TOWARDS A MACHINE LEARNING MODEL FOR EXPLICIT ALGEBRAIC REYNOLDS STRESS MODELLING USING MULTI-EXPRESSION PROGRAMMING

A. Miró[1], *S. Wallin*[2], *A. Colombo*[3], *L. Temmerman*[4], *D. Wunsch*[4], *O. Lehmkuhl*[1]

[1] *Barcelona Supercomputing Center, Barcelona, Spain*
[2] *FLOW Turbulence Lab, Eng. Mech., KTH, Stockholm, Sweden and ERCOFTAC*
[3] *Dip. di Ing. e Scienze Applicate, Università degli Studi di Bergamo, Italy*
[4] *Cadence Design Systems Belgium, Brussels, Belgium*

arnau.mirojane@bsc.es

Abstract

A novel framework for a 5-term explicit algebraic Reynolds stress model has been proposed, which is suitable for machine learning procedures. Under this framework, high-fidelity datasets have been explored and post-processed. Through multi-expression programming, models have been obtained using these data. They have been thoroughly tested and present all the desirable features of a model. Implementation of these models in a numerical code has also been possible. Results have shown that they present a small upgrade over classical models for certain kinds of flows, thus showing that the physics have been captured well. It is proved that the present workflow can obtain valid expressions with similar to better performance than the baseline model.

1 Introduction

Traditionally, RANS turbulence modelling is about predicting/estimating the Reynolds stresses from known states of the flow. While eddy-viscosity two-equation Reynolds averaged turbulence models are still highly popular in industrial applications, Reynolds stress models, such as the explicit algebraic Reynolds stress models (EARSM), have the potential for dealing with complex flows and geometries as pointed by Wallin and Johansson (2000). These models, which are derived from Differential Reynolds Stress Models, replace the eddy-viscosity relation in the two-equation models, however, they present non-trivial complexities when forming algebraic approximations of the anisotropy transport equations.

The framework for Machine-Learning (ML) aided model development from a turbulence modeller point of view would consist of constitutive relations and model equations with explicit terms representing the different physical aspects of turbulent flow modelling. The framework would comply with physical and mathematical constraints such as dimensionality, invariance, realizability and "known" fix point behaviour avoiding case specific measures utilising the essence of Human-Learned (HL) knowledge. From a ML perspective, the framework might be very different. Here, all available data in the solution field could be potential input to the learning process. Several previous proposals in this direction, Weatheritt and Sandberg (2016, 2018); Weatheritt et al. (2020) to mention a few, have based their expression on the strain- and rotation rate tensors, normalized by the turbulence time scale $\tau = K/\varepsilon$. Under this formulation, the tensors and their invariants are all dependent on the dissipation rate ε, and are not limited in the rapid limit when normalized by τ. Hence, when introduced as a framework for ML, the optimization problem is not well posed and realizability of the anisotropy cannot easily be preserved. Other works have focused in using "black-box" methodologies such as artificial neural networks (ANNs) as in Xie et al. (2021), physics-informed neural networks (PINNs) as in Wang et al. (2017) or random forest (RF) as in Kaandorp and Dwight (2020) as closure of the nonlinear terms in the Reynolds Averaged equations.

The present work results from a joint effort carried out within the European Project HiFi-TURB. It proposes a framework for ML based on a five-term EARSM. This choice is motivated by the need to work with an existing mathematical and physical framework that has sufficient degree of freedom and flexibility for model improvements. In this study, the aim is for a general model capable of capturing the essence of the turbulence statistics from the flow conditions only, local as well as non-local. Hence, general data of sufficient complexity is used and variety for training models applicable for different turbulent flows. The generated models are not formally EARSMs and should be seen as non-linear eddy-viscosity relations, or NL-EVMs. For the sake of simplicity, we shall label these models EARSMs in this work.

The approach that we followed is to enable all practically possible degrees of freedom in the model formulation restricted by physical and mathematical constraints as well as physical intuition. The ML ma-

chinery is not be used for producing the "optimal" model out of the box. In our approach, ML is used to find recalibrations and new terms with strong correlation with observations, which are further studied, refined and explained for model improvements. Using this approach, a working model has been obtained through multi-expression programming (MEP), proposed by Oltean and Groşan (2003), and using high-fidelity simulations as training data.

2 Methodology

The starting point of the use of ML for data driven turbulence modelling is the existing EARSM framework firstly introduced by Pope (1975) and later used by Wallin and Johansson (2000). This framework has a number of associated problems when used as a ML framework. For this reason, a novel EARSM framework is presented mitigating or eliminating most of these problems. This novel framework should be used together with the Menter (1994) BSL $K-\omega$ model for the scale-determining quantities.

The novel EARSM formulation

We propose to replace $\tau = K/\varepsilon$, as the normalization of the classical EARSM, by s, the L2 norm of the velocity gradient (in plane shear $s = dU/dy$)

$$s^2 = ||\Delta U||^2 = II_S - II_\Omega, \tag{1}$$

thus, the strain- and rotation rate tensors are normalized with s. Then, $II'_S - II'_\Omega = 1$ and r is introduced so that

$$II'_S = \frac{II_S}{s^2} = 1 - r, \;\; II'_\Omega = \frac{II_\Omega}{s^2} = -r, \tag{2}$$

where prime denotes normalized quantities. Now, r is limited between 0 and 1 so that $r = 1/2$ is for parallel shear, $r = 0$ for irrotational straining and $r = 1$ for pure (solid body) rotation. All these four invariants, including the higher order ones ($III'_S = III_S/s^3$, $IV' = IV/s^3$, $V' = V/s^4$ are limited between -1 and 1. Moreover, they are independent of the turbulence scales and can be considered as *structure parameters*.

The turbulence scales are used only for the equilibrium parameter $\sigma = sK/\varepsilon$. It measures the state of equilibrium by relating the turbulence to mean-flow time scales. Flow in equilibrium has σ of order unity and the rapid distortion limit is characterized by $\sigma \to \infty$. Hence, any dependency on σ must approach a constant for $\sigma \to \infty$. Additionally, some extra invariants are introduced to account for non-local effects. These are

$$\sigma_{vk} = \frac{L_{vk}s}{\sqrt{K}}, \;\; L_{vk} = \frac{2\kappa|S_{ij}|}{|\nabla U_k|}, \tag{3}$$

$$g_T = \frac{Re_T}{10 + Re_T}, \;\; Re_T = \frac{0.09\sigma K^2}{\nu\varepsilon}. \tag{4}$$

The 5-term representation (Pope, 1975; Wallin and Johansson, 2000) has been found to be practical sufficient and not over-complex. Then, any algebraic relation of the anisotropy a_{ij} can be mapped on these 5 terms $a_{ij} = \sum_{k=0}^{5} \beta_k T_{ij}^{(k)}$, where $T_{ij}^{(k)}$ are the tensor groups. The different β_k are now functions of σ, r and the other invariants (III'_S, IV', V', σ_{vk}, g_T). The first term β_1 corresponds to the eddy-viscosity part with an effective coefficient $C_\mu^{eff} = -1/(2\sigma)\beta_1$.

Interestingly, the effective eddy viscosity is not directly dependent of ε, $\mu_t^{eff} = -0.5\rho K\beta_1/s$, however, the dependency might (and will) be introduced through β_1 being a function of σ.

The production is related to the invariants as

$$-\frac{P}{sK} = \frac{a_{ij}S_{ij}}{s} = \beta_1(1-r) + \beta_3 IV' + 2\beta_4 V'. \tag{5}$$

To obtain these functions, a regularization problem is proposed as

$$\beta_k = B_{kl}^{-1} a_{ij} T_{ji}^{(l)}; \quad B_{kl} = T_{ij}^{(k)} T_{ji}^{(l)} + \lambda\delta_{kl}, \tag{6}$$

where δ_{ij} denotes the Dirac delta, while a_{ij}, $T_{ij}^{(k)}$ and the rest of the parameters are obtained from the datasets. A good compromise for the regularization parameter λ has been found to be 0.01.

The role of the ML procedure is then to minimize the error ($\epsilon = ||\beta_k - \hat{\beta}_k||$) between the dataset and the modelled coefficient $\hat{\beta}$ as a function of the aforementioned invariants.

The $K-\omega$ framework

The constitutive relation from EARSM should be used together with the Menter BSL model, as proposed in Menter et al. (2012). The ML procedure is included in the computation of the momentum diffusion and the production terms of the K and ω equations. In addition, the turbulence equations need to be balanced by introducing residuals defined as

$$\rho R_K = P_K - \rho\varepsilon + \frac{\partial}{\partial x_k}\left(\mu + \sigma_K \mu_t \frac{\partial K}{\partial x_k}\right) - \frac{DK}{Dt}, \tag{7}$$

$$\rho R_\omega = P_\omega - \beta\rho\omega^2 + \frac{\partial}{\partial x_k}\left(\mu + \sigma_\omega \mu_t \frac{\partial \omega}{\partial x_k}\right) + \frac{\sigma_d \rho}{\omega}\frac{\partial K}{\partial x_k}\frac{\partial \omega}{\partial x_k} - \frac{D\rho\omega}{Dt}, \tag{8}$$

which are normalized as

$$r_K = \frac{R_K}{\max(\varepsilon_{cut}, \varepsilon)}; \quad r_\omega = \left(\frac{K}{\max(\varepsilon_{cut}, \varepsilon)}\right)^2 R_\omega. \tag{9}$$

where ε_{cut} defines a threshold for the dissipation to avoid zero division. Then, the ML procedure becomes minimizing the errors $\epsilon_K = [|r_K - \hat{r}_k|, \lambda\hat{r}_k]$, $\epsilon_\omega = [|r_\omega - \hat{r}_\omega|, \lambda\hat{r}_\omega]$ with $\lambda = 0.01$ as a regularization parameter so to avoid a null solution. Here, two additional invariants are considered

$$d_K = \frac{1}{\varepsilon}\frac{\partial}{\partial x_k}\left(\frac{K}{\omega}\frac{\partial K}{\partial x_k}\right), \tag{10}$$

$$d_\omega = \left(\frac{K}{\varepsilon}\right)^2 \frac{\partial}{\partial x_k}\left(\frac{K}{\omega}\frac{\partial \omega}{\partial x_k}\right). \tag{11}$$

The rest of the procedure is analogous to that of the EARSM.

Datasets

The aforementioned parameters are computed for three datasets: a DNS of the Stanford diffuser, a LES of a 2-D curved backward-facing step and synthetic data generated by applying the standard EARSM. The "smart" combination of these three datasets yields a uniform exploration space for r and σ, the main invariants of this novel formulation.

Synthetic EARSM data. This dataset corresponds to the evaluation of the standard 3D EARSM model of Wallin and Johansson (2000). Since the model is analytical it can be evaluated for a range of uniform random distribution of r and σ to recover the coefficients β_1 to β_5. While this dataset does not bring any new model improvement from the ML procedure, it is however useful to complement the data gaps of the two other datasets. Moreover, it gives a first guess for the ML procedure.

2-D curved backward-facing step. This dataset is freely available and comes from a LES of a 2-D separating flow over a curved ramp run with an incompressible code (Lardeau and Leschziner, 2011; Bentaleb et al., 2012). A selection of this dataset has been done by combining the effects on the dissipation and wall distance (excluding the area where $y^+ < 30$), excluding the boundaries. On this dataset β_1 to β_3 find values while β_4 and β_5 are zero.

Stanford diffuser. This dataset is also freely available and comes from a DNS of the Stanford diffuser at $Re = 10,000$ run with an incompressible code (Ohlsson et al., 2010; Miró et al., 2023, Preprint). The dataset is processed volumetrically to obtain β_1 to β_5, and the invariants. Due to the large number of data points, data extraction has been performed by random sampling, "smart" sampling or using a variational auto-encoder (VAE). Since a huge portion of the points are in $r = 1/2$, VAE or "smart" sampling (i.e., sampling within a limited range of r) have yielded datasets with uniform r, while $0 < \sigma < 10$ has been obtained naturally.

Baseline model constraints

The definition of baseline model constraints should be based on mathematical and physical principles. Moreover, additional constraints based on empirical knowledge, such as the log-law, should be applied more restrictively and carefully, to allow for the ML process to find novel model ideas.

First, mathematical and physical strong constraints: 1) dimensionally correctness, which implies scale-invariance, 2) frame invariance and symmetries, 3) Galilean invariance, and 4) a well-posed resulting model. Moreover: a) Exclude the acceleration – (and the related pressure gradient). Pressure gradient and acceleration is connected through the momentum equation. The influence on turbulence is through straining and compression by the velocity gradient. b) Exclude case-dependent and non-local information, such as different kind of Reynolds numbers, free-stream values, reference length scales, wall skin friction, boundary layer thickness. c) Wall distance might be used. Neighbouring walls influence the local turbulence strongly. Without other more physical measures of non-locality, the inclusion of the wall distance might be useful as one measure of "non-locality" effects. d) Avoid over-complex formulations for preserving physical intuition. Also, for the sake of well-posedness and avoiding spurious multiple solutions. Complex algebra is not a problem by itself and can easily be handled by use of computer algebra.

3 Results and discussion

ML model definition

The complete model was developed using the framework and datasets previously defined. Data with $y^+ < 30$ was excluded. The MEP procedure is then used to generate a surrogate model for the set β_k as functions of r and σ. The next step is then to optimize the coefficients in this surrogate model based on the data from the 3-D diffuser, the 2-D back step as well as the artificial EARSM data. The resulting model is named SyMEP-1 and its β_k coefficients are

$$\beta_1 = -C_{\beta_{1,1}} \frac{\cosh\left(\sqrt{r^6\sigma}\right)\tanh\left(\sigma\right)}{\cosh^2\left(\sqrt{r^6\sigma}\right)}, \quad (12)$$

$$\beta_2 = -C_{\beta_{2,1}}\left(C_{\beta_{2,2}} - C_{\beta_{2,3}} e^{\left(-C_{\beta_{2,4}}\sigma\right)}\right)^2 + \frac{\left[-C_{\beta_{2,5}} r + C_{\beta_{2,6}}\left(C_{\beta_{2,7}} r - C_{\beta_{2,8}}\right)^2 + C_{\beta_{2,9}}\right]\sigma^2}{C_{\beta_{2,a}}\sigma^2 + 1}, \quad (13)$$

$$\beta_3 = -C_{\beta_{3,1}} IV\sigma \tanh\left(r\right)\tanh\left(r\sigma - \sigma\tanh\left(r\right)\right), \quad (14)$$

$$\beta_4 = -\frac{1}{2}\sinh\left[\cosh\left(\tanh\left(\sigma\right)\right) - r^{\left(r^2\right)^\sigma}\right], \quad (15)$$

$$\beta_5 = C_{\beta_{5,1}}\left[r(r\sqrt{log(\sigma)} + log(\sigma))\right]^{r^{-C_{\beta_{5,2}} r\sqrt{\log(\sigma)}}}, \quad (16)$$

where implicitly the following safe operations are in place: $(\cdot)^x = (\max(10^{-5}, \cdot))^x$; $\sqrt{\cdot} = \sqrt{\max(0, \cdot)}$; $\cdot/\cdot = \cdot/\max(10^{-5}, \cdot)$; $\log(\cdot) = \log(\max(10^{-5}, \cdot))$.

The coefficient $C_{\beta_{1,1}}$ is directly related to the effective C_μ in plane flow and, hence, the flat plate skin friction magnitude. Flat-plate with zero pressure gradient (ZPG) a-posteriori tests revealed that the flat-plate skin friction was under-predicted and, hence, $C_{\beta_{1,1}}$ was manually tuned to $C_{\beta_{1,1}} = C_{\beta_{1,1-\text{tuned}}} = 0.63$, as seen in Figure (2). This is the value to be used for the definition of SyMEP-1. The original value from the MEP optimization was $C_{\beta_{1,1}} = C_{\beta_{1,1-\text{orig}}} = 0.586964$. The rest of the coefficients are: $C_{\beta_{2,1}} = 0.526566$, $C_{\beta_{2,2}} = 0.891788$, $C_{\beta_{2,3}} = 0.880494$, $C_{\beta_{2,4}} = 0.661698$, $C_{\beta_{2,5}} = 0.10875$, $C_{\beta_{2,6}} = 0.0813025$, $C_{\beta_{2,7}} = 1.83837$, $C_{\beta_{2,8}} = 0.880494$, $C_{\beta_{2,9}} = 0.008045$, $C_{\beta_{2,a}} = 0.248196$, $C_{\beta_{3,1}} = 5.45337$, $C_{\beta_{5,1}} = 0.359138$, $C_{\beta_{5,2}} = 0.870968$.

Figure (1) shows that the effective C_μ in equilibrium ($P = \varepsilon$) for the original and tuned version are around the expected value of 0.09, while other models in the literature do not respect this constraint.

The last piece of modelling is to introduce the correction of the $K - \omega$ model equations. The residual terms r_K and r_ω were extracted from the curved backward-facing step data and used for the ML training. The reason is that the curved backward-facing contains a clear developing shear-layer where there is a transition in length scale that most RANS models cannot correctly capture. Hence, the von Kármán length scale is introduced through σ_{vk} as a parameter. Their mathematical expressions read

$$r_K = \tanh\left[\left(\sigma_{vk}^{d_K - \sigma_{vk}} - 1\right)\tanh\left(d_K\right)\right] - 0.0588804, \quad (17)$$

$$r_\omega = d_\omega r + 2\sigma + \sigma_{vk} - 11.7919. \quad (18)$$

Note that the r_K and r_ω corrections are functions of the diffusion terms as well. The linear term $(d_\omega r)$ in r_ω can ap-

proximately be included into the diffusion of ω, but the corresponding term in r_K is strongly non-linear and not easily approximated.

The corrected $K-\omega$ model equation are then solved with the SyMEP-1 constitutive relation with one important difference. After testing and validation for flat plate boundary layers (BLs) it was clear that the original value of $C_{\beta_{1,1}}$ could be used together with the added residual terms and that the previous recalibration was not consistent.

Further stability problems and delayed numerical transition to turbulence for the flat plate BL related to the large variation of the von Kármán parameter σ_{vk} in the residual terms. Hence, a limiting function is introduced as

$$\sigma^*_{vk} = c\tanh\left(\frac{\sigma_{vk}}{c}\right), \tag{19}$$

which limits σ_{vk} to $\pm c$ and approximately keep the linear relation in between. The coefficient c was tuned to be 1.5, as shown for the friction coefficient on the flat plate in Figure (2) (top).

The final definition of the SyMEP-1 model is then concluded as: 1) Use the SyMEP-1 relation in Equations (12) and (16) for the β coefficients. If using only these terms use $C_{\beta_{1,1}} = C_{\beta_{1,1-tuned}} = 0.63$. 2) For SyMEP-1-Res, use the residual terms in Equations (17) and (18) to correct the Menter BSL $K-\omega$ model equation. In that case, use $C_{\beta_{1,1}} = C_{\beta_{1,1-orig}} = 0.586964$ and correct σ_{vk} using Equation (19) with $c = 1.5$.

Verification and validation

Further validation of the ML model has been performed by implementing it in two different solvers: 1) $Fidelity^{TM}DBS$, a density-based unstructured finite-volume (FV) solver with multi-grid acceleration and local-time stepping. Low-Mach number and incompressible flows are treated with preconditioning, using a second-order scheme with scalar dissipation. 2) $MIGALE$, a Discontinuous Galerkin (DG) solver, where SyMEP-1 is coupled with the Wilcox (2006) $K-\omega$ model instead of the BSL, cf. Bassi et al. (2014).

The following three cases have been considered for validation: 1) flow on a flat plate; 2) 2-D curved backward-facing step; 3) Stanford diffuser.

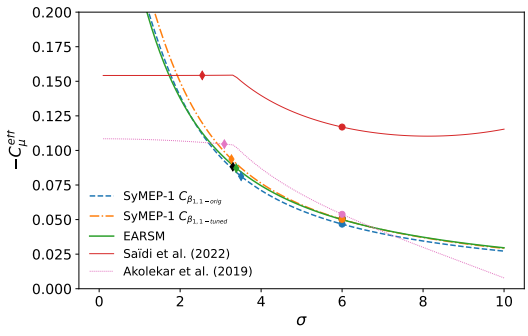

Figure 1: Effective C_μ vs. σ in parallel shear ($r = 1/2$) for the original and tuned SyMEP-1 models compared with Akolekar et al. (2019), Saïdi et al. (2022) and experimental results (in black) from Weighardt and Tillmann (1951). Diamonds are the log (equilibrium $P = \varepsilon$) solutions, while circles are the shear layer solutions.

Flat plate. A fully turbulent flat plate has been defined with reference data from Weighardt and Tillmann (1951) and Klebanoff (1955) ($L = 4m$, $U_\infty = 68.79m/s$, $\nu = 3\times 10^{-5}m^2/s$) treated as weakly compressible.

The introduction of SyMEP-1 revealed an under-predicted skin friction without considering recalibration. The residual model shows a somewhat different behavior than the previous models and reference models. The initial part of the flat plate is better captured in terms of C_f, but the mean velocity somewhat under-predicted up to $y^+ \approx 100$, as seen in Figure (2).

2-D curved backward-facing step. This verification case of Bentaleb et al. (2012) is the same data in which the model was trained. The Reynolds number based in the step height and free-stream velocity is 13,700. Inlet conditions are set by imposing profiles for U_x, U_y, K and ε upstream of the step.

Table (1) shows the predicted separation and attachment positions compared with the LES of Bentaleb et al. (2012).

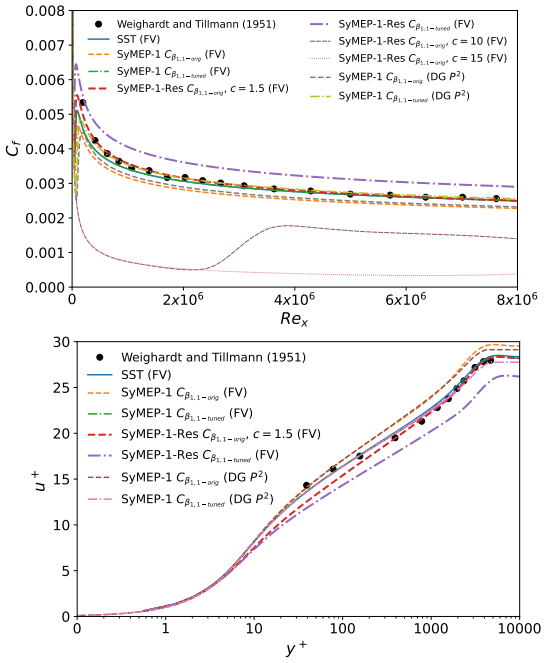

Figure 2: Flat-plate (ZPG) skin friction (top) and mean velocity (bottom) at $x = 3.487$ for the original and tuned SyMEP-1 models.

Table 1: 2-D curved backward-facing step predicted separation and attachment postions for the reference data, S-BSL EARSM and the ML models.

	$(x/H)_{sep}$	$(x/H)_{reat}$
Bentaleb et al. (2012)	0.83	4.36
S-BSL EARSM	1.24	5.0
SyMEP-1	0.95	5.02
SyMEP-1-Res	1.07	4.74

Both models capture the skin friction very similar to the reference model BSL-EARSM. The most complete model, SyMEP-1-Res, captures the skin friction slightly better than the reference model S-BSL EARSM though. In particular, the size of the separation bubble is slightly less overpredicted by SyMEP-1-Res.

Marginal improvements are observed with the SyMEP-1-Res model, which is not able to capture the rapid growth of the Reynolds stress in the initial shear layer. The added residual term is not sufficient, although there is some small tendency in the right direction seen in Figure (3). This might be due to the quite aggressive limiting function for the von Kármán parameter σ_{vk}. Less limitation resulted in delayed numerical transition to turbulence for the flat plate and convergence problems.

The convergence of the ML model is slower than for the reference S-BSL EARSM, but still acceptable. The additional residual terms do not significantly influence the convergence although they are quite complex and non-linear.

Stanford diffuser. This verification case is based on the geometry experimentally studied in Cherry et al. (2008). High-fidelity data was later computed in Miró et al. (2023, Preprint) and used for training this ML model. The case is set so that $Re = 10,000$ based on inlet bulk velocity.

The results for the pressure coefficient are seen in Figure (4) show that the ML models delivered results better

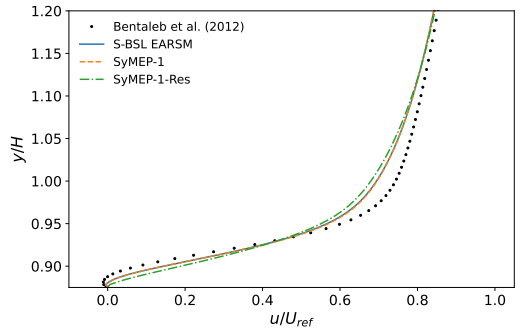

Figure 3: Back step velocity profiles for SyMEP-1 and SyMEP-1-Res compared with S-BSL EARSM and LES from Bentaleb et al. (2012).

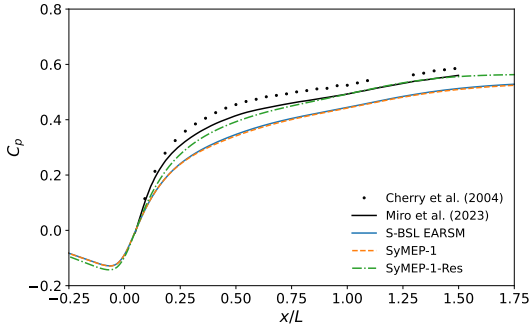

Figure 4: Distribution of the pressure coefficient along the bottom wall for the 3D Stanford 1 for SyMEP-1 and SyMEP-1-Res.

or equivalent to results obtained with current state-of-the-art models. Moreover, the application of the new model on the slightly different geometry of the Stanford diffuser 2 (not shown) gives an indication that the new EARSM models can be ported on different geometries, albeit close ones, in the present case.

4 Conclusions

When deriving the methodologies for data-driven RANS modelling the aim is focused on models of general validity, not specific models for a limited class of problems. This is a big challenge and not usually the aim for data-driven approaches. Moreover, the search for mathematical expressions that can be further analyzed for improved understanding and, more importantly, be implemented in general-purpose CFD tools lead us to the use of MEP.

The novel EARSM framework is aiding in better controllability of specific features of turbulent flows that are more directly related to the flow physics. The resulting ML models performed very well in parallel flows with two major achievements concerning the first requirement: a) The model captures the established log layer solution ($P = \varepsilon$) and, more noticeable, it scales as expected for very high Reynolds numbers, many orders of magnitudes larger than the flow considered, with a von Kármán constant $\kappa \approx 0.38$, which is close to established values. b) The model did also capture the quite different state of developed free shear layer, which is far from equilibrium ($P \approx 1.8\varepsilon$). A standard EVM with constant C_μ will fail, but this state will be captured by full DRSMs or consistent EARSMs. Also Menter SST $K-\omega$ is consistent with this state by a semi-empirical approach. These two achievements must be considered as extremely remarkable since these states are not explicitly introduced in any sense in the procedure. It is believed that the success is strongly linked to the particular new EARSM framework, the ML problem definition and the effective MEP algorithm.

When analysing the data in terms of the newly introduced more physical invariants/parameters, it became clear that the parameter space is poorly covered from data. Most of the data is almost parallel ($r \approx 0.5$) and locally 2D (3D terms are small). This might rise a question, though. How can we avoid bias from old traditional modelling when we are searching for novel correlations and model terms? A good answer is still not available, nevertheless, it is a fact that the resulting SyMEP-1 performs very similar as the baseline EARSM.

For the final model, SyMEP-1-Res, residual terms were added for balancing the $K-\omega$ model equations from data. The idea was to better capture the length scale in the initial free shear layer for separated flows resulting in the typical underprediction of the Reynolds stress and overpredicted separation length. The MEP procedure found dependencies on the von Kármán length scale σ_{vk}, which is a novel ingredient to the baseline modelling. The residual terms needed quite substantial limitation for convergence to a physical solution which might degenerate the expected behavior. In further work, the residual terms in the $K-\omega$ model equations should be complemented with (or replaced by) turbulence transport terms for better capturing the missing physics.

Acknowledgments

The research leading to this work has received funding within the HiFi-TURB project from the European Union's Horizon2020 research and innovation programme under

grant agreement No 814837. Oriol Lehmkuhl has been partially supported by a Ramon y Cajal postdoctoral contract (Ref: RYC2018-025949-I). We also acknowledge the Barcelona Supercomputing Center for awarding us access to the MareNostrum IV machine based in Barcelona, Spain.

References

Akolekar, H. D., Weatheritt, J., Hutchins, N., Sandberg, R. D., Laskowski, G. and Michelassi, V. (2019), Development and use of machine-learnt algebraic Reynolds stress models for enhanced prediction of wake mixing in low-pressure turbines, *J. Turbomach.*, Vol. 141 No. 4, pp. 041010.

Bassi, F., Botti, L., Colombo, A., Ghidoni, A. and Rebay, S. (2014), Implementation of an Explicit Algebraic Reynolds Stress Model in an Implicit Very High-Order Discontinuous Galerkin Solver, *in* 'Spectral and High Order Methods for Partial Differential Equations - ICOSAHOM 2012', pp. 111–123.

Bentaleb, Y., Lardeau, S. and Leschziner, M. A. (2012), Large-eddy simulation of turbulent boundary layer separation from a rounded step, *J. Turbul.*, pp. 1–28.

Cherry, E., Elkins, C. and Eaton, J. (2008), Geometric sensitivity of three-dimensional separated flows, *Int. J. Heat Fluid Fl.*, Vol. 29, pp. 803–811.

Kaandorp, M. L. and Dwight, R. P. (2020), Data-driven modelling of the Reynolds stress tensor using random forests with invariance, *Comput. Fluids*, Vol. 202, pp. 104497.

Klebanoff, P. S. (1955), Characteristics of turbulence in boundary layer with zero pressure gradient, *in* 'NACA-TR-1247'.

Lardeau, S. and Leschziner, M. (2011), The interaction of round synthetic jets with a turbulent boundary layer separating from a rounded ramp, *J. Fluid Mech.*, Vol. 683, pp. 172–211.

Menter, F. R. (1994), Two-equation eddy-viscosity turbulence models for engineering applications, *AIAA J.*, Vol. 32, pp. 1598–1605.

Menter, F. R., Garbaruk, A. V. and Egorov, Y. (2012), Explicit algebraic reynolds stress models for anisotropic wall-bounded flows, *in* 'Progress in Flight Physics', Vol. 3, pp. 89–104.

Miró, A., Eiximeno, B., Rodríguez, I. and Lehmkuhl, O. (2023, Preprint), DNS and POD analysis of separated flow in a three-dimensional diffuser, *Flow Turbul. Combust.*, Vol. -, pp. –.

Ohlsson, J., Schlatter, P., Fischer, P. and Henningson, D. (2010), Direct numerical simulation of separated flow in a three-dimensional diffuser, *J. Fluid Mech.*, Vol. 650, pp. 307–318.

Oltean, M. and Groşan, C. (2003), Evolving Evolutionary Algorithms Using Multi Expression Programming, *in* 'Advances in Artificial Life', pp. 651–658.

Pope, S. B. (1975), A more general effective-viscosity hypothesis, *J. Fluid Mech.*, Vol. 72, pp. 331–340.

Saïdi, I. B. H., Schmelzer, M., Cinnella, P. and Grasso, F. (2022), CFD-driven symbolic identification of algebraic Reynolds-stress models, *J. Comput. Phys.*, Vol. 457, pp. 111037.

Wallin, S. and Johansson, A. V. (2000), An explicit algebraic Reynolds stress model for incompressible and compressible turbulent flows, *J. Fluid Mech.*, Vol. 403, pp. 89–132.

Wang, J.-X., Wu, J.-L. and Xiao, H. (2017), Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data, *Phys. Rev. Fluids*, Vol. 2, pp. 034603.

Weatheritt, J. and Sandberg, R. (2016), A novel evolutionary algorithm applied to algebraic modifications of the RANS stress–strain relationship, *J. Comput. Phys.*, Vol. 325, pp. 22–37.

Weatheritt, J. and Sandberg, R. D. (2018), Improved Junction Body Flow Modeling Through Symbolic Regression, *in* '32nd Symposium on Naval Hydrodynamics', p. 15.

Weatheritt, J., Zhao, Y., Sandberg, R. D., Mizukami, S. and Tanimoto, K. (2020), Data-driven scalar-flux model development with application to jet in cross flow, *Int. J. Heat Mass Tran.*, Vol. 147, pp. 118931.

Weighardt, K. and Tillmann, W. (1951), On the turbulent friction layer for rising pressure, *in* 'NACA-TM-1314'.

Wilcox, D. C. (2006), *Turbulence modeling for CFD*, Vol. 3, DCW industries La Canada, CA.

Xie, C., Xiong, X. and Wang, J. (2021), Artificial neural network approach for turbulence models: A local framework, *Phys. Rev. Fluids*, Vol. 6, pp. 084612.