

Sujet de stage : Méthodes bayésiennes approchées pour le calage de modèles de simulation numériques, application à des modèles de sûreté



Pilotes : Merlin Keller, Michaël Baudin, Pascal Borel (EDF R&D)

Contexte : Pour un industriel comme EDF, la simulation numérique est un outil incontournable pour prédire le comportement de systèmes physiques critiques, que ce soit en situation d'exploitation normale, ou dans des situations extrêmes. Dans tous les cas, la question de la validité et de la précision des résultats d'une simulation vis-à-vis de la réalité qu'elle est censée reproduire est cruciale. Celle-ci fait l'objet d'une démarche maintenant bien établie, connue sous l'acronyme VVUQ (Verification, Validation and Uncertainty Quantification) [NRC, 2012]. En particulier, nous nous intéressons à l'étape de calage (ou calibration) de cette démarche, qui fait partie intégrante de la validation. En quelques mots, le calage consiste à réduire l'écart entre les simulations et les mesures expérimentales disponibles de la sortie d'intérêt, en ajustant les valeurs des paramètres physiques mal connus du modèle de simulation.

En pratique, la problématique du calage se retrouve dans un grand nombre de cas pratiques, et avec une grande diversité, que ce soit en terme de régularité, de complexité du modèle de simulation numérique, des dimensions d'entrées et ou de sortie du modèle, ainsi que du nombre de paramètres incertains à caler. Le cadre bayésien fournit un cadre théorique unifié pour formaliser l'opération de calage, et il permet également d'orienter la recherche de méthodologies génériques, permettant d'aborder les cas rencontrés dans toute leur diversité [Damblin, 2015]. Mais celles-ci restent très coûteuses à mettre en œuvre, car elles impliquent d'échantillonner la loi, dite *a posteriori*, de tous les paramètres à caler, compatibles en un certain sens avec les mesures expérimentales disponibles. Les méthodes usuelles, de type Monte-Carlo par chaîne de Markov (MCMC) [Robert and Casella, 2004], restent limitées car non parallélisables, sujettes à convergence vers un extrêum local de la fonction de vraisemblance, et gourmandes en évaluation de cette même vraisemblance.

Enjeu : L'essor récent des méthodes Approximate Bayesian Computation (ABC) [Marin et al, 2012] fournit une alternative très intéressante aux algorithmes MCMC : au prix d'une approximation, dont on peut maîtriser le degré, sur la caractérisation de la loi *a posteriori* des paramètres à caler, elles permettent d'échantillonner cette loi, d'une façon facilement parallélisable, et en explorant tout l'espace des paramètres, donc sans risque de convergence vers un extrêum local. Enfin, ces méthodes ne nécessitent aucune évaluation de la fonction de vraisemblance. Le but de ce stage est de tester l'applicabilité de telles techniques à des problèmes de calage, en les comparant aux approches MCMC plus classiques.

Contenu : La première partie du stage sera consacrée à une étude bibliographique des méthodes ABC, en particulier les versions les plus avancées, issues d'un couplage avec les techniques MCMC [Marjoram et al, 2003] puis de Monte-Carlo séquentiel (SMC) [Del Moral et al., 2009, Dau and Chopin, 2022], qui semblent fournir une solution générique et très performante aux problématiques de calage. Dans un second temps, le ou la stagiaire proposera une implémentation de ces méthodes et les testera sur plusieurs jeux de données, simulés et/ou issus de cas d'application industriels concrets. Il ou elle les comparera aux méthodes MCMC disponibles par ailleurs, et consignera les résultats dans le rapport final.

Perspectives : Les méthodes développées lors du stage ont vocation à venir enrichir la librairie de quantification d'incertitudes Open TURNS développée en consortium par EDF et d'autres acteurs du monde industriel [Baudin et al. 2012].

Profil recherché : M2 ou dernière année d'école d'ingénieur, avec des bases solides en probabilités et statistiques. Une connaissance de la problématique de calage de modèles physiques, et / ou des méthodes d'inférence bayésienne seraient un plus. Enfin, l'ouverture d'esprit, la curiosité et le goût pour le travail en équipe disciplinaire sont importants pour ce sujet à cheval sur plusieurs champs d'expertise. Cadre informatique privilégié : Linux, Python.

Durée du stage : 6 mois

Date de début de stage : 01/04/2026 (modifiable si besoin)

Lieu du stage : EDF Lab Chatou (Yvelines)

Contact : merlin.keller@edf.fr

Références :

NRC (2012). Assessing the Reliability of Complex Models : Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification. Washington, D.C. : The National Academies Press.

Damblin, G. (2015). Contributions statistiques au calage et à la validation des codes de calcul. Thèse de l'Université Paris-Saclay.

C.P. Robert and G. Casella (2004). Monte Carlo statistical methods. Springer Verlag,

Jean-Michel Marin, Pierre Pudlo, Christian P. Robert, and Robin J. Ryder (2012). Approximate Bayesian computational methods. Statistics and Computing, 22(6):1167–1180.

Pierre Del Moral, Arnaud Doucet, and Ajay Jasra (2009). An adaptive sequential Monte Carlo method for approximate bayesian computation. Statistics and Computing, 22, 01

Michaël Baudin, Anne Dutfoy, Bertrand Iooss, and Anne-Laure Popelin (2017). OpenTURNS: An Industrial Software for Uncertainty Quantification in Simulation R. Ghanem et al. (eds.), Handbook of Uncertainty Quantification, 6_64-1

Paul Marjoram, John Molitor, Vincent Plagnol and Simon Tavaré (2003). Markov-chain Monte-Carlo without likelihoods, PNAS, 100(26), 15324-15328.

Hai-Dang Dau, Nicolas Chopin (2022). Waste-Free Sequential Monte-Carlo. JRSSB, 84(1):114-148