

La simulation d'écoulements en milieu poreux est de fondamentale pour la gestion optimale des réservoirs de stockage souterrains naturels (énergie fossile, géothermie, extraction d'hydrogène) ou artificiels (stockage de CO₂). L'objectif de la thèse proposée est l'amélioration de la performance de tels simulateurs en terme de réduction du temps de calcul. Plus généralement, le développement de méthodes d'apprentissage pour accélérer les simulations numériques est un domaine de recherche en plein essor sur lequel le LJLL et IFPEN ont les compétences pour innover au plus haut niveau. L'accent sera mis sur la prise en compte des phénomènes d'ouverture et de fermeture de puits, qui sont bien souvent déterminants pour le temps de calcul. Mais les avancées méthodologiques et algorithmiques de cette thèse pourront également bénéficier à terme à d'autres types de simulation que les milieux poreux pour lesquels existent des évènements ponctuels déstabilisateurs d'une dynamique établie (phénomènes d'injection dans les moteurs, rotation rapide d'une éolienne sur son axe, etc.).

La simulation numérique des écoulements en milieu poreux a fait l'objet de nombreux travaux de recherches ces dernières décennies, avec des avancées notables portant sur les schémas numériques ainsi que les solveurs linéaires et non-linéaires. Ceci a notamment conduit à des gains importants en précision et en temps de calcul pour la simulation des phénomènes physiques de plus en plus complexes (écoulements réactifs, écoulements en milieux fracturés, modèles avancés de puits, etc.) sur des domaines de plus en plus grands.

Ces avancées en modélisation et en simulation permettent désormais d'exploiter les simulateurs à d'autres fins : la quantification d'incertitude, l'optimisation sous contraintes ou encore le contrôle optimal sont désormais incontournables pour les ingénieurs menant des études. De tels modes d'utilisation nécessitent un grand nombre de simulations numériques dans des conditions variées (paramètres du modèle physique, conditions initiales, conditions aux limites). Le temps de calcul de chaque simulation est encore aujourd'hui le paramètre le plus contraignant pour ce type d'utilisation.

Les ouvertures de puits forment des évènements qui viennent déstabiliser de manière extrêmement forte la dynamique du système et imposent en général des coupures brutales de pas de temps afin de capturer les phénomènes transitoires à petite échelle dans les abords du puits. Or, les scénarios d'injection de CO₂ impliquent un grand nombre d'ouvertures et de fermetures de puits afin d'éviter des phénomènes de bouchage des puits dus à la réactivité du CO₂ avec la roche. Sur certaines simulations, les ouvertures ou fermetures de puits divisent le pas de temps par 100 et peuvent ainsi être la cause de plus de 50% du temps de calcul. Les approches de simulation traditionnelles évoquées jusqu'ici n'exploitent aucunement les résultats obtenus lors d'autres simulations similaires. Ces données sont pourtant essentielles car elles permettent d'apprendre le comportement du simulateur lors de l'ouverture d'un puits et ainsi aider à la prédiction du simulateur lors d'une nouvelle ouverture de puits dans des conditions proches de celles observées dans des simulations similaires.

Cette thèse a donc pour but de mettre en place une méthodologie d'apprentissage pour exploiter les données liées aux ouvertures et fermetures de puits afin d'accélérer la résolution de ces évènements par les solveurs numériques traditionnels.

Positionnement de la proposition par rapport à l'état de l'art et apport original attendu

L'exploitation de données de simulation pour apprendre les variétés de dimension faible sur lesquelles évoluent les solutions d'équations différentielles partielles (EDP) paramétriques n'est pas une idée nouvelle : les méthodes de réduction de modèle pour la dynamique des fluides ont pris leur essor à la fin des années 80 [1]. Cependant, ces méthodes perdent en précision lorsque les solutions présentent de fortes non-linéarités ou lorsque des phénomènes multiphysiques complexes (tels que l'ouverture d'un puits en modèle hydrostatique) sont pris en compte.

L'approche envisagée dans cette thèse est d'étudier et de quantifier l'apport des méthodes de machine learning (ML) et en particulier d'apprentissage profond (Deep Learning, DL) au problème précédent. Le principe de ces méthodes consiste à construire un prédicteur d'un problème non linéaire à partir d'une base d'apprentissage. La performance des méthodes de Deep Learning n'est plus à prouver pour des tâches spécifiques de traitement d'images (classification, segmentation) ou de traitement du langage (traduction automatique de textes, compréhension de la parole). Plus récemment, de nombreux travaux de recherche tentent d'étendre leur domaine d'application au cadre de la simulation numérique, avec des résultats prometteurs lorsque l'objectif est de prédire des quantités d'intérêt macroscopiques (fatigue cumulée, production du puits [7], etc.).

Cependant, de nombreuses difficultés se présentent lorsqu'il s'agit d'utiliser les méthodes d'apprentissage pour aider à prédire les solutions d'un système d'EDP d'un pas de temps à un autre. L'approche la plus simple consiste à utiliser le solveur numérique pour résoudre de manière précise le système d'EDP, et à stocker dans une base de données les solutions ainsi obtenues. Ensuite, on fait apprendre à un modèle statistique le lien entre les solutions à un instant donné t et leur évolution à $t + \Delta t$:

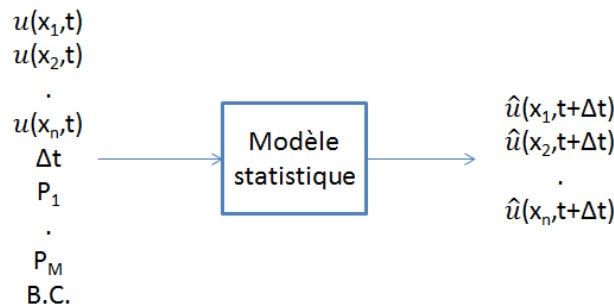


Figure 1 : Formulation simple du problème d'apprentissage. La solution calculée par le solveur numérique sur une grille composée de n éléments est u et la solution prédite est \hat{u} . Les paramètres P_1, \dots, P_M représentent les paramètres du système d'équation différentielle, et B.C. représente la variabilité dans les conditions aux limites du problème. Le modèle est obtenu en minimisant la fonction cout $\|u - \hat{u}\|^2$.

Cette approche simpliste souffre de nombreux défauts : d'une part, le modèle ne contient aucune information sous-jacente à la physique du problème. Les prédictions ne sont donc pas garanties de respecter les principes physiques intégrés dans les EDP (conservation de la masse, conservation de l'énergie, etc.). D'autre part, l'espace d'entrée du modèle est de grande dimension. Or sans ajout d'informations liées à la physique du domaine, la quantité de données nécessaires à l'établissement d'un modèle représentatif est très importante et la construction d'une base d'apprentissage représentative de la variabilité de ces entrées devient trop difficile.

Pour résoudre ces problèmes, plusieurs approches différentes ont été récemment proposées :

- **Approche intégrée** : Dans cette approche, des modules statistiques représentant une partie spécifique du solveur numérique sont entraînés dans une phase séparée puis intégrés au solveur dans le but d'accélérer celui-ci. On peut citer par exemple les travaux de J. Thompson et al. [1] qui, à l'intérieur d'un solveur des équations d'Euler, remplacent le module de résolution pour la pression (solveur de Poisson) par un réseau de neurones convolutif, ou encore les travaux de D. Ray [2] qui remplacent les traditionnels limiteurs dans les solveurs d'équations de conservation par un simple réseau de neurones (perceptron multi-couches).
- **Approche directe** : L'idée est ici de directement remplacer les schémas numériques permettant la résolution d'EDP par des réseaux de neurones profonds. Les pistes de recherche se concentrent sur l'intégration de connaissances a priori du problème au modèle statistique : intégration des équations de conservation de la masse ou du résidu dans la fonction cout à minimiser [3] ou intégration de lois physiques spécifiques dans un modèle génératif simulant une EDP [4].
- **Approche intrusive** : L'objectif de ces approches n'est pas de remplacer le schéma numérique ni de remplacer des étapes de celui-ci, mais plutôt de l'améliorer. On peut citer les travaux de S. Mishra [5] dans lesquels les coefficients du schéma numérique s'adaptent à chaque instant à la solution actuelle et au pas de temps suivant, ou encore les travaux de J. Albright [6] concernant l'apprentissage des coefficients de viscosité artificielle pour les schémas numériques de résolution d'écoulements compressibles avec chocs.

L'ensemble de ces approches se fonde sur les connaissances de la physique sous-jacente au problème ou encore sur la connaissance du schéma numérique de résolution : ce champ de recherche est ainsi souvent qualifié de « Physics Informed Deep Learning ».

A ce jour, aucune étude sur les écoulements de type Darcy n'a encore été menée alors que des résultats prometteurs existent déjà sur des écoulements de type Navier-Stokes. L'apport original de cette thèse est donc d'exploiter ces diverses méthodes pour développer un algorithme performant spécifique à la problématique de l'ouverture de puits en milieu poreux.

Les difficultés rencontrées par ces méthodes sont leur capacité de généralisation : comme exposé ci-dessus, ces méthodes sont développées spécifiquement pour une classe de problèmes dédiés et n'offrent une amélioration des performances que dans des configurations restreintes. Les travaux envisagés pour cette thèse n'y font pas exception, et les aspects de généralisation à des configurations plus complexes et à d'autres classes de problèmes feront l'objet d'une attention particulière.

[1] J. Tompson, K. Schlachter, P. Sprechmann and K. Perlin, *Accelerating Eulerian fluid simulation with convolutional networks*, Proceedings of the 34th International Conference on Machine Learning, PMLR 70:3424-3433, 2017.

- [2] D. Ray and J. S. Hesthaven, *An artificial neural network as a troubled-cell indicator*, Journal of Computational Physics, Vol. 367 (15), pp 166-191, 2018.
- [3] M. Raissi and G. E. Karniadakis, *Hidden physics models: machine learning of nonlinear partial differential Equation*, J. Comput. Phys., 357, 2018, 125-141.
- [4] Y. Yang and P. Perdikaris, *Physics-informed deep generative models*, NeurIPS 2018 workshop on Bayesian Deep Learning, 2018.
- [5] Siddhartha Mishra, *A machine learning framework for data driven acceleration of computations of differential equations*, Mathematics in Engineering, 1, 1, 118, 146, 2018.
- [6] Albright, Jason & Shashkov, Mikhail & Urban, Nathan. (2019). *Data-driven optimization strategies for staggered-grid Lagrangian Methods*. 10.13140/RG.2.2.10281.13929.
- [7] Amini, Shohreh & Mohaghegh, Shahab. (2019). Application of Machine Learning and Artificial Intelligence in Proxy Modeling for Fluid Flow in Porous Media. Fluids. 4. 126. 10.3390/fluids4030126

Stratégie de recherche envisagée

L'approche générale consiste à exploiter les différentes méthodes exposées ci-dessus pour aider à la simulation des phénomènes transitoires rapides dus à l'ouverture de puits et ainsi réduire le temps de calcul de la simulation.

Plus précisément, nous considérerons dans un premier temps un modèle d'écoulement de Darcy diphasique (eau/gaz) incompressible caractérisé par les équations suivantes :

$$\varphi \frac{\partial S_\alpha}{\partial t} + \text{div}(-K\nabla P) = Q_\alpha$$

Les inconnues sont les saturations de chaque phase S_w et S_g (avec $S_w + S_g = 1$) et la pression P , tandis que les paramètres de porosité φ et de perméabilité K sont supposés constants. La difficulté réside dans le terme source Q_α qui représente la quantité injectée lors de l'ouverture de puits.

Le modèle de puits le plus simple consiste à considérer une injection constante localisée : le terme source Q_α prend une valeur non-nulle en un point du domaine pendant un intervalle de temps donné. Dans ce modèle, on obtient alors une équation de Poisson sur la pression (en sommant les deux EDP pour l'eau et le gaz) :

$$\text{div}(-K\nabla P) = Q_w + Q_g$$

Pour résoudre cette équation, nous proposons de développer les approches intégrées et en particulier de s'appuyer sur les travaux de J. Thompson et al. [1] cités ci-dessus. L'objectif sera alors d'entraîner un modèle d'apprentissage par réseaux de convolution pour prédire la pression résultant de la présence de termes sources d'injection variables à la fois en débit mais aussi en position. Le problème peut croître graduellement en difficulté en considérant également les variations spatiales du tenseur de perméabilité, puis en intégrant la variation des conditions aux limites dans le modèle d'apprentissage. Une fois le terme de pression obtenue, le calcul de la saturation est immédiat.

Par la suite, il s'agira de travailler sur des modèles de puits complexifiés proches des modèles réels utilisés dans les simulateurs aujourd'hui.

Il est important de préciser que dans toutes les approches ci-dessus, il est possible de se servir des modèles statistiques soit pour simuler de manière approximative les phénomènes d'ouverture de puits, soit pour servir d'initialisation au solveur numérique traditionnel dans le but d'accélérer la résolution des systèmes non-linéaires associés, retrouvant ainsi la même solution mais à un coût moindre. Les données d'apprentissage sont les solutions supposées exactes du solveur numérique. Cependant, ces solutions sont obtenues avec un schéma numérique et un ensemble de pas de temps donnés. Ces pas de temps sont en général très faibles lors d'évènements d'ouverture de puits. Une difficulté survient alors lorsqu'il s'agit de prédire la solution issue d'un grand pas de temps lors d'une ouverture de puits : la solution du schéma numérique avec grand pas de temps peut être très éloignée de la solution du schéma avec une série de faibles pas de temps sur laquelle le modèle est entraîné. Ainsi, il est possible que même avec un modèle « parfait » permettant de prédire la solution exacte, cette solution représente une mauvaise approximation de la solution du solveur numérique pour un grand pas de temps, et ne permette ainsi pas d'accélérer la simulation. L'approche envisagée dans ce cas précis consiste alors à changer le schéma numérique d'intégration en temps de manière dynamique pour garder une précision importante (sur une approche similaire aux travaux de S. Mishra [5]), ce qui permet alors de se servir du modèle statistique pour initialiser la solution.

L'étudiant(e) sera à 80% au LJLL (Sorbonne Université) et 20% à IFPEN et bénéficiera des mêmes opportunités que les étudiants en thèse au LJLL (participation aux congrès, écoles, ...).