

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/331343246>

Turbulence Model Development using CFD-Driven Machine Learning

Preprint · February 2019

CITATIONS

0

READS

1,464

5 authors, including:



Yaomin Zhao

Peking University

29 PUBLICATIONS 240 CITATIONS

SEE PROFILE



Harshal Akolekar

University of Melbourne

14 PUBLICATIONS 123 CITATIONS

SEE PROFILE



Jack Weatheritt

University of Melbourne

20 PUBLICATIONS 511 CITATIONS

SEE PROFILE



Vittorio Michelassi

Baker Hughes Incorporated

153 PUBLICATIONS 2,133 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Lagrangian studies on wall-bounded transitional flows [View project](#)



Reduced-order Modeling for Membrane Wings [View project](#)

Turbulence Model Development using CFD-Driven Machine Learning

Yaomin Zhao^{1†}, Harshal D. Akolekar¹, Jack Weatheritt¹, Vittorio Michelassi², Richard D. Sandberg^{1‡}

¹Department of Mechanical Engineering, University of Melbourne, VIC 3010, Australia

²Baker Hughes, a GE Company, Via Felice Matteucci 2, Firenze 50127, Italy

(Received xx; revised xx; accepted xx)

This paper presents a novel CFD-driven machine learning framework to develop Reynolds-averaged Navier-Stokes (RANS) models. For the CFD-driven training, the gene expression programming (GEP) method (Weatheritt & Sandberg, *J. Comput. Phys.*, **325**, 22-37 (2016)) uses RANS calculations in an integrated way to evaluate the fitness of candidate models. The resulting model, which is the one providing the most accurate CFD results at the end of the training process, is thus expected to show good performance in RANS calculations. To demonstrate the potential of this new approach, the CFD-driven machine learning is applied to develop a model for improved prediction of wake mixing in turbomachines. A new model is trained based on a high-pressure turbine training case with particular physical features. The developed model is shown to have a more compact functional form than models trained without CFD assistance. Furthermore, the trained model has been evaluated *a posteriori* for the training case and three additional test cases with different physical flow features, and the predicted wake mixing profiles are significantly improved in all cases. With the present framework, the model equation is explicitly given and available for analysis, thus it could be deduced that the enhanced wake prediction is due to the extra diffusion introduced by the CFD-driven model.

Key words: Turbulence Modelling, Wakes

1. Introduction

Reynolds-averaged Navier-Stokes (RANS) remains the primary tool in most industrial applications due to its low computational cost compared to high-fidelity (Hi-Fi) simulations such as direct numerical simulation (DNS) and large-eddy simulation (LES). However, RANS suffers from low-fidelity in prediction especially for complex geometries and flow fields (Hunt & Savill 2005), which is mainly due to the limiting assumptions used to model the Reynolds stress tensor τ_{ij} . Therefore, there has been continuous interest in both academia and industry to improve the predictive accuracy of RANS models routinely applied to aerodynamic and aerothermal design verifications, especially in the aerospace field.

Recently, machine learning has gained popularity in turbulence modelling (Duraissamy *et al.* 2019). Various machine-learning techniques have been applied to modify traditional

[†] Email address for correspondence: yaomin.zhao@unimelb.edu.au

[‡] Email address for correspondence: richard.sandberg@unimelb.edu.au

RANS models to enhance the flow prediction, including modifying model parameters with Bayesian methods (Edeling *et al.* 2014), introducing a correction factor for the turbulence production term using neural networks (Zhang & Duraisamy 2015), and adding a spatially distributed correction field via field inversion and Gaussian Process (Parish & Duraisamy 2016), etc. However, the various types of corrections trained for traditional RANS models are usually not physically interpretable.

Other than simply adding corrections to existing model parameters, more comprehensive efforts have been made to construct new Reynolds stress closures via physics-informed machine learning. In particular, many studies have focused on improving the Boussinesq hypothesis

$$\tau_{ij} = \frac{2}{3}\rho k\delta_{ij} - 2\mu_t S'_{ij}, \quad (1.1)$$

which assumes a linear relation between the anisotropy of the Reynolds stress $a_{ij} = \tau_{ij} - \frac{2}{3}\rho k\delta_{ij}$ and the deviatoric strain-rate tensor S'_{ij} . Such linear relation only holds valid in limited portions of the flow field, while it is questionable in complex flow topologies. Here, ρ denotes density, k denotes turbulence kinetic energy, and μ_t represents turbulence viscosity. A random forest method was used in Wang *et al.* (2017) to train Reynolds stress discrepancy functions based on the a_{ij} tensor eigenvalues. The trained discrepancy functions were applied in test cases to show improved prediction of τ_{ij} , but the performance of mean quantities needs to be further validated. With basis tensors and scalar invariants decomposed from a_{ij} , Ling *et al.* (2016) used a deep neural network to train the Reynolds stress tensor conserving Galilean invariance. Nevertheless, the resulting neural network from the training is not straightforward to be implemented in RANS solvers.

In a series of studies, the gene-expression programming (GEP) method has been introduced (Weatheritt & Sandberg 2016) to develop explicit algebraic (Reynolds) stress models (EASM) based on the stress tensor decomposition proposed by Pope (1975). EASM-like turbulence closures are represented in the form

$$a_{ij} = \sum_k f^{(k)}(I_1, I_2, \dots, I_n) V_{ij}^k, \quad (1.2)$$

which can be taken as a solution to the shortcomings of the Boussinesq hypothesis. This implies the non-dimensionalized anisotropic stress tensor a_{ij} is represented by a combination of tensor bases V_{ij}^k and scalar invariants I_j . Models following equation (1.2) can be explicitly provided through symbolic regression by GEP, which is used to determine the coefficients $f^{(k)}$ based on training data. Applied in RANS solvers, these GEP-trained models have shown improved predictive accuracy in several *a posteriori* tests such as rectangular ducts (Weatheritt *et al.* 2017) and turbomachinery flows (Weatheritt & Sandberg 2017; Akolekar *et al.* 2018).

Although machine learning for turbulence model development is becoming a growing trend, obstacles still exist in implementing the trained models to engineering applications. One of the major concerns is that most of the training methods aim at accurately modelling the Reynolds stress tensor from Hi-Fi data, while the inherent inconsistency between RANS modelling and high-fidelity data has been neglected (Duraisamy *et al.* 2019). As shown in previous studies (Thompson *et al.* 2016; Poroseva *et al.* 2016), RANS prediction can remain unsatisfactory even if Reynolds stresses from DNS are used in the CFD calculation. Such difficulty was also documented by Parneix *et al.* (1996), who tuned a full Reynolds stress model with the aid of the DNS of a backward facing step, but the application of the tuned model to RANS did not confirm the expected improvements. Therefore, it is questionable whether existing model development methods targeting

only to accurately model τ_{ij} from Hi-Fi data can provide accurate mean flow fields. Rather than training models exclusively based on high-fidelity data, model development strategies more adaptive to RANS calculations are needed to provide models applicable for pragmatic CFD.

In order to train models adaptive to RANS environments, a novel model development framework named CFD-driven training is introduced in the present work. By integrating RANS calculations in the GEP model training method introduced by Weatheritt & Sandberg (2016), the objective is to develop Reynolds stress closures of the form given in equation (1.2), which are i) straightforward to be implemented in RANS solvers; ii) robust in a sense of providing stable solutions in RANS calculations; and iii) showing improved predictive accuracy in different *a posteriori* tests.

The present paper is structured as follows. The CFD-driven training frameworks along with the basics for symbolic regression via the GEP method are introduced in § 2. A brief description of the case setups are also listed in § 2. Then the CFD-driven training is applied to turbine wake mixing cases, and the developed models are tested across different cases in § 3. Finally, conclusions are provided in § 4.

2. Methodology and Case Setup

2.1. GEP for model development

GEP, first introduced for turbulence model development by Weatheritt & Sandberg (2016), is applied in the present study to train an EASM-like Reynolds stress relation by adding an extra anisotropic stress a_{ij}^x as

$$\tau_{ij} = \frac{2}{3}\rho k\delta_{ij} - 2\mu_t S'_{ij} + a_{ij}^x. \quad (2.1)$$

The a_{ij}^x is represented as a linear combination of tensor bases V_{ij}^k and scalar invariants I_k as in equation (1.2). To focus on statistically two-dimensional flows which are representative of the mid-span section of turbine blades, the invariants and basis tensors considered in the present work are

$$\begin{aligned} V_{ij}^1 &= s_{ij}, V_{ij}^2 = s_{ik}\omega_{kj} - \omega_{ik}s_{kj}, \\ V_{ij}^3 &= s_{ik}s_{kj} - \frac{1}{3}\delta_{ij}s_{mn}s_{nm}, \\ I_1 &= s_{mn}s_{nm}, I_2 = \omega_{mn}\omega_{nm}. \end{aligned} \quad (2.2)$$

Here, V_{ij}^k and I_k are functions of the non-dimensionalized strain rate tensor $s_{ij} = t_I S'_{ij}$ and the rotation rate tensor $\omega_{ij} = t_I \Omega_{ij}$, where t_I is the turbulence time scale.

The GEP model training starts from a randomly generated population consisting of candidate EASMs a_{ij}^{GEP} . The population of models evolves in the GEP iterations by mutations and combinations until the best model in the generation provides an acceptable approximation to pre-set target functions. Through this non-deterministic evolution by GEP, symbolic regression is achieved to find an optimal model that best fits the training data (Ferreira 2001).

While more details about the GEP method are given in Weatheritt & Sandberg (2016), we want to stress here that the model development characteristics are mainly determined by the selection process. The fitness of every generated model is evaluated by its cost function $J(a_{ij}^{GEP})$, which describes the deviation between the model and training data. The general principal for the definition of the cost function is that candidate models that better fit training data should have lower cost values, which result in better chances to

be selected for the next generation. We note that the specific form of the cost function usually depends on the training framework and the cases, which will be discussed in detail in § 2.2 and § 3.1.

2.2. Training frameworks

The training framework, in particular how to evaluate candidate models during evolution, is critical to the quality of the final model. Previous GEP studies (Weatheritt & Sandberg 2016, 2017; Akolekar *et al.* 2018), along with other existing machine-learning driven model development work (e.g. Ling *et al.* 2016; Parish & Duraisamy 2016), typically evaluate the models depending on time-averaged Hi-Fi data. This is denoted as ‘frozen’ training in the present study, with the ‘frozen’ here meaning the pre-processed Hi-Fi data remains unchanged in the training. It is noted that the inherent inconsistency between Hi-Fi data and RANS environment, such as discrepancies of the turbulent dissipation rate providing turbulence scales for non-dimensionalization, limits the applicability of the models trained with frozen training (Duraisamy *et al.* 2019). Therefore, a novel framework that includes CFD calculations in model training iterations is proposed in the present study. The CFD-driven training, aiming to train models well adapted to the RANS mean velocity and turbulence scales, is introduced following a brief description of the frozen training in this section.

2.2.1. Frozen training

For frozen training, candidate models generated by GEP are evaluated based on Hi-Fi data. With the tensor bases V_{ij}^k and scalar invariants I_k processed from a Hi-Fi flow field, the a_{ij}^{GEP} can be calculated using the candidate model equations from GEP. On the other hand, the a_{ij}^{HiFi} directly obtained from Hi-Fi data is considered as ‘accurate’ Reynolds stress anisotropy. The cost functions are defined by the deviation between a_{ij}^{GEP} and a_{ij}^{HiFi} . An example used in the present study is

$$J(a_{ij}^{GEP}) = \frac{1}{N} \sum_{n=1}^N \sum_{i,j} \frac{|a_{ij}^{HiFi} - a_{ij}^{GEP}|}{|a_{ij}^{HiFi}|}, \quad (2.3)$$

where N denotes the number of training data points. We can infer from equation (2.3) that candidate models which better fit Hi-Fi data will have a lower cost. The cost functions are then used to direct the selection and reproduction of the models. Through a series of iterations, the final model with minimized cost functions will be obtained.

We can see that the models are trained exclusively based on Hi-Fi data. The frozen-trained model generally provide accurate a_{ij} predictions as long as Hi-Fi basis tensors and scalar invariants are available as input. However, once the model is implemented into RANS, good performance is not guaranteed. In fact RANS will predict a flow field that may deviate from the Hi-Fi flow field. The impact of such variations on the GEP trained model is non-linear and difficult to forecast.

2.2.2. CFD-driven training

As shown by the schematic for the CFD-driven training in figure 1, the candidate models are evaluated by running RANS calculations that will provide feedback in the GEP training loop. For every generation of candidate models, the model equations are dynamically read into the RANS solver, and then a series of real-time CFD calculations are performed to test different models in parallel. The cost functions are quantified based on the RANS mean flow fields. If required, a new generation of models is formed through

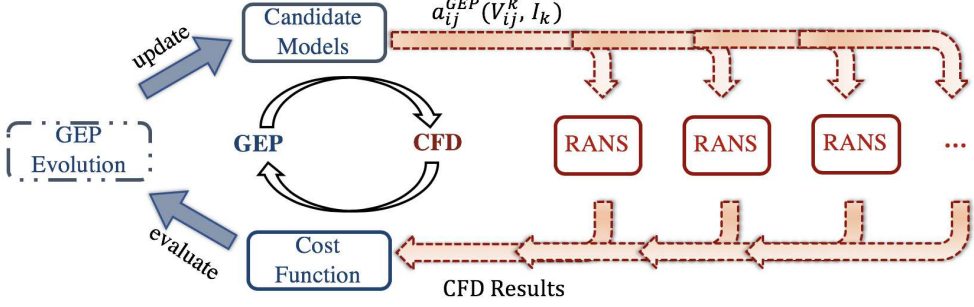


FIGURE 1. Schematic for CFD-driven training. The CFD calculations represented by red symbols are integrated in the GEP training loop, taking turbulence stress model equations as input and providing cost functions of candidate models for the evolution algorithm.

GEP evolution. Then the updated models will be evaluated again through the in-loop CFD calculations.

We remark that the CFD-driven training is executable with the GEP method because the model equations are explicitly given and can be instantly implemented into RANS solvers in iterations. With RANS calculations performed in the training loop, the CFD-driven models thus are ready to be implemented into industrial design tools. Another advantage of CFD-driven training is that the definition of cost function is more flexible compared to frozen training. Rather than being restricted to quantities that are part of the closure terms in frozen training, the CFD-driven cost functions can be defined based on any important flow feature of interest to designers, and an example for enhanced wake mixing predictions will be shown in § 3.1.

3. Numerical Results

3.1. Case setup for turbine wake mixing

As key components of turbomachines, high-pressure turbine (HPT) and low-pressure turbine (LPT) flows are challenging for RANS calculations due to their extremely complex flow features. Four different turbine cases, including the VKI LS89 HPT (Arts *et al.* 1990) and the T106A LPT (Stadtmüller & Fottner 2001) at different Reynolds numbers, are selected to test the CFD-driven training framework. The geometries of the HPT and LPT blades are shown in figure 2, with periodic boundary conditions implemented in the pitchwise direction denoted by y , and inflow/outflow conditions applied at the boundaries in the axial direction by x . The blade geometries and the inflow angles indicated by the black dashed arrow in figure 2 show the considerable differences between the cases. Furthermore, as listed in table 1, the flows in these cases are at different Reynolds and Mach numbers, and the flow features of the boundary layers on the blade surface also deviate significantly. We use the $Re = 0.57 \times 10^6$ HPT case, denoted as case A, as the training case, while the other three cases are used as testing cases for cross-validation, to assess whether the model trained in case A is suitable for a broad parameter space.

The Hi-Fi data for these cases is provided from previous DNS and high-resolution LES using code HiPSTAR (Sandberg *et al.* 2015; Michelassi *et al.* 2015; Pichler *et al.* 2017), and RANS calculations are conducted with the RANS solver TRAF (Arnone & Pacciani 1995). In the RANS calculations required for model development and testing, the $k - \omega$ SST with $\gamma - Re_\theta$ transitional correlation (Langtry & Menter 2009) is applied as the

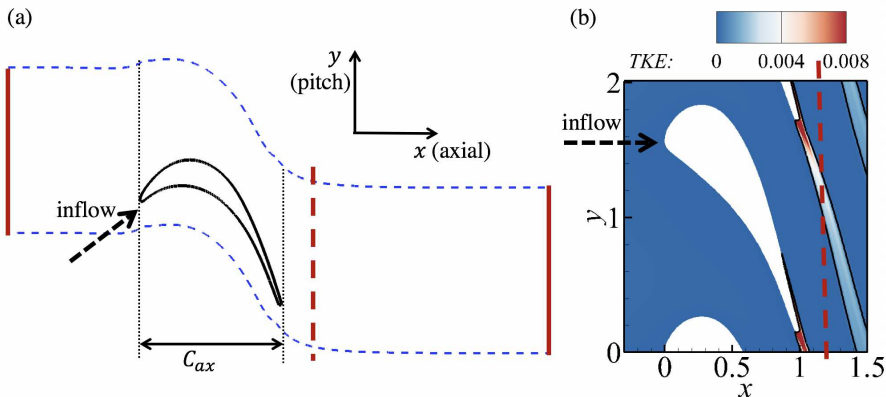


FIGURE 2. Setup for cases: (a) LPT blade geometry; (b) HPT blade geometry and the wake region indicated by TKE contour, with the blade leading edge at $x = 0$.

Cases	Re	Ma_{exit}	Flow features	Purposes
HPT A	0.57×10^6	0.9	transition & shocks	training
HPT B	1.1×10^6	0.9	transition & shocks	testing
LPT C	0.6×10^5	0.4	transition & open separation	testing
LPT D	1.0×10^5	0.4	transition & closed separation	testing

TABLE 1. Parameters for turbine flow test cases.

baseline model. A thorough grid independence study has also been conducted for the RANS calculation to ensure results are grid-independent.

As traditional RANS models applied in turbine flows share a common weakness in correctly capturing wake mixing (Pichler *et al.* 2016), the wake profile prediction, which is of critical interest for aerodynamic performance and characteristics of downstream blade boundary layers, is set as the optimisation target. Accordingly, the GEP fitness function for the CFD-driven training is defined as

$$\begin{aligned}
 \omega(y) &= \frac{p_t^i - p_t(y)}{p_t^i - p^o}, \\
 \Delta_x &= \frac{1}{L_y} \int_0^{L_y} \left(\frac{\omega_{HiFi}(y) - \omega_{RANS}(y)}{\max_y(\omega_{HiFi})} \right)^2 dy, \\
 J &= \Delta_{x_1} + \Delta_{x_2}.
 \end{aligned} \tag{3.1}$$

Here, $\omega(y)$ represents the kinetic loss profile along the pitchwise axis y as represented by the red dashed lines downstream of the blade in figure 2, with p_t^i , p_t^o , and p^o denoting inlet total pressure, outlet total pressure, and outlet static pressure, respectively. Moreover, Δ_x stands for normalised deviation between wake loss profiles from Hi-Fi and RANS results. In order to train models for the whole wake region rather than just a profile at one x location, the final fitness function is defined as a summation of Δ_x at two axial locations $x_1 = 1.15C_{ax}$ and $x_2 = 1.25C_{ax}$, as these two axial locations represent the range of most probable downstream blade leading edge position.

To focus on improving model accuracy in wake mixing, training data is extracted from the turbine wake. As indicated by the turbulence kinetic energy (TKE) contour in figure 2(b), the wake region is masked by the criterion $k > 5\%k_{max}$. Moreover, another

constraint $x > 1.04$ is implemented to avoid the boundary layer and near trailing edge interference (Weatheritt & Sandberg 2017; Akolekar *et al.* 2018). Note that the trained Reynolds stress closure is only implemented in the masked wake region for both model development and testing.

3.2. Model development

The CFD-driven training framework is applied in the HPT case A to develop a new Reynolds stress closure for turbine wake mixing. Due to the RANS calculations required in the training loop, the CFD-driven training in the present study typically requires 1000 CPU hours which is more computationally expensive than the frozen training. Nevertheless, this is still acceptable as it is only required once, and the trained models can then be used for subsequent RANS calculations with no additional cost. The resulting model is presented as τ_{ij}^{CFD} . In addition, a frozen-trained model τ_{ij}^{fro} from previous work in Weatheritt *et al.* (2017) is shown for comparison

$$\begin{aligned}\tau_{ij}^{CFD} &= \frac{2}{3}\rho k\delta_{ij} - 2\mu_t S'_{ij} + 2\rho k[\\ &\quad (-3.57 + I_1)V_{ij}^1 + 4.0V_{ij}^2 + (-0.11 + 0.09I_1I_2 + I_1I_2^2)V_{ij}^3], \\ \tau_{ij}^{fro} &= \frac{2}{3}\rho k\delta_{ij} - 2\mu_t S'_{ij} + 2\rho k[\\ &\quad (-1.334 + 0.438I_1 + 2.653I_2 + 0.0102I_1^2 - 1.021I_2^2 + 12.280I_1I_2)V_{ij}^1 \\ &\quad + (0.573 - 1.096I_1 + 8.985I_2 - 0.1102I_1^2 + 2.876I_2^2 + 90.633I_1I_2)V_{ij}^2 \\ &\quad + (12.861 - 25.094I_1 + 6.449I_2 + 1.020I_1^2 - 304.979I_1I_2 - 184.519I_2^2)V_{ij}^3].\end{aligned}\tag{3.2}$$

It is noted that the CFD-driven model exhibits a much simpler form compared to the frozen model, as the coefficient functions contain less high-order invariants. This is because that the higher-order terms in the frozen model, like $90.633I_1I_2$, usually result in very high-amplitude values on some points in RANS calculations, reducing numerical stability. Therefore, such terms fail to survive in the CFD-driven evolution, and the training via running RANS calculations in the loop tends to produce a more robust model. As non-linear turbulence model closures are known to come with stability issues, the capability of the CFD driven training to filter out numerically stiff models is very important.

3.3. Prediction results

The GEP models in equation (3.2) are first tested *a posteriori* for the training HPT case A. The trained model equations are implemented into the RANS solver, and the wake mixing profiles from RANS calculations with these models are shown in figure 3(a). In addition, profiles from Hi-Fi data and baseline models without any additional stress tensors are presented for comparison. We can see that the baseline model relying on a linear stress-strain relationship significantly over-predicts the wake peak and under-predicts the wake width. While the frozen model reduces the wake peak, the CFD-driven model provides an even better agreement with the reference Hi-Fi data, in terms of both the peak value and the profile width.

The wake prediction improvement from the trained models can be analysed by considering the leading-order terms in equation (3.2) because the scalar invariants I_1 and I_2 are less important compared to the constants (Akolekar *et al.* 2018). With $\mu_t S'_{ij} = \rho k V_{ij}^1$, we

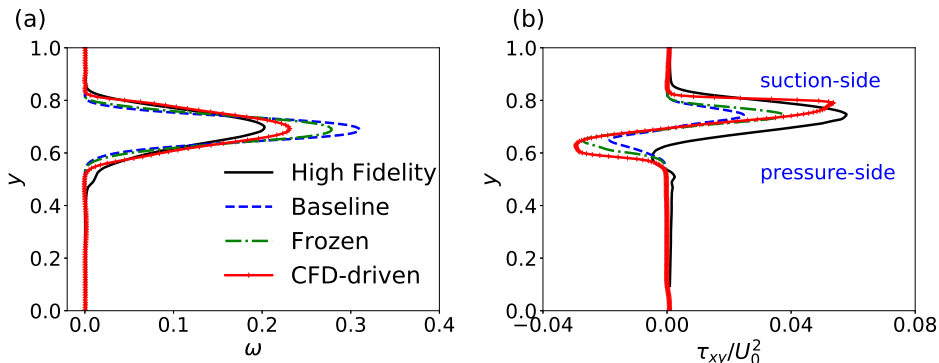


FIGURE 3. Profiles at 20% axial chord downstream of blade trailing edge for HPT case A, (a) kinetic wake loss; (b) shear stress term τ_{xy} .

have

$$\begin{aligned}\tau_{ij}^{CFD-L} &= \frac{2}{3}\rho k\delta_{ij} - 2\mu_t(1.0 + 3.57)S'_{ij}, \\ \tau_{ij}^{fro-L} &= \frac{2}{3}\rho k\delta_{ij} - 2\mu_t(1.0 + 1.334)S'_{ij}.\end{aligned}\tag{3.3}$$

Compared to the baseline model, the frozen model increases diffusion by roughly 133%, while the increase of the CFD-driven training is even larger at 357%. With the additional diffusion and extra anisotropic terms from V_{ij}^2 and V_{ij}^3 , the CFD-driven model allows for a stronger mixing and thus a better prediction of the wake profile spreading as seen in figure 3(a).

It is noted that improving predictions for mean flow quantities like kinetic wake loss is of critical interest for turbine designers, which is inherently guaranteed by the CFD-driven training as the fittest models are the ones with the smallest cost values. Improvement in Reynolds stress prediction, however, is also important for capturing the correct physics. As shown in figure 3(b), the CFD-driven model enhances the prediction of the shear stress profile, especially in terms of the suction-side peak value. Nevertheless, discrepancies exist on the pressure-side, which is presumably because steady RANS has limitations in modelling the coherent vortex shedding under the strong pressure-gradient near the blade trailing edge (Wheeler *et al.* 2016).

3.4. Cross-validation

The CFD-driven model in equation (3.2) is now applied to the other cases listed in table 1 to assess its applicability and generality in wake mixing flows with diverse flow parameters, geometries and key flow features. The kinetic wake loss profiles from this cross-validation are presented in figure 4, with the Hi-Fi data and baseline RANS available for comparison. As expected, the baseline RANS calculations based on the Boussinesq hypothesis result in over-prediction of the wake peak and under-prediction of the wake width across the different cases. In contrast, the CFD-driven model presents improved accuracy of the wake prediction in all test cases due to the additional diffusion and anisotropy added in the Reynolds stress. Considering the test cases cover a reasonably wide parameter space, this cross-validation underscores the robustness of the CFD-driven training.

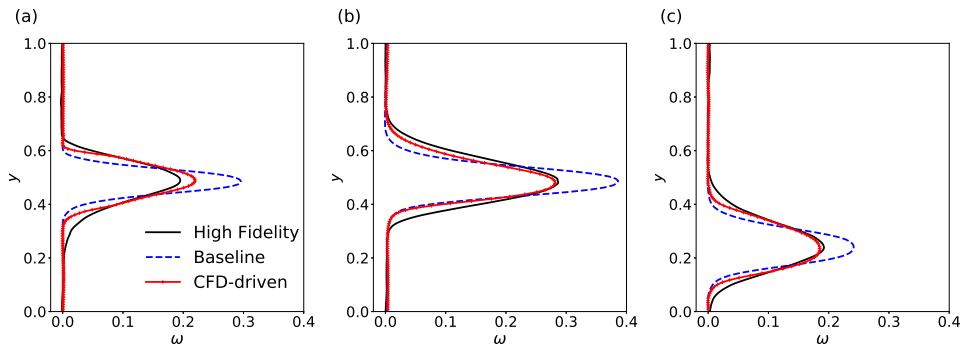


FIGURE 4. Kinetic wake loss profiles in test cases in table 1, (a) HPT case B; (b) LPT case C; (c) LPT case D.

4. Conclusions

A novel machine learning framework named CFD-driven training is introduced based on the GEP method to develop turbulence models. As the candidate EASM-like models are explicitly given via symbolic regression in GEP, the model equations can be implemented into a RANS solver, and the cost functions of the models are evaluated by running CFD in the training loop. By integrating the RANS calculations, the CFD-driven training is able to develop models adapted to the RANS environment and straightforward to be implemented into existing RANS solvers.

The CFD-driven framework was applied to one HPT case to train an EASM-like model for wake mixing in turbomachines. Compared to the traditional model training, the CFD-driven training produces a model that is relatively simple. Furthermore, the generated model is tested *a posteriori* for the training case and three other cases covering different flow configurations. Due to the extra diffusion introduced by the CFD-driven model, the results show overall good agreement with the kinetic loss profiles obtained from Hi-Fi data in all cases.

While the model trained in the present study is limited to wake mixing, the CFD-driven machine learning is shown to be a promising framework for general turbulence model development. Nevertheless, introducing RANS calculations in the model development iterations might increase the computational cost of training. Therefore, an efficient CFD solver needs to be integrated in the training loop.

This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725. This work was also supported by the resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia.

REFERENCES

- AKOLEKAR, H. D., WEATHERITT, J., HUTCHINS, N., SANDBERG, R. D., LASKOWSKI, G. & MICHELASSI, V. 2018 Development and use of machine-learnt algebraic Reynolds stress models for enhanced prediction of wake mixing in LPTs. *J. Turbomach.* (accepted manuscript).
- ARNONE, A. & PACCIANI, R. AND SESTINI, A. 1995 Multigrid computations of unsteady rotor-stator interaction using the Navier-Stokes equations. *J. Fluids Eng.* **117** (4), 647–652.
- ARTS, T., LAMBERTDEROUVOIT, M. & RUTHERFORD, A. W. 1990 Aero-thermal investigation

- of a highly loaded transonic linear turbine guide vane cascade. A test case for inviscid and viscous flow computations. *NASA STI/Recon Technical Report N* **91**.
- DURAISAMY, K., IACCARINO, G. & XIAO, H. 2019 Turbulence modeling in the age of data. *Annu. Rev. Fluid Mech.* **51**, 357–377.
- EDELING, W. N., CINNELLA, P., DWIGHT, R. P. & BIJL, H. 2014 Bayesian estimates of parameter variability in the k - ϵ turbulence model. *J. Comput. Phys.* **258**, 73–94.
- FERREIRA, C. 2001 Algorithm for solving gene expression programming: a new adaptive problems. *Complex Syst.* **13** (2), 87–129.
- HUNT, J. C. R. & SAVILL, A. M. 2005 Guidelines and criteria for the use of turbulence models in complex flows. In *Prediction of Turbulent Flows*, pp. 291–343. Cambridge University Press.
- LANGTRY, R. B. & MENTER, F. R. 2009 Correlation-based transition modeling for unstructured parallelized computational fluid dynamics codes. *AIAA J.* **47** (12), 2894–2906.
- LING, J., KURZAWSKI, A. & TEMPLETON, J. 2016 Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. *J. Fluid Mech.* **807**, 155–166.
- MICHELASSI, V., CHEN, L., PICHLER, R. & SANDBERG, R. D. 2015 Compressible direct numerical simulation of low-pressure turbines—Part II: Effect of inflow disturbances. *J. Turbomach.* **137** (7), 071005.
- PARISH, E. J. & DURAISAMY, K. 2016 A paradigm for data-driven predictive modeling using field inversion and machine learning. *J. Comput. Phys.* **305**, 758–774.
- PARNEIX, S., LAURENCE, D. & DURBIN, P. 1996 Second moment closure analysis of the backstep flow database. In *Proceedings of the Summer Program*, pp. 47–62. Center for Turbulence Research, Stanford University.
- PICHLER, R., SANDBERG, R. D., LASKOWSKI, G. & MICHELASSI, V. 2017 High-fidelity simulations of a linear HPT vane cascade subject to varying inlet turbulence. *ASME Paper GT2017-63079*.
- PICHLER, R., SANDBERG, R. D., MICHELASSI, V. & BHASKARAN, R. 2016 Investigation of the accuracy of RANS models to predict the flow through a low-pressure turbine. *J. Turbomach.* **138** (12), 121009.
- POPE, S. B. 1975 A more general effective-viscosity hypothesis. *J. Fluid Mech.* **72** (2), 331–340.
- POROSEVA, S. V., COLMENARES, J. D. & MURMAN, S. M. 2016 On the accuracy of RANS simulations with DNS data. *Phys. Fluids* **28** (11), 115102.
- SANDBERG, R. D., MICHELASSI, V., PICHLER, R., CHEN, L. & JOHNSTONE, R. 2015 Compressible direct numerical simulation of low-pressure turbines Part I: Methodology. *Journal of Turbomachinery* **137** (5), 051011.
- STADTMÜLLER, P. & FOTTNER, L. 2001 A test case for the numerical investigation of wake passing effects on a highly loaded LP turbine cascade blade. *ASME Paper 2001-GT-0311*.
- THOMPSON, R. L., SAMPAIO, L. E. B., DE BRAGANÇA ALVES, F. A. V., THAIS, L. & MOMPEAN, G. 2016 A methodology to evaluate statistical errors in DNS data of plane channel flows. *Comput. Fluids* **130**, 1–7.
- WANG, J.-X., WU, J.-L. & XIAO, H. 2017 Physics-informed machine learning approach for reconstructing reynolds stress modeling discrepancies based on DNS data. *Phys. Rev. Fluids* **2** (3), 034603.
- WEATHERITT, J., PICHLER, R., SANDBERG, R. D., LASKOWSKI, G. & MICHELASSI, V. 2017 Machine learning for turbulence model development using a high-fidelity hpt cascade simulation. *ASME Paper GT2017-63497*.
- WEATHERITT, J. & SANDBERG, R. D. 2016 A novel evolutionary algorithm applied to algebraic modifications of the rans stress-strain relationship. *J. Comput. Phys.* **325**, 22–37.
- WEATHERITT, J. & SANDBERG, R. D. 2017 The development of algebraic stress models using a novel evolutionary algorithm. *Int. J. Heat Fluid Fl.* **68**, 298–318.
- WHEELER, A. P. S., SANDBERG, R. D. AND SANDHAM, N. D., PICHLER, R. & MICHELASSI, V. 2016 Direct Numerical Simulations of a High-Pressure Turbine Vane. *J. Turbomach.* **138** (7), 071003.
- ZHANG, Z. J. & DURAISAMY, K. 2015 Machine learning methods for data-driven turbulence modeling. In *AIAA Computational Fluid Dynamics Conf.*, pp. 2015–2460.