Learning turbulent model from sparse data through data assimilation and machine learning.

R. Villiers[*], V. Mons[*†], D. Sipp[*†], E. Lamballais[‡] and M. Meldi[§]

The use of Machine Learning (ML) for learning turbulent models is an active research topic and presents promising opportunities for improving CFD models[1]. The training of such data-driven models often relies on well-resolved data. In the case of Reynolds-Averaged Navier-Stokes (RANS) models, these data typically correspond to full mean flows from scale-resolving simulations[1]. This may form a bottleneck to gather large datasets and train ML models for complex flow configurations especially for unsteady turbulence models (e.g. URANS), as the training would require unsteady data and at various conditions (in terms of Reynolds number, geometry, ...). In the present work, we investigate the possibility in learning unsteady model corrections based on limited data, i.e. data that are sparse in space and/or in time, thus relaxing the need of performing expensive high-fidelity simulations and possibly paving the way for the use of experimental data to train ML (turbulence) models. Based on the methodological contribution[2], we here rely on Data Assimilation (DA) techniques like Ensemble Klaman Filter (EnKF) to bridge the limited data with the ML phase. In this paper the DA method is used to correct a corrupted Ginzburg-Landau (GL) model prediction (said as a baseline model) from sparse observations, providing a full state to train a ML model in order to improve the baseline model. By evaluating the Expected-Maximization (EM) algorithm on a GL model, we can assess its potential and limitations in modeling turbulent unsteady flows from sparse and noisy observations. In Fig 1 the EM algorithm has shown good results in dealing with sparse observations, offering an opportunity to implement this approach for correcting Large Eddy Simulation (LES) and URANS models using Direct Numerical Simulation (DNS). Its ability to handle complex and nonlinear models makes it a promising candidate for improving the accuracy of these models in real-world applications.

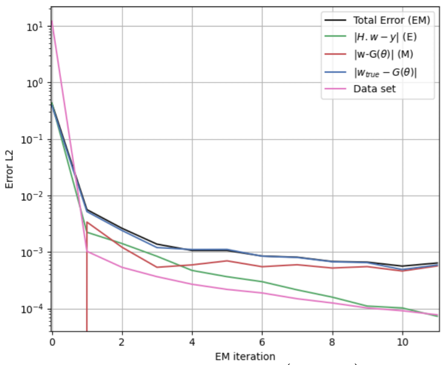

Figure 1: Expectation (green), Maximization (red), EM (black), model prediction (blue) and target model (pink) errors monitoring over EM iteration.

---

[*]DAAA, ONERA, F-92190 Meudon, France
[†]Université Paris Saclay, Saclay, France
[‡]Université de Poitiers, Poitiers, France
[§]ENSAM, Lille, France

[1]Volpiani et al., *Physical Review Fluid* **6**, 064607 (2021).
[2]Bocquet et al., *Fond. of Data Science* **1050**, 27 (2020).