

A neural network approach for the blind deconvolution of turbulent flows

R. Maulik¹ and O. San^{1,†}

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

(Received 3 June 2017; revised 31 August 2017; accepted 1 September 2017)

We present a single-layer feed-forward artificial neural network architecture trained through a supervised learning approach for the deconvolution of flow variables from their coarse-grained computations such as those encountered in large eddy simulations. We stress that the deconvolution procedure proposed in this investigation is blind, i.e. the deconvolved field is computed without any pre-existing information about the filtering procedure or kernel. This may be conceptually contrasted to the celebrated approximate deconvolution approaches where a filter shape is predefined for an iterative deconvolution process. We demonstrate that the proposed blind deconvolution network performs exceptionally well in the *a priori* testing of two-dimensional Kraichnan, three-dimensional Kolmogorov and compressible stratified turbulence test cases, and shows promise in forming the backbone of a physics-augmented data-driven closure for the Navier–Stokes equations.

Key words: computational methods, turbulence modelling

1. Introduction

In the past few decades, an exponential increase in computational power, algorithmic advances and experimental data collection strategies have seen an explosion in modelling efforts which leverage information obtained from physical data. The ability to learn from the physics of an experiment represents an attractive augmentation to the laws that have been painstakingly developed from the first principles of fluid mechanics (Schmidt & Lipson 2009). As highlighted by Kutz (2017), data-driven learning combined with well-established governing laws represent an extraordinary opportunity for the mitigation of the challenges faced by many modelling mechanisms that are purely developed from first principles.

The use of machine learning techniques have traditionally been limited to the flow control community (Lee *et al.* 1997; Milano & Koumoutsakos 2002; Gautier *et al.* 2015) but there has been recent interest in their utilization for subgrid-scale modelling for high-Reynolds-number (Re) flows. For instance, artificial neural networks (ANNs) have recently been utilized for the purpose of subgrid-scale model classification (Gamahara & Hattori 2017) and Reynolds stress anisotropy tensor prediction (Ling, Kurzawski & Templeton 2016b; Wang, Wu & Xiao 2017). Together with other

[†] Email address for correspondence: osan@okstate.edu

machine-learning-based techniques such as Gaussian process regression (Zhang & Duraisamy 2015; Raissi & Karniadakis 2016, 2017; Wang, Wu & Xiao 2016), field inversion (Duraisamy, Zhang & Singh 2015; Parish & Duraisamy 2016), symbolic regression (Brunton, Proctor & Kutz 2016; Weatheritt & Sandberg 2016) and several different algorithms (Bright, Lin & Kutz 2013; Tracey, Duraisamy & Alonso 2013; Ling & Templeton 2015; Ling, Jones & Templeton 2016a), data-driven approaches such as ANNs are poised to form the backbone of the next great leap in closure modelling for nonlinear conservation laws.

A chief motivation for this research lies in the recent advances in the image processing community which exploit the general body of machine learning techniques used for the reconstruction of noisy or blurred images (Cichocki & Amari 2002). In particular we endeavour to justify the implementation of ANN-based machine learning strategies for the purpose of subfilter-scale recovery in turbulence closure modelling. In this work, we have developed a single-layer feed-forward ANN to identify a nonlinear relationship between the low-pass spatially filtered and coarse-grained (but unfiltered) field variables for settings in two-dimensional (2D) and three-dimensional (3D) homogeneous isotropic turbulence as well as a stratified turbulence case exhibiting moderate compressibility in the limit of infinite Reynolds numbers. A fundamental basis for the selection of a single hidden layer is to reduce training durations and to constrain the degrees of freedom of our trained network (by reducing the number of weights). We emphasize that the performance demanded from our ANN is statistical in nature and generalization is vital. Our trained ANN, once deployed, performs a data recovery procedure utilizing the statistical relationship between filtered and true data. Our argument for the choices presented above are given by the fact that a successful implementation of the proposed idea would connect potential applications across many different time and length scales.

The homogeneous and isotropic turbulence framework is chosen from the point of view of a well-established understanding of the energy cascade mechanism in the inertial range. In essence, the statistical recovery of scaling behaviour in the inertial range shall represent our contribution from the first principles of fluid mechanics. The compressible stratified turbulence test case serves a dual purpose. Not only is it utilized to evaluate the performance of our proposed architecture in a perfectly *a priori* fashion, but it is also used for testing the suitability of the trained ANN for deconvolution across different flow physics. To that end, we generate training data sets from one of the 3D homogeneous isotropic or compressible stratified turbulence simulations and test it on the other. This may be considered to be more challenging than a purely *a priori* analysis and would expose the universal nature of the training in our data-driven framework. These tests reveal further information related to the suitability of the data-driven closure across different physical regimes.

In addition to the various subfilter recovery analyses outlined above, the ability of our proposed framework to stabilize aliasing error is also studied for the different test cases. For this purpose, we investigate the predictive performance of the proposed procedure by solving a noisy data recovery problem in each setting. This noisy data recovery problem requires the estimation of a relationship between inputs which are perturbed by noise and their true counterparts to examine its suitability for coarse-grained large eddy simulations which often exhibit severe aliasing error. Besides the cross-validation within the framework of the different test cases for this noisy data problem, we also utilize training and testing data from different flow physics (for instance training between homogeneous isotropic and stratified turbulence test cases) to test the universality of this regularization behaviour.

In particular, the approach outlined in our study may be considered to be analogous to the approximate deconvolution methodology (Stolz & Adams 1999) to recover subfilter contributions of low-pass spatially filtered flow fields. The chief difference is the lack of assumption of any filtering kernel (Gaussian or otherwise); a fact that necessitates the utilization of additional regularization in case of flows which may exhibit distinctly non-Gaussian distributions (Stolz, Adams & Kleiser 2001). A potential advantage to emphasize here is that a data-driven closure could be incorporated to learn the nonlinear deconvolution procedure for challenging flows. This could, conceptually, aid in a significant increase of the accuracy of reconstructed subfilter contributions to the turbulence shear stresses. We propose this hypothesis since it has been rigorously established that single and multilayered neural networks are capable of acting as universal function approximators (Hornik, Stinchcombe & White 1989) and could represent a continuous underlying ‘natural’ low-pass spatial filter shape effectively. We validate our hypothesis by testing the ability of the trained ANN to capture inertial length scales accurately (i.e., while respecting the Kolmogorov (1941) and Kraichnan (1967) scaling laws) and by examining the probability density functions of the true, distorted (through low-pass spatial filtering or noise) and reconstructed field variables. Apart from this, our approach is also cross-examined with state-of-the-art structural closures which require an explicitly specified low-pass spatial filter (Layton & Rebholz 2012). A cross-validation strategy is also detailed and performed to ensure that the performance of the proposed framework is physics-based and not an artefact of data-localization.

2. Artificial neural networks

2.1. Architecture

The basic structure of the simple feed-forward ANN consists of layers possessing a predefined number of unit cells called neurons. Each of these layers has an associated transfer function and each unit cell has an associated bias. Any input to the neuron has a bias added to it followed by activation through the transfer function. To describe this process using matrix operations, we have for a single neuron in the l th layer receiving a vector of inputs \mathbf{S}^l from the $(l-1)$ th layer given by (Demuth *et al.* 2014)

$$\mathbf{S}^l = \mathbf{W}^l \mathbf{X}^{l-1}, \quad (2.1)$$

where \mathbf{W}^l stands for a matrix of weights linking the $l-1$ and l layers with \mathbf{X}^{l-1} being the output of the $(l-1)$ th layer. The output of the l th layer is now given by

$$\mathbf{X}^l = f(\mathbf{S}^l + \mathbf{B}^l), \quad (2.2)$$

where \mathbf{B}^l is the vector of biasing parameters for the l th layer. Every node (or unit cell) has an associated transfer function which acts on its input and bias to produce an output which is fed forward through the network. The nodes which take the raw input value of our training data set (i.e., the nodes of the first layer in the network) perform no computation as they do not have any biasing or activation through a transfer function. The output layers generally have a linear activation function with a bias which implies a simple summation of inputs incident to a unit cell with its associated bias. In this investigation, we have used one hidden layer of neurons between the set of inputs and targets with a Tan-Sigmoid activation function which can be expressed as

$$f(a) = \frac{1 - e^{-2a}}{1 + e^{-2a}}, \quad (2.3)$$

where the transfer function f calculates the neuron's output given its net input. In theory, any differentiable function can qualify as an activation function (Zhang, Patuwo & Hu 1998); however, only a small number of functions which are bounded, monotonically increasing and differentiable are used for this purpose.

2.2. Extreme learning machine

In this section we detail the extreme learning machine (ELM) approach to generalized single-layer feed-forward ANN training. This methodology was proposed in Huang, Zhu & Siew (2004) for extremely fast training of a single-layer feed-forward ANN based on the principles of the least-squares approximation. We would like to note that the ELM training approach has previously been utilized for image deblurring but using the principles of classification in a classical deep convolutional network as against the principles of regression utilized here (Wang *et al.* 2011). For the ease of description, let us define a few matrices for the single-layer feed-forward network. This is a generalization of the architecture introduced in §2.1. Our input matrix is

$$\mathbf{X}^0 = [\mathbf{x}_1^0 \quad \mathbf{x}_2^0 \quad \dots \quad \mathbf{x}_{n_s}^0], \quad (2.4)$$

where \mathbf{x}_i^0 is the i th sample (out of a total of n_s samples) of a multidimensional input column vector. Our weights connecting the inputs to the middle (hidden) layer and the biases associated with the hidden layer are given by

$$\mathbf{W}^1 = \begin{bmatrix} \mathbf{w}_1^1 \\ \mathbf{w}_2^1 \\ \vdots \\ \mathbf{w}_Q^1 \end{bmatrix}, \quad \mathbf{B}^1 = \begin{bmatrix} \mathbf{b}_1^1 \\ \mathbf{b}_2^1 \\ \vdots \\ \mathbf{b}_Q^1 \end{bmatrix}, \quad (2.5)$$

where \mathbf{w}_i^1 and \mathbf{b}_i^1 represent row vectors corresponding to the weights and biases related to the i th neuron (out of a total of Q neurons) in the hidden layer. These are initialized to be small non-zero random numbers to enforce generalization. It is important to remark here that the extreme learning machine methodology prescribes biases only for hidden layer neurons (i.e., the output layer biases $\mathbf{B}^2 = 0$). The output of the hidden layer neurons becomes

$$\mathbf{H}^T = f(\mathbf{W}^1 \mathbf{X}^0 + \mathbf{B}^1), \quad (2.6)$$

where $f(\mathbf{X})$ implies a Tan-Sigmoid activation procedure on each element of a matrix \mathbf{X} and the superscript T implies a matrix transpose. The weight matrix of the second layer may be given as

$$\mathbf{W}^2 = [\mathbf{w}_1^2 \quad \mathbf{w}_2^2 \quad \dots \quad \mathbf{w}_Q^2], \quad (2.7)$$

where \mathbf{w}_i^2 now refers to the column vectors related to the hidden layer neurons. Our outputs of the ELM may thus be represented as

$$\mathbf{S}^2 = \mathbf{W}^2 \mathbf{H}^T, \quad (2.8)$$

as there is a linear activation in the outer layer with no bias. This output must be trained against a set of targets corresponding to each input vector given by

$$\mathbf{T} = [\mathbf{t}_1 \quad \mathbf{t}_2 \quad \dots \quad \mathbf{t}_{n_s}], \quad (2.9)$$

where \mathbf{t}_i is a column vector of the targets corresponding to the i th sample. The ELM training mechanism is given as follows. In order to calculate the matrix \mathbf{W}^2 , we must recognize that its optimal solution should satisfy

$$\mathbf{W}_{op}^2 \mathbf{H}^\top = \mathbf{T}, \quad (2.10)$$

or by taking a transpose of both sides

$$\mathbf{H} \mathbf{W}_{op}^{2\top} = \mathbf{T}^\top, \quad (2.11)$$

which leads us to the following expression for the optimal weights

$$\mathbf{W}_{op}^{2\top} = \mathbf{H}^\dagger \mathbf{T}^\top. \quad (2.12)$$

The weights \mathbf{W}^1 and biases \mathbf{B}^1 are generated initially using small random numbers. The utilization of small random numbers promotes generalization and prevents overfitting (Demuth *et al.* 2014). The matrix given by \mathbf{H}^\dagger is calculated using a generalized Moore–Penrose pseudoinverse (Albert 1972; Serre 2002). Once the optimal weights of the second layer are obtained, our network is trained for deployment. Due to the random number values chosen for the weights in the first layer, our network is well suited to highly effective generalization of the training process due to a smaller degree of freedom of the overall ANN. The proposed direct training procedure substitutes the gradient-based iterative optimization algorithms such as the Bayesian regularization method (MacKay 1992; Foresee & Hagan 1997) in favour of the calculation of a pseudoinverse for the least-squares solution of optimal weights. This gives us the ability to obtain generalized predictive networks through training times which are several orders of magnitude lower in comparison with standard iterative techniques (Huang, Zhu & Siew 2006). Algorithms 1 and 2 describe the training and deployment process for the single-layer ANN trained using ELM. Effectively, Step 4 in Algorithm 1 represents the computational expense related to our supervised training.

Algorithm 1 Extreme Learning Machine: Training

- | | |
|--|--|
| 1: Given \mathbf{X}^0 and \mathbf{T} | ▷ Given inputs and targets |
| 2: Initialize \mathbf{W}^1 and \mathbf{B}^1 | ▷ Initialize non-zero random parameters |
| 3: Calculate $\mathbf{H}^\top = f(\mathbf{W}^1 \mathbf{X}^0 + \mathbf{B}^1)$ | ▷ Tan-Sigmoid activation |
| 4: Calculate pseudoinverse \mathbf{H}^\dagger | ▷ Moore–Penrose pseudoinverse |
| 5: Calculate $\mathbf{W}_{op}^{2\top} = \mathbf{H}^\dagger \mathbf{T}^\top$ | ▷ Least-squares solution for optimal weights |
-

Algorithm 2 Extreme Learning Machine: Deployment

- | | |
|--|---|
| 1: Given \mathbf{X}_{test}^0 | ▷ Given testing data |
| 2: Given \mathbf{W}^1 , \mathbf{B}^1 and \mathbf{W}_{op}^2 | ▷ Given trained network from Algorithm 1 |
| 3: Obtain $\mathbf{S}^1 = \mathbf{W}^1 \mathbf{X}_{test}^0$ | ▷ Calculate input at Layer 1 neurons |
| 4: Obtain $\mathbf{X}^1 = f(\mathbf{S}^1 + \mathbf{B}^1)$ | ▷ Activate to obtain output of Layer 1 neurons |
| 5: Obtain $\mathbf{X}^2 = \mathbf{W}_{op}^2 \mathbf{X}^1$ | ▷ Obtain output of the second layer: prediction |
-

3. Blind deconvolution

The primary motivation for the development of the aforementioned training and deployment procedures for our ANN architecture is described in this section. We are motivated by the approximate deconvolution framework (Stolz & Adams 1999), which uses an iterative resubstitution procedure known as the Van Cittert iterations for reconstructing the contribution of subfilter-scale content. This procedure may be represented tensorially as:

$$\tau_{ij} = \bar{u}_i \bar{u}_j - \overline{u_i u_j}, \quad (3.1)$$

where we obtain an approximation of these subfilter stresses as (Germano 2015)

$$\tau_{ij} = \bar{u}_i \bar{u}_j - \overline{u_i^* u_j^*}, \quad (3.2)$$

in which the asterisk superscript indicates an approximately deconvolved variable. An important caveat of this mechanism is a user-defined low-pass spatial filtering kernel. While the general approach is to implement various versions of Gaussian blur kernel, it is quite possible that complex flow configurations may actually exhibit natural filter shapes which are much more contorted. Fortunately, the vast wealth of data that may be collected by modern experimental techniques represents an opportunity for us to bypass this user-defined heuristic stage. In essence, blind deconvolution refers to the estimation of the underlying blur kernel without any knowledge or closed form estimate of its true nature. We must emphasize here that while there exist several algorithms that are truly blind in nature (for instance in Dabov *et al.* (2007)), this investigation proposes a framework which is purely data-driven. To summarize, our data may be considered to be an aid in implicitly estimating the shape of the blur kernel through the form of a trained neural network.

Our supervised learning framework requires training data which encapsulates the relationships between the coarsened true data and its perturbed versions. Note that, as is common in data-driven strategies, we normalize all our data to a range between -1 and 1 . This choice of the normalized data range is due to the fact that the Tan-Sigmoid activation function provides outputs between these limits as well. The perturbation may be introduced through a low-pass spatial filtering procedure:

$$u'_i = G(\mathbf{x}; \sigma) u_i, \quad (3.3)$$

which utilizes the following Gaussian kernel with standard deviation σ (held at a default value of 1.0):

$$G(\mathbf{x}; \sigma) = \frac{1}{(\sqrt{2\pi}\sigma)^d} \exp\left(-\frac{|\mathbf{x}|^2}{2\sigma^2}\right), \quad (3.4)$$

where d is the number of dimensions of the field being blurred. The use of a low-pass spatially filtered field tests the ability of the proposed ANN architecture for inertial range recovery and becomes a true test of the blind deconvolution ability of a neural network.

Another perturbation which is examined in this investigation is the addition of a normally distributed noise with an amplitude on the order of a tenth of the maximum value of the field variable which can be expressed as:

$$u'_i = u_i + \mu \kappa, \quad (3.5)$$

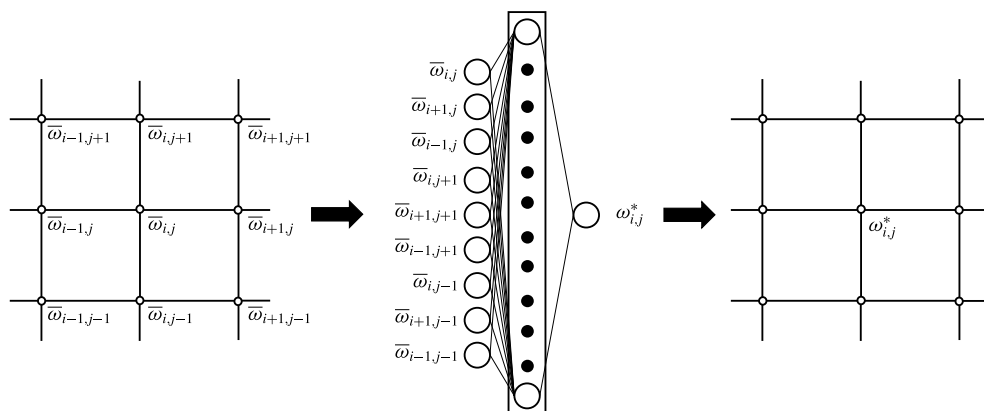


FIGURE 1. A schematic of the proposed blind deconvolution strategy for a two-dimensional test case where we aim to recover ω^* from $\bar{\omega}$. Note that each hidden neuron is associated with a unique bias.

where κ is a normally distributed random number between -1 and 1 and the coefficient μ (with a default value of 0.2) corresponds to the magnitude of the noise added to the field. The purpose of training relationships between inputs with noisy perturbations and the true solution is to test the stability of this approach for high-wavenumber energy accumulation. This is particularly important for coarse-grained large eddy simulations which are susceptible to numerical overflow. Thus, our supervised learning framework utilizes u'_i as our training inputs and u_i as our targets. Note that we train separate networks for each type of perturbation. The ELM training mechanism leads to tractable training times even for cases with several hundred neurons in the hidden layer. The results obtained in this study have employed 100 neurons for training and testing data obtained from different realizations of field values. A schematic of the blind deconvolution methodology integrated into the one-layer ELM-training-based ANN is shown for a two-dimensional test case in figure 1.

4. Cross-validation

An important assessment of a data-driven modelling approach is a cross-validation procedure to estimate the performance of the model for situations which it has not been trained for. In other words, it is necessary to ascertain that the data-driven model is not localized to the training data and provides a similar performance for different solution fields exhibiting similar physics but different numerical arrangements or magnitudes. Therefore, it is vital to test the performance of the trained ANN on a set of data which it has not seen previously (during training). For the purpose of data generation, we utilize high-fidelity simulations for our previously mentioned test cases so that their flow physics may be accurately resolved. Our 3D numerical experiments are generated using 512^3 degrees of freedom while our 2D case is generated using 2048^2 grid points. A coarse-grained large eddy simulation is mimicked through the subsequent subsampling of these high-quality data sets. Our coarse-graining procedure leaves us with a subsampled field of 64^3 and 256^2 degrees of freedom in the 3D and 2D test cases, respectively. Therefore, our coarse-graining procedure involves the selection of every eighth grid point in the high-fidelity uniform grid data.

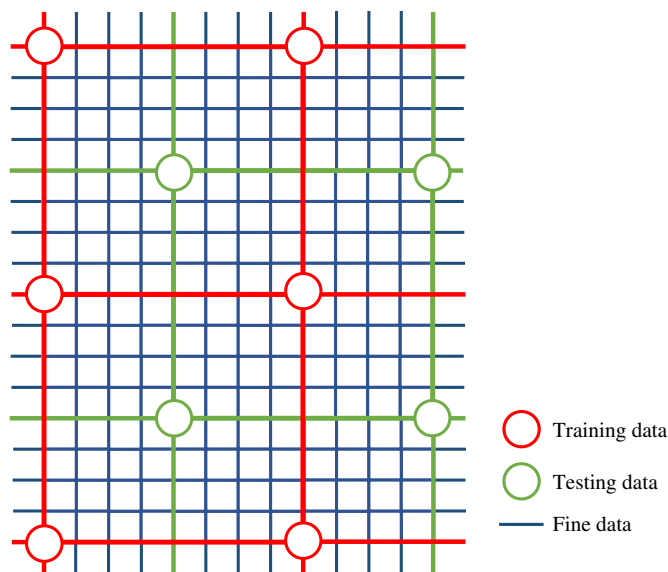


FIGURE 2. (Colour online) A schematic of the spatial shifting strategy for a simplified two-dimensional grid showing two different data sets. Several different data sets may be constructed using this technique which are numerically different physically similar.

We may devise several shifted data sets (each possessing 64^3 degrees of freedom in 3D and 256^2 degrees of freedom in 2D) but which continue to represent the physics of the fine-grained dataset. Missing data points at boundaries may be reconstructed through the use of the periodic boundary conditions for the given test cases. A simplified shifting schematic is shown in figure 2, where a simple technique of generating two coarse data sets from fine data is demonstrated. In a nutshell, the coarse-graining and shifting procedures allow us to devise up to 63 and 511 completely different data sets in two and three dimensions, respectively. For the purpose of cross-validation we randomly choose any four of these multiple data sets for the generation of three sets of testing data and one set of training data.

Once our four different randomly generated data sets are identified, we add perturbations corresponding to the type of behaviour demanded of the proposed artificial neural network. To test the deconvolution ability of our network, our training data (i.e., one of the four data sets) is filtered with an appropriate Gaussian smoothing (for our inputs to the network) and unfiltered training data are utilized as outputs to the network. Three testing data sets are generated in a similar manner, with one of these data sets being filtered with the same filter radius and the others being filtered with a 10% larger and 10% smaller filter radius, respectively. The trained ANN is then utilized to recover deconvolved approximations to the true field for these three test data sets. This ensures that the trends of the trained network are not due to overfitting or ‘data-memory’ but through an implicit learning of the inverse filtering. A similar procedure is also utilized to cross-validate the regularization ability of the closure wherein the randomly chosen data sets are perturbed through Gaussian noise. A table describing the filter radii and magnitudes of noise for our training and testing data sets is shown in table 1. Note that these data sets are all generated from the same high-fidelity solution field and correspond to a perfectly *a priori* analysis.

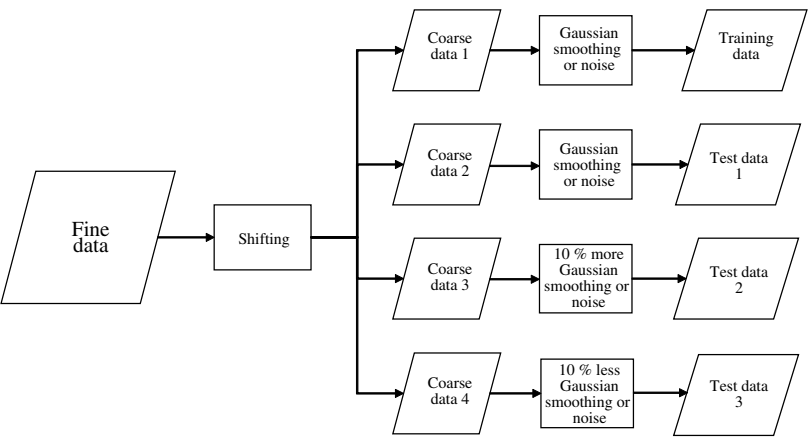


FIGURE 3. A flow chart detailing the generation of cross-validation data sets using the shifting operation as well as different perturbations. The network is trained using Algorithm 1 and is then tested on test data 1, 2 and 3.

Deconvolution		Regularization	
Data set	Filter radius (σ)	Data set	Noise (μ)
Training data	1.0	Training data	0.2
Test data 1	1.0	Test data 1	0.2
Test data 2	1.1	Test data 2	0.22
Test data 3	0.9	Test data 3	0.18

TABLE 1. Cross-validation data sets for the proposed data-driven blind deconvolution closure.

A concise summary of this cross-validation procedure is given in a flow chart in figure 3.

In addition to the *a priori* cross-validation outlined above, we also assess our proposed architecture by utilizing training and testing data across different flow physics. This procedure is carried out to test the universal nature of the learning for the purpose of both deconvolution and regularization abilities. This may be assumed to be another challenging tier for cross-validation. We elaborate this step in further detail in § 6.

5. Results

In this section, we outline the results of our proposed framework for three different benchmark flows showing distinct differences in physics. Our test cases are given by a two-dimensional homogeneous isotropic turbulent flow which exhibits a cascade of enstrophy from its integral to dissipative length scales, a canonical Taylor–Green decaying vortex corresponding to homogeneous isotropic turbulent flow with a cascade of kinetic energy from integral to dissipative length scales, and an inviscid three-dimensional turbulent flow developed from stratified Kelvin–Helmholtz instability.

5.1. Kraichnan turbulence

A periodic square domain is used to simulate a canonical flow configuration which exhibits the characteristics of decaying, homogeneous and isotropic 2D turbulence (San & Staples 2012). We utilize 2048^2 degrees of freedom for high-fidelity data, which is coarsened to 256^2 grid points to mimic a considerably coarser large eddy simulation. A Reynolds number of $Re = 32\,000$ is utilized to ensure a considerable scale separation and to display a prominent inertial range with its associated k^{-3} Kraichnan scaling (Kraichnan 1967). The 2D turbulence simulation is undertaken through the implementation of the Navier–Stokes equations in their vorticity-streamfunction formulation. The results shown here are obtained through the data-driven blind deconvolution of the vorticity. Our ANN closure was implemented by training a relationship between a nine-point stencil consisting of the point at which the deconvolution is desired along with its immediate neighbours (i.e., see figure 1). Training times were (on average) of the order of 0.01 s for approximately 65 000 data samples.

Figure 4 shows the performance of the proposed methodology in terms of statistical assessments given by angle-averaged kinetic energy spectra and probability density functions (PDF) of the vorticity field when the deconvolution aspect of the proposed framework is tested. For subfilter-scale reconstruction assessment, it can readily be observed that the proposed architecture manages to capture a far greater region of the inertial range in accordance with the k^{-3} scaling. This may also be observed from the PDF comparisons where the narrow distribution caused by the low-pass spatial filtering is successfully flattened to its original spread. A similar performance is exhibited when the testing data exhibits slightly different physics. As mentioned previously, this is simulated through the utilization of slightly larger and smaller filter radii to mimic physical behaviour in a local neighbouring range around the training data. The proposed framework successfully displays a similar performance for all the testing data versions, thereby demonstrating a good generalization. Field reconstruction attempts from noisy data display marginal benefits in terms of inertial range capture (as shown in figure 5) but the regularizing nature of the ELM training ensures that the cutoff length scale pile up of energy (or aliasing error) is successfully stabilized. This behaviour is also replicated for different testing data. As mentioned previously, this is very promising for coarse-grained large eddy simulations, particularly for non-conservative underlying numerical schemes.

Figures 6 and 7 visually represent the results of our proposed architecture. The enhancement of smaller features (due to a higher retention of the inertial range) in the case of the filtered inputs is clearly visible. On the other hand, a blurring effect can be discerned for the noisy perturbations which correspond to the stabilization effect of the framework. The distribution of normalized vorticity values seen in the PDF diagrams show a significant reconstruction of the true trends using an optimally trained ANN. A quantitative description of the performance of our data-driven deconvolution and regularization architecture for the Kraichnan turbulence test case is shown in table 2, where mean-squared-error values of the perturbed and reconstructed fields (with respect to the true field) are tabulated. It can be seen that the proposed architecture performs remarkably well for the reduction of mean-squared-error for all three testing datasets.

5.2. Kolmogorov turbulence

To further validate our claims about the potential of the proposed approach, the Taylor–Green vortex (TGV) problem was simulated in a periodic box 1, following

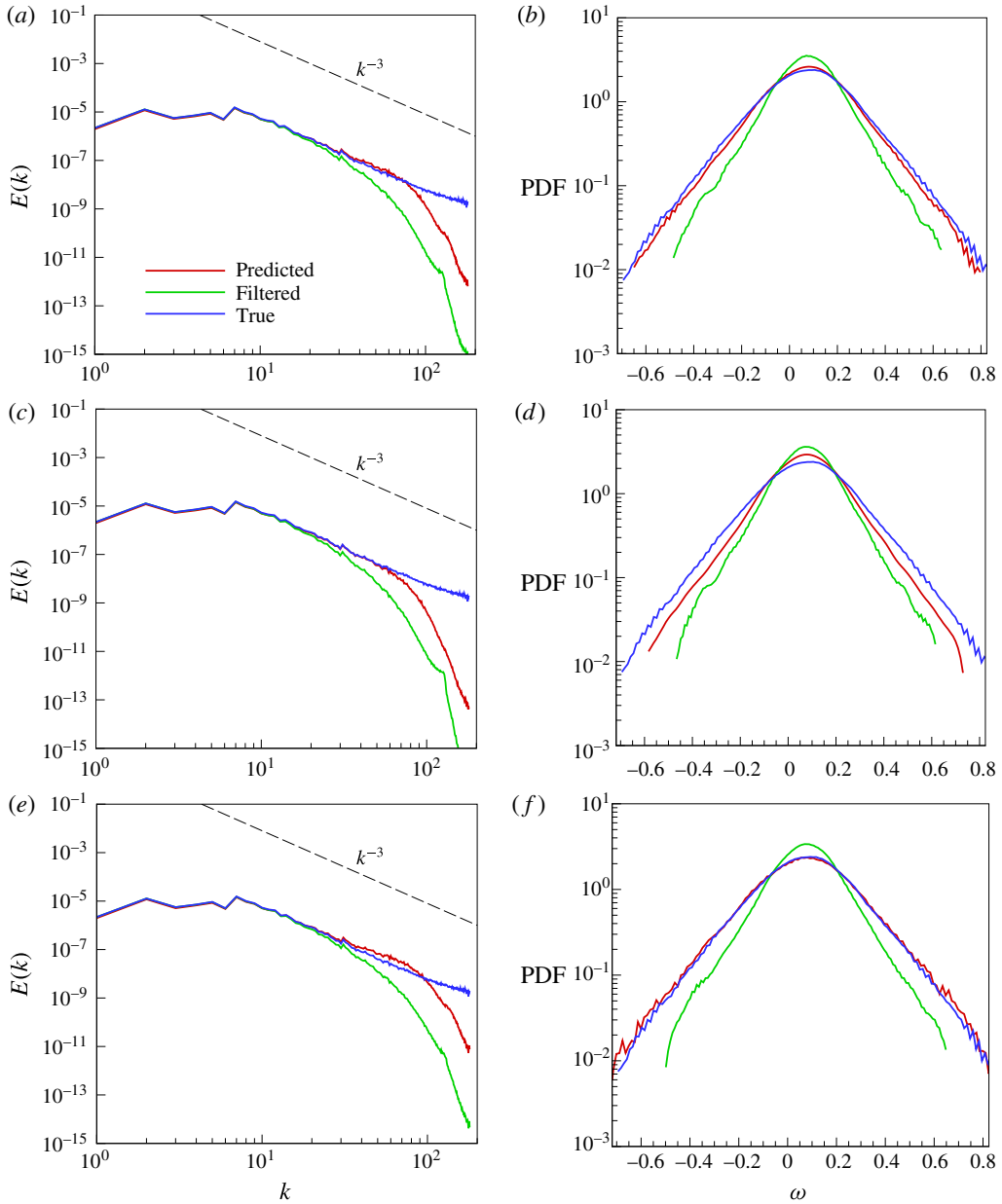


FIGURE 4. (Colour online) *A priori* results of the kinetic energy spectra (a,c,e) and PDF of the vorticity (b,d,f) for Kraichnan turbulence. Results for three different deconvolution test data sets shown. (a,b) Test data 1. (c,d) Test data 2. (e,f) Test data 3.

Bull & Jameson (2015), to exhibit the properties of decaying isotropic homogeneous 3D turbulence with the Kolmogorov scaling given by $k^{-5/3}$ (Frisch 1996). In a manner similar to the 2D turbulence test case, high-fidelity DNS data for $Re = 1600$ was generated on a uniform grid with 512^3 degrees of freedom from which coarser data at a resolution of 64^3 grid points was extracted for our studies. Three

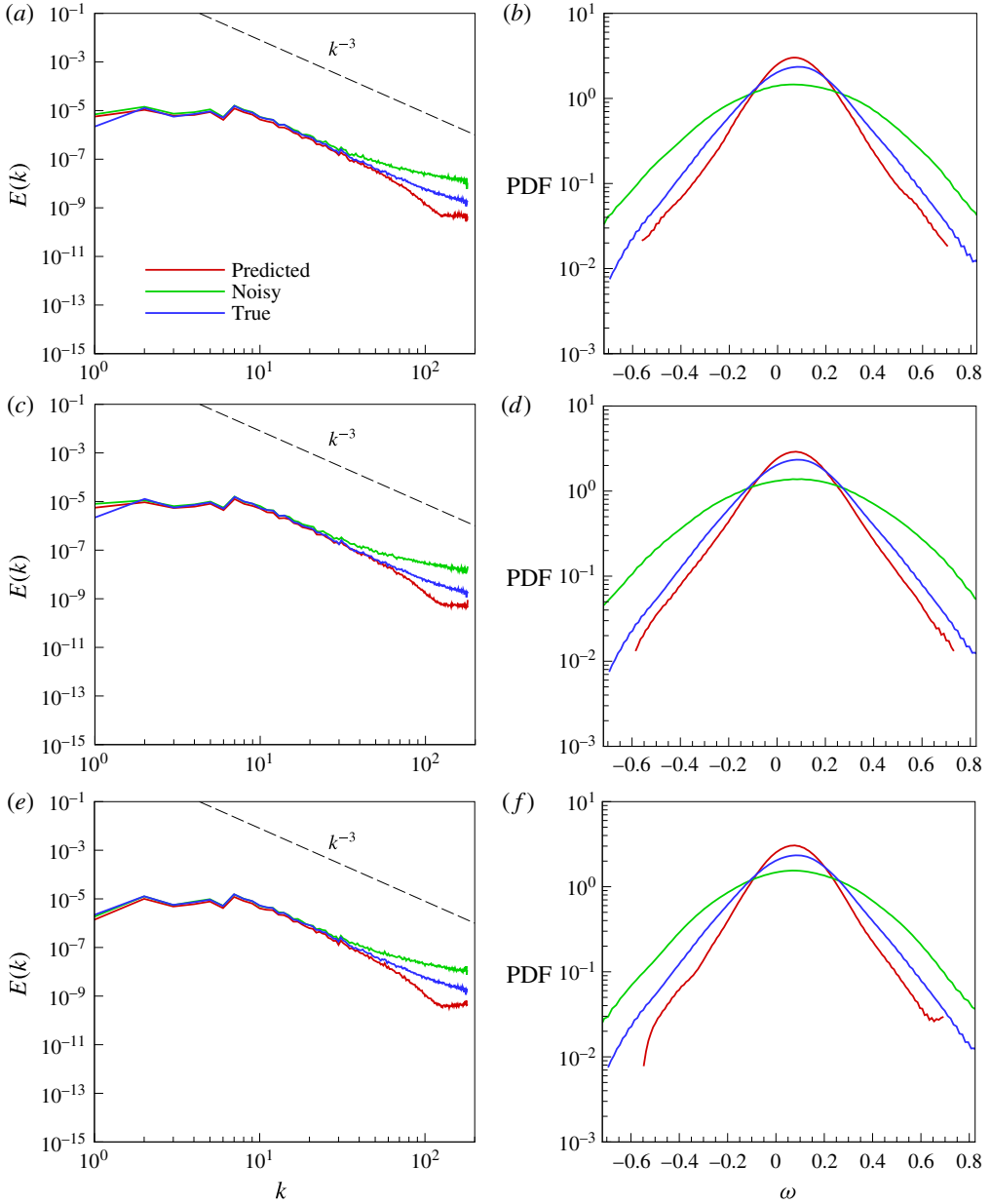


FIGURE 5. (Colour online) *A priori* results of the kinetic energy spectra (a,c,e) and PDF of the vorticity (b,d,f) for Kraichnan turbulence. Results for three different regularization test data sets shown. (a,b) Test data 1. (c,d) Test data 2. (e,f) Test data 3.

networks are trained for each velocity component and each network is trained using approximately 260 000 samples with inputs consisting of a stencil of 27 points (i.e., a three-dimensional version of the stencil described for Kraichnan turbulence) which represents a much larger computational task than the previous test case. However, the ELM approach is able to successfully train these three networks in under 10 seconds.

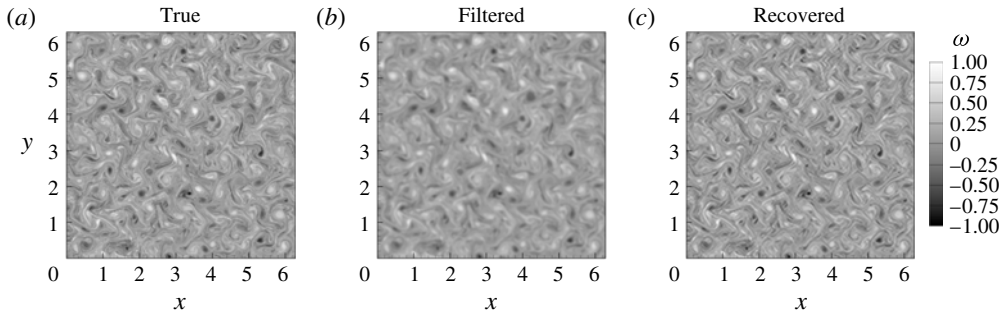


FIGURE 6. *A priori* results for vorticity recovery from low-pass spatially filtered inputs for Kraichnan turbulence. Data shown for deconvolution test data 1: (a) true coarse-grained fields, (b) coarse-grained fields with Gaussian smoothing, and (c) coarse-grained fields reconstructed using the proposed framework.

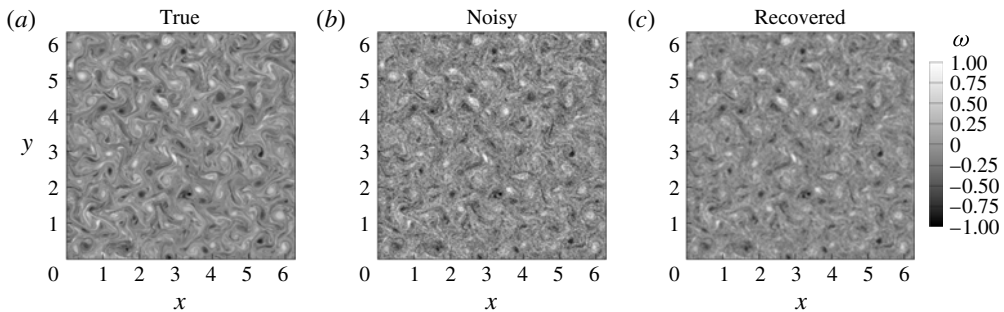


FIGURE 7. *A priori* results for vorticity recovery from noisy perturbation inputs for Kraichnan turbulence. Data shown for regularization test data 1: (a) true coarse-grained fields, (b) coarse-grained fields with Gaussian noise, and (c) coarse-grained fields reconstructed using the proposed framework.

Figure 8 demonstrates our statistical assessments for the low-pass spatially filtered test cases. As per our validation mechanism, we display results from three different versions of testing data where similar results are obtained qualitatively. Our proposed framework recovers a significant enhancement in the approximation of the true inertial range in the spherically averaged kinetic energy spectrum. It can also be seen that the peak and the tail of the true probability distribution function of the z component of velocity is captured quite accurately from the modified distribution obtained through low-pass spatial filtering. The PDF comparison between true and recovered fields is another indication of the ability of the proposed ANN architecture for reconstructing the true underlying trends from the filtered input data. It may be observed that the proposed architecture performs better for test data 3 as against test data 2 (although both trends are positive) since much more subfilter-scale information must be reconstructed than was trained for. This is possibly due to the data being rather nonlinear in physical space as against the linear parameter space we attempt to explore. Essentially, increasing the filter radius by 10% leads to some physical behaviour that the framework has not fully been exposed to in training. Addressing issues of predicting multiple perturbations using a singly trained network will require in-depth investigation of sampling strategies. Vitally, the behaviour exhibited by our

Kraichnan turbulence		
Deconvolution		
Dataset	Filtered	Deconvolved
Test data 1	1.02×10^{-2}	4.38×10^{-3}
Test data 2	1.14×10^{-2}	5.73×10^{-3}
Test data 3	9.11×10^{-3}	3.71×10^{-3}
Regularization		
Dataset	Noisy	Regularized
Test data 1	4.03×10^{-2}	1.26×10^{-2}
Test data 2	4.88×10^{-2}	1.38×10^{-2}
Test data 3	3.25×10^{-2}	1.14×10^{-2}

TABLE 2. Mean-squared-error values for deconvolved and regularized fields obtained from the proposed architecture. Data shown from the two-dimensional Kraichnan turbulence test case. Note that the mean-squared-error values are obtained from the vorticity magnitudes of the field.

proposed framework is consistent with our observations for the Kraichnan turbulence test case.

When trained to estimate the relationship between noisy inputs and true outputs, the proposed architecture demonstrates an exceptional generalization ability through the reduction of high-wavenumber content as shown in figure 9. While inertial range enhancement is once again marginal, a general improvement in the statistical trends of the recovered variable is observed through the PDF, where the peaks of the true PDF are recovered accurately. However, it is seen that an accurate capture of the tails of the PDF of the true field remains elusive for this particular architecture. Studies are underway to determine if the architecture of the ANN may aid in improving this inaccuracy since a larger number of neurons may offset the rather stringent generalization behaviour of its current implementation. This statistical generalization ability is also observed in the 2D test case, where tails of the PDF tend to be captured with lower accuracy.

We clarify that the trends in the PDF of other primitive variables are similar. The blind deconvolution ability of the proposed methodology is shown in figure 10, where one can clearly observe the presence of finer structures in the reconstructed field. Also, an examination of the noisy inputs in figure 11 show a remarkable reconstruction of data. It is apparent that the proposed approach proves adept at a partial deconvolution of the field when trained for the relationship using filtered inputs. This represents a viable opportunity for integrating data-driven closures into mainstream LES methodologies in combination with knowledge from first principles. In fact, it may even be possible to devise turbulence closures where phenomenological or heuristic arguments can be eschewed completely in favour of data-driven subfilter-scale estimations. This holds promise for inhomogeneous flows where distinct deviations are observed from Gaussian statistics in terms of the location of the integral length scales. We may effectively let experimental data determine the shape of a particular filter in a particular region of the flow. Machine-learning-inspired blind deconvolution may thus hold the key to a greater enhancement of the uniformity of an applicable turbulence closure. Table 3 offers a quantitative insight into the performance of the proposed closure, where it is once again observed that the

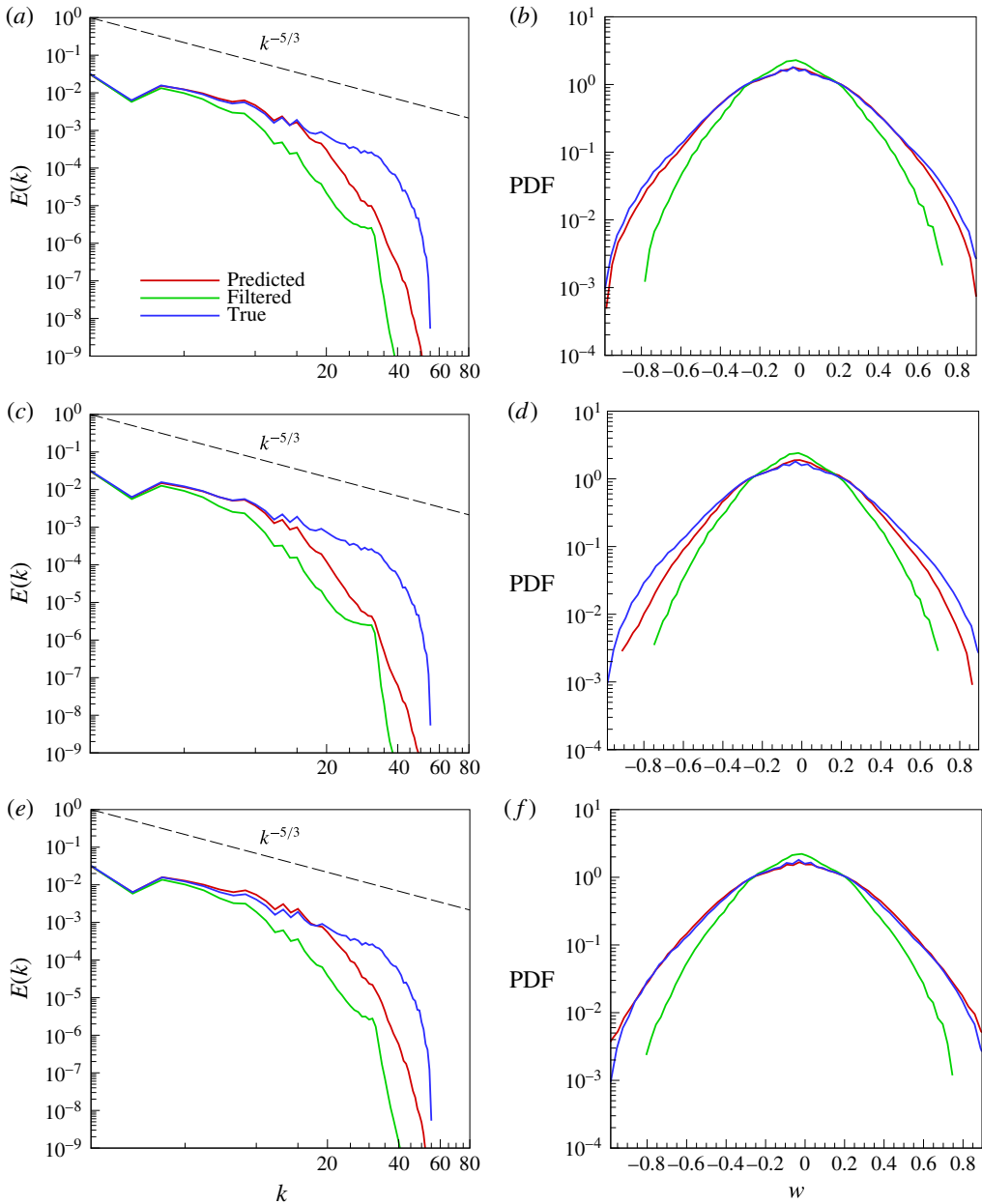


FIGURE 8. (Colour online) *A priori* results of the kinetic energy spectra (*a,c,e*) and PDF of the *z* component of velocity (*b,d,f*) for Kolmogorov turbulence. Results for three different deconvolution test data sets shown. (*a,b*) Test data 1. (*c,d*) Test data 2. (*e,f*) Test data 3.

framework is sufficiently capable in reducing the mean-squared-error of the perturbed fields for all testing datasets.

A comparison of the proposed closure with the state-of-the-art structural closure models is shown in figure 12 by using probability density functions for an assessment

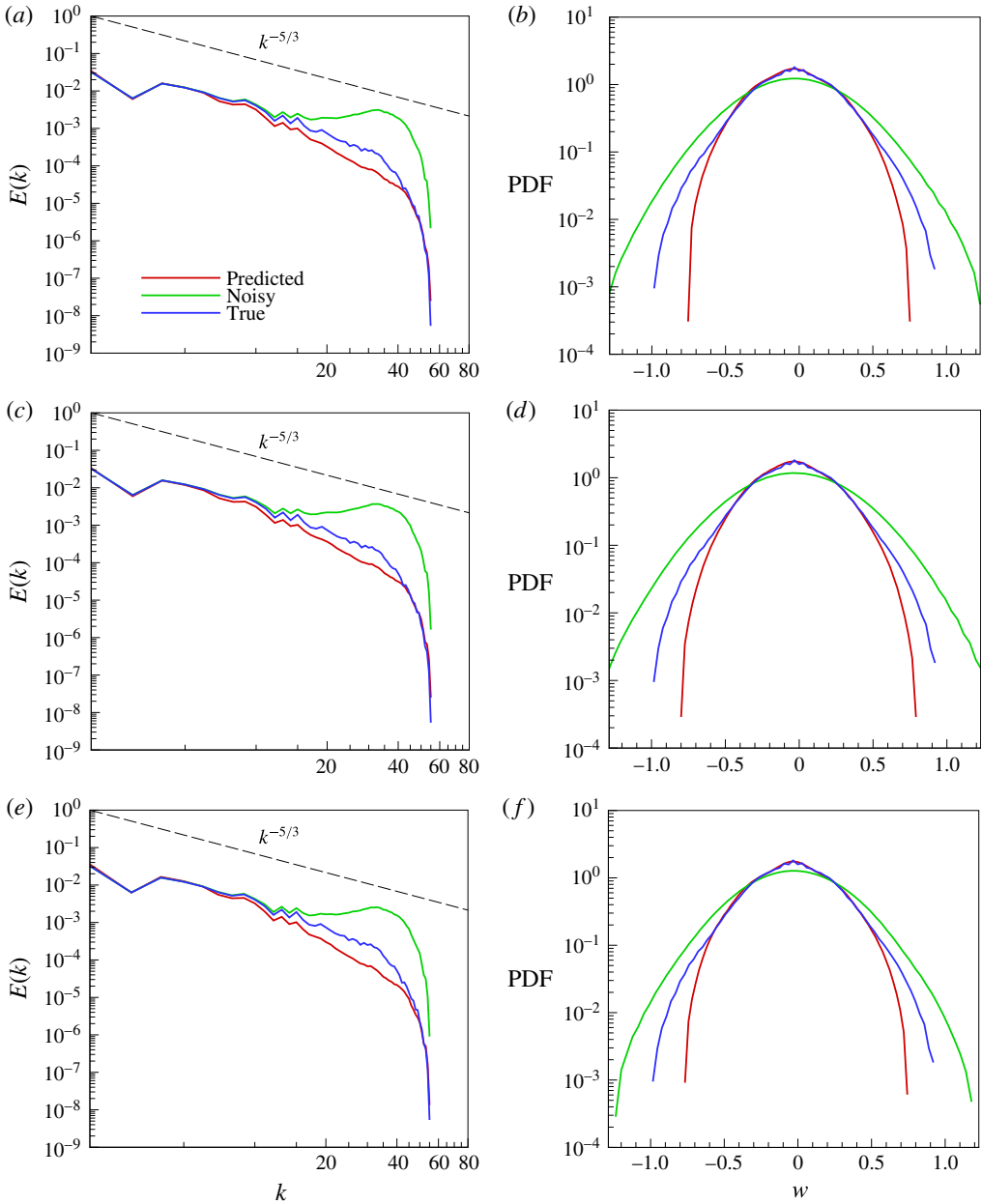


FIGURE 9. (Colour online) *A priori* results of the kinetic energy spectra (*a,c,e*) and PDF of the z component of velocity (*b,d,f*) for Kolmogorov turbulence. Results for three regularization different test data sets shown. (*a,b*) Test data 1. (*c,d*) Test data 2. (*e,f*) Test data 3.

of deviatoric subfilter stress recovery performance. For the purpose of benchmarking we utilize a set of closures including the scale-similarity approach proposed by Bardina, Ferziger & Reynolds (1980) (denoted SS), the approximate deconvolution methodology given by Stolz & Adams (1999) with three iterative deconvolutions

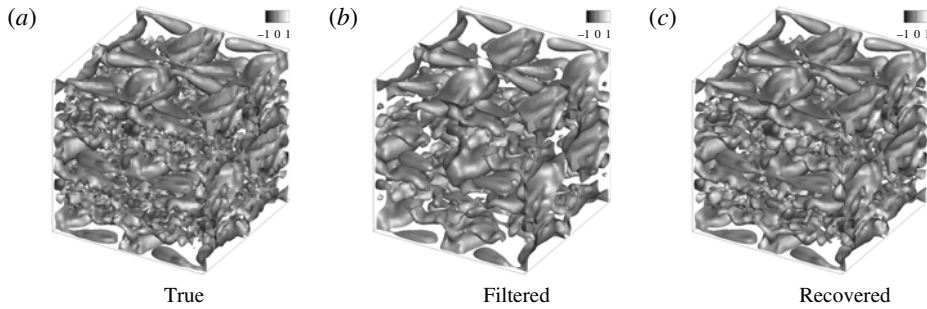


FIGURE 10. *A priori* results for velocity field recovery from low-pass spatially filtered perturbations for Kolmogorov turbulence. Isosurfaces for x -component of the velocity coloured by z -component are shown. Data shown for deconvolution test data 1: (a) true coarse-grained fields (b) coarse-grained fields with Gaussian smoothing, and (c) coarse-grained fields reconstructed using the proposed framework.

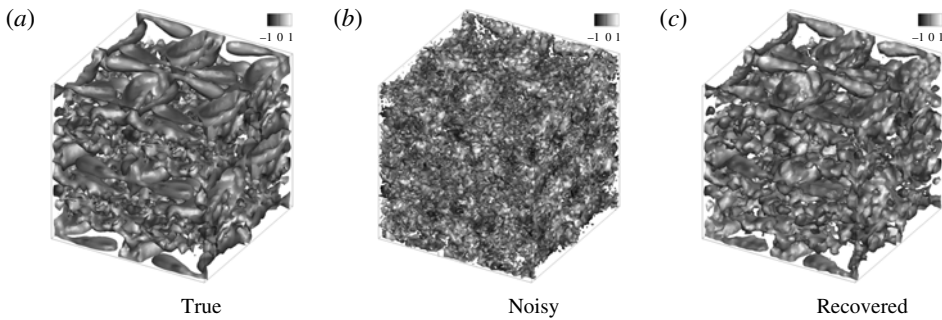


FIGURE 11. *A priori* results for velocity field recovery from noisy perturbation inputs for Kolmogorov turbulence. Isosurfaces for x -component of the velocity coloured by z -component are shown. Data shown for regularization test data 1: (a) true coarse-grained fields, (b) coarse-grained fields with Gaussian noise, and (c) coarse-grained fields reconstructed using the proposed framework.

(which forms the conceptual analogue of our proposed architecture and is denoted as AD^3) as well as the scale-similarity approach proposed by Layton & Lewandowski (2003), which may be interpreted to be an approximate deconvolution methodology with simply one iteration (and hence denoted AD^1). We clarify that while the decision to utilize three iterative deconvolutions for the approximate deconvolution approach (i.e., AD^3) is rather arbitrary, past studies (Maulik & San 2018) have shown that a choice of the number of iterations between 3 and 5 is usually sufficient for satisfactory subfilter recovery.

Our performance assessments indicate that the proposed architecture (denoted ANN) performs in a similar fashion to other widely utilized structural subfilter modelling strategies. The AD^3 approach can be seen to perform better (on average) than our proposed framework due to the fact that the specified filter utilized for the iterative deconvolution is the same as the one used for convolving the field. In comparison, we must emphasize that the proposed data-driven blind deconvolution performance is quite exceptional since no spatial filter shape is assumed. In some cases, we can also observe a slight improvement over the AD^1 and SS *a priori* implementations as well

Kolmogorov turbulence		
Deconvolution		
Dataset	Filtered	Deconvolved
Test data 1	9.56×10^{-3}	3.57×10^{-3}
Test data 2	1.16×10^{-2}	4.64×10^{-3}
Test data 3	8.10×10^{-3}	3.04×10^{-3}
Regularization		
Dataset	Noisy	Regularized
Test data 1	4.20×10^{-2}	8.28×10^{-3}
Test data 2	5.06×10^{-2}	9.18×10^{-3}
Test data 3	3.39×10^{-2}	7.58×10^{-2}

TABLE 3. Mean-squared-error values for deconvolved and regularized fields obtained from the proposed architecture. Data shown from the three-dimensional Kolmogorov turbulence test case. Note that the mean-squared-error values are obtained from the z component of the velocity field.

Model	$\tau_{11} \times 10^{-5}$	$\tau_{12} \times 10^{-5}$	$\tau_{13} \times 10^{-5}$	$\tau_{22} \times 10^{-5}$	$\tau_{32} \times 10^{-5}$	$\tau_{33} \times 10^{-5}$
ANN	8.00	3.60	3.51	7.77	3.64	6.82
SS	6.76	5.62	5.91	6.76	5.91	7.95
AD ¹	31.34	35.82	21.13	31.33	21.12	36.82
AD ³	2.46	1.69	1.82	2.46	1.83	2.91

TABLE 4. Mean-squared-error values for deviatoric subfilter-scale components with respect to the true subfilter-scale stresses for Kolmogorov turbulence.

(both of which require the specification of a low-pass spatial filter). A quantitative characterization of the subfilter stress recovery is shown in table 4, where it may be seen that the proposed method results in mean-squared-errors with respect to true subfilter stresses with values near those of the popular closure models examined in this study.

5.3. Stratified turbulence

Another three-dimensional test case is given by a stratified turbulence data set obtained using 512^3 degrees of freedom. The difference between this test case and the Kolmogorov turbulence test case (TGV) is that this particular simulation is carried out using an implicit large eddy simulation of the inviscid Euler equations (see Maulik & San (2017) for details on numerics). It is well known that this framework provides a good estimation for the Navier–Stokes equations in the limit of infinite Reynolds numbers (Sytime *et al.* 2000; Bos & Bertoglio 2006; Zhou *et al.* 2014). Therefore this test case may be assumed to be representative of a different fluid physics in comparison to the Kolmogorov turbulence problem.

Our stratified turbulence test case is obtained through the simulation of a 3D analogue of the study presented in Maulik & San (2017) to generate a Kelvin–Helmholtz instability (KHI) which eventually transitions through nonlinear interactions to fully developed compressible turbulence. Figure 13 displays density contours of the evolution of the system from its initial condition to a completely turbulent field. The system is evolved from a stratified initial condition with a denser

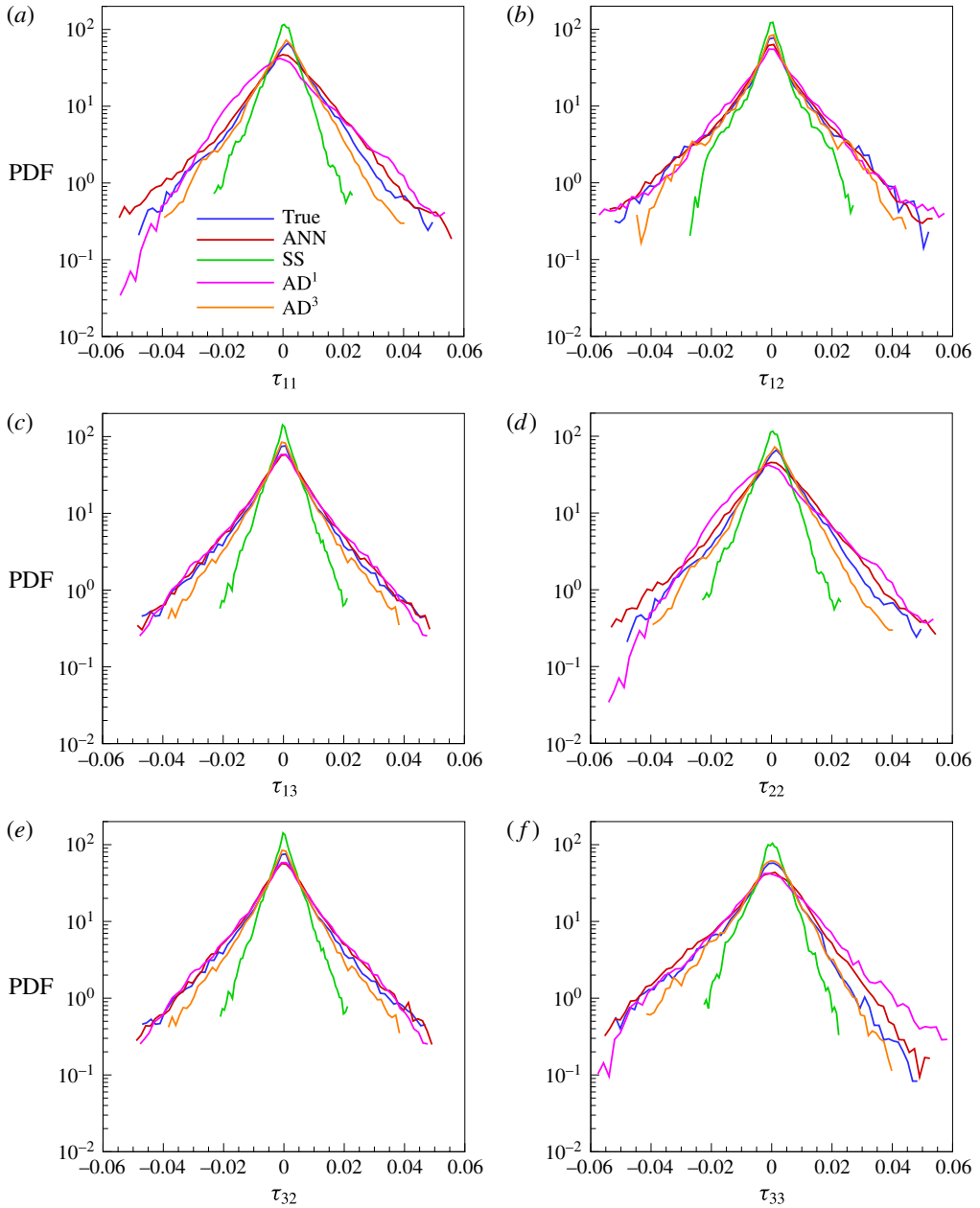


FIGURE 12. (Colour online) *A priori* results for Kolmogorov turbulence subfilter stress predictions by the proposed architecture. Probability density functions for different subfilter stress components along with predictions by state-of-the-art structural closures. Our data-driven architecture performs in a manner similar to these well-established closure strategies without any explicit definition of a low-pass spatial filter.

fluid layer sandwiched between two lighter layers. The middle layer is given an initial velocity (in the negative x direction) and the upper and lower layers are given an equal velocity magnitude in the opposite direction. The shearing velocities in

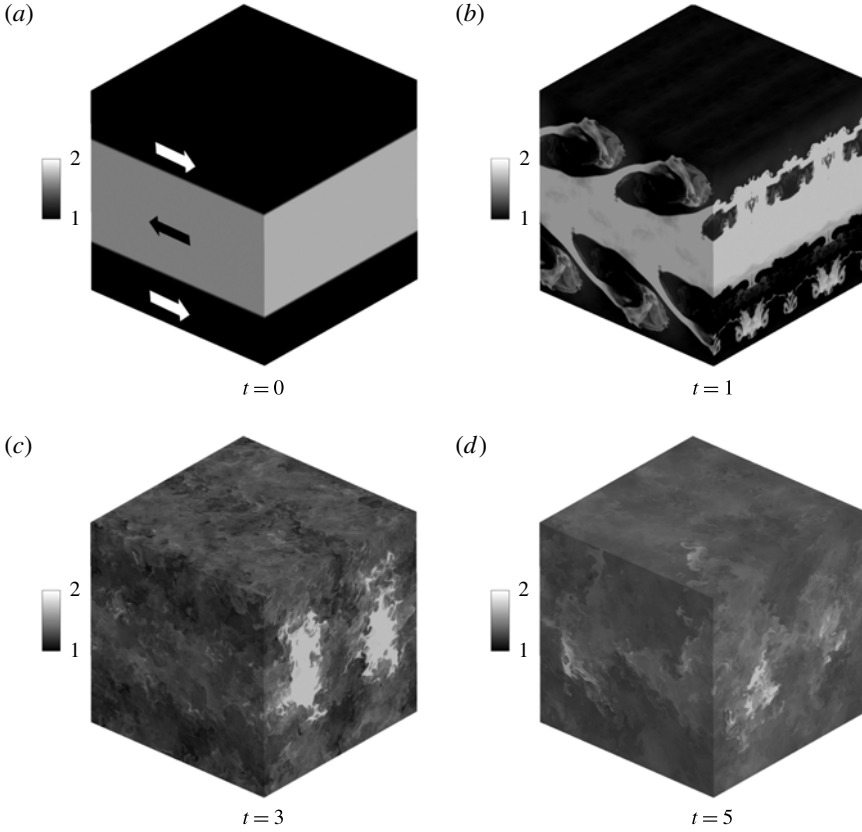


FIGURE 13. Evolution of density field for the stratified turbulence problem.

both layers are specified to ensure the characteristics of a moderately compressible turbulent field (i.e., initial aggregate Mach number values of 0.54 and 0.75 in double shear layers).

We perform assessments similar to our previous test cases with three sets of testing data each for the examination of both deconvolution and noise reduction performance. Figure 14 shows the deconvolution performance of the proposed framework, where we once again witness a notable recovery of the inertial range from the low-pass spatially filtered test data. This is also consistently observed for the differently perturbed test cases. PDF trends for the z component of the velocity are examined as previously and reveal expected trends in high-frequency recovery. We once again observe that the proposed framework performs better for test data 1 and 3 as compared to test data 2, where capture of the tails of the PDF is marginal. We may attribute this to the same reasons detailed in the previous subsection. We remark, however, that even for test data 2 our closure captures the peak of the PDF and the bulk of its distribution about the mean rather well (which leads to excellent inertial range capture).

The regularization performance of the proposed framework is tested for this particular test case with results described in figure 15. Once again, we observe similar trends to those observed before, with marginal improvement in inertial range recovery but effective capture of the peaks of the probability distribution function. The regularization ability of the ELM training procedure, however, leads to the now

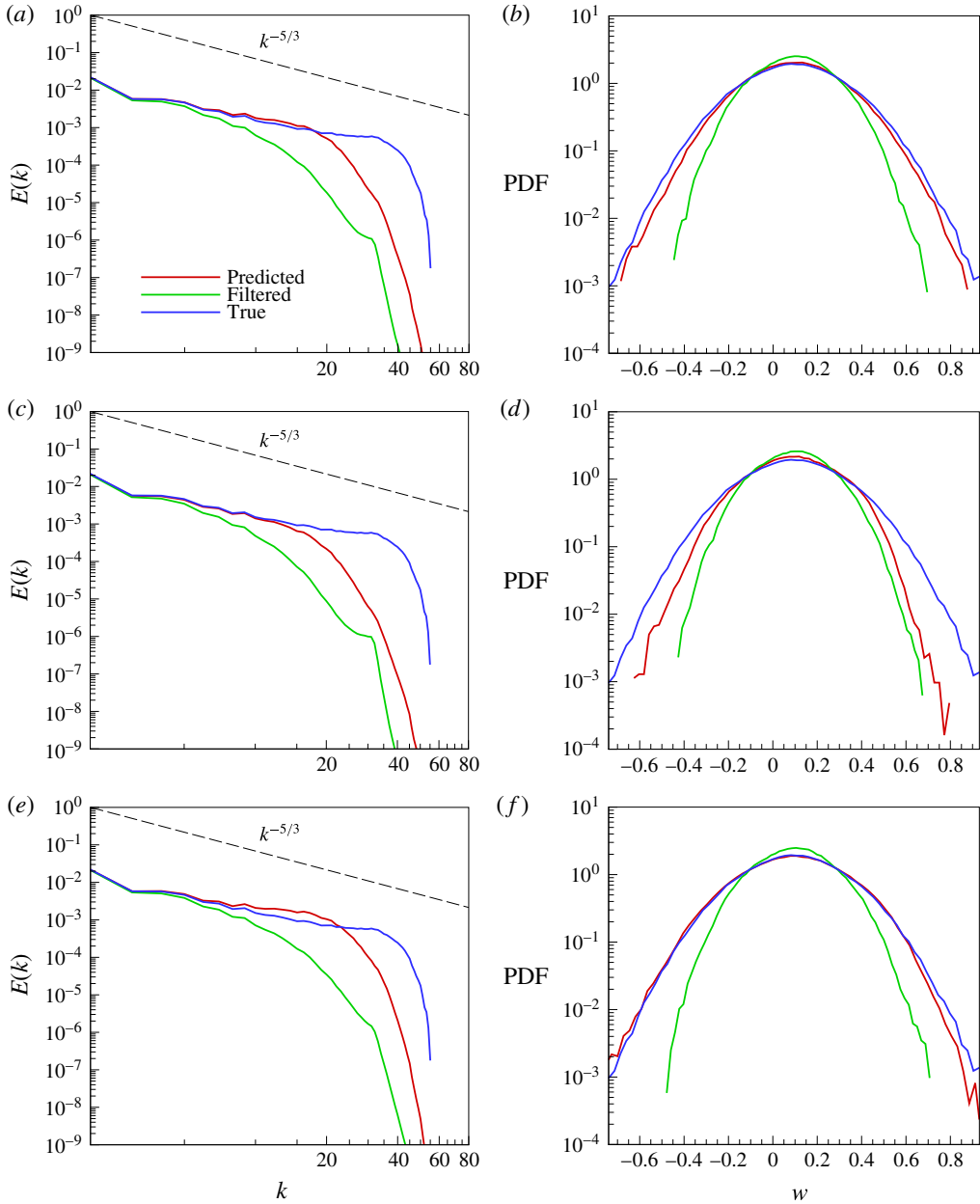


FIGURE 14. (Colour online) *A priori* results of the kinetic energy spectra (*a,c,e*) and PDF of the *z* component of velocity (*b,d,f*) for stratified turbulence. Results for three different deconvolution test data sets shown. (*a,b*) Test data 1. (*c,d*) Test data 2. (*e,f*) Test data 3.

familiar reduction in tail capture accuracy. Aliasing errors are effectively removed as expected. Figures 16 and 17 qualitatively describe the deconvolution and regularization performance of the proposed framework, where it can be seen that good recoveries of the true 3D field are obtained in both cases. A quantitative assessment of the performance of the closure for this framework may be observed in table 5, where

Stratified turbulence		
Deconvolution		
Dataset	Noisy	Regularized
Test data 1	4.03×10^{-2}	1.26×10^{-2}
Test data 2	4.88×10^{-2}	1.38×10^{-2}
Test data 3	3.25×10^{-2}	1.14×10^{-2}
Regularization		
Dataset	Noisy	Regularized
Test data 1	4.19×10^{-2}	1.09×10^{-2}
Test data 2	5.07×10^{-2}	1.18×10^{-2}
Test data 3	3.38×10^{-2}	1.00×10^{-2}

TABLE 5. Mean-squared-error values for deconvolved and regularized fields obtained from the proposed architecture. Data shown from the three-dimensional stratified turbulence test case. Note that the mean-squared-error values are obtained from the z component of the velocity field.

Model	$\tau_{11} \times 10^{-5}$	$\tau_{12} \times 10^{-5}$	$\tau_{13} \times 10^{-5}$	$\tau_{22} \times 10^{-5}$	$\tau_{32} \times 10^{-5}$	$\tau_{33} \times 10^{-5}$
ANN	6.02	2.60	3.12	3.82	1.82	5.15
SS	3.15	2.04	2.27	2.76	2.02	3.62
AD ¹	15.91	10.15	11.91	10.22	7.16	12.7
AD ³	2.01	1.18	1.29	1.60	1.12	2.05

TABLE 6. Mean-squared-error values for deviatoric subfilter-scale components with respect to the true subfilter-scale stresses for stratified turbulence.

trends similar to those observed for the Kraichnan and Kolmogorov test cases are recovered.

We outline a comparison with other structural closures in figure 18, where trends similar to those observed in the Kolmogorov test case are witnessed. The AD³ approach performs in a superior manner as expected, due to the prior specification of the Gaussian filter for both smoothing and iterative deconvolution. The data-driven closure, however, performs quite well in comparison to the chosen closure strategies and validates its application for the purpose of adequate data-driven deconvolution. A quantitative characterization of the subfilter stress recovery is shown in table 6, where it is once again observed that the proposed method results in mean-squared-errors with values near those of the popular structural closure models examined in this study, as observed in the Kolmogorov turbulence test case. We remark, once again, that the observed results for our approach are significant due to no prior specification of a filter kernel.

6. Universality

The true value of the implementation of a data-driven framework emerges from an intelligent combination of our knowledge of physics from first principles with the robust techniques inherent to data-driven modelling. One of the motivations of studying homogeneous isotropic turbulence test cases in this investigation was to link the universal nature of the physics (as evidenced in the $k^{-5/3}$ scaling observed in the averaged kinetic energy spectra) through our proposed framework across flows

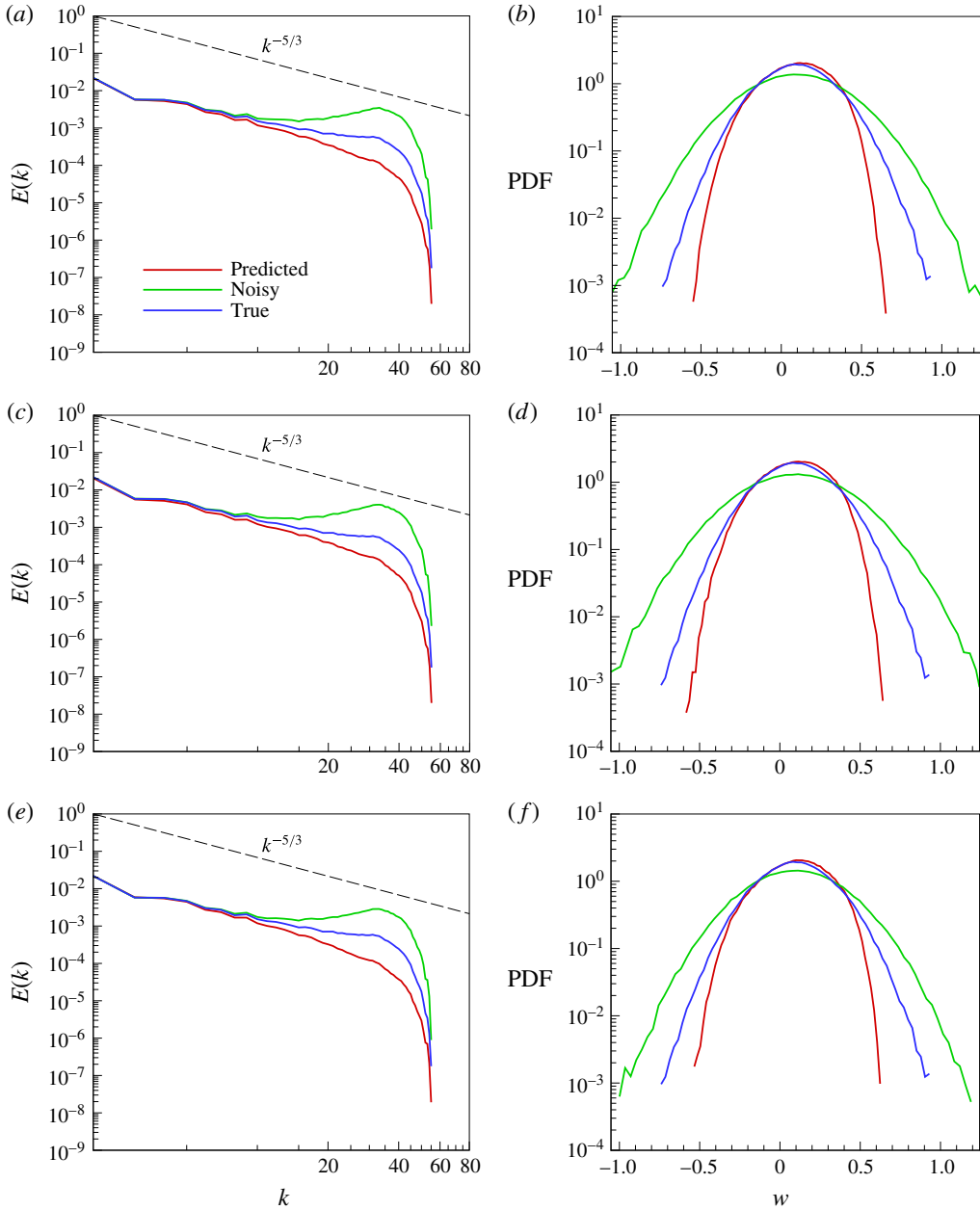


FIGURE 15. (Colour online) *A priori* results of the kinetic energy spectra (*a,c,e*) and PDF of the *z* component of velocity (*b,d,f*) for stratified turbulence. Results for three different regularization test data sets shown. (*a,b*) Test data 1. (*c,d*) Test data 2. (*e,f*) Test data 3.

exhibiting distinctly different physics. For instance, a great utility of any data-driven modelling framework would be to utilize training data from lower Reynolds numbers or shorter (high-fidelity) simulations to devise data-driven closures for runs with higher Reynolds numbers or longer simulations which are computationally intractable.

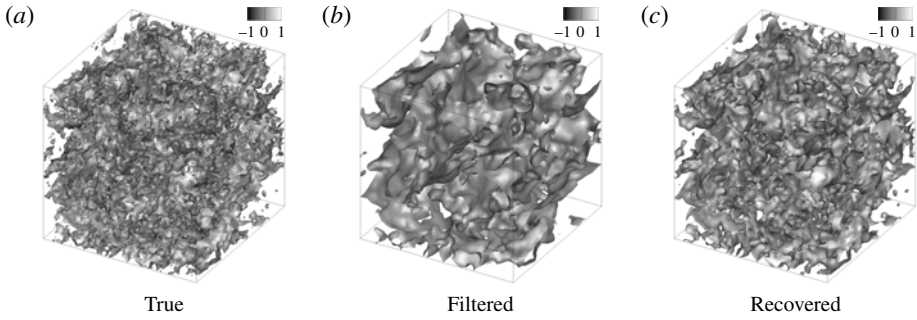


FIGURE 16. *A priori* results for velocity field recovery from low-pass spatially filtered perturbations for stratified turbulence. Data shown for deconvolution test data 1. Isosurfaces for x component of the velocity coloured by z component are shown: (a) true coarse-grained fields, (b) coarse-grained fields with Gaussian smoothing, and (c) coarse-grained fields reconstructed using the proposed framework.

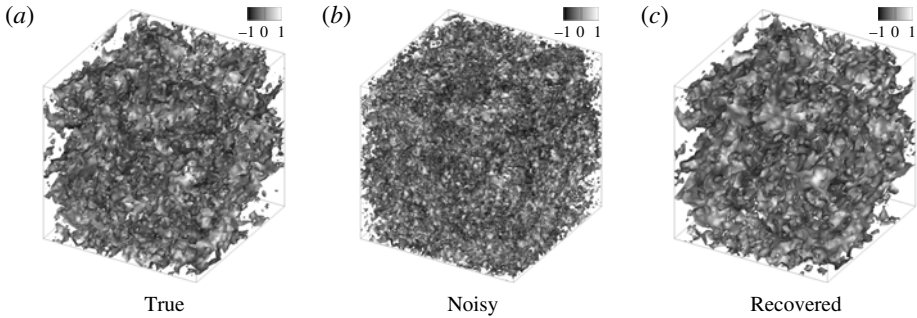


FIGURE 17. *A priori* results for velocity field recovery from noisy perturbation inputs for stratified turbulence. Data shown for regularization test data 1. Isosurfaces for x component of the velocity coloured by z component are shown: (a) true coarse-grained fields, (b) coarse-grained fields with Gaussian noise, and (c) coarse-grained fields reconstructed using the proposed framework.

We first examine the ability of the closure modelling strategy by using training data from a high-fidelity Taylor–Green vortex simulation at $Re = 1600$ to recover true fields at $Re = 5000$ as shown in figure 19. We observe that the framework is successfully able to replicate closure performance as shown in previous test cases. It tells us that the deconvolution (or regularization) ability has been learnt without violating (as it appears) the underlying physics of homogeneous isotropic turbulence. Our next assessment shown in figure 20 utilizes the closure modelling performance of the framework when training data is utilized at time $t = 15$ for a Taylor–Green vortex test case to obtain true field reconstructions at $t = 20$ (for the same simulation). It can once again be observed that the closure recovery is exceptional, with trends similar to those exhibited by previous test cases.

Next, we present cross-validation results for different flows. Here, the proposed blind deconvolution ANN closure is trained from a completely different simulation data (i.e., testing with different flow configurations). Figure 21 demonstrates the ability of the data-driven closure to predict subfilter terms for the stratified KHI turbulence test case where TGV simulation data at $Re = 1600$ is utilized to train

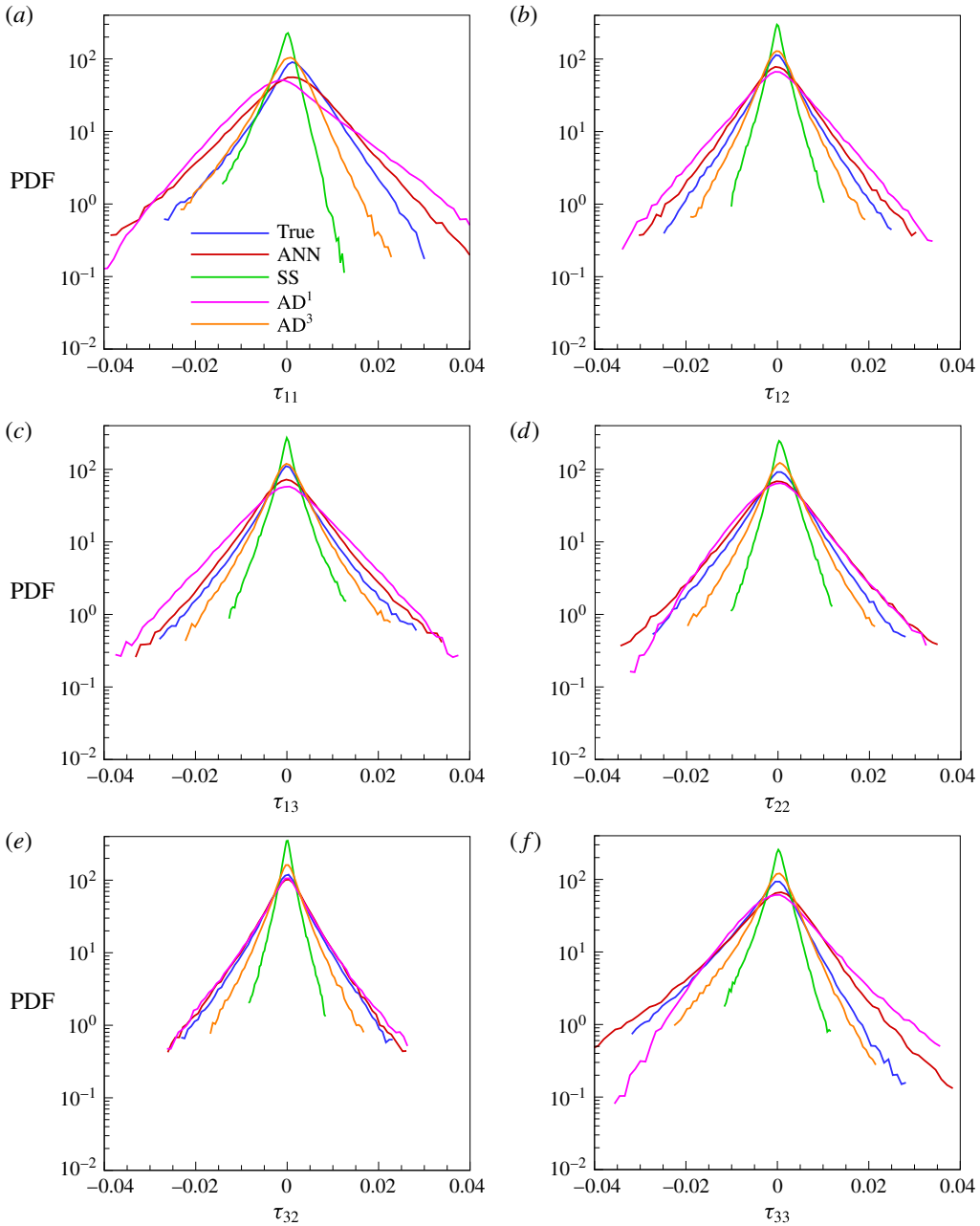


FIGURE 18. (Colour online) *A priori* results for stratified turbulence subfilter stress predictions by the proposed architecture. Probability density functions for different subfilter stress components along with predictions by state-of-the-art structural closures. Our data-driven architecture performs in a manner similar to these well-established closure strategies without any explicit definition of a low-pass spatial filter.

our ANN. In this case we highlight that the only link between training and testing data is the $k^{-5/3}$ cascade and the magnitude of perturbations to the coarse-grained field. It can be seen that a remarkable inertial range recovery and regularization

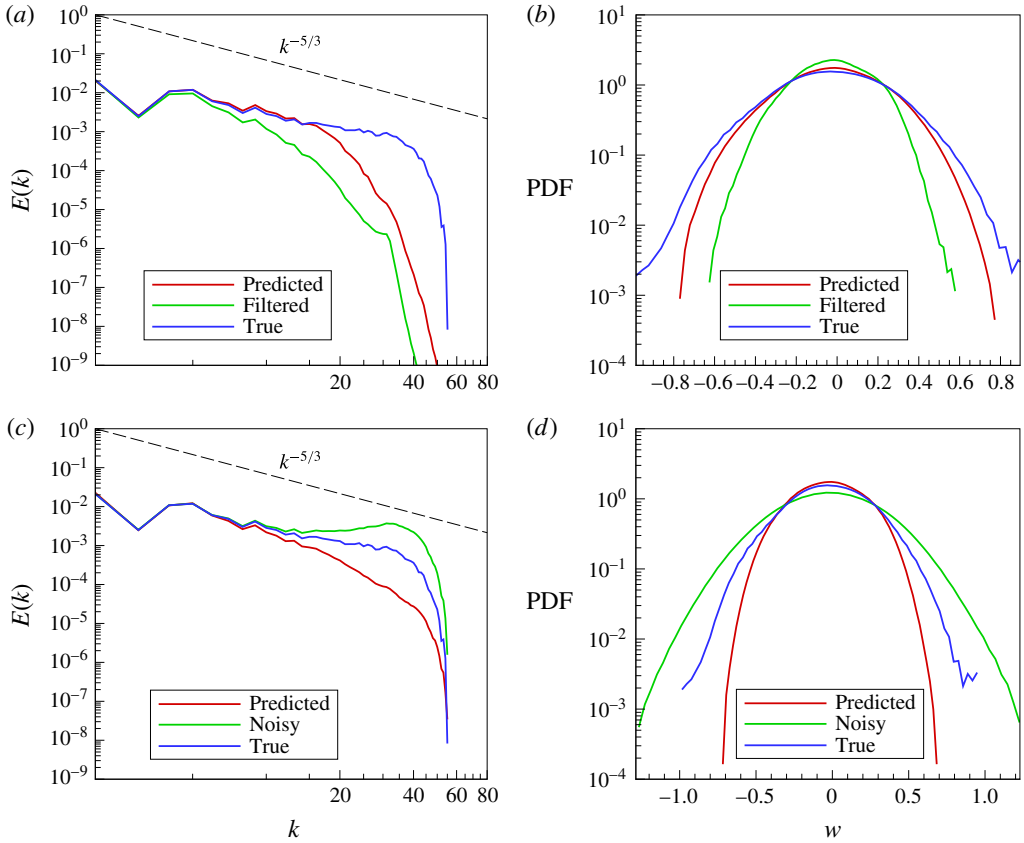


FIGURE 19. (Colour online) *A priori* results of the kinetic energy spectra (a,c) and PDF of the vorticity (b,d) for Kolmogorov turbulence. (a,b) Filtered inputs. (c,d) Noisy inputs. Here we utilize DNS data for the Taylor–Green vortex at $Re = 1600$ to reconstruct an approximation to the true field for $Re = 5000$.

performance is still observed for both the low-pass spatially filtered and noisy testing data. Analogously, in figure 22, we invert our training and testing data with stratified turbulence data being used to train the ELM, which is further used for deconvolution and regularization of an appropriately perturbed solution field given by a Taylor–Green vortex simulation at $Re = 1600$. Once again, previously observed trends in the output of the closure model are recovered. From the evidence provided above, we may tentatively conclude that for flows exhibiting similar cascade characteristics, subfilter recovery and high-wavenumber regularization are reliant on the filter radius and magnitude of high-frequency perturbation alone. This allows for exciting possibilities for 3D turbulence closures, for instance, to implicitly learn filter shapes from physical experiments and reproduce accurate trends through coarse-grained numerical simulations within a wide range of Reynolds numbers.

7. Concluding remarks

An artificial neural network architecture is proposed for the data-driven deconvolution and regularization of low-pass spatially filtered turbulence fields. Both 2D and 3D

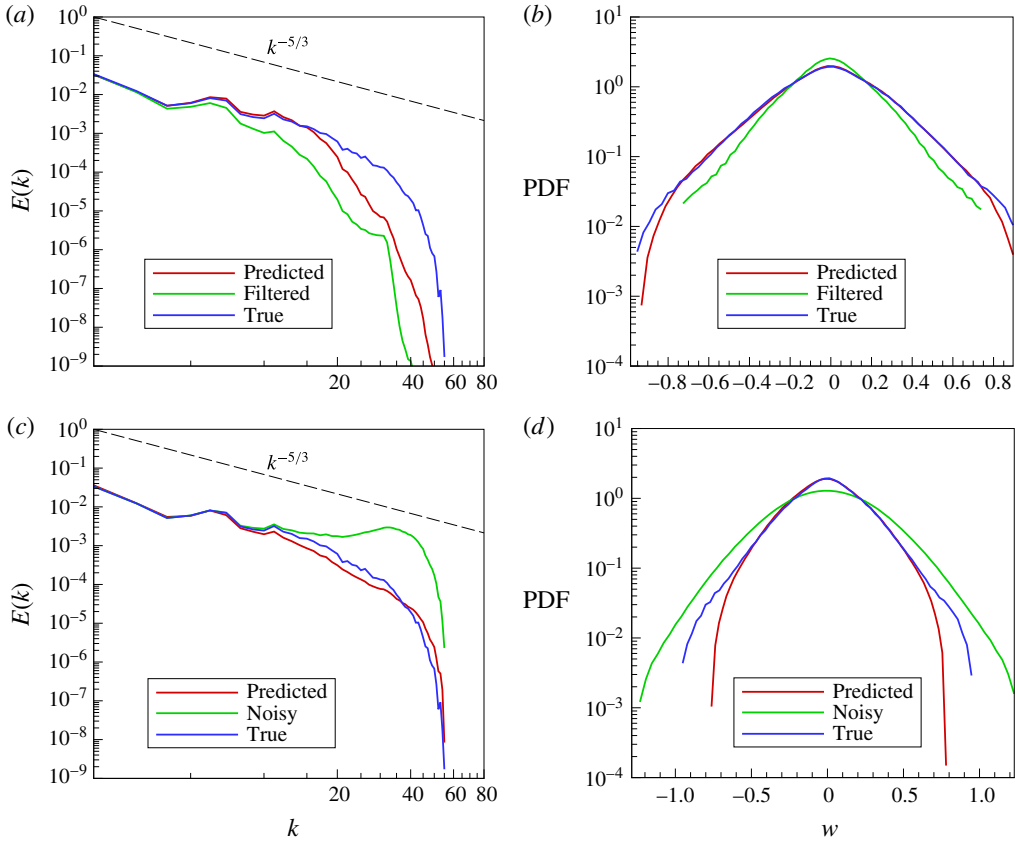


FIGURE 20. (Colour online) *A priori* results for the kinetic energy spectra (a,c) and PDF of the vorticity (b,d) for Kolmogorov turbulence. (a,b) Filtered inputs. (c,d) Noisy inputs. Here we utilize high-fidelity data from the Taylor–Green vortex at time $t = 15$ to obtain a reconstruction for the same test case at time $t = 20$ for a $Re = 1600$.

test cases are examined with training data sets obtained through the coarsened and perturbed versions of high-fidelity simulations for canonical homogeneous isotropic turbulence and stratified compressible turbulence problems. Two types of perturbations are tested, one in which the true data is filtered using a Gaussian kernel and the other which has a certain quantum of noise added to the field to represent high-wavenumber aliasing errors. Testing data sets are generated through a spatial shifting procedure as well as through the utilization of slightly different magnitudes of filter radius and noise. This ensures that deconvolution and regularization performance of the proposed architecture remains localized to physics and not numerical artefact. The proposed architecture is tested independently on Kraichnan, Kolmogorov and stratified compressible turbulence test cases and is able to provide an estimate of the deconvolved variable as examined through the energy spectra and probability density functions of the recovered fields. For noisy data, the innate regularization of the ELM training approach results in smoother predictions for the recovered variable. For our training mechanism, the ELM approach is chosen due to its exceptional speed of training in comparison with traditional gradient-based methods as well as its excellent generalization ability (which avoids the problem of overfitting). The

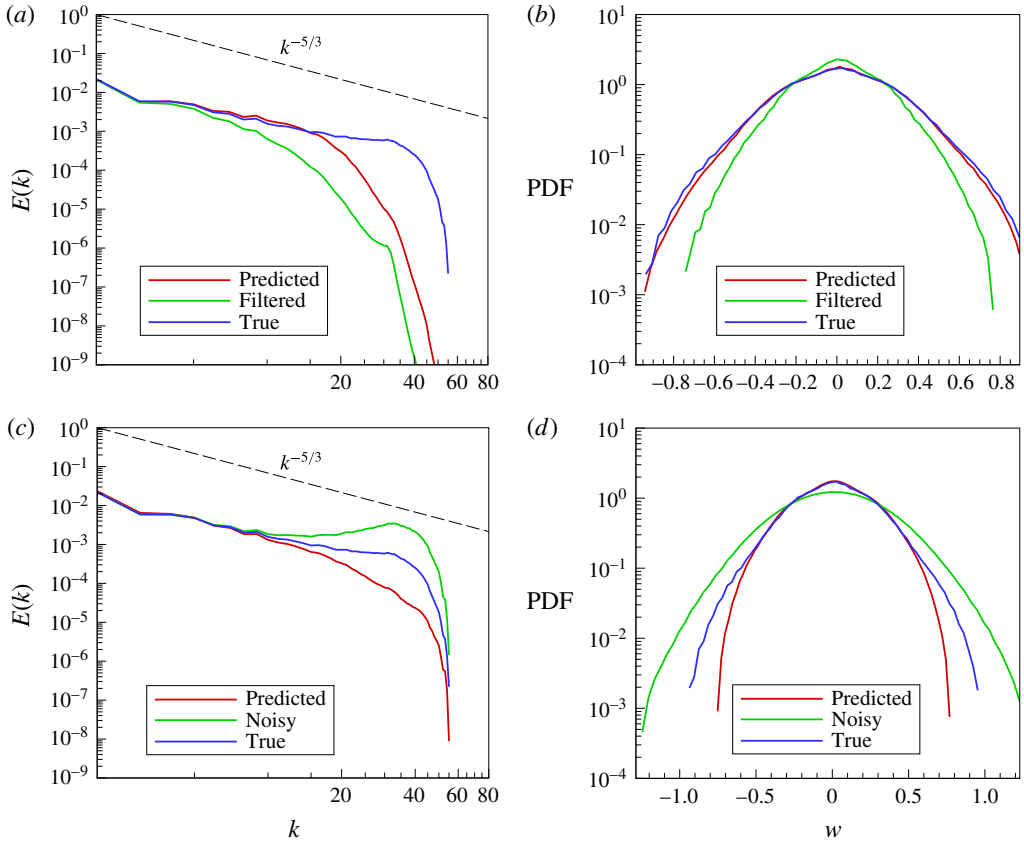


FIGURE 21. (Colour online) *A priori* results of the kinetic energy spectra (a,c) and PDF of the vorticity (b,d) stratified turbulence. (a,b) Filtered inputs. (c,d) Noisy inputs. Here we utilize DNS data for the Taylor–Green vortex at $Re = 1600$ to reconstruct an approximation to the true field for the stratified turbulence test case generated from the inviscid Euler equations.

ability of the proposed approach for subfilter-scale content recovery is benchmarked against several popular structural closure modelling strategies. It is observed that our the data-driven framework yields a similar deconvolution performance without the explicit specification of a filter kernel. In addition, we also test the universality of our data-driven closure by utilizing testing and training data sets from different simulations displaying the same classical Kolmogorov cascade. This leads to an exciting observation, as described in § 6, which tells us that the proposed framework may be utilized to leverage our understanding of the cascade of energy in three-dimensional turbulence to obtain reconstructions for flows across different physics. Indeed, the blind deconvolution procedure appears to be linked across a wide range of physics solely through the underlying filter radius that arises in LES flow computations due to coarse-graining and implicit (or explicit) numerical dissipation.

A natural follow-up to this investigation is to test our proposed approach in a fully a posteriori analysis. One of our primary goals in subsequent investigations is also to address the issue of sampling for training data. Sampling strategies must be devised to ensure that the proposed framework is exposed to data from many

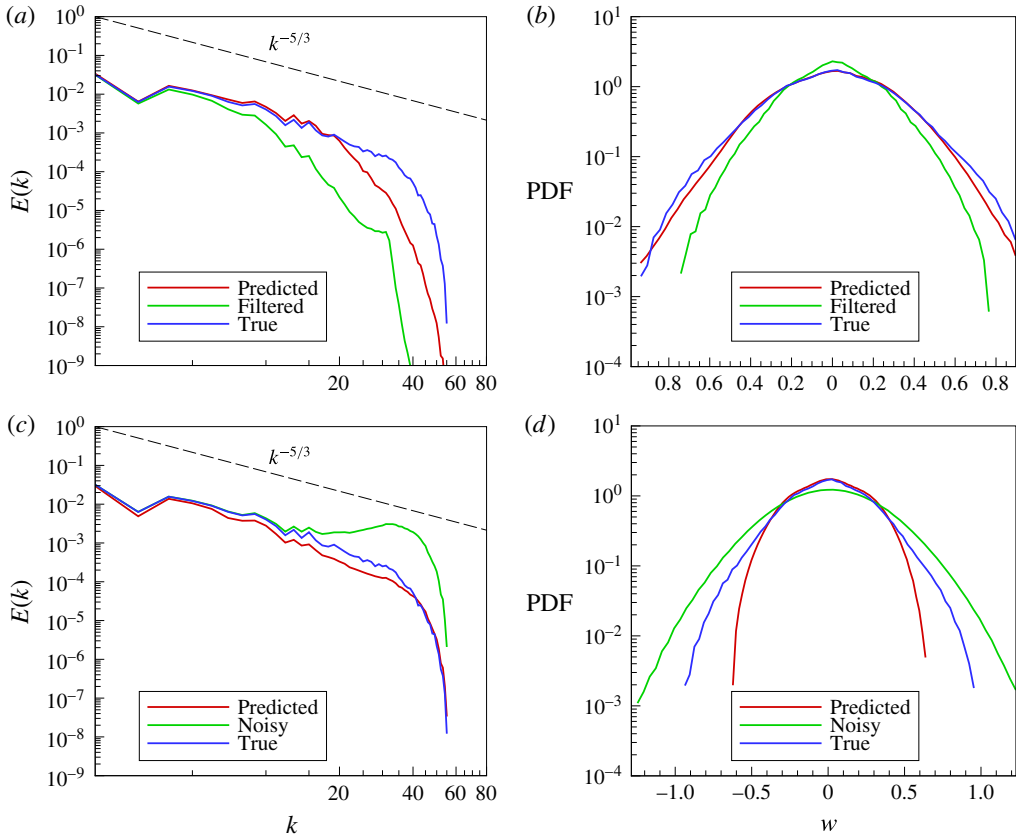


FIGURE 22. (Colour online) *A priori* results of the kinetic energy spectra (a,c) and PDF of the vorticity (b,d) for Kolmogorov turbulence. (a,b) Filtered inputs. (c,d) Noisy inputs. Here we utilize high-fidelity data from the stratified turbulence problem to train a prediction for the Taylor–Green vortex at $Re = 1600$.

physical regimes (as opposed to indiscriminate selections in parameter regimes). This is an important distinction that has significant implications on the performance of any proposed data-driven modelling framework. Basically, a data-driven model is only as good as the data it has been trained on and can only reproduce physical behaviours similar to those it has seen in training. In addition to sampling strategies, it is also important to develop outlier identification systems for noisy data. Data preprocessing, an active area of research in the data science community, must therefore be integrated into our framework. From this point of view, the fast training times of the proposed architecture suggest the use of multiple networks trained to act in a ‘committee’ for aggregate subfilter predictions of flow datasets with variable spatial and temporal characteristics. This ensures robustness towards outliers, since aggregate predictions are likely to be more accurate than those from solely one trained network.

Since, apart from its training data, an ANN’s performance is heavily dependant on its architecture, it is necessary to estimate the performance of the proposed framework for several different configurations in terms of network architectures (number of inputs, number of outputs, number of neurons, activation functions etc). Fortunately, the speed of training aids us in generating regularized network weights

for extremely large datasets with considerable ease. In conclusion, this work features a preliminary glimpse at an exciting avenue for the next generation in data-driven turbulence closures.

REFERENCES

- ALBERT, A. 1972 *Regression and the Moore–Penrose Pseudoinverse*. Academic.
- BARDINA, J., FERZIGER, J. H. & REYNOLDS, W. C. 1980 Improved subgrid-scale models for large-eddy simulation. *AIAA Paper* 80-1357.
- BOS, W. J. & BERTOGLIO, J. 2006 Dynamics of spectrally truncated inviscid turbulence. *Phys. Fluids* **18** (7), 071701.
- BRIGHT, I., LIN, G. & KUTZ, J. N. 2013 Compressive sensing based machine learning strategy for characterizing the flow around a cylinder with limited pressure measurements. *Phys. Fluids* **25** (12), 127102.
- BRUNTON, S. L., PROCTOR, J. L. & KUTZ, J. N. 2016 Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl Acad. Sci. USA* **113** (15), 3932–3937.
- BULL, J. R. & JAMESON, A. 2015 Simulation of the Taylor–Green vortex using high-order flux reconstruction schemes. *AIAA J.* **53**, 2750–2761.
- CICHOCKI, A. & AMARI, S. 2002 *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. John Wiley & Sons.
- DABOV, K., FOI, A., KATKOVNIK, V. & EGIAZARIAN, K. 2007 Color image denoising via sparse 3D collaborative filtering with grouping constraint in luminance-chrominance space. In *IEEE International Conference on Image Processing*, pp. 313–316.
- DEMUTH, H. B., BEALE, M. H., DE JESS, O. & HAGAN, M. T. 2014 *Neural Network Design*. Martin Hagan.
- DURAISAMY, K., ZHANG, Z. J. & SINGH, A. P. 2015 New approaches in turbulence and transition modeling using data-driven techniques. *AIAA Paper* 2015-1284.
- FORESEE, F. D. & HAGAN, M. T. 1997 Gauss–Newton approximation to Bayesian learning. In *IEEE International Conference on Neural Networks*, pp. 1930–1935.
- FRISCH, U. 1996 *Turbulence: the Legacy of A. N. Kolmogorov*. Cambridge University Press.
- GAMAHARA, M. & HATTORI, Y. 2017 Searching for turbulence models by artificial neural network. *Phys. Rev. Fluids* **2** (5), 054604.
- GAUTIER, N., AIDER, J. L., DURIEZ, T., NOACK, B. R., SEGOND, M. & ABEL, M. 2015 Closed-loop separation control using machine learning. *J. Fluid Mech.* **770**, 442–457.
- GERMANO, M. 2015 The similarity subgrid stresses associated to the approximate Van Cittert deconvolutions. *Phys. Fluids* **27** (3), 035111.
- HORNIK, K., STINCHCOMBE, M. & WHITE, H. 1989 Multilayer feedforward networks are universal approximators. *Neural Netw.* **2** (5), 359–366.
- HUANG, G., ZHU, Q. & SIEW, C. 2004 Extreme learning machine: a new learning scheme of feedforward neural networks. In *IEEE International Joint Conference on Neural Networks*, pp. 985–990.
- HUANG, G., ZHU, Q. & SIEW, C. 2006 Extreme learning machine: theory and applications. *Neurocomput.* **70** (1), 489–501.
- KOLMOGOROV, A. N. 1941 The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Dokl. Akad. Nauk SSSR* **30**, 301–305.
- 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.
- LAYTON, W. & LEWANDOWSKI, R. 2003 A simple and stable scale-similarity model for large eddy simulation: energy balance and existence of weak solutions. *Appl. Maths Lett.* **16** (8), 1205–1209.

- LAYTON, W. J. & REBHOLZ, L. G. 2012 *Approximate Deconvolution Models of Turbulence: Analysis, Phenomenology and Numerical Analysis*. Springer.
- LEE, C., KIM, J., BABCOCK, D. & GOODMAN, R. 1997 Application of neural networks to turbulence control for drag reduction. *Phys. Fluids* **9** (6), 1740–1747.
- LING, J., JONES, R. & TEMPLETON, J. 2016a Machine learning strategies for systems with invariance properties. *J. Comput. Phys.* **318**, 22–35.
- LING, J., KURZAWSKI, A. & TEMPLETON, J. 2016b 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.
- MACKAY, D. J. C. 1992 Bayesian interpolation. *Neural Comput.* **4** (3), 415–447.
- MAULIK, R. & SAN, O. 2017 Resolution and energy dissipation characteristics of implicit LES and explicit filtering models for compressible turbulence. *Fluids* **2** (2), 14.
- MAULIK, R. & SAN, O. 2018 Explicit and implicit LES closures for Burgers turbulence. *J. Comput. Appl. Maths* **327**, 12–40.
- MILANO, M. & KOUMOUTSAKOS, P. 2002 Neural network modeling for near wall turbulent flow. *J. Comput. Phys.* **182** (1), 1–26.
- 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.
- RAISSI, M. & KARNIADAKIS, G. E. 2016 Deep multi-fidelity Gaussian processes. [arXiv:1604.07484](https://arxiv.org/abs/1604.07484).
- RAISSI, M. & KARNIADAKIS, G. E. 2017 Hidden physics models: Machine learning of nonlinear partial differential equations. [arXiv:1708.00588](https://arxiv.org/abs/1708.00588).
- SAN, O. & STAPLES, A. E. 2012 High-order methods for decaying two-dimensional homogeneous isotropic turbulence. *Comput. Fluids* **63**, 105–127.
- SCHMIDT, M. & LIPSON, H. 2009 Distilling free-form natural laws from experimental data. *Science* **324** (5923), 81–85.
- SERRE, D. 2002 *Matrices: Theory and Applications*. Springer.
- STOLZ, S. & ADAMS, N. A. 1999 An approximate deconvolution procedure for large-eddy simulation. *Phys. Fluids* **11** (7), 1699–1701.
- STOLZ, S., ADAMS, N. A. & KLEISER, L. 2001 The approximate deconvolution model for large-eddy simulations of compressible flows and its application to shock-turbulent-boundary-layer interaction. *Phys. Fluids* **13** (10), 2985–3001.
- SYTINE, I. V., PORTER, D. H., WOODWARD, P. R., HODSON, S. W. & WINKLER, K. 2000 Convergence tests for the piecewise parabolic method and Navier–Stokes solutions for homogeneous compressible turbulence. *J. Comput. Phys.* **158** (2), 225–238.
- TRACEY, B., DURAISAMY, K. & ALONSO, J. 2013 Application of supervised learning to quantify uncertainties in turbulence and combustion modeling. *AIAA Paper* 2013-0259.
- WANG, J., WU, J. & XIAO, H. 2016 Physics-informed machine learning for predictive turbulence modeling: Using data to improve RANS modeled Reynolds stresses. [arXiv:1606.07987](https://arxiv.org/abs/1606.07987).
- WANG, J., WU, J. & XIAO, H. 2017 Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data. *Phys. Rev. Fluids* **2** (3), 034603.
- WANG, L., HUANG, Y., LUO, X., WANG, Z. & LUO, S. 2011 Image deblurring with filters learned by extreme learning machine. *Neurocomput.* **74** (16), 2464–2474.
- WEATHERITT, J. & SANDBERG, R. 2016 A novel evolutionary algorithm applied to algebraic modifications of the RANS stress–strain relationship. *J. Comput. Phys.* **325**, 22–37.
- ZHANG, G., PATUWO, B. E. & HU, M. Y. 1998 Forecasting with artificial neural networks: the state of the art. *Int'l J. Forecast.* **14** (1), 35–62.
- ZHANG, Z. J. & DURAISAMY, K. 2015 Machine learning methods for data-driven turbulence modeling. *AIAA Paper* 2015-2460.
- ZHOU, Y., GRINSTEIN, F. F., WACHTOR, A. J. & HAINES, B. M. 2014 Estimating the effective Reynolds number in implicit large-eddy simulation. *Phys. Rev. E* **89** (1), 013303.