



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

### Certification d'algorithmes de simulation et d'estimation d'événements rares : analyse de mesures de risques adaptées et application à des systèmes industriels critiques

<b>Entreprise</b>	EDF SA (EDF R&D)
<b>Entité d'accueil</b>	Département PRISME <sup>1</sup> / Groupe P17 GAIA <sup>2</sup>
<b>Lieu principal</b>	EDF Lab Chatou, 6 Quai Watier, 78401 Chatou
<b>Co-encadrement</b>	Vincent CHABRIDON (Ing.-Chercheur) Julien PELAMATTI (Ing.-Chercheur)
<b>Partenariat académique</b>	Guillaume PERRIN <sup>3</sup> (Chercheur) Université Gustave Eiffel, 77420 Champs-sur-Marne
<b>Date de début et durée du stage</b>	Début possible > 01/03/2024   Durée de 6 mois
<b>Mots-clés</b>	Quantification d'incertitudes, Mesures de risques, Certification, Robustesse, Analyse de sensibilité, Événements rares

**Environnement de recherche industrielle à EDF R&D / PRISME.** Au sein d'EDF R&D, le département PRISME a pour mission de proposer des solutions innovantes pour une exploitation plus performante des différents moyens de production d'électricité du groupe EDF : maîtrise des risques, amélioration de la sûreté, optimisation des performances technico-économiques et maîtrise de la durée de vie des installations. La thématique du **traitement des incertitudes** (*uncertainty quantification*) désigne l'ensemble des méthodes probabilistes et statistiques dédiées à la modélisation des incertitudes, à leur quantification et à leur propagation dans les codes de simulation numérique. En complément de cette thématique, dans le cas où des données d'observations (mesures expérimentales, données de fonctionnement) sont à disposition, des approches de **traitement de la donnée** (apprentissage statistique ou *machine learning*) peuvent être utilisées afin de conforter la modélisation physique et la modélisation probabiliste des variables d'entrée. Ces deux thématiques, traitement des incertitudes et traitement de la donnée, font l'objet de travaux de recherche intenses au sein du département PRISME et jouent un rôle central et transverse au sein de nombreuses activités du groupe EDF.

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

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

3. <https://pagespro.univ-gustave-eiffel.fr/guillaume-perrin>

**Contexte général et problématique industrielle.** Dans le contexte de l'évaluation de la fiabilité des systèmes complexes de production d'électricité (nucléaires, hydrauliques, éoliens, etc.) à forts enjeux de sûreté, l'exploitant EDF doit mener des analyses de risques en tenant compte des différentes sources d'incertitudes afin de garantir, par la simulation numérique, des marges de sûreté vis-à-vis des autorités de tutelle. Par exemple, les démarches historiques d'évaluation des marges de sûreté pour les scénarios accidentels d'un cœur de réacteur nucléaire étaient basées sur des approches "déterministes" et des modèles simplifiés visant à trouver un "pire cas", c'est-à-dire, à déterminer une simulation la plus pénalisante afin de déterminer une marge déterministe pour la sûreté. De profondes améliorations de la fidélité des codes de simulation numérique, ainsi que la prise en compte des diverses sources d'incertitudes, ont amené les exploitants et les autorités à mettre en place des approches dites "BEPU" pour "*Best Estimate Plus Uncertainty*" (PROŠEK et MAVKO, 2007 ; LARGET, 2019).

D'un point de vue méthodologique, toutes ces études de sûreté s'articulent généralement en quatre étapes, conformément à la démarche méthodologique traditionnellement adoptée dans le domaine du traitement des incertitudes dans les codes de calcul (DE ROCQUIGNY et al., 2008 ; SULLIVAN, 2015) :

- 1) la définition du périmètre relevant du modèle de simulation numérique (code multi-physique), noté  $G : \mathbb{R}^d \rightarrow \mathbb{R}$ , de son vecteur des variables d'entrée, noté  $\mathbf{x} \in \mathbb{R}^d$ , ainsi que d'une (ou plusieurs, comme une sortie vectorielle ou fonctionnelle) variable d'intérêt en sortie notée  $y = G(\mathbf{x})$  ;
- 2) la quantification des différentes sources d'incertitudes en entrée consiste à construire un modèle probabiliste sur les entrées (on notera  $\mathbf{X} \sim \mathbb{P}_{\mathbf{X}}$  où  $\mathbb{P}_{\mathbf{X}}$  est la loi de probabilité conjointe des entrées). Par construction, le problème déterministe initial est remplacé par un problème du type  $Y = G(\mathbf{X})$  (où la sortie est elle-même une variable aléatoire, supposée scalaire ici pour plus de simplicité). Dès lors, on définit la quantité statistique d'intérêt (*quantity of interest*, QoI), typiquement une mesure de risque sur la sortie  $\mathcal{R}(Y)$  qui caractérise la défaillance du système. Dans la pratique,  $\mathcal{R}(Y)$  est soit une probabilité de dépasser un seuil  $y_{th} \in \mathbb{R}$  donné, i.e.,  $\mathcal{R}(Y) := \mathbb{P}(Y > y_{th})$ , soit un quantile d'ordre  $\alpha \in ]0, 1[$  tel que  $\mathcal{R}(Y) := q_{\alpha}(Y)$ . Toutefois, d'autres mesures de risques existent et peuvent avoir un intérêt pour les applications industrielles (ROCKAFELLAR et al., 2014 ; ROCKAFELLAR et ROYSET, 2015) ;
- 3) la propagation des incertitudes des entrées vers la sortie, et l'estimation de la QoI visée en sortie consiste généralement à mettre en œuvre des algorithmes d'échantillonnage aléatoire (typiquement, à base de méthodes Monte-Carlo, tels que les techniques de réduction de variance ou *Importance Sampling*, et d'échantillonnage multi-niveaux ou *Splitting*) adaptés à la mesure de risque choisie ainsi qu'aux contraintes imposées par la complexité du code de calcul (coût de simulation unitaire, parallélisation, nonlinéarité, stochasticité, etc.) (RUBINSTEIN et KROESE, 2008 ; MORIO et BALESSENT, 2015) ;
- 4) enfin, une analyse de sensibilité (approche inverse) peut être réalisée dans le but de déterminer quelles sont, parmi les variables d'entrée incertaines, celles qui influencent le plus la variabilité de la QoI en sortie (LOOSS et LEMAÎTRE, 2015 ; DA VEIGA et al., 2021). En pratique, ces différentes étapes peuvent s'enchaîner linéairement ou bien être combinées à travers différents bouclages entre elles. Un panorama récapitulatif de toutes ces étapes peut être trouvé dans BLANCHARD et al. (2023).

**Enjeux.** Dans les études de fiabilité et de sûreté industrielles, on observe une certaine disparité dans les pratiques en fonction du domaine d'application, notamment dans le choix de la formalisation du problème (par exemple, dans le choix de la mesure de risque) ou dans la mise en œuvre des algorithmes d'estimation associés. Certains domaines (la sûreté nucléaire, par exemple) visent un degré de conservatisme et s'appuient beaucoup sur la notion de quantile de Wilks (WILKS, 1941) qui, historiquement, a permis d'apporter une notion de "marge au risque". Une autre stratégie consiste aussi à pénaliser certaines variables d'entrée afin de rendre le résultat de la simulation le plus pessimiste pos-

sible (CHARIGNON et al., 2016 ; LARGET, 2019 ; MARREL et al., 2022). Dans d'autres domaines (par exemple, la sûreté des installations hydroélectriques), c'est la notion de probabilité de défaillance qui est utilisée et les études visent plutôt à établir un consensus dans l'estimation statistique de ces probabilités de ruine très rares (AJENJO et al., 2023). Un point commun ressort de ces deux exemples : il est bien souvent difficile de donner des garanties fortes sur l'estimation de la mesure de risque considérée (quantile ou probabilité de défaillance) tant ces quantités sont très sensibles à la définition des variables incertaines, au scénario étudié, à la variabilité statistique des estimateurs, à la technique d'échantillonnage utilisée ou tout simplement, au budget de calcul requis pour l'étude. Dès lors se pose la question de la certification de ces analyses de risque au regard d'une autorité de tutelle. Comment donner, non pas une preuve absolue, mais un faisceau de garanties permettant d'avoir confiance dans le résultat et dans sa significativité afin de permettre une prise de décision mieux informée ?

L'enjeu principal de ce stage est donc de travailler autour de la formalisation des problèmes d'analyse de risques et de fiabilité, de leur unification formelle, et de la proposition d'une démarche robuste permettant d'aller jusqu'à la certification d'un calcul de mesure de risque. Ainsi, l'apport méthodologique de ce travail viserait à proposer un argumentaire (mathématique et informatique) consolidé permettant de valider la démarche, et ce, de manière la plus générique qui soit au regard de la mesure de risque étudiée, du contexte algorithmique et de la finalité applicative. Afin de tester la démarche proposée, le/la stagiaire aura à disposition plusieurs cas d'usage industriels typiques (simulation d'un scénario accidentel pour la sûreté nucléaire, analyse d'un risque de défaillance en hydraulique, analyse de fiabilité d'une éolienne) qu'il/elle devra s'approprier tout au long du stage, avec l'appui des ingénieurs-chercheurs d'EDF R&D.

**Contenu et planning prévisionnel.** Le déroulement de ce stage pourra s'organiser en 4 étapes :

- (i) Analyse bibliographique autour de la formulation de problème d'analyse de risques :
  - Formulation mathématique des différentes mesures de risque (par exemple, quantile, probabilité de dépassement de seuil, super-quantile, etc.), analyse des propriétés théoriques des estimateurs et des algorithmes dédiés (LABOPIN-RICHARD, 2016) ;
  - Proposition d'un critère de certification associé à une telle mesure de risque (ARIEF et al., 2021)
- (ii) Travail autour de la quantification des différentes sources d'incertitudes et de leur influence sur la mesure de risque considérée :
  - Inventaire des différentes sources d'incertitudes entrant en jeu dans une analyse de risques (incertitudes aléatoires liées à la variabilité des grandeurs physiques, incertitudes épistémiques potentiellement réductibles – par exemple, incertitude statistique liée à l'estimation des lois d'entrée, incertitudes liées au scénario étudié ainsi qu'aux modélisations, erreurs d'approximation et de réduction de dimension, etc.) ;
  - Développement d'une démarche d'analyse de sensibilité afin de quantifier l'importance des différentes sources d'incertitudes sur la QoI (via des techniques adaptées aux mesures de risque) et éventuelle stratégie de réduction des incertitudes épistémiques (si possible) (CHABRIDON, 2018 ; PERRIN et DEFAUX, 2019) ;
- (iii) Vérification et validation vis-à-vis du critère de décision :
  - Définition d'une stratégie d'estimation de la quantité d'intérêt et quantification de la marge vis-à-vis du critère de décision (lien avec les différents niveaux de risque en tests statistiques).
- (iv) Mise en œuvre informatique et application à un cas d'usage industriel :

- Développements informatiques à l'aide d'outils open source (langage Python et plateforme OpenTURNS) (BAUDIN et al., 2017) ;
- Application de la méthodologie à un ou plusieurs (en fonction de l'avancement du stage) cas d'usage industriels identifiés.

**Profil recherché.** Étudiant.e de M2 (mathématiques appliquées / probabilités-statistiques / modélisation & simulation numérique) ou d'écoles d'ingénieur.e.s (avec majeure en mathématiques appliquées / probabilités-statistiques).

**Comment postuler ?** Envoyer un CV détaillé et une lettre de motivation sur <https://www.edf.fr/edf-recrute> ou via les adresses suivantes, en précisant l'objet de votre envoi :

- vincent dot chabridon at edf dot fr
- julien dot pelamatti at edf dot fr
- guillaume dot perrin at univ-eiffel dot fr

## Références

- AJENJO, A., E. ARDILLON, V. CHABRIDON, S. COGAN et E. SADOULET-REBOUL (2023). "Robustness evaluation of the reliability of penstocks combining line sampling and neural networks". In : *Reliability Engineering and System Safety* 234, p. 109192.
- ARIEF, M., Y. BAI, W. DING, S. HE, Z. HUANG, H. LAM et D. ZHAO (2021). *Certifiable Deep Importance Sampling for Rare-Event Simulation of Black-Box Systems*. arXiv : 2111.02204 [stat.ME].
- BAUDIN, M., A. DUTFOY, B. IOOSS et A.-L. POPELIN (2017). "OpenTURNS: An Industrial Software for Uncertainty Quantification in Simulation". In : *Handbook of Uncertainty Quantification*. Sous la dir. de R. GHANEM, D. HIGDON et H. OWHADI. Cham, Switzerland : Springer International Publishing, p. 2001-2038. DOI : 10.1007/978-3-319-11259-6\_64-1.
- BLANCHARD, J.-B. et al. (2023). "Fiches pédagogiques sur le traitement des incertitudes dans les codes de calcul (I3P : EDF - CEA - Framatome)". In : *hal-04205632*, p. 1-36.
- CHABRIDON, V. (2018). "Reliability-oriented sensitivity analysis under probabilistic model uncertainty – Application to aerospace systems". Thèse de doct. Université Clermont Auvergne.
- CHARIGNON, C., J.-C. LECOY et J.-Y. SAUVAGE (2016). "CathSBI, a new methodology for the revised French LOCA rules". In : *Proc. of the 11th International Topical Meeting on Nuclear Reactor Thermal Hydraulics, Operations and Safety (NUTHOS-11)*. Gyeongju, Korea.
- DA VEIGA, S., F. GAMBOA, B. IOOSS et C. PRIEUR (2021). *Basics and Trends in Sensitivity Analysis: Theory and Practice in R*. Society for Industrial et Applied Mathematics.
- DE ROCQUIGNY, E., N. DEVICTOR et S. TARANTOLA (2008). *Uncertainty in industrial practice: a guide to quantitative uncertainty management*. Wiley.
- IOOSS, B. et P. LEMAÎTRE (2015). "A Review on Global Sensitivity Analysis Methods". In : *Uncertainty Management in Simulation-Optimization of Complex Systems: Algorithms and Applications*. Sous la dir. de G. DELLINO et C. MELONI. Boston, MA : Springer US. Chap. 5, p. 101-122.
- LABOPIN-RICHARD, T. (2016). "Méthodes statistiques et d'apprentissage pour l'estimation de quantiles et de superquantiles dans des modèles de codes numériques ou stochastiques". Thèse de doct. Université de Toulouse.

- LARGET, V. (2019). "How to bring conservatism to a BEPU analysis". In : *Proc. of the 18th International Topical Meeting on Nuclear Reactor Thermal Hydraulics (NURETH-18)*. Portland, OR, USA.
- MARREL, A., B. IOOSS et V. CHABRIDON (2022). "The ICSCREAM Methodology: Identification of Penalizing Configurations in Computer Experiments Using Screening and Metamodel – Applications in Thermal-Hydraulics". In : *Nuclear Science and Engineering* 196.3, p. 301-321.
- MORIO, J. et M. BALESSENT (2015). *Estimation of Rare Event Probabilities in Complex Aerospace and Other Systems: A Practical Approach*. Woodhead Publishing, Elsevier.
- PERRIN, G. et G. DEFAUX (2019). "Efficient Evaluation of Reliability-Oriented Sensitivity Indices". In : *Journal of Scientific Computing*, p. 1-23.
- PROŠEK, A. et B. MAVKO (2007). "The State-of-the-Art Theory and Applications of Best-Estimate Plus Uncertainty Methods". In : *Nuclear Technology* 158, p. 69-79.
- ROCKAFELLAR, R. T. et J. O. ROYSET (2015). "Engineering Decisions under Risk Averseness". In : *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* 1.2, p. 1-12.
- ROCKAFELLAR, R. T., J. O. ROYSET et S. I. MIRANDA (2014). "Superquantile regression with applications to buffered reliability, uncertainty quantification, and conditional value-at-risk". In : *European Journal of Operational Research* 234.1, p. 140-154.
- RUBINSTEIN, R. Y. et D. P. KROESE (2008). *Simulation and the Monte Carlo Method*. Second ed. Wiley.
- SULLIVAN, T. J. (2015). *Introduction to Uncertainty Quantification*. T. 63. Texts in Applied Mathematics. Springer International Publishing Switzerland.
- WILKS, S. S. (1941). "Determination of Sample Sizes for Setting Tolerance Limits". In : *The Annals of Mathematical Statistics* 12.1, p. 91-96.