



Sujet de thèse DynFluid/Safran Tech

Modélisation par *Machine Learning* d'écoulements turbulents dans les turbomachines

Contacts :

Xavier MERLE (xavier.merle@ensam.eu), Paola CINNELLA (paola.cinnella@sorbonne-universite.fr),
Grégory DERGHAM (gregory.dergham@safrangroup.com)

Type de financement : dispositif CIFRE

Entreprise : Safran Tech

Laboratoire de rattachement : DynFluid, ENSAM, Paris (www.dynfluid.eu)

Durée : septembre 2021 – août 2024

Mots clés : écoulements turbulents, modèles RANS, *Machine Learning*, régression, classification, environnement bayésien, *Culstered Bayesian Averaging*, turbomachines.

1 Contexte

Portée par un accroissement considérable de la puissance de calcul, la Simulation Numérique en Mécanique des Fluides a énormément progressé au cours des décennies passées. Ainsi, les méthodes "haute-fidélité" pour le calcul d'écoulements turbulents (DNS et LES) ont permis quelques avancées certaines. Cependant, aujourd'hui encore, le coût de ces méthodes les limite à des écoulements caractérisés par des nombres de Reynolds relativement modérés, très loin des impératifs d'un milieu industriel qui est pourtant contraint à toujours plus de précision. L'utilisation des méthodes RANS (basées sur les équations de Navier-Stokes moyennées à la Reynolds) reste la norme malgré de nombreuses lacunes. Basées, à leurs débuts, sur une modélisation purement mécaniste de la turbulence, ces approches peuvent aujourd'hui tirer profit des importants progrès enregistrés par l'apprentissage automatique au cours des vingt dernières années pour combler ces lacunes et améliorer leurs prédictions.

2 Mélanges Bayésiens de modèles et apprentissage automatique (*Machine Learning*)

D'un point de vue scientifique, améliorer la prédiction des modèles RANS nécessite de mieux comprendre la physique des écoulements turbulents pour ajuster la structure mathématique des modèles. C'est un processus long et complexe qui fait l'objet du travail d'une communauté importante de dynamiciens des fluides. Ainsi, de nombreux modèles, ou variantes de modèles, ont été développés au cours des dernières décennies, sans pour autant faire consensus. En effet, la complexité des écoulements turbulents et le choix de modéliser toutes les échelles en ne résolvant que le champ moyen rend très difficile, voire impossible, la définition d'un modèle universel, c'est-à-dire valable pour toute configuration d'écoulement. Par ailleurs, dans la pratique industrielle, les modèles RANS sont bien souvent appliqués au delà de leurs limites théoriques, par exemple, pour le calcul de configurations massivement décollées et instationnaires.

Ainsi, à défaut de pouvoir développer des modèles « universels », une alternative prometteuse a vu le jour ces dernières années, qui consiste à prévoir un écoulement donné à partir d'un mélange pondéré de plusieurs modèles de turbulence alternatifs. Les poids du mélange sont calibrés à partir d'un ensemble de données haute-fidélité (LES, DNS) ou expérimentales. Ce mélange, de surcroît, peut s'opérer dans un cadre stochastique (*Bayesian Model Averaging*[3]), et pas seulement déterministe, dotant alors les ingénieurs de bureaux d'études CFD « d'indices de confiance » quant à la qualité de leurs simulations. Les encadrants de la thèse et Safran Tech collaborent depuis plusieurs années au développement et à l'application de ces méthodes à la prédiction d'écoulements d'intérêt industriel (aérodynamique externe, compresseurs...). En l'état actuel de la recherche, cette approche présente malgré tout deux limitations :

(i) le mélange Bayésien attribue aux modèles les mêmes poids en tout point de l'espace et (ii) le calcul des poids du mélange est effectué à partir de grandeurs aérodynamiques (ou thermodynamiques) indexées spatialement. Le premier point limite nécessairement l'ampleur de l'amélioration que peut apporter le mélange puisqu'il ne tient pas compte des disparités locales de l'écoulement (un modèle peut être meilleur qu'un autre dans une certaine zone de l'écoulement mais moins bon dans une autre). Le second point circonscrit les prédictions à de nouveaux écoulements dont la topologie doit rester proche de ceux utilisés pour la calibration des poids du mélange.

Pour autant, de récents travaux conjoints entre l'équipe encadrante et Safran Tech montrent que ces verrous peuvent être levés par l'utilisation du *Machine Learning*. En effet, les différents outils de régression qu'offre l'apprentissage automatique doivent permettre de facilement calculer les poids du mélange en se basant sur des grandeurs locales de l'écoulement (des *features* ; par exemple les invariants du tenseur de déformation moyen local) ce qui permet à la fois de régionaliser les pondérations (utilisation d'un algorithme de type *Clustered Bayesian Averaging*[5], par exemple) tout en étendant les prédictions à des configurations d'écoulements davantage éloignés de ceux par lesquels l'apprentissage a été mené. Mieux encore, les outils de classification (*clustering*) peuvent permettre de limiter les ressources de calcul (nombre d'appels au code de calcul RANS, qui reste une contrainte importante pour l'utilisation du *Machine Learning* en simulation numérique d'écoulements fluides), en sélectionnant de manière pertinente les données à prendre en compte pour le calcul des pondérations. Les travaux mentionnés sont encore préliminaires, et il s'agira, dans un premier temps de choisir les outils les plus adaptés à la CFD parmi ceux classiquement utilisés en *Machine Learning*[4] (arbres de décision, *Support Vector Machine*, réseaux de neurones...) tout en tenant compte des contraintes déjà mentionnées (en particulier la mesure d'incertitude sur la prédiction de nouveaux écoulements et la « frugalité » de la méthode). Le choix des *features* (en genre et en nombre), par exemple par un outil de classification, tout comme celui des données d'entrée (analyse de sensibilité) devra également faire l'objet d'une analyse approfondie. Enfin, la quantification de l'erreur de modèle[1] (lié à sa structure mathématique) et l'utilisation d'algorithmes promouvant la parcimonie[2] (via des méthodes de régularisation par exemple) pour limiter le nombre de modèles dans le mélange devront être intégrés dans les premières étapes de la réflexion.

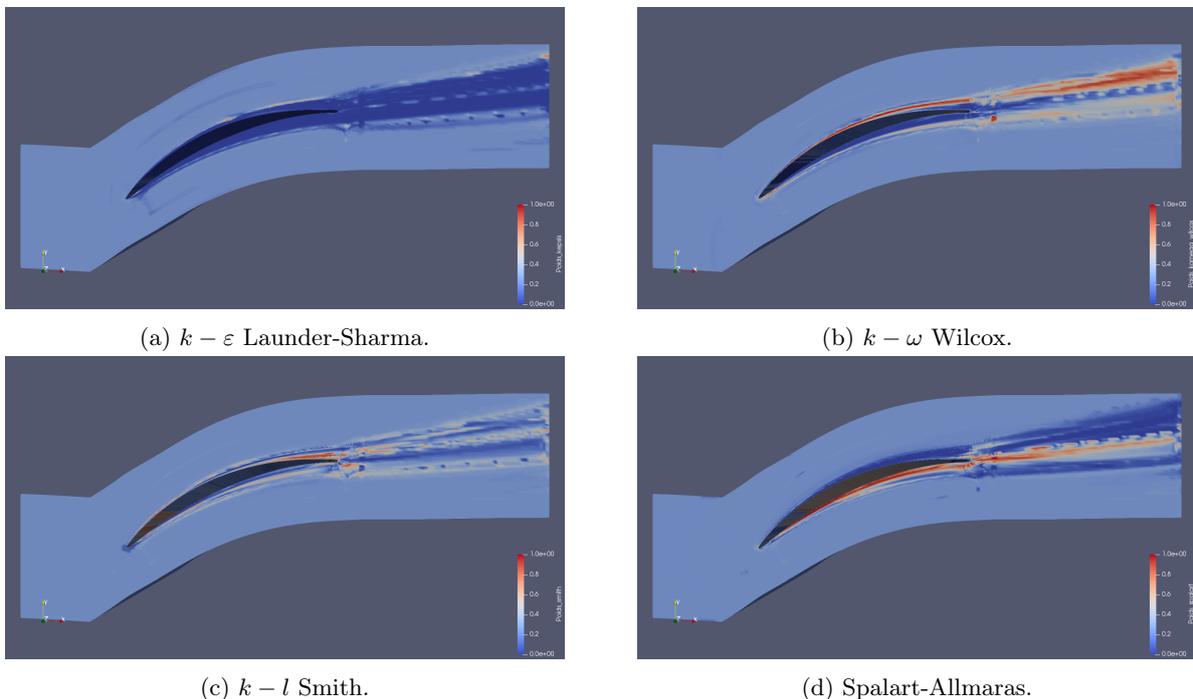


FIGURE 1 – Cartographie de poids d'un mélange de 4 modèles RANS sur une aube de compresseur.

3 Objectifs de la thèse

La thèse ambitionne de fournir une méthode permettant à Safran de prédire des écoulements turbulents avec une précision accrue, tout en ayant une estimation aussi fiable que possible des intervalles

de confiance. Le doctorant sera amené à intervenir dans les codes elsA de l'ONERA (code de calcul de référence chez Safran), ainsi que dans des codes de recherche pour lesquels une bonne maîtrise de l'environnement Linux et des langages python, FORTRAN et C/C++ est requise.

Le déroulement de la thèse se fera en deux temps :

1. la première partie sera consacrée au développement et à la validation de la méthode incluant :
 - sélection et validation des différents algorithmes de *Machine Learning* intervenant dans les processus de régression et de classification,
 - sélection des *features*,
 - sélection et validation des algorithmes/méthodes pour le traitement des points annexes (parcimonie, incertitude de modèle),
 - sélection des données d'entrée,
 - validation de l'approche pour des cas d'écoulement 2D turbulents.
2. la seconde partie de la thèse sera consacrée à l'extension de la méthode en vue de son application à des cas d'écoulement 3D turbulents (compresseur par exemple) incluant :
 - développement d'une stratégie réaliste liée à la limitation des ressources informatiques (décomposition du domaine 3D par exemple),
 - production de données de référence (via des bases de données existantes ou la production éventuelle de données synthétiques).

Cette thèse se situe au croisement du *Machine Learning* et de la Mécanique des Fluides Numérique. Elle représente donc un défi scientifique fort puisque les méthodes d'apprentissage automatique classiques ne peuvent pas être directement transposées, sans adaptation de la stratégie, à l'univers de la simulation numérique d'écoulements. Le doctorant partagera son temps entre Sorbonne Université, l'ENSAM Paris et le site de Safran Tech (Paris Saclay). Il pourra interagir avec des experts Safran en *Machine Learning*, mais aussi en CFD pour les turbomachines. Les travaux pourront être valorisés auprès des diverses sociétés du groupe Safran, qui seront particulièrement attentives aux apports de l'apprentissage automatique sur la CFD.

Références

- [1] Xiao H, Wu J, Wang J, Roy CJ. Quantifying and reducing model-form uncertainties in Reynolds averaged Navier-Stokes equations : A data-driven, physics-informed Bayesian approach. *J Comp Phys* 324 :115-136 (2016)
- [2] Schmelzer, Martin, Richard P. Dwight, Paola Cinnella. Discovery of Algebraic Reynolds-Stress Models Using Sparse Symbolic Regression. *Flow, Turbulence and Combustion* 104, n° 2-3 (march 2020) : 579-603. <https://doi.org/10.1007/s10494-019-00089-x>.
- [3] Zordo-Banliat, M. de, X. Merle, G. Dergham, P. Cinnella. Bayesian Model-Scenario Averaged Predictions of Compressor Cascade Flows under Uncertain Turbulence Models. *Computers & Fluids* 201 (april 2020) : 104473. <https://doi.org/10.1016/j.compfluid.2020.104473>.
- [4] Azencott, Chloé-Agathe. Introduction au Machine Learning. Dunod. InfoSup, 2019. <https://www.dunod.com/sciences-techniques/introduction-au-machine-learning-0>.
- [5] Yu, Qingzhao, Steven N. MacEachern, et Mario Peruggia. Clustered bayesian model averaging. *Bayesian Analysis* 8, n° 4 (2013) : 883-908. <https://doi.org/10.1214/13-BA859>.