

Reynolds-averaged Navier–Stokes equations with explicit data-driven Reynolds stress closure can be ill-conditioned

Jinlong Wu¹, Heng Xiao^{1,†}, Rui Sun¹ and Qiqi Wang²

¹Kevin T. Crofton Department of Aerospace and Ocean Engineering, Virginia Tech, Blacksburg, VA 24060, USA

²Department of Aeronautics and Astronautics, MIT, Cambridge, MA 02139, USA

(Received 3 September 2018; revised 6 December 2018; accepted 10 March 2019; first published online 29 April 2019)

Reynolds-averaged Navier–Stokes (RANS) simulations with turbulence closure models continue to play important roles in industrial flow simulations. However, the commonly used linear eddy-viscosity models are intrinsically unable to handle flows with non-equilibrium turbulence (e.g. flows with massive separation). Reynolds stress models, on the other hand, are plagued by their lack of robustness. Recent studies in plane channel flows found that even substituting Reynolds stresses with errors below 0.5 % from direct numerical simulation databases into RANS equations leads to velocities with large errors (up to 35 %). While such an observation may have only marginal relevance to traditional Reynolds stress models, it is disturbing for the recently emerging data-driven models that treat the Reynolds stress as an explicit source term in the RANS equations, as it suggests that the RANS equations with such models can be ill-conditioned. So far, a rigorous analysis of the condition of such models is still lacking. As such, in this work we propose a metric based on local condition number function for *a priori* evaluation of the conditioning of the RANS equations. We further show that the ill-conditioning cannot be explained by the global matrix condition number of the discretized RANS equations. Comprehensive numerical tests are performed on turbulent channel flows at various Reynolds numbers and additionally on two complex flows, i.e. flow over periodic hills, and flow in a square duct. Results suggest that the proposed metric can adequately explain observations in previous studies, i.e. deteriorated model conditioning with increasing Reynolds number and better conditioning of the implicit treatment of the Reynolds stress compared to the explicit treatment. This metric can play critical roles in the future development of data-driven turbulence models by enforcing the conditioning as a requirement on these models.

Key words: computational methods, turbulence modelling, turbulence simulation

1. Introduction

Reynolds-averaged Navier–Stokes (RANS) simulations play an important role in industrial simulations of turbulent flows. The state-of-the-art eddy-viscosity models,

† Email address for correspondence: hengxiao@vt.edu

e.g. the k - ε , k - ω and Spalart–Allmaras (SA) models (Launder & Sharma 1974; Wilcox 1988; Spalart & Allmaras 1994; Menter 1994), are based on the assumption that the turbulence production and dissipation are in equilibrium, and thus they perform poorly in flows with non-equilibrium turbulence (Speziale & Xu 1996; Hamlington & Dahm 2008; Hamlington & Ihme 2014), e.g. flows with massive separations or abrupt mean flow changes. On the other hand, Reynolds stress models (RSMs), also referred to as second-moment closures, take into account the transport of Reynolds stresses and thus have better performance than eddy-viscosity models for flows with non-equilibrium turbulence (Pope 2000). NASA's *CFD Vision 2030 Study* white paper identified advanced turbulence modelling based on Reynolds stress models as a priority for aeronautic technological advancement in the coming decades (Slotnick *et al.* 2014). However, so far Reynolds stress models have not been widely used in engineering applications despite their theoretical superiority. A key shortcoming of Reynolds stress models is their lack of stability and robustness, i.e. the difficulty in achieving convergence (Basara & Jakirlic 2003; Maduta & Jakirlic 2017) and the high sensitivity to the modelling of unclosed terms (particularly the pressure–rate-of-strain tensor) in the Reynolds stress transport equations.

Among the leading causes of numerical instability of Reynolds stress models is the intricate coupling as dictated by the Reynolds stress transport equations (Pope 2000, chap. 7). First, the shear stress τ_{xy} is responsible for the production of streamwise fluctuations τ_{xx} , where x and y denote streamwise and wall-normal coordinates, respectively. Then, the turbulent kinetic energy in τ_{xx} is redistributed by the pressure–rate-of-strain to the other two normal components of the Reynolds stress tensor, including τ_{yy} , which in turn generates the turbulent shear stress τ_{xy} through interactions with the mean strain field. Another possible cause for the numerical instability is the ill-conditioning of the RANS momentum equations themselves. Here model conditioning refers to the sensitivity of the solved quantities (e.g. mean velocity and pressure fields) to the modelled terms (Reynolds stresses). In computational fluid dynamics codes that solve RANS equations and turbulence transport equations in a segregated manner, ill-conditioned systems lead to numerical instability due to the sensitivity of solved velocities to residuals in the Reynolds stress from iteration to iteration. While the chain coupling mechanism described above is well known in the turbulence modelling literature, the conditioning of RANS equations has rarely been mentioned. This is probably due to the fact that the two causes are closely intertwined in traditional models based on Reynolds stress transport equations and the former dominated.

1.1. *Unique challenges in data-driven Reynolds stress closures*

Recently, data-driven turbulence modelling has emerged as a promising field of research with a number of approaches having been proposed in the past few years (e.g. Ling, Kurzawski & Templeton 2016; Singh & Duraisamy 2016; Weatheritt & Sandberg 2016; Wang, Wu & Xiao 2017). Of particular relevance to the present discussions are the data-driven Reynolds stress closures (Ling *et al.* 2016; Wang *et al.* 2017; Geneva & Zabarar 2019) in which the Reynolds stresses are obtained directly from machine learning models trained on high-fidelity simulation databases without solving partial differential equations (PDEs) or using explicit algebraic models. In such models, the chain coupling mechanism related to the Reynolds stress transport equations as described above is not a relevant cause of numerical instability, and thus the possible ill-conditioning of RANS equations is placed under the spotlight.

Moreover, these data-driven Reynolds stress models do not have explicit expressions for the Reynolds stress (Ling *et al.* 2016; Wang *et al.* 2017), which makes it difficult to treat the Reynolds stresses implicitly in the RANS equations to improve model conditioning. For example, Wang *et al.* (2017) used random forest regression to predict Reynolds stresses based on direct numerical simulation (DNS) databases and reported that the solved mean velocity field does not improve over the original RANS simulations, even though the predicted Reynolds stresses showed significant improvements. Such a paradox clearly demonstrates the gap between *a priori* and *a posteriori* performances in turbulence models based on data-driven Reynolds stress closures. That is, an ill-conditioned model can amplify small *a priori* errors in the modelled terms to large *a posteriori* errors in the solved quantities, which defeats all efforts in the improvement of the closure models. Similar gaps have been observed in data-driven subgrid-scale (SGS) stress models for large-eddy simulations (LES). For example, Gamahara & Hattori (2017) reported that their neural network model predicted better SGS stresses than the Smagorinsky model for turbulent channel flows, but the mean velocity predictions were much less satisfactory. Given the large number of efforts in developing data-driven Reynolds stresses model (e.g. Ling *et al.* 2016; Wang *et al.* 2017; Zhu *et al.* 2019) and SGS models (e.g. King, Hamlington & Dahm 2016; Maulik & San 2017; Vollant, Balarac & Corre 2017; Maulik *et al.* 2019) for turbulent flows, the conditioning in such models must be closely examined along with their *a priori* predictive performances. Note, however, that some data-driven approaches, e.g. those based on correcting turbulence transport equations (Parish & Duraisamy 2016; Singh, Medida & Duraisamy 2017) or discovering analytical turbulent constitutive relations by using symbolic regression (Weatheritt & Sandberg 2016), are less affected by the conditioning discussed here.

1.2. DNS data as the ideal scenario for data-driven Reynolds stress closures

Using Reynolds stress obtained from DNS data to solve the RANS equations for velocity can be considered the ideal scenario for data-driven Reynolds stress closure, at least for those without analytical expressions. Solving for mean velocities with a given Reynolds stress field is referred to as ‘propagation’ in this work, i.e. propagation of Reynolds stresses to mean velocities by solving the RANS equations. Such a methodology represents an absolute upper limit of performances for data-driven Reynolds stress models as pursued by Ling *et al.* (2016), Wang *et al.* (2017) and Zhu *et al.* (2019). It allows us to concentrate on the conditioning of this class of machine-learning-based Reynolds stress models without considering the *a priori* performance of any specific model.

Somewhat surprisingly, even solving RANS equations with Reynolds stresses from DNS data can lead to large errors in the velocities, which has been demonstrated in the recent work of Thompson *et al.* (2016). They propagated Reynolds stresses obtained from DNS to mean velocities for turbulent channel flows at a wide range of Reynolds numbers. These DNS were performed with extreme caution by reputable groups (Del Alamo & Jiménez 2003; Bernardini, Pirozzoli & Orlandi 2014; Lee & Moser 2015), and it was verified that the errors in the reported Reynolds stresses were indeed very small, typically less than 0.5% as shown in table 1. Thompson *et al.* (2016) showed that the solved mean velocity has an unsatisfactory agreement with the DNS data at high frictional Reynolds numbers (e.g. $Re_\tau = 5200$). Poroseva *et al.* (2016) also confirmed such observations. To motivate our work, we reproduced the two studies of solving for mean velocities by using the DNS Reynolds stresses

Frictional Reynolds number (Re_τ)	180	550	1000	2000	5200
Error in turbulent shear stresses (%)					
Volume averaged	0.17	0.21	0.03	0.15	0.31
Maximum	0.43	0.38	0.07	0.23	0.41
Errors in mean velocities (%)					
Volume averaged	0.25	1.61	0.17	2.85	21.6
Maximum	0.36	2.70	0.25	5.48	35.1

TABLE 1. Summary of the results of the channel flow test, showing percentage errors in the turbulent shear stresses and the propagated mean velocities. The large errors in the high-Reynolds-number case $Re_\tau = 5200$ are highlighted in bold. The true Reynolds shear stresses are obtained by analytically integrating the RANS equation with the DNS mean velocities (Thompson *et al.* 2016).

obtained from Lee & Moser (2015), and the results are summarized in table 1. Although the solved mean velocities shown here are accurate up to $Re_\tau = 1000$, researchers have reported that large errors in the propagated velocities can be found at Reynolds number as low as $Re_\tau = 395$ depending on the DNS data used (S. Poroseva, 2017, personal communication).

The percentage errors shown in table 1 are computed by comparing the DNS Reynolds stresses and the mean velocities propagated therefrom against their respective truths, which are obtained according to Thompson *et al.* (2016). Specifically, the true velocities are taken as the DNS mean velocities, and the true Reynolds shear stresses are computed based on the analytical solution

$$\tau = \left(1 - \frac{y}{h}\right) \tau_w - \rho \nu \frac{dU}{dy} \quad (1.1)$$

from the DNS mean velocities and wall shear stresses τ_w , where y denotes the wall distance ($y = 0$ indicates the wall), h denotes the half channel width, and ρ and ν represent density and kinematic viscosity, respectively.

It can be seen in table 1 that the errors in the Reynolds stresses are less than 0.5%. The errors in the Reynolds stress can be attributed to different sources, e.g. statistical sampling errors (due to inadequate averaging of the DNS results), iterative errors (due to lack of convergence when solving the linear equations in DNS) and interpolation errors (when interpolating the DNS data to the RANS mesh). All these errors are inevitable in any DNS data. It is noted that errors in the Reynolds stresses vary non-monotonically with respect to the Reynolds number. Such a coincidental trend should not be overly or literally interpreted, since the data shown here are a result of complex interactions that lead to superposition and cancellation among various sources as discussed above. It suffices to point out that all the DNS were obtained with extreme caution, as is evident from the very small errors in the Reynolds stresses. Nevertheless, the errors in the mean velocity solved from the Reynolds stresses clearly increase monotonically with the Reynolds number, because they are dominated by the conditioning of RANS equations, as we will show later. Specifically, while the errors in the solved mean velocity are also less than 0.5% at low Reynolds number ($Re_\tau = 180$), such errors can be as high as 35% at high Reynolds number ($Re_\tau = 5200$). This observation suggests that the RANS equations can be ill-conditioned by directly substituting the modelled Reynolds stress into the equations, which is a common

practice in current data-driven RANS modelling. Such results raise several critical questions on Reynolds-stress-based data-driven turbulence models, specifically the following:

- (i) How to explain the deteriorated conditioning (i.e. increased amplification of errors in Reynolds stresses to velocities) with increasing Reynolds numbers observed in these studies?
- (ii) Is there a quantitative metric that can characterize the conditioning of RANS equations with a given turbulence model?
- (iii) What are the implications of the above observations to data-driven turbulence models that treat Reynolds stresses as source terms in the RANS equations?

1.3. Summary and contribution of present work

Our present work aims to answer the above questions by proposing a quantitative metric and elucidating the relevance of the above-mentioned studies to data-driven turbulence modelling. Throughout this study, no turbulence models are used. Rather, Reynolds stresses obtained from DNS databases are used to represent the ideal performance for any data-driven Reynolds stress closures. A local condition number function based on the work of Chandrasekaran & Ipsen (1995) is derived as a metric to assess the conditioning property for turbulence models. Numerical tests on turbulent channel flows at various Reynolds numbers and two more complex flows suggest that the proposed metric can adequately explain observations in previous studies, i.e. deteriorated model conditioning with increasing Reynolds number, and better conditioning of the implicit treatment of the Reynolds stress compared to the explicit treatment. As an application of the proposed approach, Wu, Xiao & Paterson (2018) improved the conditioning of the data-driven Reynolds stress model of Wang *et al.* (2017) by training machine learning models for the linear and nonlinear parts of the Reynolds stress separately. The obtained Reynolds stress model achieved good conditioning and satisfactory predictive accuracy simultaneously.

The rest of this paper is organized as follows. Section 2 introduces the global condition number and shows that it fails to explain the deteriorated conditioning with increasing Reynolds numbers. A local condition number function is derived to achieve such an objective. In §3, the local condition number is used to evaluate the conditioning of RANS equations with data-driven Reynolds stress closures. Section 4 discusses the conditioning of RANS equations with traditional Reynolds stress transport models and the monolithic coupling for both traditional and data-driven models. Finally, conclusions are presented in §5.

2. Methodology

Consider the steady-state RANS equations for incompressible, constant-density fluids:

$$\mathbf{u} \cdot \nabla \mathbf{u} - \nu \nabla^2 \mathbf{u} + \nabla p - \nabla \cdot \boldsymbol{\tau} = 0, \quad (2.1)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (2.2)$$

where \mathbf{u} is the mean flow velocity, ν is molecular viscosity, p is the pressure normalized by the constant density of the fluid, and $\boldsymbol{\tau}$ is the Reynolds stress tensor, which needs to be modelled. For simplicity, we first consider a Reynolds-stress-based model where $\boldsymbol{\tau}$ is obtained by solving a transport equation in a segregated manner

with the RANS equations or by a data-driven function (see e.g. Ling *et al.* 2016). The objective of this work is to investigate the sensitivity of the obtained mean velocity with respect to small perturbations on the Reynolds stress.

For simplicity of notation, we introduce nonlinear operator \mathcal{N} to include the convection and diffusion terms with

$$\mathcal{N}(\mathbf{u}) = \mathbf{u} \cdot \nabla \mathbf{u} - \nu \nabla^2 \mathbf{u}. \quad (2.3)$$

The RANS momentum equation in (2.1) can be written as

$$\mathcal{N}(\mathbf{u}) = \nabla \cdot \boldsymbol{\tau} - \nabla p. \quad (2.4)$$

In numerical solvers, the convection term is first linearized around the current velocity $\overline{\mathbf{U}}_0$ to obtain the linearized RANS equations

$$\mathcal{L}(\mathbf{u}) = \nabla \cdot \boldsymbol{\tau} - \nabla p, \quad (2.5)$$

where \mathcal{L} is the linearized operator of \mathcal{N} , i.e.

$$\mathcal{L}(\mathbf{u}) = \mathbf{u}_0 \cdot \nabla \mathbf{u} - \nu \nabla^2 \mathbf{u}. \quad (2.6)$$

The linearized equation (2.5) is then discretized on a computational fluid dynamics (CFD) mesh to obtain a linear system of the following form:

$$\mathbf{A}\mathbf{U} = \mathbf{b}. \quad (2.7)$$

Here we have denoted $\mathbf{b} = [\nabla \cdot \boldsymbol{\tau} - \nabla p]$ as the imbalance between the two forces, pressure gradient and Reynolds stress divergence; and $\mathbf{U} = [\mathbf{u}]$ is the discretized velocity field to be solved for. Both \mathbf{b} and \mathbf{U} are $n \times 1$ vectors, where n is the number of cells or grid points in the mesh. The matrix \mathbf{A} with dimension $n \times n$ comes from the implicit discretization of the linearized convection term and the molecular diffusion term. In this work, we focus on the conditioning of linearized RANS equations. This is because most CFD codes deal with the linearized RANS equations in each iteration when solving the nonlinear RANS equations. Therefore, the conditioning metrics studied here are valid within each iteration, even though the flow of concern may deviate from the linearized RANS equations.

2.1. Matrix norm as a measure of model conditioning

We first show the derivation of the traditional matrix-norm-based condition number and explain why it fails to distinguish the sensitivities of solving for mean velocity at different Reynolds numbers as shown in table 1. Following the definition of matrix norm, the norm of the error in the velocity is bounded as follows:

$$\frac{\|\delta \mathbf{U}\|}{\|\mathbf{U}\|} \leq \mathcal{K}_{\mathbf{A}} \frac{\|\delta \mathbf{b}\|}{\|\mathbf{b}\|}, \quad (2.8)$$

where

$$\mathcal{K}_{\mathbf{A}} \equiv \|\mathbf{A}\| \|\mathbf{A}^{-1}\| \quad (2.9)$$

denotes the condition number of matrix \mathbf{A} (see e.g. Strang 2016).

As explained in the notation, the norms $\|\overline{\mathbf{U}}\|$ and $\|\mathbf{b}\|$ are taken of the discretized vectors $[\overline{\mathbf{U}}]$ and $[\mathbf{b}]$, respectively, with the brackets inside the norm omitted for

clarity. Such a brief notation does not cause confusion, because norms in this work are always taken for the discretized vectors or matrices with dimensions of $n \times 1$ or $n \times n$, respectively, and never for the velocity or force vectors at any particular location.

Considering that the objective is to assess the effects of Reynolds stress perturbation $\delta\boldsymbol{\tau}$ on the velocities, the inequality in (2.8) above is formulated as

$$\frac{\|\delta\mathbf{U}\|}{\|\mathbf{U}\|} \leq \mathcal{K}_\tau \frac{\|\nabla \cdot \delta\boldsymbol{\tau}\|}{\|\nabla \cdot \boldsymbol{\tau}\|}, \tag{2.10}$$

where

$$\mathcal{K}_\tau = \mathcal{K}_\mathbf{A} \frac{\|\nabla \cdot \boldsymbol{\tau}\|}{\|\mathbf{b}\|}. \tag{2.11}$$

Detailed derivations are omitted here for brevity and are presented in §A.1. It can be seen that the model condition number \mathcal{K}_τ consists of the condition number of the matrix \mathbf{A} and the ratio

$$\bar{\alpha} = \|\nabla \cdot \boldsymbol{\tau}\|/\|\mathbf{b}\|. \tag{2.12}$$

For plane channel flows, the convective term disappears, and thus \mathbf{b} is the force due to the divergence of the viscous stress, i.e. $\mathbf{b} = \nabla \cdot \nu(\nabla\mathbf{u} + (\nabla\mathbf{u})^T) = \nabla \cdot \boldsymbol{\tau}_{vis}$. Consequently, the ratio $\bar{\alpha}$ indicates the overall relative importance of the forces due to Reynolds stress and viscous stress.

The proposed condition number \mathcal{K}_τ based on matrix condition number $\mathcal{K}_\mathbf{A}$ is a natural first attempt in explaining the increasing sensitivity of the velocities to the Reynolds stress with increasing Reynolds numbers as shown in table 1. However, surprisingly, it turns out that the condition number \mathcal{K}_τ is more or less the same across all Reynolds numbers from $Re_\tau = 180$ to 5200, which is shown in figure 1. This observation suggests that the matrix-based condition number \mathcal{K}_τ cannot explain the ill-conditioning of the $Re_\tau = 5200$ case and the better conditioning of the lower-Reynolds-number cases as observed in table 1.

The following two factors explain why the matrix-based condition number \mathcal{K}_τ is almost the same at different Reynolds numbers:

- (i) The matrix condition number $\mathcal{K}_\mathbf{A}$ is constant for all Reynolds numbers, because the matrix \mathbf{A} itself is independent of the Reynolds number.
- (ii) The ratios $\bar{\alpha} = \|\nabla \cdot \boldsymbol{\tau}\|/\|\mathbf{b}\|$ are very similar at vastly different Reynolds numbers, because both norms (which involve integration or square sums) are dominated by the viscous wall regions of each flow. It is well known that the Reynolds number only determines the thickness of this region in outer coordinates, and the Reynolds-number effect is weak here. Each factor above will be detailed as follows.

First, it can be established through simple algebra that for plane channel flows the matrix \mathbf{A} is independent of the Reynolds number but depends on the discretization scheme and the mesh used. Since the convection term disappears for plane channel flows, the matrix \mathbf{A} results solely from the discretization of the diffusion operator $\nu\nabla^2(\cdot)$. When discretized with central difference on a uniform mesh of n cells,

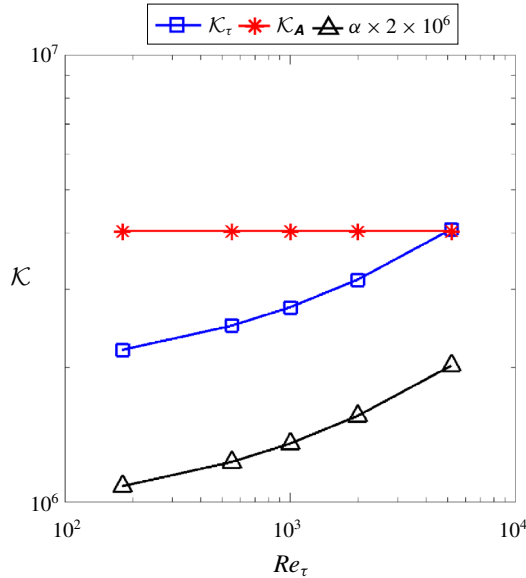


FIGURE 1. (Colour online) Conditioning measure of Reynolds-stress-based turbulence models based on \mathcal{K}_A and the ratio $\bar{\alpha}$ as defined in (2.12). The Reynolds stress is computed from DNS data to study the ideal scenario of the RANS modelling.

matrix \mathbf{A} can be written as follows (Strang 2016):

$$\mathbf{A} = \nu \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & \ddots & \\ & & \ddots & \ddots & -1 \\ & & & -1 & 2 \end{bmatrix}. \tag{2.13}$$

The condition number for matrix \mathbf{A} is $\mathcal{K}_A = 4n^2/\pi^2$. Therefore, \mathcal{K}_A does not explicitly depend on the viscosity or the Reynolds number. This analysis is confirmed by the results shown in figure 1, which shows that \mathcal{K}_A is strictly constant for all five flows at different Reynolds numbers. Moreover, \mathcal{K}_A depends on the mesh size n , which is a critical shortcoming of the matrix-norm-based condition number as a measure of the conditioning property of a turbulence model. To exclude the influences of the mesh, we used the same mesh with 1040 cells in all the flows at different Reynolds numbers presented in figure 1.

Second, the change of Reynolds number has little influence on the ratio $\bar{\alpha}$, as is shown in figure 1. We examine the profile of turbulent shear ($\nabla \cdot \boldsymbol{\tau}$) and viscous shear ($\nabla \cdot \boldsymbol{\tau}_{vis}$) in the channel in figure 2 for the two cases, $Re_\tau = 180$ and 5200. In most of the channel outside the viscous wall region, both forces (and thus the ratio) are fairly uniform. Nevertheless, in the viscous wall region, the two forces are of the same order of magnitude but with opposite signs. In contrast, outside the viscous region, the pressure gradient is the main driving force while the Reynolds shear stress is the resistance, with the viscous shear having negligible effects. The forces in the two distinct regions are illustrated schematically in figure 3. However, when

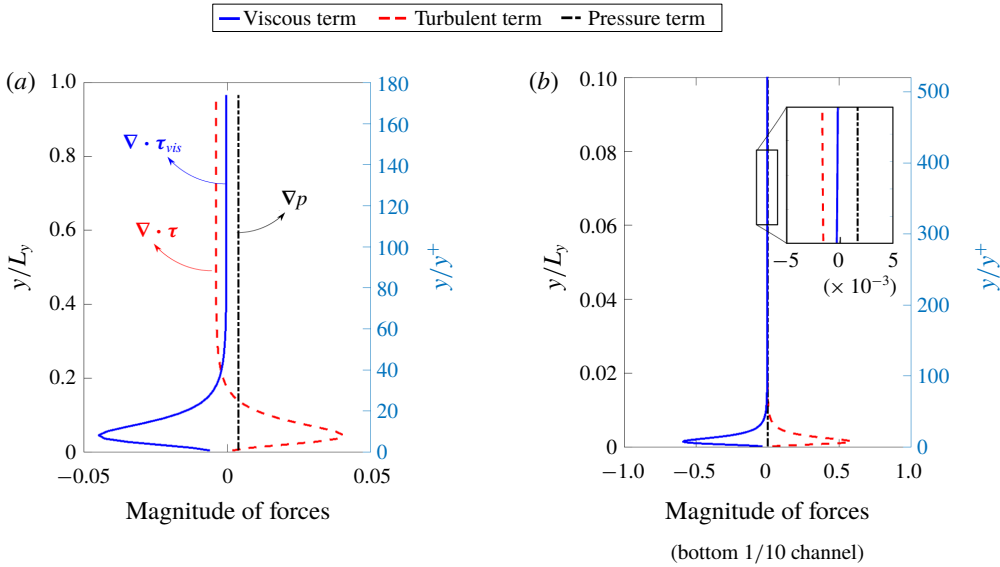


FIGURE 2. (Colour online) The balance among forces due to turbulent shear stress ($\nabla \cdot \boldsymbol{\tau}$), viscous shear stress ($\nabla \cdot \boldsymbol{\tau}_{vis}$) and pressure gradient (∇p) for two plane channel flows at frictional Reynolds numbers (a) $Re_\tau = 180$ and (b) $Re_\tau = 5200$. The right vertical axis denotes the inner coordinates (y^+).

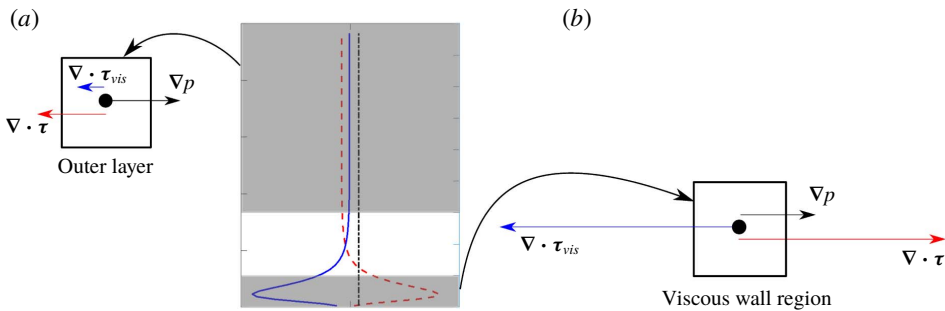


FIGURE 3. (Colour online) Force balance of the plane channel flow in (a) the outer layer and (b) the viscous wall region.

calculating the ratio $\bar{\alpha} = \|\nabla \cdot \boldsymbol{\tau}\|/\|\mathbf{b}\|$ of the two norms, values within the viscous wall region clearly dominate the calculation of both norms, which involve integration of the functions squared. It can be seen that in both figure 2(a) ($Re_\tau = 180$) and figure 2(b) ($Re_\tau = 5200$) the areas enclosed by the blue solid curve and red dashed curve (with the vertical zero line) are similar. Squaring the function places even more weights on the regions of larger function values, i.e. the viscous wall region. This observation suggests that the ratio $\|\nabla \cdot \boldsymbol{\tau}\|/\|\mathbf{b}\|$ is of $O(1)$ for both cases, as confirmed by examining figure 1. Consequently, the computed norm mostly reflects the values of the forces in the viscous region and not the outer layer. It is well known that the Reynolds-number effects are not pronounced within the viscous wall region. Increasing the Reynolds number merely extends the outer layer in terms of inner coordinates ($y^+ = y/y^*$, where $y^* = \nu/\sqrt{\tau_w/\rho}$ is the viscous unit and τ_w is the

wall shear stress). This explains why the ratio $\bar{\alpha}$ does not vary significantly (much less than proportionally) with the Reynolds number as can be seen in figure 1. If the viscous region is neglected, the factor $\bar{\alpha}$ will be higher for the larger Reynolds number, and thus it makes the global condition number a better indicator of model conditioning. However, the matrix condition number \mathcal{K}_A depends on the mesh size and thus the global condition number is still not an ideal choice of evaluating model conditioning of RANS equations.

In summary, the matrix-based condition number \mathcal{K}_τ as derived in (2.11) is not able to explain the increasing sensitivity of the velocities with respect to Reynolds stresses with increasing Reynolds number. In addition, the matrix condition number \mathcal{K}_A has another critical drawback of being mesh-dependent. The mesh dependence is highly undesirable, as the condition number is to measure the conditioning property of turbulence models at the PDE level, not any particular numerical discretization thereof. These drawbacks clearly call for a better metric for measuring the conditioning property of Reynolds-stress-based turbulence models.

2.2. Proposed metric as a measure of model conditioning

In order to address the deficiency of the global condition number as presented in § 2.1, we derive a metric based on a local condition number function to measure the sensitivity of the solved mean velocity \mathbf{u} at a given location \mathbf{x} with respect to perturbation $\delta\boldsymbol{\tau}$ on the Reynolds stress field $\boldsymbol{\tau}$. Such a local condition number is formally defined as the following bound:

$$\frac{|\delta\mathbf{u}(\mathbf{x})|}{U_\infty} \leq \mathcal{K}(\mathbf{x}) \frac{\|\nabla \cdot \delta\boldsymbol{\tau}\|_\Omega}{\|\nabla \cdot \boldsymbol{\tau}\|_\Omega}, \quad (2.14)$$

where $\mathcal{K}(\mathbf{x})$ is the local condition number function defined as

$$\mathcal{K}(\mathbf{x}) = \frac{\|G(\mathbf{x}, \boldsymbol{\xi})\|_\Omega \|\nabla \cdot \boldsymbol{\tau}\|_\Omega}{U_\infty}. \quad (2.15)$$

Function G is the Green's function corresponding to the linear operator \mathcal{L} (see e.g. Lanczos 1996), such that the solution to the linearized RANS equation (2.5) can be written formally as

$$\mathbf{u}(\mathbf{x}) = \mathcal{L}^{-1}[\mathbf{b}(\mathbf{x})] \equiv \int_\Omega G(\mathbf{x}; \boldsymbol{\xi}) \mathbf{b}(\boldsymbol{\xi}) \, d\boldsymbol{\xi}, \quad (2.16)$$

where $\mathcal{L}^{-1}[\cdot]$ is the inverse operator of \mathcal{L} . Green's function $G(\mathbf{x}; \boldsymbol{\xi})$ indicates the contribution of the source $\mathbf{b}(\boldsymbol{\xi})$ at location $\boldsymbol{\xi}$ to the solution \mathbf{u} at location \mathbf{x} . The norm $\|f(\boldsymbol{\xi})\|_\Omega$ of function $f(\boldsymbol{\xi})$ is an integration on domain Ω defined as (Debnath & Mikusiński 2005)

$$\|f(\boldsymbol{\xi})\|_\Omega = \sqrt{\int_\Omega |f(\boldsymbol{\xi})|^2 \, d\boldsymbol{\xi}}. \quad (2.17)$$

The detailed derivations to obtain (2.15) are presented in § A.2.

For functions discretized on a CFD mesh of n cells (e.g. those in RANS simulations), the function norm $\|\cdot\|_\Omega$ in (2.15) can be approximated by the norm of the discretized n -vector through numerical quadrature. That is,

$$\|G(\mathbf{x}, \boldsymbol{\xi})\|_\Omega \approx \|\mathbf{r}_j\|_n \quad \text{and} \quad \|\nabla \cdot \boldsymbol{\tau}\|_\Omega \approx \|[\nabla \cdot \boldsymbol{\tau}]\|_n, \quad (2.18a,b)$$

where \mathbf{r}_j is the j th row of the matrix \mathbf{A}^{-1} . Recall that $[\nabla \cdot \boldsymbol{\tau}]$ indicates discretization of field $\nabla \cdot \boldsymbol{\tau}$ on the CFD mesh, but the bracket can be omitted inside a vector norm $\|\cdot\|_n$ without ambiguity. In this work, the norm $\|\cdot\|_n$ is defined by the L^2 -norm as

$$\|\mathbf{v}\|_n = \left(\sum_i v_i^2 \right)^{1/2} \quad (2.19)$$

and the discretized condition number n -vector corresponding to $\mathcal{K}(\mathbf{x})$ in (2.15) is thus

$$\mathcal{K}_j = \frac{\|\mathbf{r}_j\|_n \|\nabla \cdot \boldsymbol{\tau}\|_n}{U_\infty} \quad \text{with } j = 1, 2, \dots, n, \quad (2.20)$$

which implies that the location \mathbf{x} is the coordinate of the j th cell in the CFD mesh.

The proposed local condition number function $\mathcal{K}(\mathbf{x})$ has two important merits compared to the global matrix-based condition number \mathcal{K}_τ :

- (i) First, $\mathcal{K}(\mathbf{x})$ provides a tighter upper bound than the matrix-norm condition number \mathcal{K}_τ . The main reason is that the upper bound of \mathcal{K}_τ can only be achieved when the following conditions are satisfied simultaneously: (1) the discretized mean velocity field vector \mathbf{U} is aligned with the principal axis of the coefficient matrix \mathbf{A} , and (2) the perturbation vector $\delta\mathbf{b}$ is aligned with the principal axis of \mathbf{A}^{-1} . In contrast, the derivation of $\mathcal{K}(\mathbf{x})$ does not assume any conditions on the discretized mean velocity field \mathbf{U} . Consequently, the bound provided by $\mathcal{K}(\mathbf{x})$ is a more precise assessment of the sensitivity $\delta\mathbf{u}$ with respect to Reynolds stress perturbations.
- (ii) The discretization \mathcal{K}_j of function $\mathcal{K}(\mathbf{x})$ is mesh-independent, which is an important property considering that this metric aims to measure the conditioning property of data-driven Reynolds stress models. Detailed analytical derivations to obtain (2.20) and numerical results (see figure 17 in the appendix) used to demonstrate the mesh independence of the local condition number \mathcal{K}_j are presented in appendix B.

While the local condition metrics proposed here are generally applicable to any linearized PDE and its discretization, it is important to emphasize that these conditioning metrics faithfully embody the physics described by the underlying PDEs. Taking the RANS equations as an example, the mean flow field contributes to the linearized differential operator and thus is reflected in the analysis of local conditioning. Therefore, these local condition metrics are flow-specific properties and thus reflect the mean flow physics. More detailed discussion can be found in §§ 3.2.1 and 3.2.2 on the complex, two-dimensional flows.

Finally, we comment briefly on the numerical implementation and computational complexity of conditioning metrics proposed below. The local condition number \mathcal{K}_j does not require a full inversion of the matrix \mathbf{A} but only needs the j th row of \mathbf{A}^{-1} to be computed. This can be achieved by solving the equation $\mathbf{A}^T \mathbf{r}_j = \mathbf{l}_j$ based on the identity $\mathbf{A}^{-1} \mathbf{A} = \mathbf{I}$, where \mathbf{l}_j denotes the j th column of the identity matrix \mathbf{I} . Solving this equation is a standard, inexpensive routine available in CFD codes, e.g. with a computational complexity of $O(n \log n)$ if a multigrid linear solver is used, where n indicates the number of elements in \mathbf{r} . Therefore, it can be estimated that even obtaining the full condition number field only has a computational complexity of $O(n^2 \log n)$ with a standard multigrid linear solver, which is much lower than the complexity of $O(n^3)$ for typical algorithms of matrix inversions.

As the local condition number $\mathcal{K}(\mathbf{x})$ is a spatial function, and its discretization is an n -vector, it is desirable to obtain a scalar quantity to provide an integral, more straightforward measure of model conditioning property similar to the global condition number \mathcal{K}_τ . To this end, we define a volume-averaged local condition number $\bar{\mathcal{K}}_x$ defined as

$$\bar{\mathcal{K}}_x = \frac{\sum_{j=1}^n [\mathcal{K}_j][\Delta V_j]}{V}, \quad (2.21)$$

where ΔV_j denotes the volume of the j th cell in the CFD mesh, and V is the total volume of the computational domain. This volume-averaged local condition number $\bar{\mathcal{K}}_x$ has a similar interpretation to \mathcal{K}_τ but preserves the merits of \mathcal{K}_j , i.e. tighter bounds and mesh independence.

In the derivations above, the Reynolds stress term is substituted directly into the RANS equation and is treated explicitly. When the Reynolds stress term is treated implicitly as in many practical implementations of Reynolds stress models (e.g. Basara & Jakirlic 2003; Maduta & Jakirlic 2017), the corresponding local condition number of the model can be obtained similarly, except that the Green's function is modified to account for the implicit modelling of the linear part of the Reynolds stress with the eddy-viscosity model. Specifically, the general form of implicit treatment of the Reynolds stress can be written as

$$\boldsymbol{\tau} = 2\nu_t \mathbf{S} + \boldsymbol{\tau}^\perp, \quad (2.22)$$

where ν_t represents the eddy viscosity, $\mathbf{S} = \frac{1}{2}(\nabla \mathbf{u} + (\nabla \mathbf{u})^\top)$ denotes the strain-rate tensor and $\boldsymbol{\tau}^\perp$ denotes the nonlinear part. In this work, the eddy viscosity ν_t is obtained from projecting DNS Reynolds stress onto DNS strain-rate tensor and not from RANS simulations. With such an optimal eddy viscosity ν_t^m , equation (2.22) treats the linear part of the Reynolds stress tensor implicitly to enhance the conditioning of RANS equations. Consequently, the Green's function \tilde{G} corresponding to the linear operator

$$\tilde{\mathcal{L}}(\mathbf{u}) = \mathcal{L}(\mathbf{u}) - \nu_t^m \nabla^2 \mathbf{u} = \mathbf{u}_0 \cdot \nabla \mathbf{u} - (\nu + \nu_t^m) \nabla^2 \mathbf{u} \quad (2.23)$$

should be used in (2.15) and (2.20), with \tilde{G} related to $\tilde{\mathcal{L}}$ in a similar way as G is to \mathcal{L} in (2.16). The optimal eddy viscosity ν_t^m is computed by minimizing the discrepancy between the linear eddy-viscosity model and the DNS Reynolds stress data, i.e.

$$\nu_t^m(\mathbf{x}) = \arg \min_{\nu_t} \|\boldsymbol{\tau}^{DNS} - 2\nu_t(\mathbf{x}) \mathbf{S}^{DNS}\|, \quad (2.24)$$

where $\boldsymbol{\tau}^{DNS}$ and \mathbf{S}^{DNS} denote the Reynolds stress and the strain-rate tensor from the DNS database, respectively. Here we emphasize that the optimal eddy viscosity is location-dependent. The detailed derivations are presented in § A.3. It is noted that the eddy viscosity ν_t in (2.22) only quantifies the amount of Reynolds stress being treated implicitly and is not necessarily the optimal eddy viscosity ν_t^m . By specifying different ν_t , the amount of implicit treatment of Reynolds stress and the amount of the nonlinear part of the Reynolds stress vary accordingly, leading to different conditioning of RANS equations. Therefore, equation (2.22) provides a general form of the implicit treatment of Reynolds stresses. The optimal eddy viscosity is capped

to be positive for numerical stability of solving the RANS equations. Therefore, the implicit treatment always leads to better conditioning of the system, but the difference between the implicit and explicit treatments becomes smaller in the regions where the linear part of the Reynolds stress is less dominant. In the extreme (albeit unlikely) situation where the Reynolds stress tensor is orthogonal to the strain-rate tensor (i.e. optimal eddy viscosity is zero across the flow domain), the conditioning with implicit and explicit treatments would be equivalent.

In summary, we proposed a local condition number to assess the sensitivity of local mean velocities with regard to data-driven Reynolds stress models. It has the following three forms: the spatial function $\mathcal{K}(\mathbf{x})$ (i.e. condition number function), an n -vector \mathcal{K}_j obtained by discretizing $\mathcal{K}(\mathbf{x})$ on the CFD mesh, and a scalar $\overline{\mathcal{K}}_x$ obtained by integration of $\mathcal{K}(\mathbf{x})$. This metric is applicable to different types of data-driven Reynolds stress models. The main contributions of this work are (1) highlighting the importance of the conditioning of PDE-governed systems with data-driven closure models, and (2) providing a quantitative metric for assessing such conditioning. Data-driven modelling is becoming an emerging area in the fluid mechanics community. However, all existing works of data-driven modelling focus on the accuracy of the model itself. Our work demonstrates that an accurate data-driven model does not necessarily guarantee satisfactory predictions of quantities of interest. Therefore, ensuring good conditioning of the problem formulation (i.e. PDEs with data-driven closure) is as important as improving the accuracy of data-driven model. We envision a standard practice in the future where all developed data-driven closures are presented along with corresponding conditioning analysis, in the same way in which experimental data reported nowadays are accompanied by their associated uncertainties.

3. Results

We first use the proposed local condition number to explain the model conditioning of RANS equations when using an explicit treatment with fixed Reynolds stress. Specifically, turbulent channel flows at different Reynolds numbers are studied and the results are shown in table 1. In addition, the implicit treatment of the Reynolds stress is studied to compare with the model conditioning of RANS equations by using an explicit treatment with fixed Reynolds stress. By studying these two types of RANS modelling, we show that the proposed local condition number can be used to assess the sensitivities of RANS simulations for data-driven modelling. We also extend the study of the model conditioning and the proposed local condition number to two more complex flows, including the flow over periodic hills and the flow in a square duct. The results show that the RANS equations can be ill-conditioned for other types of flows, which can be assessed by the proposed condition number metric.

The RANS simulations are performed in a finite-volume CFD platform OpenFOAM, using a modified flow solver that allows the explicit and implicit treatments of the Reynolds stress computed from DNS data. For numerical discretizations, the second-order central difference scheme is chosen for all terms except for the convection term, which is discretized with a second-order upwind scheme. The second-order upwind scheme was used to avoid the possible numerical instability when using the central difference scheme for the convection term. For the turbulent plane channel cases, the convection term disappears at steady state and thus the results presented in § 3.1 are not influenced by the discretization of the convection term thereof. In § 3.1, the mean velocity is obtained by directly solving (2.5) since

the mean velocity and the pressure are decoupled for the RANS simulation of a fully developed plane channel flow. The convergence criterion of solving the momentum equations is set as 10^{-8} in absolute error.

3.1. Turbulent channel flow

The fully developed turbulent plane channel flows are investigated by using the local condition number \mathcal{K}_j . In this work, we consider an ideal scenario in which the Reynolds stress τ is directly computed from a DNS database at various Reynolds numbers $Re_\tau = 180, 550, 1000, 2000$ and 5200 . The numbers of mesh cells N are 36, 110, 200, 400 and 1040, respectively, for the channel flows at those Reynolds numbers. Non-uniform meshes are used and the expansion ratio is adjusted to ensure that the y^+ of the first cell centre is kept below 1. Therefore, the mesh size of the first cell in the wall-normal direction is below h/N , where h denotes the half channel height. The DNS data were obtained from the University of Texas, Austin, online database (Lee & Moser 2015). The mean velocity field is then solved by substituting the computed Reynolds stress as the closure term of RANS equations.

In the practice of RANS modelling, iterations are involved in solving RANS equations and the modelling of the Reynolds stress is updated by the mean velocity field during the iterations. Therefore, it is possible that the mean velocity field and the Reynolds stress can adjust to each other during the iterations. We employed the ratio $\delta U_{rms}/U_{rms}^{DNS}$ to assess the error of the solved mean flow field at each iteration step. Specifically, the volume-averaged root-mean-squared (r.m.s.) error of the solved mean velocity is defined as

$$\delta U_{rms} = \sqrt{\frac{\sum_{j=1}^n ([U]_j - [U^{DNS}]_j)^2 [\Delta V_j]}{V}}. \quad (3.1)$$

The volume-averaged r.m.s. DNS velocity is defined as

$$U_{rms}^{DNS} = \sqrt{\frac{\sum_{j=1}^n ([U^{DNS}]_j)^2 [\Delta V_j]}{V}}. \quad (3.2)$$

3.1.1. Reynolds stress models with explicit treatment

The Reynolds stress term is directly computed from DNS data and substituted into RANS as shown in algorithm 1. The purpose is to study most existing data-driven RANS models, in which the Reynolds stress is directly predicted by training on DNS data and then used to solve for mean velocity field. In this subsection, we only study the explicit treatment with fixed Reynolds stress, i.e. the dependence of the Reynolds stress on strain-rate tensor is not considered. It is because such a dependence has not been taken into consideration in many data-driven turbulence models, and thus we illustrate the corresponding issue here. More results of explicit treatment with dependence on strain-rate tensor can be found in appendix C, where we further show that the explicit coupling between Reynolds stress and mean velocity during the iterations can gradually amplify the errors.

Algorithm 1: Explicit treatment of Reynolds stress with the DNS Reynolds stress being fixed among iterations.

- 1 Set Reynolds stress from DNS data: $\boldsymbol{\tau} = \boldsymbol{\tau}^{DNS}$
 - 2 **for** each iteration step $i = 1, 2, \dots, N$ **do**
 - 3 | Solve RANS equations: $\mathcal{N}(\bar{\mathbf{u}}^{(i)}) = \nabla \cdot \boldsymbol{\tau} - \nabla p$ to obtain $\bar{\mathbf{u}}^{(i)}$
 - 4 **end**
-

The local condition numbers \mathcal{K}_j of the explicit treatment of the Reynolds stress are shown in figure 4. With the increase of the Reynolds number, it can be seen that the magnitude of local condition number also increases. Specifically, the local condition number of the flow at $Re_\tau = 180$ is of $O(1)$, while the local condition number of the flow at $Re_\tau = 5200$ is of $O(10^2)$. This rapid increase of local condition number agrees well with the increased error of solved mean velocity U as summarized in table 1. In addition, the local condition number is greater near the channel centre than close to the wall, especially for the high-Reynolds-number cases. Such pattern of local condition number also agrees with the spatial pattern of the error of the solved mean velocity as illustrated in figure 9(b). As demonstrated by Thompson *et al.* (2016), the error of solved channel flow mean velocity at a given point is an integration of Reynolds stress errors (which happen to be of the same sign in the entire domain, indicating a systematic nature) from the wall to that given point, i.e. the Reynolds stress errors in between are accumulated without cancellation. The local condition number in figure 4 faithfully reflects such an accumulative nature with deteriorated conditioning further away from the wall for all cases. Recall that the local condition number \mathcal{K}_j assesses the relative error of solved mean velocity at a given point with respect to the error in the whole Reynolds stress field, and not with respect to the error of the Reynolds stress at the same given point.

The averaged local condition number $\bar{\mathcal{K}}_x$ increases with the Reynolds number by using explicit treatment of the Reynolds stress, which is clearly seen in figure 7. Such increase of averaged local condition number with Reynolds number reveals the potential shortcoming of explicit modelling of the Reynolds stress, i.e. a relatively accurate but explicit modelling of the Reynolds stress does not guarantee the satisfactory mean velocity by solving RANS equations, especially for high-Reynolds-number flows. This observation has been reported in the work of Thompson *et al.* (2016), and the proposed averaged local condition number can be used as an integral indicator to estimate the extent of error propagation from the modelled Reynolds stress to the solved mean velocity field. The error propagation with iterations is presented in figure 5(a) together with the averaged local condition number in each iteration. It can be seen that the error in the solved mean velocity stays constant in every iteration step. This observation is consistent with our expectation since the convection term disappears in the channel flows and the RANS equations become linear. Therefore, the error within the solved mean velocity in a given iteration does not influence the model conditioning and thus has no effect upon the error in the solved mean velocity at the next iteration.

3.1.2. Reynolds stress models with implicit treatment

The eddy viscosity is directly computed from DNS data and substituted into RANS equations to study the ideal situation of data-driven Reynolds stress models with implicit treatment as shown in algorithm 2.

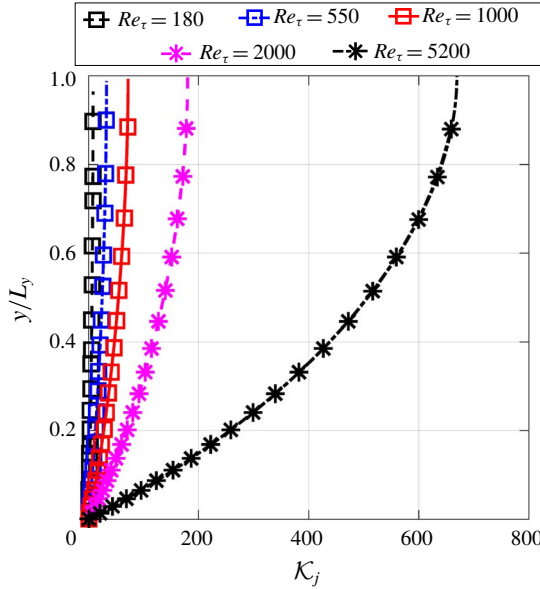


FIGURE 4. (Colour online) The profiles of local condition number \mathcal{K}_j at different Reynolds numbers by using *explicit treatment* with fixed Reynolds stress, i.e. the fixed DNS Reynolds stress is substituted into RANS equations.

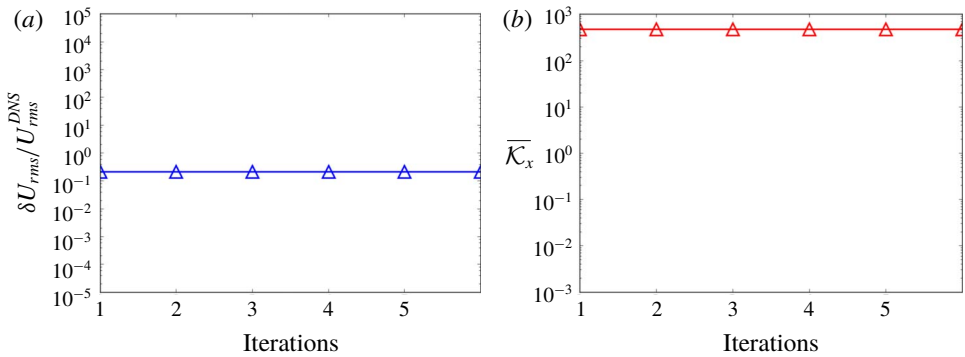


FIGURE 5. (Colour online) The error propagation analysis of solving streamwise velocity iteratively by using *explicit treatment* with fixed Reynolds stress, including (a) relative error of mean velocity and (b) volume-averaged local condition number.

Algorithm 2: Implicit treatment of Reynolds stress that depends on the strain rate among iterations.

- 1 Compute optimal eddy viscosity ν_t^m from DNS Reynolds stresses based on (2.24)
 - 2 **for** each iteration step $i = 1, 2, \dots, N$ **do**
 - 3 Compute Reynolds stress: $\boldsymbol{\tau}^{(i)} = \nu_t^m (\nabla \bar{\mathbf{u}}^{(i)} + (\nabla \bar{\mathbf{u}}^{(i)})^T) + \boldsymbol{\tau}_{DNS}^\perp$
 - 4 Solve the RANS equations: $\mathcal{N}(\bar{\mathbf{u}}^{(i)}) = \nabla \cdot \boldsymbol{\tau}^{(i)} - \nabla p$ to obtain $\bar{\mathbf{u}}^{(i)}$
 - 5 **end**
-

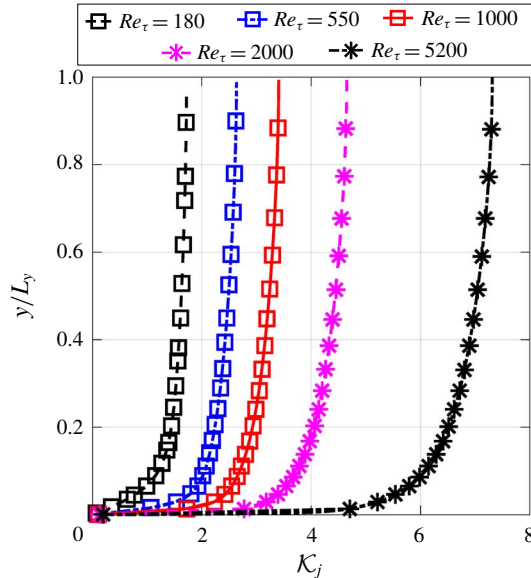


FIGURE 6. (Colour online) The local condition number at different Reynolds numbers by using *implicit treatment* of Reynolds stress, i.e. the linear part of Reynolds stress is implicitly treated by introducing an optimal eddy viscosity.

Compared to the Reynolds stress models, it is well known that eddy-viscosity models can enhance the stability and conditioning of RANS equations with turbulence closures. In the practice of traditional RSM, the modelled Reynolds stress is empirically blended with the Reynolds stress from eddy-viscosity models to achieve better convergence and conditioning (Basara & Jakirlic 2003; Maduta & Jakirlic 2017). In this work, we demonstrate that the local condition number \mathcal{K}_j can quantitatively explain the improved conditioning of the implicit treatment of the Reynolds stress by introducing an eddy viscosity. It can be seen in figure 6 that the local condition number \mathcal{K}_j is significantly reduced compared with the results of explicit treatment of the Reynolds stress as shown in figure 4, especially for high-Reynolds-number flows. Although the local condition number of high-Reynolds-number flow is still greater than that for low Reynolds number, they are of the same order of magnitude for different Reynolds numbers. The volume-averaged local condition number in figure 7 is also significantly reduced by using an implicit treatment of the Reynolds stress, demonstrating the merit of using an implicit treatment of the Reynolds stress in improving the conditioning when solving RANS equations for mean velocity field. The conditioning can be further improved by adjusting τ_{DNS}^\perp to enhance the implicit treatment part, e.g. increasing ν_t^m by $\Delta\nu_t$ and adjusting the nonlinear part of the Reynolds stress as $\tau_{DNS}^\perp - \Delta\nu_t(\nabla\bar{u}^{(i)} + (\nabla\bar{u}^{(i)})^T)$ accordingly. However, it should be noted that such a purely numerical enhancement may introduce excessive errors to iterative solvers when the chosen $\Delta\nu_t$ is too large. The proposed scheme aims to strike a balance between accuracy and conditioning.

We further show that the relative error of mean velocity is much smaller by using an implicit treatment of the Reynolds stress in RANS simulations. It can be seen in figure 8(a) that the relative error of the solved mean velocity is much smaller than the one shown in figure 5(a). In addition, the volume-averaged local condition number

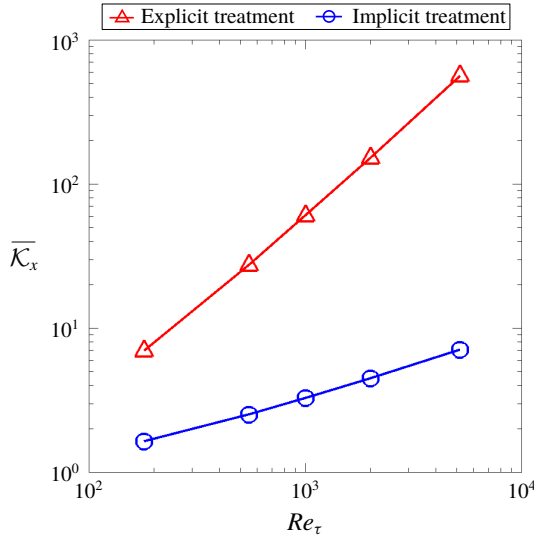


FIGURE 7. (Colour online) The volume-averaged local condition number at different Reynolds numbers for explicit and implicit treatments of Reynolds stress.

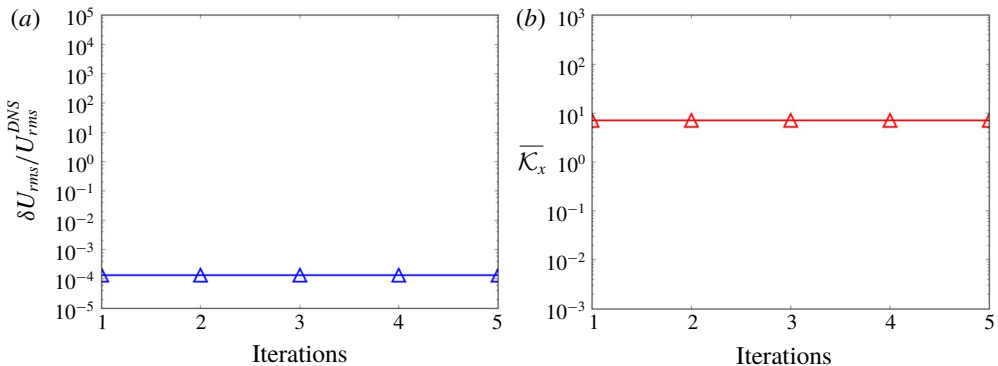


FIGURE 8. (Colour online) The error propagation analysis of solving streamwise velocity iteratively by using *implicit treatment* of Reynolds stress, including (a) relative error of mean velocity and (b) volume-averaged local condition number.

stays at $O(1)$ as shown in figure 8(b), which explains the better convergence of solving for mean velocity field by using an implicit treatment of the Reynolds stress.

The mean velocity U is solved and presented in figure 9 at Reynolds numbers $Re_\tau = 180$ and 5200 by using explicit and implicit treatments of the Reynolds stress. At Reynolds number $Re_\tau = 180$, it can be seen in figure 9(a) that the solved mean velocity U by using both kinds of treatments has a good agreement with DNS data. This demonstrates that the error propagation from Reynolds stress to mean velocity is not severe at low Reynolds number, and the percentage error of the mean velocity is comparable by using either an explicit or implicit treatment of the Reynolds stress as shown in figure 9(c). These results show good agreement with the local condition number presented in figures 4 and 6, which shows that the local condition number is of the same order for the flow at Reynolds number $Re_\tau = 180$ by using both

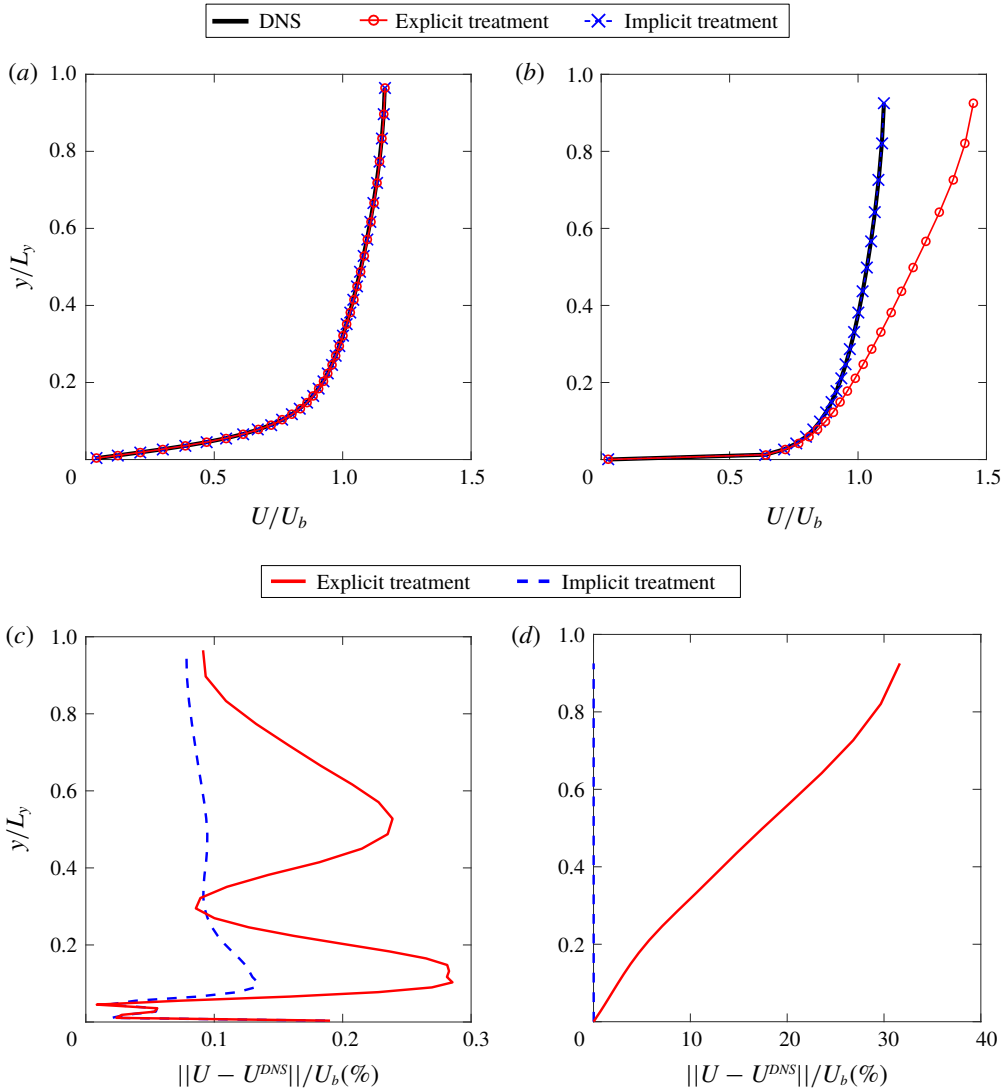


FIGURE 9. (Colour online) The comparison of solved mean velocity by using explicit and implicit treatments of Reynolds stress, including (a) mean velocity U at $Re_\tau = 180$, (b) mean velocity U at $Re_\tau = 5200$, (c) percentage error of mean velocity U at $Re_\tau = 180$ and (d) percentage error of mean velocity U at $Re_\tau = 5200$.

types of treatments. However, the solved mean velocity fields are noticeably different at high Reynolds number ($Re_\tau = 5200$) as shown in figure 9(b). Specifically, the solved mean velocity by using an explicit treatment of the Reynolds stress shows a significant difference from the DNS data, while the solved mean velocity by using an implicit treatment of the Reynolds stress still agrees well with the DNS data at $Re_\tau = 5200$. The subtle differences between the explicit and implicit treatments are further discussed in appendix D. The percentage error of solved mean velocity at $Re_\tau = 5200$ in figure 9(d) also confirms that the error in mean velocity by using an explicit treatment of the Reynolds stress is orders of magnitude higher than the error

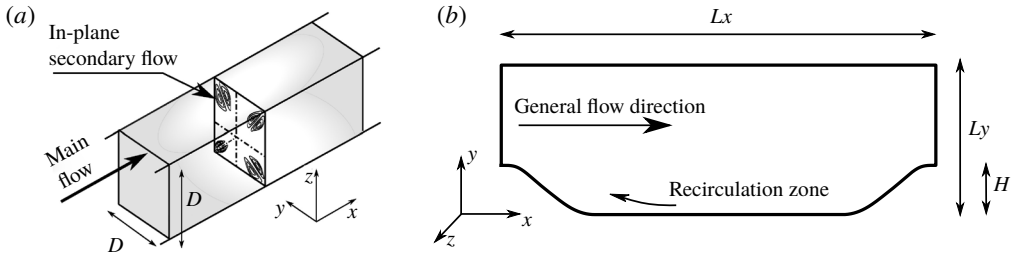


FIGURE 10. The configuration of (a) the flow in a square duct at $Re_b = 3500$ and (b) the flow over periodic hills at $Re_b = 5600$.

of using an implicit treatment of the Reynolds stress. Such comparison of solved mean velocity fields agrees well with the differences in local condition number \mathcal{K}_j , demonstrating that the proposed local condition number can be used to quantitatively assess the error propagation from Reynolds stress to mean velocity when solving RANS equations. At a high Reynolds number ($Re_\tau = 5200$), the profile of the error is dominated by the local condition number, which is much larger near the channel centre as shown in figure 4. At a low Reynolds number ($Re_\tau = 180$), the local condition number is of the same order of magnitude in the whole domain, and thus the profile of the error is dominated by the error within the DNS Reynolds stress. Therefore, the noisy pattern of the profile of the error can only be seen in figure 9(c) for $Re_\tau = 180$ but not in figure 9(d) for $Re_\tau = 5200$. Inspired by the comparison of conditioning with explicit and implicit treatments of Reynolds stress closures, Wu *et al.* (2018) further proposed an implicit treatment of the Reynolds stress for machine-learning-assisted RANS modelling to improve the conditioning when solving for mean velocity field.

3.2. Model conditioning of RANS equations for more complex flows

We further study the model conditioning of RANS equations for more complex flows including the flow in a square duct at $Re_b = 3500$ (Pinelli *et al.* 2010) and the flow over periodic hills at $Re_b = 5600$ (Breuer *et al.* 2009), where Re_b is defined by the bulk velocity U_b at the inlet. The flow configurations are shown in figure 10. In this work, we show that the RANS equations of more complex flows can also be ill-conditioned. In addition, we demonstrate that the proposed local condition number can be used to assess the model conditioning of RANS equations in these more complex flows.

3.2.1. Flow in a square duct

We first study the solved mean secondary velocity U_z by using an explicit treatment with fixed Reynolds stress and an implicit treatment of the Reynolds stress. The comparison of mean velocity profiles demonstrates that the implicit treatment of the Reynolds stress leads to solved mean velocity that shows better agreement with DNS data in figure 11. We then focus on the analysis of the errors in this work. The error is quantified by the ratio $\|U_z - U_z^{DNS}\|/U_{z,rms}^{DNS}$, where $U_{z,rms}^{DNS}$ is a volume-averaged velocity of U_z^{DNS} as defined in (3.2). It can be seen in figure 12(a) that some large errors exist in the region of the vertical symmetry plane and around the diagonal within the cross-plane. Compared to the errors as shown in figure 12(a), noticeable

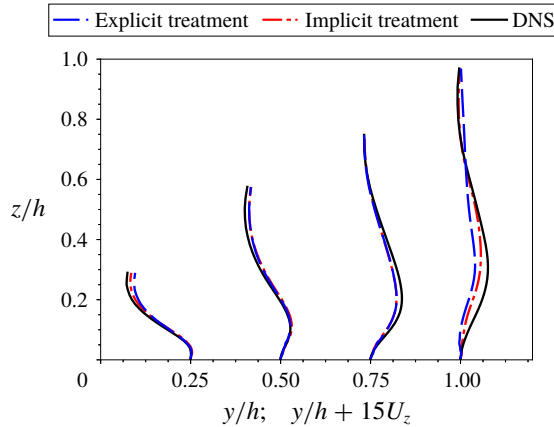


FIGURE 11. (Colour online) The comparison of solved mean velocities by using explicit treatment with fixed Reynolds stress and implicit treatment of Reynolds stress. The computational domain covers a quarter of the cross-section of the physical domain, i.e. $h = D/2$. This is due to the symmetry of the mean flow in both y and z directions as shown in figure 10(a). It should be noted that the Reynolds stress is obtained from a DNS database (Pinelli *et al.* 2010).

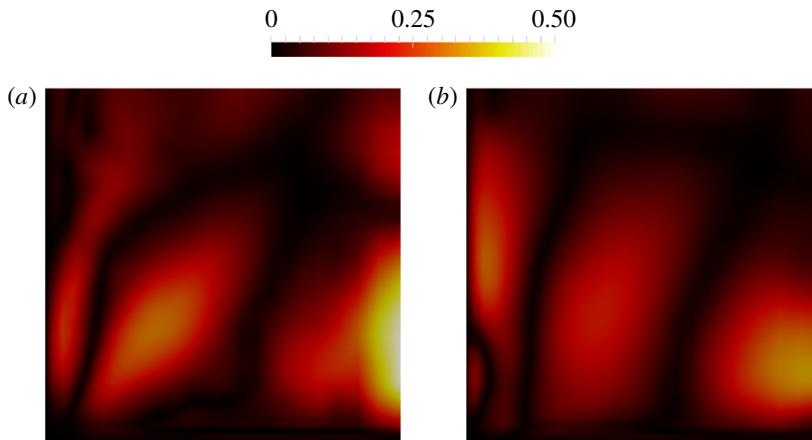


FIGURE 12. (Colour online) The normalized error of the solved secondary flow velocity U_z of the flow in a square duct by using (a) explicit treatment with fixed Reynolds stress and (b) implicit treatment of Reynolds stress.

reduction of errors can be observed in figure 12(b), where the implicit treatment of the Reynolds stress is used.

The proposed local condition number can be used to analyse the model conditioning for the flow in a square duct as demonstrated in figure 13. It can be seen in figure 13(a) that the local condition number is also large in the region of the vertical symmetry plane and around the diagonal within the cross-plane, which is consistent with the error of mean velocity as shown in figure 12(a). In addition, figure 12(b) shows that the local condition number is generally smaller across the whole domain by using an implicit treatment of the Reynolds stress. Such a reduction of local

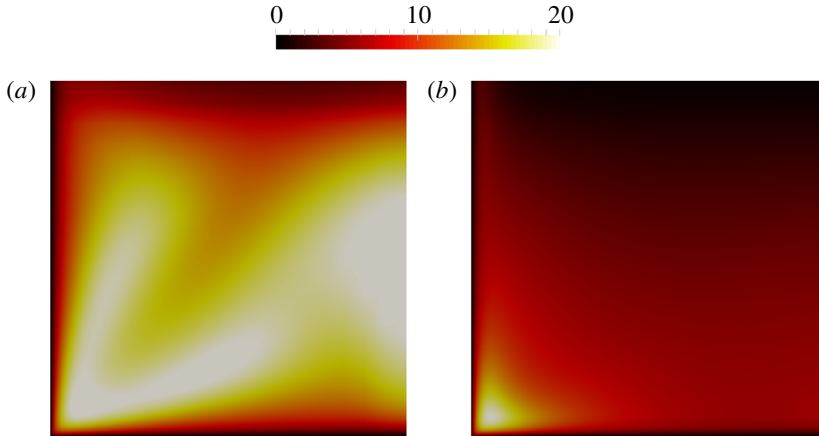


FIGURE 13. (Colour online) The local condition number of the flow in a square duct by using (a) explicit treatment with fixed Reynolds stress and (b) implicit treatment of Reynolds stress.

condition number also correlates well with the comparison of mean velocity error in figure 12. It should be noted that the mean velocity error is determined by both the local condition number and the error in Reynolds stress. Therefore, the spatial pattern of mean velocity error in figure 12 cannot be solely explained by the local condition number in figure 13. However, the analysis of local condition number can still provide some information about whether the solved mean velocity is reliable. In practical applications, the error of the Reynolds stress is usually unknown, and caution should be exercised when regions with large local condition number exist. Note that the results shown in figures 12 and 13 demonstrate that the vertical velocity U_z is not symmetric with itself about the diagonal of the domain. Rather, the diagonal symmetry of this flow is such that the velocity components U_z and U_y are symmetrical to each other about the diagonal. That is, it is expected that $U_y(\mathbf{x}) = U_z(\mathbf{x}')$ and not $U_z(\mathbf{x}) = U_z(\mathbf{x}')$, where \mathbf{x} and \mathbf{x}' denote two points symmetric about the diagonal.

3.2.2. Flow over periodic hills

Unlike the flow in a square duct, the error of mean velocity can be large across the whole domain for the flow over periodic hills. We first show the comparison of the Reynolds stress from different DNS datasets, including the data by Breuer *et al.* (2009) (dataset 1) and two datasets obtained by using Incompact3d (Laiyet & Lamballais 2009; Laiyet & Li 2011) by using two different mesh resolutions (datasets 2 and 3). The simulation details are summarized in table 2. It can be seen in figure 14(a) that the Reynolds stresses obtained from these three datasets are very close to each other. According to the comparison of these Reynolds stresses, we might intuitively expect that the solved mean velocity fields are similar to each other by substituting the fixed DNS Reynolds stresses into RANS equations. However, this is not the case, as shown in figure 14(b), where we compare the solved mean velocity with the DNS mean velocity field. It can be seen that the solved mean velocity field by using fixed DNS Reynolds stress from dataset 2 show noticeable differences across the whole domain. In contrast, the solved mean velocity fields from datasets 1 and 3 have better agreement with the DNS mean velocity field, but they are still different

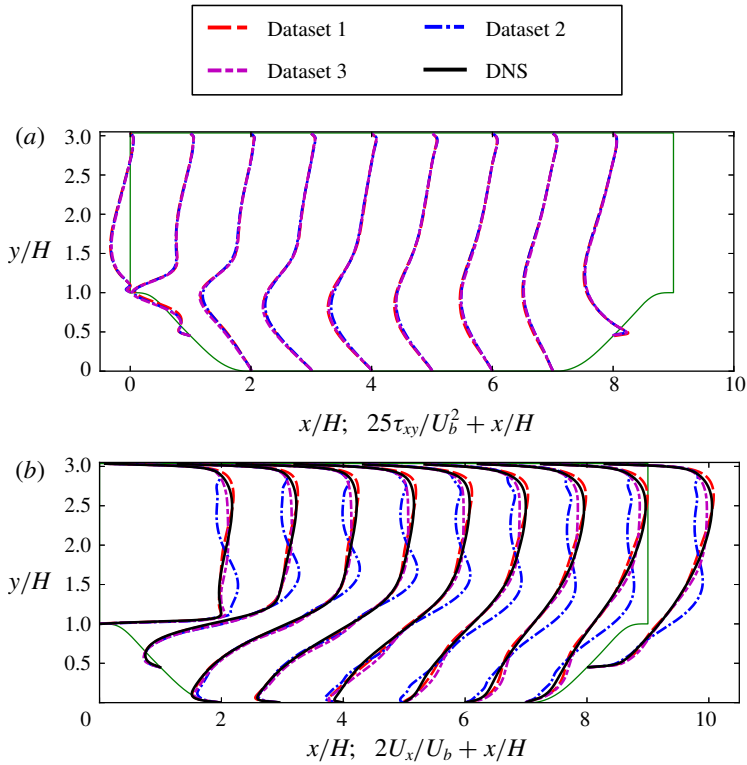


FIGURE 14. (Colour online) The comparison of (a) DNS Reynolds stress and (b) solved mean velocity between different DNS datasets. The solved mean velocity is obtained by substituting the corresponding DNS Reynolds stress and solving RANS equations. The solid black lines in panel (b) denote the DNS mean velocity as a benchmark. It should be noted that the DNS mean velocities from different datasets have no noticeable difference, and here we present the DNS mean velocity from Breuer *et al.* (2009).

Dataset	Mesh ($N_x \times N_y \times N_z$)	Solver	Methods
1	$281 \times 234 \times 200$	LESOCC (Breuer <i>et al.</i> 2009)	LES, finite volume method, body-fitting grid, second-order discretization
2	$512 \times 257 \times 128$	Incompact3d	DNS, Cartesian grid with immersed boundary method, pseudo-spectral method, sixth-order discretization
3	$768 \times 385 \times 128$	(Laiyet & Li 2011)	

TABLE 2. Summary of the datasets of the flow over periodic hills at Reynolds number $Re = 5600$, including the numerical methods, treatment of solid boundaries, accuracy of numerical discretization, mesh sizes and type. Here N_x , N_y , and N_z indicate number of grids in streamwise, wall-normal and spanwise directions, respectively.

from each other. It is expected that implicit treatment leads to better agreement of solved mean velocity with DNS data, which has been demonstrated in a related work (Wu *et al.* 2018).

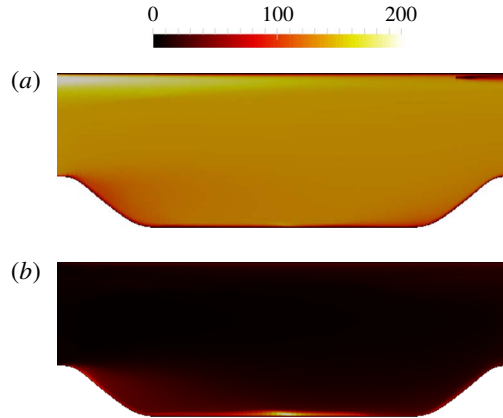


FIGURE 15. (Colour online) The local condition number \mathcal{K}_j of the flow over periodic hills at $Re = 5600$ by using (a) explicit treatment with fixed Reynolds stress and (b) implicit treatment of Reynolds stress.

The comparison in figure 14 can be explained by studying the model conditioning of RANS equations with specified Reynolds stress. Specifically, it can be seen in figure 15(a) that the local condition number by using fixed Reynolds stress is of $O(10^2)$ in most areas, indicating that the RANS equations are ill-conditioned in these regions. On the other hand, the local condition number is smaller in the near-wall and the recirculation regions. The large local condition number can explain the comparison in figure 14, i.e. the similar Reynolds stress fields lead to dramatically different mean velocity fields by solving the RANS equations. The physical justification of such a pattern of the local condition number field is that the mean velocity error is strongly correlated along the streamlines, and the periodic boundary condition of the inlet and outlet exacerbates the model conditioning for this flow. We also study the local condition number by using an implicit treatment of the Reynolds stress. It can be seen in figure 15(b) that the local condition number is much smaller than those where fixed Reynolds stress is used in figure 15(a). Therefore, it can be expected that the solved mean velocity by using an implicit treatment of DNS Reynolds stress have a better agreement with the DNS mean velocity, which has been confirmed by Wu *et al.* (2018).

In this work, the local condition number assesses the relative error of the solved mean velocity at a given point with regard to the errors in the Reynolds stress field. Therefore, the global effect of error can be captured by the local condition number. Specifically, such an effect depends on the mean flow pattern, which is embodied in the differential operator of linearized RANS equations and influences the local condition number via the Green's function $G(\mathbf{x}, \boldsymbol{\xi})$ as defined in (2.15). For instance, the errors in the solved mean velocity of the periodic hill flow is generally large across the upper channel region as shown in figure 14. This is largely due to the fact that the cyclic boundary conditions of the inlet and outlet introduce a strong correlation among errors in the upper channel region. It can be seen in figure 15 that the local condition number \mathcal{K}_j is also large across the upper channel region, demonstrating that the local condition number takes into account the mean flow pattern and thus truthfully reflects the potentially large error in the solved mean velocities therein.

4. Discussion

While this work primarily focuses on the conditioning of a particular class of data-driven Reynolds stress models, the conditioning issue is an equally important challenge for traditional Reynolds stress models based on Reynolds stress transport equations. Although solving a monolithic system of Reynolds stress transport equations and RANS equations is the most effective way to improve conditioning, it is uncommon due to increased computational costs. Moreover, monolithic coupling is by no means a panacea that guarantees well-conditioning and stability. The conditioning and stability ultimately depend on the characteristics of the turbulence model itself. For example, the popularity of the SA model in external aerodynamics is largely attributed to its excellent robustness in terms of both model conditioning and numerical stability, which many other models do not have. As of now, most commercial and open-source general-purpose CFD packages (e.g. Weller *et al.* 1998) still solve the turbulence transport equations and the RANS mean flow equations in a segregated manner, even in solvers where velocity and pressure are solved concurrently. In such segregated solvers, the modelled Reynolds stress is often updated with the mean velocity field at every iteration step, in the hope that the mean velocity field and the Reynolds stress can consistently adjust to each other during the iterations. However, if the RANS equations are ill-conditioned and the Reynolds stresses are treated explicitly as source terms, the error can be amplified within each iteration. Consequently, a small error in the modelled Reynolds stress can lead to large errors in the solved mean velocity field, which is carried over to the Reynolds stresses in the next iteration step and further amplified. Such an error amplification destabilizes the solution procedure and can potentially lead to divergence.

For the reasons outlined above, RANS simulations with Reynolds-stress-based turbulence models need to be stabilized to increase the robustness of the solvers. Examples of stabilization include (1) using the velocity solved with eddy-viscosity models as an initial condition for iterations in the RSM-based solver and (2) partial implicit treatment of the Reynolds stress, among others. In the latter category, researchers introduced a hybrid scheme of computing Reynolds stress by blending the RSM-modelled Reynolds stress with that computed from eddy-viscosity models, with the latter stabilizing the solution (Basara & Jakirlic 2003; Maduta & Jakirlic 2017). However, the choice of the blending factor is largely *ad hoc* due to the lack of a quantitative method to evaluate the model conditioning. A large blending factor improves the conditioning and stabilizes the solution of the RANS equations, but it impairs the accuracy of the solved mean velocity, since the linear eddy-viscosity assumption would be increasingly dominant. The metric proposed in this work can assess the model conditioning with any given blending factor, and thus it is possible to choose a minimum blending factor that maintains good conditioning.

A monolithic coupling for data-driven turbulence models is more challenging than for traditional PDE-based models, if it is possible at all. For example, for a neural-network-based data-driven turbulence model (e.g. Ling *et al.* 2016; Zhu *et al.* 2019), a monolithic coupling is possible, because neural network models are differentiable. However, for non-differentiable models, e.g. those based on random forests or other tree-based models (e.g. Wang *et al.* 2017), a monolithic coupling is not straightforwardly viable.

5. Conclusion

Recently, several researchers have employed DNS Reynolds stress data as the closure term and solved the RANS equations for mean velocities on turbulent channel flows. They reported unexpected results that the obtained mean velocities deviated significantly (up to 35%) from the DNS data at high Reynolds numbers. In this work, we aim to identify a metric to quantitatively assess the conditioning of RANS equations with data-driven Reynolds stress closures, i.e. how a small error in Reynolds stress can lead to large errors in the mean velocity by solving RANS equations. The turbulent channel flow is studied to evaluate the candidate metrics. Our analysis shows that the global matrix-based condition number is not able to distinguish the different sensitivity of solved mean velocities at different Reynolds numbers. A local condition number function is then derived as a more precise indicator of model conditioning. We demonstrate that such a local condition number explains the error propagation from the modelled Reynolds stress to the solved mean velocity in RANS simulations for turbulent channel flows at different Reynolds numbers. Two more complex flows are also studied to further demonstrate the capability of the proposed local condition number in evaluating the conditioning of RANS equations with data-driven Reynolds stress closures. The proposed condition number provides a quantitative metric to assess the model conditioning of RANS equations, facilitating the development of conditioning-oriented schemes in data-driven turbulence modelling.

Acknowledgements

The authors would like to thank Dr S. Laizet of Imperial College London and Professor M. Breuer of Helmut-Schmidt-Universität, Hamburg, for providing us with their high-fidelity simulation data of the flow over periodic hills. The authors would also like to thank Dr G. N. Coleman of NASA Langley and Dr F. Menter of ANSYS for helpful discussions during this research. We gratefully acknowledge Dr S. Murman of NASA Ames and Professor S. Poroseva of the University of New Mexico for their constructive comments on the first draft of the manuscript. The authors would also like to thank the reviewers for their constructive and valuable comments, which helped to improve the quality and clarity of this paper.

Appendix A. Derivations of condition numbers

A.1. Derivation of global matrix-based condition number

The global matrix-based condition number \mathcal{K}_τ is defined as follows:

$$\frac{\|\delta\mathbf{U}\|}{\|\mathbf{U}\|} \leq \mathcal{K}_\tau \frac{\|\nabla \cdot \delta\boldsymbol{\tau}\|}{\|\nabla \cdot \boldsymbol{\tau}\|}, \quad (\text{A } 1)$$

where \mathcal{K}_τ measures the sensitivity of the solved mean velocity field due to the perturbation of the Reynolds stress field, and $|\cdot|$ indicates the Euclidean norm of a vector (of all values in the discretized velocity field). To derive the formulation of \mathcal{K}_τ , the perturbation $\delta\mathbf{b}$ in (2.8) is further written as

$$\delta\mathbf{b} = \nabla \cdot \delta\boldsymbol{\tau} - \delta(\nabla p). \quad (\text{A } 2)$$

For the purpose of the sensitivity study here, it is assumed that a constant pressure gradient is imposed to drive the flow, i.e. $\delta(\nabla p) = 0$, and thus we have

$$\delta\mathbf{b} = \nabla \cdot \delta\boldsymbol{\tau}. \quad (\text{A } 3)$$

Hence,

$$\frac{\|\delta \mathbf{U}\|}{\|\mathbf{U}\|} \leq \mathcal{K}_A \frac{\|\delta \mathbf{b}\|}{\|\mathbf{b}\|} = \mathcal{K}_A \frac{\|\nabla \cdot \delta \boldsymbol{\tau}\|}{\|\mathbf{b}\|} = \underbrace{\mathcal{K}_A \frac{\|\nabla \cdot \boldsymbol{\tau}\|}{\|\mathbf{b}\|}}_{\mathcal{K}_\tau} \frac{\|\nabla \cdot \delta \boldsymbol{\tau}\|}{\|\nabla \cdot \boldsymbol{\tau}\|}. \tag{A 4}$$

Comparing the forms of (A 4) with the definition of \mathcal{K}_τ in (A 1), the matrix-norm-based condition number for Reynolds-stress-based turbulence models is thus

$$\mathcal{K}_\tau = \mathcal{K}_A \frac{\|\nabla \cdot \boldsymbol{\tau}\|}{\|\mathbf{b}\|}. \tag{A 5}$$

A.2. Derivation of local condition number function

The continuous local condition number $\mathcal{K}(\mathbf{x})$ is defined as follows:

$$\frac{|\delta \mathbf{u}(\mathbf{x})|}{U_\infty} \leq \mathcal{K}(\mathbf{x}) \frac{\|\nabla \cdot \delta \boldsymbol{\tau}\|_\Omega}{\|\nabla \cdot \boldsymbol{\tau}\|_\Omega}, \tag{A 6}$$

where $\mathcal{K}(\mathbf{x})$ measures the sensitivity of the solved mean velocity at any given location \mathbf{x} due to the perturbation of the Reynolds stress field, and U_∞ is a constant representative velocity magnitude for normalization. The function norm $\|\cdot\|_\Omega$ of function $f(\boldsymbol{\xi})$ on domain Ω is defined as in (2.17).

To derive the formulation of this local condition number, we first consider the solution \mathbf{u} at a particular location \mathbf{x}' :

$$\mathbf{u}(\mathbf{x}') = \int_\Omega G(\mathbf{x}'; \boldsymbol{\xi}) \mathbf{b}(\boldsymbol{\xi}) \, d\boldsymbol{\xi}, \tag{A 7}$$

where G represents the Green's function of the linear differential operator \mathcal{L} in the linearized RANS equations as defined in (2.6). Denoting $G_{\mathbf{x}'} = G(\mathbf{x}'; \boldsymbol{\xi})$, the perturbation of the solution is thus

$$\delta \mathbf{u}(\mathbf{x}') = \int_\Omega G_{\mathbf{x}'}(\boldsymbol{\xi}) \delta \mathbf{b}(\boldsymbol{\xi}) \, d\boldsymbol{\xi} = \langle G_{\mathbf{x}'}, \delta \mathbf{b} \rangle_\Omega, \tag{A 8}$$

where $\langle \cdot \rangle_\Omega$ is the inner product of functions defined on domain Ω .

Using the Schwartz inequality (Steele 2004; Debnath & Mikusiński 2005) leads to

$$|\delta \mathbf{u}(\mathbf{x}')| \leq \|G_{\mathbf{x}'}\|_\Omega \|\delta \mathbf{b}\|_\Omega \tag{A 9}$$

$$= \|G_{\mathbf{x}'}\|_\Omega \|\nabla \cdot \delta \boldsymbol{\tau}\|_\Omega. \tag{A 10}$$

As in § A.1, the pressure gradient is assumed constant and thus $\delta \mathbf{b} = \nabla \cdot \delta \boldsymbol{\tau}$. Finally, the sensitivity of mean velocity \mathbf{u} with respect to the Reynolds stress $\boldsymbol{\tau}$ perturbations is derived as follows:

$$\frac{|\delta \mathbf{u}(\mathbf{x}')|}{U_\infty} \leq \frac{\|G_{\mathbf{x}'}\|_\Omega \|\nabla \cdot \delta \boldsymbol{\tau}\|_\Omega}{U_\infty} \tag{A 11}$$

$$= \frac{\|G_{\mathbf{x}'}\|_\Omega \|\nabla \cdot \boldsymbol{\tau}\|_\Omega}{U_\infty} \frac{\|\nabla \cdot \delta \boldsymbol{\tau}\|_\Omega}{\|\nabla \cdot \boldsymbol{\tau}\|_\Omega}. \tag{A 12}$$

Therefore, by comparing (A 12) and (A 6), we define a local condition number function \mathcal{K} of spatial location \mathbf{x} as

$$\mathcal{K}(\mathbf{x}) = \frac{\|G_{\mathbf{x}}\|_\Omega \|\nabla \cdot \boldsymbol{\tau}\|_\Omega}{U_\infty} = \frac{\|G(\mathbf{x}, \boldsymbol{\xi})\|_\Omega \|\nabla \cdot \boldsymbol{\tau}\|_\Omega}{U_\infty}. \tag{A 13}$$

Without causing ambiguity, we have dropped the subscript of \mathbf{x}' in the equation above and in the text for simplicity of notation.

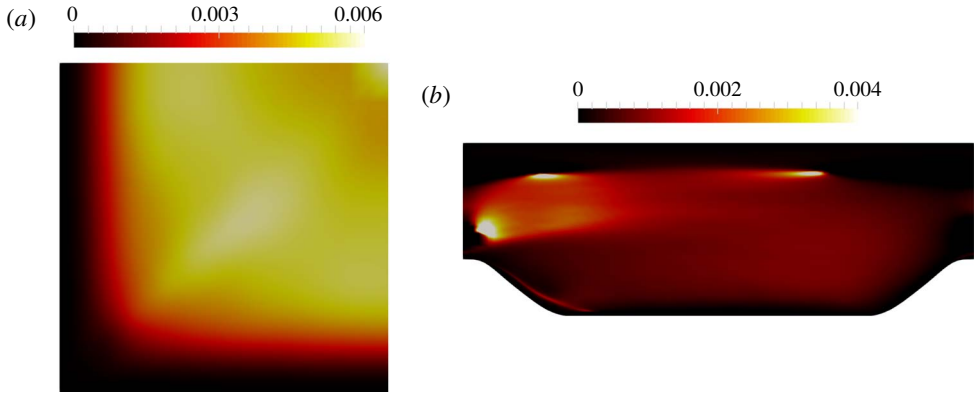


FIGURE 16. (Colour online) The optimal eddy-viscosity fields for (a) the flow in a square duct at $Re_b = 3500$ and (b) the flow over periodic hills at $Re_b = 5600$.

A.3. Local condition number for implicit treatment of Reynolds stress

In the practice of RANS modelling, eddy-viscosity models are widely used, and the modelled eddy viscosity influences the differential operator \mathcal{L} associated with RANS equations. Therefore, we extend the derivation of (2.15) to make it compatible with the implicit treatment of the Reynolds stress. According to the general form of the implicit treatment of the Reynolds stress (Pope 1975) in (2.22), the linearized RANS equations in (2.5) can be rearranged as follows:

$$\tilde{\mathcal{L}}(\mathbf{u}) = \nabla \cdot \boldsymbol{\tau}^\perp - \nabla p, \tag{A 14}$$

where $\tilde{\mathcal{L}} = \mathcal{L} - \nu_i^m \nabla^2$ is the modified linear differential operator by using an implicit treatment of the Reynolds stress. Examples of optimal eddy viscosity ν_i^m for the flow over periodic hills and the flow in a square duct are shown in figure 16. Here we only study the perturbation on the nonlinear term $\boldsymbol{\tau}^\perp$ of the Reynolds stress $\boldsymbol{\tau}$, i.e.

$$\delta \boldsymbol{\tau} = \delta \boldsymbol{\tau}^\perp. \tag{A 15}$$

Finally, we have the local condition number $\mathcal{K}(\mathbf{x}')$ in (2.15) rederived as follows for the implicit treatment of the Reynolds stress:

$$\frac{|\delta \mathbf{u}(\mathbf{x}')|}{U_\infty} \leq \frac{\|\tilde{\mathcal{G}}_{\mathbf{x}'}\|_\Omega \|\nabla \cdot \delta \boldsymbol{\tau}^\perp\|_\Omega}{U_\infty} \tag{A 16}$$

$$= \underbrace{\frac{\|\tilde{\mathcal{G}}_{\mathbf{x}'}\|_\Omega \|\nabla \cdot \boldsymbol{\tau}\|_\Omega}{U_\infty}}_{\kappa_j} \frac{\|\nabla \cdot \delta \boldsymbol{\tau}\|_\Omega}{\|\nabla \cdot \boldsymbol{\tau}\|_\Omega}, \tag{A 17}$$

where the kernel function $\|\tilde{\mathcal{G}}_{\mathbf{x}'}\|_\Omega$ represents the Green’s function that corresponds to the linear differential operator $\tilde{\mathcal{L}}$ defined in (2.23), taking into account the implicit modelling of the linear part of the Reynolds stress by introducing an eddy viscosity.

Appendix B. Mesh independence of the local condition number function

We present the numerical discretization of the proposed local condition number on a CFD mesh and show that the discretized local condition number is mesh-independent. First, the function norms $\|\cdot\|_\Omega$ are approximated in vector norms $\|\cdot\|_n$ through numerical integration on a CFD mesh of n cells. This is derived as follows:

$$\|G_{x'}\|_\Omega = \sqrt{\int_\Omega |G(x'; \xi)|^2 d\xi} \tag{B 1}$$

$$\approx \sqrt{\sum_{i=1}^n ([r_{j,i}]^2 \Delta V_i)} \tag{B 2}$$

$$= \sqrt{\sum_{i=1}^n ([r_{j,i}] \sqrt{[\Delta V_i]})^2} \tag{B 3}$$

$$= \|r_j \sqrt{[\Delta V]}\|_n = \|\tilde{r}_j\|_n, \tag{B 4}$$

with $\tilde{r}_j = r_j \sqrt{[\Delta V]}$ and \approx indicating numerical discretization of the integral involved in the function norm in (B 1); ΔV_i is the volume for the i th cell; ΔV is the n -vector consisting of volumes of cells in the mesh; and r_j is the j th row of the matrix \mathbf{A}^{-1} , with \mathbf{A} being the discretization of the operator \mathcal{L} as seen in (2.6). The numbering implies that the location x' is the coordinate of the j th cell.

Similarly, the function norm of the Reynolds stress divergence is approximated on the CFD mesh as follows:

$$\|\nabla \cdot \boldsymbol{\tau}\|_\Omega \approx \|[\nabla \cdot \boldsymbol{\tau}] \sqrt{[\Delta V]}\|_n. \tag{B 5}$$

It is clear that the function norm $\|G_{x'}\|_\Omega$ is mesh-independent as its definition does not involve any discretization mesh (quadrature), so its numerical approximation $\|r_j \sqrt{[\Delta V]}\|_n$ should also be mesh-independent on a sufficiently fine mesh. In the same way, both the function norm $\|\nabla \cdot \boldsymbol{\tau}\|_\Omega$ and its numerical approximation $\|[\nabla \cdot \boldsymbol{\tau}] \sqrt{[\Delta V]}\|_n$ are mesh-independent. The mesh independence can be further verified in the special case, where the divergence field $\nabla \cdot \boldsymbol{\tau}$ is a non-zero constant β and the mesh consists of n uniformly sized cells. In this case we have $\Delta V = |\Omega|/n$, where $|\Omega|$ is the total volume of the computational domain Ω (independent of the discretization mesh). Therefore, the vector norm, which is a numerical approximation of the function form, is as follows:

$$\|[\nabla \cdot \boldsymbol{\tau}] \sqrt{[\Delta V]}\|_n = \sqrt{\sum_{i=1}^n \left(\beta \sqrt{\frac{|\Omega|}{n}} \right)^2} \tag{B 6}$$

$$= \sqrt{n \beta^2 \frac{|\Omega|}{n}} \tag{B 7}$$

$$= \beta \sqrt{|\Omega|}, \tag{B 8}$$

which is clearly independent of the number of cells in the mesh. In order to complement and validate the derivations above, a numerical study of mesh convergence is presented in figure 17, where the local condition number \mathcal{K}_j at $Re = 5200$ is

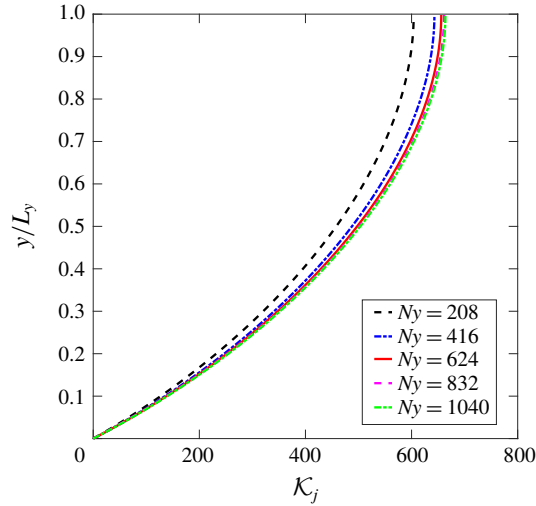


FIGURE 17. (Colour online) Mesh convergence study of local condition number \mathcal{K}_j at $Re = 5200$ by using explicit treatment with fixed Reynolds stress, i.e. the fixed DNS Reynolds stress is substituted into RANS equations.

calculated by using different mesh resolutions. It can be seen that asymptotical convergence of local condition number \mathcal{K}_j can be achieved by gradually refining the mesh resolution. Moreover, even on a coarser mesh the overall magnitude and spatial distribution of the condition number \mathcal{K}_j are correctly captured.

Appendix C. Explicit treatment of Reynolds stress with dependence on strain rate

The conditioning of RANS equations with explicit treatment of the Reynolds stress has been discussed in § 3.1.1, where the Reynolds stress is obtained from a DNS database and is fixed when solving RANS equations. If the dependence of the Reynolds stress on strain rate is considered, it can be seen in figure 18 that the conditioning of RANS equations in the first iteration is the same as the conditioning with fixed Reynolds stress as shown in figure 5. This is because the detailed explicit treatment of the Reynolds stress does not influence the model conditioning at a given state. Here we further show that the explicit coupling between Reynolds stress and mean velocity during the iterations can gradually amplify the errors and lead to divergence. Such explicit coupling is often used in the Reynolds stress transport models. Specifically, the Reynolds stress is obtained by solving its transport equations with the mean velocity field at the previous iteration step. In the following, we use a simplified example with an iterative solver as shown in algorithm 3 to illustrate the convergence issue of Reynolds stress transport models. In addition, we demonstrate that the proposed local condition number can be used to detect and explain the corresponding ill-conditioning issue during iterations.

The Reynolds stress at the i th iteration step is explicitly treated by using DNS data according to algorithm 3. Unlike the data-driven Reynolds stress modelling as shown in algorithm 1, this simplified explicit Reynolds stress treatment allows the update of the Reynolds stress at each iteration based on the solved mean velocity field at the previous iteration step. Compared to the implicit treatment of the Reynolds stress as shown in algorithm 2, the only difference of this explicit treatment is the computing

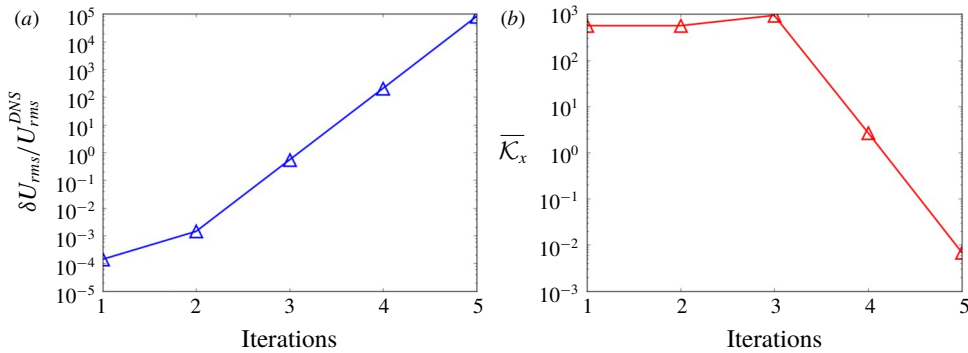


FIGURE 18. (Colour online) The error propagation analysis of solving streamwise velocity iteratively by using *explicit treatment* of Reynolds stress that depends on the strain rate, including (a) relative error of mean velocity and (b) volume-averaged local condition number.

Algorithm 3: Explicit treatment of Reynolds stress that depends on strain rate among iterations.

- 1 Compute optimal eddy viscosity ν_t^m from DNS Reynolds stresses based on (2.24)
 - 2 **for** each iteration step $i = 1, 2, \dots, N$ **do**
 - 3 Compute the Reynolds stress:

$$\boldsymbol{\tau}^{(i)} = \nu_t^m (\nabla \bar{\mathbf{u}}^{(i-1)} + (\nabla \bar{\mathbf{u}}^{(i-1)})^T) + \boldsymbol{\tau}^{\perp, DNS}$$
Solve the RANS equations:

$$\mathcal{N}(\bar{\mathbf{u}}^{(i)}) = \nabla \cdot \boldsymbol{\tau}^{(i)} - \nabla p$$
to obtain $\bar{\mathbf{u}}^{(i)}$
 - 4 **end**
-

of the Reynolds stress with the mean velocity at the previous iteration step, which is indicated by the superscript $i - 1$ at line 3 of algorithm 3.

The errors of solved mean velocity field by using an explicit treatment of the Reynolds stress is presented in figure 18(a). The DNS mean velocity is used as the initial field in RANS simulations, and thus the initial value of $\delta U_{rms}/U_{rms}^{DNS}$ is small. However, the value of $\delta U_{rms}/U_{rms}^{DNS}$ increases rapidly during the first several iteration steps. It demonstrates that the conditioning issue within each iteration can lead to error amplification, i.e. a small error in the modelled Reynolds stress can lead to noticeable errors in the solved mean velocity field and thus influence the modelled Reynolds stress in the next iteration step. Owing to such coupling of error amplification, even a small error of modelled Reynolds stress can lead to divergence of the simulation eventually. It can be seen in figure 18(b) that the volume-averaged local condition number is of $O(10^2)$ within the first three iteration steps, explaining the rapid growth of error in the solved mean velocity. The error of the solved mean velocity grows rapidly and eventually leads to divergence of the simulation. Therefore, the solved mean velocity is not presented in this work since a converged solution was not achieved. Therefore, the proposed local condition number is still of importance in traditional RANS modelling since it provides a quantitative assessment of model conditioning at every iteration step. The divergence of the mean velocity is because the Reynolds stress is updated according to the mean velocity at each iteration step. With the rapidly increased error in the mean velocity at the first several iteration steps, the error in the Reynolds stress also increases accordingly. The decrease of

the condition number as shown in figure 18(b) does not guarantee the decrease of the error in the mean velocity in such a scenario, considering that the error of mean velocity is also influenced by the error in the modelled Reynolds stress. Therefore, the small condition number needs to be interpreted with caution when the source term in RANS equations changes during the simulation. With large condition number in some previous iteration steps, it is possible that the error of the modelled Reynolds stress becomes large and the mean velocity error keeps increasing even though the condition number decreases.

Appendix D. Discrepancies in velocities obtained with different treatments

It was shown in figure 9 that the solved mean velocity can be significantly different depending on whether the DNS Reynolds stress used to solve (2.4) is treated explicitly or implicitly. In other words, solving the following two equations

$$\mathcal{L}(\mathbf{u}^{exp}) = \nabla \cdot \boldsymbol{\tau}^{exp} - \nabla p, \quad (\text{D } 1)$$

$$\mathcal{L}(\mathbf{u}^{imp}) = \nabla \cdot \boldsymbol{\tau}^{imp} - \nabla p \quad (\text{D } 2)$$

yields drastically different velocities. The superscript ‘imp’ indicates the implicit treatment of the Reynolds stress and ‘exp’ denotes the explicit treatment, i.e. $\boldsymbol{\tau}^{imp} = \nu_t(\nabla \mathbf{u}^{imp} + (\nabla \mathbf{u}^{imp})^T) + \boldsymbol{\tau}^{\perp, DNS}$ and $\boldsymbol{\tau}^{exp} = \boldsymbol{\tau}^{DNS}$. This finding apparently contradicts the common understanding in traditional CFD practice that the converged solution of the mean velocity should be the same regardless of how the Reynolds stress is treated. Indeed, the Reynolds stresses used in the two formulations in (D 1) and (D 2) are approximately equal, since $\nu_t(\nabla \mathbf{u}^{imp} + (\nabla \mathbf{u}^{imp})^T) + \boldsymbol{\tau}^{\perp, DNS} \approx \boldsymbol{\tau}^{DNS}$. More precisely, the difference between \mathbf{u}^{imp} and \mathbf{u}^{DNS} is approximately 0.1%, and the difference between $\boldsymbol{\tau}^{imp}$ and $\boldsymbol{\tau}^{exp}$ should be at a similar level. However, the condition number with regard to the nonlinear differential operator \mathcal{L} is large for the flows at high Reynolds numbers, and thus a small difference between $\boldsymbol{\tau}^{imp}$ and $\boldsymbol{\tau}^{exp}$ can lead to a large difference between the solved mean velocities \mathbf{u}^{imp} and \mathbf{u}^{exp} .

In addition, the better solution of \mathbf{u}^{imp} with the implicit treatment of the Reynolds stress can be explained by the improved model conditioning, i.e. the condition number is smaller with regard to the linear differential operator $\tilde{\mathcal{L}} = \mathcal{L} - \nu_t^m \nabla^2$ for the implicit treatment of the Reynolds stress. Specifically, we first define an optimal Reynolds stress $\boldsymbol{\tau}^{op}$ that can lead to \mathbf{u}^{DNS} by solving the RANS equations

$$\mathcal{L}(\mathbf{u}^{DNS}) = \nabla \cdot \boldsymbol{\tau}^{op} - \nabla p, \quad (\text{D } 3)$$

where $\boldsymbol{\tau}^{op}$ denotes the true Reynolds stress that provides \mathbf{u}^{DNS} by solving the RANS equations. The errors $\|\boldsymbol{\tau}^{DNS} - \boldsymbol{\tau}^{op}\|$ and $\|\boldsymbol{\tau}^{\perp, DNS} - \boldsymbol{\tau}^{\perp, op}\|$ are of the same order of magnitude. Therefore, $\|\mathbf{u}^{imp} - \mathbf{u}^{DNS}\|$ is smaller than $\|\mathbf{u}^{exp} - \mathbf{u}^{DNS}\|$ due to the smaller sensitivity of solving the mean velocity by using the implicit treatment of the Reynolds stress.

Notation

We use $[\phi]$ to indicate the n -vector obtained by discretizing the field ϕ on the mesh, where n is the number of cells/grid points; $\|[\phi]\|$ denotes the norm of vector $[\phi]$ resulting from the discretization. Since the norm is always taken on the discretized n -vector, we abbreviated $\|[\phi]\|$ as $\|\phi\|$ without ambiguity. When mentioning function norm and n -vector norm simultaneously, we used $\|\cdot\|_{\Omega}$ and $\|\cdot\|_n$ to distinguish them, with Ω denoting the domain on which the norm is defined.

\mathbf{u}	mean velocity field
\mathbf{U}	discretized mean velocity (n -vector)
$\boldsymbol{\tau}$	Reynolds stress tensor
\mathbf{S}	rate-of-strain tensor
\mathbf{b}	imbalance vector between Reynolds stress divergence and pressure gradient
\mathbf{A}	$n \times n$ coefficients matrix of discretized RANS equations
\mathcal{N}	nonlinear differential operator
\mathcal{L}	linear differential operator
G	Green's function corresponding to \mathcal{L}
$\mathcal{K}_{\mathbf{A}}$	condition number of matrix \mathbf{A}
$\bar{\alpha}$	ratio between Reynolds stress divergence and the total source term
$\mathcal{K}_{\boldsymbol{\tau}}$	matrix-norm condition number associated with Reynolds stress perturbation
\mathcal{K}_j	local condition number vector approximated on a CFD mesh (n -vector)
$\overline{\mathcal{K}}_x$	volume-averaged local condition number (scalar)
δU_{rms}	volume-averaged r.m.s. error of solved mean velocity
U_{rms}^{DNS}	volume-averaged r.m.s. DNS mean velocity
\perp	superscript indicating the nonlinear part of Reynolds stress

REFERENCES

- BASARA, B. & JAKIRLIC, S. 2003 A new hybrid turbulence modelling strategy for industrial CFD. *Intl J. Numer. Meth. Fluids* **42** (1), 89–116.
- BERNARDINI, M., PIROZZOLI, S. & ORLANDI, P. 2014 Velocity statistics in turbulent channel flow up to $Re = 4000$. *J. Fluid Mech.* **742**, 171–191.
- BREUER, M., PELLER, N., RAPP, C. & MANHART, M. 2009 Flow over periodic hills: numerical and experimental study in a wide range of Reynolds numbers. *Comput. Fluids* **38** (2), 433–457.
- CHANDRASEKARAN, S. & IPSEN, I. C. F. 1995 On the sensitivity of solution components in linear systems of equations. *SIAM J. Matrix Anal. Applics.* **16** (1), 93–112.
- DEBNATH, L. & MIKUSIŃSKI, P. 2005 *Hilbert Spaces with Applications*. Academic Press.
- DEL ALAMO, J. C. & JIMÉNEZ, J. 2003 Spectra of the very large anisotropic scales in turbulent channels. *Phys. Fluids* **15** (6), L41–L44.
- GAMAHARA, M. & HATTORI, Y. 2017 Searching for turbulence models by artificial neural network. *Phys. Rev. Fluids* **2** (5), 054604.
- GENEVA, N. & ZABARAS, N. 2019 Quantifying model form uncertainty in Reynolds-averaged turbulence models with Bayesian deep neural networks. *J. Comput. Phys.* **383**, 125–147.
- HAMLINGTON, P. E. & DAHM, W. J. A. 2008 Reynolds stress closure for nonequilibrium effects in turbulent flows. *Phys. Fluids* **20** (11), 115101.
- HAMLINGTON, P. E. & IHME, M. 2014 Modeling of non-equilibrium homogeneous turbulence in rapidly compressed flows. *Flow Turbul. Combust.* **93** (1), 93–124.
- KING, R. N., HAMLINGTON, P. E. & DAHM, W. J. A. 2016 Autonomic closure for turbulence simulations. *Phys. Rev. E* **93** (3), 031301.
- LAISET, S. & LAMBALLAIS, E. 2009 High-order compact schemes for incompressible flows: a simple and efficient method with quasi-spectral accuracy. *J. Comput. Phys.* **228** (16), 5989–6015.
- LAISET, S. & LI, N. 2011 Incompact3d: a powerful tool to tackle turbulence problems with up to $O(10^5)$ computational cores. *Intl J. Numer. Meth. Fluids* **67** (11), 1735–1757.
- LANCZOS, C. 1996 *Linear Differential Operators*. SIAM.
- LAUNDER, B. E. & SHARMA, B. I. 1974 Application of the energy-dissipation model of turbulence to the calculation of flow near a spinning disc. *Lett. Heat Mass Transfer* **1** (2), 131–137.

- LEE, M. & MOSER, R. D. 2015 Direct numerical simulation of turbulent channel flow up to $Re = 5200$. *J. Fluid Mech.* **774**, 395–415.
- LING, J., KURZAWSKI, A. & TEMPLETON, J. 2016 Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. *J. Fluid Mech.* **807**, 155–166.
- MADUTA, R. & JAKIRLIC, S. 2017 Improved RANS computations of flow over the 25°-slant-angle Ahmed body. *SAE Intl J. Passenger Cars – Mech. Syst.* **10** (2), 649–661.
- MAULIK, R. & SAN, O. 2017 A neural network approach for the blind deconvolution of turbulent flows. *J. Fluid Mech.* **831**, 151–181.
- MAULIK, R., SAN, O., RASHEED, A. & VEDULA, P. 2019 Sub-grid modelling for two-dimensional turbulence using neural networks. *J. Fluid Mech.* **858**, 122–144.
- MENTER, F. R. 1994 Two-equation eddy-viscosity turbulence models for engineering applications. *AIAA J.* **32** (8), 1598–1605.
- PARISH, E. J. & DURAISAMY, K. 2016 A paradigm for data-driven predictive modeling using field inversion and machine learning. *J. Comput. Phys.* **305**, 758–774.
- PINELLI, A., UHLMANN, M., SEKIMOTO, A. & KAWAHARA, G. 2010 Reynolds number dependence of mean flow structure in square duct turbulence. *J. Fluid Mech.* **644**, 107–122.
- POPE, S. B. 1975 A more general effective-viscosity hypothesis. *J. Fluid Mech.* **72** (2), 331–340.
- POPE, S. B. 2000 *Turbulent Flows*. Cambridge University Press.
- POROSEVA, S. V., COLMENARES, F., JUAN, D. & MURMAN, S. M. 2016 On the accuracy of RANS simulations with DNS data. *Phys. Fluids* **28** (11), 115102.
- SINGH, A. P. & DURAISAMY, K. 2016 Using field inversion to quantify functional errors in turbulence closures. *Phys. Fluids* **28**, 045110.
- SINGH, A. P., MEDIDA, S. & DURAISAMY, K. 2017 Machine-learning-augmented predictive modeling of turbulent separated flows over airfoils. *AIAA J.* **55** (7), 2215–2227.
- SLOTNICK, J., KHODADOUST, A., ALONSO, J., DARMOFAL, D., GROPP, W., LURIE, E. & MAVRIPLIS, D. 2014 CFD Vision 2030 Study: a path to revolutionary computational aerosciences. *Tech. Rep.* National Aeronautics and Space Administration, Langley Research Center, Hampton, VA.
- SPALART, P. R. & ALLMARAS, S. R. 1994 A one-equation turbulence model for aerodynamic flows. *Rech. Aerosp.* **1**, 5–21.
- SPEZIALE, C. G. & XU, X.-H. 1996 Towards the development of second-order closure models for nonequilibrium turbulent flows. *Intl J. Heat Fluid Flow* **17** (3), 238–244.
- STEELE, J. M. 2004 *The Cauchy–Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities*. Cambridge University Press.
- STRANG, G. 2016 *Introduction to Linear Algebra*, 5th edn. Wellesley–Cambridge Press.
- THOMPSON, R. L., SAMPAIO, L. E. B., DE BRAGANÇA ALVES, F. A. V., THAIS, L. & MOMPEAN, G. 2016 A methodology to evaluate statistical errors in DNS data of plane channel flows. *Comput. Fluids* **130**, 1–7.
- VOLLANT, A., BALARAC, G. & CORRE, C. 2017 Subgrid-scale scalar flux modelling based on optimal estimation theory and machine-learning procedures. *J. Turbul.* **18** (9), 854–878.
- WANG, J.-X., WU, J.-L. & XIAO, H. 2017 Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data. *Phys. Rev. Fluids* **2** (3), 034603.
- WEATHERITT, J. & SANDBERG, R. 2016 A novel evolutionary algorithm applied to algebraic modifications of the RANS stress–strain relationship. *J. Comput. Phys.* **325**, 22–37.
- WELLER, H. G., TABOR, G., JASAK, H. & FUREBY, C. 1998 A tensorial approach to computational continuum mechanics using object-oriented techniques. *Comput. Phys.* **12** (6), 620–631.
- WILCOX, D. C. 1988 Reassessment of the scale-determining equation for advanced turbulence models. *AIAA J.* **26** (11), 1299–1310.
- WU, J.-L., XIAO, H. & PATERSON, E. 2018 Physics-informed machine learning approach for augmenting turbulence models: a comprehensive framework. *Phys. Rev. Fluids* **3** (7), 074602.
- ZHU, L., ZHANG, W., KOU, J. & LIU, Y. 2019 Machine learning methods for turbulence modeling in subsonic flows around airfoils. *Phys. Fluids* **31** (1), 015105.