

Sujet de thèse - Risque d'excursion sous hétérogénéité

Caractérisation topologique de classification et apprentissage actif pour l'estimation de probabilités d'événements rares en grande dimension

C. Prieur, C. Blanchet-Scalliet, Y. Richet

2026-01-01

Contexte de la thèse

Dans le cadre de la sûreté nucléaire, la sûreté-criticité des installations repose sur l'évaluation et la gestion des risques liés à des phénomènes physiques complexes, dont la neutronique. Le coefficient de multiplication neutronique (k -effectif) est une grandeur physique centrale pour caractériser la réaction en chaîne neutronique, et en particulier la situation à éviter d'une réaction en chaîne non-contrôlée si le k -effectif dépasse la valeur 1.

Cette grandeur physique scalaire "intégrale" synthétise donc le comportement global neutronique d'un système nucléaire, en fait hétérogène et complexe. En outre, la modélisation numérique neutronique d'un tel système rationalise usuellement sa géométrie ou sa composition en les considérant majoritairement homogènes, ce qui permet de réduire la dimension du problème et d'en simplifier ainsi l'analyse de sûreté.

Toutefois, cette démarche de simplification peut s'avérer inadéquate pour des systèmes présentant des hétérogénéités fortes, qui ne seront pas représentées par ce modèle artificiellement homogénéisé, même en utilisant des hypothèses supplémentaires conservatives du risque de criticité.

Quelques exemples de telles situations "non homogénéisables" incluent :

- Les études de "crédit burn-up" des combustibles nucléaires usés, où la distribution isotopique du crayon combustible devient très hétérogène après une longue période d'irradiation et peut ainsi ne pas être correctement représentée par un modèle homogénéisé (voir figure 1 ci-dessous).
- Dans le contexte de la mise en assemblage des crayons combustibles, l'absence impromptue de quelques crayons peut avoir une effet significatif sur la criticité globale de l'assemblage (voir figure 2 ci-dessous).
- Dans le cas d'accidents nucléaires, comme celui de Fukushima, la constitution du "corium", agrégat fondu de matériaux divers, engendre une distribution très hétérogène des matériaux rendant les modèles homogénéisés inadéquats pour évaluer le risque de criticité précisément, et permettre la traitement sûr de ces déchets (voir figure 3 ci-dessous).

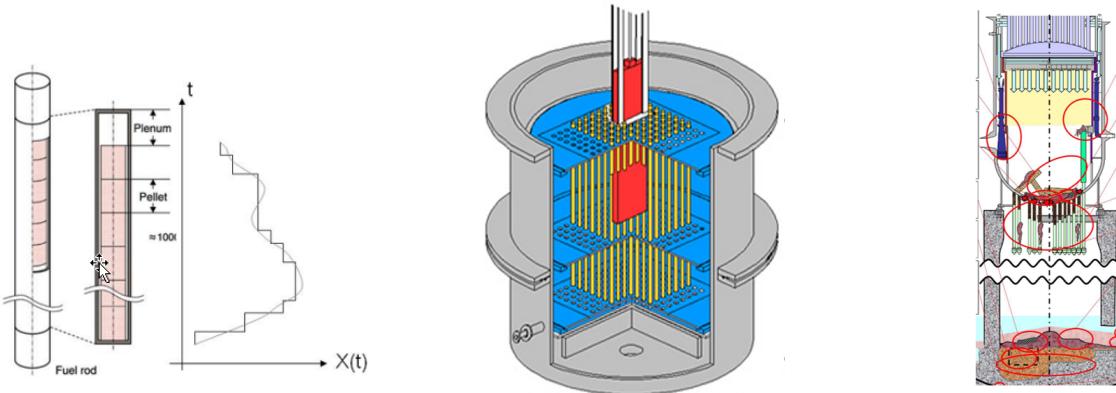


Figure 1 : Distribution d'isotopes dans un crayon combustible usé

Figure 2 : Assemblage de crayons combustibles avec crayons manquants

Figure 3 : Représentation schématique du corium de Fukushima

Finalement, cette thèse vise donc à élaborer une méthodologie de prise en compte de la complexité due aux hétérogénéités dans les modèles numériques, en s'appuyant sur des approches récentes de machine learning et de modélisation topologique. Cette méthodologie, bien qu'appliquée au contexte neutronique, pourra être transposée à d'autres domaines de la modélisation numérique où l'homogénéisation de l'objet d'étude s'avère trop réductrice.

Le défi majeur de cette thèse réside donc dans la construction d'un objet mathématique pertinent pour représenter l'hétérogénéité, à partir des données disponibles, sans information a priori forte (similaire à la problématique de "feature engineering" en machine learning). Une transformation de l'espace des entrées sera donc recherchée pour permettre de construire un méta-modèle qui sera ensuite exploité pour estimer la probabilité de dépassement d'un seuil critique (risque d'excursion) de la sortie scalaire du modèle.

Mots clés : classification (clustering), sélection des caractéristiques (feature selection), analyse topologique des données (topological data analysis), réduction dimension (dimension reduction), événements rares (rare events).

Sujet

Formalisation mathématique

Mathématiquement, le problème est formalisé comme suit. On considère un modèle qui prend en entrée un nuage de m points sur une grille régulière de taille $n \geq m$ de \mathbb{R}^d et qui donne un scalaire en sortie. Il est alors possible de coder le nuage de points par un vecteur de taille n , dont chaque composante correspond à un point de grille. Pour chaque composante, la valeur 0 est attribuée s'il n'y a pas de point du nuage sur le point de grille correspondant à la composante. Sinon, la valeur 1 est attribuée.

L'objectif est d'estimer la probabilité de dépassement d'un seuil critique, défini par la probabilité que la sortie dépasse un seuil.

Parmi les enjeux liés au cas d'application, il est important de souligner :

- Le coût de calcul élevé du modèle (chaque évaluation du modèle prend un temps non négligeable)
- La stochasticité du modèle qui correspond à un code Monte Carlo
- Le fait que la probabilité à calculer (risque d'excursion) est très faible, de l'ordre de 10^{-6}
- La taille du nuage de points en entrée varie (bien que restant inférieure à la taille de la grille)

Étapes de la thèse

La thèse comportera plusieurs étapes :

1. Revue bibliographique des approches TDA (topological data analysis) permettant d'extraire les caractéristiques topologiques d'un nuage de points
2. Développement d'une approche de classification supervisée pour repérer les caractéristiques (features) topologiques en entrée qui permettent de discriminer les valeurs "hautes" de la sortie scalaire
3. Revue bibliographique des approches d'apprentissage actif pour estimer des probabilités d'événements rares
4. Développement d'une approche d'apprentissage actif avec échantillonnage d'importance exploitant les caractéristiques (features) apprises pour estimer le risque d'excursion

Références

- CHAZAL, Frédéric et MICHEL, Bertrand. An introduction to topological data analysis: fundamental and practical aspects for data scientists. *Frontiers in artificial intelligence*, 2021, vol. 4, p. 667963.
- Zhonglai Wang, Wei Zhang, Fei Du, Shan Lu, Zhibo Geng. A novel active learning reliability analysis method based on ensemble of Kriging models and importance sampling for small failure probabilities. *Computer Methods in Applied Mechanics and Engineering*, 2025, 441.

Profil du candidat

Niveau d'études

Master 2 ou équivalent

Compétences techniques exigées

- Probabilités et statistiques
- Machine learning (modèles de régression, en particulier processus gaussiens)
- Optimisation
- Programmation en R et/ou Python
- Compétences en simulation numérique valorisables : openMC (MIT), calcul clusters, scripting...

Compétences comportementales

- Esprit d'analyse
- Curiosité intellectuelle
- Esprit de synthèse
- Rigueur
- Aisance relationnelle
- Aptitude à travailler dans un contexte pluri-disciplinaire

Compétences linguistiques

Français courant, bon niveau d'anglais requis.

Laboratoire et encadrement

La thèse sera encadrée par l’Institut Camille Jordan <https://math.univ-lyon1.fr/icj/> (Ecole Centrale de Lyon) et le laboratoire LJK de l’Université de Grenoble-Alpes. Le doctorant sera basé à l’Ecole Centrale de Lyon. Des séjours à l’ASNR et à Grenoble seront à prévoir.

Contacts

- Christophette Blanchet-Scalliet (Christophette.Blanchet@ec-lyon.fr)
- Clémentine Prieur (clementine.prieur@univ-grenoble-alpes.fr)
- Yann Richet (yann.richet@asnr.fr)