fiabilité des Réacteurs

Titre de la thèse

Développement de méthodologies de traitement des incertitudes pour les codes de simulation
stochastiques

Financement demandé : contrat CFR

Projet DEN : DSOE/Simulation/PICI2

Problématique industrielle

En ingénierie nucléaire, les modèles numériques prédictifs sont souvent d'une aide précieuse lors des phases de validation ou de conception de systèmes complexes (*e.g.* conception d'un système de sauvegarde sur un réacteur, optimisation de l'emplacement des crayons dans un assemblage combustible). Dans ce type d'études, il est indispensable d'estimer précisément les incertitudes sur les prédictions, de connaître les risques d'événements rares, voire d'optimiser des réponses sous contraintes.

Pour toutes ces raisons, le développement d'outils génériques d'analyses d'incertitudes, d'analyses de sensibilité et de calibration de modèles est d'une importance fondamentale, aussi bien pour les phases de développement et de validation, que pour celles d'utilisation des codes de calcul. L'analyse, la planification et la modélisation des expériences numériques est un sujet de recherche sur lequel le CEA et ses partenaires travaillent depuis une dizaine d'années (de Rocquigny et al., 2008, Iooss, 2009). Ces travaux de recherche débouchent à présent sur des méthodologies et outils opérationnels (*e.g.* la plate-forme « Incertitudes » de la DEN URANIE), déployables sur des problèmes industriels (*e.g.* le logiciel LEONAR qui réalise des évaluations probabilistes d'accidents graves dans les REPs, Iooss et al., 2008).

Lorsque le modèle numérique étudié est coûteux en temps de calcul (une évaluation pouvant requérir plusieurs heures de calculs par exemple), l'approche adoptée consiste à construire une fonction mathématique simple (nommée « métamodèle ») à partir de quelques évaluations bien choisies du modèle (Fang et al., 2006). Ce métamodèle est alors capable d'approximer le code numérique et de prédire de nouveaux calculs avec un temps CPU négligeable. Les métamodèles ont jusqu'ici été principalement étudiés pour des codes de calcul déterministes, c'est-à-dire pour lesquels deux calculs avec un même jeu de paramètres d'entrée procurent les mêmes réponses. Ceci a permis de résoudre par exemple des problématiques liées à la modélisation du transfert de contaminants dans les sols (Volkova et al., 2008), à la simulation de transitoires accidentels dans les REPs à l'aide de modèles thermohydrauliques complexes et coûteux (Cannamela et al., 2008), à la calibration de modèles de transfert des produits de corrosion dans les circuits REPs (Pérot, 2006).

Dans certaines applications, les codes de calcul sont eux-mêmes de nature stochastique, c'est-à-dire qu'ils intègrent un aléa intrinsèque au modèle. De tels codes sont par exemple :

- les codes de calcul basés sur la méthode de Monte Carlo pour simuler les trajectoires de neutrons dans un cœur de réacteur nucléaire,
- par extension les codes de calcul déterministes dont certaines variables d'entrée ne sont pas réduites à l'espérance de l'évaluation expérimentale, par exemple la prise en compte de l'incertitude des données nucléaires (irréductible sur une échelle de temps de l'ordre de la dizaine d'années),

- les codes de calcul d'impact dans l'environnement modélisant la diffusion et la dispersion de particules par des équations stochastiques (approche lagrangienne),
- les modèles de files d'attente (illustrés par exemple dans Kleijnen, 2008) et les modèles de croissance de population,
- d'un point de vue industriel, les codes multi-physiques intégrant une instabilité numérique due au mauvais conditionnement multi-physique intrinsèque des systèmes.

Dans ce type de modèles stochastiques, on ne cherche plus à récupérer la valeur d'une variable en sortie d'un calcul mais sa distribution de probabilités. Cette distribution de probabilités dépend des variables d'entrée du modèle, dont certaines sont elles-mêmes incertaines mais contrôlables (par opposition à l'aléa intrinsèque qui lui est considéré comme incontrôlable, Iooss et al, 2006). Ces modèles numériques sont le plus souvent extrêmement coûteux ce qui motive fortement le développement d'approche de type « métamodèles » afin de déployer des analyses d'incertitudes et de sensibilité à leur égard. Le sujet de cette thèse porte sur le développement et l'analyse de tels métamodèles, sujet qui n'a pas été traité jusqu'ici.

Cette problématique est un besoin qui apparaît dans certaines applications, par exemple dans les analyses d'incertitudes des calculs d'historique d'irradiation par les codes Monte Carlo dans le cadre de la dosimétrie des réacteurs (DER/SPEX). Les applications qui seront traitées durant la thèse concernent l'analyse de sensibilité de codes de calcul Monte Carlo utilisés par le DER/SPRC (par exemple TRIPOLI) et les calculs d'évolution radioactive du combustible par Monte Carlo (IRSN/FAR/DSU/SEC).

Approche scientifique proposée

Les aspects planification d'expériences et construction du métamodèle sont les deux axes de recherche principaux de cette thèse. Au niveau de la planification d'expériences, il conviendra de développer des stratégies optimales par rapport au métamodèle utilisé. Le doctorant pourra s'inspirer de l'imposante littérature qui existe sur le thème de la planification robuste d'expériences. Quand les expériences sont numériques, seuls quelques articles émergent (Lehman et al., 2004, Bates et al., 2006) et proposent de premières pistes de recherche. Les approches standards planifient les expériences en fonction de l'objectif de la modélisation de la moyenne et de la variance de la réponse. La variance est en fait la dispersion induite par l'aléa stochastique du modèle. Dans cette thèse, des approches innovantes devront être explorées afin de s'intéresser à toute la distribution de la réponse.

Au niveau du métamodèle, des travaux récents proposent de premières pistes de travail et il conviendra de les évaluer en premier lieu. Par exemple, Iooss et al. (2006) proposent de modéliser la moyenne et la variance par un modèle joint de type Modèle Additif Généralisé (GAM), c'est-à-dire un modèle GAM pour la moyenne et un modèle GAM pour la variance. Kleijnen (2008) et Ginsbourger et al. (2008) proposent de traiter ce problème à l'aide de métamodèles hétéroscédastiques basés sur le modèle du krigeage. Ces travaux, ne s'intéressant qu'aux deux premiers moments statistiques de la variable de sortie, des méthodologies doivent être mises en œuvre pour calculer les quantiles conditionnels (cf. Rigby & Stasinopoulos (2000) dans le cas gaussien et Zabalza-Mezghani (2000) dans le cas non gaussien). En second lieu, le doctorant essaiera de développer une approche plus ambitieuse en proposant un métamodèle permettant la reconstruction complète de la distribution de probabilité conditionnelle. Récemment, Reich et al. (2008) ont proposé une telle méthode en se basant sur un modèle de mélange de lois gaussiennes et Rigby & Stasinopoulos (2005) en se basant sur les GAM. Le doctorant pourra s'intéresser aux propriétés théoriques de ce type d'estimateurs et développera éventuellement des modèles statistiques complémentaires.

A l'issue de cette étape, il conviendra de développer des méthodes d'analyse de sensibilité globale adaptées à la situation des codes stochastiques : la réponse n'étant plus une variable scalaire mais une distribution de probabilité. Les méthodes d'analyse de sensibilité basées sur la mesure d'entropie et non plus de variance pourraient être une première piste de recherche. Il sera également pertinent de s'intéresser à l'abondante littérature développée dans le cadre de l'analyse de sensibilité locale de modèles stochastiques (Koda, 1992, Gunawan et al., 2005).

Contexte CEA

Le CEA vise à être l'un des acteurs majeurs en matière de simulation numérique. L'un des points durs (notamment dans le cadre de la valorisation industrielle des modèles numériques) actuellement identifié sur ce sujet concerne la prise en compte des incertitudes sur les variables d'entrée physiques, les paramètres des modèles, les modèles, ... Les travaux de R&D sur le traitement des incertitudes en simulation numérique constituent un sujet d'intérêt (projet DSOE/Simulation/PICI2). La plupart des travaux réalisés à ce jour concernent les codes de calcul déterministes. Les codes de calcul stochastiques constituent pourtant une part non négligeable des applications du CEA. Le temps de calcul coûteux pour ce type de codes justifie l'approche par métamodèle proposée ici.

Le DER/SESI/LCFR (Laboratoire de Conduite et de Fiabilité des Réacteurs) est particulièrement impliqué dans l'utilisation et le développement de méthodes de fiabilité, de propagation d'incertitudes et d'analyse de sensibilité appliquées à divers processus (mesures, expériences, codes de calcul). Les travaux de l'équipe Incertitudes du LCFR sont reconnus aussi bien au niveau du CEA (animation du réseau d'échanges IMPEC) qu'au niveau scientifique national (animation du GdR CNRS MASCOT-NUM). Cette thèse se situe dans ses programmes de R&D amont.

Enfin, le développement des grandes plate-formes métiers de la DEN (PLEIADES dans le domaine des combustibles, NEPTUNE en thermohydraulique, ALLIANCES pour la gestion des déchets radioactifs) s'accompagne d'un souci de prise en compte des incertitudes de manière systématique dans les modèles physiques. Cette thèse entre dans le cadre de l'application des méthodes de propagation d'incertitudes sur ces codes souvent coûteux en temps de calcul et du développement de la plate-forme « Incertitudes » URANIE. Les outils développés durant cette thèse ont pour vocation à être intégrés à terme dans URANIE, afin de répondre aux besoins des unités du CEA.

Caractère formateur du sujet et débouchés professionnels

Au cours de la thèse, l'étudiant aura la responsabilité de son projet de recherche. Celui-ci demande des connaissances solides, à la fois en statistique et en calcul numérique. L'étudiant sera amené à réaliser un important travail de synthèse afin d'intégrer les multiples techniques statistiques, mathématiques et physiques nécessaires à l'élaboration de solutions novatrices. Le doctorant développera également son esprit d'ouverture et sa culture scientifique de par la multiplicité des acteurs impliqués dans la thèse : laboratoires universitaires, laboratoires CEA et IRSN à thématiques numérique et physique, équipes industrielles d'EDF, AREVA ...

Le thème général de cette recherche est ambitieux et actuellement très porteur. Différentes initiatives de type collaboratif ont vu le jour ces dernières années :

- le Groupement de Recherche (GdR) CNRS « Méthodes d'Analyse Stochastique pour les COdes et Traitements NUMériques » (<http://www.gdr-mascotnum.fr/>) qui a pour vocation de fédérer la recherche en France sur le sujet ;
- le Groupe de Travail et de Réflexion « Incertitudes » de l'Institut de Maîtrise des Risques qui a pour objectif de susciter des échanges génériques entre les praticiens et les chercheurs, et à proposer des formations pour les ingénieurs intéressés par ces approches ;
- le réseau IMPEC du CEA (<http://www-impec.cea.fr:8000/>) qui suscite des échanges et partages d'expériences en interne CEA (principalement entre DAM et DEN) afin de promouvoir l'utilisation de ces techniques.

La gestion des incertitudes dans les processus utilisés dans le monde industriel (industries automobile, pétrolière, aéronautique, spatiale, chimie, météorologie, ...) est devenu indispensable, et le doctorant aura d'importants atouts pour poursuivre dans ce domaine. L'élargissement des compétences du doctorant devrait lui permettre de s'adapter rapidement dans de nombreuses équipes de mathématiques appliquées industrielles et universitaires. Ces dernières années, tous les doctorants ayant travaillé sur ce sujet au CEA ont trouvé immédiatement à l'issue de leur thèse un emploi dans un EPIC ou dans l'industrie (IFP, EADS, CEA/DAM, CEA/DRT, CEA/DEN).

Encadrement CEA : B. Iooss

Habilitation à diriger des Recherches : oui (janvier 2009).

Encadrement universitaire : F. Gamboa (Institut de Mathématiques de Toulouse)

Sur le thème des méthodes statistiques pour les analyses d'incertitudes et de sensibilité, le LCFR a des échanges fructueux avec certains chercheurs du Laboratoire de Statistiques et Probabilités (LSP) de Toulouse. Ce laboratoire possède, entre autres, de fortes compétences en planification d'expériences (e.g. J-M. Azaïs) et en statistique non paramétrique (B. Laurent qui a déjà dirigé une thèse co-encadrée par le LCFR). L'école doctorale de ce laboratoire est également l'école de rattachement pour l'habilitation de B. Iooss.

Collaborations :

Des échanges méthodologiques sur les métamodèles avec le service d'étude en criticité de l'IRSN/FAR, qui est concerné par le même type de problématiques, sont envisagés.

Références

- R.A. Bates, R.S. Kenett, D.M. Steinberg and H.P. Wynn (2006), Achieving robust design from computer simulations, *Quality Technology & Quantitative Management*, 3:161-177.
- C. Cannamela, J. Garnier & B. Iooss (2008), Controlled stratification for quantile estimation, *Annals of Applied Statistics*, in press.
- E. de Rocquigny, N. Devictor and S. Tarantola (eds) (2008), *Uncertainty in industrial practice*, Wiley.
- K-T. Fang, R. Li, A. Sudjianto (2006), *Design and modeling for computer experiments*, Chapman & Hall.
- D. Ginsbourger, O. Roustant and Y. Richet (2008), Kriging with heterogeneous nugget effect for the approximation of noisy simulators with tunable fidelity, *Proceedings of Joint Meeting of the Statistical Society of Canada and the Société Française de Statistique*, Ottawa, Canada.
- R. Gunawan, Y. Caio, L. Petzold and F.J. Doyle, Sensitivity analysis of discrete stochastic systems, *Biophysical journal*, 88:2530-2540.
- B. Iooss, M. Ribatet and A. Marrel (2006), Global sensitivity analysis for stochastic computer models with Generalized Additive Models, *Technometrics*, submitted.
- B. Iooss, M. Marquès, F. Gaudier, B. Spindler & B. Tourniaire (2008), Uncertainty assessments in severe accident scenarios using the URANIE software, *35rd ESReDA Seminar on "Uncertainty in Industrial Practicet"*, Marseille, France, novembre 2008.
- B. Iooss (2009), *Contributions au traitement des incertitudes en modélisation numérique*, Mémoire d'Habilitation à Diriger des Recherches, Université Paul Sabatier – Toulouse III, à paraître.
- J.P.C. Kleijnen (2008), *Design and analysis of simulation experiments*, Springer.
- J.S. Lehman, T.J. Santner and W.I. Notz (2004), Designing computer experiments to determine robust control variables *Statistica Sinica*, 14:571-590.
- M. Koda (1992), Sensitivity analysis of stochastic dynamical systems, *International Journal of Systems Science*, 23:2187-2195.
- N. Pérot (2006), *Calibration réduite du code PACTOLE suivant une méthodologie d'optimisation stochastique sur des critères statistiques*, NT CEA/DEN/CAD/DER/SESI/LCFR/NT DO 19 05/10/06
- B.J. Reich, E. Kalendra, C.B. Storlie, H.D. Bondell, M. Fuentes (2008), Variable selection for high-dimensional Bayesian density estimation: Application to human exposure simulation, *Institute for Statistical Mimeo Series #2616*.
- R.A. Rigby & D.M. Stasinopoulos (2000), Construction of reference centiles using mean and dispersion additive models, *The Statistician*, 49:41-50.
- R.A. Rigby & D.M. Stasinopoulos (2005), Generalized additive models for location, scale and shape, *Applied Statistics*, 54:507-554.
- E. Volkova, B. Iooss and F. Van Dorpe (2008), Global sensitivity analysis for a numerical model of radionuclide migration from the "RRC Kurchatov Institute" radwaste disposal site, *Stochastic Environmental Research and Risk Assessment*, 22:17-31.
- I. Zabalza-Mezghani (2000), *Analyse statistique et planification d'expérience en ingénierie de réservoir*, Thèse de l'Université de Pau et des Pays de l'Adour.