J. Fluid Mech. (2019), *vol.* 858, *pp.* 122–144.
doi:10.1017/jfm.2018.770

Subgrid modelling for two-dimensional turbulence using neural networks

R. Maulik[1], O. San[1,†], A. Rasheed[2] and P. Vedula[3]

[1]School of Mechanical and Aerospace Engineering, Oklahoma State University, Stillwater, OK 74078, USA

[2]CSE Group, Mathematics and Cybernetics, SINTEF Digital, N-7465 Trondheim, Norway

[3]School of Aerospace and Mechanical Engineering, The University of Oklahoma, Norman, OK 73019, USA

(Received 15 June 2018; revised 20 September 2018; accepted 20 September 2018; first published online 2 November 2018)

In this investigation, a data-driven turbulence closure framework is introduced and deployed for the subgrid modelling of Kraichnan turbulence. The novelty of the proposed method lies in the fact that snapshots from high-fidelity numerical data are used to inform artificial neural networks for predicting the turbulence source term through localized grid-resolved information. In particular, our proposed methodology successfully establishes a map between inputs given by stencils of the vorticity and the streamfunction along with information from two well-known eddy-viscosity kernels. Through this we predict the subgrid vorticity forcing in a temporally and spatially dynamic fashion. Our study is both *a priori* and *a posteriori* in nature. In the former, we present an extensive hyper-parameter optimization analysis in addition to learning quantification through probability-density-function-based validation of subgrid predictions. In the latter, we analyse the performance of our framework for flow evolution in a classical decaying two-dimensional turbulence test case in the presence of errors related to temporal and spatial discretization. Statistical assessments in the form of angle-averaged kinetic energy spectra demonstrate the promise of the proposed methodology for subgrid quantity inference. In addition, it is also observed that some measure of *a posteriori* error must be considered during optimal model selection for greater accuracy. The results in this article thus represent a promising development in the formalization of a framework for generation of heuristic-free turbulence closures from data.

Key words: computational methods, quasi-geostrophic flows, turbulence modelling

1. Introduction

The efficient computational modelling of energetic flows continues to remain an important area of research for many engineering and geophysical applications. Over the past few decades, coarse-grained techniques such as Reynolds-averaged Navier–Stokes (RANS) and large-eddy simulation (LES) have proven promising for

† Email address for correspondence: osan@okstate.edu

the statistically accurate prediction of the grid-resolved scales of a turbulent flow. While RANS is based on the modelling of turbulence in a temporally averaged sense, LES requires the specification of a model for the finer scales and their effect on the grid-resolved quantities. This modelling of the excluded wavenumbers in LES represents the classical closure problem which has spawned a variety of algebraic or equation-based techniques for representing the effect of these discarded scales on the resolved ones (Berselli, Iliescu & Layton 2005; Sagaut 2006). It has generally been observed that the choice of the subgrid model is physics-dependent, i.e. that different flow phenomena require different expressions for subgrid terms with *a priori* assumptions of phenomenology (Vreman 2004). We use this fact as a motivation for moving to an equation-free model for the source term through the use of an artificial neural network (ANN). Our hope, in addition to the formulation of a prediction framework, is to devise the formalization of a 'machine-learning experiment' where *a priori* model selection and *a posteriori* deployment are coupled to reveal information about the physical characteristics of a particular flow class. This not only enables the selection of computationally efficient predictive models but also reveals the importance of certain grid-resolved quantities of interest from the flow characteristics. In accordance with the recent trends of first-principles informed learning for physics inference in turbulence (Ling & Templeton 2015; Tracey, Duraisamy & Alonso 2015; Xiao *et al.* 2016; Schaeffer 2017; Singh, Medida & Duraisamy 2017; Wang *et al.* 2017*a*; Wang, Wu & Xiao 2017*b*; Weatheritt & Sandberg 2017; Mohan & Gaitonde 2018; Raissi & Karniadakis 2018; Wan *et al.* 2018; Wu, Xiao & Paterson 2018*a*), a major goal of this research is to study the combination of the traditional learning framework (inherently data-driven) and the physics-based prediction tool (based on the coarse-grained Navier–Stokes equations). We devote particular attention to the necessity for physical realizability as well as the issues faced by learning frameworks and their interactions with numerical discretization error.

Over the past decade, there have been multiple studies on the use of machine-learning tools for the reduced-order prediction of energetic flow physics. The study of these techniques has been equally popular for both severely truncated systems such as those obtained by leveraging sparsity in transformed bases (Faller & Schreck 1997; Cohen *et al.* 2003; Mannarino & Mantegazza 2014; San & Maulik 2018) as well as for modelling methodologies for coarse-grained meshes such as LES and RANS simulations (Maulik & San 2017*a*; Wang *et al.* 2017*b*; Wu, Xiao & Paterson 2018*b*). Therefore, they represent a promising direction for the assimilation of high-fidelity numerical and experimental data during the model-formulation phase for improved predictions during deployment. A hybrid formulation leveraging our knowledge of governing equations and augmenting these with machine learning represents a great opportunity for obtaining optimal LES closures for multiscale physics simulations (Langford & Moser 1999; Moser *et al.* 2009; King, Hamlington & Dahm 2016; Pathak *et al.* 2018).

From the point of view of turbulence modelling, we follow a strategy of utilizing machine-learning methods for estimating the subgrid forcing quantity such as the one utilized in Ling, Kurzawski & Templeton (2016) where a deep ANN has been described for Reynolds stress predictions in an invariant subspace. ANNs have been also implemented in Parish & Duraisamy (2016) to correct errors in RANS turbulence models after the formulation of a field-inversion step. Gamahara & Hattori (2017) detailed the application of ANNs for identifying quantities of interest for subgrid modelling in a turbulent channel flow through the measurement of Pearson correlation coefficients. Milano & Koumoutsakos (2002) also implemented these

techniques for turbulent channel flow but for the generation of low-order wall models, while Sarghini, De Felice & Santini (2003) deployed ANNs for the prediction of the Smagorinsky coefficient (and thus the subgrid contribution) in a mixed subgrid model. In Beck, Flad & Munz (2018), an ANN prediction has been hybridized with a least-squares projection onto a truncated eddy-viscosity model for LES. In these (and most) utilizations of machine-learning techniques, subgrid effects were estimated using grid-resolved quantities. Our approach is similar, wherein grid-resolved information is embedded into the input variables for predicting LES source terms for the filtered vorticity transport equation.

We outline a methodology for the development, testing and validation of a purely data-driven LES modelling strategy using ANNs which precludes the utilization of any phenomenology. However, in our framework the machine-learning paradigm is used for predicting the vorticity forcing or damping of the unresolved scales, which lends to an easier characterization of numerical stability restrictions as well as ease of implementation. Our model development and testing framework is outlined for Kraichnan turbulence (Kraichnan 1967) where it is observed that a combination of *a priori* and *a posteriori* analyses ensure the choice of model frameworks that are optimally accurate and physically constrained during prediction. Conclusions are drawn by statistical comparison of predictions with high-fidelity data drawn from direct numerical simulations (DNS).

To improve the viability of our proposed ideas, we devise our learning using extremely subsampled datasets. The use of such subsampled data necessitates a greater emphasis on physics distillation to prevent extrapolation and over-fitting during the training phase. An *a priori* hyper-parameter optimization is detailed for the selection of our framework architecture before deployment. An *a posteriori* prediction in a numerically evolving flow tests the aforementioned 'learning' of the framework for spectral scaling recovery which are compared to robust models utilizing algebraic eddy viscosities given by the Smagorinsky (1963) and Leith (1968) models. A hardwired numerical realizability also ensures viscous stability of the proposed framework in an *a posteriori* setting. Later discussions demonstrate how the proposed framework is suitable for the prediction of vorticity forcing as well as damping in the modelled scales. The proposed formulation also ensures data locality, where a dynamic forcing or dissipation of vorticity is specified spatio-temporally.

Following our primary assessments, our article proposes the use of a combined *a priori* and *a posteriori* study for optimal predictions of kinetic energy spectra as well as hyper-parameter selection prior to deployment for different flows that belong to the same class but have a different control parameter or initial conditions. It is also observed that the specification of eddy-viscosity kernels (which are devised from dimensional analyses) constrain the predictive performance of the framework for the larger scales. Results also detail the effect of data locality, where an appropriate region of influence utilized for sampling is shown to generate improved accuracy. The reader may find a thorough review of concurrent ideas in Duraisamy, Iaccarino & Xiao (2018). An excellent review of the strengths and opportunities of using artificial neural networks for fluid dynamics applications may also be found in Kutz (2017).

The mathematical background of subgrid modelling for the LES of two-dimensional turbulence may be summarized in the following. In terms of the vorticity–stream-function formulation, our non-dimensional governing equation for incompressible flow may be represented as

$$\frac{\partial \omega}{\partial t} + J(\omega, \psi) = \frac{1}{Re}\nabla^2 \omega, \tag{1.1}$$

where Re is the Reynolds number, and ω and ψ are the vorticity and streamfunction, respectively, connected to each other through the Poisson equation given by

$$\nabla^2\psi = -\omega. \tag{1.2}$$

It may be noted that the Poisson equation implicitly ensures a divergence-free flow evolution. The nonlinear term (denoted the Jacobian) is given by

$$J(\omega, \psi) = \frac{\partial\psi}{\partial y}\frac{\partial\omega}{\partial x} - \frac{\partial\psi}{\partial x}\frac{\partial\omega}{\partial y}. \tag{1.3}$$

A reduced-order implementation of the aforementioned governing laws (i.e. an LES) is obtained through

$$\frac{\partial\bar{\omega}}{\partial t} + J(\bar{\omega}, \bar{\psi}) = \frac{1}{Re}\nabla^2\bar{\omega} + \Pi, \tag{1.4}$$

where the overbarred variables are now evolved on a grid with far fewer degrees of freedom. The subgrid term Π encapsulates the effects of the finer wavenumbers which have been truncated due to insufficient grid support and must be approximated by a model. Mathematically we may express this (ideal) loss as

$$\Pi = J(\bar{\omega}, \bar{\psi}) - \overline{J(\omega, \psi)}. \tag{1.5}$$

In essence, the basic principle of LES is to compute the largest scales of turbulent motion and use closures to model the contributions from the smallest turbulent flow scales. The nonlinear evolution equations introduce unclosed terms that must be modelled to account for local, instantaneous momentum and energy exchange between resolved and unresolved scales. If these inter-eddy interactions are not properly parametrized, then an increase in resolution will not necessarily improve the accuracy of these large scales (Frederiksen, O'Kane & Zidikheri 2013; Frederiksen & Zidikheri 2016). Additionally, most LES closures are based on three-dimensional turbulence considerations primarily encountered in engineering applications. These LES models fundamentally rely on the concept of the forward energy cascade, and their extension to geophysical flows is challenging (Eden & Greatbatch 2008; Fox-Kemper *et al.* 2011; San, Staples & Iliescu 2013), due to the effects of stratification and rotation, which suppress vertical motions in the thin layers of fluid. In the following, we shall elaborate on the use of a machine-learning framework to predict the approximate value of Π in a pointwise fashion on the coarser grid and assess the results of its deployment in both *a priori* and *a posteriori* testing. Through this we attempt to bypass an algebraic or differential-equation-based specification of the turbulence closure and let the data drive the quantity and quality of subgrid forcing. We note here that the definition of the subgrid source term given in (1.5) is formulated for the LES of two-dimensional Navier–Stokes equations in the vorticity–streamfunction formulation but the framework outlined in this article may be readily extended to the primitive-variable formulation in two or higher dimensions (Mansfield, Knio & Meneveau 1998; Marshall & Beninati 2003).

2. Machine-learning architecture

In this section, we introduce the machine-learning methodology employed for the previously described regression problem. The ANN, also known as a multilayered perceptron, consists of a set of linear or nonlinear mathematical operations on an input

space vector to establish a map to an output space. Other than the input and output spaces, an ANN is also said to contain multiple hidden layers (denoted so due to the obscure mathematical significance of the matrix operations occurring here). Each of these layers is an intermediate vector in a multistep transformation which is acted on by biasing and activation before the next set of matrix operations. Biasing refers to an addition of a constant vector to the incident vector at each layer, on its way to a transformed output. The process of activation refers to an elementwise functional modification of the incident vector to generally introduce nonlinearity into the eventual map. In contrast, no activation (also referred to as 'linear' activation) results in the incident vector being acted on solely by biasing. Note that each component of an intermediate vector corresponds to a unit cell also known as the neuron. The learning in this investigation is supervised, implying label data used for informing the optimal map between inputs and outputs. Mathematically, if our input vector $\boldsymbol{p}$ resides in a P-dimensional space and our desired output $\boldsymbol{q}$ resides in a Q-dimensional space, this framework establishes a map $\mathbb{M}$ as follows:

$$\mathbb{M} : \{p_1, p_2, \ldots, p_P\} \in \mathbb{R}^P \rightarrow \{q_1, q_2, \ldots, q_Q\} \in \mathbb{R}^Q. \tag{2.1}$$

A schematic for this map may be observed in figure 1, where input, output and hidden spaces are summarized. In equation form, our default optimal map is given by

$$\begin{aligned}\mathbb{M} : \{&\bar{\omega}_{i,j}, \bar{\omega}_{i,j+1}, \bar{\omega}_{i,j-1}, \ldots, \bar{\omega}_{i-1,j-1}, \bar{\psi}_{i,j}, \bar{\psi}_{i,j+1}, \bar{\psi}_{i,j-1}, \ldots, \bar{\psi}_{i-1,j-1}, |\bar{S}|_{i,j}, |\boldsymbol{\nabla}\bar{\omega}|_{i,j}\} \\ &\in \mathbb{R}^{20} \rightarrow \{\tilde{\Pi}_{i,j}\} \in \mathbb{R}^1,\end{aligned} \tag{2.2}$$

where

$$|\bar{S}| = \sqrt{4\left(\frac{\partial^2\bar{\psi}}{\partial x \partial y}\right)^2 + \left(\frac{\partial^2\bar{\psi}}{\partial x^2} - \frac{\partial^2\bar{\psi}}{\partial y^2}\right)^2} \quad \text{and} \quad |\boldsymbol{\nabla}\bar{\omega}| = \sqrt{\left(\frac{\partial\bar{\omega}}{\partial x}\right)^2 + \left(\frac{\partial\bar{\omega}}{\partial y}\right)^2} \tag{2.3a,b}$$

are eddy-viscosity kernel information input to the framework and $\tilde{\Pi}$ is the approximation to the true subgrid source term. Note that the indices i and j correspond to discrete spatial locations on a coarse-grained two-dimensional grid. The map represented by (2.2) is considered 'default' due to the utilization of a nine-point sampling stencil of vorticity and streamfunction (corresponding to 18 total inputs) and two other inputs of the Smagorinsky and Leith kernels. The purpose of utilizing the additional information from these well-established eddy-viscosity hypotheses may be considered a data pre-processing mechanism where certain important quantities of interest are distilled and presented 'as-is' to the network for simplified architectures and reduced training durations. The motivation behind the choice of these particular kernels is discussed in later sections, where it is revealed that they also introduce a certain regularization to the optimization. We note that all our variables in this study are non-dimensionalized at the stage of problem definition and no further pre-processing is utilized prior to exposing the map to the input data for predictions. The predicted value of $\tilde{\Pi}$ is post-processed before injection into the vorticity equation as follows:

$$\Pi = \begin{cases} \tilde{\Pi}, & \text{if } (\nabla^2\bar{\omega})(\tilde{\Pi}) > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{2.4}$$

This ensures numerical stability due to potentially negative eddy viscosities embedded in the source term prediction and may be considered to be an implicit assumption of

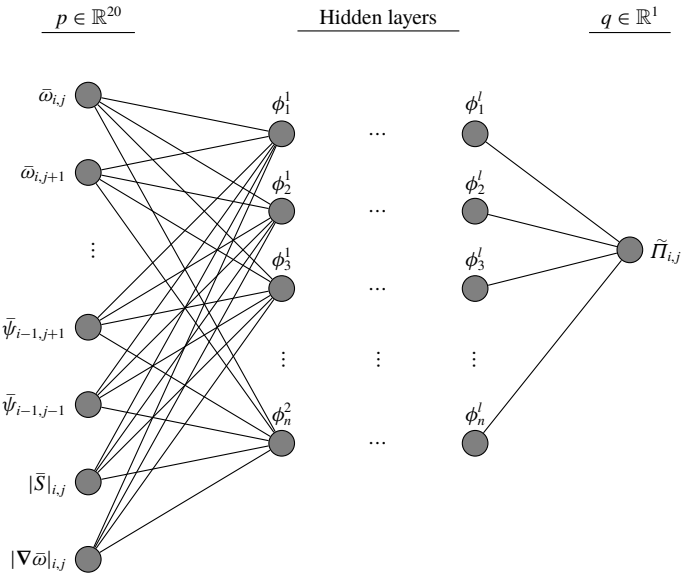

FIGURE 1. Proposed artificial neural network architecture and relation to sampling and prediction space.

the Boussinesq hypothesis for functional subgrid modelling. It is later demonstrated that the presence of this constraint does not preclude the prediction of positive or negative values of $\tilde{\Pi}$, which implies that the proposed framework is adept at predicting vorticity forcing or damping at the finer scales, respectively. The damping of vorticity at the finer scales would correspond to a lower dissipation of kinetic energy (assuming that vorticity dissipates kinetic energy in the subgrid scales). Similarly, the forcing of vorticity at the finer scales may be assumed to be a localized event of high kinetic energy dissipation. In general, (2.4) precludes the presence of a backscatter of enstrophy for strict adherence to viscous stability requirements on the coarse-grained mesh. Instead of the proposed truncation, one may also resort to some form of spatial averaging in an identifiable homogeneous direction as utilized by Germano *et al.* (1991). However, the former was chosen to remove any dependence on model forms or coefficient calculations. In what follows for the rest of this article, our proposed framework is denoted ANN-SGS. Details related to hyper-parameter selection and supervised learning of the model are provided in the appendices.

3. *A priori* validation

We first outline an *a priori* study for the proposed framework where the optimal map is utilized for predicting probability distributions for the true subgrid source term. In other words, we assess the turbulence model for a one-snapshot prediction. Before proceeding, we return to our previous discussion about the choice of Smagorinsky and Leith viscosity kernels by highlighting their behaviour for different choices of model coefficients (utilized in effective eddy-viscosity computations using mixing-length-based phenomenological arguments). The Smagorinsky or Leith subgrid-scale models may be implemented in the vorticity–streamfunction formulation via the specification of an effective eddy viscosity

$$\tilde{\Pi} = \nu_e \nabla^2 \bar{\omega}, \tag{3.1}$$

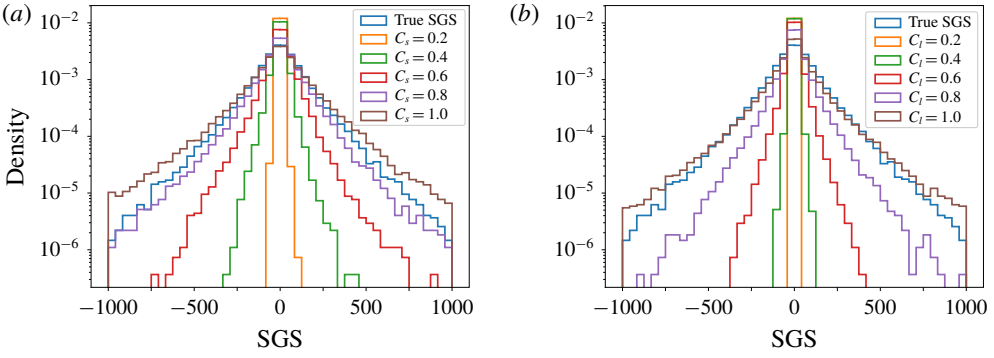

FIGURE 2. (Colour online) *A priori* performance of (*a*) Smagorinsky and (*b*) Leith models for varying model coefficients for data snapshot at $t = 2$. Here, instances refer to the probability densities of truth and prediction at different magnitudes.

where the Smagorinsky model utilizes

$$\nu_e = (C_s\delta)^2|\bar{S}|, \tag{3.2}$$

while the Leith hypothesis states

$$\nu_e = (C_l\delta)^3|\nabla\bar{\omega}|. \tag{3.3}$$

In the above relations, δ refers to the grid volume (or area in two-dimensional cases) and ν_e is an effective eddy viscosity. From figure 2, it is apparent that the choice of model-form coefficients C_s and C_l for the Smagorinsky and Leith models dictates the accuracy of the closure model in *a priori* analyses. Instances here refer to the probability densities of truth and prediction at different magnitudes. We would also like to draw the reader's attention to the fact that ideal reconstructions of the true subgrid term are with coefficients near the value of 1.0, a value that is rather different from the theoretically accepted values of C_s applicable in three-dimensional turbulence. This dependence of closure efficacy on model coefficients continues to represent a non-trivial *a priori* parameter specification task for practical utilization of common LES turbulence models particularly in geophysical applications. Later, we shall demonstrate that *a posteriori* implementations of these static turbulence models is beset with difficulties for non-stationary turbulent behaviour.

In contrast, figure 3 shows the performance of the proposed framework in predicting subgrid contributions purely through the indirect exposure to supervised data in the training process. Figure 3 shows a remarkable ability for Π reconstruction for both Re values of 32 000 and 64 000, solely from grid-resolved quantities. Performance similar to ideal model coefficients mentioned in figure 2 are also observed. The $Re = 64\,000$ case is utilized to assess model performance for 'out-of-training' snapshot data in an *a priori* sense. The trained framework is seen to lead to viable results for a completely unseen dataset with more energetic physics. We may thus conclude that the map has managed to embed a relationship between sharp spectral cutoff filtered quantities and subgrid source terms.

We also visually quantify the effect of (2.4) (described for the process of numerical realizability) in figure 4, where a hardwired truncation is utilized for precluding violation of viscous stability in the forward simulations of our learning deployment.

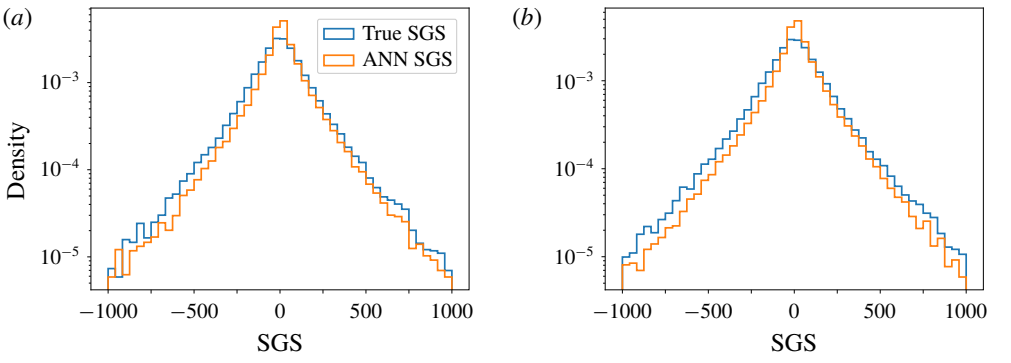

FIGURE 3. (Colour online) *A priori* results for the probability density distributions of the true and framework predicted LES source terms for (*a*) $Re = 32\,000$ and (*b*) $Re = 64\,000$. Note that the training data were generated for $Re = 32\,000$ only and prediction on $Re = 64\,000$ represents a stringent validation.

One can observe that the blue regions of figure 4, which are spatial locations of subgrid forcing ($\tilde{\Pi}$) and Laplacian $\nabla^2\bar{\omega}$ being of opposite sign, are truncated. However, we must clarify that this does not imply a constraint on the nature of forcing being obtained by our model – a negative value of the subgrid term implies a damping of vorticity and the finer scales, whereas a positive value implies production at the finer scales. Our next step is to assess the ability of this relationship to recover statistical trends in an *a posteriori* deployment. The fact that roughly half of the predicted subgrid terms are truncated matches the observations in Piomelli *et al.* (1991), where it is observed that forward- and backscatter are present in approximately equal amounts when extracted from DNS data. Studies are under way to extend some form of dynamic localization of backscatter to the current formulation along the lines of Ghosal *et al.* (1995).

4. Deployment and *a posteriori* assessment

The ultimate test of any data-driven closure model is in an *a posteriori* framework with subsequent assessment for the said model's ability to preserve coherent structures and scaling laws. While the authors have undertaken *a priori* studies with promising results for data-driven ideologies for LES (Maulik & San 2017*a*), the results of the following section are unique in that they represent a model-free turbulence model computation in temporally and spatially dynamic fashion. This test set-up is particulary challenging due to the neglected effects of numerics in the *a priori* training and assessment. In the following we utilize angle-averaged kinetic energy spectra to assess the ability of the proposed framework to preserve integral and inertial range statistics. In brief, we mention that the numerical implementation of the conservation laws is through second-order discretization for all spatial quantities (with a kinetic-energy-conserving Arakawa discretization (Arakawa 1966) for the calculation of the nonlinear Jacobian). A third-order total-variation-diminishing Runge–Kutta method is utilized for the vorticity evolution and a spectrally accurate Poisson solver is utilized for updating streamfunction values from the vorticity. Our proposed framework is deployed pointwise for approximate Π at each explicit time step until the final time of $t = 4$ is reached. The robustness of the network to the effects of numerics is thus examined.

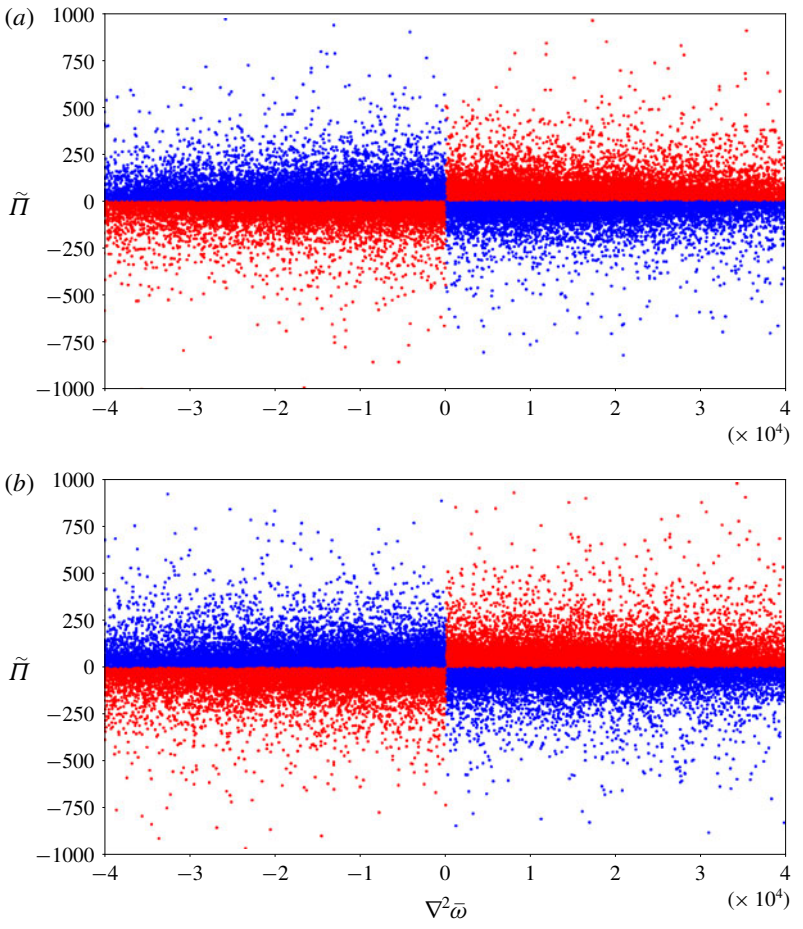

FIGURE 4. (Colour online) An *a priori* assessment of the nature of truncation given by (2.4) for $t = 2$ snapshot data at (*a*) $Re = 32\,000$ and (*b*) $Re = 64\,000$. The nature of this truncation is for the preservation of viscous stability in a coarse-grained forward simulation.

Figure 5 displays the statistical fidelity of coarse-grained simulations obtained with the deployment of the proposed framework for $Re = 32\,000$. Stable realizations of the vorticity field are generated due to the combination of our training and post-processing. For the purpose of comparison, we also include coarse-grained no-model simulations, i.e. unresolved numerical simulations (UNS), which demonstrate an expected accumulation of noise at grid cutoff wavenumbers. DNS spectra are also provided showing agreement with the k^{-3} theoretical scaling expected for two-dimensional turbulence. Our proposed framework is effective at stabilizing the coarse-grained flow by estimating the effect of subgrid quantities and preserving trends with regards to the inertial range scaling. We also demonstrate the utility of our learned map on an *a posteriori* simulation for $Re = 64\,000$ data where similar trends are recovered. This also demonstrates an additional stringent validation of the data-driven model for ensuring generalized learning. The reader may observe that Smagorinsky and Leith turbulence model predictions using static model coefficients of value 1.0

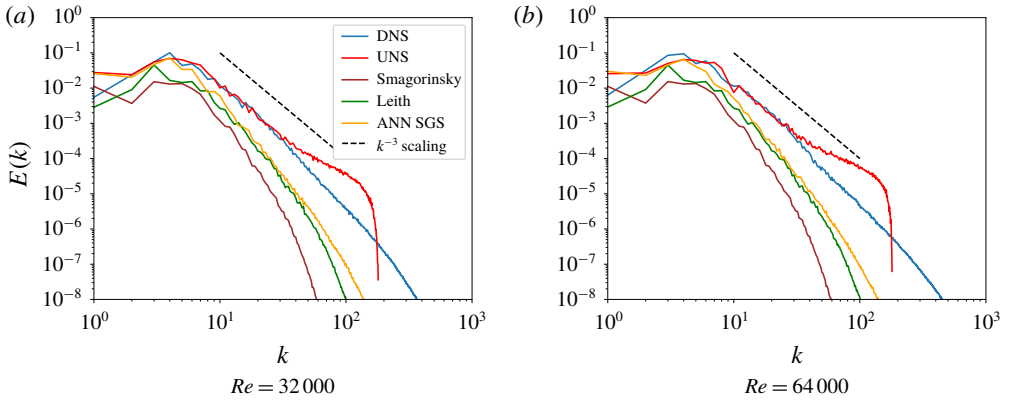

FIGURE 5. (Colour online) *A posteriori* results for the spatially averaged kinetic energy spectra for the proposed framework compared with DNS and UNS solutions. Note that only $Re = 32\,000$ training data are used for both deployments, and network is applied spatially and temporally in a dynamic manner until $t = 4$.

(i.e. $C_s = C_l = 1.0$) lead to over-dissipative results particularly at the lower (integral) wavenumbers. This trend is unsurprising, since the test case examined here represents non-stationary decaying turbulence for which fixed values of the coefficients are not recommended. Indeed, the application of the Smagorinsky model to various engineering and geophysical flow problems has revealed that the constant is not single-valued and varies depending on resolution and flow characteristics (Galperin & Orszag 1993; Canuto & Cheng 1997; Vorobev & Zikanov 2008), with higher values specifically for geophysical flows (Cushman-Roisin & Beckers 2011). In comparison, the proposed framework has embedded the adaptive nature of dissipation into its map, which is a promising outcome. Figures 6 and 7 show the performance of the Smagorinsky and Leith models, respectively, for a $Re = 32\,000$ and $Re = 64\,000$ *a posteriori* deployment for different values of the eddy-viscosity coefficients. One can observe that the choice of the model-form coefficient is critical in the capture of the lower wavenumber fidelity.

In particular, we would like to note that the choice of a coarse-grained forward simulation using a Reynolds number of 64 000 represents a test for establishing what the model has learned. This forward simulation verifies if the closure performance of the framework is generalizable and not a numerical artifact. A similar performance of the model on a different deployment scenario establishes the hybrid nature of our framework where the bulk behaviour of the governing law is retained (through the vorticity–streamfunction formulation) and the artificial intelligence acts as a corrector for statistical fidelity. This observation holds promise for the development of closures that are generalizable to multiple classes of flow without being restricted by initial or boundary conditions. To test the premise of this hypothesis, we also display ensemble-averaged kinetic energy spectra from multiple coarse-grained simulations at $Re = 32\,000$ and at $Re = 64\,000$, utilizing a different set of random initial conditions for each test case. In particular, we utilize 24 different tests for averaged spectra, which are displayed in figure 8. We would like to emphasize here that the different initial conditions correspond to the same initial energy spectrum in wavenumber space but with random vorticity fields in Cartesian space. The performance of our proposed framework is seen to be repeatable across different instances of random initial

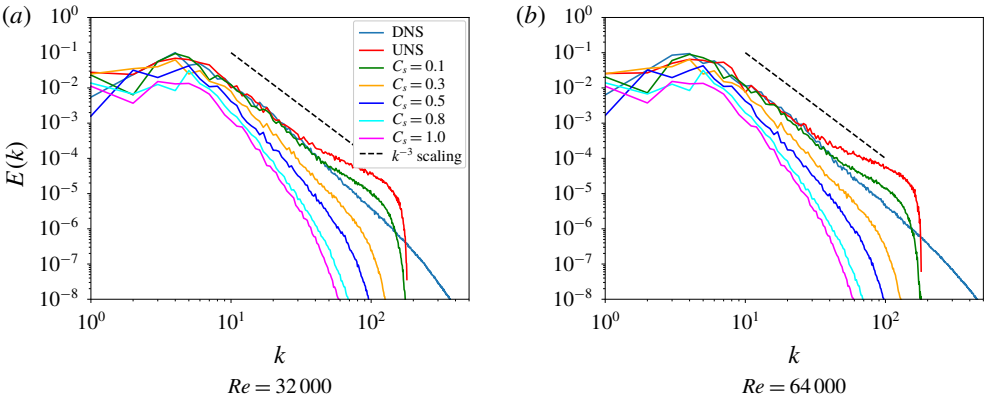

FIGURE 6. (Colour online) *A posteriori* results for the spatially averaged kinetic energy spectra for the Smagorinsky model for different values of their eddy-viscosity coefficients and for different Reynolds numbers at $t = 4$. One can observe that the capture of lower-wavenumber energy and scaling is heavily dependent on the value of these coefficients.

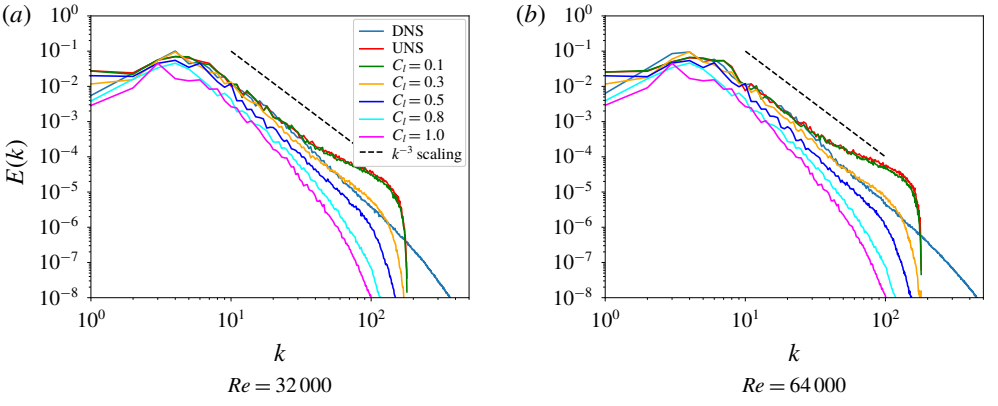

FIGURE 7. (Colour online) *A posteriori* results for the spatially averaged kinetic energy spectra for the Leith model for different values of their eddy-viscosity coefficients and for different Reynolds numbers at $t = 4$. One can observe that the capture of lower-wavenumber energy and scaling is heavily dependent on the value of these coefficients.

vorticity fields sharing the same energy spectra. Details related to the generation of these random initial conditions may be found in Maulik & San (2017*b*). In addition, we also display spectra obtained from an *a posteriori* deployment of our framework till $t = 6$ for $Re = 32\,000$ and $Re = 64\,000$, shown in figure 9, which ensures that the model has learned a subgrid closure effectively and predicts the vorticity forcing adequately in a temporal region to which it has not been exposed during training.

Figure 10 shows a qualitative assessment of the stabilization property of machine-learning framework where a significant reduction in noise can be visually ascertained

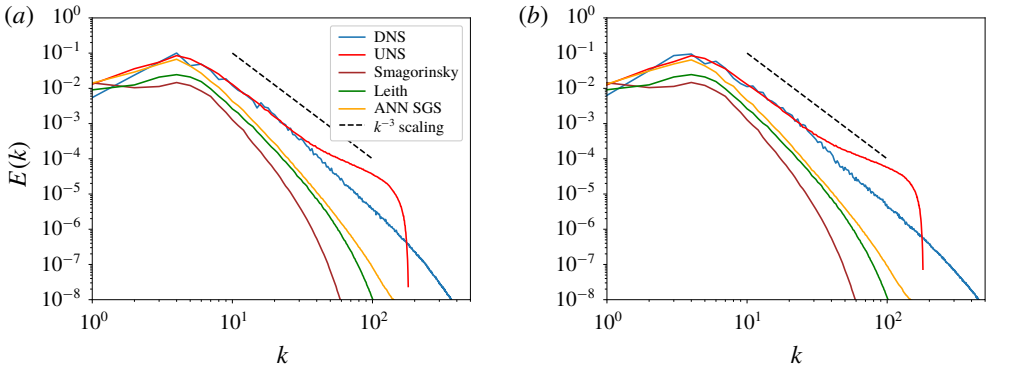

FIGURE 8. (Colour online) *A posteriori* results for 24 ensemble-averaged simulations for (*a*) $Re = 32\,000$ and (*b*) $Re = 64\,000$.

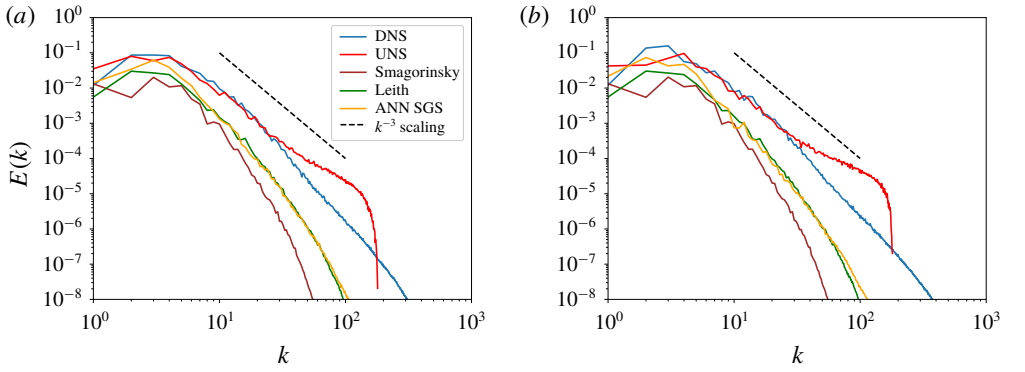

FIGURE 9. (Colour online) The deployment of our framework till $t = 6$ for (*a*) $Re = 32\,000$ and (*b*) $Re = 64\,000$ showing that a subgrid model has been learned for utility beyond the training region. We note that the training region is defined between $t = 0$ and $t = 4$ alone.

due its deployment. Coherent structures are retained successfully as against UNS results where high-wavenumber noise is seen to corrupt field realizations heavily. Filtered DNS (FDNS) data obtained by Fourier cutoff filtering of vorticity data obtained from DNS are also shown for the purpose of comparison. As discussed previously, the stabilization behaviour is observed for both $Re = 32\,000$ and $Re = 64\,000$ data. We may thus conclude that the learned model has established an implicit subgrid model as a function of grid-resolved variables. We reiterate that the choice of the eddy viscosities is motivated by ensuring a fair comparison with the static Smagorinsky and Leith subgrid models and studies are under way to increase complexity in the mapping as well as input space.

5. *A priori* and *a posteriori* dichotomy

In the previous sections, we have outlined the performance of our proposed framework according to the optimal model architecture chosen by a grid search for the number of hidden layers as well as the number of hidden-layer neurons.

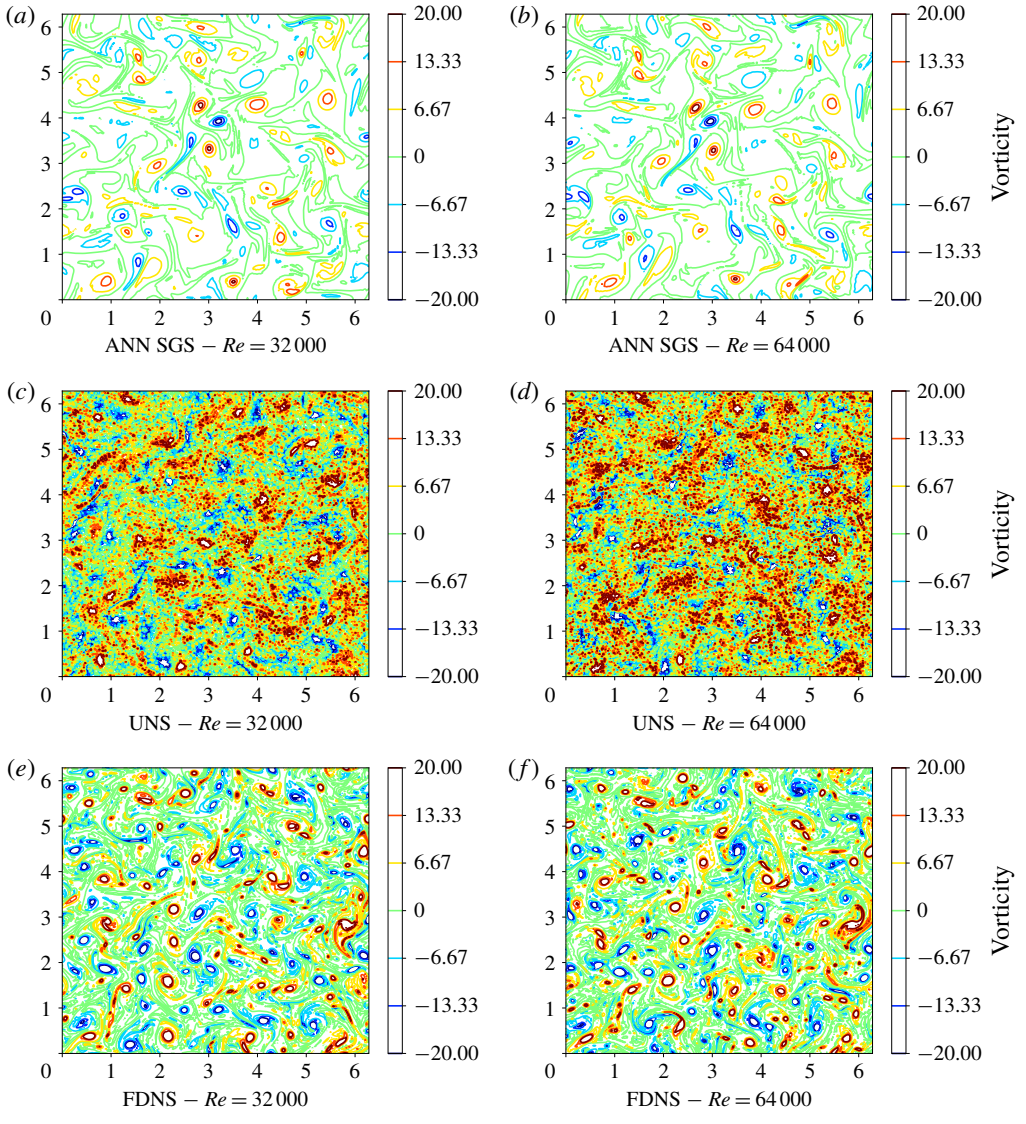

FIGURE 10. (Colour online) *A posteriori* results for the proposed framework showing vorticity fields for $Re = 32\,000$ and $Re = 64\,000$ data using coarse-grained grids (*a*,*b*). We also provide no-model simulations (*c*,*d*) and filtered DNS contours (*e*,*f*) for the purpose of comparison.

This *a priori* hyper-parameter selection is primarily devised on mean-squared-error minimization and is susceptible to providing model architectures that are less resistant to over-fitting and more prone to extrapolation. Our experience shows that an *a posteriori* prediction (such as for this simple problem) must be embedded into the model selection decision process to ensure an accurate learning of physics. We briefly summarize our observations of the *a priori* and *a posteriori* dichotomy in the following.

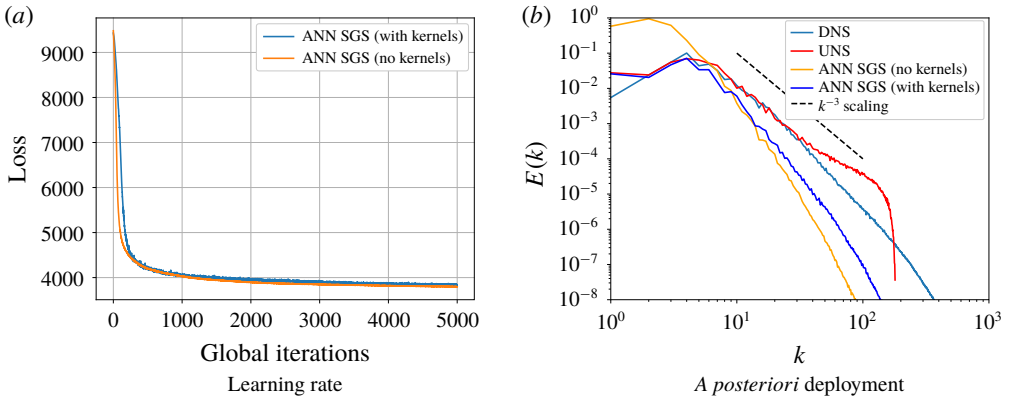

FIGURE 11. (Colour online) (*a*) *A priori* and (*b*) *a posteriori* effect of the utilization of eddy-viscosity kernel inputs in training and deployment for a two-layer 50-neuron network with a nine-point stencil. The presence of these kernels (intangible in *a priori* error minimization) leads to constrained statistical fidelity in *a posteriori* deployment at $Re = 32\,000$.

5.1. *Effect of eddy-viscosity inputs*

By fixing our optimal set of hyper-parameters (i.e. a two-layer 50-neuron network), we attempted to train a map using an input space without the choice of Smagorinsky and Leith viscosity kernels. Therefore our inputs would simply be the nine-point stencils for vorticity and streamfunction as shown in the mathematical expression given by

$$\mathbb{M} : \{\bar{\omega}_{i,j}, \bar{\omega}_{i,j+1}, \bar{\omega}_{i,j-1}, \ldots, \bar{\omega}_{i-1,j-1}, \bar{\psi}_{i,j}, \bar{\psi}_{i,j+1}, \bar{\psi}_{i,j-1}, \ldots, \bar{\psi}_{i-1,j-1}\} \in \mathbb{R}^{18} \rightarrow \{\tilde{\Pi}_{i,j}\} \in \mathbb{R}^{1}. \tag{5.1}$$

As shown in figure 11, the modification of our input space had very little effect on the training performance of our optimal network architecture. This would initially seem to suggest that the Smagorinsky and Leith kernels were not augmenting learning in any manner. However, our *a posteriori* deployment of this model, which mapped to subgrid quantities from the 18-dimensional input space, displayed an unconstrained behaviour at the larger scales with the formation of non-physical large-scale structures (also shown in figure 8). This strongly points towards an implicit regularization of our model due to the selection of input dimensions with these kernels.

We undertook the same study for a five-layer, 50-neuron ANN (one that was deemed too complex by our grid search) with results shown in figure 12. Two conclusions are apparent here – the utilization of these kernels in the learning process has prevented *a priori* reduction of training error at a much higher value, and the deployment of both networks (i.e. with and without input viscosities) has led to a constrained prediction of the k^{-3} spectral scaling. Large-scale statistical predictions remain unchanged and, indeed, a better agreement with the DNS spectrum can be observed with the deeper network with the use of the kernels.

5.2. *A posteriori informed architecture selection*

While *a priori* hyper-parameter tuning is classically utilized for most machine-learning deployments, the enforcement of physical realizability constraints (such as those given by (2.4)) and the presence of numerical errors during deployment may often

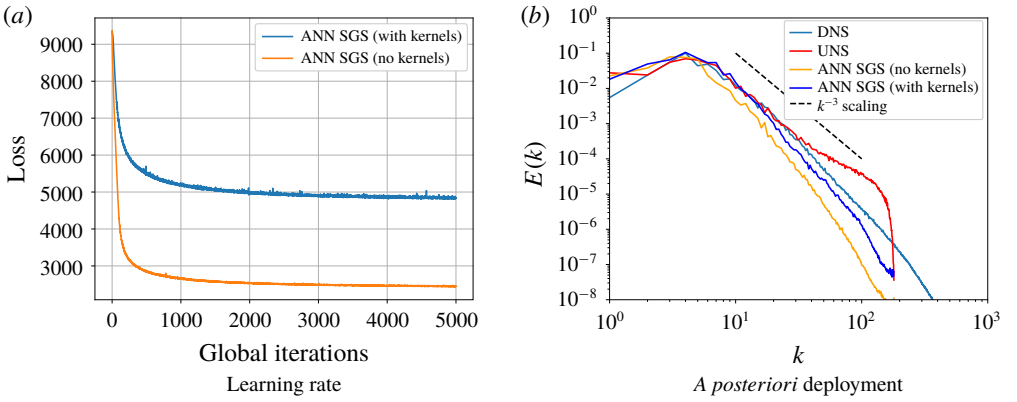

FIGURE 12. (Colour online) (*a*) *A priori* and (*b*) *a posteriori* effect of the utilization of eddy-viscosity kernel inputs in training and deployment for a five-layer 50-neuron network with a nine-point stencil. The presence of these kernels leads to higher training errors but viable statistical fidelity in *a posteriori* deployment at $Re = 32\,000$.

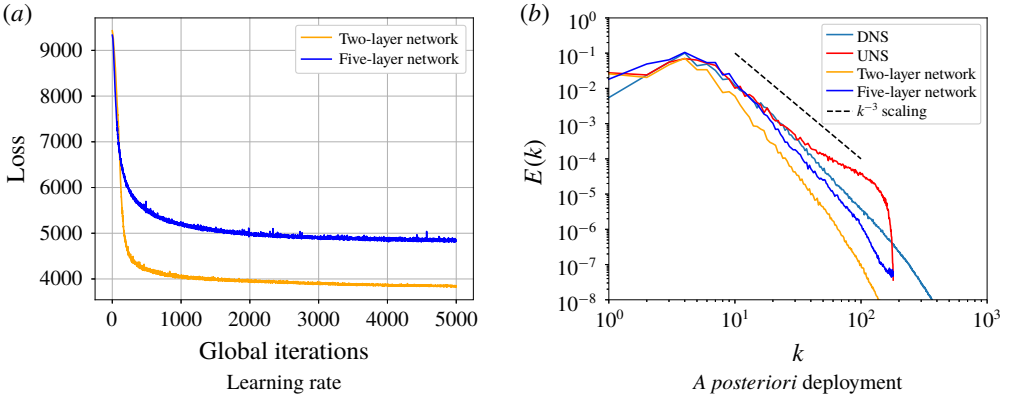

FIGURE 13. (Colour online) (*a*) *A priori* and (*b*) *a posteriori* effect of the number of hidden layers in the proposed framework. While the two-layered ANN with a nine-point stencil leads to excellent *a priori* results, the five-layered network predicts k^{-3} scaling more accurately in deployment for an *a posteriori* simulation at $Re = 32\,000$.

necessitate architectures that differ significantly during *a posteriori* deployment. This article demonstrates the fact that, while constrained predictions are obtained by our optimal two-layer network (obtained by a grid search), the utilization of a deeper network actually leads to more accurate predictions of the Kraichnan turbulence spectrum as shown in figure 13. This is despite the fact that the deeper network displays a great mean-squared-error during the training phase (which was the root cause of it being deemed ineligible in the hyper-parameter tuning). Figure 12 thus tells us that it is important to couple some form of *a posteriori* analysis during model-form selection before it is deemed optimal (physically or computationally) for deployment. We note that both networks tested in this subsection utilized the Smagorinsky and Leith eddy viscosities in their input space.

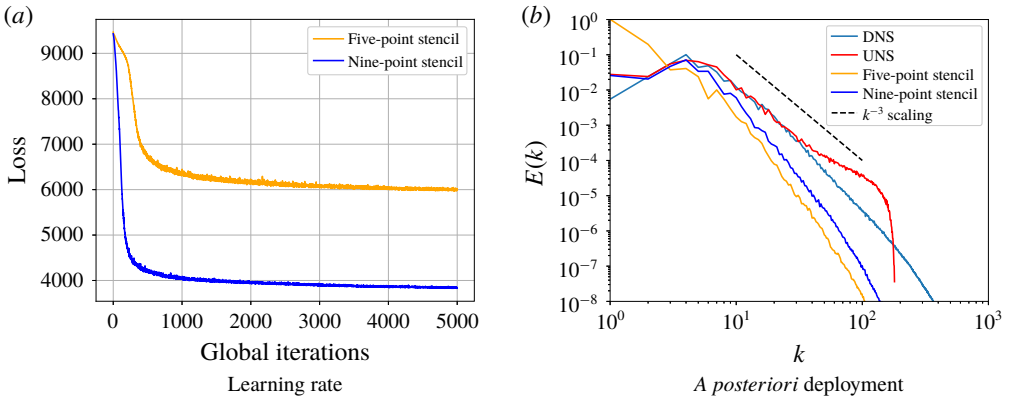

FIGURE 14. (Colour online) (*a*) *A priori* and (*b*) *a posteriori* effect of the stencil size in the two-layer, 50-neuron framework for a $Re = 32\,000$ simulation. While the nine-point stencil leads to similar *a priori* training errors, an *a posteriori* deployment at $Re = 32\,000$ reveals its limitations.

5.3. *Stencil selection*

Another comparison is made when the input dimension is substantially reduced by choosing a five-point stencil (instead of the aforementioned nine-point stencil). In this architecture, vorticity and streamfunction values are chosen only for the x and y directions (i.e. $\bar{\omega}_{i,j}$, $\bar{\omega}_{i+1,j}$, $\bar{\omega}_{i-1,j}$, $\bar{\omega}_{i,j+1}$, $\bar{\omega}_{i,j-1}$ for vorticity and similarly for streamfunction). The input eddy viscosities given by the Smagorinsky and Leith kernels are also provided to this reduced network architecture. Mathematically, this new map may be expressed as

$$\mathbb{M} : \{\bar{\omega}_{i,j}, \bar{\omega}_{i,j+1}, \bar{\omega}_{i,j-1}, \bar{\omega}_{i+1,j}, \bar{\omega}_{i-1,j}\bar{\psi}_{i,j}, \bar{\psi}_{i,j+1}, \bar{\psi}_{i,j-1}, \bar{\psi}_{i+1,j}, \bar{\psi}_{i-1,j}, |\bar{S}|_{i,j}, |\nabla\bar{\omega}|_{i,j}\} \in \mathbb{R}^{12} \rightarrow \{\tilde{\Pi}_{i,j}\} \in \mathbb{R}^{1}. \tag{5.2}$$

Figure 14 shows the performance of this set-up in training and deployment, where it can once again be observed that *a posteriori* analysis is imperative for determining a map for the subgrid terms. While training errors are more or less similar, the reduced stencil fails to capture the nonlinear relationship between the resolved and cutoff scales, with consequent results on the statistical fidelity of the lower wavenumbers. We perform a similar study related to this effect of data locality on a deeper network given by five layers and 50 neurons to verify the effect of the deeper architecture on constrained prediction. The results of this training and deployment are shown in figure 15, where it is observed that the increased depth of the ANN leads to a similar performance with a smaller stencil size. This implies that optimal data locality (in terms of the choice of a stencil) leads to a reduced number of hidden layers. Again, the *a priori* mean-squared error is not indicative of the quality of *a posteriori* prediction.

The main point to take away from this section thus becomes the fact that optimal architectures and maps for subgrid predictions require a careful *a priori* and *a posteriori* study for tractable computational problems (such as the Kraichnan turbulence case) before they may be deployed for representative flows. The effect of realizability constraints and numerical errors often leads to unexpected *a posteriori* performance and some form of lightweight deployment must be utilized for confirming model feasibility.

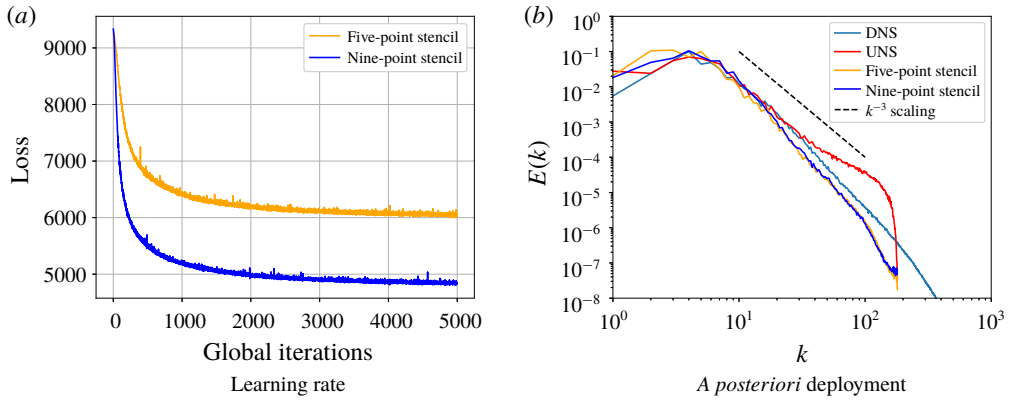

FIGURE 15. (Colour online) (*a*) *A priori* and (*b*) *a posteriori* effect of the stencil size in the five-layer, 50-neuron framework for a $Re = 32\,000$ simulation. With deeper architectures, the five- and nine-point stencils show similar statistical performance.

6. Conclusions

In this investigation, a purely data-driven approach to closure modelling utilizing artificial neural networks is detailed, implemented and analysed in both *a priori* and *a posteriori* assessments for decaying two-dimensional turbulence. An extensive hyper-parameter selection strategy is also performed prior to the selection of an optimal network architecture in addition to explanations regarding the choice of input space and truncation for numerical realizability. The motivation behind the search of a model-free closure stems from the fact that most closures utilize empirical or phenomenological relationships to determine closure strength with associated hazards of insufficient or more than adequate dissipation in *a posteriori* utilizations. To that end, our proposed framework utilizes an implicit map with inputs as grid-resolved variables and eddy viscosities to determine a dynamic closure strength. Our optimal map is determined by training an artificial neural network with extremely subsampled data obtained from high-fidelity direct numerical simulations of the decaying two-dimensional turbulence test case. Our inputs to the network are given by sampling stencils of vorticity and streamfunction in addition to two kernels utilized in the classical Smagorinsky and Leith models for eddy-viscosity computations. Based on these inputs, the network predicts a temporally and spatially dynamic closure term which is pre-processed for numerical stability before injection into the vorticity equation as a potential source (or sink) of vorticity in the finer scales. Our statistical studies show that the proposed framework is successful in imparting a dynamic dissipation of kinetic energy to the decaying turbulence problem for accurate capture of coherent structures and inertial range fidelity.

In addition, we also come to the conclusion that the effects of prediction truncation (for numerical realizability) and numerical error during forward simulation deployment necessitate the need for *a posteriori* analyses when identifying optimal architectures (such as the number of hidden layers and the input spaces). This conclusion has significant implications for the modern era of physics-informed machine learning for fluid dynamics applications where *a priori* trained learning is constrained by knowledge from first principles. Our conclusions point towards the need for coupling *a posteriori* knowledge during hyper-parameter optimization either passively (as demonstrated in this article) or through the use of custom training objective functions

which embed physics in the form of regularization. Our study basically proposes that data-driven spatio-temporally dynamic subgrid models may be developed for tractable computational cases such as Kraichnan and Kolmogorov turbulence through a combination of *a priori* and *a posteriori* study before they may be deployed for practical flow problems such as those encountered in engineering or geophysical flows. Studies are under way to extend these concepts to multiple flow classes in pursuit of data-driven closures that may prove to be more universal.

While this article represents the successful application of a proof of concept, our expectation is that further robust turbulence closures may be developed on the guidelines presented herein, with the utilization of more grid-resolved quantities such as flow invariants and physics-informed hyper-parameter optimization. In addition, network-embedded symmetry considerations are also being explored as a future enhancement for this research. Dataset pre-processing for outlier identification, not utilized in this study, is also a potential avenue for improved *a posteriori* performance and more efficient hyper-parameter selection. Our ultimate goal is to determine maps that may implicitly classify closure requirements according to inhomogeneities in a computational domain (through exposure to different flow classes) that may then be ported as predictive tools in multiscale phenomena with complex initial and boundary conditions. The results in this article indicate a promising first step in that direction.

Acknowledgements

The authors are grateful for helpful comments from the referees. This material is based upon work supported by the US Department of Energy, Office of Science, Office of Advanced Scientific Computing Research under Award Number DE-SC0019290. O.S. gratefully acknowledges their support. We also acknowledge the support of NVIDIA Corporation for supporting our research in this area. The computing for this project was partially performed at the OSU High Performance Computing Center at Oklahoma State University. O.S. and P.V. acknowledge the financial support received from the Oklahoma NASA EPSCoR Research Initiation Grant programme. A.R. acknowledges the financial support received from the Norwegian Research Council and the industrial partners of OPWIND: Operational Control for Wind Power Plants (grant no. 268044/E20).

Appendix A. Hyper-parameter optimization

In this appendix, we detail the process of *a priori* architecture selection before training and deployment. Our hidden layers have neurons that are activated by the rectified linear (ReLU) function. The choice of the ReLU activation was made for

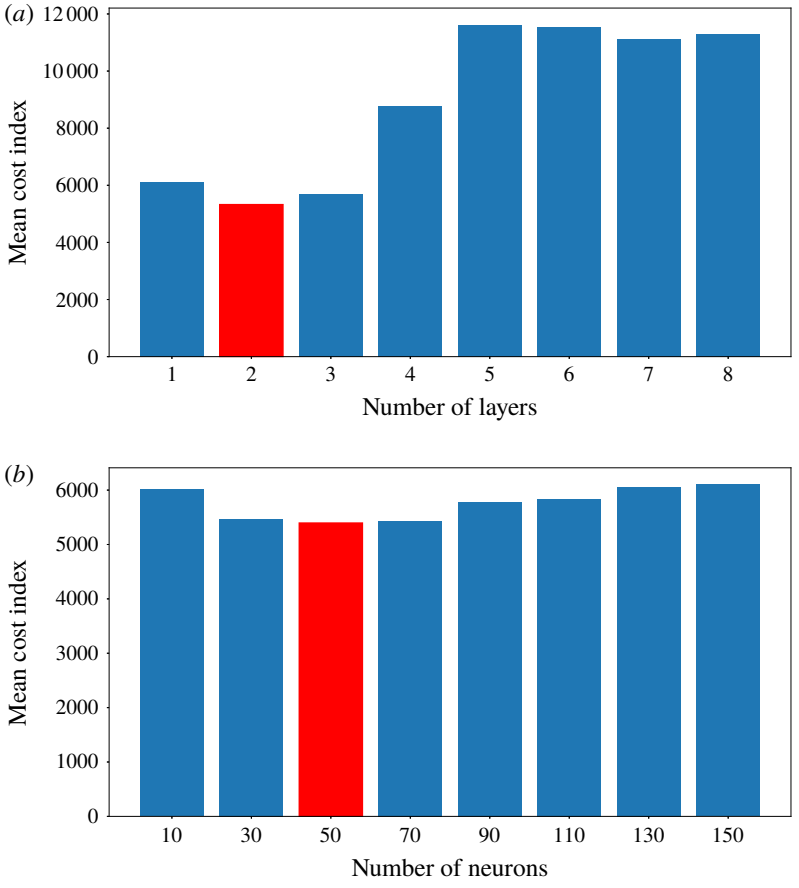

FIGURE 16. (Colour online) Quantification of hyper-parameter optimization shown for (*a*) number of layers and (*b*) number of neurons. An optimal network architecture of two layers and 50 neurons is chosen for our study.

efficient optimization of the network architecture by bypassing the problems of vanishing gradients inherent in sigmoidal activation functions (Ling *et al.* 2016).

For the purpose of optimal network architecture selection, we utilize a grid-search selection coupled with a three-fold cross-validation implemented in the open-source library Scikit-learn. In essence, a parameter space given by a grid is coupled with three trainings, tests and validations for each network through three partitions of the total training data. We first undertake our aforementioned optimization for the number of layers by utilizing a total of 1000 epochs for determining the optimal depth of the network. Each network with a particular choice of the number of layers (ranging between one and eight) is optimized three times using a three-fold cross-validation strategy and utilized for prediction on the test and validation partitions not used for weight optimization. The three networks for each hyper-parameter are then assigned a mean cost-function score, which is used for selection of the final model depth. We observe that a two-layer model outperforms other alternatives during this grid search as shown in figure 16. We note that the number of neurons in this first grid search is fixed at 50 although similar trends are recovered with varying specifications

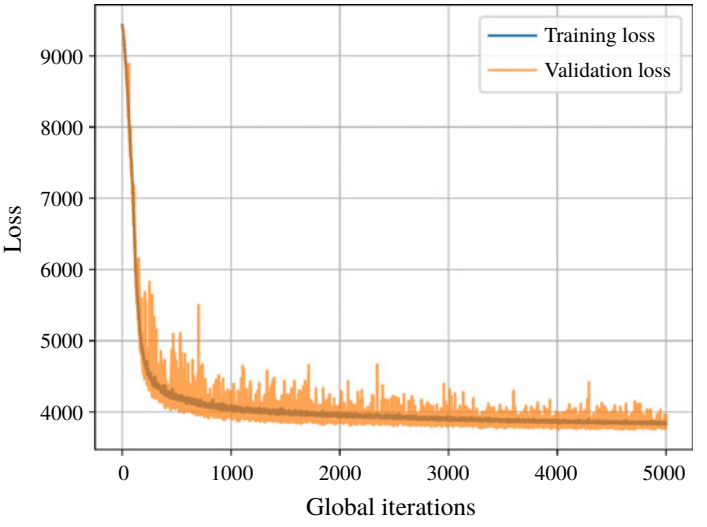

FIGURE 17. (Colour online) Learning rate of the proposed optimal model architecture. Note how training and validation loss are correlated closely for this learning problem.

between 10 and 100. Our mean cost index is given by the following expression for each location on the grid:

$$\text{mean cost index} = \frac{1}{K}\sum_{i=1}^{K} \|\Pi_K^{true} - \tilde{\Pi}_K\|_2, \tag{A 1}$$

where K refers to the training fold chosen for gradient calculation in the back-propagation within the same dataset.

A second grid search is performed with a fixed number of layers (i.e. two obtained from the previous tuning) and with a varying number of neurons. The results of this optimization are observed in figure 16, which shows that an optimal number of neurons of 50 suffice for this training. We note, however, that the choice for the number of neurons in the two-layer network does not affect the tuning score significantly. We clarify here that the model optimization may have been carried out using a multidimensional grid search for the optimal hyper-parameters or through sampling in a certain probability distribution space; however, our approach was formulated out of a desire to reduce offline training cost as much as possible. The final network was then selected for a longer duration of training (5000 epochs) till the learning rate is minimal as shown in figure 17. Details of our network optimization and dataset generation are provided in appendix B.

Appendix B. Network training

For the purpose of generating an optimal map discussed in the previous section, we utilize a supervised learning with sets of labelled inputs and outputs obtained from direct numerical simulation (DNS) data for two-dimensional turbulence (San & Staples 2012; Maulik & San 2017*b*). Our grid-resolved variables (which, we remind the reader, are denoted as overbarred quantities) are generated by a Fourier cutoff filter so as to truncate the fully resolved DNS fields (obtained at 2048^2 degrees of freedom)

to coarse-grained grid level (i.e. given by 256^2 degrees of freedom). Therefore, this procedure is utilized to generate input–output pairs for the process of training our ANN map. We also emphasize the fact that, while the DNS data generated multiple time snapshots of flow evolution, data were harvested from times $t = 0$, 1, 2, 3 and 4 for the purpose of training and validation. This represents a stringent subsampling of the total available data for map optimization. To quantify this subsampling, we note that we had potential access to 40 000 space–time snapshots of DNS data, out of which only five were chosen for training and validation data generation (0.0125 % of total data). We also note that the Reynolds number chosen for generating the training and validation datasets is given by $Re = 32\,000$ alone.

Two-thirds of the total dataset generated for optimization was utilized for training and the rest was utilized for validation assessment. Here, training refers to the use of data for loss calculation (which in this study is a classical mean-squared error) and back-propagation for parameter update. Validation was utilized to record the performance of the trained network on data to which it was not exposed during training. Similar behaviour in training and validation loss would imply a well-formulated learning problem. The final ANN (obtained post-training) would be selected according to the best validation loss after a desired number of iterations, which for this study was fixed at 5000. We also note that the error minimization in the training of the ANN utilized the Adam optimizer (Kingma & Ba 2014) implemented in the open-source ANN training platform TensorFlow. Figure 17 shows the learning rate of the proposed framework with very similar behaviour between training and validation loss, implying a successfully optimized map. We remark that, while the network may have learned the map from the data with which it has been provided for training and validation, testing would require an *a posteriori* examination, as detailed in § 4.

REFERENCES

ARAKAWA, A. 1966 Computational design for long-term numerical integration of the equations of fluid motion: two-dimensional incompressible flow. Part I. *J. Comput. Phys.* **1** (1), 119–143.

BECK, A. D., FLAD, D. G. & MUNZ, C.-D. 2018 Neural networks for data-based turbulence models. arXiv:1806.04482.

BERSELLI, L. C., ILIESCU, T. & LAYTON, W. J. 2005 *Mathematics of Large Eddy Simulation of Turbulent Flows*. Springer.

CANUTO, V. M. & CHENG, Y. 1997 Determination of the Smagorinsky–Lilly constant C_S. *Phys. Fluids* **9** (5), 1368–1378.

COHEN, K., SIEGEL, S., MCLAUGHLIN, T. & GILLIES, E. 2003 Feedback control of a cylinder wake low-dimensional model. *AIAA J.* **41** (7), 1389–1391.

CUSHMAN-ROISIN, B. & BECKERS, J.-M. 2011 *Introduction to Geophysical Fluid Dynamics: Physical and Numerical Aspects*, vol. 101. Academic.

DURAISAMY, K., IACCARINO, G. & XIAO, H. 2018 Turbulence modeling in the age of data. arXiv:1804.00183.

EDEN, C. & GREATBATCH, R. J. 2008 Towards a mesoscale eddy closure. *Ocean Model.* **20** (3), 223–239.

FALLER, W. E. & SCHRECK, S. J. 1997 Unsteady fluid mechanics applications of neural networks. *J. Aircraft* **34** (1), 48–55.

FOX-KEMPER, B., DANABASOGLU, G., FERRARI, R., GRIFFIES, S. M., HALLBERG, R. W., HOLLAND, M. M., MALTRUD, M. E., PEACOCK, S. & SAMUELS, B. L. 2011 Parameterization of mixed layer eddies. III. Implementation and impact in global ocean climate simulations. *Ocean Model.* **39** (1–2), 61–78.

FREDERIKSEN, J. S., O'KANE, T. J. & ZIDIKHERI, M. J. 2013 Subgrid modelling for geophysical flows. *Phil. Trans. R. Soc. Lond.* A **371** (1982), 20120166.

FREDERIKSEN, J. S. & ZIDIKHERI, M. J. 2016 Theoretical comparison of subgrid turbulence in atmospheric and oceanic quasi-geostrophic models. *Nonlinear Process. Geophys.* **23** (2), 95–105.

GALPERIN, B. & ORSZAG, S. A. 1993 *Large Eddy Simulation of Complex Engineering and Geophysical Flows*. Cambridge University Press.

GAMAHARA, M. & HATTORI, Y. 2017 Searching for turbulence models by artificial neural network. *Phys. Rev. Fluids* **2** (5), 054604.

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

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.

KING, R. N., HAMLINGTON, P. E. & DAHM, W. J. 2016 Autonomic closure for turbulence simulations. *Phys. Rev.* E **93** (3), 031301.

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

KRAICHNAN, R. H. 1967 Inertial ranges in two-dimensional turbulence. *Phys. Fluids* **10** (7), 1417–1423.

KUTZ, J. N. 2017 Deep learning in fluid dynamics. *J. Fluid Mech.* **814**, 1–4.

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

LEITH, C. E. 1968 Diffusion approximation for two-dimensional turbulence. *Phys. Fluids* **11** (3), 671–672.

LING, J., KURZAWSKI, A. & TEMPLETON, J. 2016 Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. *J. Fluid Mech.* **807**, 155–166.

LING, J. & TEMPLETON, J. 2015 Evaluation of machine learning algorithms for prediction of regions of high Reynolds averaged Navier–Stokes uncertainty. *Phys. Fluids* **27** (8), 085103.

MANNARINO, A. & MANTEGAZZA, P. 2014 Nonlinear aeroelastic reduced order modeling by recurrent neural networks. *J. Fluid. Struct.* **48**, 103–121.

MANSFIELD, J. R., KNIO, O. M. & MENEVEAU, C. 1998 A dynamic LES scheme for the vorticity transport equation: formulation and *a priori* tests. *J. Comput. Phys.* **145** (2), 693–730.

MARSHALL, J. S. & BENINATI, M. L. 2003 Analysis of subgrid-scale torque for large-eddy simulation of turbulence. *AIAA J.* **41** (10), 1875–1881.

MAULIK, R. & SAN, O. 2017*a* A neural network approach for the blind deconvolution of turbulent flows. *J. Fluid Mech.* **831**, 151–181.

MAULIK, R. & SAN, O. 2017*b* A stable and scale-aware dynamic modeling framework for subgrid-scale parameterizations of two-dimensional turbulence. *Comput. Fluids* **158**, 11–38.

MILANO, M. & KOUMOUTSAKOS, P. 2002 Neural network modeling for near wall turbulent flow. *J. Comput. Phys.* **182** (1), 1–26.

MOHAN, A. T. & GAITONDE, D. V. 2018 A deep learning based approach to reduced order modeling for turbulent flow control using LSTM neural networks. arXiv:1804.09269.

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.

PARISH, E. J. & DURAISAMY, K. 2016 A paradigm for data-driven predictive modeling using field inversion and machine learning. *J. Comput. Phys.* **305**, 758–774.

PATHAK, J., WIKNER, A., FUSSELL, R., CHANDRA, S., HUNT, B. R., GIRVAN, M. & OTT, E. 2018 Hybrid forecasting of chaotic processes: using machine learning in conjunction with a knowledge-based model. *Chaos* **28** (4), 041101.

PIOMELLI, U., CABOT, W. H., MOIN, P. & LEE, S. 1991 Subgrid-scale backscatter in turbulent and transitional flows. *Phys. Fluids* **3** (7), 1766–1771.

RAISSI, M. & KARNIADAKIS, G. E. 2018 Hidden physics models: machine learning of nonlinear partial differential equations. *J. Comput. Phys.* **357**, 125–141.

SAGAUT, P. 2006 *Large Eddy Simulation for Incompressible Flows: An Introduction*. Springer.

SAN, O. & MAULIK, R. 2018 Neural network closures for nonlinear model order reduction. *Adv. Comput. Math.*; doi:10.1007/s10444-018-9590-z.

SAN, O. & STAPLES, A. E. 2012 High-order methods for decaying two-dimensional homogeneous isotropic turbulence. *Comput. Fluids* **63**, 105–127.

SAN, O., STAPLES, A. E. & ILIESCU, T. 2013 Approximate deconvolution large eddy simulation of a stratified two-layer quasigeostrophic ocean model. *Ocean Model.* **63**, 1–20.

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

SCHAEFFER, H. 2017 Learning partial differential equations via data discovery and sparse optimization. *Proc. R. Soc. Lond.* A **473** (2197), 20160446.

SINGH, A. P., MEDIDA, S. & DURAISAMY, K. 2017 Machine-learning-augmented predictive modeling of turbulent separated flows over airfoils. *AIAA J.* **55** (7), 2215–2227.

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

TRACEY, B. D., DURAISAMY, K. & ALONSO, J. J. 2015 A machine learning strategy to assist turbulence model development. In *53rd AIAA Aerospace Sciences Meeting*; 5–9 January 2015, Paper no: 2015-1287, American Institute of Aeronautics and Astronautics SciTech Forum, Kissimmee, FL.

VOROBEV, A. & ZIKANOV, O. 2008 Smagorinsky constant in LES modeling of anisotropic MHD turbulence. *Theor. Comput. Fluid Dyn.* **22** (3–4), 317–325.

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

WAN, Z. Y., VLACHAS, P., KOUMOUTSAKOS, P. & SAPSIS, T. 2018 Data-assisted reduced-order modeling of extreme events in complex dynamical systems. *PloS One* **13** (5), e0197704.

WANG, J.-X., WU, J., LING, J., IACCARINO, G. & XIAO, H. 2017*a* A comprehensive physics-informed machine learning framework for predictive turbulence modeling. arXiv:1701.07102.

WANG, J.-X., WU, J.-L. & XIAO, H. 2017*b* Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data. *Phys. Rev. Fluids* **2** (3), 034603.

WEATHERITT, J. & SANDBERG, R. D. 2017 Hybrid Reynolds-averaged/large-eddy simulation methodology from symbolic regression: formulation and application. *AIAA J.* **55** (11), 3734–3746.

WU, J.-L., XIAO, H. & PATERSON, E. 2018*a* Data-driven augmentation of turbulence models with physics-informed machine learning. arXiv:1801.02762.

WU, J.-L., XIAO, H. & PATERSON, E. 2018*b* Physics-informed machine learning approach for augmenting turbulence models: a comprehensive framework. *Phys. Rev. Fluids* **3** (7), 074602.

XIAO, H., WU, J.-L., WANG, J.-X., SUN, R. & ROY, C. J. 2016 Quantifying and reducing model-form uncertainties in Reynolds-averaged Navier–Stokes simulations: a data-driven, physics-informed Bayesian approach. *J. Comput. Phys.* **324**, 115–136.