Oral presentation | Data science and AI
Data science and AI-IV
Fri. Jul 19, 2024 10:45 AM - 12:45 PM Room C

[13-C-04] Out-of-distribution prediction in RANS turbulence modelling

*Richard Dwight[1] (1. TU Delft)

Keywords: RANS closure modelling, data-driven modelling, out-of-distribution generalization

Twelfth International Conference on Computational Fluid Dynamics (ICCFD12), Kobe, Japan, July 14-19, 2024

Out-of-distribution prediction in RANS turbulence modelling

R.P. Dwight*, M. Lăcătuş* and T. Buchanan*
Corresponding author: r.p.dwight@tudelft.nl * Aerodymamics Group, Aerospace Faculty, TU Delft, The Netherlands.

1 Introduction

Data-driven modelling – in the context of RANS turbulence closures – is a class of methods for taking experimental, LES, or DNS data from a set of training flows, and producing a new RANS closure that performs accurately on these and similar flows. Generally a traditional 1-eqn or 2-eqn model is chosen as a baseline, and one or more terms are modified as a function of local, resolved flow features. In these methods, the terms corrected vary, as does the function representation, which may be a neural network, a random-forest, or a symbolic expression [1]. These varied approaches tend to show essentially similar performance: the resulting closure models are able to reproduce the training flows and similar flows better than the baseline model, but the corrections do not generalize well. Nonetheless these approaches can be valuable, e.g. for shape optimization [2], or as cheap surrogates for LES [3].

The goal of building general-purpose closures however remains. Simply extending the training set to include a large and diverse set of flows, has had disappointing results [4]. Most current work in this direction uses some form of blending multiple models, each trained on some relatively simple flow, and activated under conditions similar to the training conditions [5, 6].

2 Problem Statement

The central issue that limits generalizability may be related to what is being learnt from the data. The data – e.g. a DNS of a specific flow – necessarily contains correlations (between the mean-state and the required correction) that are caused by the geometry and inflow conditions, in addition to those of the turbulence physics itself. It is only the latter that are relevant for our closure model, but these multiple causes are not straight-forward to separate. The most basic tool of causal inference is to perform experiments varying one cause at a time; yet it is not meaningful to vary the geometry of a flow without also varying the turbulence. Another approach is to exploit invariants, notably Galilean invariance and frame-invariance, yet these are already fully exploited in existing data-driven models. As such we ask: How can we separate the causes in our datasets?

3 Approach and Findings

Our approach follows Arjovsky et al. [7] who introduced Invariant Risk Minimization (IRM). The idea is: given data from multiple training "environments" $e \in \mathcal{E}_{\text{train}}$ (in our case representing multiple flows with LES data), to separate invariants $\Phi(\cdot)$, from spurious correlations by seeking a correction model $f(\Phi)$ which is simultaneously optimal in each training flow given Φ. In particular we solve the bi-level optimization problem:

$$\min_{f,\Phi} \sum_{e' \in \mathcal{E}_{\text{train}}} \|\boldsymbol{y}^{e'} - f \circ \Phi(\boldsymbol{w}^{e'})\|^2 \tag{1}$$

$$\text{s.t.} \quad f = \arg\min_{f'} \|\boldsymbol{y}^{e} - f' \circ \Phi(\boldsymbol{w}^{e})\|^2, \;\; \forall e \in \mathcal{E}_{\text{train}}, \tag{2}$$

where $\boldsymbol{w}$ represents the local flow-state (mean velocity-gradients, T.K.E., etc.), and $\boldsymbol{y}$ is the necessary correction – originating from the frozen approach [1]. For Φ to be a solution of the above system, there must exist an f which is simultaneously optimal for all training flows.

We do not make a linear simplification of this problem, as in [7], rather we solve the bi-level optimization directly. First we reduce the search space by using symbolic regression with complexity constraints. A multi-objective genetic programming method (NSGA-II) with gradient information for constant optimization is used to address (2), in which the Pareto front is used to assess the consistency of f.

We apply the method first to some simple fundamental testcases with well-understood invariances. We demonstrate the ability of the procedure to recover tensor-invariants from tensor-component features; and the ability to recover turbulence models with Galilean- and rotational invariance. Finally the method is applied to the question of RANS modelling of separated flows. We consider a number of separated flows with varying geometries but identical topologies, and present a feature showing substantial invariance across these flows.

References

[1] Martin Schmelzer, Richard P. Dwight, and Paola Cinnella. Discovery of algebraic reynolds-stress models using sparse symbolic regression. *Flow, Turbulence and Combustion*, 104(2–3):579–603, December 2019.

[2] Yu Zhang, Richard P. Dwight, Martin Schmelzer, Javier F. Gómez, Zhong-hua Han, and Stefan Hickel. Customized data-driven rans closures for bi-fidelity les–rans optimization. *Journal of Computational Physics*, 432:110153, May 2021.

[3] Julia Steiner, Axelle Viré, and Richard P. Dwight. Classifying regions of high model error within a data-driven rans closure: Application to wind turbine wakes. *Flow, Turbulence and Combustion*, 109(3):545–570, August 2022.

[4] Christopher L. Rumsey, Gary N. Coleman, and Li Wang. In search of data-driven improvements to rans models applied to separated flows. In *AIAA SCITECH 2022 Forum*. American Institute of Aeronautics and Astronautics, January 2022.

[5] Adrián Lozano-Durán and H. Jane Bae. Machine learning building-block-flow wall model for large-eddy simulation. *Journal of Fluid Mechanics*, 963, May 2023.

[6] M. de Zordo-Banliat, G. Dergham, X. Merle, and P. Cinnella. Space-dependent turbulence model aggregation using machine learning. *Journal of Computational Physics*, 497:112628, January 2024.

[7] Martin Arjovsky, Léon Bottou, Ishaan Gulrajani, and David Lopez-Paz. Invariant risk minimization, 2020.