

## PROPOSITION DE SUJET DE THÈSE

**Intitulé : Prédiction conforme pour l'estimation d'évènements rares par apprentissage actif**

Référence : **SNA-DTIS-2026-17**  
(à rappeler dans toute correspondance)

Début de la thèse : Octobre 2026

Date limite de candidature : Juillet 2026

### Mots clés

Prédiction conforme, estimation d'évènements rares, quantification d'incertitudes, intervalle de confiance, apprentissage actif, processus gaussien.

### Profil et compétences recherchées

Profil : Master 2 et/ou dernière année d'école d'ingénieur généraliste.

Compétences : Mathématiques appliquées, probabilités, statistique, maîtrise d'un langage de programmation.

### Présentation du projet doctoral, contexte et objectif

De nombreux systèmes physiques sont schématiquement décrits par une relation du type  $Y = \phi(X)$ , où l'entrée multidimensionnelle  $X$  est supposée aléatoire et où la sortie  $Y$  est déterminée via la fonction déterministe  $\phi$ . Un exemple proéminent d'application est l'analyse d'un code de calcul boîte noire :  $\phi$  représente alors un code de calcul, tel que des calculs de contraintes sur des structures mécaniques complexes et  $X$  les conditions extérieures dans lesquelles ce calcul est effectué. On peut notamment penser à un code de type éléments finis, dont la complexité rend impossible toute étude analytique de la fonction  $\phi$  et donc de la sortie  $Y$ .

La défaillance de systèmes critiques peut engendrer des conséquences environnementales, économiques ou humaines potentiellement dramatiques. Il s'agit fort heureusement généralement d'évènements rares. Pour des raisons de sécurité et certification, il est néanmoins crucial de quantifier le risque associé à de tels systèmes : il s'agit de l'analyse de fiabilité. Mathématiquement parlant, la défaillance d'un système est modélisée par un évènement de la forme  $\{Y > t\}$ , où  $t$  est un seuil de défaillance défini par des considérations physiques. Le risque associé au système est alors quantifié par la probabilité de défaillance  $p_t = P(Y > t)$ . Cependant, de par la complexité du code de calcul  $\phi$ ,  $p_t$  ne peut pas être calculer analytiquement.

L'estimation de la probabilité de défaillance est aujourd'hui un thème de recherche qui est largement étudié dans la littérature (méthodes Monte-Carlo, échantillonnage préférentiel [1], subset simulation [2]). Par ailleurs, le code de calcul  $\phi$  étant potentiellement très coûteux en temps à évaluer, une famille répandue de méthodes d'estimation consiste à remplacer  $\phi$  par un modèle de substitution, comme un processus gaussien par exemple (algorithme AK-MCS [3]). L'erreur d'approximation du métamodèle est cependant difficile à maîtriser et ces méthodes ne permettent pas de quantifier pleinement l'erreur d'estimation de  $p_t$  induite par le métamodèle.

Dans ce contexte, **l'objectif de la thèse est de construire un intervalle de confiance valide de la probabilité de défaillance, lorsqu'elle est estimée à l'aide d'un modèle de substitution, qui prend en compte l'erreur d'approximation de  $\phi$** . La difficulté majeure de cette question provient de la méconnaissance totale du modèle  $\phi$ . Une piste de recherche intéressante est alors la prédiction conforme [4]. Il s'agit d'une méthode statistique permettant de quantifier l'incertitude de n'importe quel modèle de prédiction en construisant des ensembles de prédiction statistiquement valides. En particulier, une méthode conforme spécifique aux processus gaussiens a récemment été proposé [5,6]. En revanche, une telle approche n'a pas encore été explorée dans le cadre de la quantification de l'incertitude de l'estimation d'une quantité intégrée, telle que la probabilité de défaillance.

Une première étape de la thèse consistera à effectuer une revue bibliographique poussée sur l'analyse de fiabilité, les modèles de substitution et les méthodes de prédiction conforme. Un stage de projet de fin d'étude de niveau M2 sur un sujet proche pourra notamment servir de base de

départ à cette thèse. Puis, un premier axe de travail sera d'appliquer ou de développer une approche de prédiction conforme au méta-modèle utilisé et de la propager sur l'estimation de la probabilité de défaillance afin de construire un intervalle de confiance statistiquement valide. Un deuxième axe de travail sera de prendre en considération le caractère adaptatif des algorithmes de fiabilité, tel que l'algorithme AK-MCS. En effet, l'enrichissement actif du méta-modèle ne garantit plus l'hypothèse fondamentale d'échangeabilité des données. Des méthodes de prédiction conforme adaptées à cette situation seront à explorer [7] pour garantir la validité statistique de l'intervalle de confiance construit.

- [1] J. A. Bucklew. Introduction to rare event simulation. *Springer Series in Statistics*. Springer-Verlag, New York, 2004.
- [2] F. Cérou, P. Del Moral, T. Furon, and A. Guyader. Sequential Monte Carlo for Rare Event Estimation, *Statistics and Computing*, 22(3):795-808, 2012.
- [3] Echard, B., Gayton, N., & Lemaire, M. AK-MCS: an active learning reliability method combining Kriging and Monte Carlo simulation. *Structural safety*, 33(2), 145-154, 2011.
- [4] Angelopoulos, A. N., & Bates, S. ,A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *arXiv preprint arXiv:2107.07511*, 2021.
- [5] Jaber, E., Blot, V., Brunel, N. J., Chabridon, V., Remy, E., Iooss, B., & Leite, A. Conformal approach to gaussian process surrogate evaluation with marginal coverage guarantees. *Journal of Machine Learning for Modeling and Computing*, 6(3), 2025.
- [6] Pion, A., & Vazquez, E. Gaussian process interpolation with conformal prediction: methods and comparative analysis. In *International Conference on Machine Learning, Optimization, and Data Science*, 218-228, Springer Nature Switzerland, 2024.
- [7] Barber, R. F., Candes, E. J., Ramdas, A., & Tibshirani, R. J. Conformal prediction beyond exchangeability. *The Annals of Statistics*, 51(2), 816-845, 2023.

#### **Collaborations envisagées**

ISAE-SUPAERO/DISC (Iain Henderson)

<b>Laboratoire d'accueil à l'ONERA</b> Département : Traitement de l'information et Systèmes Lieu (centre ONERA) : Toulouse <b>Contact</b> : Julien Demange-Chrst Tél. : 05 62 25 28 11 Email : julien.demange-chrst@onera.fr	<b>Directeur de thèse</b> Nom : Jérôme Morio Laboratoire : ONERA/DTIS Tél. : 05 62 25 26 63 Email : jerome.morio@onera.fr
---	---

Pour plus d'informations : <https://www.onera.fr/rejoindre-onera/la-formation-par-la-recherche>