

# PROCESSUS GAUSSIENS ET PLANIFICATION SÉQUENTIELLE D'EXPÉRIENCES NUMÉRIQUES POUR L'OPTIMISATION ET L'INVERSION EN PRÉSENCE D'INCERTITUDES

## En bref

Ce document présente une proposition de thèse au [Laboratoire des Signaux et Systèmes](https://l2s.centralesupelec.fr/)<sup>1</sup>, unité mixte de recherche CNRS – CentraleSupélec – Université Paris-Saclay, dans un environnement stimulant (projet collaboratif ANR SAMOURAI, séminaire UQSay<sup>2</sup>, participation au groupe de recherche national MASCOT-NUM<sup>3</sup>...).

Le sujet s'adresse à un(e) étudiant(e) avec une formation en mathématiques appliquées / sciences des données / IA / statistique.

Lisez la suite pour en savoir plus !

## Contexte

L'utilisation de [simulations numériques](#) pour la conception de systèmes ou l'étude de phénomènes naturels est un enjeu très important dans l'industrie ou la recherche académique. Des simulations numériques sont, par exemple, utilisées pour maximiser les performances des systèmes, les rendre plus sûrs et respectueux de l'environnement. Mentionnons également le domaine de l'intelligence artificielle, dans lequel des simulations intensives sont utilisées pour optimiser les hyper-paramètres des algorithmes d'apprentissage statistique.

Toutefois, la modélisation numérique d'un système physique, ou l'évaluation des performances d'un algorithme d'apprentissage sur une large collection d'exemples, requiert souvent des ressources informatiques importantes. Il est alors nécessaire de conduire les simulations numériques avec parcimonie et de manière efficace.

Les méthodes de [planification séquentielle d'expériences](#), également connue sous le nom d'apprentissage actif (*active learning*) dans le domaine de l'apprentissage statistique, permettent de guider le choix des simulations à réaliser en s'appuyant sur des modèles prédictifs des quantités d'intérêts. Ces techniques sont au cœur du projet ANR SAMOURAI [13] dans le cadre duquel est proposé ce sujet de thèse. Partenaires : IFPEN, EDF R&D, Safran Tech, CEA, CentraleSupélec, EMSE et Polytechnique Montréal.

## Objectifs de la thèse

L'objectif de la thèse est développer de nouvelles stratégies de planification séquentielles d'expériences numériques pour des simulateurs numériques possédant des variables d'entrée incertaines, correspondant par exemple à des grandeurs fluctuant de manière non contrôlée.

---

1. <https://l2s.centralesupelec.fr/>

2. <https://www.uqsay.org/>

3. <https://www.gdr-mascotnum.fr/>

Considérons par exemple un problème de conception d’une éolienne. Sous l’effet du vent, du vieillissement des matériaux, etc., l’éolienne est soumise à des contraintes que l’on peut étudier via des simulations numériques en fonction des paramètres de conception. Formellement, nous pouvons modéliser une contrainte par une variable aléatoire

$$Z = f(u, X)$$

où  $f$  est la fonction correspondant au simulateur numérique,  $X$  est le vecteur aléatoire des variables incertaines dont la loi est supposée connue, et  $u$  est le vecteur des paramètres de conception de l’éolienne. Une simulation numérique correspond au fait de choisir un point de conception  $u$ , une réalisation  $x$  de la variable  $X$ , et d’évaluer  $f$  en ce point. En général, le concepteur de l’éolienne s’intéresse à des fonctions de la loi de  $Z$  en chaque point  $u$  : par exemple la probabilité que  $Z$  dépasse un seuil critique, ou encore des quantiles de  $Z$ . Des exemples de problèmes formulés de cette façon se trouvent notamment dans [6, 8–12].

Lorsque les évaluations de  $f$  sont coûteuses, il est très important de mettre en place des stratégies permettant de résoudre efficacement, à l’aide d’un budget limité d’appel au simulateur, des problèmes d’intérêt : par exemple, en supposant que la réponse d’intérêt est un quantile  $q_\alpha(u)$  de la loi de  $Z$  en  $u$ , on peut vouloir optimiser  $q_\alpha$  ou encore déterminer l’ensemble des valeurs de  $u$  pour lesquelles  $q_\alpha$  dépasse un seuil donné.

Dans la continuité de travaux menés dans l’équipe depuis une quinzaine d’année, cette thèse se concentre tout particulièrement sur des méthodes de planification séquentielles d’expériences numériques fondées sur une modélisation par [processus gaussien](#) du simulateur, et sur le paradigme de la [réduction séquentielle d’incertitude](#) (SUR) [1–3, 5, 14–16]. Dans le cas où le problème à traiter est un problème d’optimisation, de telles méthodes relèvent plus généralement du domaine de l’[optimisation bayésienne](#), un sujet qui fait l’objet de recherches actives dans plusieurs communautés scientifiques (statistique, machine learning, recherche opérationnelle. . .) depuis une vingtaine d’années.

Une difficulté centrale, du point de vue numérique, consistera à approcher les intégrales de grandes dimension apparaissant dans les critères SUR, en lien avec le nombre élevé de variables incertaines dans les problèmes considérés. Une seconde difficulté a trait à l’optimisation de ces critères, qui sont typiquement des fonctions très multi-modales. Les méthodes de Monte Carlo séquentielles (SMC), qui ont déjà fait leurs preuves dans des problèmes similaires [4, 7] sont une piste prometteuse pour résoudre ces deux problèmes conjointement.

Des cas d’application proposés par les partenaires du projet SAMOURAI, dans les domaines de l’éolien offshore (IFPEN), de la sûreté nucléaire (CEA, EDF R&D) et de la conception de systèmes aéronautiques (Safran), serviront à illustrer l’utilité pratique des méthodes développées.

## Informations pratiques

**Connaissances souhaitées.** Mathématiques appliquées, optimisation, statistiques, apprentissage. Très bonne maîtrise d’au moins un outil de programmation « numérique » standard (Matlab, R, Python. . .). Anglais niveau B2 minimum (C1 souhaité).

**Aptitudes personnelles souhaitées.** Autonomie, ouverture d’esprit, écoute, synthèse.

**Lieu de travail.** La thèse se déroulera au Laboratoire des Signaux & Systèmes (L2S), sur le campus de CentraleSupélec Paris-Saclay.

**Contact.** Envoyer curriculum vitae, lettre de motivation et relevés de notes à Julien Bect et Emmanuel Vazquez {julien.bect, emmanuel.vazquez}@centralesupelec.fr.

## References

- [1] A. Arnaud, J. Bect, M. Couplet, A. Pasanisi, and E. Vazquez. Évaluation d'un risque d'inondation fluviale par planification séquentielle d'expériences. In *42èmes Journées de Statistique*, Marseille, France, France, 2010.
- [2] J. Bect, F. Bachoc, and D. Ginsbourger. A supermartingale approach to Gaussian process based sequential design of experiments. *Bernoulli*, 25(4A):2883–2919, 2019.
- [3] J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22(3):773–793, 2012.
- [4] R. Benassi, J. Bect, and E. Vazquez. Bayesian optimization using sequential monte carlo. In *International Conference on Learning and Intelligent Optimization*, pages 339–342. Springer, 2012.
- [5] C. Chevalier, J. Bect, D. Ginsbourger, E. Vazquez, V. Picheny, and Yann Richet. Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. *Technometrics*, 56(4):455–465, 2014.
- [6] V. Dubourg, B. Sudret, and J.-M. Bourinet. Reliability-based design optimization using kriging surrogates and subset simulation. *Structural and Multidisciplinary Optimization*, 44(5):673–690, 2011.
- [7] P. Feliot, J. Bect, and E. Vazquez. A Bayesian approach to constrained single- and multi-objective optimization. *Journal of Global Optimization*, 67(1):97–133, 2017.
- [8] M. Fuhrländer and S. Schöps. A blackbox yield estimation workflow with Gaussian process regression applied to the design of electromagnetic devices. *Journal of Mathematics in Industry*, 10, 2020.
- [9] N. Lelièvre, P. Beaurepaire, C. Mattrand, N. Gayton, and A. Otsmane. On the consideration of uncertainty in design: optimization-reliability-robustness. *Structural and Multidisciplinary Optimization*, 54(6):1423–1437, 2016.
- [10] A. Marrel, B. Iooss, and V. Chabridon. Statistical identification of penalizing configurations in high-dimensional thermal-hydraulic numerical experiments: The IC-SCREAM methodology. arXiv preprint arXiv:2004.04663, 2020.
- [11] M. Moustapha, B. Sudret, J.-M. Bourinet, and B. Guillaume. Quantile-based optimization under uncertainties using adaptive kriging surrogate models. *Structural and multidisciplinary optimization*, 54(6):1403–1421, 2016.
- [12] G. Pujol, R. Le Riche, X. Bay, and O. Roustant. Minimisation de quantiles-application en mécanique. In *9ème colloque national en calcul des structure, 25–29 mai 2009, Giens, France*, 2009.
- [13] D. Sinoquet et al. SAMOURAI — Simulation Analytics and Meta-model-based solutions for Optimization, Uncertainty and Reliability Analysis. Projet ANR AAPG 2020, PRCE, débuté en 2021.
- [14] E. Vazquez and M. Piera-Martinez. Estimation du volume des ensembles d'excursion d'un processus gaussien par krigeage intrinsèque. In *39ème Journées de Statistiques Conférence Journée de Statistiques, Angers France*, 2007.
- [15] J. Villemonteix. *Optimisation de fonctions coûteuses Modèles gaussiens pour une utilisation efficace du budget d'évaluations: théorie et pratique industrielle*. PhD thesis, Université Paris Sud-Paris XI, 2008.
- [16] J. Villemonteix, E. Vazquez, and É. Walter. An informational approach to the global optimization of expensive-to-evaluate functions. *Journal of Global Optimization*, 44(4):509–534, 2009.