



Modélisation de la turbulence par combinaison d'assimilation de données et de machine learning à partir de données limitées

Turbulence modeling from sparse and noisy observations using data assimilation and machine learning

Soutenance de thèse – Raphaël VILLIERS

Mercredi 25 juin à 14H00
Salle AY.02.63 – ONERA Meudon

lien JITSi : https://rdv.onera.fr/soutenance_raphael_villiers
(Configuration recommandée : navigateur Chrome)

Devant le jury composé de :

Examineurs :

- MARIOTTI Alessandro, Professeur, University of Pisa
- MATHELIN Lionel, Chargé de recherche, CNRS et Université Paris-Saclay
- SAYADI Taraneh, Professeure, CNAM

Rapporteurs :

- GOMEZ Thomas, Professeur, Icam
- IOLLO Angelo, Professeur, Université de Bordeaux et INRIA

Directeur de thèse :

- LAMBALLAIS Éric, Professeur, Université de Poitiers

Co-Directeur de thèse :

- MARCELLO Meldi, Professeur, ENSAM et Université de Lille

Encadrants :

- MONS Vincent, Chargé de recherche, ONERA
- SIPP Denis, Directeur de recherche, ONERA

Invité :

- CORDIER Laurent, Directeur de recherche, CNRS et Université de Poitiers

Résumé / Abstract :

La modélisation de la turbulence par les approches RANS (Reynolds-Averaged Navier-Stokes) et URANS (Unsteady RANS) en particulier permet l'estimation numérique d'écoulements à des coûts modérés, mais souffre de défauts importants pour la prédiction de certains phénomènes complexes comme le décollement et le décrochage. Ces dernières années, l'apprentissage automatique est de plus en plus employé pour l'identification de corrections de modèles de turbulence. Bien qu'une telle approche soit prometteuse, la mise en œuvre de l'apprentissage automatique peut être limitée par la quantité et le caractère épars des données disponibles, en particulier dans le cas d'écoulements à haut Reynolds comme considérés en aérodynamique.

Dans ce contexte, cette thèse aborde le défi de l'apprentissage de corrections de modèles de turbulence à partir d'observations limitées et possiblement bruitées. Afin d'accomplir une telle tâche, une méthodologie mêlant assimilation de données, apprentissage automatique et l'algorithme EM (Expectation-Maximization) dans un formalisme Bayésien est décrite et évaluée. Dans cette approche, il est justifié d'alterner entre des phases d'assimilation de données pour inférer des écoulements complets à partir des données limitées disponibles, ici grâce à un filtre de Kalman d'ensemble, et des phases d'apprentissage employant ces états complets pour mettre à jour la correction de modèle considérée. Deux variantes de cette méthodologie sont développées, prenant en compte ou non des erreurs de modélisation résiduelles.

La méthodologie est d'abord appliquée à l'équation unidimensionnelle de Ginzburg-Landau. La variante de la méthodologie avec prise en compte d'erreurs de modélisation résiduelles montre une robustesse accrue lorsque les observations sont moins fréquentes en temps, tandis que les deux stratégies montrent une relative insensibilité à la densité des observations en espace. Ces résultats valident l'efficacité du cycle assimilation de données-apprentissage automatique proposé et soulignent l'importance du traitement de l'erreur du modèle lors de la reconstruction de l'état.

L'approche développée est ensuite appliquée à l'écoulement turbulent autour d'un cylindre circulaire à $Re=3900$ simulé numériquement par l'approche URANS. Des mesures de vitesse ponctuelles sont utilisées pour entraîner des termes correctifs dans les équations de conservation de la quantité de mouvement et du modèle de turbulence, permettant d'améliorer significativement la qualité des résultats prédits, y compris pour d'autres valeurs du nombre de Reynolds que celle considérée pour l'apprentissage. La robustesse des corrections de modèles obtenues par rapport à la densité de mesures disponible est également étudiée.

Dans l'ensemble, cette thèse développe une méthodologie combinant de manière optimale assimilation de données et apprentissage d'automatique pour l'identification de corrections de modèles à partir d'observations limitées, ouvrant de nouvelles voies pour l'intégration des observations expérimentales dans les stratégies de modélisation de turbulence. La robustesse démontrée face à la rareté des observations, combinée au caractère non-intrusif de l'approche développée, marque une avancée significative vers le déploiement de modèles dynamiques améliorés par apprentissage automatique dans les applications CFD (Computational Fluid Dynamics) pratiques.

Turbulence modeling is an essential tool to make the numerical prediction of industrial flows affordable. However, it may suffer from deficiencies in capturing complex flow phenomena, such as separation and stall. In recent years, the use of ML (Machine Learning) to identify corrections to turbulence models has been increasingly investigated. Despite of significant progress in this research field, the sparse character of data as available for aerodynamic applications may still hinder the application of ML techniques.

In this context, this PhD thesis investigates the possibilities in learning corrections to turbulence models from limited and possibly noisy data. To this end, a methodology that merges DA (Data Assimilation), ML and the EM (Expectation-Maximization) algorithm in a Bayesian formulation is exposed and assessed here. In this framework, it may be justified to alternate between DA stages, where full flow descriptions are inferred from the limited observations, which is here achieved through ensemble Kalman filter techniques, and ML stages, where these full states are employed to update the considered model correction. Two versions of this methodology are elaborated that account or not for residual modeling errors.

The described methodology is first applied to an one-dimensional Ginzburg-Landau model. Numerical experiments indicate that taking into account residual modeling errors may favor the quality of the inferred model and its robustness with respect to the time period of the observations, while both versions of the methodology exhibit low sensitivity with respect to the spatial density of the observations. These results confirm the efficacy of the developed methodologies and highlight the benefit of accounting for modeling uncertainties in the DA phases.

The present approach is then applied to the URANS (Unsteady Reynolds-Averaged Navier-Stokes) modeling of the turbulent flow past a circular cylinder at $Re=3900$. Pointwise observations of the velocity field are considered to learn corrective terms in the momentum equations and in the transport equation that is associated to the employed turbulence model. The application of the present methodology successfully improves the quality of the URANS results compared to the baseline model. This enhancement is shown to persist at other values of the Reynolds number for which data have been extracted. The robustness of these findings with respect to the spatial density of the data is also investigated.

Overall, a promising strategy that optimally combines the strengths of DA and ML techniques to infer model corrections from limited observations is presented and successfully applied to unsteady turbulence modeling in particular. The robustness of the discussed methodology with respect to the observation sparsity along with its non-intrusive character should favor its application to more complex and industrial flow configurations.

Mots clés / Key words :

Modélisation de la turbulence, apprentissage automatique, assimilation de données, algorithme d'Expectation-Maximization, filtre de Kalman d'ensemble

Turbulence modeling, machine learning, data assimilation, Expectation-Maximization algorithm, ensemble Kalman filter