

PROPOSITION DE STAGE

TITRE : Méthodes d'apprentissage statistique, réduction de dimension et estimation de fonctions pour optimiser les simulations numériques d'écoulements diphasiques de Cigéo

DUREE : 5 à 6 mois (début : mars-avril 2025)

LOCALISATION : Andra, 1-7 rue Jean Monnet – 92290 Châtenay-Malabry & Laboratoire Probabilités Statistique et Modélisation (LPSM), Université Paris Cité, bât. Sophie Germain, 8 place Aurélie Nemours, 75013 Paris

DESCRIPTION DU SUJET :

L'Andra (Agence nationale pour la gestion des déchets radioactifs) a pour mission de trouver de solutions de gestion des déchets à long terme afin de protéger les générations actuelles et futures du risque que présentent ces déchets. Dans le cadre du projet Cigéo, ces déchets doivent être stockés dans une couche d'argile à grande profondeur (environ 500 mètres) pour retarder et atténuer au maximum l'arrivée de la radioactivité au contact des êtres humains, et avec des impacts radiologiques maîtrisés. Les structures métalliques utilisées pour le stockage se corrodent avec le temps devenant des sources de gaz à l'origine d'une montée en pression dans la roche. Dans ce contexte, la description de l'évolution de la pression d'eau et de gaz au sein du stockage est une composante essentielle pour la conception et l'évaluation de la sûreté à long terme des installations de stockage. La simulation numérique constitue un moyen de prédiction majeur pour contribuer à cette description au regard de l'échelle de temps du million d'années requise pour l'évaluation.

Malgré les progrès récents en puissance de calcul, la modélisation des écoulements diphasiques nécessite bien souvent des durées de calcul très longues. Ces méthodes reposent sur la résolution d'équations aux dérivées partielles (EDP) non-linéaires contrôlant l'évolution temporelle de la pression et de la saturation en eau (et en gaz) au sein du stockage. Il est alors nécessaire de réaliser un nombre très important d'opérations, coûteuses en temps, pour couvrir les durées représentatives des processus à l'œuvre. Des méthodes de calcul dites de haute performance ont pourtant été développées pour pouvoir faire appel à des architectures numériques parallèles mais les limites actuelles incitent à explorer de nouvelles voies.

Une approche alternative repose sur des méthodes d'apprentissage statistique pour estimer les quantités d'intérêt à partir d'un échantillon de résultats fournis par le code de calcul de référence. En raison du nombre important de variables d'entrée, une piste d'amélioration consiste à tenter de réduire astucieusement le nombre de variables dont dépendent les quantités à estimer pour améliorer la précision de leur estimation. L'objet de ce stage sera de développer des méthodes de réduction de dimension et d'estimation de fonctions dans un cadre non paramétrique multivarié et d'étudier leur potentiel pour réduire les temps de calcul obtenus par des simulations classiques de résolution d'EDP.

À partir des résultats obtenus sur un cas simplifié, la méthode pourra être étendue sur des cas plus représentatifs des problématiques de l'Andra et appliquée à l'estimation des pressions maximales d'eau ou de gaz atteintes dans le stockage.

Le stage se déroulera en présentiel à l'Andra et/ou au LPSM. Le stage sera co-encadré par l'Andra (Direction de la Recherche et du Développement/Service Évaluation et Analyse de Performance) et le LPSM. Des échanges réguliers entre les encadrants et l'étudiant(e) permettront d'aborder efficacement ce sujet. Ces travaux seront valorisés par la rédaction d'un rapport de synthèse qui pourrait être travaillé dans l'objectif d'une publication scientifique en fonction de l'avancée des travaux.

Une poursuite en thèse financée par l'Andra est envisageable en fonction du déroulement du stage.

MOYENS MIS À DISPOSITION :

PC, cluster de calcul, codes de calculs, et base de connaissances de l'Andra et du LPSM

RÉMUNERATION : Statut stagiaire Andra

COMPÉTENCES PROFESSIONNELLES ET QUALITÉS :

Formation et expérience : étudiants de M2 ou en 3ème année d'école d'ingénieurs.

Spécialité : formation en sciences des données (data science), statistique, apprentissage statistique, machine learning.

Qualités individuelles, savoirs et savoir-faire : Méthode, rigueur et esprit d'initiative.

PERSONNES À CONTACTER :

Céline LÉVY-LEDUC, professeur de statistique (LPSM)

Email : celine.levy-leduc@lpsm.paris

Tél : 01.57.27.93.17

Page web : <https://sites.google.com/view/pagewebdecelinelevyleduc/>

Jacques WENDLING, ingénieur ANDRA

Email : jacques.wending@andra.fr

Tél : 01.46.11.81.81

Marc LECONTE, ingénieur ANDRA

Email : marc.leconte@andra.fr

Tél : 01.46.11.84.71