

Sujet de stage de fin d'études à EDF R&D :

Algorithmes d'échantillonnage pour la propagation d'incertitudes en régime de données limitées

Entreprise	EDF SA (EDF R&D)
Entité d'accueil	Département PRISME ¹ / Groupe P17 GAIA ²
Site principal	EDF Lab Chatou, 6 Quai Watier, 78401 Chatou
Co-encadrement	Elias FEKHARI (Ing.-Chercheur) Bertrand IOOSS (Chercheur Senior)
Date de début et durée du stage	T2/T3 2027 Durée de 6 mois
Mots-clés	Quantification d'incertitudes, Techniques d'échantillonnage, Plan d'expériences, Métamodèles, Quadrature bayésienne

Environnement de recherche industrielle à EDF R&D. Au sein d'EDF R&D, le département PRISME propose des solutions innovantes pour améliorer la performance et la sûreté des différents moyens de production d'électricité du groupe EDF. En particulier, le Groupe P17 GAIA héberge la compétence *Probabilités et Statistiques pour les Phénomènes Physiques* (PSPP) qui développe des méthodes et outils pour **modéliser, quantifier et propager** efficacement les **incertitudes** au sein de simulateurs numériques, ceci à des fins de maîtrise des risques, d'optimisation des performances en exploitation ou de conception robuste des moyens de production.

Motivations. Certaines études de propagation d'incertitudes liées à la démonstration de sûreté des installations nucléaires reposent sur des simulateurs numériques extrêmement coûteux en temps de calcul, par exemple des modèles de mécanique des fluides (CFD) nécessitant plusieurs jours de calcul haute performance. Ces simulateurs peuvent être vus comme une fonction « boîte noire » avec un vecteur de variables d'entrée (dont l'incertitude est modélisée par une loi de probabilité) et une variable de sortie étudiée. La propagation des incertitudes des variables d'entrée au travers du simulateur induit une incertitude sur les sorties étudiées, que l'on caractérise par des quantités statistiques (moyenne, variance, quantiles, etc.) qui peuvent alimenter ces démonstrations de sûreté. Cependant, lorsque le nombre de simulations réalisables est faible (par ex. quelques dizaines), les méthodes standard comme l'échantillonnage de Monte Carlo classique sont inefficace : les estimations obtenues sont trop imprécises pour être exploitables.

Enjeux scientifiques et objectifs du stage. Dans le domaine de la quantification d'incertitudes, une vaste littérature existe sur les méthodes d'échantillonnage permettant de réduire la variance des estimateurs Monte Carlo, notamment les plans d'expériences à bon remplissage de l'espace (e.g. les *hypercubes latins*), les techniques de *quasi-Monte Carlo* (SULLIVAN, 2015), et plus récemment la *quadrature bayésienne* (BQ) (BRIOL et al., 2019). Dans la communauté CFD, l'estimation des premiers moments d'une sortie incertaine est souvent réalisée par la

1. PRISME : Performance, Risque Industriel et Surveillance pour la Maintenance et l'Exploitation

2. GAIA : Gestion d'Actifs, Incertitudes et Apprentissage statistique

méthode du *deterministic sampling* (DS) (HESSLING, 2013), qui consiste à choisir un petit nombre de points déterministes qui reproduisent exactement les premiers moments des entrées, puis à propager ces points au travers du simulateur. Ces méthodes produisent des échantillons structurés, parfois déterministes, qui convergent plus vite que la méthode de Monte Carlo sur des fonctions suffisamment régulières, mais ne s'accompagnent pas toujours de garanties théoriques (notamment en termes d'intervalles de confiance sur les estimations obtenues). Elles sont de plus souvent conçues pour l'estimation de l'espérance des sorties d'intérêt, alors que d'autres quantités sont déterminantes en pratique (moments d'ordre supérieur, quantiles, etc.).

L'objectif de ce stage sera d'étudier plusieurs méthodes d'échantillonnage dans un contexte de faible taille d'échantillons, afin de proposer des outils de propagation d'incertitudes, à la fois efficaces et accompagnés de garanties d'estimation. Une première étape consistera à étudier le DS, et de le positionner parmi les méthodes d'échantillonnage usuelles en quantification d'incertitudes, en identifiant ses hypothèses implicites et son domaine de validité. Ensuite, un benchmark numérique sur des fonctions test analytiques comparant un panel de méthodes permettra d'évaluer leur efficacité pour estimer, au-delà de l'espérance, d'autres quantités d'intérêt (variance et quantiles), en fonction de la taille d'échantillon, de la dimension et de la régularité de la fonction. Enfin, un travail théorique visera à associer des garanties d'estimation (idéalement des intervalles de confiance valables à taille d'échantillon finie) à la méthode BQ. En effet, cette méthode s'appuie sur une hypothèse de processus gaussien sous-jacent (RASMUSSEN et WILLIAMS, 2006), ce qui fournit naturellement une incertitude a posteriori sur l'estimation (intervalle de crédibilité). Néanmoins, la fiabilité de cet intervalle dépend de l'adéquation du processus gaussien (type de noyau, hyperparamètres) à la fonction réellement étudiée.

Planning prévisionnel. Le déroulement de ce stage pourra s'organiser en 4 étapes :

- (i) Analyse bibliographique autour des méthodes d'échantillonnage citées,
- (ii) Implémentation en Python de la méthode DS,
- (iii) Réalisation d'un benchmark numérique sur des cas analytiques,
- (iv) Travail théorique sur les garanties d'estimation de la BQ.

Ce sujet s'inscrit dans le projet [GUNSROSES](#)³ de l'Agence Nationale de la Recherche. Selon les résultats obtenus, une valorisation sous forme de publication scientifique pourra être envisagée.

Le montant de la gratification s'élève à environ 1 500 euros bruts par mois. Cette somme peut être cumulée avec une indemnité de logement de 185 euros mensuels en cas de stage nécessitant un changement de région.

Profil recherché.

- Étudiant.e en dernière année d'école d'ingénieur.e.s ou M2 (stage de fin d'études ou de césure)
- Compétences en mathématiques appliquées, probabilités & statistiques, programmation Python

Comment postuler ? Envoyer un CV détaillé et une lettre de motivation par mail aux adresses des encadrants (prenom.nom@edf.fr). Pour enrichir votre candidature, n'hésitez pas à nous faire part de vos relevés de notes ou de rapports de projets/stages scientifiques pertinents.

Références

- BRIOL, F.X., C. OATES, M. GIROLAMI, M. OSBORNE et D. SEJDINOVIC (2019). "Probabilistic Integration: A Role in Statistical Computation?" In : *Statistical Science* 34.1, p. 1 -22.
- HESSLING, J.P. (2013). "Deterministic sampling for propagating model covariance". In : *SIAM/ASA Journal on Uncertainty Quantification* 1.1, p. 297-318.
- RASMUSSEN, C. E. et C. K. I. WILLIAMS (2006). *Gaussian processes for machine learning*. MIT Press.
- SULLIVAN, T. J. (2015). *Introduction to Uncertainty Quantification*. T. 63. Texts in Applied Mathematics. Springer International Publishing Switzerland.

3. *Guarantees and UNcertainties on expensive Simulations Reach Operational SafEty Studies*.