ARTICLE IN PRESS

International Journal of Heat and Fluid Flow xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

International Journal of Heat and Fluid Flow

journal homepage: www.elsevier.com/locate/ijhff

The development of algebraic stress models using a novel evolutionary algorithm

J. Weatheritt*,[1], R.D. Sandberg[2]

Department of Mechanical Engineering, University of Melbourne, Parkville, VIC, 3010, Australia

ARTICLE INFO

Keywords:
RANS modelling
Machine-learning
Evolutionary algorithm
Gene expression programming
Algebraic stress modelling

ABSTRACT

This work presents developments to a novel evolutionary framework that symbolically regresses algebraic forms of the Reynolds stress anisotropy tensor. This work contributes to the growing trend in machine-learning for modelling physical phenomena. Our framework is shown to be computational inexpensive and produce accurate and robust models that are tangible mathematical expressions. This transparency in the result allows us to diagnose issues with the regressed formulae and appropriately make amendments, as we further understand the regression tools. Such models are created using hybrid RANS/LES flow field data and a passive solving of the RANS transport equations to obtain the modelled time scale. This process shows that models can be regressed from a qualitatively correct flow field and fully resolved DNS is not necessarily required. Models are trained and tested using rectangular ducts, an example flow genus that linear RANS models even qualitatively fail to predict correctly. *A priori* and *a posteriori* testing of the new models show that the framework is a viable methodology for RANS closure development. This *a posteriori* agenda includes testing on an asymmetric diffuser, for which the new models vastly outperform the baseline linear model. Therefore this study presents one of the most rigorous and complete CFD validation of machine learnt turbulent stress models to date.

1. Introduction

For many design problems of engineering interest, flow is predicted primarily with Reynolds-Averaged Navier-Stokes (RANS) equations. This is because of the excessive computational effort required for Large Eddy (LES) and Direct Numerical Simulation (DNS) techniques (Hanjalić, 2005). That said, RANS modelling is based on very limiting assumptions that often fall down unexpectedly and with catastrophic repercussions for statistical results on moderately complex geometries (Hunt and Savill, 2005). This is because RANS model uncertainty remains poorly understood (Ling and Templeton, 2015). Particular examples include the now classical periodic hills test case (Temmerman and Leschziner, 2001; Temmerman et al., 2003; Fröhlich et al., 2005) and the asymmetric diffuser of Cherry et al. (2008); 2009). ERCOFTAC Workshops surrounding the former (Jakirlić et al., 2001; Manceau, 2003) and latter (Steiner et al., 2009) show that RANS closures fail to predict even global flow features *reliably*. For the periodic hills test case, this manifests as poor reverse flow prediction and for the diffuser, separation often occurred from the *wrong* side of the duct. Note, the periodic hills have been subject to a recent study (Weatheritt and Sandberg, 2016c), whilst the diffuser is a large focus of this paper.

Because of this high uncertainty, a plethora of RANS methods exist throughout the literature (e.g. Leschziner, 2015) and there is no general consensus on a particular approach. There perhaps may not be a universal RANS closure that outperforms all others for an arbitrary flow configuration (Spalart, 2000). With respect to industry standard, models tend to be variants of the $k-\omega$-SST (Menter, 1994), $k-\varepsilon$ (Chien, 1982) and Spalart and Allmaras (1994) turbulence closures. These three all utilise a linear stress-strain relationship, despite well known theoretical (e.g. Schmitt, 2007) and practical shortfalls (e.g. Wilcox, 1993), because of its robustness and the uncertainty in more complex approaches (Pope, 1999).

Instead, this work adheres to what (Spalart, 2015) terms, an 'Openly Empirical' approach. That is, the turbulence closure aims to correctly model physics but contains no 'explicit connection' to the exact turbulence equations. The models proposed in this work are derivatives (not in the formal sense) of high-fidelity data sets. Because of the shortfalls in the linear stress-strain relationship, this is a primary candidate for improvement when developing turbulence closures. Rather

* Corresponding author.
E-mail addresses: jack.weatheritt@unimelb.edu.au (J. Weatheritt), richard.sandberg@unimelb.edu.au (R.D. Sandberg).
[1] Post-doctoral researcher.
[2] Chair of computational mechanics.

http://dx.doi.org/10.1016/j.ijheatfluidflow.2017.09.017
Received 13 April 2017; Received in revised form 18 August 2017; Accepted 27 September 2017
0142-727X/

than assuming the task of discovering the universal relationship, we propose a *universal framework* for *specific closure* formulation. That is, for a given database one obtains a class of stress-strain relationships suitable for flows at least topologically similar. This is demonstrated in this paper. Changing model coefficients is widely practised for a calculation, using a given model, for a specific flow. In this paper, we generalise to alter the stress-strain functional form. Indeed, many turbulence models have been developed for specific flows, take the examples of jets in crossflow (Bergeles et al., 1978) and turbulent wall jets (Ljuboja and Rodi, 1981). The latter model was developed specifically to overcome the standard model's tuning on free jets. We propose new non-linear stress-strain relationships, known widely as Explicit Algebraic (Reynolds) Stress Models (EASMs or EARSMs) (Pope, 1975; Rodi, 1976), of which many such models exist (e.g. Gatski and Speziale, 1993; Wallin and Johansson, 2000; Craft et al., 2000).

The task of formulating models from data, in the manner presented here, falls under the umbrella of techniques widely known as machine-learning. Applications of such techniques for turbulence model development is a fast growing area of research and we would denote it 'Maximally Empirical', to extend the taxonomy of Spalart (2015). That is, we enforce Galilean invariance — through an integrity basis (Spencer and Rivlin, 1958; 1959) — but beyond this, the machine-learning algorithm is entirely free to formulate mathematical expressions that relate stress to strain.

Similar studies have emerged in previous years which we classify, rather broadly, into: uncertainty quantification, model development demonstrating *a priori* success and model development demonstrating *a posteriori* success. Model development infers some modification to the turbulence closure, whilst we define *a priori* demonstrations as still relying on a high fidelity database and *a posteriori* demonstrations as full CFD of a flow problem, converging from some initial condition. Uncertainty quantification, therefore, aims to understand why certain models fail and model development attempts to correct such deficiencies. *A priori* relates to the assumption that machine-learning approaches successfully minimising some objective on high-fidelity data, a process known as training, will improve CFD performance. *A posteriori* aims to demonstrate this reasoning and is consequently much harder to establish, yet vital for successful industrial implementation.

In efforts to understand uncertainty, databases can be inspected and used as a road-map (Hunt and Savill, 2005; Schmitt, 2007), to give the engineer an idea of whether a given closure is applicable for a given flow topology. More complex machine-learning approaches have had their ability to identify regions of inaccuracy assessed (Ling and Templeton, 2015). Presenting another method of general uncertainty identification, Spalart et al. (2015) have passively solved the RANS equations using a DNS mean flow to assess closure quality. Edeling et al. (2014b) used Bayesian methods for identifying the uncertainty in the values of closure coefficients, whilst Xiao et al. (2016) applied similar methods for quantifying uncertainty in predicted flow quantities.

In terms of model development, the eddy viscosity has been optimised to find the β^* coefficient in the $k - \omega$-SST turbulence model that best applies to a turbomachinery flow (Weatheritt et al., 2017b). This was done using least-squares regression in an *a priori* manner that has been previously applied many times (Muldoon and Acharya, 2006; Spalart et al., 2015; Pichler et al., 2015). Correction parameters have been learnt and added to the model transport equations by Parish and Duraisamy (2016); however, these were a function of space, limiting any *a posteriori* demonstration to an identical geometry. Such studies have elicited the argument to ensure Galilean invariance in machine-learning approaches and, when adhered to, the resulting model performs better (Ling et al., 2016). Edeling et al. (2014a) have used Bayesian methods to modify model coefficients for a full *a posteriori* demonstration. This is encouraging, yet proved to be a computationally intensive approach.

More complex regression methods, that alter more than the coefficients in the original model are possible (say adding new non-linear terms). For example, neural networks (e.g. LeCun et al., 2015) have been used to replace terms in the RANS equations. Ling et al. (2016) used deep neural networks to model the Reynolds stress and whilst showing good *a priori* agreement, the only *a posteriori* demonstration required the matrix of Reynolds stresses at each grid point to be inserted into the solver — thus limiting the model to the same flow geometry. Because of this apparent difficulty at full CFD implementation, a neural network has been used to correct a converged linear RANS solution (Weatheritt et al., 2017a) in a similar fashion to random forest regression (Ling et al., 2017). The primary issue with this approach is, that should the linear RANS model predict a vastly inaccurate flow field, as with the present study, then the modified Reynolds stresses are still a function of the incorrect velocity field. In other words, the method is unable to alter global phenomena such as separation and reattachment points, for example.

Another emerging approach to model development is via symbolic regression, which most commonly manifests itself as a variant of genetic programming (Koza, 1992). Symbolic regression aims to return a mathematical equation that best fits some high-fidelity data, both in terms of error and simplicity. No model is provided as a starting point by the user and the algorithm searches the space of *all* mathematical expressions. This has several advantages over other machine-learning techniques. Primarily, the resulting equation is tangible and can be implemented into a CFD code readily. Genetic programming is an evolutionary algorithm (e.g. Steeb, 2014), which evolves a collection of candidate solutions analogous to Darwin's (1858) theory of natural selection. Such approaches have been applied to separated flow (Weatheritt and Sandberg, 2016c); a modified version of the original gene expression programming (GEP) concept of Ferreira (2001), suitable for tensor regression, showed excellent skin friction prediction in *a posteriori* demonstrations. This was achieved by optimising a non-linear stress-strain relationship. However, the modelled time scale was not correctly accounted for during the machine-learning phase[3] and the objective function did not account for the magnitude of the modelled Reynolds stress. Further studies for RANS modelling have performed well in *a priori* studies (Weatheritt et al., 2017a; 2017b). Note, the framework is also being developed for hybrid RANS/LES methods (Weatheritt and Sandberg, 2015; 2016a; 2016b).

In this work, we present an extension of our GEP evolutionary framework to account for the proper modelling of the turbulent time scale and new objective function to account for the magnitude of the Reynolds stress. We develop this framework with the following considerations:

1. Resulting models are easy to implement into CFD codes.
2. Resulting models are Galilean invariant.
3. Demonstrate *a priori* success on flows similar to the high-fidelity training database.
4. Demonstrate *a posteriori* success on complex flows, by solving all transport equations from an initial condition. Hence validating *exactly* as a conventional turbulence model would be.
5. The entire framework is computationally feasible for industrial application.

By ensuring these points, this paper presents a major contribution to the field of machine-learning for turbulence modelling. Above, we identified the linear stress-strain relationship as a major contributor to uncertainty and so we target this for improvement, effectively producing an EASM. We apply our framework to duct flows, using hybrid RANS/LES (H-R/L) data to build new stress-strain equations. The algorithm is non-deterministic, meaning for each database we produce a class of unique models and then gain statistical information regarding algorithm performance. The use of H-R/L shows that high-resolution

[3] Although the authors argued this was unnecessary.