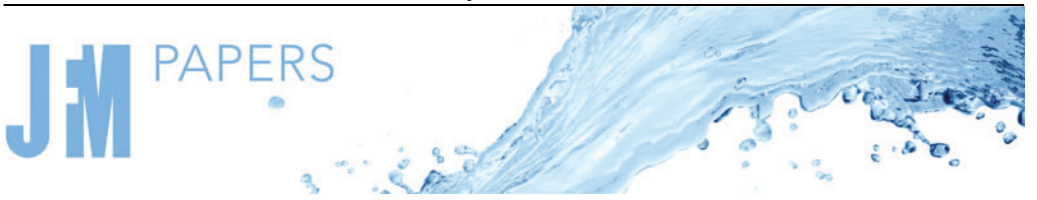
J. Fluid Mech. (2024), *vol*. 1000, A76, doi:10.1017/jfm.2024.992



A recursive neural-network-based subgrid-scale model for large eddy simulation: application to homogeneous isotropic turbulence

Chonghyuk Cho[1], **Jonghwan Park**[1] **and Haecheon Choi**[1,2],†

[1]Department of Mechanical Engineering, Seoul National University, Seoul 08826, Republic of Korea

[2]Institute of Advanced Machines and Design, Seoul National University, Seoul 08826, Republic of Korea

(Received 6 January 2024; revised 10 October 2024; accepted 10 October 2024)

We introduce a novel recursive procedure to a neural-network-based subgrid-scale (NN-based SGS) model for large eddy simulation (LES) of high-Reynolds-number turbulent flow. This process is designed to allow an SGS model to be applicable to a hierarchy of different grid sizes without requiring expensive filtered direct numerical simulation (DNS) data: (1) train an NN-based SGS model with filtered DNS data at a low Reynolds number; (2) apply the trained SGS model to LES at a higher Reynolds number; (3) update this SGS model with training data augmented with filtered LES (fLES) data, accommodating coarser filter size; (4) apply the updated NN to LES at a further higher Reynolds number; (5) go back to Step (3) until a target (very coarse) filter size divided by the Kolmogorov length scale is reached. We also construct an NN-based SGS model using a dual NN architecture whose outputs are the SGS normal stresses for one NN and the SGS shear stresses for the other NN. The input is composed of the velocity gradient tensor and grid size. Furthermore, for the application of an NN-based SGS model trained with one flow to another flow, we modify the NN by eliminating bias and introducing a leaky rectified linear unit function as an activation function. The present recursive SGS model is applied to forced homogeneous isotropic turbulence (FHIT) and successfully predicts FHIT at high Reynolds numbers. The present model trained from FHIT is also applied to decaying homogeneous isotropic turbulence and shows an excellent prediction performance.

Key words: turbulence simulation, turbulence modelling, isotropic turbulence

† Email address for correspondence: choi@snu.ac.kr

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

1. Introduction

Large eddy simulation (LES) is an effective tool for predicting turbulent flow because it does not require significant computational resources towards resolution of smaller eddies. This resource reduction is realized through modelling of eddy motions at scales smaller than the grid size (called subgrid scales). The subgrid-scale (SGS) stress from these SGS motions should be modelled to close the governing equations for LES.

The SGS models have been derived for decades based on turbulence theory and statistical approximation. A subset of these models that favour simplicity and stability are represented as linear eddy-viscosity models. Examples include the Smagorinsky (Smagorinsky 1963) and dynamic Smagorinsky (Germano *et al.* 1991) models, based on the theory that the rate of energy transfer to smaller eddies is counterbalanced by the viscous dissipation within the inertial subrange, WALE model (Nicoud & Ducros 1999), using square of the velocity gradient tensor and analysing proper near-wall scaling for the eddy viscosity, and Vreman model (Vreman 2004), rooted in the principle that the SGS stress should be reduced in near-wall region or vanish in laminar flow. On the other hand, the scale-similarity model (SSM) (Bardina, Ferziger & Reynolds 1980) assumed that the interplay between resolved and modelled eddies can be accurately delineated by the difference between the filtered and doubly filtered velocities. The gradient model (GM) (Clark, Ferziger & Reynolds 1979) was derived from the Taylor expansion of the definition of SGS stress. Nevertheless, these models contain some limitations. The linear eddy-viscosity models, in particular, show very low correlations with true SGS stresses in *a priori* tests (Liu, Meneveau & Katz 1994; Salvetti & Banerjee 1995; Park, Yoo & Choi 2005), while they successfully predict turbulence statistics for various flows in *a posteriori* tests (i.e. actual simulations) (see, for example, Germano *et al.* (1991) for turbulent channel flow, Le Ribault, Sarkar & Stanley (1999) for turbulent jet, Kravchenko & Moin (2000) for flow over a circular cylinder). In contrast, both SSM and GM exhibit improved results in *a priori* tests, but they do not sufficiently dissipate turbulent kinetic energy in actual simulations, thereby leading to numerical instability (Vreman, Geurts & Kuerten 1996, 1997).

An alternative approach for the derivation of an SGS model involves a direct employment of direct numerical simulation (DNS) data. For example, the optimal LES (Langford & Moser 1999; Völker, Moser & Venugopal 2002) adopted stochastic estimation techniques (Adrian *et al.* 1989; Adrian 1990) to diminish the discrepancy between the ideal and simulated flow variables. This approach necessitated the use of multipoint correlation data as inputs, which in turn required corresponding DNS data. Moser *et al.* (2009) presented an optimal LES that did not require DNS data, but the SGS model was created based on the assumption of isotropic flow. Another example is an employment of deep learning techniques with a particular emphasis on artificial neural networks (NNs) (Sarghini, de Felice & Santini 2003; Gamahara & Hattori 2017; Maulik *et al.* 2018). Creating an SGS model through NNs requires the accumulation of extensive data through filtering DNS data. This approach is based on an assumption that there may be a complex but well-defined relationship between the SGS stress and resolved flow variables. The pioneering application of NNs in the calculation of SGS stress, to the best of our knowledge, was implemented by Sarghini *et al.* (2003). They employed an NN to find an optimal turbulent viscosity coefficient for a hybrid model (combined Smagorinsky and similarity models) applied to turbulent channel flow with nine velocity gradients and six resolved Reynolds stresses as inputs, and demonstrated its ability to accurately approximate the coefficient.

With the rapid development of deep learning technology, more advanced techniques have been employed to develop SGS models for diverse applications. NN-based SGS

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

models may be classified into three categories: closed-form, reassembling and direct NN-based SGS models, respectively. The closed-form NN-based SGS models, analogous to the study conducted by Sarghini *et al.* (2003), postulate that SGS models may be derived in a closed form from physical, empirical or analytical tools. The coefficients of these models are adjustable, and are computed using NNs with resolved flow variables as inputs. Wollblad & Davidson (2008) identified the coefficients pertinent to proper orthogonal decomposition for turbulent channel flow through an NN. Beck, Flad & Munz (2019) and Pawar *et al.* (2020) suggested that calculating eddy viscosity could be employed as a surrogate to the direct computation of SGS stress and contribute to enhanced stability. Moreover, Xie *et al.* (2019*a*), Xie, Yuan & Wang (2020*d*) and Wang *et al.* (2021) suggested that the SGS stress can be evaluated through a polynomial function including the strain rate, rotation rate, velocity gradient and grid size, with NNs used to determine the coefficients of the polynomial. Yu *et al.* (2022) and Liu *et al.* (2023) obtained the coefficients of Smagorinsky and helicity SGS models using NN, respectively.

Reassembling NN-based SGS models extract unfiltered variables from filtered ones, and subsequently use these extracted variables for the computation of SGS stress. In this approach, NNs play a prominent role in the unfiltering or filtering process. Maulik *et al.* (2018) engaged NNs in the training for unfiltering (Maulik & San 2017) and filtering for two-dimensional decaying homogeneous isotropic turbulence (DHIT) in LES. The first NN collected filtered vorticity and stream function data across multiple grids and gauged the values of unfiltered vorticity and stream function at a single grid. The second NN was subsequently deployed to approximate the filtered nonlinear term, a crucial element for the determination of the SGS stress. Aligning with this methodology, Yuan, Xie & Wang (2020) employed an NN for the estimation of SGS stresses in three-dimensional forced homogeneous isotropic turbulence (FHIT), specifically using the NN to unfilter the velocities.

Direct NN-based SGS models derive the SGS stress or force directly from resolved flow variables through the application of NNs. Gamahara & Hattori (2017) employed NNs to compute the SGS stress as immediate outputs for turbulent channel flow from four input groups such as the strain rate, rotation rate, wall distance and velocity gradient. Previous direct NN-based SGS models have used various techniques and methods to enhance model performance and stability. These include an *ad hoc* method such as wall-damping function or clipping (Gamahara & Hattori 2017; Maulik *et al.* 2019; Zhou *et al.* 2019), derivation of inputs from multiple grids for single or multiple outputs (Beck *et al.* 2019; Maulik *et al.* 2019; Xie *et al.* 2019*b*; Zhou *et al.* 2019; Pawar *et al.* 2020; Sirignano, MacArt & Freund 2020; Xie *et al.* 2020*a*,*b*; Xie, Wang & Weinan 2020*c*; MacArt, Sirignano & Freund 2021; Stoffer *et al.* 2021; Cheng *et al.* 2022; Guan *et al.* 2022; Liu *et al.* 2022; Guan *et al.* 2023), incorporation of second derivatives of velocities (Wang *et al.* 2018; Xie *et al.* 2019*b*; Pawar *et al.* 2020; Sirignano *et al.* 2020; Xie *et al.* 2020*a*; MacArt *et al.* 2021) and consideration of wall distance or filter size as inputs (Gamahara & Hattori 2017; Zhou *et al.* 2019; Abekawa *et al.* 2023). Others (Sirignano *et al.* 2020; MacArt *et al.* 2021; Guan *et al.* 2023; Sirignano & MacArt 2023) defined the training error (or objective function) using flow variables other than the SGS stresses. Sirignano *et al.* (2020), MacArt *et al.* (2021) and Sirignano & MacArt (2023) trained NNs with adjoint-based, PDE (partial differential equation)-constrained optimization methods. Other improvements involve retraining NNs through transfer learning for higher Reynolds number flows (Subel *et al.* 2021; Guan *et al.* 2022) and training NNs in complex flows (MacArt *et al.* 2021; Sirignano & MacArt 2023; Kim, Park & Choi 2024). On the other hand, other types of techniques to improve the performance of NN-based SGS models have also been explored, such as the use of a convolutional neural network (Beck *et al.* 2019; Pawar *et al.* 2020; Guan *et al.* 2022;

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

Liu *et al.* 2022; Guan *et al.* 2023), reinforcement learning (Novati, de Laroussilhe & Koumoutsakos 2021; Kim *et al.* 2022; Kurz, Offenhäuser & Beck 2023) and a graph neural network (Abekawa *et al.* 2023).

Park & Choi (2021) developed a direct NN-based SGS model for turbulent channel flow whose input and output were the velocity gradient or strain rate and the SGS stress, respectively. They performed a comprehensive analysis of the salient characteristics inherent to direct NN-based SGS models, and demonstrated from *a posteriori* test without introducing *ad hoc* techniques such as wall-damping function or clipping that SGS models relying on a single grid input can compute turbulence statistics with a reasonably decent level of accuracy. In addition, two specific results from this study are noteworthy: first, when the LES grid size falls within the range of filter sizes for training NN, the SGS model successfully predicts the flow; second, the SGS model is also successful in predicting the flow at higher Reynolds number when the grid size in wall units is the same as that trained. Hence, these observations indicate that the applicability of direct NN-based SGS models depends on the dimensionless LES grid size relative to the trained grid (or filter) size. For the development of a general NN-based SGS model for complex flow, it is essential that it should function properly over a wide range of non-dimensional filter or grid sizes.

Despite their promising characteristics, previous direct NN-based SGS models have confronted substantial challenges when aimed for application to untrained flows. These challenges originate from the difficulty in developing universal non-dimensional input and output variables for various flows and the nonlinear property (and thus occurrence of potentially large errors in extrapolation) of NN. To achieve regularized input and output values, various techniques have been explored, including the min-max method (Xie *et al.* 2019*b*; Pawar *et al.* 2020; Wang *et al.* 2021), Gaussian normalization method (Stoffer *et al.* 2021; Cheng *et al.* 2022; Liu *et al.* 2022), normalization with root-mean-square (r.m.s.) values (Xie *et al.* 2020*a*,*b*,*c*; Guan *et al.* 2022, 2023), non-dimensionalization in wall units (Park & Choi 2021; Kim *et al.* 2022), and prescribed velocity and length scales (MacArt *et al.* 2021). However, each method has its own limitations: e.g. it requires an equilibrium state or homogeneous direction(s), does not show generalizability when switched to other flows, or proves to be infeasible as it requires DNS results. Consequently, a plausible approach towards general direct NN-based SGS models would be to employ only local variables for non-dimensionalization, like those described by Prakash, Jansen & Evans (2022) and Abekawa *et al.* (2023). On the other hand, two strategies may be proposed for the issue of extrapolation errors of NN. The first strategy is to discover normalized flow variables and SGS stresses that are bounded within a certain range for various flows. The second strategy is to incorporate all accessible data to cover an exhaustive range of flow fields. The former presents considerable challenges, because it requires a formidable task of pinpointing normalized flow variables that remain invariant with respect to factors like the Reynolds number, filter size and flow topology. The latter may demand DNS data of various flows at fairly high Reynolds numbers and thus imposes a huge computational burden.

Therefore, the objective of the present study is to devise a new NN-based SGS model designed to overcome the limitations of existing direct NN-based SGS models. We modify the structure of an NN and formulate a recursive procedure to generate an SGS model valid for a wide range of grid sizes. Here, an NN is trained using fDNS data at low Reynolds number and then updated by data at higher Reynolds number collected through filtered LES data. To validate the performance of the present recursive NN-based SGS model, several SGS models are employed to simulate FHIT at various Reynolds numbers. In addition, two DHIT cases are simulated using the NN trained with FHIT. Section 2 provides the LES framework, NN-based SGS models under consideration, FHIT, and

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

filtering method employed. Section 3 describes the recursive algorithm for constructing recursive NN-based SGS models, and provides the results of *a priori* and *a posteriori* tests. In § 4, the present SGS models are applied to FHIT at high Reynolds numbers and DHIT, and the results are discussed, followed by conclusions in § 5.

2. Numerical details

2.1. *Large eddy simulation*

In LES, the flow variables are filtered using the following operation (Pope 2000):

$$\bar{\phi}(\boldsymbol{x}, t) = \int \bar{G}(\boldsymbol{r})\phi(\boldsymbol{x} - \boldsymbol{r}, t)\,\mathrm{d}\boldsymbol{r}, \tag{2.1}$$

$$\int \bar{G}(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} = 1, \tag{2.2}$$

where $\bar{\phi}$ is a filtered flow variable, $\bar{G}$ is a filter function, $\boldsymbol{x}$ and $\boldsymbol{r}$ denote position vectors, and t is time. The spatially filtered continuity and Navier–Stokes equations for incompressible flows are

$$\frac{\partial \bar{u}_i}{\partial x_i} = 0, \tag{2.3}$$

$$\frac{\partial \bar{u}_i}{\partial t} + \frac{\partial \bar{u}_i \bar{u}_j}{\partial x_j} = -\frac{1}{\rho}\frac{\partial \bar{p}^*}{\partial x_i} + \nu\frac{\partial^2 \bar{u}_i}{\partial x_j \partial x_j} - \frac{\partial \tau_{ij}^r}{\partial x_j}, \tag{2.4}$$

where x_i are the coordinates (x, y, z), u_i are the corresponding velocities (u, v, w), and ρ and ν are the fluid density and kinematic viscosity, respectively (Pope 2000). The effect of the SGS eddy motions is modelled as the anisotropic part of SGS stress, τ_{ij}^r:

$$\tau_{ij} = \overline{u_i u_j} - \bar{u}_i \bar{u}_j, \tag{2.5}$$

$$\tau_{ij}^r = \tau_{ij} - \tfrac{1}{3}\tau_{kk}\delta_{ij}, \tag{2.6}$$

where δ_{ij} is the Kronecker delta. The filtered pressure $\bar{p}^*$ includes the isotropic components of SGS stress:

$$\bar{p}^* = \bar{p} + \tfrac{1}{3}\rho\tau_{kk}. \tag{2.7}$$

In the case of HIT, the filtered continuity and Navier–Stokes equations can be transformed in a spectral space. The pseudo-spectral method is used for spatial discretization, and the zero-padding method, augmented with the 3/2 rule, is applied to control aliasing errors. For temporal integration, a third-order Runge–Kutta method is used for the convection term, and the second-order Crank–Nicolson method is applied to the diffusion term.

To evaluate the performance of an NN-based SGS model relative to those of traditional SGS models, LESs with the constant Smagorinsky model (CSM, Smagorinsky 1963), dynamic Smagorinsky model (DSM, Germano *et al.* 1991; Lilly 1992), gradient model (GM, Clark *et al.* 1979) and dynamic mixed model (DMM, Anderson & Meneveau 1999) are carried out. The Smagorinsky models determine the anisotropic component of SGS stress from

$$\tau_{ij}^r = -2(C_s\bar{\Delta})^2|\bar{S}|\bar{S}_{ij}, \tag{2.8}$$

where C_s is the Smagorinsky model coefficient, $\bar{\Delta}$ is the grid size, $\bar{S}_{ij} = (1/2)(\partial\bar{u}_i/\partial x_j + \partial\bar{u}_j/\partial x_i)$ and $|\bar{S}| = \sqrt{2\bar{S}_{ij}\bar{S}_{ij}}$. For CSM, C_s is a fixed value of 0.17 (Pope 2000). For DSM,

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

the Smagorinsky model coefficient is obtained as

$$C_s^2 = \max[\langle L_{ij}^r M_{ij}\rangle_h / \langle M_{ij} M_{ij}\rangle_h, 0], \tag{2.9}$$

$$L_{ij} = \widetilde{\bar{u}_i \bar{u}_j} - \tilde{\bar{u}}_i \tilde{\bar{u}}_j, \tag{2.10}$$

$$L_{ij}^r = L_{ij} - \tfrac{1}{3} L_{kk} \delta_{ij}, \tag{2.11}$$

$$M_{ij} = -2\tilde{\Delta}^2 |\tilde{\bar{S}}| \tilde{\bar{S}}_{ij} + 2\bar{\Delta}^2 \widetilde{|\bar{S}| \bar{S}_{ij}}, \tag{2.12}$$

where $\overline{(\cdot)}$ and $\widetilde{(\cdot)}$ correspond to the grid- and test-filtering operations, respectively, and $\tilde{\Delta} = 2\bar{\Delta}$. The notation $\langle\cdot\rangle_h$ denotes an instantaneous averaging over homogeneous directions. The SGS stress in GM is derived through the application of Taylor expansions of the filtered velocity (Clark *et al.* 1979; Vreman *et al.* 1996) as

$$\tau_{ij}^r = \frac{1}{12}\left(\beta_{ij} - \frac{1}{3}\beta_{kk}\delta_{ij}\right), \tag{2.13}$$

$$\beta_{ij} = \sum_l \bar{\Delta}_l^2 \frac{\partial \bar{u}_i}{\partial x_l} \frac{\partial \bar{u}_j}{\partial x_l}, \tag{2.14}$$

where $\bar{\Delta}_l$ is the grid size in the x_l direction. The DMM (Anderson & Meneveau 1999) is a combination of CSM and GM with the model coefficients C_s and C_g that are dynamically determined:

$$\tau_{ij}^r = -2(C_s \bar{\Delta})^2 |\bar{S}| \bar{S}_{ij} + C_g(\beta_{ij} - \tfrac{1}{3}\beta_{kk}\delta_{ij}), \tag{2.15}$$

where

$$C_s^2 = \max\left[\frac{\langle L_{ij}^r M_{ij}\rangle_h \langle N_{ij} N_{ij}\rangle_h - \langle L_{ij}^r N_{ij}\rangle_h \langle M_{ij} N_{ij}\rangle_h}{\langle M_{ij} M_{ij}\rangle_h \langle N_{ij} N_{ij}\rangle_h - \langle M_{ij} N_{ij}\rangle_h \langle M_{ij} N_{ij}\rangle_h}, 0\right], \tag{2.16}$$

$$C_g = \frac{\langle L_{ij}^r N_{ij}\rangle_h \langle M_{ij} M_{ij}\rangle_h - \langle L_{ij}^r M_{ij}\rangle_h \langle M_{ij} N_{ij}\rangle_h}{\langle M_{ij} M_{ij}\rangle_h \langle N_{ij} N_{ij}\rangle_h - \langle M_{ij} N_{ij}\rangle_h \langle M_{ij} N_{ij}\rangle_h}, \tag{2.17}$$

$$N_{ij} = \left(B_{ij} - \frac{1}{3} B_{kk}\delta_{ij}\right) - \left(\tilde{\beta}_{ij} - \frac{1}{3}\tilde{\beta}_{kk}\delta_{ij}\right), \tag{2.18}$$

$$B_{ij} = \sum_l \tilde{\Delta}_l^2 \frac{\partial \tilde{\bar{u}}_i}{\partial x_l} \frac{\partial \tilde{\bar{u}}_j}{\partial x_l}. \tag{2.19}$$

2.2. *NN-based SGS model*

So far, most NN-based SGS models have optimized the weights and biases of the NNs, despite using different deep learning techniques (see § 1). For example, Park & Choi (2021) implemented an NN consisting of two hidden layers and 128 neurons per hidden layer to

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

compute six components of the SGS stress:

$$\left.\begin{aligned}
h_i^{(1)} &= \bar{q}_i && (i = 1, 2, \ldots, N_q),\\
h_j^{(2)} &= \max\left[0, \gamma_j^{(2)}\left(\sum_{i=1}^{N_q} W_{ij}^{(1)(2)} h_i^{(1)} + b_j^{(2)} - \mu_j^{(2)}\right)/\sigma_j^{(2)} + \beta_j^{(2)}\right] && (j = 1, 2, \ldots, 128),\\
h_k^{(3)} &= \max\left[0, \gamma_k^{(3)}\left(\sum_{j=1}^{128} W_{jk}^{(2)(3)} h_j^{(2)} + b_k^{(3)} - \mu_k^{(3)}\right)/\sigma_k^{(3)} + \beta_k^{(3)}\right] && (k = 1, 2, \ldots, 128),\\
h_l^{(4)} &= s_l = \sum_{k=1}^{128} W_{kl}^{(3)(4)} h_k^{(3)} + b_l^{(4)} && (l = 1, 2, \ldots, 6).
\end{aligned}\right\} \tag{2.20}$$

Here, $\bar{\boldsymbol{q}}$ is the grid-filtered input, N_q is the number of the input components, $\boldsymbol{W}^{(m)(m+1)}$ is the weight matrix between mth and $(m+1)$th layers, $\boldsymbol{b}^{(m)}$ is the bias vector of the mth layer, $\boldsymbol{s}$ is the output (six components of the SGS stress), and $\boldsymbol{\gamma}^{(m)}$, $\boldsymbol{\mu}^{(m)}$, $\boldsymbol{\sigma}^{(m)}$ and $\boldsymbol{\beta}^{(m)}$ are the parameters for batch normalization (Ioffe & Szegedy 2015). During an NN training process, the parameters ($\boldsymbol{W}^{(m)(m+1)}$, $\boldsymbol{b}^{(m)}$, $\boldsymbol{\gamma}^{(m)}$, $\boldsymbol{\mu}^{(m)}$, $\boldsymbol{\sigma}^{(m)}$ and $\boldsymbol{\beta}^{(m)}$) are optimized to minimize the training error.

In the present study, we construct an NN-based SGS model using a dual NN architecture (figure 1), where the output of one NN (NN_N) is the SGS normal stresses (τ_{11}^r, τ_{22}^r, τ_{33}^r) and that of the other (NN_S) is the SGS shear stresses (τ_{12}^r, τ_{13}^r, τ_{23}^r). The reason for using two NNs is that the ranges of the normal and shear SGS stresses are quite different from each other, and thus separate treatments of these SGS stresses increase the prediction capability of the present NN-based SGS model. Each of the present NNs consists of two hidden layers and 64 neurons per hidden layer, and the output of the mth layer, $\boldsymbol{h}^{(m)}$, is computed as

$$\left.\begin{aligned}
h_i^{(1)} &= \bar{q}_i && (i = 1, 2, \ldots, 9),\\
h_j^{(2)} &= \max[0.02 r_j^{(2)}, r_j^{(2)}], \quad r_j^{(2)} = \sum_{i=1}^{9} W_{ij}^{(1)(2)} h_i^{(1)} && (j = 1, 2, \ldots, 64),\\
h_k^{(3)} &= \max[0.02 r_k^{(3)}, r_k^{(3)}], \quad r_k^{(3)} = \sum_{j=1}^{64} W_{jk}^{(2)(3)} h_j^{(2)} && (k = 1, 2, \ldots, 64),\\
h_l^{(4)} &= s_l = \sum_{k=1}^{64} W_{kl}^{(3)(4)} h_k^{(3)} && (l = 1, 2, 3).
\end{aligned}\right\} \tag{2.21}$$

Note that we remove the bias ($\boldsymbol{b}$) and batch normalization parameters ($\boldsymbol{\gamma}$, $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\beta}$). Also, we use the leaky rectified linear unit function (leaky ReLU) as an activation function: $f(x) = \max[ax, x]$, which was proposed by Maas, Hannun & Ng (2013) to overcome the vanishing gradient problem (Bengio, Simard & Frasconi 1994). The negative slope is determined to be $0.01 \leq a \leq 0.2$ (Xu *et al.* 2015), and we choose $a = 0.02$. Note that one of the reasons for introducing the bias $\boldsymbol{b}$ is to produce non-zero output even if the input is zero (Goodfellow, Bengio & Courville 2016). For the present flow problem, the bias is not necessarily needed because the output should be zero when the input is zero. The effects of removing the bias and batch normalization parameters and replacing ReLU with leaky ReLU on the results of *a priori* and *a posteriori* tests are discussed in Appendix A, where

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

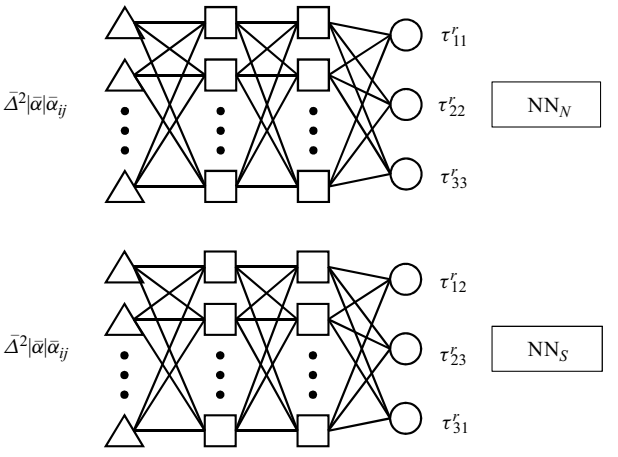


Figure 1. Schematic diagram of the present NN-based SGS model that consists of a dual NN: NN_N and NN_S predict the SGS normal and shear stresses, respectively.

we show that setting $\boldsymbol{b} = \boldsymbol{\mu} = \boldsymbol{\beta} = 0,\ \boldsymbol{\gamma} = \boldsymbol{\sigma} = 1$ and replacing ReLU with leaky ReLU do not deteriorate the numerical solutions for the present flow.

Without the bias and batch normalization parameters and with the use of leaky ReLU, the present NN complies with the following conditions:

$$\mathbf{NN}(c\boldsymbol{A}) = c\mathbf{NN}(\boldsymbol{A}) \quad \text{only if } c \geq 0, \tag{2.22}$$

$$\mathbf{NN}(\boldsymbol{A} + \boldsymbol{B}) \neq \mathbf{NN}(\boldsymbol{A}) + \mathbf{NN}(\boldsymbol{B}), \tag{2.23}$$

where c is a positive scalar, and $\boldsymbol{A}$ and $\boldsymbol{B}$ are arbitrary tensors ($\boldsymbol{A} \neq \boldsymbol{B}$). For example, for $c = 2$ and $\boldsymbol{B} = 2\boldsymbol{A}$, $\mathbf{NN}(c\boldsymbol{A}) = c\mathbf{NN}(\boldsymbol{A})$ and $\mathbf{NN}(\boldsymbol{A} + \boldsymbol{B}) = \mathbf{NN}(\boldsymbol{A}) + \mathbf{NN}(\boldsymbol{B})$, but for $c = 2$ and $\boldsymbol{B} = -2\boldsymbol{A}$, $\mathbf{NN}(c\boldsymbol{A}) = c\mathbf{NN}(\boldsymbol{A})$ and $\mathbf{NN}(\boldsymbol{A} + \boldsymbol{B}) \neq \mathbf{NN}(\boldsymbol{A}) + \mathbf{NN}(\boldsymbol{B})$ due to the leaky ReLU used. Thus, the present NN still retains its nonlinearity, making it suitable for nonlinear regression between the SGS stresses and local flow variables. Equation (2.22) allows us to apply an NN trained from one flow to another flow. Let us show an example for turbulent channel flow and flow over a circular cylinder. Assume that the input and output variables are $\bar{\Delta}^2|\bar{\alpha}|\bar{\alpha}_{ij}$ and τ^r_{ij} for these two flows, where $\bar{\Delta} = (\bar{\Delta}_x\bar{\Delta}_y\bar{\Delta}_z)^{1/3}$, $\bar{\alpha}_{ij} = \partial\bar{u}_i/\partial x_j$ and $|\bar{\alpha}| = \sqrt{\bar{\alpha}_{ij}\bar{\alpha}_{ij}}$. An NN is constructed from turbulent channel flow with input and output variables normalized by the wall-shear velocity u_τ: i.e. $\tau^r_{ij}/u^2_\tau = \mathbf{NN}(\bar{\Delta}^2|\bar{\alpha}|\bar{\alpha}_{ij}/u^2_\tau)$. Then, this trained NN can be used for flow over a circular cylinder with input and output variables normalized by the free-stream velocity u_∞ from the following relation:

$$\frac{\tau^r_{ij}}{u^2_\infty} = \frac{u^2_\tau}{u^2_\infty}\frac{\tau^r_{ij}}{u^2_\tau} = \frac{u^2_\tau}{u^2_\infty}\mathbf{NN}\left(\frac{\bar{\Delta}^2|\bar{\alpha}|\bar{\alpha}_{ij}}{u^2_\tau}\right) = \mathbf{NN}\left(\frac{u^2_\tau}{u^2_\infty}\frac{\bar{\Delta}^2|\bar{\alpha}|\bar{\alpha}_{ij}}{u^2_\tau}\right) = \mathbf{NN}\left(\frac{\bar{\Delta}^2|\bar{\alpha}|\bar{\alpha}_{ij}}{u^2_\infty}\right). \tag{2.24}$$

During the training process, the weight parameters $\boldsymbol{W}^{(m)(m+1)}$ are optimized to minimize the mean square error (loss) given by

$$L = \frac{1}{3}\frac{1}{N_b}\sum_{l=1}^{3}\sum_{n=1}^{N_b}(s^{fDNS}_{l,n} - s_{l,n})^2, \tag{2.25}$$

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

where N_b denotes the minibatch size of 256, and $s_{l,n}^{fDNS}$ corresponds to the anisotropic components τ_{ij}^r obtained from fDNS. Here, the loss of each minibatch is calculated and then the weights are updated for every minibatch.

We apply the Adam algorithm (a type of gradient descent, Kingma & Ba 2014) and learning rate annealing method (Simonyan & Zisserman 2014; He *et al.* 2016) to optimize the weights and enhance the training speed, respectively. Early stopping (Goodfellow *et al.* 2016) is employed in the training process to avoid overfitting. The learning rate is initially 0.025, and is subsequently reduced by a factor of 10 if no improvement in the training error is observed over five epochs. If the learning rate is decreased three times but there is no reduction in the training error over five epochs, training is stopped. The training requires 150 to 300 epochs. The entire process to train and execute the present NN is carried out with the PyTorch open-source library in the Python programming environment.

We develop an NN-based SGS model whose input and output are $\bar{\Delta}^2|\bar{\alpha}|\bar{\alpha}_{ij}$ and τ_{ij}^r, respectively (NN-based velocity gradient model, called NNVGM hereafter). Note that the dimensions of input and output are matched to be the same. The velocity gradient tensor as an input has been used by many previous studies (Gamahara & Hattori 2017; Wang *et al.* 2018; Xie *et al.* 2019*b*; Zhou *et al.* 2019; Pawar *et al.* 2020; Prat, Sautory & Navarro-Martinez 2020; Xie *et al.* 2020*a*,*b*,*c*; Park & Choi 2021; Kim *et al.* 2022; Abekawa *et al.* 2023; Kim *et al.* 2024), and the same form of input was considered by Jamaat & Hattori (2022) for one-dimensional Burgers turbulence. The NNVGM obtains inputs from each grid point and computes corresponding SGS stresses.

2.3. *Forced homogeneous isotropic turbulence (FHIT) and filtering method*

FHIT involves an additional forcing term in the Navier–Stokes equations at low wavenumbers (Ghosal *et al.* 1995; Rosales & Meneveau 2005; Park *et al.* 2006):

$$f_i = \epsilon_t \frac{u_i}{\sum_{0<|\boldsymbol{k}|<2} |\hat{u}_j(\boldsymbol{k})|^2}, \tag{2.26}$$

where f_i is the forcing term, $\hat{u}_i$ is the Fourier coefficient of u_i, ϵ_t is the prescribed mean total dissipation rate determining the turbulent energy injection rate, $|\mathbf{k}| = \sqrt{k_x^2 + k_y^2 + k_z^2}$, and k_i is the wavenumber in the i direction. Note that ϵ_t encompasses both the resolved and modelled dissipation.

In FHIT, two distinct Reynolds numbers, $Re_\lambda = u_{rms}\lambda/\nu$ and $Re_L = UL/\nu$, can be defined. Here, u_{rms} is the r.m.s. velocity fluctuations, $\lambda(=\sqrt{15u_{rms}^2\nu/\epsilon_t})$ is the Taylor microscale, L is associated with the computational domain size of $2\pi L \times 2\pi L \times 2\pi L$, and U is the characteristics velocity such that $U = (\epsilon_t L)^{1/3}$. The Taylor microscale Reynolds number Re_λ has been widely used in turbulence studies. Nevertheless, the extraction of u_{rms} requires intricate experimental data or DNS. In contrast, Re_L does not require such preliminary outcomes, and can be deduced from $\epsilon_t L/U^3 = 1$, $\eta k_{max,DNS}$ and N_{DNS}, where η is the Kolmogorov length scale, and $k_{max,DNS}$ and N_{DNS} are the largest wavenumber and number of grid points in each direction in DNS, respectively. According to Pope (2000), $\eta k_{max,DNS} = 3/2$ or $\Delta_{DNS}/\eta = 2\pi/3$ is set to achieve sufficiently high resolution in DNS, where Δ_{DNS} is the grid size in DNS. Then, $L = \Delta_{DNS}N_{DNS}/(2\pi) = \eta N_{DNS}/3$. With these relations together with $\eta = (\nu^3/\epsilon_t)^{1/4}$, one can obtain $Re_L = (N_{DNS}/3)^{4/3}$. Table 1 lists the cases of DNS at various Reynolds numbers by changing the number of grid points for FHIT, where each case is named as DNSN_{DNS}. Note that the cases listed in table 1 do not necessarily require actual DNS except for the case of DNS128 (the recursive procedure

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

Case	N_{DNS}	Re_L	Re_λ
DNS128	128	149.09	93.03
DNS256	256	375.69	154.17
DNS512	512	946.67	—
DNS1024	1024	2385.46	—
DNS2048	2048	6010.98	—
DNS4096	4096	15146.71	—
DNS8192	8192	38167.31	—
DNS16384	16 384	96175.60	—
DNS32768	32 768	242347.33	—

Table 1. Cases of DNS at various Reynolds numbers for FHIT. Here, $Re_L = (N_{DNS}/3)^{4/3}$ and $\Delta_{DNS}/\eta = 2\pi/3$ ($N_{DNS}\Delta_{DNS} = 2\pi L$). DNSs are performed for DNS128 and DNS256, and the corresponding Re_λ values are given in this table. Here, Δ_{DNS} and N_{DNS} are the size and number of grid points in each direction in DNS, respectively.

Case	N_c	Re_L	$\bar{\Delta}/\eta$
fDNS32/128 or LES32/128	32	149.09	8.38
fDNS64/128	64	149.09	4.19
fDNS32/256 or LES32/256	32	375.69	16.76
fDNS64/256 or LES64/256	64	375.69	8.38
fDNS128/256	128	375.69	4.19
LES64/512	64	946.67	16.76
LES64/1024	64	2385.46	33.51

Table 2. Cases of fDNS and LES at various Reynolds numbers for FHIT. Here, N_c is the number of grid points in each direction for fDNS or LES, and $\bar{\Delta}$ denotes the filter size for fDNS ($\bar{\Delta}_{fDNS}$) or the grid size for LES ($\bar{\Delta}_{LES}$). The name of each case indicates fDNSN_c/N_{DNS} or LESN_c/N_{DNS}. Note that $\Delta_{DNS}/\eta = 2\pi/3$.

introduced below requires DNS only at a low Reynolds number), and LESs are conducted for higher Reynolds number cases through the recursive procedure.

Table 2 shows the cases considered for fDNS and LES of FHIT. The fDNS data are obtained by applying a transfer function $\hat{G}(\mathbf{k})$ to the Fourier-transformed DNS data $\hat{u}_i(\mathbf{k})$:

$$\hat{\bar{u}}_i(\boldsymbol{k}) = \hat{u}_i(\boldsymbol{k})\hat{\bar{G}}(\boldsymbol{k}), \tag{2.27}$$

$$\hat{\bar{G}}(\boldsymbol{k}) = \int \exp(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r})\bar{G}(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r}. \tag{2.28}$$

The fDNS data can be obtained by applying a spectral cutoff filter or Gaussian filter in the spectral space. However, these filterings have limitations when the filtered variables are compared with LES results. The spectral cutoff filtering removes the Fourier coefficients $\hat{u}_i$ at $k > k_c$, and allows filtered variables to be placed on coarser grids of LES, where $k_c(= \pi/\bar{\Delta})$ is the cutoff wavenumber. However, this filtering fails to satisfy the realizability condition due to negative weights in the physical space (Vreman, Geurts & Kuerten 1994). Additionally, the spectral cutoff filtering may exhibit non-local oscillations (Meneveau & Katz 2000; Pope 2000). On the other hand, the Gaussian filtering is free from these drawbacks, and requires filtering even when the wavenumber exceeds the LES limit (i.e. $k > \pi/\bar{\Delta}_{LES}$). Although the filter size may coincide with the LES grid size, the Gaussian filtered data are allocated at DNS grids, not at LES grids.

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

Filter	$\hat{\bar{G}}(\mathbf{k})$	Largest wavenumber of filtered variables
Spectral cutoff	$H(k_c - k)$	k_c
Gaussian	$\exp(-k^2\bar{\Delta}^2/24)$	$k_{DNS}(=\pi/\Delta_{DNS})$
Cut-Gaussian	$\exp(-k^2\bar{\Delta}^2/24)$	k_c

Table 3. Comparison of three filters. Here, H is the Heaviside step function.

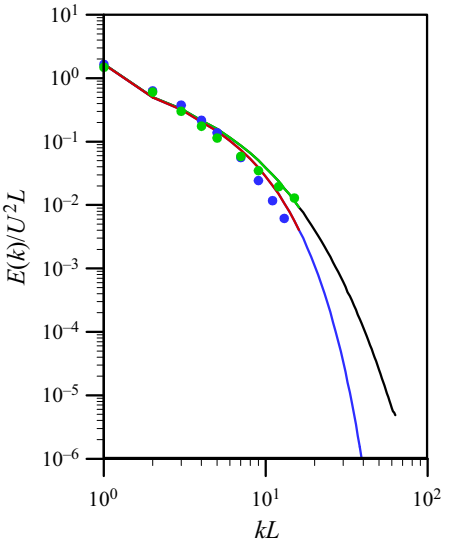


Figure 2. Three-dimensional energy spectra: black line, DNS128 ($Re_L = 149.09$; $Re_\lambda = 93.03$); red line, fDNS32/128 with cut-Gaussian filtering; blue line, fDNS32/128 with Gaussian filtering; green line, fDNS32/128 with spectral cutoff filtering; blue ○, LES32/128 with DSM; green ○, LES32/128 with GM.

In the present study, we use the cut-Gaussian filtering (Zanna & Bolton 2020; Guan *et al.* 2022, 2023; Pawar *et al.* 2023) which retains the transfer function of Gaussian filtering but truncates the filtered variables at the wavenumbers exceeding the cutoff wavenumber k_c. Table 3 shows three different filters in the spectral space, corresponding transfer functions and largest wavenumbers of filtered variables, respectively. Figure 2 shows the energy spectra from DNS128, fDNS32/128 using cut-Gaussian, Gaussian and spectral cutoff filters, and LES32/128 with DSM and GM, respectively. The energy spectrum with GM is in excellent agreement with that of fDNS using the spectral cutoff filter. In contrast, the energy spectrum with DSM is closer to that of fDNS using the cut-Gaussian filter. Given the increased adaptability of DSM over GM, we measure the prediction capability of the SGS models based on the cut-Gaussian filtered DNS data.

3. A recursive NN-based SGS model (NNVGM)

3.1. *A recursive algorithm*

The present study employs a recursive procedure to construct an NNVGM adapted to various grid sizes normalized by the Kolmogorov length scale from the following steps:

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

(i) obtain training data (fDNS data) by filtering DNS data at a low Reynolds number (Re_L);
(ii) train an NNVGM with the fDNS data;
(iii) apply the NNVGM to LES at a higher Reynolds number, where the ratio of LES grid size to the Kolmogorov length scale ($\bar{\Delta}_{LES}/\eta$) is equal to that of filter size to the Kolmogorov length scale ($\bar{\Delta}_{fDNS}/\eta$) (see Appendix B for the effect of different filter to grid ratios);
(iv) filter the LES data and include the filtered LES (fLES) data in the training dataset, where the new filter size is twice the LES grid size;
(v) train the NNVGM using the augmented training data;
(vi) apply the updated NNVGM to LES at even higher Reynolds number, where the LES grid size is equal to the filter size defined in Step (iv);
(vii) repeat steps (iv)–(vi) for LES at higher Reynolds numbers.

Steps (i) and (ii) are similar to those used in constructing previous NN-based SGS models. The present recursive procedure aims at improving the performance of an NNVGM at a wider range of non-dimensional grid sizes by recursively performing LES and training it using fLES data.

3.2. *NNVGM trained with fDNS data: a priori and a posteriori tests*

First, fDNS data are used to train an NN. The Reynolds number and number of grid points in each direction for DNS are $Re_L = 149.09$ and $N_{DNS} = 128$, respectively (table 1). The filter sizes considered are $\bar{\Delta}_{fDNS}/\eta = 8.39$ and 4.19, corresponding to the cases of fDNS32/128 and fDNS64/128, respectively (table 2). Including two different filter sizes in constructing fDNS data helps to improve the prediction capability of NNVGMs (see, for example, Park & Choi 2021) by broadening the ranges of the input and output, when they are appropriately normalized. In the present study, we train a dual NN (NN_N and NN_S in figure 1) with the training data (input and output) normalized by r.m.s. SGS normal ($\tau^r_{11,rms}$) and shear ($\tau^r_{12,rms}$) stress fluctuations, respectively. Note that $\tau^r_{11,rms} (= \tau^r_{22,rms} = \tau^r_{33,rms})$ and $\tau^r_{12,rms} (= \tau^r_{13,rms} = \tau^r_{23,rms})$ are obtained from (2.5) and (2.6) with DNS data. During actual LES, however, the SGS normal and shear stress fluctuations are *a priori* unknown, and thus providing normalized input data to the NN is not possible. Hence, in actual LES, we can only provide the input normalized by the characteristic velocity scale ($\bar{\Delta}^2|\bar{\alpha}|\bar{\alpha}_{ij}/U^2_{LES}$). Thanks to the important property of the present NN, (2.22), we obtain the following relation (no summation on i and j):

$$\frac{\tau^r_{ij}}{U^2_{LES}} = \frac{\tau^r_{ij,rms}}{U^2_{LES}} \frac{\tau^r_{ij}}{\tau^r_{ij,rms}} = \frac{\tau^r_{ij,rms}}{U^2_{LES}} \mathbf{NN}\left(\frac{\bar{\Delta}^2|\bar{\alpha}|\bar{\alpha}_{ij}}{\tau^r_{ij,rms}}\right) = \mathbf{NN}\left(\frac{\bar{\Delta}^2|\bar{\alpha}|\bar{\alpha}_{ij}}{U^2_{LES}}\right). \tag{3.1}$$

This relation indicates that, during actual LES, one can provide the input normalized by U_{LES} to the NN trained with the input normalized by the r.m.s. SGS stresses.

However, the SGS stresses obtained from DNS data have a problem of imbalanced distribution because they tend to be mainly distributed around zero. So, we choose the normalized output and corresponding input through the process known as undersampling (Drummond & Holte 2003; Liu, Wu & Zhou 2008). During this process, we do not use all the filtered data as the training one, but choose data such that the possibility of choosing data as the training one decreases when its magnitude is near to zero: i.e. the probability

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

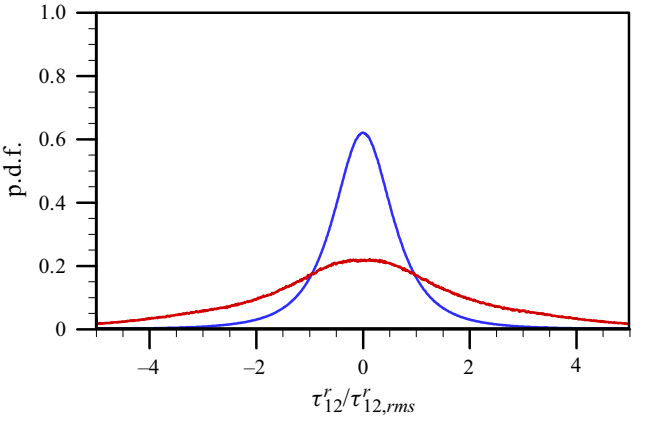


Figure 3. Probability density function (p.d.f.) of normalized τ^r_{12} before (blue) and after (red) undersampling.

($\mathcal{P}$) to choose an SGS shear stress and corresponding inputs as training data is

$$\mathcal{P} = \begin{cases} \sin^2\theta & \text{if } 0 \le \theta < \pi/2 \\ 1 & \text{if } \theta \ge \pi/2 \end{cases}, \tag{3.2}$$

$$\theta = \frac{\pi}{8}\frac{\sqrt{\{(\tau^r_{12})^2 + (\tau^r_{23})^2 + (\tau^r_{13})^2\}/3}}{\tau^r_{12,rms}}. \tag{3.3}$$

Figure 3 illustrates the effect of undersampling on the probability density function (p.d.f.) of training data. This undersampling reduces the probability of zero SGS stresses and enhances the occurrence of strong SGS stresses, leading to generate high SGS dissipation and thus improving the performance of an NNVGM. The SGS normal stresses are also similarly processed. For each fDNS dataset, approximately 600 000 pairs of $\bar{\Delta}^2|\bar{\alpha}|\bar{\alpha}_{ij}$ and τ^r_{ij} are used for training. Other types of the probability $\mathcal{P}$ for undersampling are also considered and the results from different choices of $\mathcal{P}$ are discussed in Appendix C.

We perform *a priori* tests with the present NNVGMs and traditional models using fDNS32/128 ($Re_L = 149.09$) and fDNS64/256 ($Re_L = 375.69$), respectively. Table 4 provides the correlations of the SGS normal and shear stresses and SGS dissipation, and the magnitude of the SGS dissipation from various SGS models. Here, NNVGM(32+64)/128 is trained using fDNS32/128 and fDNS64/128 datasets, and NNVGM(64+128)/256 is done using fDNS64/256 and fDNS128/256 datasets. For both Reynolds numbers, NNVGMs provide the highest correlations of the SGS stresses and SGS dissipation, and the magnitudes of the SGS dissipation closest to those of fDNS data. It is notable that NNVGM(32+64)/128 trained at $Re_L = 149.09$ successfully predicts the correlations and SGS dissipation even when it is applied to a higher Reynolds number of $Re_L = 375.69$.

Figure 4(*a*,*b*) shows the p.d.f.s of ϵ_{SGS}, τ^r_{11} and τ^r_{12} from various SGS models for $Re_L = 149.09$ and 375.69, respectively. As shown, the p.d.f.s of ϵ_{SGS} from NNVGM and GM agree very well with fDNS data. The p.d.f.s from both CSM and DSM appear only at positive SGS dissipation, but do not agree well with fDNS data. However, DMM provides large positive SGS dissipation (its curves nearly fall on those of CSM), leading to overestimated SGS dissipation (see table 4). Similarly, NNVGM and GM predict SGS stresses better than other models. It should be noted that NNVGM(32+64)/128

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

Re_L	SGS model	$R(\tau_{11}^r)$	$R(\tau_{12}^r)$	$R(\epsilon_{SGS})$	ϵ_{SGS}
149.09	fDNS32/128	—	—	—	0.430
	NNVGM(32+64)/128	0.661	0.692	0.647	0.447
	CSM32/128	0.191	0.204	0.569	1.077
	DSM32/128	0.191	0.204	0.569	0.859
	GM32/128	0.653	0.682	0.633	0.315
	DMM32/128	0.649	0.681	0.637	0.958
375.69	fDNS64/256	—	—	—	0.439
	NNVGM(32+64)/128	0.664	0.695	0.662	0.461
	NNVGM(64+128)/256	0.664	0.695	0.662	0.452
	CSM64/256	0.186	0.201	0.578	1.118
	DSM64/256	0.186	0.201	0.577	0.892
	GM64/256	0.656	0.686	0.647	0.324
	DMM64/256	0.654	0.686	0.653	0.965

Table 4. Statistics from *a priori* tests at $Re_L = 149.09$ and 375.69 with $\bar{\Delta}_{fDNS}/\eta = 8.38$: Pearson correlation coefficients (R) of τ_{ij}^r and SGS dissipation $\epsilon_{SGS}(= -\tau_{ij}^r \bar{S}_{ij})$, and magnitude of ϵ_{SGS}.

trained at $Re_L = 149.09$ predicts the p.d.f.s accurately at $Re_L = 375.69$ as much as NNVGM(64+128)/256 does.

Now, we perform *a posteriori* tests (actual LES) with the present NNVGMs and traditional models for $Re_L = 149.09$ and 375.69, and compare the results with fDNS32/128 and fDNS64/256 data, respectively. Figures 5 and 6 show the three-dimensional energy spectra, and p.d.f.s of the SGS dissipation, SGS stresses, strain rates ($\bar{S}_{11}$ and $\bar{S}_{12}$) and rotation rate ($\bar{\Omega}_{23} = 0.5(\partial\bar{u}_3/\partial x_2 - \partial\bar{u}_2/\partial x_3)$), respectively. The energy spectra from NNVGM and GM agree well with fDNS data, whereas CSM, DSM and DMM provide rapid fall-off at high wavenumbers (figure 5). For p.d.f.s of the SGS dissipation and stresses, NNVGM, GM and DMM provide accurate predictions, whereas CSM and DSM do not. For the strain and rotation rates, NNVGM and GM provide the most accurate results (figure 6). These results indicate that LES with NNVGM trained at a Reynolds number predicts turbulence statistics very well for a different (higher) Reynolds number flow if $\bar{\Delta}_{LES}/\eta = \bar{\Delta}_{fDNS}/\eta$. A similar conclusion was also made by Park & Choi (2021), where the grid size was non-dimensionalized in wall units rather than by the Kolmogorov length scale.

3.3. *NNVGM trained with fDNS and fLES data: a priori and a posteriori tests*

As mentioned before, we train the NN with fLES data as well as fDNS data. The theoretical basis for filtering LES data can be attributed to the test filter introduced by Germano *et al.* (1991):

$$\tilde{\phi}(\boldsymbol{x}, t) = \int \phi(\boldsymbol{x} - \boldsymbol{r}, t)\tilde{G}(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r}, \tag{3.4}$$

where $\tilde{G}$ is a filter function and $\tilde{\bar{G}} = \tilde{G}\bar{G}$. The test-filtered SGS stresses are written as

$$T_{ij} = \widetilde{\bar{u}_i\bar{u}_j} - \tilde{\bar{u}}_i\tilde{\bar{u}}_j, \tag{3.5}$$

$$T_{ij}^r = T_{ij} - \tfrac{1}{3}T_{kk}\delta_{ij}, \tag{3.6}$$

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

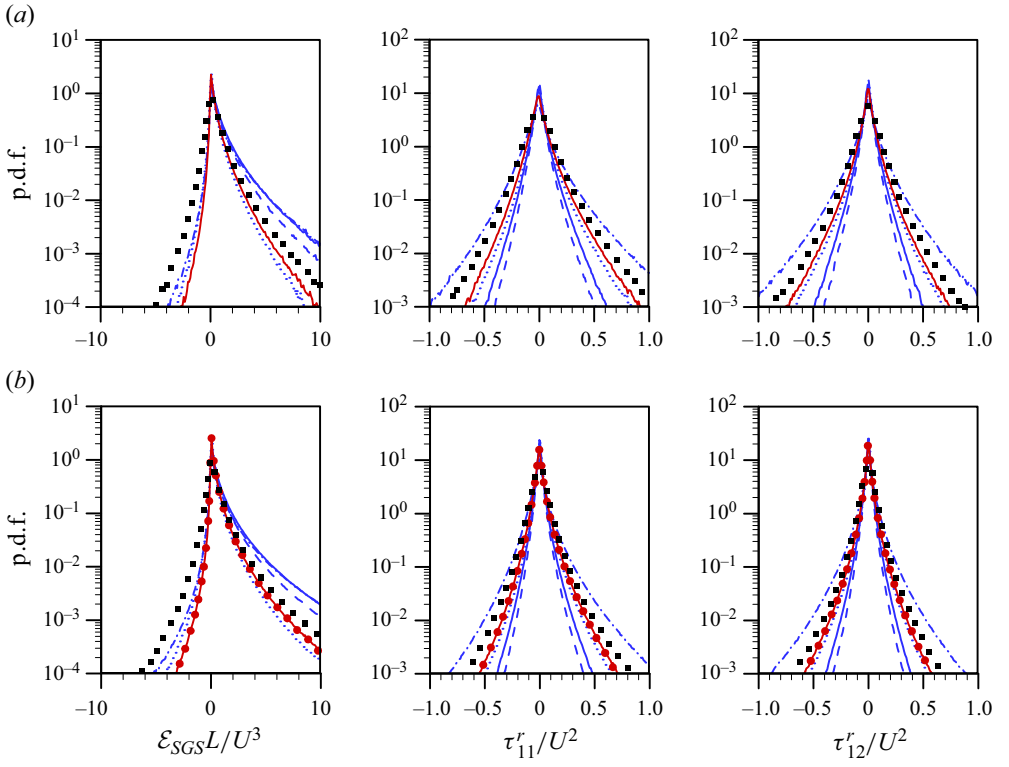


Figure 4. Probability density functions of the SGS dissipation and SGS normal and shear stresses (*a priori* test with $\bar{\Delta}_{fDNS}/\eta = 8.38$): (*a*) $Re_L = 149.09$; (*b*) $Re_L = 375.69$. In panel (*a*), ■, fDNS32/128; red line, NNVGM(32+64)/128; blue line, CSM32/128; blue dashed line, DSM32/128; blue dotted line, GM32/128; blue dash-dotted line, DMM32/128. In panel (*b*), ■, fDNS64/256; red ○, NNVGM(64+128)/256; red line, NNVGM(32+64)/128; blue line, CSM64/256; blue dashed line, DSM64/256; blue dotted line, GM64/256; blue dash-dotted line, DMM64/256.

and T^r_{ij} is obtained by the following relation:

$$T^r_{ij} = \tilde{\tau}^r_{ij} + L^r_{ij}, \tag{3.7}$$

where

$$L_{ij} = \widetilde{\tilde{\bar{u}}_i \tilde{\bar{u}}_j} - \tilde{\tilde{\bar{u}}}_i \tilde{\tilde{\bar{u}}}_j, \tag{3.8}$$

$$L^r_{ij} = L_{ij} - \tfrac{1}{3} L_{kk} \delta_{ij}. \tag{3.9}$$

The present NNVGM is updated through the accumulation of new datasets with the input of $\tilde{\Delta}^2 |\tilde{\bar{\alpha}}| \tilde{\bar{\alpha}}_{ij}$ and the output of T^r_{ij}. The test filter size is twice the LES grid size, i.e. $\tilde{\Delta} = 2\bar{\Delta}_{LES}$. The filtered LES data, referred to as fLES32/256, are obtained by filtering LES64/256 data. The fLES data are selected using the same normalization and undersampling techniques described in § 3.2. The fLES32/512, fLES32/1024, …, fLES32/32768 data can be created by filtering LES64/512, LES64/1024, …, LES64/32768, respectively, during the recursive procedure. Table 5 summarizes the present NNVGMs considered and corresponding training data.

Let us conduct *a priori* and *a posteriori* tests for $Re_L = 375.69$ with $\bar{\Delta}/\eta = 16.76$ (twice that of LES64/256 and LES32/128). During the training process,

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

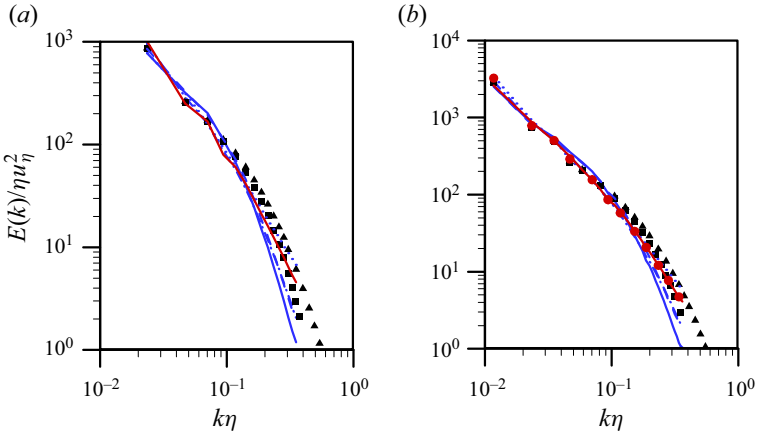


Figure 5. Three-dimensional energy spectra (*a posteriori* test; LES32/128 and LES64/256 for $Re_L = 149.09$ and 375.69, respectively, with $\bar{\Delta}_{LES}/\eta = 8.38$): (*a*) $Re_L = 149.09$; (*b*) $Re_L = 375.69$. In panel (*a*), ▲, DNS128; ■, fDNS32/128; red line, NNVGM(32+64)/128; blue line, CSM32/128; blue dashed line, DSM32/128; blue dotted line, GM32/128; blue dash-dotted line, DMM32/128. In panel (*b*), ▲, DNS256; ■, fDNS64/256; red ○, NNVGM(64+128)/256; red line, NNVGM(32+64)/128; blue line, CSM64/256; blue dashed line, DSM64/256; blue dotted line, GM64/256; blue dash-dotted line, DMM64/256. Here, u_η is the Kolmogorov velocity scale.

SGS model	Training data
NNVGM128D	fDNS32/128, fDNS64/128
NNVGM256DD	fDNS64/256, fDNS128/256
NNVGM(128D+256D)	fDNS32/128, fDNS64/128, fDNS32/256
NNVGM(128D+256L)	fDNS32/128, fDNS64/128, fLES32/256
NNVGM(128D+512L)	fDNS32/128, fDNS64/128, fLES32/256, fLES32/512
NNVGM(128D+1024L)	fDNS32/128, fDNS64/128, fLES32/256, fLES32/512, fLES32/1024
NNVGM(128D+2048L)	fDNS32/128, fDNS64/128, fLES32/256, fLES32/512, fLES32/1024, fLES32/2048
NNVGM(128D+4096L)	fDNS32/128, fDNS64/128, fLES32/256, fLES32/512, fLES32/1024, fLES32/2048, fLES32/4096
NNVGM(128D+8192L)	fDNS32/128, fDNS64/128, fLES32/256, fLES32/512, fLES32/1024, fLES32/2048, fLES32/4096, fLES32/8192
NNVGM(128D+16384L)	fDNS32/128, fDNS64/128, fLES32/256, fLES32/512, fLES32/1024, fLES32/2048, fLES32/4096, fLES32/8192, fLES32/16384
NNVGM(128D+32768L)	fDNS32/128, fDNS64/128, fLES32/256, fLES32/512, fLES32/1024, fLES32/2048, fLES32/4096, fLES32/8192, fLES32/16384, fLES32/32768

Table 5. NNVGMs and corresponding training data. Here, NNVGM128D and NNVGM256DD are the same as NNVGM(32+64)/128 and NNVGM(64+128)/256 discussed in § 3.2, respectively. A directory with the NNs (NNVGM128D, NNVGM(128D+256L), NNVMG(128D+512L), . . ., NNVMG(128D+32768L)) and a customizable Jupyter notebook can be accessed at https://www.cambridge.org/S0022112024009923/JFM-Notebooks/files/Table_5/Table_5.ipynb. Here, the weights $W_{ij}^{(1)(2)}$, $W_{jk}^{(2)(3)}$ and $W_{kl}^{(3)(4)}$ in (2.21) are also accessible for NNVGM128D to NNVGM(128D+32768L). The SGS stresses τ_{ij}^r/U^2 from NNVGMs are calculated with user's input values of $\bar{\alpha}_{ij}L/U$.

NNVGM(128D+256D) employs fDNS data only (fDNS32/128, fDNS64/128 and fDNS32/256), but NNVGM(128D+256L) uses a combination of fDNS and fLES data (fDNS32/128, fDNS64/128 and fLES32/256). Table 6 and figure 7 show the results of *a priori* tests. For the correlation coefficients, the performances of NNVGMs are

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

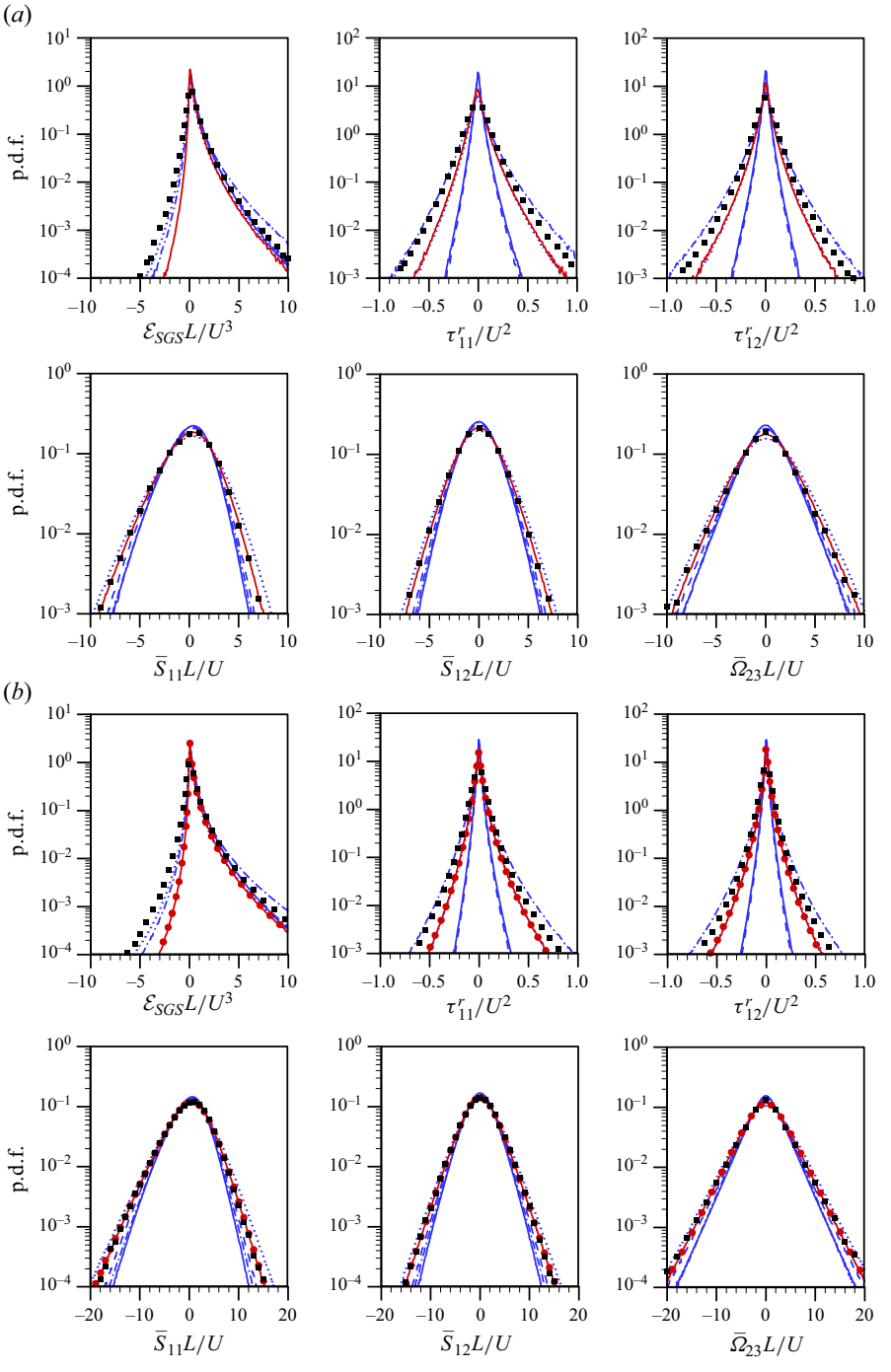


Figure 6. Probability density functions of the SGS dissipation, SGS stresses, strain rates and rotation rate (*a posteriori* test; LES32/128 and LES64/256 for $Re_L = 149.09$ and 375.69, respectively, with $\bar{\Delta}_{LES}/\eta = 8.38$): (*a*) $Re_L = 149.09$; (*b*) $Re_L = 375.69$. In panel (*a*), ■, fDNS32/128; red line, NNVGM(32+64)/128; blue line, CSM32/128; blue dashed line, DSM32/128; blue dotted line, GM32/128; blue dash-dotted line, DMM32/128. In panel (*b*), ■, fDNS64/256; red ○, NNVGM(64+128)/256; red line, NNVGM(32+64)/128; blue line, CSM64/256; blue dashed line, DSM64/256; blue dotted line, GM64/256; blue dash-dotted line, DMM64/256.

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

SGS model	$R(\tau^r_{11})$	$R(\tau^r_{12})$	$R(\epsilon_{SGS})$	ϵ_{SGS}
fDNS32/256	—	—	—	0.734
NNVGM(128D+256D)	0.528	0.563	0.540	0.653
NNVGM(128D+256L)	0.527	0.563	0.539	0.706
CSM32/256	0.196	0.221	0.489	1.397
DSM32/256	0.196	0.221	0.489	1.316
GM32/256	0.511	0.544	0.507	0.378
DMM32/256	0.506	0.543	0.524	1.683

Table 6. Statistics from *a priori* test at $Re_L = 375.69$ with $\bar{\Delta}/\eta = 16.76$: Pearson correlation coefficients (R) of τ^r_{ij} and ϵ_{SGS}, and magnitude of ϵ_{SGS}.

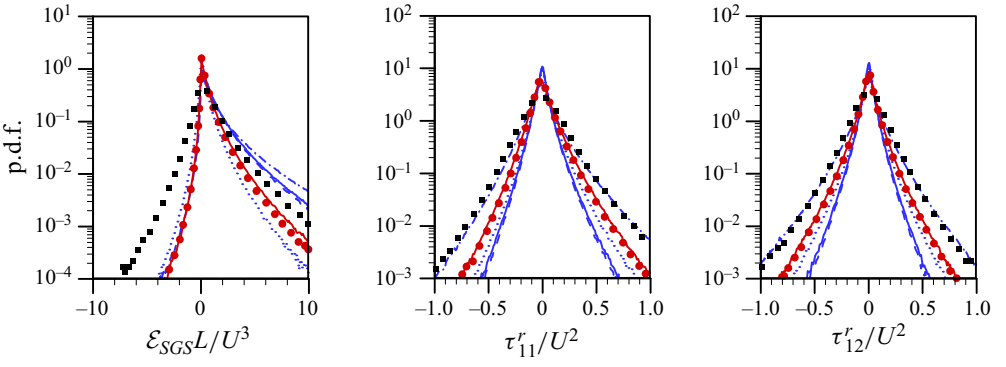


Figure 7. Probability density functions of the SGS dissipation and stresses (*a priori* test at $Re_L = 375.69$ with $\bar{\Delta}/\eta = 16.76$): ■, fDNS32/256; red ○, NNVGM(128D+256D); red line, NNVGM(128D+256L); blue line, CSM32/256; blue dashed line, DSM32/256; blue dotted line, GM32/256; blue dash-dotted line, DMM32/256.

similar to those of GM and DMM, and are much better than those of CSM and DSM, whereas the SGS dissipation is best predicted by NNVGMs (table 6). For p.d.f.s (figure 7), similar predictions as discussed in figure 4 are observed, and thus we do not repeat here. A notable observation is that the results of NNVGM(128D+256L) are nearly identical to those obtained from NNVGM(128D+256D), implying that fLES data can replace fDNS data in constructing NNVGMs. Currently, we reach this conclusion only from numerical simulation, but a more mathematical analysis may be required to confirm the usefulness of fLES data in constructing the NN.

Figures 8 and 9 show the results of energy spectra and p.d.f.s from *a posteriori* tests, where the result from GM32/256 is not shown because the simulation diverged. Again, NNVGM(128D+256D) and NNVGM(128D+256L) perform well among the SGS models considered, but show underpredictions at intermediate wavenumbers. The DMM predicts p.d.f.s very well, but its energy spectrum rapidly decreases at high wavenumbers. Thus, we perform additional LES with NNVGM(128D+1024L) whose trained filter sizes are $4.19 \leq \bar{\Delta}_{fDNS}/\eta$ and $\bar{\Delta}_{fLES}/\eta \leq 67.02$ (note that the filter sizes for NNVGM(128D+256L) are $4.19 \leq \bar{\Delta}_{fDNS}/\eta$ and $\bar{\Delta}_{fLES}/\eta \leq 16.76$). The energy spectrum with NNVGM(128D+1024L) is in excellent agreement with fDNS32/256, suggesting that one should expect successful LES when $\bar{\Delta}_{LES}/\eta$ is within the range of trained filter sizes (see § 4.2 for further discussion). It is also important to note that the present NNVGMs

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

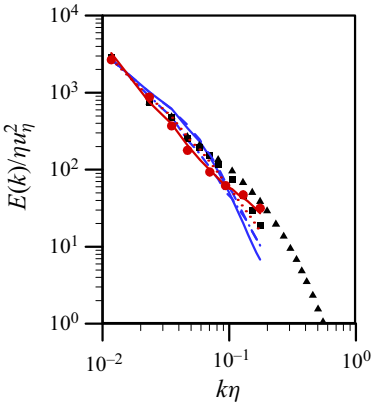


Figure 8. Three-dimensional energy spectra (*a posteriori* test at $Re_L = 375.69$ with $\bar{\Delta}_{LES}/\eta = 16.76$; $N^3 = 32^3$): ▲, DNS256; ■, fDNS32/256; red ○, NNVGM(128D+256D); red line, NNVGM(128D+256L); red dotted line, NNVGM(128D+1024L); blue line, CSM32/256; blue dashed line, DSM32/256; blue dash-dotted line, DMM32/256.

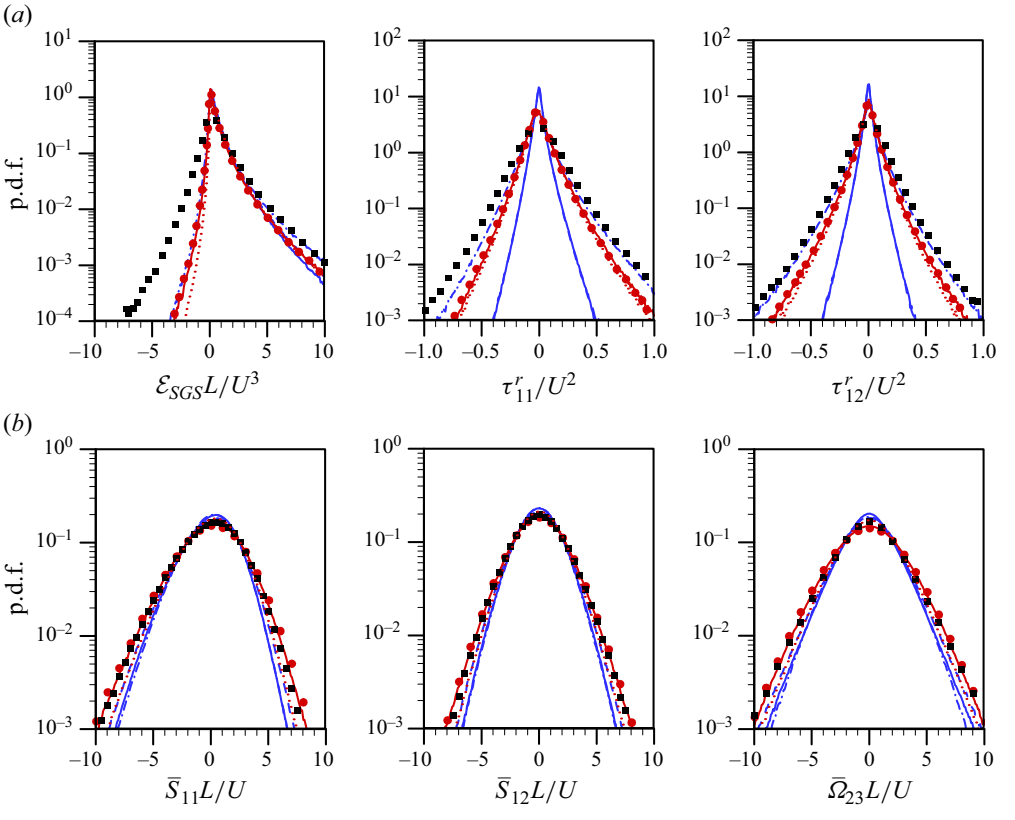


Figure 9. Probability density functions (*a posteriori* test at $Re_L = 375.69$ with $\bar{\Delta}_{LES}/\eta = 16.76$; $N^3 = 32^3$): (*a*) SGS dissipation and stresses; (*b*) strain and rotation rates. ■, fDNS32/256; red ○, NNVGM(128D+256D); red line, NNVGM(128D+256L); red dotted line, NNVGM(128D+1024L); blue line, CSM32/256; blue dashed line, DSM32/256; blue dash-dotted line, DMM32/256.

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

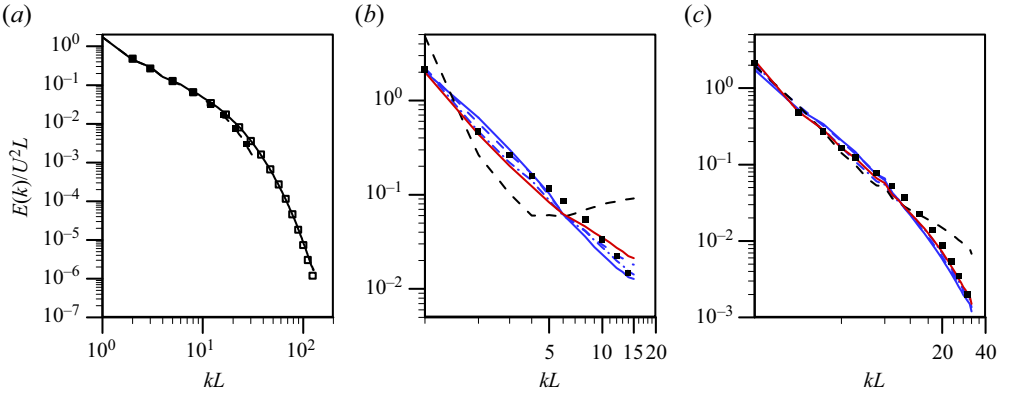


Figure 10. Three-dimensional energy spectra (*a posteriori* test at $Re_L = 375.69$ using CD2 and box filter): (*a*) DNS256 (black line, spectral method; □, CD2) and fDNS64/256 (dashed line, spectral method and cut-Gaussian filter; ■, CD2 and box filter); (*b*) *a posteriori* test with $N^3 = 32^3$; (*c*) *a posteriori* test with $N^3 = 64^3$. In panels (*b*) and (*c*), ■, fDNS; red line, NNVGM(128D+32765L); blue line, CSM; blue dashed line, DSM; blue dash-dotted line, DMM; dashed line, no SGS model (diverge with $N^3 = 32^3$).

do not show the inconsistency between *a priori* and *a posteriori* tests observed from the traditional SGS models (Park *et al.* 2005).

3.4. *A posteriori test using a finite difference method with a box filter*

In this subsection, we apply a trained NNVGM (from a spectral method with the cut-Gaussian filter) to FHIT using the second-order central difference method (CD2) with a box filter, to see if the NNVGM trained from one numerical method can be successfully applied to LES using a different numerical method. All the spatial derivative terms in the filtered continuity and Navier–Stokes equations (2.3) and (2.4) are discretized using CD2 in staggered grids, and the filtered flow variable is obtained using a box filter in the physical space as

$$\bar{\phi}(\boldsymbol{x}, t) = \int_{-\infty}^{\infty} \bar{G}(\boldsymbol{r})\phi(\boldsymbol{x} - \boldsymbol{r}, t)\,\mathrm{d}\boldsymbol{r}, \tag{3.10}$$

$$\bar{G}(\boldsymbol{r}) = \begin{cases} 1/\bar{\Delta}^3 & \text{if } |\boldsymbol{r}| \le \bar{\Delta}/2 \\ 0 & \text{if } |\boldsymbol{r}| > \bar{\Delta}/2 \end{cases}. \tag{3.11}$$

We perform DNS of FHIT at $Re_L = 375.69$ with $N^3 = 256^3$ and LESs with $N^3 = 32^3$ and 64^3 using NNVGM(128D+32768L) and traditional SGS models, respectively. The results are given in figure 10. The energy spectra from DNS and fDNS using the CD2 and box filter are in excellent agreements with those using the spectral method and cut-Gaussian filter (figure 10*a*). The predictions of the energy spectrum by SGS models with $N^3 = 32^3$ are quite similar among themselves but show some deviations from the fDNS data (figure 10*b*). On the other hand, with $N^3 = 64^3$, the predictions are much better than those with $N^3 = 32^3$, and reasonably agree with the fDNS data (figure 10*c*). Hence, the prediction using the NNVGM trained from the spectral method is quite accurate, even though LES is performed with the CD2 and box filter, which may indicate the potential of the present recursive procedure to flow over/inside a complex geometry.

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

4. Applications to FHIT at high Reynolds numbers and DHIT

To assess the applicability of the present NNVGMs (table 5), we conduct LESs of FHIT at high Reynolds numbers and of DHIT. During simulation, a clipping is applied to the model coefficients of DSM and DMM when they are negative (see (2.9) and (2.16)), whereas NNVGMs do not use any clipping.

4.1. *Forced homogeneous isotropic turbulence at high Reynolds numbers*

LESs of FHIT with the number of grid points of 64^3 (LES64/256, LES64/512, LES64/1024, LES64/2048, LES64/4096, LES64/8192, LES64/16384 and LES64/32768) are conducted at eight different Reynolds numbers ($Re_L = 375.69$–242347.33) using nine different SGS models (NNVGM128D, NNVGM(128D+512L), NNVGM(128D+2048L), NNVGM(128D+8192L), NNVGM(128D+32768L), CSM, DSM, GM and DMM). The computational time step is set at $\Delta tU/L = 0.0025$, at which the Courant–Friedrichs–Lewy (CFL) numbers are below 0.3.

Figure 11 shows the three-dimensional energy spectra at $Re_L = 946.67$, 6010.98, 38167.31 and 242347.33 using NNVGM128D, NNVGM(128D+512L), NNVGM(128D+2048L), NNVGM(128D+8192L) and NNVGM(128D+32768L), respectively. The energy spectra predicted from NNVGMs trained with low-Reynolds-number data such as NNVGM128D and NNVGM(128D+512L) show energy pile-up at high wavenumbers for high-Reynolds-number simulations. This energy pile-up at high wavenumbers decreases significantly by NNVGMs trained with higher-Reynolds-number data. The results from NNVGM128D indicate an inherent limitation of the non-recursive NN-based SGS model, when the LES grid size is bigger than the filter sizes of fDNS employed for training the NN. It is interesting to note that the energy spectrum for the case of $Re_L = 242347.33$ can be reasonably predicted by NNVGM(128D+2048L) and NNVGM(128D+8192L) that includes the training data up to $Re_L = 6010.98$ and 38167.31, respectively. This result indicates that, when the training data contain the length scales in the inertial range large enough compared with the dissipation scale, the present SGS model should predict isotropic turbulence even at the Reynolds number higher than that trained.

Figure 12 shows the three-dimensional energy spectra at eight Reynolds numbers from various SGS models including traditional ones. First, LESs with GM reveal instability at high Reynolds numbers (see also Vollant, Balarac & Corre 2016). LESs without SGS model exhibit non-physical energy pile up at high wavenumbers (figure 12*f*) due to insufficient dissipation, as observed from LESs with NNVGM128D and NNVGM(128D+512L) in figure 11. NNVGM(128D+32768L) suitably calculates the energy spectra that are quite similar to those of CSM, DSM and DMM. The predicted energy spectra from NNVGM, CSM, DSM and DMM follow the Kolmogorov energy spectrum at low and intermediate wavenumbers.

In §§ 3.2 and 3.3, we observed that GM and NNVGM performed similarly in *a priori* tests for small filter sizes, whereas, in actual LES, GM diverged but NNVGM performed stably for large grid sizes. One of the reasons for these different results is that the two models compute the SGS dissipation differently. As shown in table 6, for a large filter size, GM predicts the SGS dissipation significantly smaller than fDNS data, while NNVGM computes it accurately. Another reason may be that the error of the SGS stress prediction by GM grows exponentially with increasing grid size (Vreman *et al.* 1996), since GM is derived from Taylor expansion. However, NNVGM is trained with the SGS stresses from small to large filter sizes, and thus the error of the SGS stress prediction by NNVGM does not increase with increasing grid size.

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

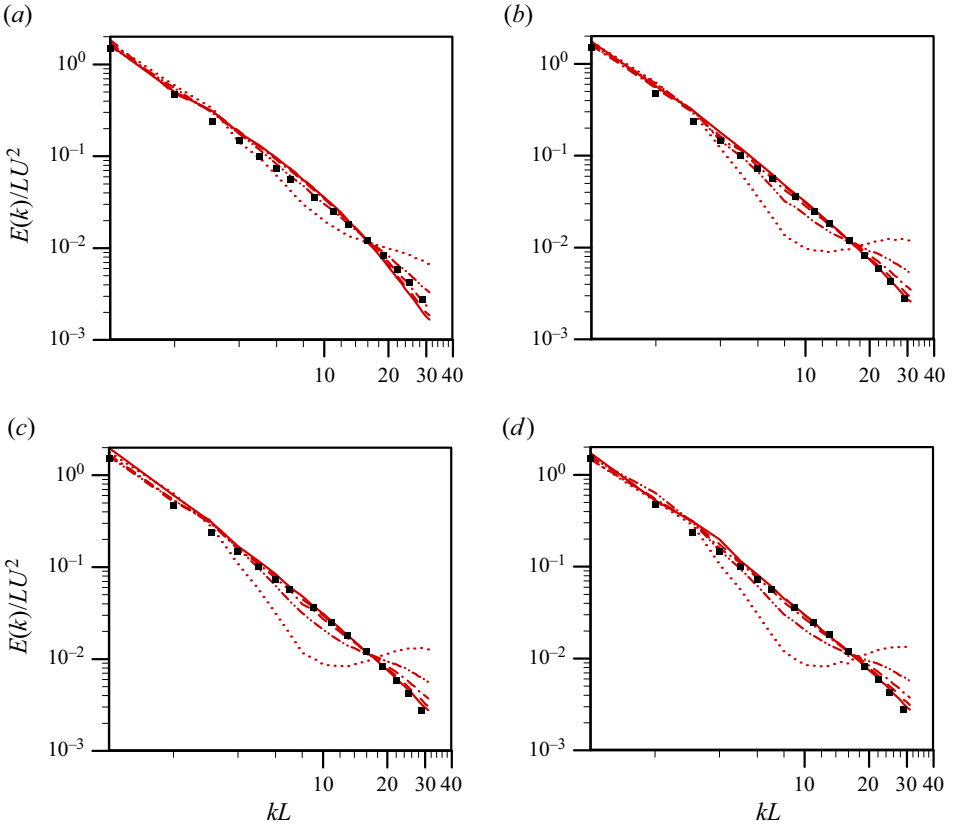


Figure 11. Three-dimensional energy spectra from various NNVGMs ($N^3 = 64^3$): (*a*) $Re_L = 946.67$; (*b*) 6010.98; (*c*) 38167.31; (*d*) 242347.33. ■, filtered Kolmogorov energy spectrum, $E(k) = \frac{3}{2}\epsilon^{2/3}k^{-5/3}\exp(-k^2\bar{\Delta}^2/12)$; red dotted line, NNVGM128D; red dash-dot-dotted line, NNVGM(128D+512L); red dash-dotted line, NNVGM(128D+2048L); red dashed line, NNVGM(128D+8192L); red line, NNVGM(128D+32768L).

4.2. *Decaying homogeneous isotropic turbulence*

The present approach is based on a hypothesis that the NN-based SGS models in LES successfully predict the turbulence statistics when the LES grid size normalized by the Kolmogorov length scale is within the range of filter sizes used for training (see also Park & Choi 2021). In figure 13, we indicate the filter sizes (normalized by the Kolmogorov length scale) used for fDNSs and fLESs of FHIT with NNVGMs and the normalized grid sizes used for LESs of two DHITs. This figure should be understood as follows: for DHIT of Comte-Bellot & Corrsin (1971), NNVGM(128D+1024L) or NNVGM(128D+32768L) can be used for successful LES, but not NNVGM128D and NNVGM(128D+256L); for DHIT of Kang, Chester & Meneveau (2003), only NNVGM(128D+32768L) can be used for successful LES. To examine this hypothesis, LESs are performed for two DHITs (untrained flows) of Comte-Bellot & Corrsin (1971) and Kang *et al.* (2003) with the present NNVGMs trained only with FHIT data. This task should be a notable challenge due to the transient nature of DHIT as opposed to FHIT.

The first DHIT to simulate is the experiment by Comte-Bellot & Corrsin (1971) (called CBC hereafter). In CBC, turbulence data were experimentally generated through grid turbulence using a mesh size of $M = 5.08$ cm and free-stream velocity of $U_0 =$

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

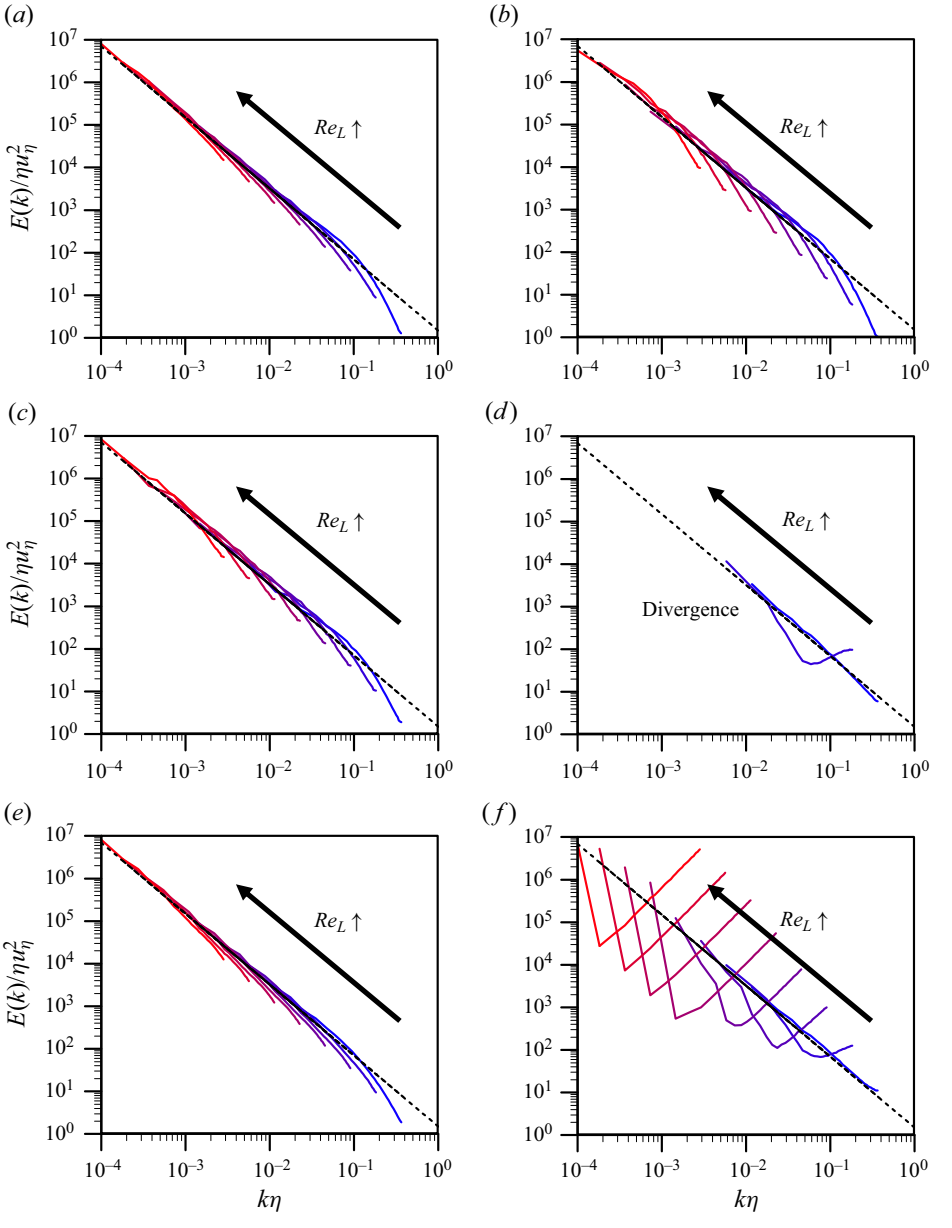


Figure 12. Three-dimensional energy spectra at $Re_L = 375.69$–242347.33 from various SGS models ($N^3 = 64^3$): (*a*) NNVGM(128D+32768L); (*b*) CSM; (*c*) DSM; (*d*) GM; (*e*) DMM; (*f*) no SGS model. The black dashed line denotes the Kolmogorov energy spectrum, $E(k) = \frac{3}{2}\epsilon^{2/3}k^{-5/3}$.

10 m s^{-1}. The Reynolds number based on the Taylor microscale was $Re_\lambda = 71.6$ at $tU_0/M = 42$. For LES, flow variables are non-dimensionalized by $L = 11M/(2\pi)$ and $U = \sqrt{3/2}u_{rms}|_{tU_0/M=42}$ (Lee, Choi & Park 2010). The numbers of grid points tested are $N^3 = 32^3$ and 64^3, respectively. A divergence-free initial field is obtained using the logistic polynomial approximation method (Knight *et al.* 1998) and rescaling method (Kang *et al.* 2003). LESs are performed 50 times by varying initial fields to obtain ensemble averages.

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

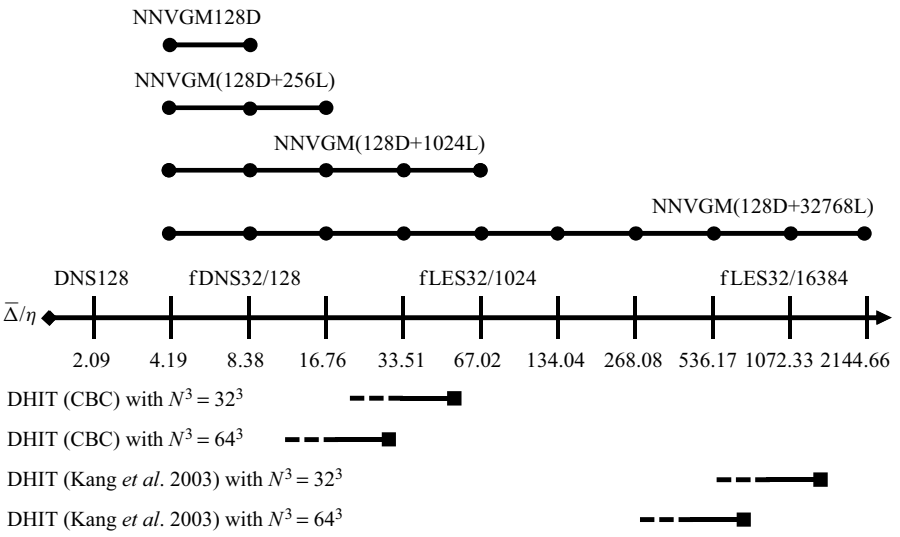


Figure 13. Illustration of the filter-size ranges of fDNS and fLES used to generate each NNVGM (FHIT), and the grid-size ranges required for the simulations of DHITs (Comte-Bellot & Corrsin 1971; Kang *et al.* 2003). On the top of this figure (FHIT), each solid circle (●) denotes the filter size of fDNS or fLES used for generating training data. For each NNVGM, the line connecting the first and last solid circles indicates the range of grid sizes for successful LES. On the bottom of this figure (DHIT), each solid square (■) indicates the grid size normalized by the initial Kolmogorov length scale. As time goes by, the Kolmogorov length scale increases and thus $\bar{\Delta}_{LES}/\eta$ decreases denoted as a thick dashed line.

The computational time step is fixed to be $\Delta tU/L = 0.01$. As turbulence decays in time, the Kolmogorov length scale η increases in time, and thus $\bar{\Delta}_{LES}/\eta = 60.2$ and 26.5 ($N^3 = 32^3$), and 30.1 and 13.2 ($N^3 = 64^3$) at $tU_0/M = 42$ and 171, respectively. Each LES starts from large $\bar{\Delta}_{LES}/\eta$ (denoted as a solid square in figure 13) and moves to smaller $\bar{\Delta}_{LES}/\eta$ due to increased η in time (denoted as a thick dashed line).

The LES results on CBC DHIT with $N^3 = 32^3$ (starting from $\bar{\Delta}_{LES}/\eta = 60.2$) are given in figure 14. Without SGS model, the turbulent kinetic energy decays very slowly and energy pile up occurs at high wavenumbers. With GM, resolved kinetic energy first decreases and increases later in time, and simulation finally diverges, as shown by Vreman *et al.* (1996). DSM and CSM are relatively good at predicting turbulence decay and energy spectra, but they overestimate the energy spectra at intermediate wavenumbers within the inertial range. However, NNVGM(128D+32768L) performs best in predicting energy spectra, while NNVGM(128D+1024L) and DMM also perform better than other types of SGS models (DMM shows energy fall-off at high wavenumbers), whereas NNVGM(128D+1024L) shows energy pile up at high wavenumbers, possibly because $\bar{\Delta}_{LES}/\eta$ (≤ 60.2) is only slightly smaller than $\bar{\Delta}_{fLES}/\eta = 67.02$ (fLES32/1024). Note also that the prediction with NNVGM128D ($\bar{\Delta}_{LES}/\eta \gg \bar{\Delta}_{fDNS}/\eta = 8.38$) is not good at all and worse than CSM and DSM, validating our conjecture on the choice of NNVGM for successful LES (figure 13). With $N^3 = 64^3$ (starting from $\bar{\Delta}_{LES}/\eta = 30.1$), NNVGM(128D+1024L) also performs very good as much as NNVGM(128D+32768L) does (figure 15), because $\bar{\Delta}_{LES}/\eta$ (≤ 30.1) is sufficiently smaller than $\bar{\Delta}_{fLES}/\eta = 67.02$, while other SGS models (except DMM) still show some deviations from experimental data.

The second DHIT to simulate is the experimental one by Kang *et al.* (2003). The Reynolds number is much higher ($Re_\lambda = 716$ at $tU_0/M = 20$) than that of CBC DHIT,

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

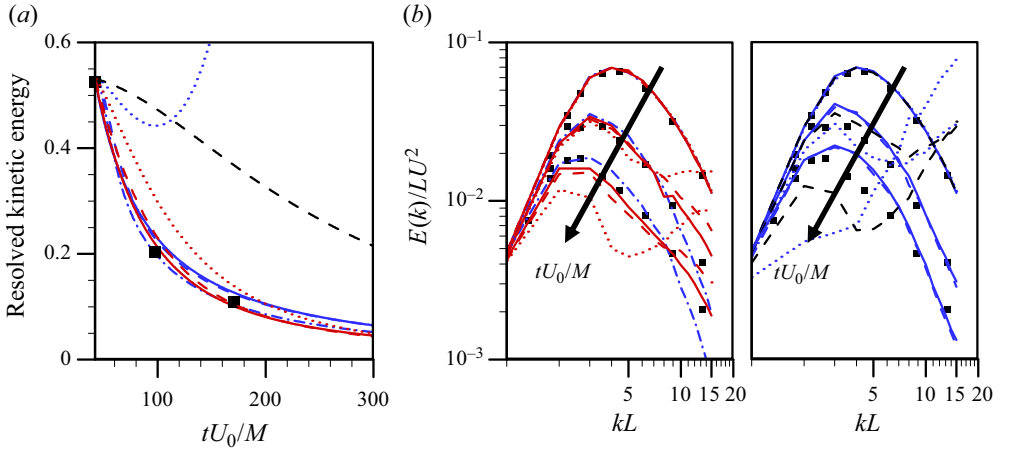


Figure 14. LES of CBC DHIT with $N^3 = 32^3$: (*a*) resolved turbulent kinetic energy; (*b*) $E(k)$ at $tU_0/M = 42$, 98 and 171. ■, filtered CBC data; red dotted line, NNVGM128D; red dashed line, NNVGM(128D+1024L); red line, NNVGM(128D+32768L); blue line, CSM; blue dashed line, DSM; blue dotted line, GM; blue dash-dotted line, DMM; dashed line, no SGS model.

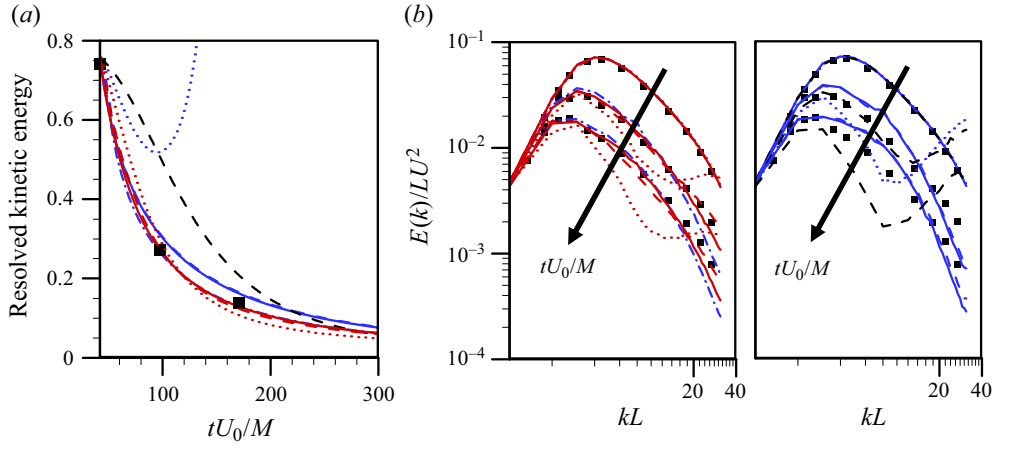


Figure 15. LES of CBC DHIT with $N^3 = 64^3$: (*a*) resolved turbulent kinetic energy; (*b*) $E(k)$ at $tU_0/M = 42$, 98 and 171. ■, filtered CBC data; red dotted line, NNVGM128D; red dashed line, NNVGM(128D+1024L); red line, NNVGM(128D+32768L); blue line, CSM; blue dashed line, DSM; blue dotted line, GM; blue dash-dotted line, DMM; dashed line, no SGS model.

where $U_0 = 11.2$ m s^{-1} and $M = 0.152$ m. We consider the present NNVGMs and traditional SGS models. LESs are performed 50 times by varying initial fields to obtain ensemble averages. The flow variables are non-dimensionalized by $L = 33.68M/(2\pi)$ and $U = \sqrt{3/2}u_{rms}|_{tU_0/M=20}$. The numbers of grid points tested are $N^3 = 32^3$ and 64^3. The flows are initiated using the energy spectrum and rescaling method as done by Kang *et al.* (2003). For $N^3 = 32^3$, $\bar{\Delta}_{LES}/\eta = 1454.5$ and 888.9 at $tU_0/M = 20$ and 48, respectively. The trained grid sizes of NNVGM(128D+32768L) ($4.19 \leq \bar{\Delta}_{fDNS}/\eta$ and $\bar{\Delta}_{fLES}/\eta \leq 2144.66$) include this range of $\bar{\Delta}_{LES}/\eta$, whereas those of NNVGM128D ($4.19 \leq \bar{\Delta}_{fDNS}/\eta \leq 8.38$) do not (see figure 13), suggesting that NNVGM(128D+32768L)

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

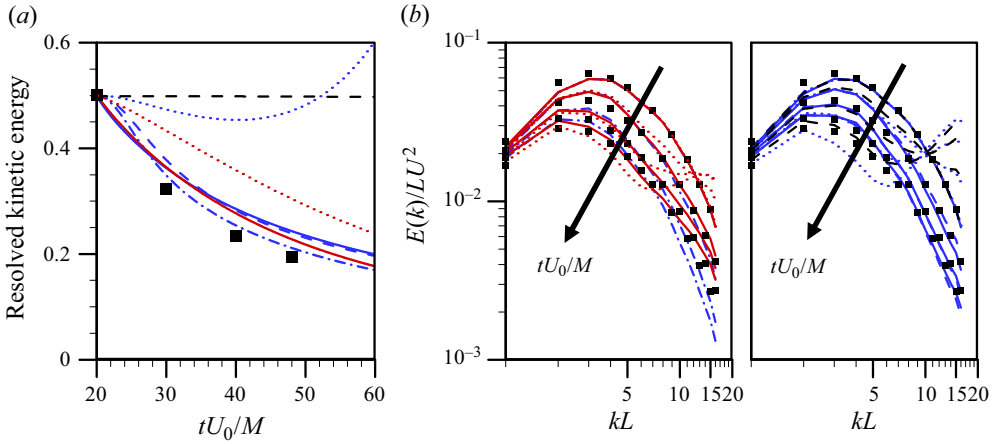


Figure 16. LES of DHIT by Kang *et al.* (2003) with $N^3 = 32^3$: (*a*) resolved turbulent kinetic energy; (*b*) $E(k)$ at $tU_0/M = 20$, 30, 40 and 48. ■, filtered experimental data; red dotted line, NNVGM128D; red line, NNVGM(128D+32768L); blue line, CSM; blue dashed line, DSM; blue dotted line, GM; blue dash-dotted line, DMM; dashed line, no SGS model.

should properly predict turbulence decay and variation of the energy spectra of DHIT by Kang *et al.* (2003).

The LES results on DHIT by Kang *et al.* (2003) with $N^3 = 32^3$ are shown in figure 16. While CSM and DSM overpredict the energy spectra even at intermediate wavenumbers and thus overestimate turbulent kinetic energy, DMM predicts the resolved kinetic energy and energy spectra most accurately. NNVGM(128D+32768L) slightly overestimates the resolved turbulent kinetic energy due to its slight overprediction of energy at high wavenumbers. This may be because $\bar{\Delta}_{fLES}/\eta = 2144.66$ (fLES32/32768) is only slightly larger than $\bar{\Delta}_{LES}/\eta$ (≤ 1454.5). By increasing the number of grid points ($N^3 = 64^3$), the predictions by NNVGM(128D+32768L) become much better (figure 17). It is interesting to see that the level of kinetic energy without SGS model is almost constant in time due to the energy pile up at high wavenumbers. This is because the LES grid size ($\bar{\Delta}_{LES}/\eta \approx 10^3$) is so large that the corresponding eddies do not properly dissipate energy at these length scales, and thus total kinetic energy is nearly conserved without SGS model.

Finally, the effects of the present NNVGMs on the prediction of the resolved kinetic energy and spectrum of DHIT of Kang *et al.* (2003) are shown in figure 18. As shown before, NNVGM128D does not provide an accurate energy spectrum, because it is trained with the data from $4.19 \leq \bar{\Delta}_{fDNS}/\eta \leq 8.38$ and does not have sufficient information on $8.38 < \bar{\Delta}_{LES}/\eta \leq 727.3$ (see figure 13 for the grid size range suggested for LES of DHIT of Kang *et al.* 2003). As the NNVGM is recursively updated (from NNVGM128D to NNVGM(128D+32768L)), its prediction becomes increasingly accurate. NNVGM(128D+2048L) provides quite an accurate energy spectrum albeit showing energy pile up at high wavenumbers. This is because NNVGM(128D+2048L) is trained with the data from $4.19 \leq \bar{\Delta}_{fDNS}/\eta$ and $\bar{\Delta}_{fLES}/\eta \leq 134.04$ and contains large length scales in the inertial range. As expected, NNVGM(128D+32768L) provides accurate energy spectra at all wavenumbers.

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

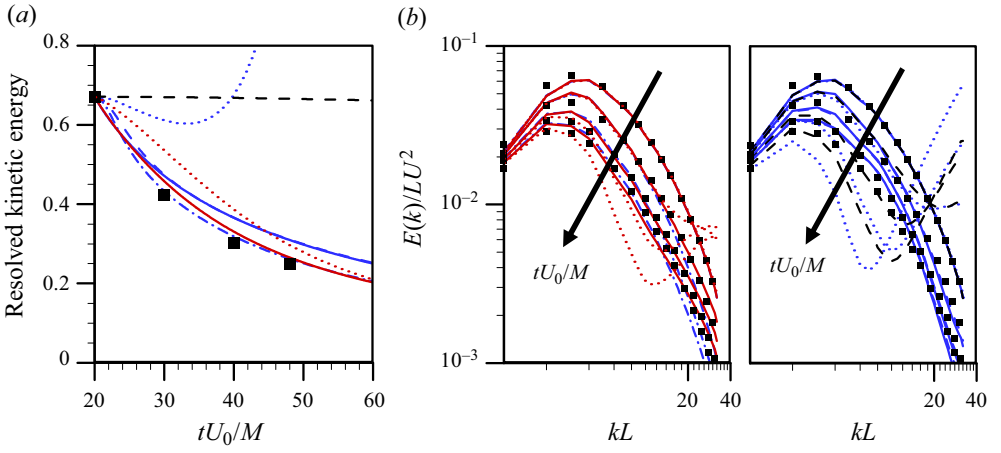


Figure 17. LES of DHIT (Kang *et al.* 2003) with $N^3 = 64^3$: (*a*) resolved turbulent kinetic energy; (*b*) $E(k)$ at $tU_0/M = 20$, 30, 40 and 48. ■, filtered experimental data; red dotted line, NNVGM128D; red line, NNVGM(128D+32768L); blue line, CSM; blue dashed line, DSM; blue dotted line, GM; blue dash-dotted line, DMM; dashed line, no SGS model.

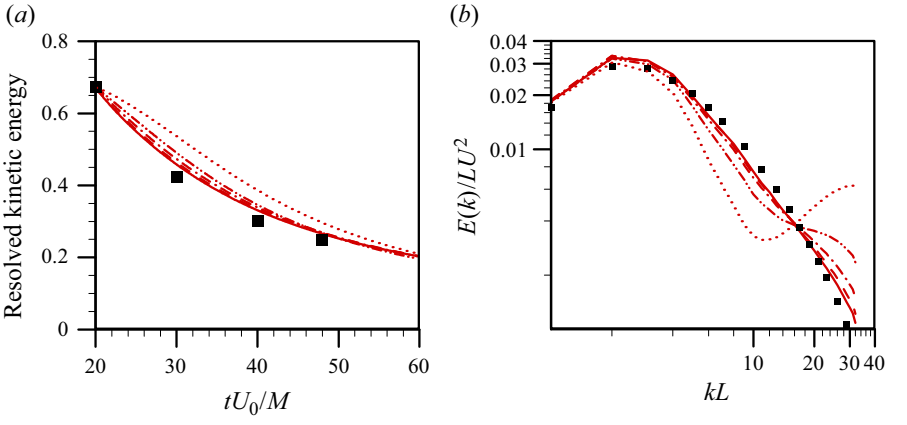


Figure 18. Effects of NNVGMs on the prediction of DHIT (Kang *et al.* 2003) with $N^3 = 64^3$: (*a*) resolved turbulent kinetic energy; (*b*) $E(k)$ at $tU_0/M = 48$. ■, filtered experimental data; red dotted line, NNVGM128D; red dash-dot-dotted line, NNVGM(128D+512L); red dash-dotted line, NNVGM(128D+2048L); red dashed line, NNVGM(128D+8192L); red line, NNVGM(128D+32768L).

4.3. *Computational cost*

The amounts of CPU time required for estimating the SGS stresses and advancing one computational time step for the traditional and NN-based SGS models are given in table 7. With the same number of grids ($N^3 = 64^3$), the amounts of CPU time required to obtain the SGS stresses by DSM, DMM and NNVGM are approximately 3.3, 5.1 and 2.8 times that by CSM, respectively. Therefore, the computational cost of NNVGM is higher than CSM but lower than those of DSM and DMM, which is consistent with the report of Kang, Jeon & You (2023) studying a similar HIT. Note, however, that the total amounts of CPU time for advancing one computational time step are relatively similar among themselves. The reason that the CPU time of DSM or DMM for determining the SGS stresses is higher than that of NNVGM is because DSM and DMM require Fourier and inverse Fourier

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

	SGS model	FHIT ($Re_L = 149.09$)	DHIT (CBC)	DHIT (Kang *et al.* 2003)
Number of grid points	—	64^3	64^3	64^3
SGS stresses	CSM	0.042	0.043	0.042
	DSM	0.138	0.140	0.140
	DMM	0.217	0.214	0.214
	NNVGM (128D+32768L)	0.123	0.116	0.115
One computational time step	CSM	0.386	0.393	0.381
	DSM	0.483	0.479	0.484
	DMM	0.560	0.560	0.553
	NNVGM (128D+32768L)	0.474	0.455	0.453

Table 7. Amounts of CPU time (second) required for estimating the SGS stresses and advancing one computational time step, respectively. Simulations are performed using eight CPU cores (Intel(R) Core(TM) i9-11900KF CPU @ 3.50 GHz), and the amounts of CPU time are obtained by averaging over 50 computational time steps. In the present simulation, τ_{ij} was obtained using one GPU (GeForce RTX 3060) for NNVGM(128D+32768L), but, for comparison with traditional models, the amounts of CPU time using eight CPU cores are provided in this table.

transforms to determine the dynamic coefficient(s). However, for flow over a complex geometry, such transforms are not required based on a finite difference method. In that case, an NN-based SGS model may require more CPU time than DSM or DMM (see Kim *et al.* 2024 as an example).

5. Conclusions

In the present study, we developed an NN-based SGS model (NN-based velocity gradient model; NNVGM) using a dual NN architecture (the output of one NN is the SGS normal stresses and that of the other is the SGS shear stresses) with the input of $\bar{\Delta}^2|\bar{\alpha}|\bar{\alpha}_{ij}$, where α_{ij} is the velocity gradient tensor. We aimed at overcoming the difficulties encountered during the SGS model development, i.e. how to obtain high-Reynolds-number data for training and how to apply a trained NN to untrained flow. By eliminating bias and batch normalization, and employing the leaky ReLU function as an activation function within the NNs, the present NN retains its nonlinearity but satisfies $\mathbf{NN}(c\boldsymbol{A}) = c\mathbf{NN}(\boldsymbol{A})$ for a positive scalar c. This important property of NN allows us to apply the NN to a flow different from the trained one, because the parameters used for non-dimensionalization can be included in the scalar c. To generate training data, we adopted a cut-Gaussian filter rather than the Gaussian or spectral cutoff filter, considering its application to LES with coarser grids and realizability condition. We also designed a recursive procedure which consisted of the following steps: (1) conducting DNS; (2) training an NN with fDNS data; (3) conducting LES at a higher Reynolds number; (4) training the NN with augmented data including fLES data; (5) going to Step (3) for higher Reynolds number flows. For the present FHIT, the grid and filter sizes were normalized by the Kolmogorov length scale, and these normalized sizes became double at every recursive procedure. The NN trained through this recursive procedure contained a wide range of filter sizes normalized by the Kolmogorov length scale such that the LES grid size required could be located within this range of filter sizes. Testing NNVGMs on FHIT showed that fLES data can be a practical

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

alternative to fDNS data for training an NN, and one can avoid using costly DNS to extract training data.

We conducted LESs of forced and decaying homogeneous isotropic turbulence (FHIT and DHIT) with NNVGMs and traditional SGS models, respectively. In FHIT, the same number of grid points was used for different Reynolds numbers. Among the SGS models considered, NNVGM constructed through the recursive procedure performed very well for FHIT. In contrast, NNVGM trained only with fDNS data at a low Reynolds number showed energy pile up at high wavenumbers. We considered two different DHIT, one by Comte-Bellot & Corrsin (1971) and the other at much higher Reynolds number by Kang *et al.* (2003). In both DHIT, the recursive NNVGM predicted the decay of resolved turbulent kinetic energy and its energy spectra in time most accurately among the SGS models considered, while most other SGS models provided excessive energy spectra at intermediate or high wavenumbers.

In the present study, we applied the NNVGM to FHIT and DHIT, but this procedure may not be applicable to laminar and inhomogeneous turbulent flows. Hence, the next step is to extend the present approach to those flows by employing the dynamic approach used in DSM or training NN with more flows having different topology. This work is being carried out in our group (Cho & Choi 2023; Kim & Choi 2023).

Supplementary material. Computational Notebook files are available as supplementary material at https://doi.org/10.1017/jfm.2024.992 and online at https://www.cambridge.org/S0022112024009923/JFM-Notebooks.

Funding. This work is supported by the National Research Foundation through the Ministry of Science and ICT (nos. 2022R1A2B5B0200158612 and 2021R1A4A1032023). The computing resources are provided by the KISTI Super Computing Center (no. KSC-2023-CRE-0197).

Declaration of interests. The authors report no conflict of interest.

Author ORCIDs.
Chonghyuk Cho https://orcid.org/0009-0009-0882-565X;
Jonghwan Park https://orcid.org/0009-0005-7786-926X;
Haecheon Choi https://orcid.org/0000-0003-0696-847X.

Appendix A. Effect of the bias, batch normalization and activation function on numerical solutions

The effects of excluding the bias and batch normalization and modifying the activation function (from ReLU to leaky ReLU) on the results of *a priori* and *a posteriori* tests are examined. Table 8 shows four other cases together with the present case for this purpose. The five cases share same NN structure, training data (extracted from fDNS32/128) and undersampling method. However, the training data are normalized with U^2 instead of $\tau^r_{ij,rms}$, because the cases of ReLUOO, leReLUOO and leReLUOX do not satisfy (3.1) and $\tau^r_{ij,rms}$ is *a priori* unknown in the *a posteriori* test (see § 3.2 for the details).

We perform *a priori* and *a posteriori* tests for five cases and compare the results with fDNS32/128 data. Table 9 shows the statistics from *a priori* tests at $Re_L = 149.09$. Except for the case of XXX, the four other cases provide very similar statistics and overpredict the SGS dissipation. Note that the case of leReLUXX is same as that of NNVGM32/128, and its correlation coefficients are similar to those of NNVGM(32+64)/128 but its predicted SGS dissipation is less accurate than that of NNVGM(32+64)/128 ($\epsilon_{SGS} = 0.447$; see table 4), indicating that combining more databases is indeed more beneficial. Figures 19 and 20 show the results (p.d.f. and energy spectrum) from both *a priori* and *a posteriori* tests for the five cases. As shown, four cases except XXX provide nearly the same

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

Case	Activation function	Bias ($\boldsymbol{b}$)	Batch normalization ($\boldsymbol{\gamma}$, $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, $\boldsymbol{\beta}$)
ReLUOO	ReLU	O	O
leReLUOO	leaky ReLU	O	O
leReLUOX	leaky ReLU	O	X
leReLUXX (present)	leaky ReLU	X	X
XXX	X	X	X

Table 8. Cases with different activation functions and with/without the bias and batch normalization. In this table, the case of leReLUXX is the same as NNVGM32/128, ReLU and leaky ReLU correspond to $f(x) = \max[0, x]$ and $\max[0.02x, x]$, respectively, and O and X indicate with and without using corresponding function (or parameters), respectively.

Case	$R(\tau^r_{11})$	$R(\tau^r_{12})$	$R(\epsilon_{SGS})$	ϵ_{SGS}
fDNS32/128	—	—	—	0.430
ReLUOO	0.654	0.686	0.634	0.602
leReLUOO	0.655	0.681	0.630	0.609
leReLUOX	0.654	0.687	0.637	0.616
leReLUXX	0.661	0.691	0.642	0.536
XXX	0.197	0.210	0.586	0.642

Table 9. Statistics from *a priori* tests at $Re_L = 149.09$ with $\bar{\Delta}_{fDNS}/\eta = 8.38$: Pearson correlation coefficients (R) of τ^r_{ij} and SGS dissipation $\epsilon_{SGS}(= -\tau^r_{ij}\bar{S}_{ij})$, and magnitude of ϵ_{SGS}.

results. These results indicate that the use of bias and batch normalization is unnecessary to accurately predict the current flow, but the introduction of ReLU or leaky ReLU is necessary. However, as discussed in § 3.2, the use of bias and batch normalization in the NN structure prevents the trained NN from being applied to untrained flows.

Appendix B. Effect of the filter to grid ratio

In the present study, the filter to grid ratio (FGR) of 1 (i.e. $\bar{\Delta}_{LES} = \bar{\Delta}_{filter}$) is used to develop NNVGMs. However, some previous studies (Ghosal 1996; Chow & Moin 2003; Meyers, Geurts & Baelmans 2003) suggested to use a filter size larger than the LES grid size to decrease the discretization and aliasing errors when LES is performed with a finite difference method. Although the present LESs are conducted with a spectral method, we analyse in this appendix how the performance of NNVGM changes when FGR is 2 (i.e. $2\bar{\Delta}_{LES} = \bar{\Delta}_{filter}$). The recursive procedure introduced in § 3.1 is changed as follows (Steps (i), (ii), (v) and (vii) are the same):

(iii) apply the NNVGM to LES at a higher Reynolds number, where the ratio of LES grid size to the Kolmogorov length scale ($\bar{\Delta}_{LES}/\eta$) is *half* that of filter size to the Kolmogorov length scale ($\bar{\Delta}_{fDNS}/\eta$);

(iv) filter the LES data and include the filtered LES (fLES) data in the training dataset, where the new filter size is *four times* the LES grid size;

(vi) apply the updated NNVGM to LES at even higher Reynolds number, where the LES grid size is *half* the filter size defined in Step (iv).

First, fDNS32/128 and fDNS64/128 are used for training NNVGM(32+64)/128. Now, LES128/256 (instead of LES64/256) is conducted with this NNVGM. The test filter is applied to the results of LES128/256 with $\tilde{\Delta} = 4\bar{\Delta}_{LES}$: the resulting filtered data are called fLES32/128-FGR2. Then, the NN is trained again with the augmented training

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

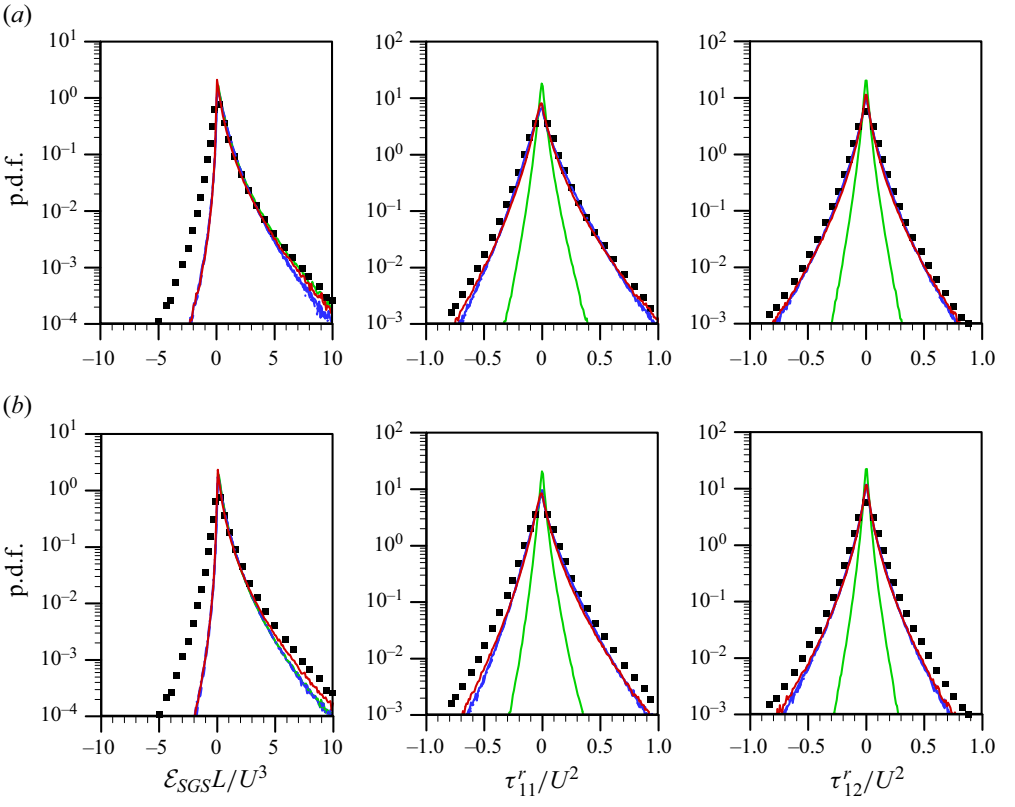


Figure 19. Probability density functions of the SGS dissipation and SGS stresses; (*a*) *a priori* test with fDNS32/128; (*b*) *a posteriori* test with LES32/128. ■, fDNS32/128; blue dash-dotted line, ReLUOO; blue dashed line, leReLUOO; blue line, leReLUOX; red line, leReLUXX; green line, XXX.

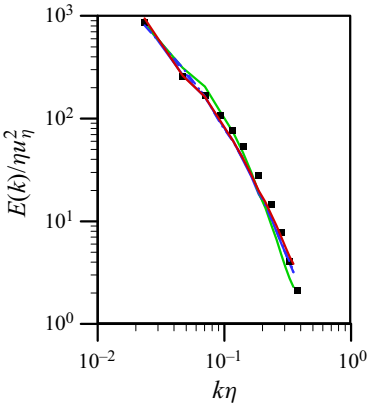


Figure 20. Three-dimensional energy spectra (*a posteriori* test at $Re_L = 149.09$; LES32/128) ■, fDNS32/128; blue dash-dotted line, ReLUOO; blue dashed line, leReLUOO; blue line, leReLUOX; red line, leReLUXX; green line, XXX.

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

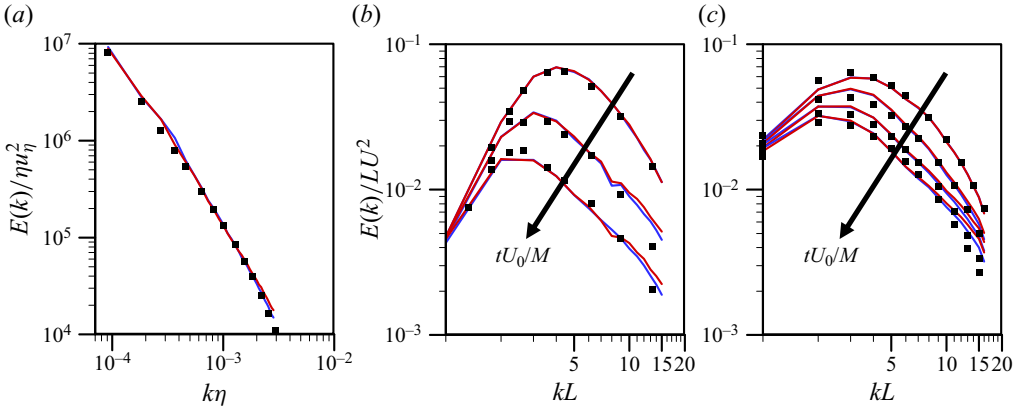


Figure 21. Three-dimensional energy spectra (*a posteriori* test; LES64/32768): (*a*) FHIT ($Re_L = 242347.33$; $N^3 = 64^3$); (*b*) DHIT at $tU_0/M = 42$, 98 and 171 (CBC; $N^3 = 32^3$); (*c*) DHIT at $tU_0/M = 20$, 30, 40 and 48 (Kang *et al.* 2003; $N^3 = 32^3$). ■, filtered Kolmogorov energy spectrum (in panel *a*) or filtered experimental data (in panels *b* and *c*); blue line, FGR = 1; red line, FGR = 2.

data of fDNS64/128, fDNS32/128 and fLES32/128-FGR2. The trained NN is applied to LES128/512. This recursive procedure is continued until the NN is trained with fLES32/32768-FGR2.

Two NNVGMs for FGR = 1 and 2 are applied to LES of FHIT at $Re_L = 242347.33$, and to DHITs of Comte-Bellot & Corrsin (1971) and Kang *et al.* (2003), respectively. As shown in figure 21, the FGRs of 1 and 2 provide very similar results.

Appendix C. Choice of the probability $\mathcal{P}$ for undersampling

For the purpose of undersampling of training data, we introduce the probability $\mathcal{P}$ to the training data such that the possibility of choosing data as the training one decreases when its magnitude is near to zero. We further consider the following probability functions ($\mathcal{P}$):

$$\mathcal{P} = \begin{cases} g(\theta) & \text{if } 0 \le \theta < \pi/2 \\ 1 & \text{if } \theta \ge \pi/2 \end{cases}, \tag{C1}$$

$$\theta = \frac{\pi}{8} \frac{\sqrt{\{(\tau_{12}^r)^2 + (\tau_{23}^r)^2 + (\tau_{13}^r)^2\}/3}}{\tau_{12,rms}^r}. \tag{C2}$$

We consider six different functions for $g(\theta)$ and they are listed in table 10. Here, case Pc does not use undersampling, and four other cases (Ps1, Ps2, Ps3 and Ps4) use sine functions, because sine function rapidly decays to 0 when θ approaches zero and reaches 1 at $\theta = \pi/2$. The last case Pl uses a linear function for $0 \le \theta < \pi/2$. For all the cases, the numbers of original training data are the same (4 900 000), but the numbers of training data resulting from undersampling become significantly different, as shown in table 10.

These six different probability functions are applied to LESs of FHIT at $Re_L = 242347.33$, and of DHITs of Comte-Bellot & Corrsin (1971) and Kang *et al.* (2003), and their results are given in figure 22. It is noticeable that undersampling affects the energy spectra at high wavenumbers. Without undersampling (case Pc), the energy spectrum is overpredicted at high wavenumbers. However, the prediction performance improves when undersampling is applied, while five probability functions provide similar

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

Case	$g(\theta)$	Approximate number of training data (undersampling)
Pc	1	4 900 000
Ps1	$\sin\theta$	1 500 000
Ps2	$\sin^2\theta$	600 000
Ps3	$\sin^3\theta$	300 000
Ps4	$\sin^4\theta$	200 000
Pl	$2\theta/\pi$	1 000 000

Table 10. Probability functions for undersampling and corresponding numbers of each fDNS or fLES training data (i.e. fDNS64/128, fDNS32/128, fLES32/256, fLES32/512, etc.) during the recursive procedure. Here, NNVGM(128D+32768L) is used for the SGS model.

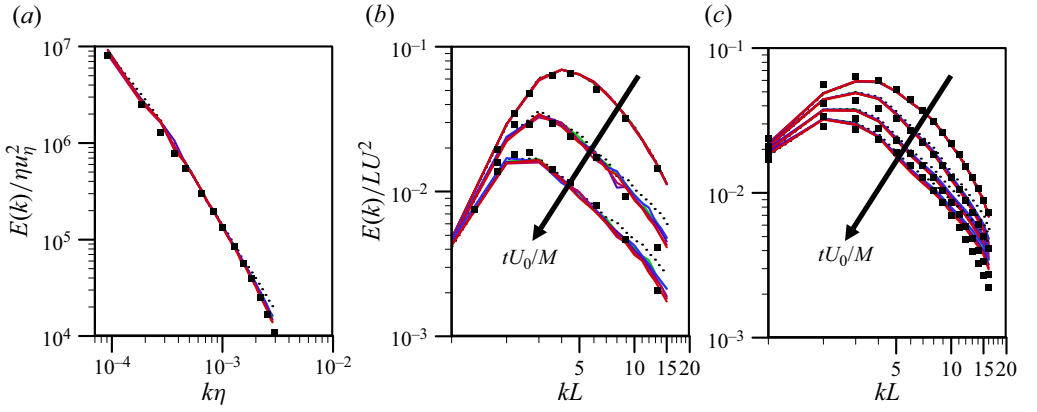


Figure 22. Three-dimensional energy spectra (*a posteriori* test with NNVGM(128D+32768L)): (*a*) FHIT ($Re_L = 242347.33$; $N^3 = 64^3$); (*b*) DHIT at $tU_0/M = 42$, 98 and 171 (CBC; $N^3 = 32^3$); (*c*) DHIT at $tU_0/M =$ 20, 30, 40 and 48 (Kang *et al.* 2003; $N^3 = 32^3$). ■, filtered Kolmogorov energy spectrum (in panel *a*) or filtered experimental data (in panels *b* and *c*); dotted line, Pc; blue line, Ps1; blue purple line, Ps2; red purple line, Ps3; red line, Ps4; green line, Pl.

results, indicating that the probability function itself has little influence on the prediction performance, once it is well designed to provide balanced distribution of training data.

References

Abekawa, A., Minamoto, Y., Osawa, K., Shimamoto, H. & Tanahashi, M. 2023 Exploration of robust machine learning strategy for subgrid scale stress modeling. *Phys. Fluids* **35** (1), 015162.

Adrian, R. 1990 Stochastic estimation of sub-grid scale mations. *Appl. Mech. Rev.* **43**, S214–S218.

Adrian, R.J., Jones, B.G., Chung, M.K., Hassan, Y., Nithianandan, C.K. & Tung, A.T.-C. 1989 Approximation of turbulent conditional averages by stochastic estimation. *Phys. Fluids* A **1** (6), 992–998.

Anderson, R. & Meneveau, C. 1999 Effects of the similarity model in finite-difference LES of isotropic turbulence using a lagrangian dynamic mixed model. *Flow Turbul. Combust.* **62**, 201–225.

Bardina, J., Ferziger, J. & Reynolds, W.C. 1980 Improved subgrid-scale models for large-eddy simulation. In *13th Fluid and Plasmadynamics Conference*, p. 1357. AIAA.

Beck, A., Flad, D. & Munz, C.-D. 2019 Deep neural networks for data-driven LES closure models. *J. Comput. Phys.* **398**, 108910.

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

BENGIO, Y., SIMARD, P. & FRASCONI, P. 1994 Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks* **5** (2), 157–166.

CHENG, Y., *et al.* 2022 Deep learning for subgrid-scale turbulence modeling in large-eddy simulations of the convective atmospheric boundary layer. *J. Adv. Model. Earth Syst.* **14** (5), e2021MS002847.

CHO, C. & CHOI, H. 2023 A dynamic recursive neural-network-based subgrid-scale model for large eddy simulation. In *76th Annual Meeting of the Division of Fluid Dynamics, Washington DC, USA*.

CHOW, F.K. & MOIN, P. 2003 A further study of numerical errors in large-eddy simulations. *J. Comput. Phys.* **184** (2), 366–380.

CLARK, R.A., FERZIGER, J.H. & REYNOLDS, W.C. 1979 Evaluation of subgrid-scale models using an accurately simulated turbulent flow. *J. Fluid Mech.* **91** (1), 1–16.

COMTE-BELLOT, G. & CORRSIN, S. 1971 Simple Eulerian time correlation of full-and narrow-band velocity signals in grid-generated, 'isotropic' turbulence. *J. Fluid Mech.* **48** (2), 273–337.

DRUMMOND, C. & HOLTE, R.C. 2003 C4. 5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling. In *Workshop on Learning from Imbalanced Datasets II*, vol. 11, pp. 1–8. International Conference on Machine Learning.

GAMAHARA, M. & HATTORI, Y. 2017 Searching for turbulence models by artificial neural network. *Phys. Rev. Fluids* **2**, 054604.

GERMANO, M., PIOMELLI, U., MOIN, P. & CABOT, W.H. 1991 A dynamic subgrid-scale eddy viscosity model. *Phys. Fluids* A **3** (7), 1760–1765.

GHOSAL, S. 1996 An analysis of numerical errors in large-eddy simulations of turbulence. *J. Comput. Phys.* **125** (1), 187–206.

GHOSAL, S., LUND, T.S., MOIN, P. & AKSELVOLL, K. 1995 A dynamic localization model for large-eddy simulation of turbulent flows. *J. Fluid Mech.* **286**, 229–255.

GOODFELLOW, I., BENGIO, Y. & COURVILLE, A. 2016 *Deep Learning*. MIT Press.

GUAN, Y., CHATTOPADHYAY, A., SUBEL, A. & HASSANZADEH, P. 2022 Stable a posteriori LES of 2D turbulence using convolutional neural networks: backscattering analysis and generalization to higher re via transfer learning. *J. Comput. Phys.* **458**, 111090.

GUAN, Y., SUBEL, A., CHATTOPADHYAY, A. & HASSANZADEH, P. 2023 Learning physics-constrained subgrid-scale closures in the small-data regime for stable and accurate LES. *Physica* D **443**, 133568.

HE, K., ZHANG, X., REN, S. & SUN, J. 2016 Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (ed. L. O'Conner), pp. 770–778. IEEE.

IOFFE, S. & SZEGEDY, C. 2015 Batch normalization: accelerating deep network training by reducing internal covariate shift. In *Proceedings of 32nd International Conference on Machine Learning* (ed. F. Bach & D. Blei), vol. 37, pp. 448–456. Journal of Machine Learning Research.

JAMAAT, A.G.T. & HATTORI, B.Y. 2022 Development of subgrid-scale model for LES of Burgers turbulence with large filter size. *Phys. Fluids* **34** (4), 045120.

KANG, H.S., CHESTER, S. & MENEVEAU, C. 2003 Decaying turbulence in an active-grid-generated flow and comparisons with large-eddy simulation. *J. Fluid Mech.* **480**, 129–160.

KANG, M., JEON, Y. & YOU, D. 2023 Neural-network-based mixed subgrid-scale model for turbulent flow. *J. Fluid Mech.* **962**, A38.

KIM, J., KIM, H., KIM, J. & LEE, C. 2022 Deep reinforcement learning for large-eddy simulation modeling in wall-bounded turbulence. *Phys. Fluids* **34** (10), 105132.

KIM, M. & CHOI, H. 2023 Large eddy simulation of flow over a circular cylinder using a neural-network-based subgrid-scale model and its application to complex turbulent flows. In *76th Annual Meeting of the Division of Fluid Dynamics, Washington DC, USA*.

KIM, M., PARK, J. & CHOI, H. 2024 Large eddy simulation of flow over a circular cylinder with a neural-network-based subgrid-scale model. *J. Fluid Mech.* **984**, A6.

KINGMA, D.P. & BA, J. 2014 Adam: A method for stochastic optimization. arXiv:1412.6980

KNIGHT, D., ZHOU, G., OKONG'O, N. & SHUKLA, V. 1998 Compressible large eddy simulation using unstructured grids. In *36th AIAA Aerospace Sciences Meeting and Exhibit*, p. 535. AIAA.

KRAVCHENKO, A.G. & MOIN, P. 2000 Numerical studies of flow over a circular cylinder at $Re_D = 3900$. *Phys. Fluids* **12** (2), 403–417.

KURZ, M., OFFENHÄUSER, P. & BECK, A. 2023 Deep reinforcement learning for turbulence modeling in large eddy simulations. *Intl J. Heat Fluid Flow* **99**, 109094.

LANGFORD, J.A. & MOSER, R.D. 1999 Optimal LES formulations for isotropic turbulence. *J. Fluid Mech.* **398**, 321–346.

LEE, J., CHOI, H. & PARK, N. 2010 Dynamic global model for large eddy simulation of transient flow. *Phys. Fluids* **22** (7), 075106.

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

LE RIBAULT, C., SARKAR, S. & STANLEY, S.A. 1999 Large eddy simulation of a plane jet. *Phys. Fluids* **11** (10), 3069–3083.

LILLY, D.K. 1992 A proposed modification of the germano subgrid-scale closure method. *Phys. Fluids* A **4** (3), 633–635.

LIU, B., YU, H., HUANG, H., LIU, N. & LU, X. 2022 Investigation of nonlocal data-driven methods for subgrid-scale stress modeling in large eddy simulation. *AIP Adv.* **12** (6), 065129.

LIU, S., MENEVEAU, C. & KATZ, J. 1994 On the properties of similarity subgrid-scale models as deduced from measurements in a turbulent jet. *J. Fluid Mech.* **275**, 83–119.

LIU, W., QI, H., SHI, H., YU, C. & LI, X. 2023 Helical model based on artificial neural network for large eddy simulation of compressible wall-bounded turbulent flows. *Phys. Fluids* **35** (4), 045120.

LIU, X.Y., WU, J. & ZHOU, Z.H. 2008 Exploratory undersampling for class-imbalance learning. *IEEE Trans. Syst. Man Cybern.* B *Cybern.* **39** (2), 539–550.

MAAS, A.L., HANNUN, A.Y. & NG, A.Y. 2013 Rectifier nonlinearities improve neural network acoustic models. In *Proceedings of 30th International Conference on Machine Learning* (ed. S. Dasgupta & D. McAllester), vol. 28, p. 3. Journal of Machine Learning Research.

MACART, J.F., SIRIGNANO, J. & FREUND, J.B. 2021 Embedded training of neural-network subgrid-scale turbulence models. *Phys. Rev. Fluids* **6**, 050502.

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. 2018 Data-driven deconvolution for large eddy simulations of Kraichnan turbulence. *Phys. Fluids* **30** (12), 125109.

MAULIK, R., SAN, O., RASHEED, A. & VEDULA, P. 2019 Subgrid modelling for two-dimensional turbulence using neural networks. *J. Fluid Mech.* **858**, 122–144.

MENEVEAU, C. & KATZ, J. 2000 Scale-invariance and turbulence models for large-eddy simulation. *Annu. Rev. Fluid Mech.* **32** (1), 1–32.

MEYERS, J., GEURTS, B.J. & BAELMANS, M. 2003 Database analysis of errors in large-eddy simulation. *Phys. Fluids* **15** (9), 2740–2755.

MOSER, R.D., MALAYA, N.P., CHANG, H., ZANDONADE, P.S., VEDULA, P., BHATTACHARYA, A. & HASELBACHER, A. 2009 Theoretically based optimal large-eddy simulation. *Phys. Fluids* **21** (10), 105104.

NICOUD, F. & DUCROS, F. 1999 Subgrid-scale stress modelling based on the square of the velocity gradient tensor. *Flow Turbul. Combust.* **62** (3), 183–200.

NOVATI, G., DE LAROUSSILHE, H.L. & KOUMOUTSAKOS, P. 2021 Automating turbulence modelling by multi-agent reinforcement learning. *Nat. Mach. Intell.* **3** (1), 87–96.

PARK, J. & CHOI, H. 2021 Toward neural-network-based large eddy simulation: application to turbulent channel flow. *J. Fluid Mech.* **914**, A16.

PARK, N., LEE, S., LEE, J. & CHOI, H. 2006 A dynamic subgrid-scale eddy viscosity model with a global model coefficient. *Phys. Fluids* **18** (12), 125109.

PARK, N., YOO, J.Y. & CHOI, H. 2005 Toward improved consistency of a priori tests with a posteriori tests in large eddy simulation. *Phys. Fluids* **17** (01), 015103.

PAWAR, S., SAN, O., RASHEED, A. & VEDULA, P. 2020 A priori analysis on deep learning of subgrid-scale parameterizations for Kraichnan turbulence. *Theor. Comput. Fluid Dyn.* **34**, 429–455.

PAWAR, S., SAN, O., RASHEED, A. & VEDULA, P. 2023 Frame invariant neural network closures for Kraichnan turbulence. *Physica* A **609**, 128327.

POPE, S.B. 2000 *Turbulent Flows*. Cambridge University Press.

PRAKASH, A., JANSEN, K.E. & EVANS, J.A. 2022 Invariant data-driven subgrid stress modeling in the strain-rate eigenframe for large eddy simulation. *Comput. Meth. Appl. Mech. Engng* **399**, 115457.

PRAT, A., SAUTORY, T. & NAVARRO-MARTINEZ, S. 2020 A priori sub-grid modelling using artificial neural networks. *Intl J. Comput. Fluid Dyn.* **34** (6), 397–417.

ROSALES, C. & MENEVEAU, C. 2005 Linear forcing in numerical simulations of isotropic turbulence: physical space implementations and convergence properties. *Phys. Fluids* **17** (9), 095106.

SALVETTI, M.V. & BANERJEE, S. 1995 *A priori* tests of a new dynamic subgrid-scale model for finite-difference large-eddy simulations. *Phys. Fluids* **7** (11), 2831–2847.

SARGHINI, F., DE FELICE, G. & SANTINI, S. 2003 Neural networks based subgrid scale modeling in large eddy simulations. *Comput. Fluids* **32** (1), 97–108.

SIMONYAN, K. & ZISSERMAN, A. 2014 Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556

SIRIGNANO, J. & MACART, J.F. 2023 Deep learning closure models for large-eddy simulation of flows around bluff bodies. *J. Fluid Mech.* **966**, A26.

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press

SIRIGNANO, J., MACART, J.F. & FREUND, J.B. 2020 DPM: a deep learning PDE augmentation method with application to large-eddy simulation. *J. Comput. Phys.* **423**, 109811.

SMAGORINSKY, J. 1963 General circulation experiments with the primitive equations: I. The basic experiment. *Mon. Weath. Rev.* **91** (3), 99–164.

STOFFER, R., VAN LEEUWEN, C.M., PODAREANU, D., CODREANU, V., VEERMAN, M.A., JANSSENS, M., HARTOGENSIS, O.K. & VAN HEERWAARDEN, C.C. 2021 Development of a large-eddy simulation subgrid model based on artificial neural networks: a case study of turbulent channel flow. *Geosci. Model. Dev.* **14**, 3769–3788.

SUBEL, A., CHATTOPADHYAY, A., GUAN, Y. & HASSANZADEH, P. 2021 Data-driven subgrid-scale modeling of forced Burgers turbulence using deep learning with generalization to higher Reynolds numbers via transfer learning. *Phys. Fluids* **33** (3), 031702.

VÖLKER, S., MOSER, R.D. & VENUGOPAL, P. 2002 Optimal large eddy simulation of turbulent channel flow based on direct numerical simulation statistical data. *Phys. Fluids* **14** (10), 3675–3691.

VOLLANT, A., BALARAC, G. & CORRE, C. 2016 A dynamic regularized gradient model of the subgrid-scale stress tensor for large-eddy simulation. *Phys. Fluids* **28** (2), 025114.

VREMAN, A.W. 2004 An eddy-viscosity subgrid-scale model for turbulent shear flow: algebraic theory and applications. *Phys. Fluids* **16** (10), 3670–3681.

VREMAN, B., GEURTS, B. & KUERTEN, H. 1994 Realizability conditions for the turbulent stress tensor in large-eddy simulation. *J. Fluid Mech.* **278**, 351–362.

VREMAN, B., GEURTS, B. & KUERTEN, H. 1996 Large-eddy simulation of the temporal mixing layer using the Clark model. *Theor. Comput. Fluid Dyn.* **8** (4), 309–324.

VREMAN, B., GEURTS, B. & KUERTEN, H. 1997 Large-eddy simulation of the turbulent mixing layer. *J. Fluid Mech.* **339**, 357–390.

WANG, Y., YUAN, Z., XIE, C. & WANG, J. 2021 Artificial neural network-based spatial gradient models for large-eddy simulation of turbulence. *AIP Adv.* **11** (5), 055216.

WANG, Z., LUO, K., LI, D., TAN, J. & FAN, J. 2018 Investigations of data-driven closure for subgrid-scale stress in large-eddy simulation. *Phys. Fluids* **30** (12), 125101.

WOLLBLAD, C. & DAVIDSON, L. 2008 Pod based reconstruction of subgrid stresses for wall bounded flows using neural networks. *Flow Turbul. Combust.* **81**, 77–96.

XIE, C., WANG, J., LI, H., WAN, M. & CHEN, S. 2019*a* Artificial neural network mixed model for large eddy simulation of compressible isotropic turbulence. *Phys. Fluids* **31** (8), 085112.

XIE, C., WANG, J., LI, H., WAN, M. & CHEN, S. 2020*a* Spatial artificial neural network model for subgrid-scale stress and heat flux of compressible turbulence. *Theor. Appl. Mech. Lett.* **10** (1), 27–32.

XIE, C., WANG, J., LI, H., WAN, M. & CHEN, S. 2020*b* Spatially multi-scale artificial neural network model for large eddy simulation of compressible isotropic turbulence. *AIP Adv.* **10**, 015044.

XIE, C., WANG, J., LI, K. & MA, C. 2019*b* Artificial neural network approach to large-eddy simulation of compressible isotropic turbulence. *Phys. Rev.* E **99**, 053113.

XIE, C., WANG, J. & WEINAN, E. 2020*c* Modeling subgrid-scale forces by spatial artificial neural networks in large eddy simulation of turbulence. *Phys. Rev. Fluids* **5**, 054606.

XIE, C., YUAN, Z. & WANG, J. 2020*d* Artificial neural network-based nonlinear algebraic models for large eddy simulation of turbulence. *Phys. Fluids* **32**, 115101.

XU, B., WANG, N., CHEN, T. & LI, M 2015 Empirical evaluation of rectified activations in convolutional network. arXiv:1505.00853

YU, C., YUAN, Z., QI, H., WANG, J., LI, X. & CHEN, S. 2022 Kinetic-energy-flux-constrained model using an artificial neural network for large-eddy simulation of compressible wall-bounded turbulence. *J. Fluid Mech.* **932**, A23.

YUAN, Z., XIE, C. & WANG, J. 2020 Deconvolutional artificial neural network models for large eddy simulation of turbulence. *Phys. Fluids* **32** (11), 115106.

ZANNA, L. & BOLTON, T. 2020 Data-driven equation discovery of ocean mesoscale closures. *Geophys. Res. Lett.* **47** (17), e2020GL088376.

ZHOU, Z., HE, G., WANG, S. & JIN, G. 2019 Subgrid-scale model for large-eddy simulation of isotropic turbulent flows using an artificial neural network. *Comput. Fluids* **195**, 104319.

https://doi.org/10.1017/jfm.2024.992 Published online by Cambridge University Press