



Sujet de thèse

Quantification robuste de l'incertitude sur une mesure de risque issue d'un code de calcul

CONTEXTE :

Le Département Performance, Risque Industriel, Surveillance et Maintenance pour l'Exploitation (PRISME) d'EDF R&D participe à une mission essentielle du Groupe EDF : apporter un appui scientifique aux directions de production et d'ingénierie pour optimiser les performances des installations de production d'électricité, tout en garantissant leur sûreté.

Dans cette optique, l'estimation de quantiles extrêmes de lois de probabilités représentant l'incertitude sur les grandeurs physiques critiques pour la sûreté d'une installation est une problématique récurrente, dans le but de démontrer que de tels quantiles demeurent en deçà des bornes réglementaires.

Un tel calcul statistique s'appuie nécessairement sur des hypothèses concernant les différentes sources d'incertitudes à prendre en compte, hypothèses régulièrement remises en causes par les autorités régulatrices destinataires des dossiers de sûreté.

C'est pourquoi EDF souhaite désormais développer une approche robuste du traitement de l'incertitude, permettant de s'affranchir autant que possible d'hypothèses fortes sur les sources d'incertitudes mal connues (incertitudes dites « épistémiques »). Les applications à court terme seront issues de problématiques hydrauliques (calcul de crues avec Telemac 2D) et thermohydrauliques (avec code CATHARE). Les méthodes développées pourront néanmoins servir dans tout dossier de sûreté s'appuyant sur des argumentaires statistiques.

OBJECTIFS DE LA THESE :

Le but de la thèse est de développer une méthode de quantification robuste pour le calcul de bornes sur un indicateur statistique extrême d'une variable d'intérêt. Celle-ci est la sortie d'un code de calcul (potentiellement coûteux), dont les entrées sont modélisées par des variables aléatoires. Les lois de ces variables d'entrée ne sont qu'imparfaitement connues, la seule information dont on dispose étant qu'elles satisfont à certaines contraintes sur leurs moments.

Pour ce faire, on envisage de transposer les techniques bien établies pour l'établissement de bornes robustes relatives aux probabilités de dépassement de seuil, en inversant ces bornes pour encadrer les quantiles correspondants. Cette approche nécessitant de nombreux appels au code, l'utilisation d'un métamodèle est également envisagée afin de diminuer les temps de calculs.

L'indicateur statistique utilisé dans la pratique est le quantile. Ce dernier peut n'être cependant pas adapté à toutes les situations. Un autre objectif de la thèse sera donc d'investiguer la piste du superquantile comme mesure de risque. Sur ce sujet peu investigué en analyse d'incertitudes de modèle, il conviendra de développer des méthodologies d'estimation, de contrôle de l'erreur d'estimation et d'analyse de sensibilité. Le cadre décisionnel sous-jacent sera également analysé.

DEROULEMENT :

Dans un premier temps, une étude bibliographique permettra au doctorant de s'approprier les techniques d'inférence robuste, telles que l' 'Optimal Uncertainty Quantification', ainsi que la construction de métamodèles par krigeage. On s'intéressera en particulier aux métamodèles séquentiels dits « goal-oriented », adaptés à l'optimisation d'un critère ou à l'identification d'ensembles d'excursion. Cette étude devra conduire dans un deuxième temps à l'établissement d'une méthode de construction de bornes robustes sur un quantile incertain. Sur le sujet des superquantiles, une étude bibliographique devra être faite dans le domaine des calculs de risque financier, qui manipule ce type de quantités depuis longtemps. L'utilisation de données expérimentales pour réduire l'incertitude sur les lois des variables d'entrée pourra être envisagée, dans le cadre de l'inférence bayésienne robuste.

Les méthodes proposées seront ensuite implémentées, en s'appuyant sur les outils existants, tels qu'OpenTURNS bien entendu, mais aussi les packages 'Mystic' (Python) pour l'optimisation robuste, 'sensitivity' (R) pour l'analyse de sensibilité, et 'Dice-Kriging' (R), 'scikit-learn' ou 'Open-TURNS' (Python) pour le krigeage. Ces méthodes seront validées sur des données simulées, avant d'être appliquées sur un ou plusieurs cas industriels.

PROFIL DE L'ETUDIANT :

Etudiants de M2 probabilités / statistiques ou grandes écoles.

ENCADREMENT :

Thèse CIFRE réalisée en collaboration entre EDF R&D (Merlin Keller, Bertrand Iooss, Thibault Delage) et l'Institut de Mathématiques de Toulouse de l'Univ. Toulouse III (Fabrice Gamboa).

LIEU DE REALISATION DE LA THESE :

Localisation principale à EDF Lab Chatou. Des séjours de travail à l' Institut de Mathématiques de Toulouse seront mis en place.

COMPETENCES SOUHAITEES :

Cursus solide en probabilités et statistiques. Des compétences relatives à l'optimisation et à l'inférence robuste seraient un plus. Aisance en informatique.

QUELQUES REFERENCES BIBLIOGRAPHIQUES :

- Owhadi, H., Scovel, C., Sullivan, T. J., McKerns, M., and Ortiz, M. (2013). *Optimal Uncertainty Quantification*. SIAM Rev., 2(55), 271-345.
- Rios Insua, D. and Ruggeri, F. (2000). *Robust Bayesian Analysis*. Springer-Verlag.
- Donald R. Jones, Matthias Schonlau, and William J. Welch. (1998). *Efficient Global Optimization of Expensive Black-Box Functions*. *J. of Global Optimization* 13(4), 455-492.
- Tyrrell Rockafellar, R. and O. Royset, Johannes (2013). *Superquantiles and Their Applications to Risk, Random Variables, and Regression*, Theory Driven by Influential Applications, Chapter 8, 151-167.

CONTACTS :

merlin.keller@edf.fr et bertrand.iooss@edf.fr