Accepted Manuscript

A Framework to Develop Data-Driven Turbulence Models for Flows with Organised Unsteadiness

Chitrarth Lav, Richard D. Sandberg, Jimmy Philip
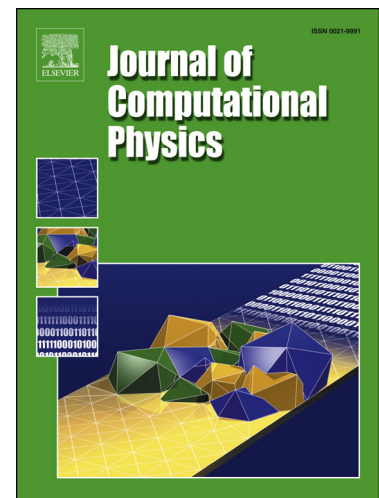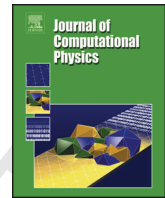
Please cite this article in press as: C. Lav et al., A Framework to Develop Data-Driven Turbulence Models for Flows with Organised Unsteadiness, *J. Comput. Phys.* (2019), https://doi.org/10.1016/j.jcp.2019.01.022

Highlights

- A novel machine-learnt framework for flows with organised unsteadiness
- Turbulence closure developed from DNS data for the stochastic part of the anisotropy
- Gene-Expression Programming used to develop mathematical expression for anisotropy
- Developed model reveals turbulence diffusion needs to be significantly reduced
- Closures used in URANS and PANS show significant improvement of results for a wake

A Framework to Develop Data-Driven Turbulence Models for Flows with Organised Unsteadiness

Chitrarth Lav[a,*], Richard D. Sandberg[a,1], Jimmy Philip[a]

[a]*Department of Mechanical Engineering, University of Melbourne, Parkville, Victoria 3010, Australia*

ARTICLE INFO

Article history:

ABSTRACT

Turbulence modelling development has received a boost in recent years through assimilation of machine learning methods and increasing availability of high-fidelity datasets. This paper presents an approach that develops turbulence models for flows exhibiting organised unsteadiness. The novel framework consists of three parts. First, using triple decomposition, the high-fidelity data is split into organised motion and stochastic turbulence. A data-driven approach is then used to develop a closure only for the stochastic part of turbulence. Finally, unsteady calculations are conducted, which resolve the organised structures and model the unresolved turbulence using the developed bespoke turbulence closure. A case study of a wake with vortex shedding behind a normal flat plate, at a Reynolds number of 2,000, based on plate height and freestream velocity, is considered to demonstrate the method. The approach shows significant improvement in mean velocity and Reynold stress profiles compared with standard turbulence models.

© 2019 Elsevier Inc. All rights reserved.

1. Introduction

1.1. Unsteady flows and their prediction

Most flows encountered in engineering applications are unsteady and the importance of their accurate prediction cannot be overemphasised. A subset of these flows includes those that exhibit some form of organised unsteadiness. This organised unsteadiness often manifests itself through large-scale coherent structures. Existence of coherent structures has been known for more than three decades, yet there is still little consensus over a holistic definition. One possible attempt at defining them is that they are a "large-scale turbulent fluid mass with a phase-correlated vorticity over its spatial extent, i.e. underlying the three-dimensional random vorticity fluctuations characterising turbulence, there is an organised component of the vorticity which is phase correlated over the extent of the structure" [1]. This means that the organised component is embedded inside a randomly distributed field; however, the coherent structures

*Corresponding author:
e-mail: chitrarth.lav@gmail.com (Chitrarth Lav)
[1]richard.sandberg@unimelb.edu.au

associated with this component are not easily identified or characterised [2]. To simplify the identification and characterisation process, a smaller subset can thus be considered, wherein the organised unsteadiness is deterministic in nature. Examples include vortex shedding behind a bluff body, external fluid forcing and unsteadiness due to motion of turbine/compressor blades.

Prediction of unsteady flows for industrial applications still relies heavily on Reynolds-Averaged Navier Stokes (RANS) simulations. However, a review by Tucker [3] found that RANS as a tool for simulating unsteady turbomachinery flows has a catalogue of poor predictions. The most notable of these failures is in predicting separated flows, such as those behind a bluff body. Tucker [4] cites several studies, including those of Franke & Rodi [5], Johansson *et al.* [6], Durbin [7] and Przulj & Younis [8], which observed these prediction issues. Growing computational resources have allowed use of high-fidelity (HiFi) simulations, such as Direct Numerical Simulations (DNS), but these calculations remain too expensive for industrial design analyses, and are mostly restricted to relatively simple geometries and low to moderate Reynolds numbers (Re). An alternative are scale-resolving simulations (SRS) which use cut-off filters to resolve and model parts of the turbulence spectrum, such as Large Eddy Simulations (LES), Detached Eddy Simulations (DES) and Partially Averaged Navier-Stokes (PANS). As the name suggests, these simulations resolve a portion of the turbulent length scales, and as a result are a good compromise between the computational demand of a DNS and the lack of consideration of coherent structure influence of a RANS [2]. Of these, PANS is of particular interest because it serves as a bridging method that spans the range of resolved scales, from all scales (DNS) to no scales (RANS) [9]. First introduced by Girimaji & Abdol-Hammid [10], PANS has found use in multiple areas of research over the years such as the lid-driven cavity flow [11, 12], flow past square and circular cylinders [13, 14, 15] and wall bounded flows [16, 17, 18]. The drawback of RANS modelling of not accounting for the coherent structures can thus be overcome by using PANS. However, PANS by itself is not a sufficient tool for improved prediction as it still relies on the same turbulence closures as RANS.

A recent 2030 CFD Vision report highlighted concerns of the aerospace industry that turbulence modelling efforts have reached a plateau [19]. These concerns can be attributed to the state of current turbulence models, which are old compared with the increasing computational power and recent advances in CFD technology, particularly in the areas of solution fidelity and speed [20]. Bringing machine learning into the centrefold for advancing turbulence model development is the logical next step. Furthermore, combining SRS such as PANS with these new turbulence models could propel the unsteady flow prediction efforts forward.

1.2. Machine learning as a tool

One of the most widely used turbulence closure for RANS calculations is the Boussinesq approximation [21], where the modelled anisotropy tensor (a_{ij}^{mod}) is related linearly to the traceless mean strain rate tensor ($S_{ij} = 0.5(\partial U_i/\partial x_j + \partial U_j/\partial x_i)$) through the eddy viscosity (ν_t) and turbulence kinetic energy (TKE, k):

$$a_{ij}^{mod} = -\frac{\nu_t}{k} S_{ij}, \tag{1}$$

This model for the anisotropy can be used to obtain the modelled Reynolds stress (τ_{ij}^{mod}), through:

$$\tau_{ij}^{mod} = \frac{2}{3} k \delta_{ij} - 2\nu_t S_{ij}. \tag{2}$$

Numerous studies using this closure have reported differences in the quantities of interest such as the mean velocity and Reynolds stress profiles compared with experimental and HiFi data [22, 23, 24, 25, 26, 27]. To better understand the reason for differences, the validity of the Boussinesq hypothesis was, for example, tested by Schmitt [28] through computing the alignment of the modelled anisotropy tensor with the reference one given by

$$\rho_{RS} = \frac{a_{ij}^{ref} a_{ji}^{mod}}{\sqrt{a_{mn}^{ref} a_{nm}^{ref} a_{pq}^{mod} a_{qp}^{mod}}}, \tag{3}$$

$$a_{ij}^{ref} = \frac{\tau_{ij}}{2k} - \frac{\delta_{ij}}{3}. \tag{4}$$

The alignment ρ_{RS} is defined as the normalised inner product between the anisotropy tensor from the reference database (Eqn. 4) and that from the closure model. Values for both these tensors are obtained from the reference

database. A completely valid model would exhibit perfect alignment with a value of 1.0, while models that are orthogonal would have a value of 0.0 and models with a value of -1.0 are oppositely aligned. For testing the Boussinesq approximation, a_{ij}^{mod} is obtained from Eqn. (1). It was found that this universally used closure was not valid for most simple flows [28]. A natural extension then, to the linear approximation hypothesis, is a non-linear approximation hypothesis. Numerous attempts have been made to develop non-linear constitutive relationships between the anisotropy and tensor basis functions and scalar invariants with moderate success [29, 30, 31].

However, with the recent abundance of HiFi datasets, research into extracting useful information from them to develop better non-linear closures has taken centrestage [32]. These studies rely on data-driven machine learning algorithms to obtain the information and generate models, such as those of Duraisamy [32] or Tracey *et al.* [20]. One such approach of interest is that of Weatheritt & Sandberg [33, 34], which uses Gene Expression Programming (GEP) to develop non-linear closures for the anisotropy tensor. The non-linear relationship obtained through GEP is a function of tensor basis (V_{ij}^n) and scalar invariant functions (I_n) i.e. $a_{ij}^{mod} = f(V_{ij}^1, V_{ij}^3, V_{ij}^3, I_1, I_2)$.

$$
\begin{aligned}
V_{ij}^1 &= \frac{S_{ij}}{\omega}, \qquad I_1 = S_{mn}S_{nm}, \qquad I_2 = \Omega_{mn}\Omega_{nm}, \\
V_{ij}^2 &= \frac{S_{ik}\Omega_{kj} - \Omega_{ik}S_{kj}}{\omega}, \\
V_{ij}^3 &= \frac{S_{ik}S_{kj} - \frac{1}{3}\delta_{ij}S_{mn}S_{nm}}{\omega}.
\end{aligned}
\tag{5}
$$

Here, $S_{ij} = 0.5(\partial U_i/\partial x_j + \partial U_j/\partial x_i)$, $\Omega_{ij} = 0.5(\partial U_i/\partial x_j - \partial U_j/\partial x_i)$, $\omega = \epsilon/k$ are the mean strain rate, mean rotation rate and specific dissipation rate, respectively. k refers to the turbulent kinetic energy while ϵ to the dissipation. For a more detailed discussion of the machine learning process, the reader is referred to Weatheritt & Sandberg [33]. A brief explanation of this process will be outlined in Section 2.3. Throughout this text, the words "training" and GEP refer to the machine learning process.

In this study, a proof-of-concept is offered that is built upon the machine learning algorithm of Weatheritt & Sandberg [33], extending the application to flows that exhibit organised unsteadiness. Section 2 presents the three prongs of the approach and explains how they interlink with one another to produce a novel technique for unsteady flow prediction. Section 3 applies this approach via a case study and discusses other areas of applications and it's limitations.

2. Conceptual approach

2.1. Segregation of length scales

The idea of length scale segregation, in the context of deterministic unsteadiness, is not new, and was first considered by Hussain & Reynolds [35, 36]. It states that a given flow variable such as velocity could be decomposed into a time independent mean, stochastic unsteadiness and deterministic unsteadiness (Eqn. (6)). Combining these two types of unsteadiness gives rise to the well known Reynolds decomposition while their individual representation is known as triple decomposition.

$$
u(\mathbf{x}, t) = \overline{u}(\mathbf{x}) + u'(\mathbf{x}, t) = \overline{u}(\mathbf{x}) + \widetilde{u}(\mathbf{x}, t) + u''(\mathbf{x}, t).
\tag{6}
$$

where $\overline{u}$ represents the time-independent mean, the stochastic component (u'') represents the random part of turbulence while the deterministic component ($\widetilde{u}$) represents the organised wave motion whose frequencies one knows or can extract easily. Because these organised motions have particular frequencies, the term "periodic" would be used interchangeably to represent this component. The Reynolds stress tensor (as well as the turbulence kinetic energy and the anisotropy tensor) can be decomposed using the result from Eqn. (6) as follows:

$$
\begin{aligned}
\tau_{ij} &= \overline{u_i'u_j'}, \\
&= \overline{(\widetilde{u}_i + u_i'')(\widetilde{u}_j + u_j'')}, \\
&= \overline{\widetilde{u}_i\widetilde{u}_j} + \overline{u_i''u_j''}, \\
&= \widetilde{\tau}_{ij} + \tau_{ij}''.
\end{aligned}
\tag{7}
$$

The first term above can be considered as a "periodic" Reynolds stress while the second term can be considered as a "turbulent" Reynolds stress. It can be seen that the total Reynolds stress is a sum of these two stresses. The "periodic" component is the result of organised wave motion on the flow field while the "turbulent" component is the result of the stochastic unsteadiness. Since these organised motions are usually limited in bandwidth to a small number of frequencies, it might be possible to segregate the length scales associated with the organised motion and turbulence. Assuming such a split is possible, one can study the impact of each component separately, which may assist in developing better turbulence models. An example of such a split is shown in detail in Section 3, where vortex shedding from the wake of a bluff body is responsible for the organised motion.

2.2. Resolution of length scales

A RANS calculation models all length scales, which for unsteady flows results in poor predictions. Improved prediction can be sought if scales not contributing to turbulence were resolved instead of being modelled. Following the segregation of scales, the "periodic" scales can be resolved through a PANS calculation. The underlying equations involved in PANS are quite similar to those used in RANS. The difference between the two stems from the modified turbulence equations, such as the $k - \epsilon$ or $k - \omega$ equations. For this study, the standard $k - \omega$ SST model [37] is used as a baseline for comparison and modification, as it is widely used in industrial calculations. PANS has been adopted over other SRS for this study due to the availability of explicit cut-off parameters f_k and f_ω, defined as,

$$f_k = \frac{k_u}{k}; \quad f_\omega = \frac{\omega_u}{\omega}. \tag{8}$$

The subscript u represents unresolved/modelled quantities. The cut-off f_ω can be further simplified to:

$$f_\omega = \frac{\omega_u}{\omega} = \frac{\epsilon_u}{\epsilon} \frac{k}{k_u},$$
$$= \frac{f_\epsilon}{f_k} = \frac{1}{f_k}. \tag{9}$$

Here, the ratio f_ϵ is set to 1.0 which signifies all dissipation must be modelled. This is a valid choice since dissipation is a small scale phenomenon and resolving it would amount to conducting a HiFi simulation, which is not the aim of this approach.

With the correct choice of f_k as an input parameter, the PANS equations can resolve the length scales associated with organised motion. Recent research [15] has identified with this idea, demonstrating the extent of resolution of coherent structures, associated with vortex shedding, with different spatially constant values of f_k. However, to resolve only the relevant length scales, f_k must vary spatially instead, which is permissible since its time scale of change is smaller than those of turbulence quantities [17]. The spatially varying f_k field (referred to as f_k^* to differentiate from the commonly used f_k in literature) can be extracted from the HiFi dataset once a decomposition has been performed according to Eqn. (7). Comparing the spatially varying "periodic" and "turbulent" TKE from the HiFi dataset to the resolved and unresolved TKE contributions from PANS, the "periodic" component should be equivalent to the resolved contribution and the "turbulent" component to the modelled/unresolved contribution. This means $k_u = k''$, so

$$f_k^* = \frac{k''}{k}. \tag{10}$$

2.3. Modelling of unresolved length scales

As mentioned in Section 1.2, the standard closure used in RANS modelling is the Boussinesq approximation (Eqn. 2), where the TKE is obtained through solving the k equation, in conjunction with the equation for ω in the case of using the $k - \omega$ SST model [37]:

$$\frac{\partial k}{\partial t} + U_j \frac{\partial k}{\partial x_j} = P_k - \beta^* k\omega + \frac{\partial}{\partial x_j}\left[\nu_t \sigma_k \frac{\partial k}{\partial x_j}\right], \tag{11}$$

$$\frac{\partial \omega}{\partial t} + U_j \frac{\partial \omega}{\partial x_j} = \alpha S^2 - \beta \omega^2 + \frac{\partial}{\partial x_j}\left[\nu_t \sigma_\omega \frac{\partial \omega}{\partial x_j}\right] + 2(1 - F_1)\sigma_{\omega 2}\frac{1}{\omega}\frac{\partial k}{\partial x_i}\frac{\partial \omega}{\partial x_i}. \tag{12}$$
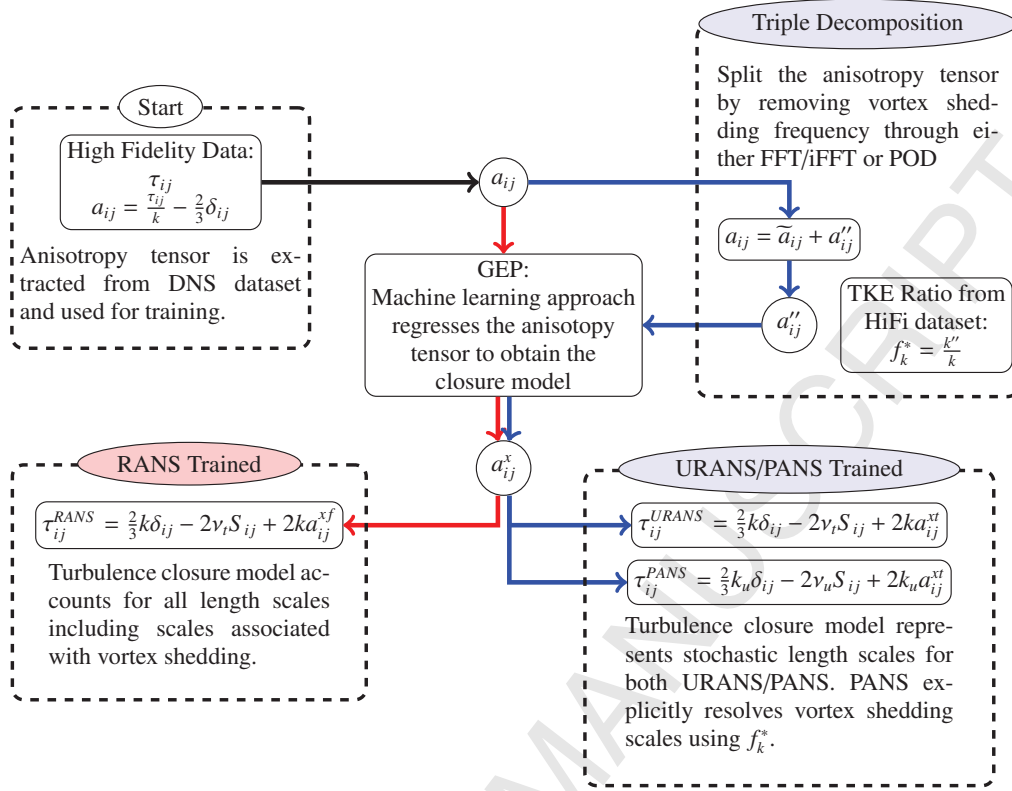
Fig. 1: Schematic of model training for steady (following red arrows) and unsteady flows (following blue arrows)

Solving these equations for k and ω then allows construction of the turbulent eddy viscosity, as required by the Boussinesq approximation in Eqn. (1):

$$\nu_t = \frac{a_1 k}{\max(a_1 \omega, S F_2)}.$$

The terms F_1, F_2 and $CD_{k\omega}$ are defined as:

$$F_1 = \tanh \left[\left(\min \left[\max \left(\frac{\sqrt{k}}{\beta^* \omega y}, \frac{500\nu}{y^2 \omega} \right), \frac{4\sigma_{\omega 2} k}{CD_{k\omega} y^2} \right] \right)^4 \right],$$

$$F_2 = \tanh \left[\left[\max \left(\frac{2\sqrt{k}}{\beta^* \omega y}, \frac{500\nu}{y^2 \omega} \right) \right]^2 \right],$$

$$CD_{k\omega} = \max \left(2\sigma_{\omega 2} \frac{1}{\omega} \frac{\partial k}{\partial x_i} \frac{\partial \omega}{\partial x_i}, 10^{-10} \right).$$

while the model coefficients (say Φ) are obtained by blending the coefficients of the $k-\epsilon$ and $k-\omega$ models using: $\Phi = \Phi_1 F_1 + (1-F_1)\Phi_2$. Thus, the coefficients α, β, σ_k and σ_ω are blends of $(\alpha_1 = 5/9, \alpha_2 = 0.44)$, $(\beta_1 = 3/40, \beta_2 = 0.0828)$, $(\sigma_{k1} = 0.85, \sigma_{k2} = 1.0)$ and $(\sigma_{\omega 1} = 0.5, \sigma_{\omega 2} = 0.856)$ respectively, while $\beta^* = 0.09, a_1 = 0.31$. $S = \sqrt{2S_{ij}S_{ij}}$ refers to the mean strain rate, y is the distance to the nearest wall and P_k to the production of the modelled TKE:

$$P_k = \tau_{ij} \frac{\partial U_i}{\partial x_j}. \tag{13}$$

This model in its current form ($k - \omega$ SST with the Boussinesq closure) is responsible for modelling all the length scales in the flow, and has been shown in literature to be a poor form of representation of turbulence [21, 28]. To improve upon the modelling, the machine learning approach of Weatheritt & Sandberg [33] uses HiFi datasets to

regress the anisotropy tensor and obtain a new closure model instead of the linear Boussinesq model for the anisotropy. The regression undertaken by the algorithm assumes no particular functional form for the anisotropy closure model. Instead, an initial population of models is generated randomly, with the algorithm driving towards a final model by minimising the error between the reference and model anisotropies:

$$J(a_{ij}^{mod}) = \frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{3} \sum_{j \leq i} [w_{ij}(a_{ij}^{ref} - a_{ij}^{mod})]^2. \tag{14}$$

Here, J refers to the cost function and w_{ij} are weighting functions, set in this study to 0.25 i.e. $w_{11} = w_{22} = w_{33} = w_{12} = 0.25$, while N refers to the number of grid points used from the reference dataset to fit the model. This number depends on the spatial domain chosen for the training and the grid resolution in the chosen domain. The algorithm begins a filtering process, analogous to the survival-of-the-fittest approach, by rejecting models that have high error values. The surviving models are then genetically mutated by incrementally and randomly changing the coefficients of the tensor basis functions, providing a mutated population of models. Successive iterations of the algorithm result in populations with decreasing error values, with the final model selected from the last iteration with the lowest error value. The GEP is not a deterministic algorithm, i.e., the final model obtained is dependent on the initial population, amongst other factors. To offset bias, the algorithm is executed multiple times with different initialisations and the final model chosen is an ensemble average of the models that occur most frequently. This can be obtained by statistically representing the obtained models on a histogram. An example of the use of the histogram for model selection will be shown in Sec. 3.5.

The GEP implementation described above has been used extensively for RANS model development [33, 34, 38, 39], where the "full" anisotropy tensor (a_{ij}^{ref}) was used to produce a model (a_{ij}^{xf}) for the extra anisotropy which is then added to the modified Boussinesq approximation as a new closure:

$$\tau_{ij}^{RANS} = \frac{2}{3}k\delta_{ij} - 2\nu_t S_{ij} + 2k a_{ij}^{xf}. \tag{15}$$

To conduct a RANS calculation with a trained model, the Reynolds stress was constructed from Eqn. (15), with the production term in Eqn. (13) being modified as well, using the expression for τ_{ij}^{RANS}. For steady RANS calculations, the trained closure above has shown improvements in the prediction of the mean flow statistics [33, 34, 39]. In this paper, for the first time, the GEP algorithm is being evaluated for and applied to unsteady RANS (URANS) calculations. In that context, we propose a modification to the turbulence closure in Eqn. (15). As explained previously, this closure has been obtained from the "full" reference anisotropy tensor (a_{ij}^{ref}), which uses all the length scales, including the contribution from the deterministic ones, i.e. $\widetilde{a_{ij}}\big|_{ref}$. A URANS calculation of an inherently unsteady flow usually captures the large-scale low-frequency unsteadiness. So, using the closure in Eqn. (15) would amount to accounting for these deterministic scales twice: once via the closure and once via the resolution of scales. The new approach proposed here, develops a turbulence closure (a_{ij}^{xt}) using only the reference stochastic anisotropy, i.e. $a_{ij}''\big|_{ref}$, while the URANS equations capture the deterministic unsteadiness. The resulting turbulence closure from such an endeavour is given by:

$$\tau_{ij}^{URANS} = \frac{2}{3}k\delta_{ij} - 2\nu_t S_{ij} + 2k a_{ij}^{xt}. \tag{16}$$

As done in the case of RANS, the Reynolds stress term in Eqn. (13) is now replaced with τ_{ij}^{URANS}. While the new turbulence closure (Eqn. (16)) is a true representation of the stochastic unsteadiness, the same cannot be said about the resolution of the deterministic unsteadiness in the URANS sense. To obtain the true determinstic representation, the turbulence equations for URANS require modification to their equivalent PANS counterparts, i.e, a modification of the existing $k - \omega$ SST model to the $k_u - \omega_u$ SST model. This allows the modified equations to explicitly solve for the unresolved (modelled) fields i.e., k_u, ω_u instead of k, ω. The total fields can be reconstructed using the ratios f_k and f_ω and the unresolved fields k_u, ω_u from Eqn. (8). Using the correct value of these ratios can specify the level of resolution acheived. For the deterministic scales, the extracted ratio f_k^* in Eqn. (10) is the correct choice.

The equations for the $k_u - \omega_u$ SST model are given by Eqns. (17), (18), where the ratio f_ω has been replaced with
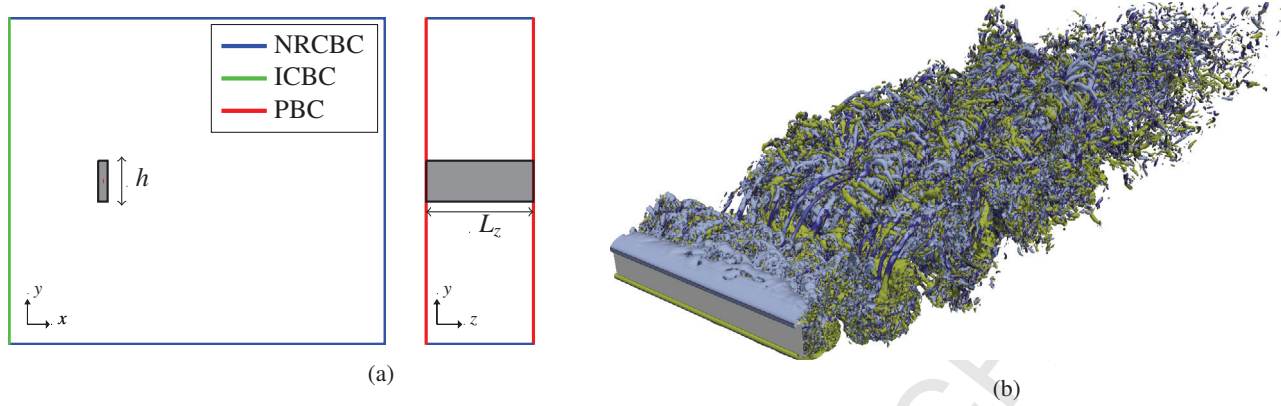
Fig. 2: (a) Schematic of the domain for the DNS calculation, with the boundary conditions (b) Contours of the Q-criterion for the DNS calculation

the definition in Eqn. (9):

$$\frac{\partial k_u}{\partial t} + U_j \frac{\partial k_u}{\partial x_j} = P_{k_u} - \beta^* k_u \omega_u + \frac{\partial}{\partial x_j}\left[\nu_u \sigma_{ku} \frac{\partial k_u}{\partial x_j}\right], \tag{17}$$

$$\frac{\partial \omega_u}{\partial t} + U_j \frac{\partial \omega_u}{\partial x_j} = \alpha \frac{P_{k_u}}{\nu_u} - \left(\alpha\beta^* \frac{k_u}{\nu_u} - \alpha\beta^* f_k^* \frac{k_u}{\nu_u} + \beta f_k^* \omega_u\right)\omega_u$$

$$+ \frac{\partial}{\partial x_j}\left[\nu_u \sigma_{\omega u} \frac{\partial \omega_u}{\partial x_j}\right] + 2(1 - F_1)\sigma_{\omega 2} \frac{1}{f_k^{*2}\omega_u} \frac{\partial k_u}{\partial x_i} \frac{\partial \omega_u}{\partial x_i}. \tag{18}$$

The first and second terms to the right hand side in the ω_u equation have been modified following the corrections as in the formulations of Pereira *et al.* [40, 15]. In these equations, ν_u, σ_{ku} and $\sigma_{\omega u}$ are given by:

$$\nu_u = \frac{a_1 k_u}{\max(a_1 \omega_u, S F_2)}, \quad \sigma_{ku} = \frac{\sigma_k}{f_k^{*2}}, \quad \sigma_{\omega u} = \frac{\sigma_\omega}{f_k^{*2}}.$$

Through these modifications, the resulting closure is given by:

$$\tau_{ij}^{PANS} = \frac{2}{3}k_u \delta_{ij} - 2\nu_u S_{ij} + 2k_u a_{ij}^{xt}. \tag{19}$$

and the unresolved production term P_{k_u} using the above closure is expressed as:

$$P_{k_u} = \tau_{ij}^{PANS} \frac{\partial U_i}{\partial x_j}. \tag{20}$$

Figure 1 shows a schematic that summarises the three approaches described above. The red arrows show the path taken in the training process for steady flows [33] and what resulting term needs to be implemented in a RANS solver, while the blue arrows show the novel training process, including the new term that needs to be implemented into a URANS and a PANS solver. This novel training process, it's subsequent implementation and use as a calculation tool is termed as the Data-driven Stochastic Closure Simulation (DSCS), owing to the data-driven turbulence closure developed using the stochastic length scales. The calculations performed with the machine-learnt closures will be referred to as trained calculations (T) while the baseline calculations utilising the Boussinesq approximation will be referred to as untrained calculations (UT).

3. Case study

3.1. Reference Dataset

The test case considered for demonstration for the new approach is that of a flat plate normal to the flow, which acts as a wake generating body (WGB). A schematic of the geometry is shown in Figure 2a. The flow is near incompressible (Mach ∼ 0.3) at Re = 2,000, based on the plate height and freestream velocity. The wake generated is
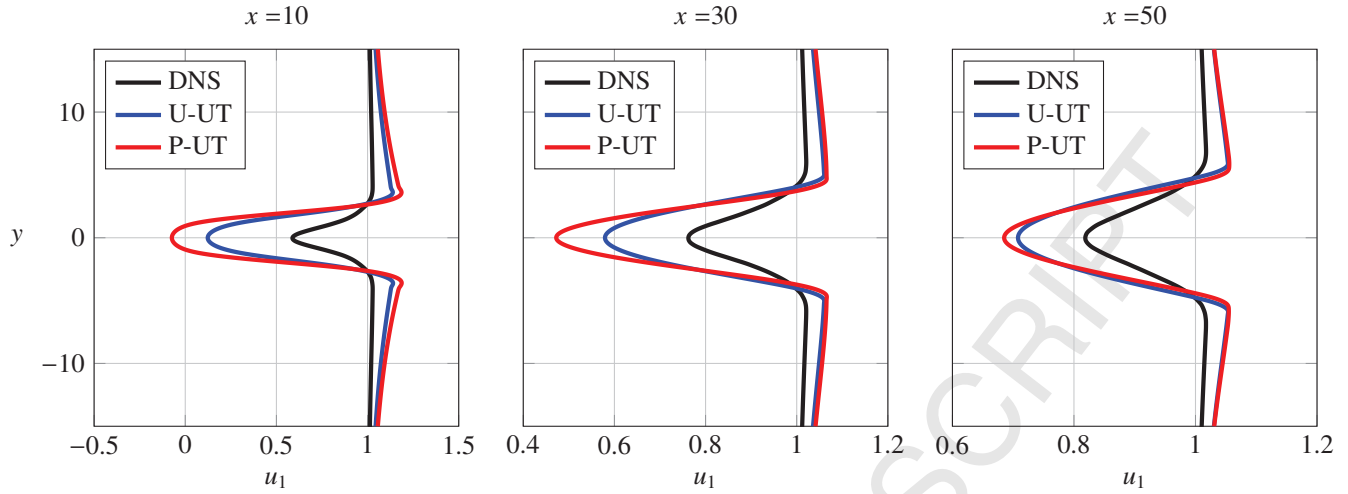
Fig. 3: Mean velocity profiles at 3 different streamwise locations for the untrained calculations. U: URANS, P: PANS, UT: Untrained

allowed to evolve in a zero pressure gradient environment. A DNS is conducted for this case to generate the high-fidelity dataset. The calculation is performed using HiPSTAR, which is a structured finite difference compressible Navier-Stokes solver employing a fourth-order accurate stencil for spatial discretisation and a fourth-order accurate Runge-Kutta fourth order scheme for temporal discretisation [41]. Due to the compressible nature of the code, characteristic boundary conditions, as proposed by Kim & Lee [42, 43], are employed to prevent spurious reflections from the inflow/outflow boundaries. The outflow boundaries are set to the Non-Reflecting Characteristic (NRCBC) while the inflow boundary is set to the Integrated Characteristic (ICBC). The outflow boundary is also coupled with a zonal boundary condition, which acts as a sponge to damp out any reflections back into the domain [44]. The spanwise direction is solved in Fourier space to reduce computational effort and as such the spanwise boundaries are periodic. A grid resolution of $1,056 \times 600$ points in the $x - y$ plane with 64 Fourier modes (or equivalently 130 points) in the span wide direction was found to be sufficiently well resolved, for a plate of dimensions 1.0×0.16 units with a spanwise extent (L_z) of 8.0 units. The spanwise extent was chosen after a domain convergence study was conducted and was found to be larger than the extents chosen by other HiFi numerical studies [45, 46]. The domain in the $x - y$ plane extends from -15.0 to 85.0 units in the x and -50.0 to 50.0 in the y directions. Grid convergence tests were conducted on the configuration to ensure resolution of the smallest length scales in the flow. This was ascertained from the TKE equation budget terms, by comparing the sum of budgets from all the terms except the dissipation to the dissipation budget. It was found that the difference between the two quantities was negligible, giving confidence in the DNS resolution.

An instantaneous snapshot of this calculation is presented in Figure 2b which shows the Q-criterion [47]. It can be clearly seen that the flow is highly three dimensional and close to the WGB, the flow exhibits large coherent structures which undergo complete breakdown as the flow moves further downstream. Since the training algorithm requires the anisotropy tensor from the high-fidelity dataset, spanwise-averaged mean statistics were collected from the DNS, following a time-convergence test. The final data collected spanned 600 time units, which corresponds to over 6 flow-through times for the given domain. The test case was also simulated using the WALE [48] LES model, with a grid resolution of 672×456 in the $x - y$ plane and 64 Fourier modes. This calculation was performed to compare the computational costs associated with the different calculations undertaken in this study, outlined later in Table 1.

3.2. Baseline

Next, a URANS calculation was performed using the standard $k - \omega$ SST model in OpenFOAM to assess the differences with the DNS results. It also serves as a baseline to compare the modified closure results with. The domain extents in the $x - y$ plane were the same as those for the DNS. Two different in-plane ($x - y$) grid resolutions were considered, $1,056 \times 600$ and 672×456, which correspond to the DNS and LES resolutions. Both two and
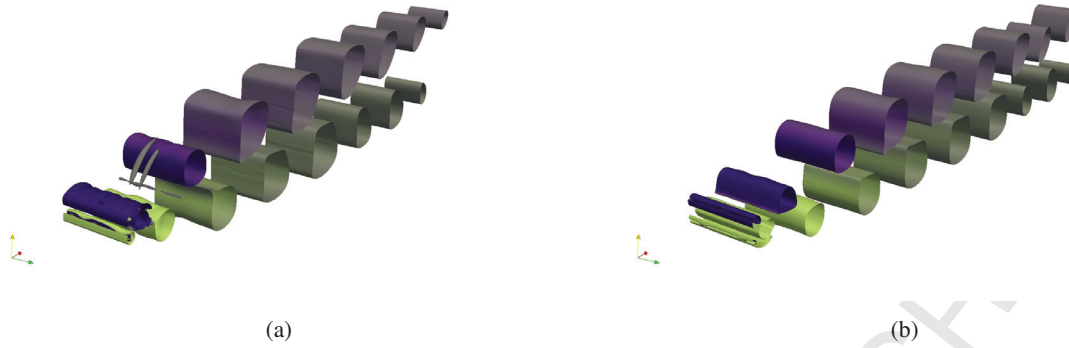
(a)												(b)

Fig. 4: Q-criterion isosurfaces coloured by the spanwise vorticity for (a) URANS UT (b) PANS UT

three-dimensional calculations were then conducted where it was observed that the differences in the mean profiles for the two-dimensional and three-dimensional simulations were insignificant. This finding is not surprising and has been observed in other wake flow calculations [49] for this Re, on account of no spanwise variation. This is also evident from the Q-criterion for the three-dimensional URANS calculation in Figure 4a. The spanwise extent for the study was limited to 4 units, matching other three-dimensional URANS wake flow studies [50, 51].

Figure 3 shows the grid independent mean velocity profiles for the baseline URANS, i.e, the untrained URANS (using the Boussinesq approximation), at three different streamwise locations downstream of the WGB for the 672×456 grid. These results indicate that the baseline model fails to correctly predict the wake evolution, even for a simple zero pressure gradient wake flow. Other parameters such as the Reynolds stress, wake half-width and wake velocity deficit were overpredicted as well, but are not shown here. There are two schools of reasoning that can account for this overprediction: insufficient turbulent diffusion and excess turbulent diffusion. Both lines of reasoning, though contradictory, are equally valid, depending on the proportion of modelling to resolving. From a RANS perspective, where all length scales are modelled, more diffusion added to the closure would correct the overprediction. If a proportion of the length scales is to be resolved, lower turbulent diffusion can correct the overprediction as the resolved scales would provide a more realistic representation of the Reynolds stresses. This is similar to an LES, which has less modelled turbulent diffusion, through a smaller eddy viscosity. In other words, the more length scales are resolved, the smaller the turbulence viscosity needs to be. As will be shown later, the novel approach improves prediction by choosing the latter line of reasoning.

3.3. Triple Decomposition

As discussed in Section 2.1, the segregation of the length scales is performed here using triple decomposition. The flow is shown to exhibit vortex shedding and as such the vortex shedding frequency is expected to be the scale responsible for the organised wave motion. The vortex shedding corresponds to a Strouhal Number (St) of 0.145, which is most pronounced in the DNS energy spectra at the location $(x, y) = (10, 2)$ in Figure 5a. The figure also shows that the shedding frequency is not present at locations further downstream, indicating breakdown of the coherent vortices.

Dynamic mode decomposition (DMD) [52] can be used to verify this as well using a section of the flow field downstream of the WGB. Following the mean statistics collection, the DNS was run further to collect instantaneous data (containing the three velocity components). 400 time samples were collected at an interval of 0.5 time units (over 2 flow-through times) such that 29 vortex shedding cycles at a frequency of 0.145 were recorded. Each DMD mode from the collected data corresponds to a fixed frequency and these frequencies (or modes) can be organised according to the normalised energy content ($\|\Phi\|$), as shown in Figure 5b, where only the positive frequencies are shown. As can be seen, the mode corresponding to the frequency of 0.145 has the largest energy content. The collected instantaneous time samples are also used to perform a proper orthogonal decomposition (POD) [53], since the first 2 POD modes are known to be responsible for vortex shedding [54, 55].

A spatial comparison of the first POD mode and the DMD mode corresponding to the vortex shedding frequency is shown in Figure 6. It can be seen that the two modes are spatially alike, which is in agreement with the known literature cited above. With the point of split identified, the Reynolds stress tensor can be split according to Eqn.
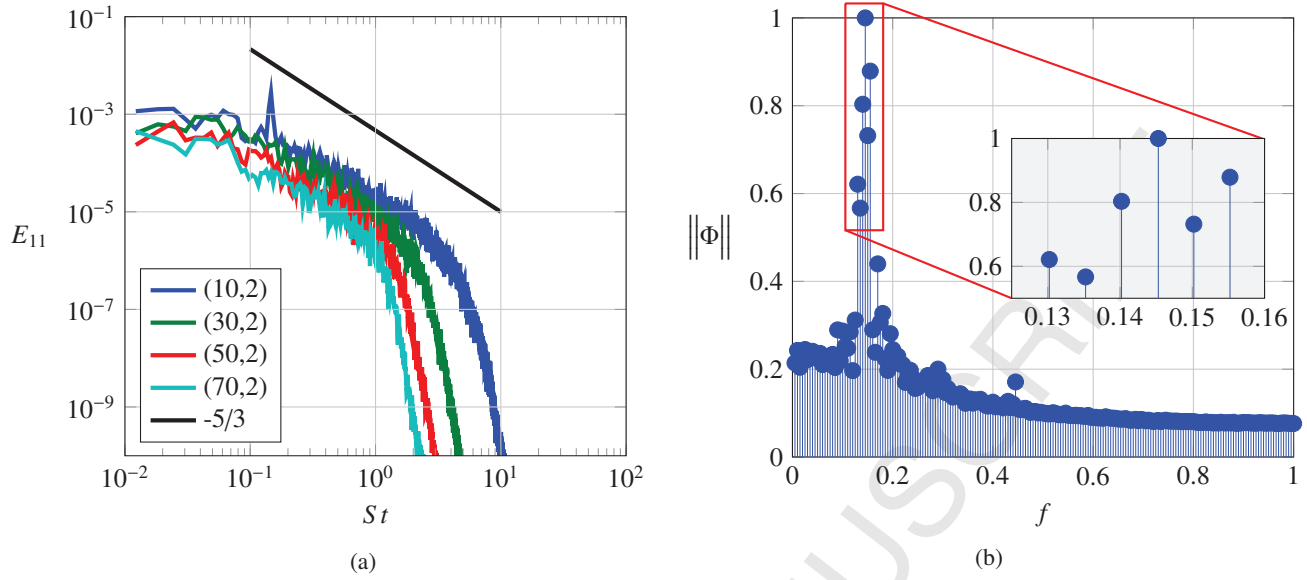
Fig. 5: (a) Turbulence energy spectra at 4 different probe locations (b) energy content in DMD modes v/s the frequencies of each mode.

(7). Two different approaches are used here: FFT and POD. This is done to verify the split stresses from two different methods and also gives an idea about the processing time between them. In the FFT approach, each grid point's velocity time series was transformed to the frequency space where the Fourier coefficients were multiplied with a Gaussian distribution centered around the vortex shedding frequency (f_{vs})

$$H(f) = \exp\left(\frac{-(f - f_{vs})^2}{2\sigma^2}\right). \tag{21}$$

σ represents the extent of frequencies retained by the Gaussian, taken here to be 8 frequencies on either side of f_{vs} spanning 6σ. This series was then inverse transformed to the time domain. A Gaussian distribution was chosen instead of an ideal filter centered at f_{vs} to prevent ringing. This time series is a representation of instantaneous $\widetilde{u}_i$. Consequent spanwise and time averaging of the instantaneous field $\widetilde{u}_i\widetilde{u}_j$ would provide $\widetilde{\tau}_{ij}$.

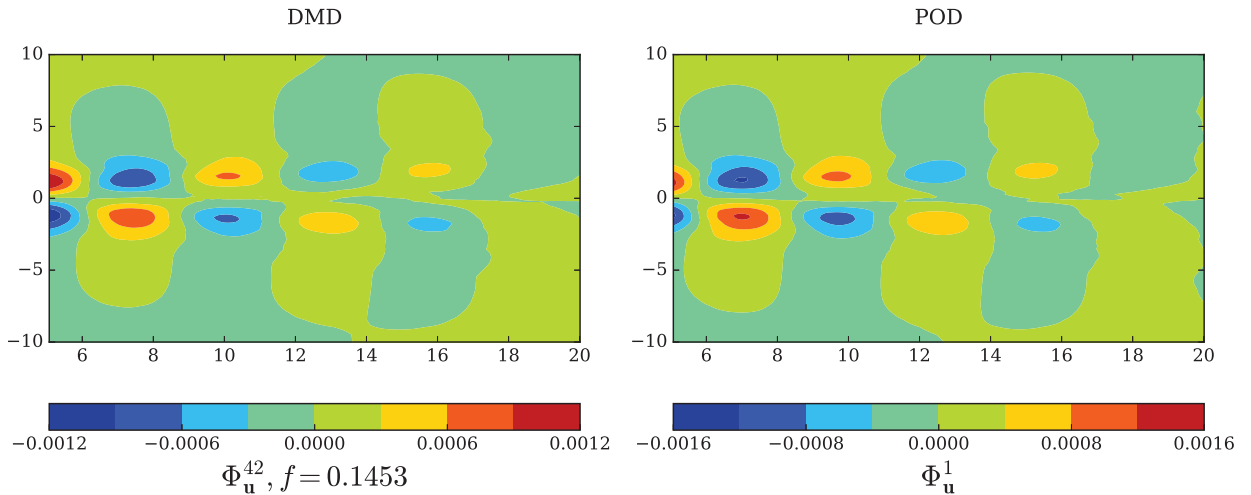Fig. 6: Spatial mode ($x - y$ plane) Φ for u_1 component, corresponding to the vortex shedding mode from DMD and first mode from POD.
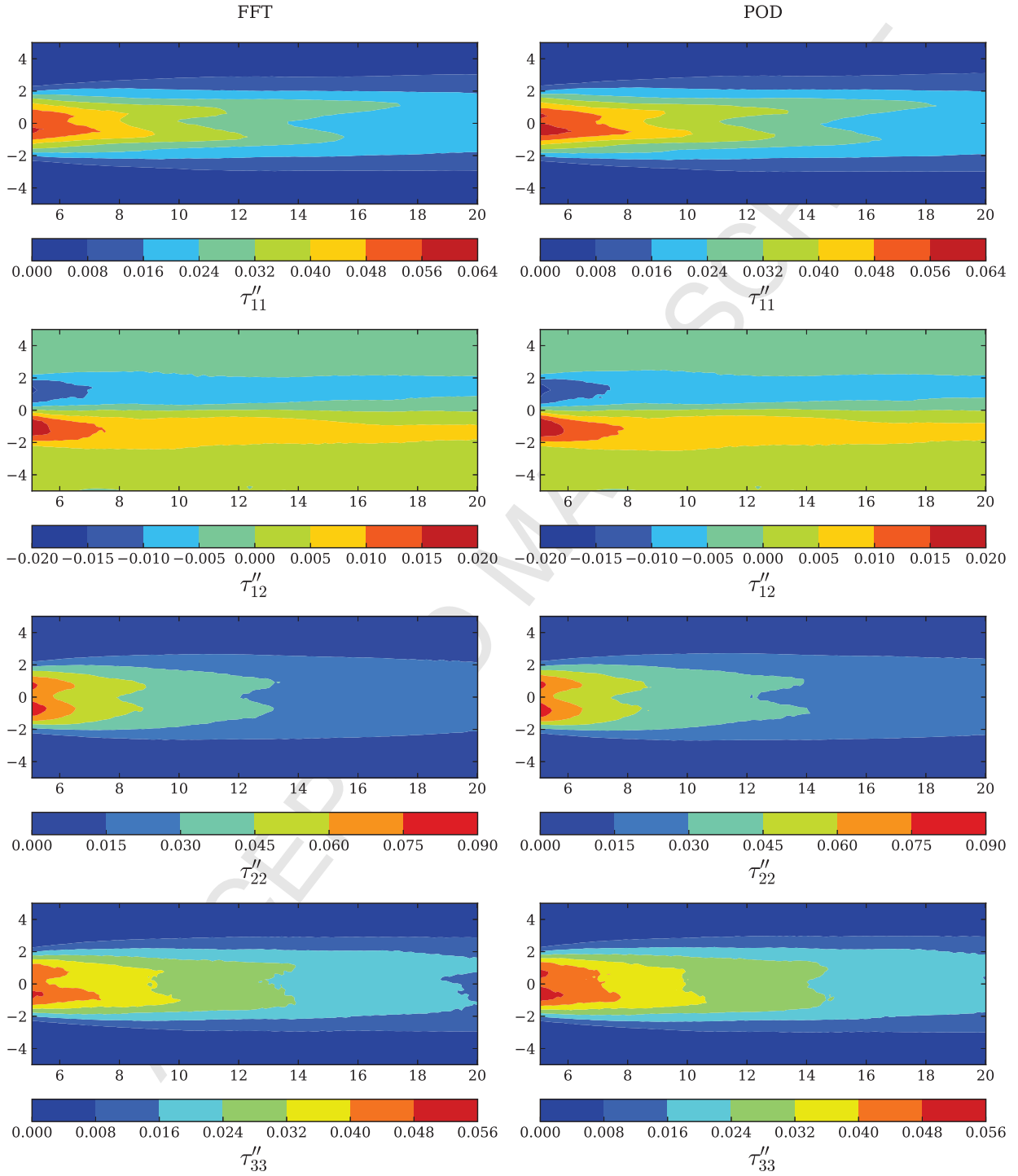
FFT   POD

Fig. 7: Comparison of spatial ($x - y$ plane) τ''_{ij} field obtained through FFT and POD approaches.

τ''_{ij} can be obtained either by subtracting the Gaussian distribution from Fourier coefficients and doing the same as above or by subtracting the $\widetilde{\tau}_{ij}$ field from the time averaged τ_{ij} field collected separately. The latter option is chosen for processing because the τ_{ij} field is collected over a larger number of flow-through times, i.e, 6, whereas the FFT processing is done on data collected over a shorter number of flow-through times, i.e., 2. The $\widetilde{\tau}_{ij}$ field can be created using this shorter time period since we have a sufficient number of vortex shedding cycles (∼29) to average the information.

In the POD approach, the first two POD modes are used to reconstruct the instantaneous velocity, which is again a representation of $\widetilde{u}_i$. Figure 7 shows the spatial distribution of the "turbulent" Reynolds stresses obtained by the two approaches considered. It can be seen that both approaches produce remarkably similar distribution of the stochastic turbulence fields. There are some subtle differences in the absolute values, particularly closer to the WGB, but there is good overall agreement. The differences can be attributed to the choice of the Gaussian distribution in the FFT approach and the first 2 modes in the POD approach. The Gaussian distribution included contribution from other frequencies while the POD modes also include multiple frequencies. This is also true for the periodic fields, which have not been shown here to avoid redundancy. The processing times between the two methods showed that the FFT approach was significantly faster on a single compute node compared with the POD approach (∼ 5 − 10 minutes as opposed to ∼ 3 hours). In the FFT approach, each grid point's time series is subjected to the transformation independently whereas in the POD approach, all grid points are assembled in a matrix and undergo a Singular Value Decomposition, which is more time consuming. Consequently, the FFT approach could realise even faster processing times by simply parallelising the distribution of grid points across multiple compute nodes.

3.4. Selective Scale Resolution

The next step after splitting the field is to obtain the spatially varying f_k^* field. As was discussed in Section 2.2, f_k^* in our approach is obtained from the DNS dataset by computing k''/k, rather than being dependent on the grid resolution of the SRS as is the case for the standard PANS [9]. The resulting field (labelled f_k^*) is given in Figure 8. It can be observed from the figure that the values of f_k^* are smaller than 1.0 closer to the WGB and progressively increase on moving further downstream. This is in line with the observation of breakdown of coherent structures further downstream, signifying increased presence of turbulence. The figure shows only a clipped region because beyond $x = 20.0$, the values of f_k^* are quite close to 1.0. This is readily verified from the energy spectra for locations beyond $x = 20.0$ in Figure 5a, which shows the lack of a peak associated with the vortex shedding frequency. It must be pointed out that this distribution indicates the spatial value of f_k^* at a given location irrespective of the grid resolution at that spatial location. However, when running a PANS calculation with this f_k^*, it is essential to have a grid that can support this spatial distribution, else, the resolution of scales would introduce errors. Figure 8 also shows the distribution of f_k obtained from the baseline URANS calculation, which provides an indication of the values of f_k^* the given RANS grid can support [10]

$$f_k \geq \frac{1}{k\sqrt{C_\mu}}\big(\epsilon\Delta\big)^{2/3}. \tag{22}$$

Here, Δ is the grid cell dimension, chosen to be the maximum cell edge length and $C_\mu = 0.09$. The values for k and ϵ are the mean values taken from the baseline URANS calculation. The figure shows that the chosen grid (672 × 456 grid) is sufficiently well resolved in the wake region to run the f_k^* values obtained from the DNS dataset. Relying explicity on the definition in Eqn. (22) to assess grid quality for a PANS calculation is not recommended though, rather, a separate grid independence study. As will be discussed later, the baseline URANS has a significant overprediction in the TKE, particularly close to the WGB, which can skew the value of f_k. A grid independence study with a fixed spatial distribution of f_k, such as with our f_k^* is a better choice to assess grid quality since any further grid refinements from the optimal grid would not result in more resolution than is allowed by f_k^*. All PANS studies use either a constant value of f_k or a field that is grid dependent through Eqn. (22), whereas this study sets apart the use of PANS through resolving selective scales, obtained from a HiFi dataset. For the given grid, the f_k^* field is fed to the turbulence equations for k_u and ω_u (Eqns. (17), (18)) as an input. The f_k^* field dictates the proportion of length scales to be modelled v/s resolved. The unsteady RANS calculations can be run using $f_k^* = 1.0$ since doing so will recover the original turbulence equations (11), (12). The PANS calculations were run in two and three dimensions as well, using the grid resolutions as the URANS calculation in Section 3.2, since resolving of scales is expected to produce three-dimensional structures, as observed in other wake flow studies [14, 18]. For the three-dimensional calculations, a spanwise extent of 4 units was chosen, with resolutions of 16 and 32 points. The Q-criterion for the three-dimensional field (Figure 4b) shows a lack of spanwise instabilities, which would've produced a breakdown
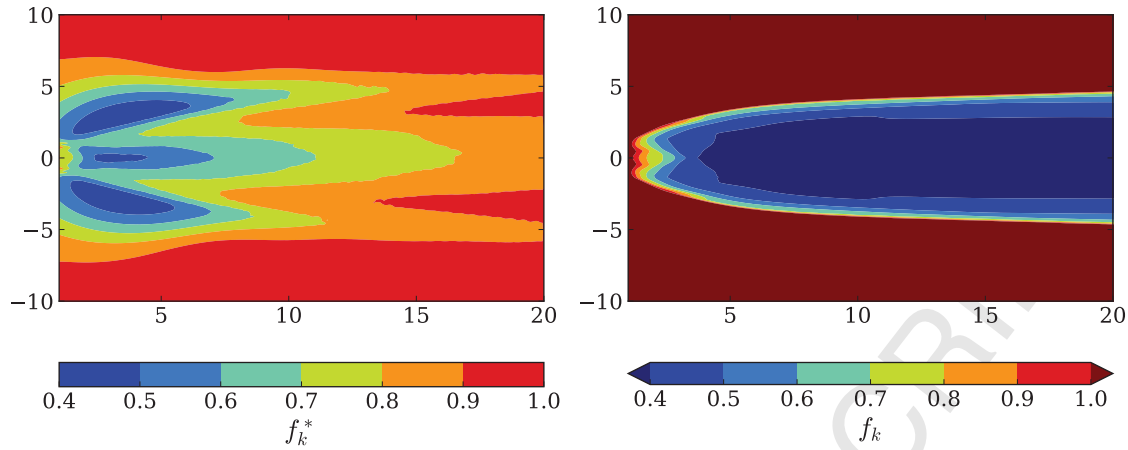
Fig. 8: f_k^* field obtained from the triple decomposition split of the TKE. f_k field obtained from the $k - \omega$ SST model by Eqn. (22)

of the visible rollers. As a result, the profiles for the two and three-dimensional calculations show an insignificant difference in the mean profiles (not shown here). Comparing with the other PANS studies, it can be pointed out that the resolution of the scales observed there are due to the extremely low values of f_k adopted, i.e. ~ 0.1-0.2, whereas the lowest value of f_k^* in the calculations conducted here was 0.44, which does not produce the expected behaviour. As will be seen in the next section, the Boussinesq approximation is responsible for adding excess diffusion, damping out any instabilities that could've lead to a three-dimensional field. This can be explained by considering the total eddy viscosity (ν_t) and unresolved eddy viscosity (ν_u) which are related as $\nu_u = f_k^{*2} \nu_t$. For a given ν_t, reducing f_k^* reduces the diffusion in the PANS equations quadratically. In relative terms, if the reduction is not significant, resolution of length scales to the desired extent is not achieved and the insufficiently reduced diffusion causes an overprediction of the flow properties. This is observed in the mean velocity plots for the untrained PANS case (Figure 3). Thus, modification of the closure approximation is needed to have the correct balance of scale resolving and modelling.

3.5. Machine Learnt Turbulence Closures

Utilizing the machine learning approach of Weatheritt & Sandberg [33], new closures are developed from the DNS dataset. Previous work has shown that the obtained closure can depend on the extent of the spatial domain chosen [33, 34]. The purpose of this work is not to conduct a sensitivity study of the training region but to bring attention to a
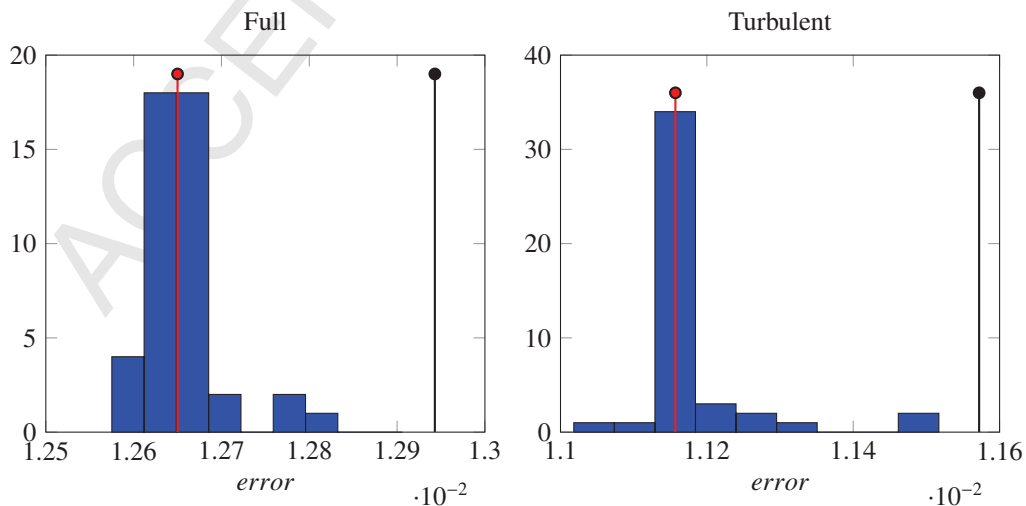

Fig. 9: Histogram of 50 full and turbulent anisotropy models. Red: model from arithmetic average of models lying within most frequently occuring bracket. Black: fitness of the Boussinesq approximation model. The lines represent the fitness value and not the frequency.
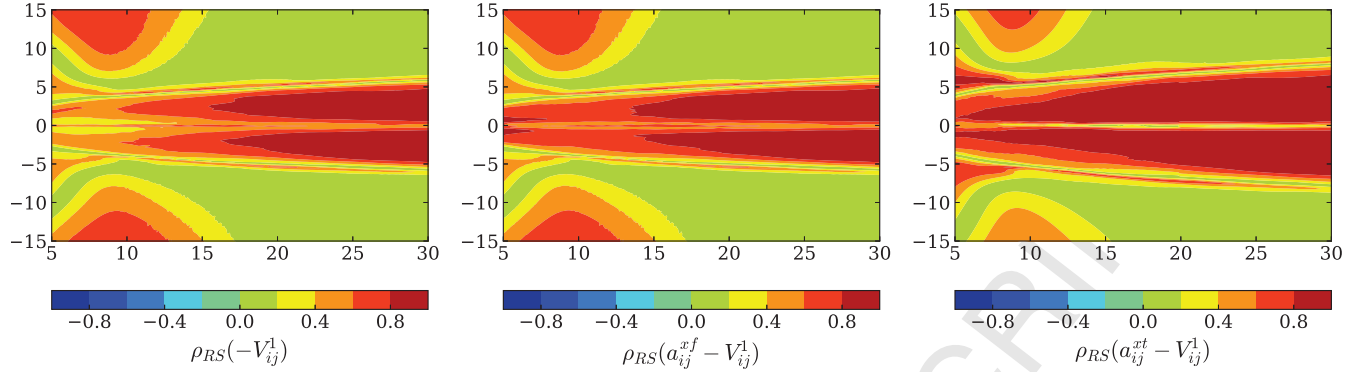
Fig. 10: Alignment plots ($x - y$ plane) for the Boussinesq approximation, "full" anisotropy model and "turbulent" anisotropy model. The term in the brackets corresponds to a_{ij}^{mod} in Eqn. (3).

novel form of training for unsteady flows. Hence, only one domain is considered for this case study, which spans 5.0 to 65.0 units downstream of the WGB in the x direction and -20.0 to 20.0 units in the y direction. This domain is chosen because it contains the wake in y, is downstream of the mean recirculation bubble (located ~2.2 units downstream) and includes a range of length scales from large ones close to the WGB to small ones far downstream. Two closures are developed for this domain, one using the "full" anisotropy tensor and the other using only the "turbulent" anisotropy tensor. In order to do that, the GEP algorithm is run 50 times with different intialisations for each case and the fitness values for each run are plotted as a histogram in Figure 9. This figure shows the frequency of the commonly occurring models according to their fitness values. Each GEP run is characterised by a different initialised models that have been generated randomly, which provides the machine learning algorithm different start points as it attempts to drive towards an optimised model. The resulting models from the 50 runs are not too dissimilar in the fitness values and the coefficients of the tensor basis functions (Eqn. (5)).

The final closure models for each case are obtained by taking the arithmetic average of all the models that fall within the highest frequency bracket for each case i.e. models within the fitness range of [0.012612, 0.012685] for the full anisotropy and [0.011129, 0.011184] for the "turbulent" anisotropy (Eqns. (23), (24)). The superscripts f, t represent models for full and turbulent cases, also shown in Figure 9.

$$a_{ij}^{xf} = 0.872V_{ij}^1 + (0.0037 + 0.00053I_1 - 0.00028I_2)V_{ij}^2 + 0.032V_{ij}^3. \tag{23}$$

$$a_{ij}^{xt} = 0.869V_{ij}^1 + 0.003V_{ij}^2 + (0.008 - 0.000034I_2)V_{ij}^3. \tag{24}$$

The two models show that the dominant tensor basis function is V_{ij}^1, which corresponds to the mean strain rate (Eqn. (5)). This tensor basis function is responsible for diffusion, with large negative coefficients providing more diffusion. The standard Boussinesq approximation has a coefficient of -1.0 ($-\nu_t S_{ij}/k = -S_{ij}/\omega = -V_{ij}^1$ for the $k - \omega$ SST model) and has a worse fitness value than the averaged models. Since these models are added to the Eqns. (16),(19), which contain the Boussinesq approximation, the effective coefficients for V_{ij}^1 are -0.128 and -0.131 for the "full" and "turbulent" anisotropies. This shows that for both the cases, the actual diffusion is much smaller (~8 times) than when using the Boussinesq approximation.

An assessment of the alignment for the models can be made by using Eqn. (3) whereby a_{ij}^{mod} is replaced by $a_{ij}^{xf} - V_{ij}^1$ and $a_{ij}^{xt} - V_{ij}^1$ for the "full" and "turbulent" anisotropies (see Figure 10). The alignment for the Boussinesq approximation is also shown in Figure 10, where a_{ij}^{mod} is replaced by $-V_{ij}^1$. This figure shows that for regions beyond $x = 20.0$, the Boussinesq approximation is well aligned with the anisotropy. This is understandable since the flow considered exhibits free shear, for which the approximation was developed [56]. However, for $x < 20.0$, the Boussinesq approximation doesn't do as well, with values successively decreasing further upstream. The GEP training process shows an improvement in the alignment for the "full" anisotropy case, with better alignment values upstream. However, the alignment values of 0.7-0.8 in the regions $x = 5.0 - 15.0$ can affect the solution. Comparing this to the "turbulent" anisotropy alignment, there is a much better distribution for $x < 20.0$, which shows that a calculation with
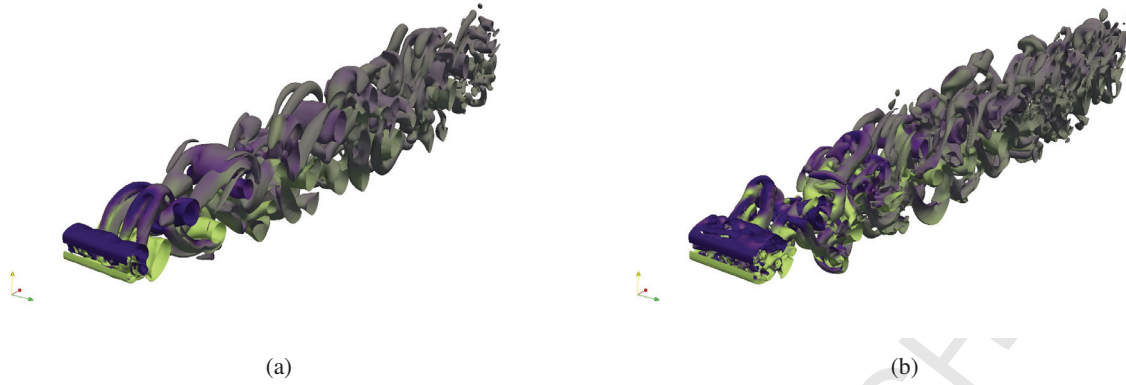
Fig. 11: Q-criterion isosurfaces coloured by the spanwise vorticity for (a) URANS T (b) PANS T

the proposed approach would result in better prediction of the mean flow statistics. The orthogonal alignment in the freestream in all the plots is not a significant issue, given there are no gradients in the freestream.

Based on these *a priori* results, the "turbulent" closure was implemented in the PANS turbulence solver for an *a posteriori* analysis. It must be pointed out that even though the models obtained when training for "full" anisotropy tensor or for the "turbulent" anisotropy tensor are similar, unsteady calculations are performed using only the "turbulent" anisotropy model, for the reason discussed in Section 2.3. Calculations were once more performed in both two and three-dimensions. For the trained calculation with $f_k = 1.0$ everywhere (also referred to as URANS trained), there were significant differences observed between the two and three-dimensional results. The Q-criterion plot in Figure 11a shows presence of some three-dimensional structures, through the breakdown of the spanwise rollers. This three-dimensionality shows significant differences in the mean velocity profiles, as shown in Figure 12 for the two and three-dimensional calculations. The low diffusion from the model results in an overprediction of profiles in two dimensions but an underprediction in three dimensions, which highlights the point discussed in Section 3.2. In two dimensions, resolution of any length scale is not possible and only modelling must represent the mean quantities. A low diffusion coefficient in this context would cause an overprediction of the profiles. The same diffusion coefficient in three dimensions, however, provides room for resolution, which would result in better mean flow profiles. However, further improvements can be sought by specifying the scales that need resolving.

For the trained calculation with the f_k^* distribution (referred to as PANS trained) calculation, spanwise instabilities were observed along with resolution and breakdown of structures (see Figure 11b). Since the trained calculations
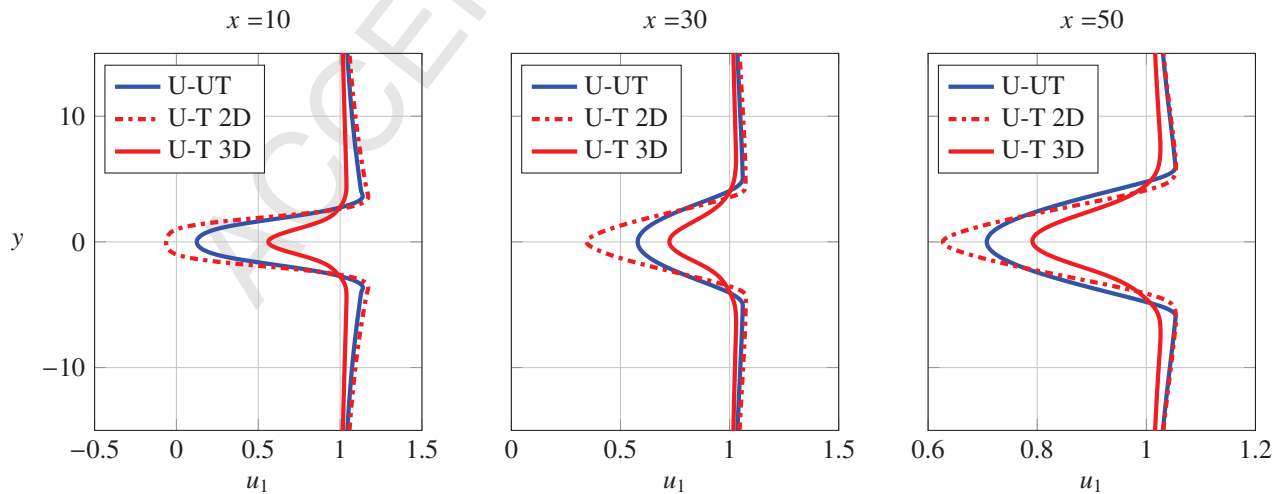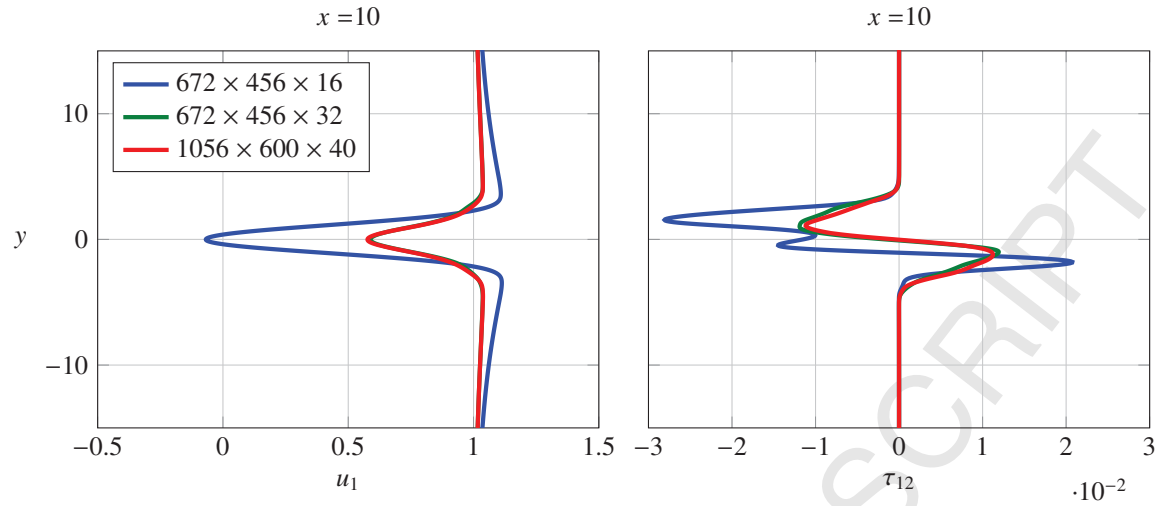

Fig. 12: Two v/s three dimensional comparison for the URANS trained case. Also shown is the baseline URANS untrained result for reference.

Fig. 13: Grid dependence study for PANS Trained case at $x = 10.0$.

exhibited a higher proportion of three-dimensional structures, compared with the untrained calculations performed so far, a grid independence test was conducted by comparing the results using three different grids: $672 \times 456 \times 16$, $672 \times 456 \times 32$ and $1056 \times 600 \times 40$. A grid dependence study is particularly important for the trained PANS case since additional length scales are being resolved and it is important to ensure the grid is capable of supporting such scales. It was discussed in Section 3.4 that the scales being resolved are pre-determined though a fixed spatial f_k^*; as such, grid refinements beyond the optimal resolution should, in principle, produce no change in the mean statistics. Figure 13 shows the comparison of the three different grids considered for the mean velocity and mean shear stress at $x = 10.0$. The plots show that the grid $672 \times 456 \times 32$ is an optimum resolution and further refinement does not change the mean flow, which supports the statement made about the f_k^* field and how it is different from the standard f_k used in other PANS studies. This suggests the model is robust to different resolutions once the optimal resolution is reached, i.e. adapts itself. Using this grid resolution, all other OpenFOAM calculations were conducted and the mean statistics are plotted against the DNS result (see Figure 14).

The grid independent results show the trained calcuations perform significantly better than the baseline URANS. There is a remarkable all round improvement in all the mean flow statistics, which is a strong statement in support of using machine learning for developing better turbulence closures. Between the two trained calculations, the PANS trained does slightly better than the URANS trained, which can be attributed to the low diffusion coefficient allowing a resolution of a proportion of length scales and the f_k^* field dictating the amount of scales needing resolving. Perhaps, better profiles to highlight the difference between these two calcuations are the centerline velocity and TKE profiles (Figure 15).

The trained PANS does better at capturing the recirculation bubble and also has better prediction further downstream. This improvement can be explained by considering how the total Reynolds stresses are obtained through the two approaches. Figure 16 shows the split of the resolved and modelled components of the shear stress and TKE profiles. The actual stress can be obatined by summing the modelled and resolved contributions, which was shown in Figure 14. Results are shown for only one streamwise location ($x = 30.0$) since the trends dicussed here are observed at other locations as well.

The PANS trained case shows a smaller proportion of modelling compared with the URANS trained case. For the normal stresses (and the TKE), the ratio of modelling to resolving is > 1.0 for URANS trained and < 1.0 for PANS trained. However, for the shear stress, the ratio is < 1.0 for both cases, which would explain the significantly better profiles URANS trained generates compared to the baseline. The additional improvement obtained with PANS can be seen by comparing the actual values of the ratios. In case of the TKE, the ratio, for instance at the centerline, is 0.62 and 1.17 for the PANS and URANS trained cases, respectively. In the shear stress, at the location with the maximum magnitudes, the ratios were found to be 0.17 and 0.49 for the PANS and URANS trained cases respectively. This shows that the proposed method of the PANS trained calculation is better at resolving more scales on the same grid, which influences the quality of the solution.
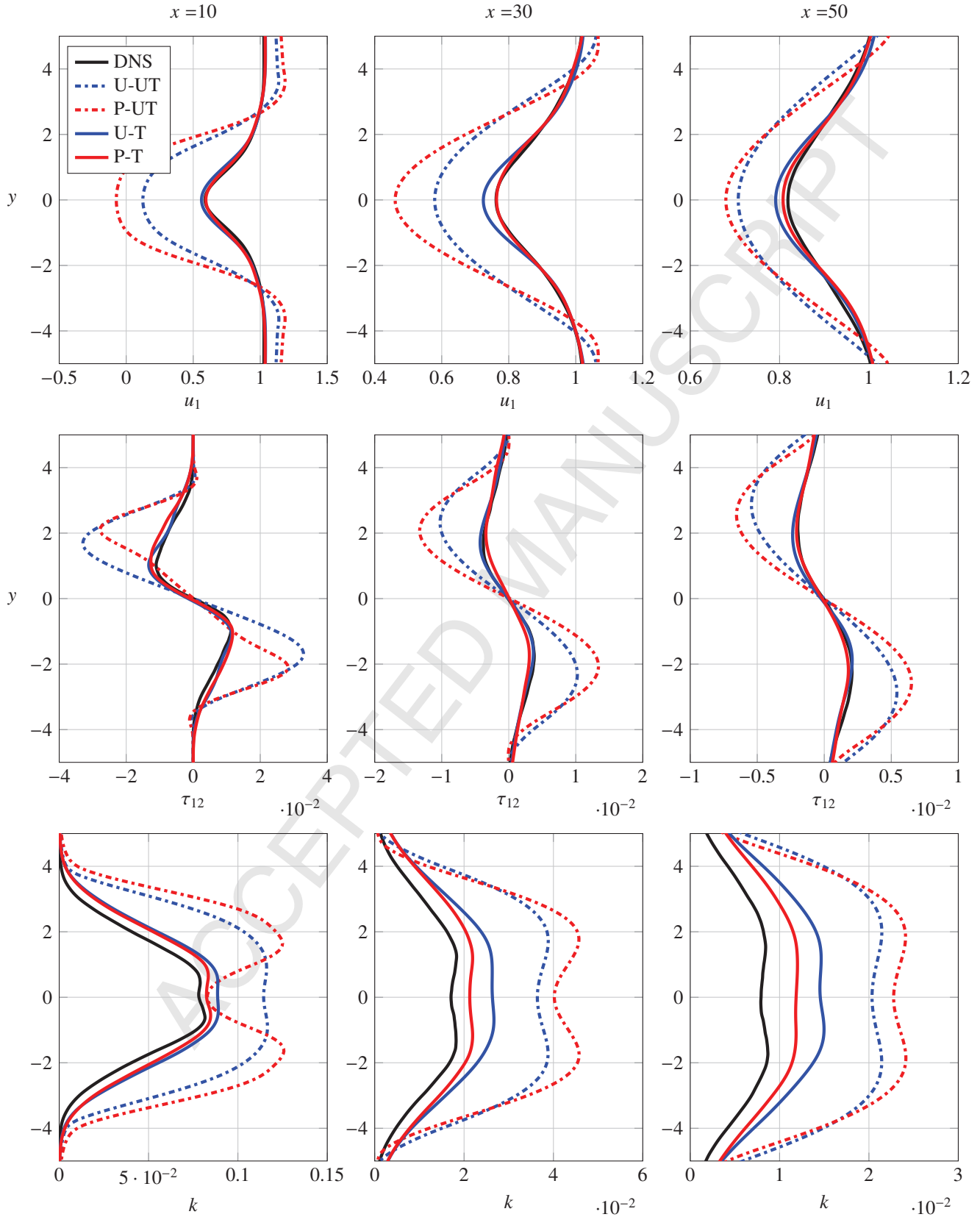
Fig. 14: Mean velocity, shear and turbulence kinetic energy profiles at 3 different locations. U: URANS, P: PANS, UT: Untrained, T: Trained
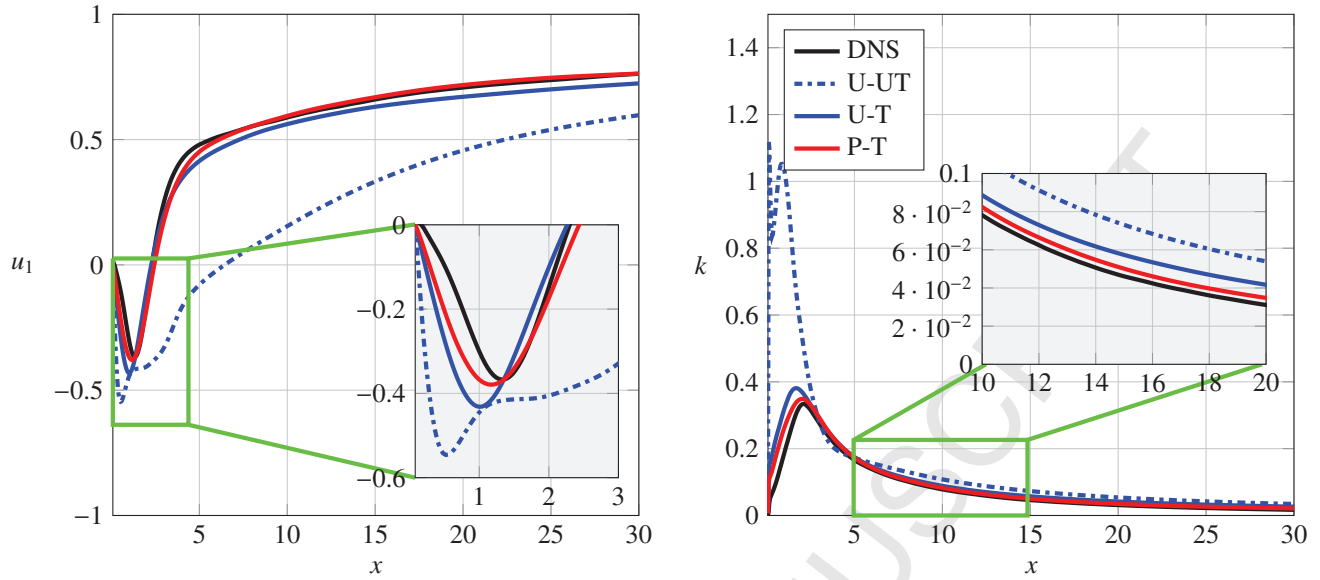
Fig. 15: Mean velocity and turbulence kinetic energy profiles at the centerline. U: URANS, P: PANS, UT: Untrained, T: Trained

With the results obtained, it is also important to identify the computational cost associated with these calculations. The juxtaposition of the computational costs incurred for the different calculations is shown in Table 1. Direct comparison of the cost of the new URANS and PANS approaches implemented in OpenFOAM with the reference simulations using HiPSTAR is not appropriate, since HiPSTAR is a high-fidelity code, purpose-built for optimised running. A more representative cost reduction factor is obtained by comparing with an OpenFOAM LES. This calculation was undertaken to match the HiPSTAR LES in terms of the grid resolution and time step, however, the calculation was not performed to a time-converged solution, instead to estimate the wall time of one flow-through time, which was then extrapolated to the same time interval as the PANS convergence. The extrapolated core-hours shows a reduction of 69.4 times using the proposed approach. Note, however, that the OpenFOAM LES does not guarantee the same level of agreement with the DNS due to the more dissipative numerical scheme. Thus, an accurate LES conducted in OpenFOAM might be of even higher cost and the reduction of cost with the proposed PANS-trained approach even more significant. This shows that the proposed approach is a significantly cheaper scale resolving alternative providing high predictive accuracy.
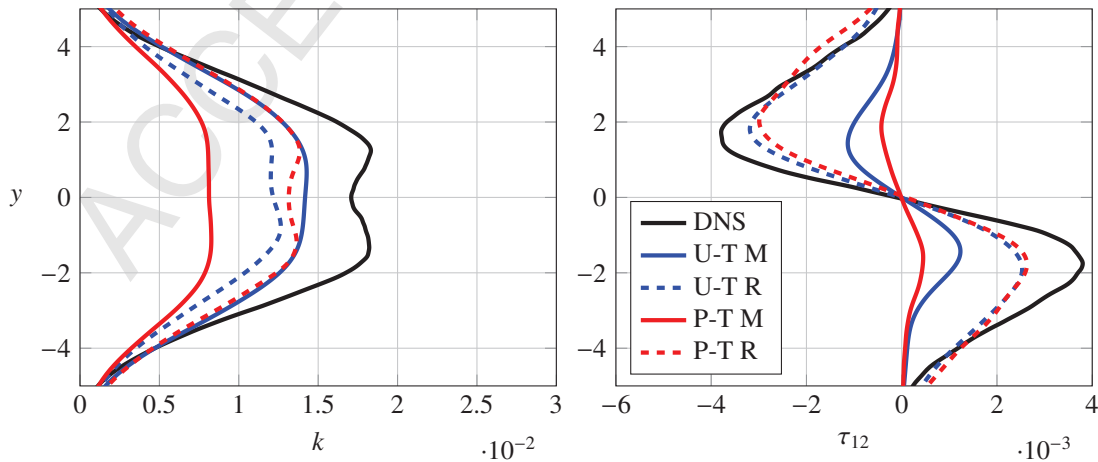

Fig. 16: Mean shear and turbulence kinetic energy profiles at $x = 30.0$. U: URANS, P: PANS, T: Trained, M: Modelled, R: Resolved

Simulation Type	Code	Grid Size	Cores	Hours	Core-Hours
DNS	HiPSTAR	$1056 \times 600 \times 130$	2640	17.5	46,200
LES	HiPSTAR	$672 \times 456 \times 130$	576	13.5	7,760
LES	OpenFOAM	$672 \times 456 \times 128$	576	212	122,112
URANS/PANS	OpenFOAM	$672 \times 456 \times 1$	240	4	960
URANS/PANS	OpenFOAM	$672 \times 456 \times 32$	240	7.3	1,760

Table 1: Computational expense for the different calculations undertaken. The hours represent the time taken for a time-converged solution.

4. Limitations

With the successful demonstration of the proof-of-concept, it is important to consider the limitations of this approach and whether those can be addressed. It is straight-forward to see that this method is dependent on the availability of a high-fidelity dataset, which is not the norm in most scenarios. However, this study was presented as an extension to the work of Weatheritt & Sandberg [33], which also has a dependence on high-fidelity data. Besides, the current crop of turbulence closures and turbulence model coefficients were obtained through calibration with past data, both numerical and experimental. And, as outlined in Duraisamy [32], the purpose of data-driven turbulence modelling is to improve not only the prediction capabilities of RANS models, but also our understanding of the physics the non-linear turbulence closure provides. The case study adopted in this work is a prime example of that, where the balance between f_k^* and reduced diffusion leads to better prediction. The understanding of the physics unearthed by the different calculations undertaken provides us with a framework that should be adopted in the future for flows that exhibit similar behaviour, which brings us to the next advantage of data-driven turbulence modelling. The algorithm produces a tangible equation that gets added to the modelled Reynolds stress equation, resulting in models that are easily re-usable on other similar flows, such as flows with similar geometries, or similar Re ranges. This is because the model, once implemented, can be run as any other model and the solver does not need access to the training data or the training framework. Research on re-usability of the developed models has shown promise with comparable improvements on other similar cases [57]. This offers confidence that the method adopted here should also demonstrate improvements. This is currently under investigation and would be included in future studies.

Conclusions

A new method, Data-driven Stochastic Closure Simulation (DSCS), was presented to improve the prediction accuracy of URANS-type calculations for turbulent flows containing organised unsteadiness. This new method achieves improvement by resolving the length scales that are responsible for the organised unsteadiness, and modelling the remaining length scales. Crucially, the modelling is based on a bespoke non-linear turbulence closure, obtained through a machine-learning algorithm.

A case study, of a normal flat plate at Re = 2,000, was considered to demonstrate the applicability of the method as a proof-of-concept. A baseline URANS of the case study showed an overprediction in the mean profiles of the streamwise velocity component and of the Reynolds stresses. For the new method, three different techniques (DMD, POD and FFT) were contrasted to segregate the vortex shedding length scales and ensure the correct decomposition of the flow field. It was found that the FFT approach was significantly faster than the other two techniques, while essentially producing the same results. Having separated the large-scale unsteadiness, to be resolved, from the total turbulent flow field, an explicit ratio (f_k^*) was extracted from the high-fidelity dataset, the spatial distribution of which dictates the amount of the "turbulent", or stochastic length scales that require modelling. In the near wake ($x = 1.0 - 20.0$), the values of f_k^* were smaller than the values further downstream, which indicated a significant proportion of length scales associated with vortex shedding ($\sim 20 - 60\%$) in this region.

A non-linear turbulence closure could then be developed based on the "turbulent" part of the DNS anisotropy. Using this model closure for URANS calculations of the same Re = 2,000 wake showed a significant improvement in the mean flow statistics compared with the baseline URANS. The developed closure features a significantly reduced diffusion coefficient of the linear term, which when conducting CFD calculations results in less damping of instabilities and hence in more structures being present. Further improvements were obtained when the length scale resolution was explicitly defined using f_k^*. This was achieved by converting the turbulence equations of URANS to

PANS, where the parameter f_k^* serves as a cut-off filter between modelling and resolving. The improvements in the results in the case of the PANS with the non-linear closure calculation over the URANS with the non-linear closure was due to the extra resolution of length scales, on account of the filter f_k^*.

Since this approach is a modification of the machine-learning process for steady flows, there is a dependence of the method on the availability of a high-fidelity dataset. However, the closure obtained with this method can be reused for similar flow configurations (i.e. flows with similar geometries and/or Reynolds numbers). This has been demonstrated successfully in the case of steady flows [33, 34, 39] and is currently under investigation for the current case study.

Acknowledgments

References

[1] A. K. M. F. Hussain, Coherent structures - reality and myth, Phys. Fluids 26 (1983) 2816.
[2] J.-p. Bonnet, J. Delville, Review of Coherent Structures in Turbulent Free Shear Flows and Their Possible Influence on Computational Methods, Flow, Turbul. Combust. 66 (2001) 333–353.
[3] P. G. Tucker, Computation of unsteady turbomachinery flows: Part 2-LES and hybrids, Prog. Aerosp. Sci. 47 (2011) 546–569.
[4] P. G. Tucker, Computation of Unsteady Internal Flows, 16, 1 ed., Kluwer Academic Publishers, 2001.
[5] R. Franke, W. Rodi, Calculation of vortex shedding past a square cylinder with various turbulence models, in: Turbul. Shear Flows 8, Springer, 1993, pp. 189–204.
[6] S. H. Johansson, L. Davidson, E. Olsson, Numerical simulation of vortex shedding past triangular cylinders at high Reynolds number using a k-epsilon turbulence model, Int. J. Numer. Methods Fluids 16 (1993) 859–878.
[7] P. A. Durbin, Separated flow computations with the k-epsilon-v-squared model, AIAA J. 33 (1995) 659–664.
[8] V. Przulj, B. A. Younis, Some aspects of the prediction of turbulent vortex shedding, in: Fluids Eng. Conf., Publ by ASME, 1993.
[9] S. S. Girimaji, Partially-Averaged Navier-Stokes Model for Turbulence: A Reynolds-Averaged Navier-Stokes to Direct Numerical Simulation Bridging Method, J. Appl. Mech. 73 (2006) 413.
[10] S. S. Girimaji, K. Abdol-Hamid, Partially-Averaged Navier Stokes Model for Turbulence: Implementation and Validation, in: 43rd AIAA Aerosp. Sci. Meet. Exhib., January, 2005, pp. 1–14. URL: http://arc.aiaa.org/doi/abs/10.2514/6.2005-502. doi:10.2514/6.2005-502.
[11] B. Akula, P. Roy, P. Razi, S. Anderson, S. S. Girimaji, Partially-Averaged Navier-Stokes (PANS) Simulations of Lid-Driven Cavity Flow—Part 1: Comparison with URANS and LES, in: Prog. Hybrid RANS-LES Model., volume 130, 2015, pp. 359–369. URL: http://link.springer.com/10.1007/978-3-642-31818-4. doi:10.1007/978-3-642-31818-4.
[12] P. Razi, V. Venugopal, S. Jagannathan, Partially-Averaged Navier-Stokes (PANS) Simulations of Lid-Driven Cavity Flow—Part II: Flow Structures, in: Prog. Hybrid RANS-LES Model., volume 130, 2015, pp. 421–430. URL: http://link.springer.com/10.1007/978-3-642-31818-4. doi:10.1007/978-3-642-31818-4.
[13] D. Luo, C. Yan, H. Liu, R. Zhao, Comparative assessment of PANS and DES for simulation of flow past a circular cylinder, J. Wind Eng. Ind. Aerodyn. 134 (2014) 65–77.
[14] P. Ranjan, A. Dewan, Partially Averaged Navier Stokes simulation of turbulent heat transfer from a square cylinder, Int. J. Heat Mass Transf. 89 (2015) 251–266.
[15] F. S. Pereira, G. Vaz, L. Eca, S. S. Girimaji, Simulation of the flow around a circular cylinder at Re=3900 with Partially-Averaged Navier-Stokes equations, Int. J. Heat Fluid Flow 69 (2018) 234–246.
[16] B. Basara, S. Krajnovic, S. S. Girimaji, PANS vs. LES for computations of the flow around a 3D bluff body, in: Proc. ERCOFTAC 7th Int. Symp., volume 2-3, Cyprus, 2008, pp. 548–554.
[17] B. Basara, S. Krajnovic, S. S. Girimaji, Z. Pavlovic, Near-Wall Formulation of the Partially Averaged Navier Stokes Turbulence Model, AIAA J. 49 (2011) 2627–2636.
[18] S. Lakshmipathy, V. Togiti, Assessment of alternative formulations for the specific-dissipation rate in RANS and variable-resolution turbulence models, in: 20th AIAA Comput. Fluid Dyn. Conf., 2011, pp. 1–20. URL: http://arc.aiaa.org/doi/abs/10.2514/6.2011-3978. doi:10.2514/6.2011-3978.
[19] J. Slotnick, A. Khodadoust, J. Alonso, D. Darmofal, W. Gropp, E. Lurie, D. Mavriplis, CFD Vision 2030 Study: A Path to Revolutionary Computational Aerosciences, Technical Report March, NASA, Langley Research Center, Hampton, Virginia, 2014. URL: http://ntrs.nasa.gov/search.jsp?R=20140003093. doi:10.1017/CBO9781107415324.004.
[20] B. D. Tracey, K. Duraisamy, J. J. Alonso, A Machine Learning Strategy to Assist Turbulence Model Development, 53rd AIAA Aerosp. Sci. Meet. (2015) 1–23.
[21] D. Wilcox, Turbulence modeling for CFD, volume 93, 2 ed., Canada, 1998.
[22] M. J. Tummers, D. M. Passchier, R. A. W. M. Henkes, Experimental investigation of an adverse pressure gradient wake and comparison with calculations, Exp. Therm. Fluid Sci. 14 (1997) 17–24.
[23] R. Hoffenberg, J. Sullivan, Measurement and simulation of wake deceleration, 36th AIAA Aerosp. Sci. Meet. Exhib. (1998) 1–10.
[24] J. R. Carlson, N. Duquesne, C. L. Rumsey, T. B. Gatski, Computation of turbulent wake flows in variable pressure gradient, Comput. Fluids 30 (2001) 161–187.

[25] M. J. Tummers, K. Hanjalić, D. M. Passchier, R. A. W. M. Henkes, Computations of a turbulent wake in a strong adverse pressure gradient, Int. J. Heat Fluid Flow 28 (2007) 418–428.

[26] J. D. Denton, Some Limitations of Turbomachinery CFD, ASME Turbo Expo 2010 Power Land, Sea Air (2010) 1–11.

[27] R. Pichler, J. Kopriva, G. Laskowski, V. Michelassi, R. D. Sandberg, Highly Resolved LES of a Linear HPT Vane Cascade Using Structured and Unstructured Codes, Turbomach. Tech. Conf. Expo. GT2016 (2016) 1–10.

[28] F. G. Schmitt, About Boussinesq's turbulent viscosity hypothesis: historical remarks and a direct evaluation of its validity, Comptes Rendus - Mec. 335 (2007) 617–627.

[29] S. B. Pope, A more general effective-viscosity hypothesis, J. Fluid Mech. 72 (1975) 331–340.

[30] W. Rodi, A new algebraic relation for calculating the Reynolds stresses, in: Gesellschaft Angew. Math. und Mech. Work. Paris Fr., volume 56, 1976.

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

[32] K. Duraisamy, G. Iaccarino, H. Xiao, Turbulence Modeling in the Age of Data, Annu. Rev. Fluid Mech. (2018) 1–23.

[33] J. Weatheritt, R. D. Sandberg, A novel evolutionary algorithm applied to algebraic modifications of the RANS stress-strain relationship, J. Comput. Phys. 325 (2016) 22–37.

[34] J. Weatheritt, R. D. Sandberg, G. Laskowski, V. Michelassi, Machine Learning For Turbulence Model Development Using A High-Fidelity HPT Cascase Simulation, in: Turbomach. Tech. Conf. Expo. GT2017-63497, Charlotte, North Carolina, USA, 2017, pp. 1–12.

[35] A. K. M. F. Hussain, W. C. Reynolds, The mechanics of an organized wave in turbulent shear flow, J. Fluid Mech. 41 (1970) 241–258.

[36] W. C. Reynolds, A. K. M. F. Hussain, The mechanics of an organized wave in turbulent shear flow. Part 3. Theoretical models and comparisons with experiments, J. Fluid Mech. 54 (1972) 263–288.

[37] F. R. Menter, M. Kuntz, R. Langtry, Ten years of industrial experience with the SST turbulence model, Turbul. Heat Mass Transf. 4 (2003) 625–632.

[38] J. Weatheritt, R. D. Sandberg, J. Ling, G. Saez, J. Bodart, A Comparative Study of Contrasting Machine Learning Frameworks Applied to RANS Modeling of Jets in Crossflow, in: Turbomach. Tech. Conf. Expo. GT2017-63403, Charlotte, North Carolina, USA, 2017, pp. 1–12. URL: http://dx.doi.org/10.1115/GT2017-63403.

[39] R. D. Sandberg, R. Tan, J. Weatheritt, A. Ooi, A. Haghiri, V. Michelassi, G. Laskowski, Applying Machine Learnt Explicit Algebraic Stress and Scalar Flux Models To a Fundamental Trailing Edge Slot, J. Turbomach. 140 (2018) 1–13.

[40] F. Pereira, G. Vaz, L. Eca, An assessment of Scale-Resolving Simulation models for the flow around a circular cylinder, Turbul. Mass Heat Transf. 8 (2015) 1–17.

[41] R. D. Sandberg, V. Michelassi, R. Pichler, L.-w. Chen, R. Johnstone, Compressible direct numerical simulation of low-pressure turbines: part I - methodology, J. Turbomach. 137 (2015) 051011–1–10.

[42] J. W. Kim, D. J. Lee, Generalized Characteristic Boundary Conditions for Computational Aeroacoustics, AIAA J. 38 (2000) 2040–2049.

[43] J. W. Kim, D. J. Lee, Generalized Characteristic Boundary Conditions for Computational Aeroacoustics, Part 2, AIAA J. 42 (2004) 47–57.

[44] R. D. Sandberg, N. D. Sandham, Nonreflecting Zonal Characteristic Boundary Condition for Direct Numerical Simulation of Aerodynamic Sound, AIAA J. 44 (2006) 402–405.

[45] V. D. Narasimhamurthy, H. I. Andersson, Numerical simulation of the turbulent wake behind a normal flat plate, Int. J. Heat Fluid Flow 30 (2009) 1037–1043.

[46] A. Hemmati, D. H. Wood, R. J. Martinuzzi, Characteristics of distinct flow regimes in the wake of an infinite span normal thin flat plate, Int. J. Heat Fluid Flow 62 (2016) 423–436.

[47] J. Jeong, F. Hussain, On the identification of a vortex, J. Fluid Mech. 285 (1995) 69–94.

[48] F. Nicoud, F. Ducros, Subgrid-Scale Stress Modelling Based on the Square of the Velocity Gradient Tensor, Flow, Turbul. Combust. 62 (1999) 183–200.

[49] C. Mannini, A. Šoda, G. Schewe, Unsteady RANS modelling of flow past a rectangular cylinder: Investigation of Reynolds number effects, Comput. Fluids 39 (2010) 1609–1624.

[50] W. Rodi, Comparison of LES and RANS calculations of the flow around bluff bodies, J. Wind Eng. Ind. Aerodyn. 69-71 (1997) 55–75.

[51] W. Wienken, J. Stiller, A. Keller, A method to predict cavitation inception using large-eddy simulation and its application to the flow past a square cylinder, J. Fluids Eng. 128 (2006) 316–325.

[52] P. J. Schmid, Dynamic mode decomposition of numerical and experimental data, J. Fluid Mech. 656 (2010) 5–28.

[53] G. Berkooz, P. Holmes, J. L. Lumley, The Proper Orthogonal Decomposition in the Analysis of Turbulent Flows, Annu. Rev. Fluid Mech. 25 (1993) 539–575.

[54] R. Perrin, M. Braza, E. Cid, S. Cazin, A. Barthet, A. Sevrain, C. Mockett, F. Thiele, Obtaining phase averaged turbulence properties in the near wake of a circular cylinder at high Reynolds number using POD, Exp. Fluids 43 (2007) 341–355.

[55] Q. Zhang, Y. Liu, S. Wang, The identification of coherent structures using proper orthogonal decomposition and dynamic mode decomposition, J. Fluids Struct. 49 (2014) 53–72.

[56] L. Michael, Statistical turbulence modelling for fluid dynamics-demystified: an introductory text for graduate engineering students, World Scientific, 2015.

[57] J. Weatheritt, R. D. Sandberg, The development of algebraic stress models using a novel evolutionary algorithm, Int. J. Heat Fluid Flow 68 (2017) 298–318.