Sujet de recherche :
Algorithmes d'estimation de probabilité de défaillance utilisant des méta-modèles multi-fidélité, dans un contexte d'hydrodynamique fluviale

13 septembre 2023

LARTISSTE
research initiative on
uncertainty quantification
@paris-saclay

Porteurs : Sophie Ricci (CERFACS) & Paul Mycek (CERFACS) & Romain Espoeys (CERFACS/ONERA), Mathieu Balesdent (ONERA), Loïc Brévault (ONERA), Didier Lucor (LISN)

Stade de Maturité : analyse bibliographique déjà faite (méta-modèles multi-fidélité, enrichissement de surrogate), analyse de méthodes existantes, disponibilité de cas d'étude jouets et industriels

1 Sujet de recherche pour le stage

La quantification d'incertitudes de systèmes modélisés par des solveurs coûteux en temps de calcul est une tâche difficile. L'analyse de fiabilité consiste à estimer la probabilité de défaillance d'un système en tenant compte de différentes sources d'incertitude, c'est à dire, estimer une probabilité pour des événements rares. Dans le cadre du stage, on s'intéresse à la probabilité de dépassement d'une digue de protection d'une plaine d'inondation, étant donné des incertitudes inhérentes à la description du frottement dans le lit de la rivière et dans la plaine ainsi qu'à la description de la condition limite amont prescrite par un débit d'apport constant. On propose ici d'estimer la probabilité de défaillance à l'aide de méta-modèles par processus gaussiens enrichis par active learning, avec une approche multi-fidélité, c'est à dire en s'appuyant sur une hiérarchie de modèles de fidélités différentes. Ce sujet s'inscrit dans la thématique plus large de construction de métamodèle pour des modèles complexes et de l'obtention de garantie sur ces métamodèles.

2 Contexte

Soit $F(\cdot)$ un solveur haute-fidélité (précis mais coûteux à évaluer) permettant de simuler le phénomène physique étudié. La défaillance du système peut être définie par les valeurs de sortie de $F : \mathbb{R}^d \to \mathbb{R}$ au-dessus d'un seuil donné T. $F(\cdot)$ est souvent appelée "fonction d'état limite" et $F(\cdot) = T$ est appelé "état limite", définissant la frontière entre état de sécurité et état de défaillance. Dans la littérature, les incertitudes sont décrites à l'aide d'un formalisme probabiliste [17, 19]. Dans ce cadre, en considérant un vecteur de paramètres incertains $\mathbf{X}$ modélisé par une densité de probabilité (PDF) jointe $\phi_{\mathbf{X}}(\cdot)$, l'objectif est de déterminer aussi précisément que possible la probabilité de défaillance P_f sur le domaine incertain Ω de telle sorte que

$$P_f = \int_{\Omega} \mathbf{1}_{F(\mathbf{x})>T} \phi_{\mathbf{X}}(\mathbf{x}) \mathrm{d}\mathbf{x}, \tag{1}$$

où $\mathbf{1}_{F(\mathbf{x})>T}$ est la fonction indicatrice qui prend la valeur 1 si $F(\mathbf{x}) > T$ et 0 dans le cas contraire. Cette intégrale peut être estimée par échantillonnage Monte-Carlo (MCS) [11, 9] comme une somme finie telle que

$$P_f \approx \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{F(\mathbf{X}^{(i)})>T}. \tag{2}$$

où N est la taille de l'échantillon et $\mathbf{X}^{(i)}$, $i = 1, \ldots, N$ sont indépendants et identiquement distribués. Un coefficient de variation noté COV_{P_f} est associé à cet estimateur [13], tel que :

$$\text{COV}_{P_f} = \sqrt{\frac{1 - P_f}{P_f * N}}. \tag{3}$$

Pour obtenir une estimation précise de la probabilité de défaillance, le nombre de simulations du modèle haute fidélité $F(\cdot)$ peut devenir très important, rendant cette estimation inabordable compte tenu du coût de calcul de $F(\cdot)$. Par exemple, pour estimer une probabilité de l'ordre de 10^{-2} avec $\text{COV}_{P_f} = 0.1$, 10^4 échantillons sont nécessaires. Dans ce contexte, d'autres méthodes telles que l'Importance Sampling (IS) [10, 14] ou le Subset Simulation (SS) [21, 2], ont été développées afin de diminuer le budget de calcul et de réduire la variance de l'estimateur (Eq. 2) tout en fournissant une estimation précise de P_f. Malgré ces techniques alternatives d'échantillonnage, le coût de calcul associé à l'analyse de la fiabilité pour des codes couteux reste associé à un budget de calcul conséquent que l'on souhaite réduire tout en conservant une bonne précision sur l'estimation de la défaillance. En effet dans la plupart des systèmes complexes, le budget de calcul ne permet l'évaluation que de quelques dizaines ou centaines de simulations.

3 Problème scientifique

Dans le but d'effectuer des analyses de fiabilité avec un budget de calcul limité, on souhaite remplacer le code haute fidélité par un méta-modèle, pour ensuite pouvoir utiliser notre estimateur (Eq. 2). Une approche classique dans la littérature consiste à construire un méta-modèle par processus Gaussien (GP) [18], puis à enrichir ce méta-modèle au cours de l'analyse de fiabilité, autour du point de l'espace des entrées susceptible de mener à la défaillance. Cette construction permet d'affiner le surrogate uniquement dans les zones d'intérêt vis à vis de l'estimation de la probabilité de défaillance. Un tel processus est appelé "active learning".

Suivant ce principe, plusieurs méthodes d'active learning pour l'analyse de fiabilité à l'aide de GP ont été développées dans la littérature [3, 7]. Dans le contexte de systèmes complexes, obtenir un méta-modèle précis nécessite un grand nombre d'appels au code haute fidélité, entraînant un coût de calcul conséquent. Il s'avère pertinent de faire appel à des solveurs de fidélités intermédiaires et inférieures, par conséquent moins précis, mais aussi moins coûteux. Un moyen de réduire les coûts de l'analyse de fiabilité et de garantir une estimation précise des probabilités consiste ainsi à utiliser des modèles de substitution multi-fidélité exploitant les informations de différents niveaux de fidélité. On investigue ici la construction du méta-modèle multi-fidélité, et son enrichissement par ajout de donnée, en un point qu'il convient d'identifier, issue d'une simulation réalisée avec un niveau de fidélité qu'il convient aussi de déterminer.

La méthode la plus largement utilisée pour construire des modèles multi-fidélité par GP est l'Auto-Regressive Model (AR1) [12], adapté lorsque les différentes fidélités ont une relation linéaire entre elles [4]. D'autres modèles multi-fidélité permettent d'exprimer d'autres relations entre sources d'information. Par exemple, le Linear Model of Coregionalization (LMC) [1] définit un GP multi-output dans lequel chaque sortie représente une fidélité (adaptée en cas de relation linéaire entre les niveaux de fidélité), et le Non-linear Auto-Régressive Multi-fidelity Gaussian Process (NARGP) [15] est une extension de l'AR1 (adapté en cas de relations non linéaires entre les niveaux de fidélité).

Récemment, différentes techniques d'analyse de fiabilité basées sur l'AR1 ont été proposées dans la littérature [16]. Ces méthodes utilisent des critères d'enrichissement permettant de sélectionner des nouveaux points de calcul à ajouter au plan d'expériences pour améliorer la prédiction du GP multi-fidélité dans les zones d'intérêt. Ce choix du nouveau point et de la source d'information pour l'évaluer est réalisé en une seule étape (par exemple, avec le critère mfEGRA [5, 6]) ou en deux étapes (avec le critère CLF [20] par exemple). L'efficacité de la stratégie d'active learning en combinant ces différentes méthodes multi-fidélité et critères d'enrichissement a été comparée sur divers cas tests de complexités variées dans le cadre de la thèse de Romain Espoeys

(ONERA, CERFACS, cite Chap Livre RA). En conclusion de ce benchmark, plusieurs recommandations sur le choix d'une combinaison modèle/critère ont été formulées.

4 Applications

Le stage propose d'appliquer ces techniques d'active learning en multi-fidélité à un cas test d'hydrodynamique fluviale, en considérant l'état de défaillance comme le dépassement d'une digue en plaine.

Dans les travaux décrits précédemment, une première illustration a été proposée pour l'équation de Manning appliquée à un canal rectangulaire à pente constante, en considérant comme source d'incertitude la pente (altitude du fond amont et aval), l'apport amont et le frottement. La défaillance se produit lorsque l'élévation de la surface libre excède la hauteur de la digue sur le tronçon. Les différentes méthodes évoquées précédemment ont été évaluées pour 2 niveaux de fidélité, la basse fidélité provenant de l'introduction d'une erreur modèle dans le calcul de la hauteur d'eau. Il apparaît que pour cette configuration de test, la méthode NARGP qui permet de prendre en compte les non linéarités entre les niveaux de fidélité donne de meilleurs résultats qu'AR1 ou LMC, et ce quelque soit le critère d'enrichissement choisi.

Ce travail doit être étendu à un cas hydrodynamique plus complexe que l'actuel, bien qu'idéalisé, en élaborant les aspects suivants :

— On utilisera avec le code de calcul Saint-Venant Telemac [1] qui résout les équations hydrodynamiques en 2D pour des écoulements stationnaires ou instationnaires, en prenant en compte une description fine de la géométrie de la rivière, sur un maillage triangulaire non structuré.

— On supposera que les incertitudes proviennent de la description des frottements dans le lit de la rivière et dans la plaine d'inondation ainsi que du débit d'apport. On suppose que les frottements sont décrits par zones uniformes par des scalaires. On supposera que le débit d'apport est constant et que l'on simule donc un régime permanent qui établit, au bout d'un temps correspondant au temps de transfert du réseau, une hauteur d'eau constante (mais pas uniforme) dans la rivière et les plaines d'inondation. On supposera alors que l'incertitude est décrite par des variables aléatoires scalaires dont on spécifiera les pdfs.

— L'extension de l'espace scalaire des entrée incertaines à des variables incertaines fonctionnelles est envisageable, notamment via la perturbation du champ de bathymétrie/topographie. La génération de géométrie perturbée peut se faire de manière paramétrique plus ou moins simples, permettant une réduction de l'espace incertain via des méthodes type décomposition en modes propres ou Karhuren Loeve. On précise que ces perturbations doivent préserver le maillage pour les aspects multi-fidélité évoqués ci dessous.

— On supposera que la quantité d'interêt est le champ 2D de hauteur d'eau décrit sur le maillage. La définition du critère de défaillance reste à établir. On peut

1. opentelemac.org

par exemple considérer le dépassement de la digue en un point, ou considérer le max de la hauteur le long de l'ouvrage. La description du réseau d'infrastructure (une digue ou plusieurs digues) conditionnera aussi la définition du critère de défaillance. La prise en compte d'une sortie fonctionnelle constitue une évolution notable par rapport au travail réalisé à ce jour. La construction du méta modèle pour le champ de hauteur d'eau pourra s'inspirer des travaux de recherche réalisés au CERFACS proposant la construction d'un mélange d'experts de type polynomes du chaos avec une étape de réduction de dimension et classification [8].
— On supposera que les niveaux de fidélité se distinguent par la résolution spatiale du maillage, en préservant une imbrication des grilles telle que tout point du maillage grossier existe dans le maillage fin. Il conviendra d'évaluer la relation entre les niveaux de fidélité définis en fonction du maillage et éventuellement introduire une définition alternative du niveau de fidélité.

Il est à noter qu'au delà de ce cas idéalisé, il est envisagé d'utiliser un cas test de référence réel sur la Garonne Marmandaise.

5 Programme de travail

Le programme de travail pour le stage s'articule en 4 étapes principales.

— Etude bibliographique sur les méthodes de méta modélisation, avec enrichissement et multifidélité (bibliographie déjà élaborée dans le cadre de la thèse de R. Espoeys). Extension de la bibliographie au domaine de l'hydrodynamique fluviale et dimensionnement des ouvrages.
— Prise en main des codes et méthodes sur les méthodes de méta modélisation (par GP), d'active learning ainsi que le code existant pour les GP multifidélité et active learning
— Prise en main du code d'hydrodynamque fluvial et du cas test idéalisé (mise en place en amont du stage). Constitution d'une base de données établie pour éprouver les méthodes classique de méta modélisation
— Choix des critères de défaillance et implémentation des GP multifidélité avec active learning pour le cas T2D.

6 Type de collaboration souhaitée

— Stage de recherche à réaliser à partir de **mars 2024** entre CERFACS et ONERA. A discuter avec les co-proposants.
— Utilisation des codes développés dans le cadre de la thèse de Romain Espoeys (PhD ONERA, CERFACS)
— Utilisation de cas test développés dans le cadre de la thèse de Quention Bonassies (PhD CERFACS, CNES, CLS)
— Utilisation du cas test Garonne Marmandaise mis en place par EDF (Convention d'usage EDF-CERFACS dans un cadre recherche) et en cours d'upgrade au CERFACS (collab. CNES).

7 Chercheurs intéressés et responsabilités souhaitées

- Encadrement par les équipes du CERFACS, ONERA.
- Encadrement par Romain Espoeys (PhD ONERA, CERFACS)
- Encadrement Didier Lucor (LISN)

8 Section ouverte aux commentaires

Recherche d'un profil de candidat math appli, programation python. Une sensibilité à la thématique mécanique des fluides et hydrodynamique est souhaitée. Le déroulement du stage aura lieu à Toulouse, au CERFACS.

Références

[1] Mauricio A Alvarez, Lorenzo Rosasco, Neil D Lawrence, et al. Kernels for vector-valued functions : A review. *Foundations and Trends® in Machine Learning*, 4(3) :195–266, 2012.

[2] Siu-Kui Au and James L Beck. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic engineering mechanics*, 16(4) :263–277, 2001.

[3] Barron J Bichon, Michael S Eldred, Laura Painton Swiler, Sandaran Mahadevan, and John M McFarland. Efficient global reliability analysis for nonlinear implicit performance functions. *AIAA journal*, 46(10) :2459–2468, 2008.

[4] Loïc Brevault, Mathieu Balesdent, and Ali Hebbal. Overview of Gaussian process based multi-fidelity techniques with variable relationship between fidelities, application to aerospace systems. *Aerospace Science and Technology*, 107 :106339, 2020.

[5] Anirban Chaudhuri, Alexandre N Marques, and Karen Willcox. mfegra : Multifidelity efficient global reliability analysis through active learning for failure boundary location. *Structural and Multidisciplinary Optimization*, 64(2) :797–811, 2021.

[6] Anirban Chaudhuri and Karen Willcox. Generalized multifidelity active learning for gaussian-process-based reliability analysis. In *Dynamic Data Driven Application Systems*, 2022.

[7] Benjamin Echard, Nicolas Gayton, and Maurice Lemaire. AK-MCS : an active learning reliability method combining Kriging and monte carlo simulation. *Structural Safety*, 33(2) :145–154, 2011.

[8] Siham El Garroussi, Sophie Ricci, Matthias De Lozzo, Nicole Goutal, and Didier Lucor. Tackling random fields non-linearities with unsupervised clustering of

polynomial chaos expansion in latent space : application to global sensitivity analysis of river flooding. *Stochastic Environmental Research and Risk Assessment*, 36(3) :693–718, 2022.

[9] Donald L Ermak and Helen Buckholz. Numerical integration of the langevin equation : Monte carlo simulation. *Journal of Computational Physics*, 35(2) :169–182, 1980.

[10] John Geweke. Bayesian inference in econometric models using monte carlo integration. *Econometrica : Journal of the Econometric Society*, pages 1317–1339, 1989.

[11] John Geweke. Monte carlo simulation and numerical integration. *Handbook of computational economics*, 1 :731–800, 1996.

[12] Marc C Kennedy and Anthony O'Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1) :1–13, 2000.

[13] Jérôme Morio and Mathieu Balesdent. *Estimation of rare event probabilities in complex aerospace and other systems : a practical approach*. Woodhead publishing, 2015.

[14] Radford M Neal. Annealed importance sampling. *Statistics and computing*, 11 :125–139, 2001.

[15] Paris Perdikaris, Maziar Raissi, Andreas Damianou, Neil D Lawrence, and George Em Karniadakis. Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling. *Proceedings of the Royal Society A : Mathematical, Physical and Engineering Sciences*, 473(2198) :20160751, 2017.

[16] Bharath Pidaparthi and Samy Missoum. A Multi-Fidelity Approach for Reliability Assessment Based on the Probability of Classification Inconsistency. *Journal of Computing and Information Science in Engineering*, 23(1) :011008, 02 2023.

[17] Rüdiger Rackwitz. Reliability analysis—a review and some perspectives. *Structural safety*, 23(4) :365–395, 2001.

[18] Carl Edward Rasmussen. Gaussian processes in machine learning. In *Summer school on machine learning*, pages 63–71. Springer, 2003.

[19] Shelemyahu Zacks. *Introduction to reliability analysis : probability models and statistical methods*. Springer Science & Business Media, 2012.

[20] Chi Zhang, Chaolin Song, and Abdollah Shafieezadeh. Adaptive reliability analysis for multi-fidelity models using a collective learning strategy. *Structural Safety*, 94 :102141, 2022.

[21] Konstantin Zuev. Subset simulation method for rare event estimation : an introduction. *arXiv preprint arXiv :1505.03506*, 2015.