centralelille

Laboratoire de Mécanique des Fluides de Lille
Kampé de Fériet

MASTER 2 (M2) RESEARCH INTERNSHIP: Investigation of Machine Learning strategies for *a priori* Subgrid-Scale Modelling in Large Eddy Simulation

Large Eddy Simulations (LES), which directly resolve large, energy-containing eddies while accounting for unresolved, dissipative small scales through subgrid-scale (SGS) closures, have generated considerable enthusiasm and attracted extensive study in the last 50 years. Nevertheless, they still have significant drawbacks in robustness and computational cost that prevent widespread use in realistic industrial configurations. These include limited knowledge of energy-transfer dynamics between large and small scales, assumptions of isotropy and homogeneity for unresolved scales, and lack of awareness of the spatial filter used to obtain the resolved flow variables (commonly referred to as filtered quantities $\overline{(\cdot)}$), among others. In fact, for the latter, spatially filtering the Navier-Stokes equations for the velocity u yields a term, the so-called SGS stress tensor $\tau_{ij} = \overline{u_i u_j} - \overline{u_i}\,\overline{u_j}$, which accounts for the effect of the smallest, non-resolved scales on the resolved flow field, and which needs to be modeled.

It is generally accepted that assessing LES models should consider both *a priori* and *a posteriori* performance. The first involves evaluating the hypotheses underlying SGS closures by comparing the predicted SGS stress tensor τ_{ij} with a reference tensor τ_{ij}^{ref}, generally estimated from a Direct Numerical Simulation (DNS), but also eventually from experimental data. Since DNS resolve all flow scales, we can estimate the exact SGS tensor from the filtered DNS data. The second approach involves analysing the time evolution of filtered flow quantities when injecting the subgrid closure into a dynamical model. Within this framework, much knowledge already exists about the validity of classical subgrid closures. For instance, eddy-viscosity models, such as the classical Smagorinsky model [1], are known to show important limitations concerning the sole transfer of energy from the largest to the smallest scales (forward cascade), exhibiting important limitations in *a priori* assessments; however, the excessive dissipation they generate ensures the model's robustness. On the other hand, models not based on the eddy-viscosity assumption, such as scale-similarity models [2], which assume that the structure of the largest unresolved scales is similar to the smallest resolved scales and can reproduce energy backscatter (energy transfer from subfilter to resolved scales), generally result in better correlations during the *a priori* tests, but lack the necessary dissipation to be numerically stable. In this context, data-driven SGS models show promise. Machine Learning (ML)-based architectures may provide more flexible representations of SGS stresses and energy transfer. A recent review [3] shows how neural network (NN)-based architectures have been explored to improve functional models such as those discussed above by learning free parameters to improve their adaptability to different flows and turbulence properties or to learn the full SGS tensor directly. These studies differ considerably in network architecture, input and output variables, training strategies, and flow configurations, whose turbulence characteristics can strongly influence the resulting closure performance. Moreover, no clear procedure has been established for determining which aspects to assess to judge whether an SGS model performs satisfactorily in both *a priori* and *a posteriori* evaluation. Much of the existing literature has demonstrated improved performance for specific test cases, but comparatively little attention has been devoted to why a given methodology succeeds or fails, which physical and numerical ingredients control its generalisation, and how to exploit these aspects to design more robust and transferable SGS closures.

The **objective of the current internship** is to **propose and evaluate an appropriate ML architecture and strategy to optimise the *a priori* prediction in LES.** Extensive DNS databases available at the *Laboratoire de Mécanique des Fluides de Lille* (LMFL) of statistically stationary forced Homogeneous Isotropic Turbulence (HIT) at $Re_\lambda = 87$ and $Re_\lambda = 160$ will provide the investigation framework, as it offers a controlled environment in which the relationships between resolved structures, unresolved-scale dynamics, filtering and interscale energy transfer can be systematically analysed, giving a foundation for future tests in more complex configurations with inhomogeneity in any of the

directions. The spatial filtering of the databases will provide both the resolved flow quantities used as model inputs and the exact SGS closures. These data will enable the *a priori* evaluation of classical closures, including the Smagorinsky and scale-similarity models, alongside data-driven alternatives, without requiring additional LES simulations.

The *a priori* performance of the models will be assessed through complementary quantities: the SGS stress tensor τ_{ij}; the local interscale energy transfer $\Pi = -\tau_{ij}\overline{S}_{ij}$ between resolved and subfilter scales, where $\overline{S}_{ij} = \frac{1}{2}\left(\frac{\partial \overline{u}_i}{\partial x_j} + \frac{\partial \overline{u}_j}{\partial x_i}\right)$ is the filtered strain-rate tensor; and the SGS forcing $-\partial\,\tau_{ij}/\partial\,x_j$, which is the quantity directly entering the filtered Navier-Stokes momentum equations.

To do so, with the support of the SigMA team at CRIStAL[1], we would like to explore different aspects to gain expertise in understanding the turbulent structures and interscale interactions:

- Doing an exhaustive literature review of the state-of-the-art approaches and preparing filtered DNS datasets for training ML architectures.
- Training local NN-based architectures to study aspects such as the influence of input and output variables, network complexity, spatial filter type and width, and the numerical resolution used to represent filtered quantities on LES grids.
- Understanding which kind of physics we need to impose on our architecture, e.g., symmetries, Galilean and rotational invariances… (through invariants [4, 5] and/or loss functions [6]) to better address inter-scale energy transfer and, potentially, transferability.
- As a first exploratory objective, analysing the *a posteriori* behaviour from a few time steps of several runs, allowing for the assessment of the dynamical behaviour.
- As a second exploratory objective, assessing the role of non-locality incorporating spatial information through, for example, graph-based SGS formulations and neural operators, which are currently being explored in the literature.

Supervisors: Miguel MARTINEZ VALERO (miguel.martinezvalero@univ-lille.fr) – Jean-Philippe LAVAL (jean-philippe.laval@univ-lille.fr)

Keywords: Machine Learning; Large Eddy Simulations; turbulence

Location: *Laboratoire de Mécanique des Fluides de Lille* (LMFL), *Bâtiment M6 Cité Scientifique*, 59650, Villeneuve d'Ascq

Duration & start date: 6 months beginning February-March 2027, according to the availability of the candidate.

Application limit: Applications (CV and motivation letter) may be submitted to the email addresses provided above until **31 December 2026**. However, applications will be reviewed on a rolling basis, and we reserve the right to select a candidate before the deadline.

Expected skills and knowledge: Building on LMFL's established expertise in turbulence, we seek a candidate with a strong background in Big Data and Machine Learning, including experience with frameworks such as *PyTorch* or *TensorFlow*. Knowledge of High-Performance Computing (HPC), particularly GPU-accelerated computing, would be a valuable asset. The successful candidate will complement the laboratory's expertise by developing and applying advanced data-driven methodologies to exploit its extensive high-fidelity turbulence databases. The candidate could also benefit from access to the computational resources of the IDRIS national supercomputing centre to support their research activities if needed.

[1] https://www.cristal.univ-lille.fr/equipes/sigma/

[1] J. Smagorinsky, General circulation experiments with the primitive equations: I. The basic experiment, *Monthly Weather Review*, 91, n°99, (1963).
[2] J. Bardina and H. J. Ferziger and W. C. Reynolds, Improved subgrid-scale models for large-eddy simulation, *Proceedings of the 13th Fluid and Plasmadynamics Conference*, (1980).
[3] A. Prakash; K. E. Jansen and J. A. Evans, Perspective on machine-learning-based large-eddy simulation, *Physical Review Fluids*, 10, 110701, (2025).
[4] A. Prakash; K. E. Jansen and J. A. Evans, Invariant data-driven subgrid stress modeling in the strain-rate eigenframe for large eddy simulation, *Computer methods in applied mechanics and engineering*, 399, 115457, (2022).
[5] A. Prakash; K. E. Jansen and J. A. Evans, Invariant data-driven subgrid stress modeling on anisotropic grids for large eddy simulation, *Computer methods in applied mechanics and engineering*, 422, 116807, (2024).
[6] A. Wu and S. K. Lee, Two neural network Unet architecture for subfilter stress modeling, *Physical Review Fluids*, 10, 014601, (2025).