Data-driven RANS closures for three-dimensional flows around bluff bodies

Huijing, Jasper P.; Dwight, Richard P.; Schmelzer, Martin

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Contents lists available at ScienceDirect

Computers and Fluids

Data-driven RANS closures for three-dimensional flows around bluff bodies

Jasper P. Huijing, Richard P. Dwight*, Martin Schmelzer

Aerodynamics Group, Faculty of Aerospace Engineering, TU Delft. Kluyverweg 2, 2629HT Delft, The Netherlands

ARTICLE INFO

ABSTRACT

In this short note we apply the recently proposed data-driven RANS closure modelling framework of *Schmelzer* et al.*(2020)* to fully three-dimensional, high Reynolds number flows: namely wall-mounted cubes and cuboids at $Re = 40,000$, and a cylinder at $Re = 140,000$. For each flow, a new RANS closure is generated using sparse symbolic regression based on LES or DES reference data. This new model is implemented in a CFD solver, and subsequently applied to prediction of the other flows. We see consistent improvements compared to the baseline $k - \omega$ SST model in predictions of mean-velocity in complete flow domain.

1. Introduction

Reynolds averaged Navier-Stokes (RANS) models are notoriously inaccurate in the presence of massive flow separation, for example in the wake of bluff bodies. The steady RANS paradigm – of modelling fluctuations at all scales – is fundamentally ill-suited to representing the complex unsteady dynamics in the wake of e.g. a cylinder or wall-mounted cube. Nonetheless, in a wide variety of industrial applications (notably the automotive industry) it would be extremely valuable to have access to RANS closures that give reasonably accurate predictions in wakes.

Data-driven turbulence modelling uses data from high-fidelity simulations (LES, DNS) or experiments to semi-automatically derive new closure models, see the surveys [1,2]. The methods of supervised machine-learning are used to represent and fit models, e.g. neural networks [3], random-forests [4–6], gene-expression programming [7], and sparse symbolic regression [8,9]. The latter two methods have the advantage of generating concise expressions for closure models, which can be inspected, analysed and implemented easily.

In this work we apply the Sparse Regression of Turbulence Anisotropy (SpaRTA) framework first developed by Schmelzer et al. [8]. The models produced are explicit algebraic Reynolds-stress models, based on $k - \epsilon$ or $k - \omega$ closures, but correcting both the turbulence anisotropy, and the t.k.e. production. The required correction fields are solved for by injecting DES, LES or DNS data into the RANS equations; and a model is obtained by regressing these corrections against mean-flow quantities available to RANS. This model can then be applied to predict a flow for which no reference data is available.

Though the objective of predicting bluff body flows with RANS might seem optimistic, SpaRTA has already been demonstrated with success for flows with significant separation [10]. Also note: our intention here is not to derive new *general purpose* closure models. At a minimum that would require a more diverse set of training flows. Rather, we wish to demonstrate that our framework has the capability of constructing RANS closures that generalize acceptably for massively detached flows. As such they may be useful as components of larger, general purpose models.

While using the framework of [8] and [10], this work extends those articles in several ways. We consider flows at significantly higher Reynolds numbers ($Re = 140,000$ vs $Re \simeq 10,000$ in [8]); in 3D vs only 2D previously; and we use DES data for the first time. The 3D LES data source means that optimization of the symbolic regression for large data-sets is required, and we introduce a practical technique for library reduction.

2. Methodology: SpaRTA

The objective is to generate a RANS closure from reference data, that improves predictions for some class of flows. For extended details of our methodology, see [8]. In brief: the incompressible RANS $k - \omega$ SST equations are modified with correction terms b^{Δ} and R

---

* Corresponding author.
  E-mail address: r.p.dwight@tudelft.nl (R.P. Dwight).

to:

$$U_j \partial_j U_i = \partial_j \left[-\frac{1}{\rho} P + \nu \partial_j U_i + \nu_T \partial_j U_i - 2k b_{ij}^{\Delta}, \right],$$

$$U_j \partial_j k = P_k + R - \beta^* \omega k + \partial_j \left[(\nu + \sigma_k \nu_t) \partial_j k \right],$$

$$U_j \partial_j \omega = \frac{\gamma}{\nu_t} (P_k + R) - \beta \omega^2 + \partial_j \left[(\nu + \sigma_\omega \nu_t) \partial_j \omega \right]$$
$$+ CD_{k\omega}, \tag{1}$$

with ρ the fluid density, U, P the mean velocity and generalized pressure, ν, $\nu_T = \nu_T(k, \omega)$ the molecular- and turbulence viscosity, k, $P_k = -2k(b_{ij} + b_{ij}^{\Delta}) \partial_j U_i$ the turbulence kinetic energy and its production rate, and ω the specific dissipation rate. See [11] for details of remaining terms and coefficients. The baseline SST model uses the Boussinesq assumption as a fundamental modelling premise, namely

$$b_{ij} \simeq -\frac{\nu_t}{k} S_{ij}, \quad S_{ij} := \frac{1}{2} (\partial_i U_j + \partial_j U_i).$$

where b_{ij} is the normalized turbulence anisotropy. The purpose of b^{Δ} is to relax this assumption by allowing for deviations from Boussinesq. Similarly the SST k-equation is a model for the true t.k.e. equation; and the new term R is placed to allow for modelling errors here. In practice R allows for control of turbulence intensity, and b^{Δ} for control of turbulence anisotropy.

Solving for corrective fields Given full-field LES or DES data for flow A, our preliminary objective is to find corrective fields $R(\boldsymbol{x})$, $b^{\Delta}(\boldsymbol{x})$ – i.e. as functions of the spatial coordinates $\boldsymbol{x} := (x, y, z)$ – such that when the system (1) is solved for the same flow A, the resulting U and k correspond well to LES/DES mean values. This is achieved by injecting frozen LES/DES quantities into (1), and solving for the remaining unknowns R, b^{Δ} and ω. This procedure is named *k-corrective-frozen-RANS* [8] and can be seen as a generalization of the "frozen" method for estimating t.k.e. dissipation rate from LES data [12]. We use DES/LES estimates of the Reynolds stress tensor and t.k.e. which include contributions from both modelled and resolved stresses. Note that we deliberately conflate the k of the $k - \omega$ model, and the true t.k.e. at this step. The objective is to obtain a system of equations which is predictive of true t.k.e., at the cost of perhaps larger correction terms.

Model Regression To make predictions it is necessary to generalize these corrective fields by building a closure model. We achieve this by now finding expressions for $R(S, \Omega)$, $b^{\Delta}(S, \Omega)$ as functions of the mean strain-rate tensor S and rotation rate tensor $\Omega_{ij} := \frac{1}{2} (\partial_j U_i - \partial_i U_j)$; quantities available to the RANS solver. In particular we follow Pope's integrity basis formulation [13], which specifies that the most general functional form (under modest assumptions) for $b^{\Delta}(S_{ij}, \Omega_{ij})$ can be written:

$$b_{ij}^{\Delta}(S_{ij}, \Omega_{ij}) = \sum_{k=1}^{10} T_{ij}^{(k)} \alpha_k(\lambda_1, \dots, \lambda_5), \tag{2}$$

where $T^{(1)} := S$, $T^{(2)} := S\Omega - \Omega S$, etc. are basis tensors, λ_l are the five invariants of S and Ω [14], and $\alpha_k : \mathbb{R}^5 \to \mathbb{R}$ are ten, arbitrary scalar-valued functions. By this construction b^{Δ} is always symmetric and traceless, the map $b^{\Delta}(\cdot)$ is invariant under rotations, and Galilean invariant by virtue of the dependence only on $\partial_j U_i$. In this work we use only $T^{(1)}, \dots, T^{(4)}$ and λ_1, λ_2, thereby restricting our search to *quadratic* nonlinear eddy-viscosity models, following e.g. [15,16].

Aiming for simple algebraic expressions, we represent each $\alpha_k(\lambda_1, \lambda_2)$ as a linear combination of a large *library* of basis functions $\mathcal{L} = (\phi^1, \dots, \phi^L)$:

$$\alpha_k(\lambda_1, \lambda_2) = \sum_{l=0}^{L} c_k^l \phi^l(\lambda_1, \lambda_2), \quad k \in \{1, \dots, 4\}, \tag{3}$$

where the library is generated from products and powers of the inputs $\mathcal{L} = (1, \lambda_1, \lambda_1^2, \dots, \lambda_2, \lambda_2^2, \dots, \lambda_1 \lambda_2, \dots)$, see for example [17]. To avoid models with large numbers of terms, as well as overfitting, we apply elastic net regression [18], which encourages sparsity (most coefficients $c_k^l = 0$), to find the model form. Let $\boldsymbol{c} \in \mathbb{R}^{4 \times L}$ represent the vector of all model coefficients, then we solve

$$\min_{\boldsymbol{c}} \| b^{\Delta}(\boldsymbol{x}) - b^{\Delta}(\boldsymbol{c}; S(\boldsymbol{x}), \Omega(\boldsymbol{x})) \|^2 + \theta \rho \|\boldsymbol{c}\|_1 + \theta (1 - \rho) \|\boldsymbol{c}\|_2^2,$$
$$\tag{4}$$

where in practice the first norm is estimated using the points of the mesh used to obtain the corrective field $b^{\Delta}(\boldsymbol{x})$. The term $\|\boldsymbol{c}\|_1 := \sum |c_i|$ encourages sparsity of $\boldsymbol{c}$ and the term $\|\boldsymbol{c}\|_2^2 := \sum c_i^2$ controls the magnitude of nonzero coefficients. Both are blended by $\rho \in [0, 1]$ and $\theta \in \mathbb{R}^+$ ultimately controls the extent of regularization. Using path elastic net [19], a large number of candidate models for various ρ, θ are obtained, and a second ridge-regression step is used to select coefficients. A similar procedure is applied to model $R(S, \Omega)$: we assume the form $R = 2k \partial_j U_i \hat{R}_{ij}(S, \Omega)$, and use the base-tensor series to model $\hat{R}_{ij}$.

With the extension from 2D cases in [8] to 3D cases here the size of the data-set in (4) has increased significantly, and symbolic regression becomes a significant memory bottleneck due to storage of the library $\mathcal{L}$ evaluated on the full data-set. We introduce a *cliqueing* procedure motivated by the high multi-collinearity observed in $\mathcal{L}$. Specifically we compute the correlation coefficient between all pairs of library functions ϕ^i, and sort them into *cliques* whose correlation within a clique always exceeds a cut-off (of 0.99). Efficiently finding cliques is an established problem in graph theory [20]. We then select the algebraically simplest member of the clique to represent the clique, and discard the remainder. Although ϕ^i are nonlinear in the inputs, a linear measure of correlation is adequate, as they are combined linearly in (3). This method is reminiscent of *elite basis regression* [21], except that we are not concerned with correlation with the target, only with basis functions amongst themselves.

3. Results

We examine three flows: a wall-mounted cube in a channel (Flow A), a wall-mounted cuboid (length:width:height ratio 3:2:2) also in a channel (Flow B), and an infinite circular cylinder (Flow C). Flows A, B are at Re = 40, 000 based on bulk velocity and cube/cuboid height h, with the channel of height $2h$, and are well-used experimental [22] and numerical test-cases [23]. Flow C is at Re = 140, 000 based on cylinder diameter [24]. All three flows include massive separation, resulting in unsteady wakes with a wide range of time- and length-scales.

Ground-truth reference fields for Flows A, B are obtained with DES [25] simulating a domain of size $14.5h \times 9h \times 2h$, the obstacle centred at $x = 4h$, with periodic boundary-conditions in the cross-channel direction, and synthetic channel turbulence at the inflow plane [26], and an averaging time of $\simeq 12.4$ flow-throughs. The RANS part of the DES is based on the Spalart-Allmaras one-equation model [27], in contrast to our use of $k - \omega$ SST in our enhanced RANS.

For Flow C ground-truth comes from a wall-resolved LES. The cylinder has diameter d and simulated length $\pi \cdot d$ with periodic boundaries. The flow is fully turbulent, with synthetic turbulence at the inlet plane. For all flows, velocity profiles were compared with existing published results, and found to be sufficiently accurate for the application.

The "Cube" model We first apply the SpaRTA methodology to Flow A. Mean velocity profiles for the DES reference and baseline RANS are shown in Fig. 1. The baseline significantly overestimates the size of the recirculation region on the top and sides of the
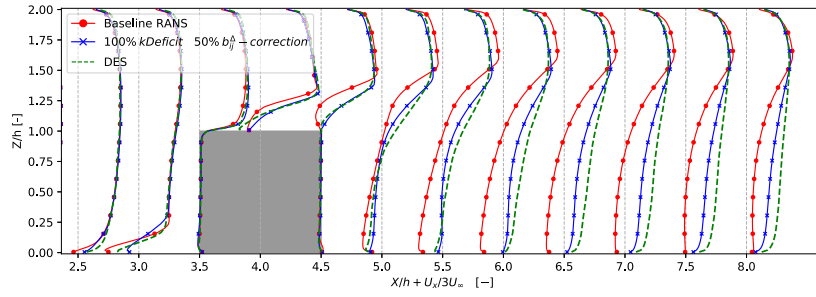
(a) Velocity (U_x) in symmetry plane.
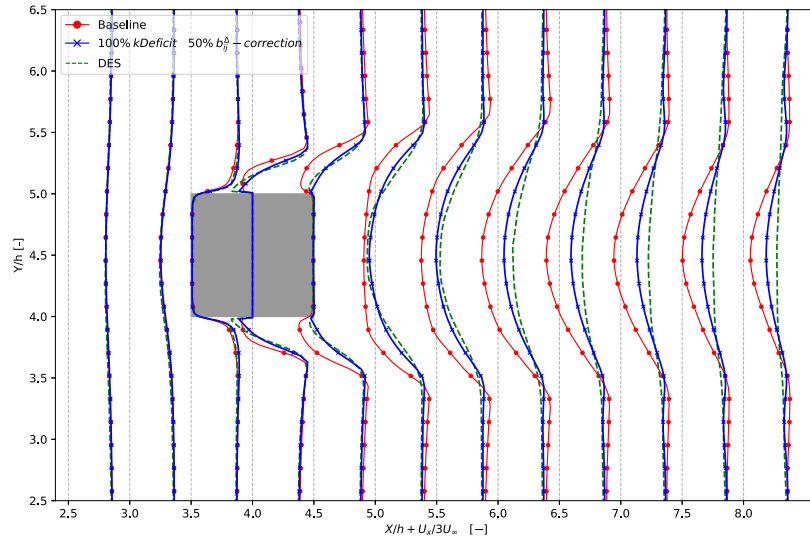
(b) Velocity (U_x) at mid-height of the cube $z = 0.5h$.
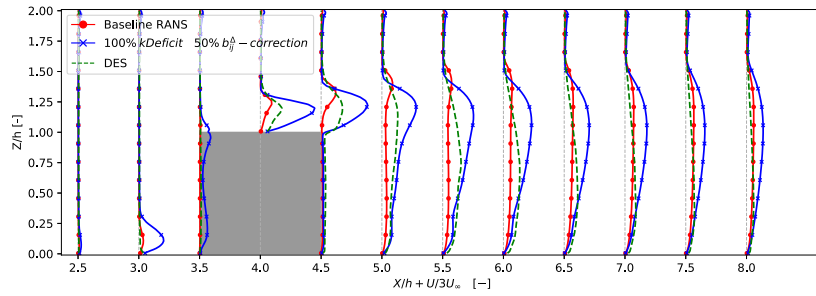
(c) T.k.e. (k) in symmetry plane.

Fig. 1. Flow A: Wall-mounted cube comparison.

cube, and the wake recovery is very significantly delayed. Frozen corrective fields are obtained, and symbolic regression is used to reduce these to a closure model, giving the following correction to $k - \omega$ SST:

$$b^{\Delta}_{\text{Cube}} = T^{(1)}(28.68\lambda_1 + 4.717\lambda_2 - 0.3560)$$
$$+ T^{(2)}(-88.99\lambda_1 + 68.77\lambda_2 + 13.80)$$
$$+ T^{(3)}(20.59\lambda_1 - 0.8594)$$
$$+ T^{(4)}(11.46\lambda_1 - 2.770\lambda_2 - 0.94412),$$

$$\hat{R}_{\text{Cube}} = T^{(1)}(-35.74\lambda_1\lambda_2 - 69.58\lambda_1$$
$$+ 39.74\lambda_2^2 + 7.573\lambda_2 + 3.739)$$
$$+ T^{(3)}(5.867\lambda_1 + 0.3755\lambda_2^2 - 6.784\lambda_2 - 16.48)$$
$$+ T^{(4)}(-2.304\lambda_1 - 0.2182\lambda_2^2 + 3.136\lambda_2 - 4.822).$$

These corrections are implemented in our solver, a modified version of `simpleFoam` in OpenFOAM, as a new turbulence model – this is straightforward and can be completely automated. The RANS solver is run with the new model for Flow A as a verification check. The resulting mean-flow is also shown in Fig. 1, where the flow in the entire domain is seen to better represent the reference. In particular the flow near the stagnation point, the separation on the top and sides, and the wake recovery are all closer to the reference – although wake recovery is still somewhat under-predicted.

The k-profiles in Fig. 1c show substantial overestimation of k by the model. To explain this observation, note that to obtain stable RANS simulations it was necessary to relax b^{Δ} by 50% to gain stability, and this factor was used in all simulations in this work. Furthermore we observed that the R correction term acted to *increase* the production of k, while the contribution of b^{Δ} acted to *decrease* production – especially in the wake. In the frozen corrective fields these effects cancelled to some extent, but the same cancellation
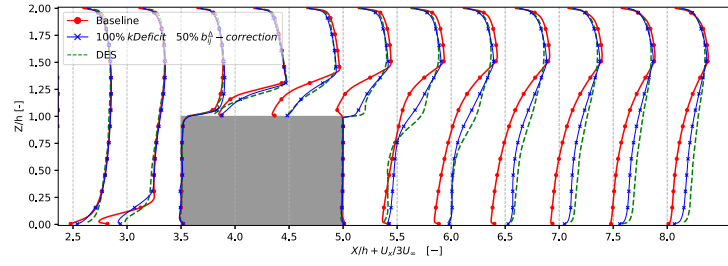
(a) Velocity (U_x) in symmetry plane.
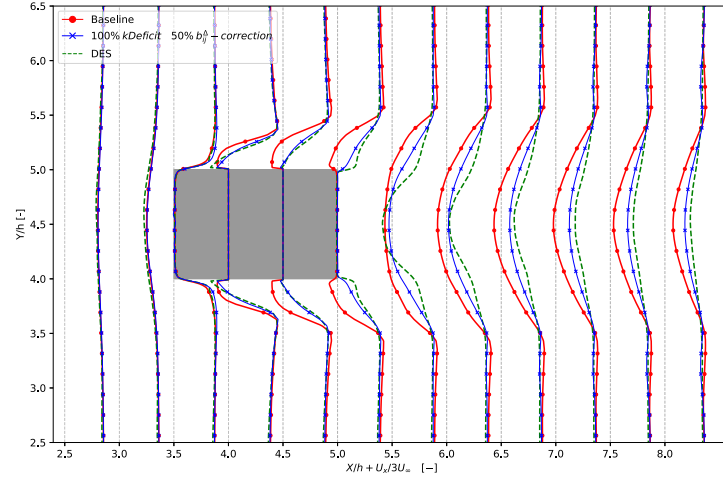
(b) Velocity (U_x) at mid-height of the cube $z = 0.5h$.
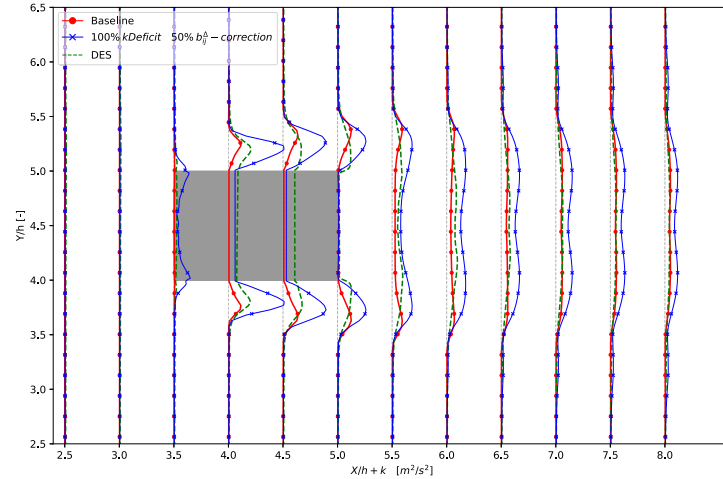
(c) T.k.e. (k) at mid-height of the cube $z = 0.5h$.

Fig. 2. Flow B: Extended cube comparison.

was less apparent in the discovered models, and this lead to instability of the solution. By introducing 50% relaxation of b^Δ we tend to increase k, and the eddy-viscosity beyond that recommended by the frozen corrective fields, and thereby stabilize the simulation, at the cost of overpredicting k. Note that the k-omega SST production limiter was used, applied to the production term net of the b^Δ correction, but not including the R correction.

In our experience instabilities are a major issue in data-driven RANS modelling. They are mitigated to some extent by our use of simple expressions for the closure – and are much worse with random-forest or neural network closures. We recommend further

work analysing instability mechanisms and adding instability penalties to the loss-function.

To make a prediction, the "Cube" model is applied to Flow B. The mean velocity profiles are shown in Fig. 2. The flow is topologically similar to Flow A, as such it is reasonable to hope that the model generalizes well. This is indeed the case, with significant improvements over the RANS baseline visible everywhere in the domain. Notably, on the longer top and sides of the obstacle, the reproduction of the ground-truth is good – while the flow separation is quite different. A weakness can be seen in the near wake, where in Fig. 2b a much narrower shear-layer is visible in the ref-
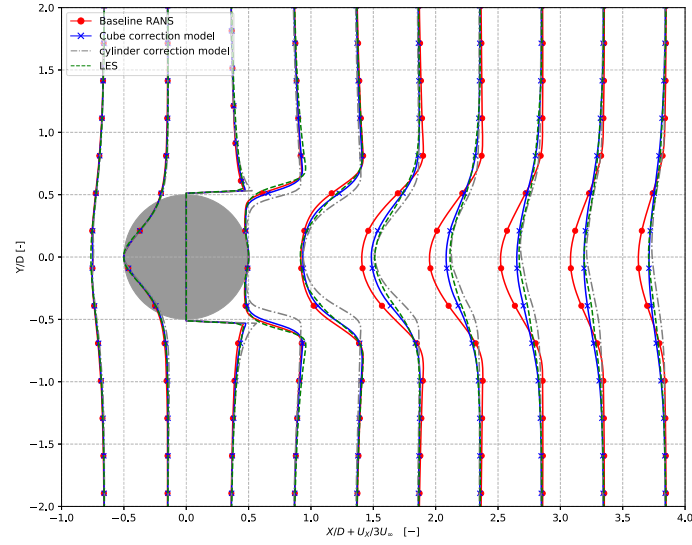
(a) Velocity (U_x).
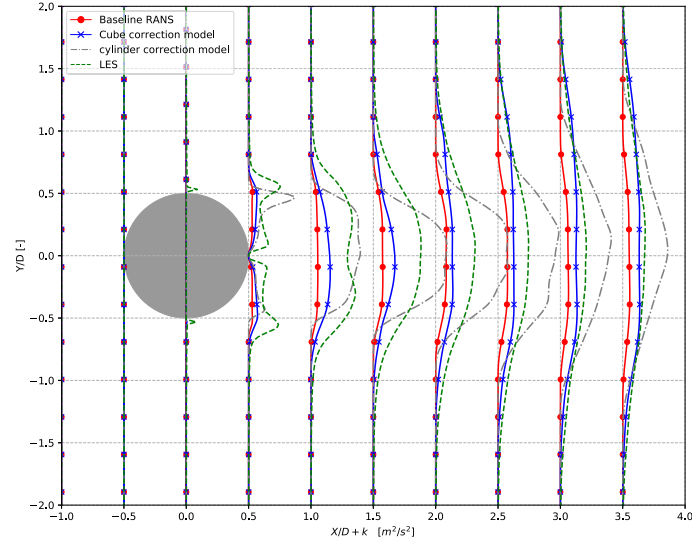
(b) T.k.e. (k).

Fig. 3. Flow C: Infinite cylinder comparison.

erence compared to the corrected model, however the wake recovery is good.

As a second prediction, the Cube model is applied to Flow C, see Fig. 3. In this flow the baseline RANS predicts the flow upstream of the obstacle very well, and the correction model does not modify the flow here. Once again however the baseline predicts much too slow wake recovery, and this is corrected almost completely by our model based only on data from Flow A.

The "Cylinder" model To investigate the sensitivity of the data-source on the resulting model and its performance, we apply the SpaRTA methodology using the wall-resolved LES data of Flow C. Again the frozen correction fields are found readily, and the symbolic regression approach gives:

$$b^\Delta_{\text{Cylinder}} = T^{(1)}(19.24\lambda_1 + 57.86\lambda_2 + 2.939)$$
$$+ 9.695 \cdot T^{(2)} + 7.805 \cdot T^{(3)} + 1.171 \cdot T^{(4)}, \quad (5)$$

$$\hat{R}_{\text{Cylinder}} = T^{(1)}(-2.823\lambda_1 + 33.82\lambda_2 + 2.586)$$
$$+ 11.17 \cdot T^{(3)} - 3.107 \cdot T^{(4)}. \quad (6)$$

In this case it was possible to find a significantly sparser model that nonetheless reduced the regression error to an acceptable level. This is likely at least partially thanks to the two-dimensionality of the mean-flow. Again the model was implemented in our OpenFOAM-based solver, and Flow C was predicted as a verification exercise – see Fig. 3. Once more the flow is significantly more accurate than the baseline (and especially in terms of wake recovery) except in a small region in the near wake. Notably the "Cylinder" model performs worse for this verification exercise, than the "Cube" model does in prediction. Again we suspect an issue with the stability of the RANS solver. The asymmetry observed in the cylinder correction model is due to the RANS solver not converging to a fully steady-state; the plotted profiles represent the solution at a single iteration (not averaged). Another factor is the

relatively diverse flow content of the cube data-set compared to the cylinder – the choice of training data will be a subject of future work.

Finally, in Fig. 3b k-profiles for Flow C are compared. Again both correction models improve dramatically on the baseline – which massively under-predicts k in the wake – however neither are particularly accurate. A likely cause is the large-scale dynamics of the Kármán vortex street, which contribute to the LES estimate of t.k.e. but are missing in RANS. Once more however the free-stream is unaltered.

4. Conclusions

We've demonstrated the ability to construct custom explicit algebraic Reynolds stress models for bluff-body flows from LES and DES data. We've shown the models have a generalization capability, with respect to different geometries and different flows, while consistently outperforming baseline models. This study shows the possibility of using RANS for massively separated flows using suitable machine-learning techniques, though many questions remain. Specifically: Is the method robust to training-data choice? Is symbolic regression general enough to capture the necessary corrections? How do the b^Δ and R corrections interact in k-production? Further work will focus on investigating the stability of automatically generated models, and scaling up training data from one flow, to a large database of flows.

Declaration of Competing Interest

No.

References

[1] Duraisamy K, Iaccarino G, Xiao H. Turbulence modeling in the age of data. Annu Rev Fluid Mech 2019;51:1–23. doi:10.1146/annurev-fluid.

[2] Xiao H, Cinnella P. Quantification of model uncertainty in RANS simulations: a review. Prog Aerosp Sci 2019;108:1–31. doi:10.1016/j.paerosci.2018.10.001.

[3] Ling J, Templeton J. Evaluation of machine learning algorithms for prediction of regions of high Reynolds averaged Navier Stokes uncertainty. Phys Fluids 2015;27(8):085103. doi:10.1063/1.4927765.

[4] Wang J-X, Wu J-L, Xiao H. Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data. Phys Rev Fluids 2017;2(3):034603. doi:10.1103/PhysRevFluids.2.034603.

[5] Kaandorp MLA, Dwight RP. Data-driven modelling of the Reynolds stress tensor using random forests with invariance. Comput Fluids 2020;202:104497. doi:10.1016/j.compfluid.2020.104497.

[6] Luan Y, Dwight RP. Influence of turbulence anisotropy on RANS predictions of wind-turbine wakes. J Phys Conf Ser 2020;1618:062059. doi:10.1088/1742-6596/1618/6/062059.

[7] Zhao Y, Akolekar HD, Weatheritt J, Michelassi V, Sandberg RD. RANS turbulence model development using CFD-driven machine learning. J Comput Phys 2020;411:109413. doi:10.1016/j.jcp.2020.109413.

[8] Schmelzer M, Dwight RP, Cinnella P. Discovery of algebraic Reynolds-stress models using sparse symbolic regression. Flow Turbul Combust 2020;104(2):579–603. doi:10.1007/s10494-019-00089-x.

[9] Steiner J, Dwight R, Viré A. Data-driven turbulence modeling for wind turbine wakes under neutral conditions. J Phys Conf Ser 2020;1618:062051. doi:10.1088/1742-6596/1618/6/062051. Publisher: IOP Publishing.6596/1618/6/062051

[10] Zhang Y., Dwight R.P., Schmelzer M., Gomez J.F., Hickel S., Han Z.h.. Customized data-driven RANS closures for bi-fidelity LES-RANS optimization. ArXiv:2004.03003.

[11] Menter FR. Two-equation eddy-viscosity turbulence models for engineering applications. AIAA J 1994;32(8):1598–605. doi:10.2514/3.12149. Tex.eprint

[12] Weatheritt J, Sandberg RD. The development of algebraic stress models using a novel evolutionary algorithm. Int J Heat Fluid Flow 2017;68:298–318. doi:10.1016/j.ijheatfluidflow.2017.09.017.

[13] Pope SB. Turbulent flows. Cambridge: Cambridge University Press; 2000. doi:101017/CBO9780511840531.

[14] Johnson RW. Handbook of fluid dynamics. CRC Press; 2016. doi:101201/b19031.

[15] Speziale CG. On nonlinear k-l and $k − \epsilon$ models of turbulence. J Fluid Mech 1987;178:459–75. doi:10.1017/S0022112087001319.

[16] Shih T-H, Liou WW, Shabbir A, Yang Z, Zhu J. A new $k − \epsilon$ eddy viscosity model for high Reynolds number turbulent flows. Comput Fluids 1995;24(3):227–38. doi:10.1016/0045-7930(94)00032-T.

[17] Korns MF, Riolo R, McConaghy T, Vladislavleva E. Abstract expression grammar symbolic regression. In: Genetic programming theory and practice VIII, Genetic and evolutionary computation. New York, NY: Springer; 2011. p. 109–28. doi:10.1007/978-1-4419-7747-2_7.

[18] Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Ser B (Statistical Methodology) 2005;67(2):301–20. doi:10.1111/j.1467-9868.2005.00503.x.

[19] Friedman JH, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33(1):1–22. doi:10.18637/jss.v033.i01.

[20] Alba RD. A graph- theoretic definition of a sociometric clique. J Math Sociol 1973;3(1):113–26. doi:10.1080/0022250X.1973.9989826.

[21] Chen C, Luo C, Jiang Z. Elite bases regression: a real-time algorithm for symbolic regression. In: 2017 13th International conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD); 2017. p. 529–35. doi:10.1109/FSKD.2017.8393325.

[22] Martinuzzi R, Tropea C. The flow around surface-mounted, prismatic obstacles placed in a fully developed channel flow (data bank contribution). J Fluids Eng 1993;115(1):85–92. doi:10.1115/1.2910118.

[23] Alfonsi G, Restano C, Primavera L. Coherent structures of the flow around a surface-mounted cubic obstacle in turbulent channel flow. J Wind Eng Ind Aerodyn 2003;91(4):495–511. doi:10.1016/S0167-6105(02)00429-4.

[24] Breuer M. A challenging test case for large eddy simulation: high Reynolds number circular cylinder flow. Int J Heat Fluid Flow 2000;21(5):648–54. doi:10.1016/S0142-727X(00)00056-4.

[25] Spalart PR, Deck S, Shur ML, Squires KD, Strelets MK, Travin A. A new version of detached-eddy simulation, resistant to ambiguous grid densities. Theor Comput Fluid Dyn 2006;20(3):181. doi:10.1007/s00162-006-0015-0.

[26] Poletto R, Craft T, Revell A. A new divergence free synthetic eddy method for the reproduction of inlet flow conditions for LES, flow. Turbul Combust 2013;91(3):519–39. doi:10.1007/s10494-013-9488-2.

[27] Spalart P, Allmaras S. A one-equation turbulence model for aerodynamic flows. 30th Aerospace sciences meeting and exhibit, aerospace sciences meetings, American Institute of Aeronautics and Astronautics; 1992. doi:102514/61992-439.